

Partial neural correlations suggest detailed interactions in cortical 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: yatsenko@cns.bcm.edu

Abstract

Ambitious projects aim to record the activity of ever larger and denser subsets of neurons *in vivo*. Correlations in neural activity measured in such recordings can reveal important aspects of the functional organization of neural circuits. However, estimating and interpreting large correlation matrices obtained from finite recordings is challenging. Estimation can be improved by regularization, *i.e.* by imposing a structure on the estimate. The amount of improvement depends on how closely the assumed structure represents dependencies in the data. Therefore, the selection of the most efficient correlation matrix estimator for a given neural circuit must be determined empirically. Furthermore, the identity and structure of the most efficient estimator informs about the types of dominant dependencies governing the system.

We sought the most statistically efficient estimators of neural correlation matrices in recordings from large, dense groups of cortical neurons. Using fast 3D random-access laser scanning microscopy of calcium signals, we recorded the activity of nearly every neuron in volumes of about 200 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 most efficiently represented as the combination of a sparse matrix of pairwise partial correlations representing local interactions and a low-rank component representing common fluctuations and external inputs. Indeed, in cross-validation tests, the covariance matrix estimator with such structure consistently outperformed other regularized estimators. The connectivity patterns inferred from the sparse component of the estimate manifested a strong relationship to the geometrical arrangement and orientation tuning properties of cells: the density of positive ‘excitatory’ connections decreased rapidly with geometric distance and the difference in orientation preference whereas the negative ‘inhibitory’ connections were less selective. Because of its superior performance, this estimator likely provides a more physiologically relevant representation of the functional connectivity in dense recordings than sample correlations.

Author Summary

Correlations in the activity of neural populations have proven useful as descriptions of the functional connectivity in the embedding circuit. Correlations may arise from various underlying processes that reflect differently on the structure of the correlation matrix: Common network fluctuations or external inputs are well approximated by low-rank representations whereas linear effects between specific pairs of neurons are well approximated by fitting their pairwise partial correlations. What kinds of correlation effects dominate the population activity in cortical microcircuits? To answer this question, we imposed various correlation structures on empirical estimates of neural correlation matrices and evaluated which

of them produced greatest estimation improvement. In fast 3D two-photon imaging of calcium signals of large and dense groups of neurons in mouse visual cortex, the greatest improvement was produced by decomposing the correlation matrix as a sparse network of partial correlations and a low-rank component. Since it leads to the best estimates, we propose that this structure provides a better picture of the functional connectivity than that obtained from sample correlations. As an application of this approach, we analyzed how the inferred connectivity related to distances between cells and to differences in their preferred orientations and found basic resemblance to results from studies of synaptic connectivity.

Introduction

Neural correlations Pearson correlations between the spiking activity of pairs of neurons, or simply *neural correlations*, are among the most familiar descriptive statistics of neural population activity [1–5]. For example, *noise correlations*, *i.e.* the correlations of trial-to-trial response variability between pairs of neurons, have a profound impact on stimulus coding [1, 2, 6–11]. In addition, noise correlations and correlations in spontaneous activity have been hypothesized to reflect key features of the functional connectivity in neural circuits [12]. This interpretation is supported by a series of discoveries of nontrivial relationships between neural correlations and other aspects of circuit organization such as the physical distances between neurons [13, 14], their synaptic connectivity [15], stimulus response similarity [3–5, 15–21], cell types [22], cortical layer specificity [23, 24], progressive changes in development and in learning [25, 26], changes due to sensory stimulation and global brain states [5, 27–29] among others.

However, neural correlations do not submit to ready or unambiguous mechanistic interpretation. Theoretical studies and simulations have shown that neural correlations on various temporal scales may arise from combinations of multiple mechanisms including direct synaptic interactions, common inputs or correlated inputs, chains of multiple synaptic connections, oscillations, top-down modulation, and background network fluctuations [30–35].

Structure of correlation matrices Modern multineuronal recordings allow measuring entire correlation matrices from increasingly large and dense populations of highly interconnected neurons. Multivariate analysis techniques based on large correlation matrices provide information that could not be extracted from an equivalent number of isolated measurements of pairwise correlations. For example, when the correlation matrix is approximated in a form with a reduced number of variables, this *reduced representation* can isolate an important aspect of the functional connectivity in neural circuits and possibly point to underlying physiological mechanisms and computational principles.

One reduced representation of the correlation matrix is its low-rank approximation. Low-rank approximations are particularly suitable for capturing common fluctuations and common external inputs and can be extracted using principal component analysis, independent component analysis, factor analysis, or [36–39].

Another reduced representation of the correlation matrix can be obtained by *covariance selection*, *i.e.* by selecting a subset of the pairwise *partial correlations* that are sufficient to reproduce the entire correlation matrix [40, 41]. The partial correlation between a pair of neurons is a measure of their linear association after the effects of the rest of the network have been taken into account [42]. Partial correlations are particularly meaningful when all effects between the variables are nearly linear and a large fraction of the variables are observed; in this case, partial correlations approximate conditional dependency. A lack of conditional dependency between two neurons suggests a lack of direct interaction between them. Networks of partial correlations (sometimes called *association networks*) have been used to infer gene interaction networks [43, 44] and interactions between brain regions in brain imaging studies [45, 46].

Improving estimation through regularization Imposing a structure on an estimate of the covariance matrix can improve its estimation performance.

Estimation of covariance matrices from large populations presents a number of numerical challenges: The amount of recorded data grows only linearly with population size whereas the number of estimated coefficients increases quadratically. This mismatch leads to an increase in spurious correlations, overestimation of common activity (*i.e.* overestimation of large eigenvalues) [47], and poorly conditioned partial correlations [43].

In neuroscience studies, the true correlation matrix is usually estimated by the *sample correlation matrix*, which is rarely distinguished from the true correlation matrix. The sample correlation matrix is unbiased but, because it has many free parameters, it is highly sensitive to sampling noise. As a result, on average, the sample correlation matrix is far from the true correlation matrix. In contrast, reduced estimates are typically less susceptible to sampling noise although can introduce their respective biases.

In statistics, the technique of deliberately imposing a structure on an estimate in order to improve its performance is known as *regularization* [43, 48]. To ‘impose structure’ on an estimate means to bias (‘shrink’) it toward a reduced representation known as the shrinkage target. The selection of the optimal target estimate and the optimal amount of shrinkage can be deduced from the training data either analytically or by cross-validation. Some regularization schemes only perform target selection (*dimensionality reduction, variable selection, or feature selection*) and replace the estimate completely. Others only perform shrinkage toward a single target estimate (*shrinkage estimators*). Yet others can do both target selection and shrinkage toward the target.

Although regularized covariance matrix estimates have become commonplace in other fields such as finance [49], functional genomics [43], and brain imaging [46], surprisingly little work has been done to identify optimal regularization of neural correlation matrices.

Illustration of estimation of a neural correlation matrix To illustrate the challenges and techniques of estimation of the correlation matrix from a finite sample, we consider a regularization scheme based on *covariance selection* (Figure 1) [40, 41]. Here the estimate is produced by fitting only an optimal subset of the coefficients of the precision matrix (inverse of the covariance matrix) while setting the rest to zero. Zeros in the off-diagonal elements of the precision matrix indicate zero partial correlations between the corresponding pair of neurons. Therefore, covariance selection amounts to finding an approximation of the covariance matrix that maximizes the number of neuron pairs that are linearly independent of each other when conditioned on the activity of all the other observed cells.

We estimated the covariance matrix of somatic calcium signals of a group of 298 neurons in mouse visual cortex using both the sample covariance matrix (Fig. 1 A, B) and covariance selection (Fig. 1 C). Due to the low-pass filtration effect of calcium dye kinetics, correlations in unprocessed calcium signals are higher than typical firing rate correlations on shorter temporal scales. The close similarity of the two estimates of the correlation matrix (Fig. 1 B and C) belies the utterly different partial correlation structure (Fig. 1 D and E): The regularized estimate is produced from the same data by setting to zero 31,501 (71.2%) of the off-diagonal coefficients of the precision matrix and fitting the remaining 12,752 coefficients.

Estimation by covariance selection is attractive for several reasons: First, with fewer free parameters it is less susceptible to sampling noise, yielding estimates that are, on average, closer to the truth. Additionally, if we assume that partial correlations do indeed reflect elements of the functional connectivity in the circuit, the second estimate also appears more plausible: in neural circuits, each neuron only interacts with a fraction of the entire population.

Conceptual questions and basic approach 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

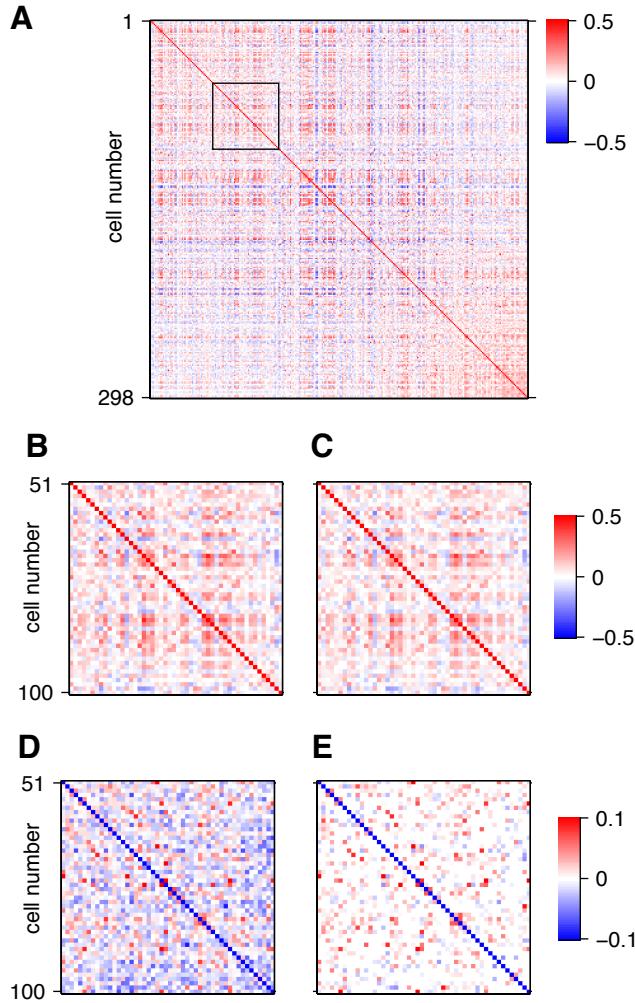


Figure 1. Illustration of regularized estimation of partial correlations. **A.** The sample correlation matrix of unprocessed somatic calcium signals from a population of cells in mouse visual cortex. The outlined square fragment is magnified in **B**. **C.** The same fragment of another estimate of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

within the recorded microcircuit?

To address these questions, we evaluated five covariance matrix estimators based on different assumptions about the underlying correlation structure:

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

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

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

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

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

First, we demonstrated that, in simulations with known true correlation structures, estimators that imposed a matching structure were generally more efficient than the other estimators. The estimators relative performance was evaluated by cross-validation. The logic of this study relies on the converse of

this relationship: the covariance matrix estimator that is found to be most efficient on empirical data is proposed to provide a better picture of the correlation structure than other estimators. Caveats for this type of inference are reviewed in the Discussion.

Therefore, we 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 fast 3D random-access two-photon imaging of calcium signals. We found that $C_{\text{sparse+latent}}$ consistently outperformed the other estimators. This estimator produced a sparse network of partial correlations between the observed neurons combined with correlations between several inferred latent units. We examined the structure of the sparse partial correlations inferred by the estimate with respect to the neurons geometrical positions and preferred orientations and found that stronger dependencies than those of sample correlations. In Discussion, we outline possible approaches to further corroborate the relationship between the structure of the functional connectivity revealed by such methods to the anatomical organization of the circuit.

Results

Covariance estimation The icovariance matrix defined as

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

where $\mathbb{E} [\cdot]$ denotes expectation; x is the $p \times 1$ vector of real-valued instantaneous firing rates in bins of duration Δt ; and μ is the vector of mean firing rates. In the case of noise correlations, μ depends on the stimulus condition. Usually, neural covariance matrices have been estimated as the *sample covariance matrix* C_0 calculated from the empirical sample of observations $x(t)$, $t = n(1, \dots, n)$ as

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

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

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

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

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

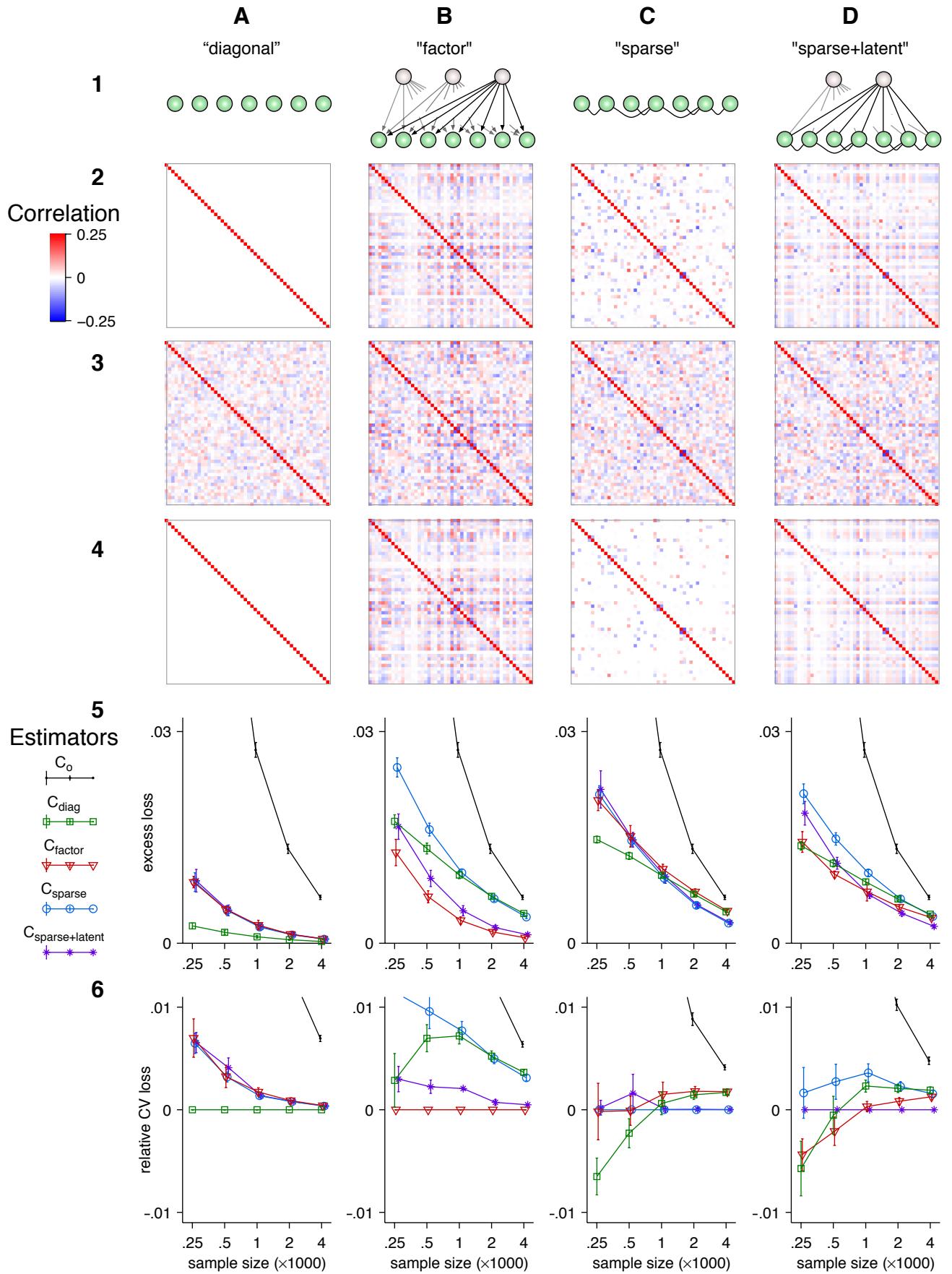
$$P = -(I \circ C^{-1})^{-\frac{1}{2}} C^{-1} (I \circ C^{-1})^{-\frac{1}{2}} \quad (4)$$

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

We considered four regularized estimators based on distinct families of low-dimensional target estimates: C_{diag} , C_{factor} , C_{sparse} , and $C_{\text{sparse+latent}}$. In probabilistic models with exclusively linear dependencies, their targets estimates can be described by distinct families of graphical models (Fig. 2 row 1).

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

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



The structure of target D favors the absence of linear associations between the activity of observed neurons (Fig. 2 A). If sample correlations arise from spurious or contingent effects, C_{diag} should outperform the other estimators.

The target of estimator C_{factor} is the factor model $L + \Psi$, where L is the $p \times p$ positive definite matrix of rank d , 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 number of factors d and shrinkage intensity λ . The target estimate $L + \Psi$ has the structure that arises when correlated fluctuations in population activity are linearly driven by a number of latent factors that affect many cells while direct interactions between cells are insignificant (Fig. 2 B).

Estimator C_{sparse} is based on the assumption that correlations result from direct linear associations between pairs of observed neurons and that only a fraction of neuron pairs interact (Fig. 2 C). This assumption is enforced by fitting only a subset of the off-diagonal elements of the precision matrix while setting the rest to zero. The estimator takes the form

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

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

Estimator $C_{\text{sparse+latent}}$ allow the assumptions of both C_{factor} and C_{sparse} . Its correlation structure can be explained by linear effects exerted between some pairs of observed neurons and several latent units (Fig. 2 D). The estimator has the form [50, 51]:

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

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

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

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

An evaluation of the performance of a covariance matrix estimator C requires a real-valued *loss function* $\mathcal{L}(C, \Sigma)$. The loss function quantifies the deviation of the estimate C from truth Σ and attains its minimum when $C = \Sigma$.

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

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

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

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

Although the two functions appear similar in form, they differ conceptually: The log-likelihood is a function of the parameter Σ given the sample covariance matrix C_0 whereas the loss function expresses the distance from an arbitrary estimate C to the given truth Σ .

The normalization by $\frac{1}{p}$ is not standard but serves to make the values of the loss functions comparable across different population sizes.

The choice of loss function is motivated by mathematical convenience. We expect that the main conclusions of our study will not change qualitatively with other well behaved loss functions, such as the Frobenius norm of the difference $\|C - \Sigma\|_F$ [43, 47], Stein's entropy loss, or quadratic loss [52, 53].

We define the *excess loss* as

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

which assumes zero at its minimum.

We computed the excess losses for 30 samples for each sample size $n = 250, 500, 10000, 2000$, and 4000 for each of the five estimators and for each ground truth (Fig. 2 row 5). For each estimator, hyperparameters were estimated by nested cross-validation (See Methods.). As expected, estimators whose structure matched that of the true model typically outperformed the other estimators. There were two exceptions: First, for some smaller samples, the data were insufficient to reveal the true correlation structure and the estimators with the simpler structure, C_{diag} and C_{factor} outperformed the estimators with matching structure (Fig. 2 columns C and D, rows 5 and 6). Second, $C_{\text{sparse+latent}}$ often performed equally well to the sparse estimator even when the ground truth did not have any latent units. In these cases, it inferred the right number of latent units, zero, and estimated optimally.

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

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

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

Then the cross-validation loss is in fact an unbiased estimate of loss.

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

The property of additivity does not hold for other popular loss functions such as Stein's entropy loss or various quadratic losses. With non-additive loss functions, validation loss is a biased estimate of loss, but it is still generally appropriate to be used in lieu of loss.

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

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

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, allowing for other loss functions to be used.

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

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

In simulation, CV-loss accurately reproduced the differences between the estimators’ excess losses, although with greater variability (Fig. 2 Row 6). For each kind of ground truth (Row 2), relative CV-losses were computed with respect to the estimator whose regularization target matched the structure of the respective ground truth.

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

Covariance estimation in neural data We recorded calcium activity in dense populations of neurons in the supragranular layers in primary visual cortex of anesthetized mice using fast random-access 3D scanning two-photon microscopy [54–56]. We presented 300 repetitions of full-field drifting gratings with two directions of motion or 150 repetitions with five directions (Fig. 3 A) on one side of the visual field of anesthetized mice. Groups of cells loaded with a calcium-sensitive fluorescent dye were imaged and localized in 3D space (Fig. 3 B and E) in the visual cortex on the contralateral side to the stimulus. Using acousto-optic deflectors (AOD) to steer the laser in 3D, we recorded the somatic calcium activity from the located cells with concurrent motion detection [56]. This technique allowed fast sampling (100–150 Hz) from large numbers (150–350) of cells in a small volume of cortical tissue ($200 \times 200 \times 100 \mu\text{m}^3$) in layers 2/3 and 4 (Fig. 3 C). After downsampling the signal to 20 Hz, relative firing rates were inferred using sparse nonnegative deconvolution [57] (Fig. 3 C and D). Only cells that produced detectable calcium activity were included in the analysis (See Methods). The average stimulus response was subtracted from each trial; the remaining signals were further downsampled into 150 ms bins to compute the noise correlation matrix (Fig. 3 F and G). The durations of recordings dedicated to estimating the noise correlations ranged between 15–20 minutes. Besides the high-repetition stimulus protocol used to estimate noise correlations, an orientation tuning stimulus protocol was used to map the orientation tuning of each cell (Fig. 3 E). Overall, 31 imaged sites from 24 animals were included in the study.

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

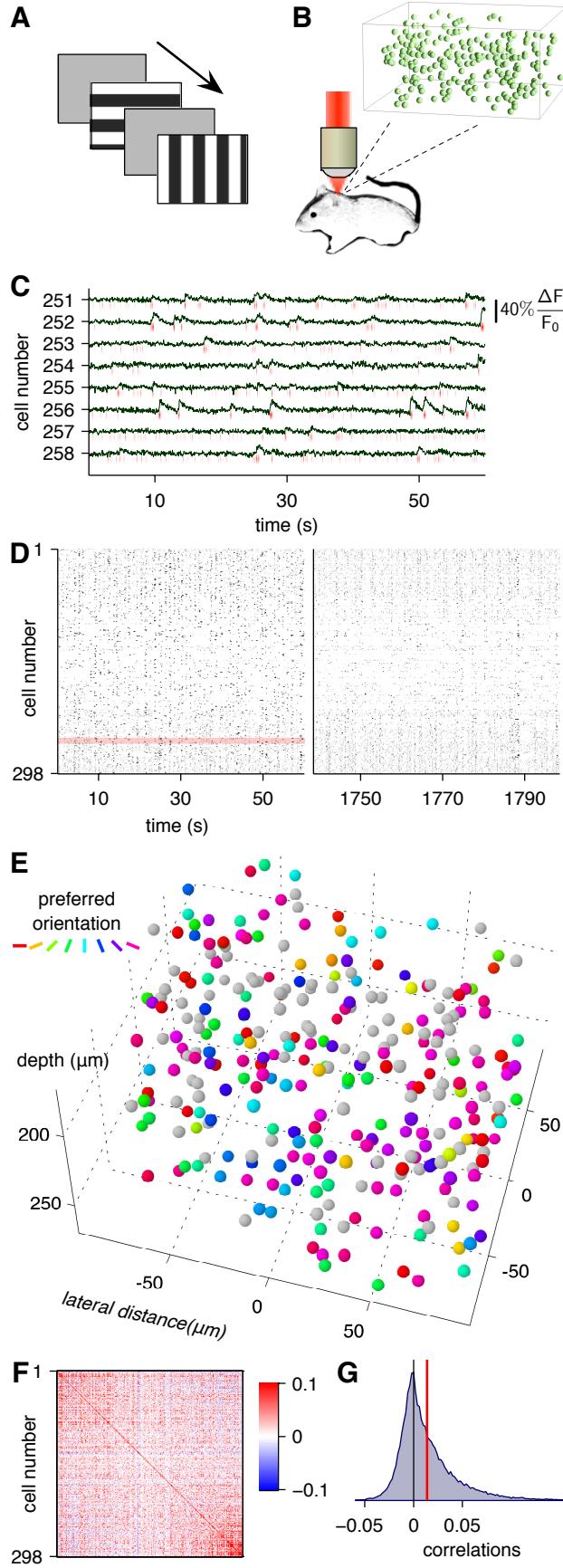


Figure 3. Acquisition of neural signals for the estimation of noise correlations. Visual stimuli comprising full-field drifting gratings interleaved with blank screens (A) were presented to anesthetized mice while two-photon recordings of somatic calcium signals were collected using fast 3D random-access microscopy (B). The visual stimulus included an initial period with 16 directions of motion for orientation tuning followed by a longer (15–20 min) period of stimulation with only 2 or 5 directions of motion for the computation of the noise correlation matrix. C. Representative calcium signals from eight cells out of 298 cells downsampled to 20 Hz. The inferred firing rate binned in 150 ms intervals are indicated by red ticks below each trace. D. The raster plot of the inferred firing rates, binned in 150 ms intervals, from the entire population from the first (left) and last (right) minute of the entire recording. The traces from C are highlighted in red. E. The spatial arrangement and orientation tuning of the 298 cells from the imaged site. F. The noise correlation matrix of the activity of the neural population. G. Histogram of noise correlation coefficients. The red line indicates the mean.

estimates biased toward low-rank correlation structure (*e.g.* C_{factor} , $C_{\text{sparse+latent}}$). Thus it was *a priori* unclear which estimator would perform best.

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

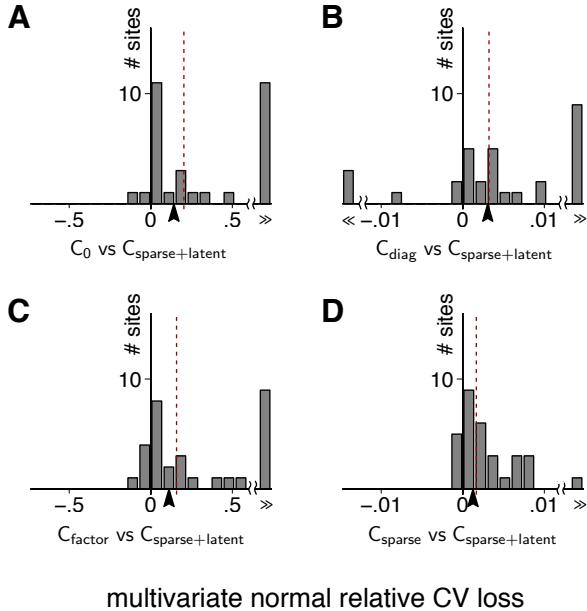
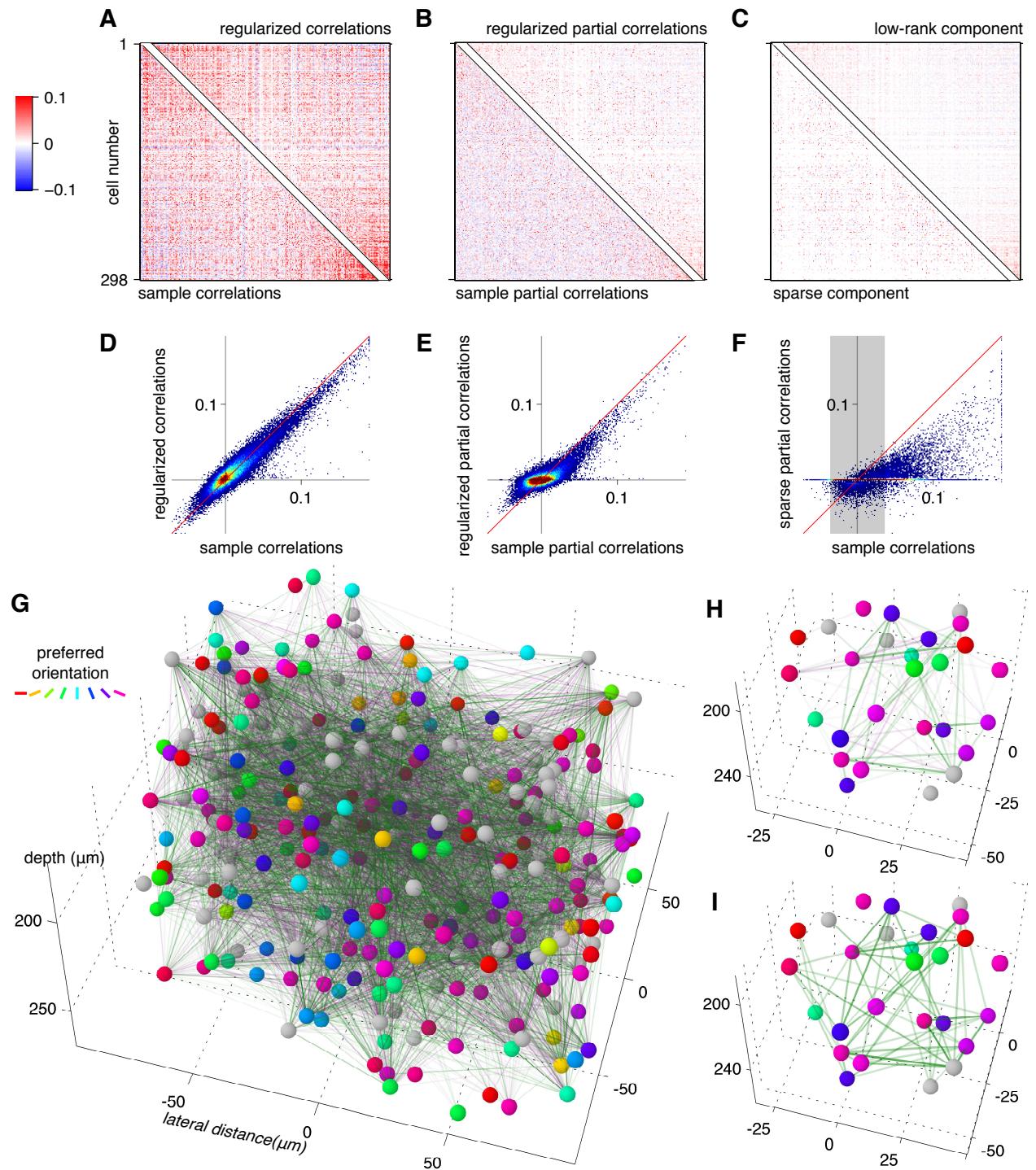


Figure 4. The sparse+latent estimator $C_{\text{sparse+latent}}$ outperforms the other estimators on neural data. **A–D.** Histograms of average cross-validation loss differences of the respective estimators C_0 , C_{diag} , C_{factor} , and C_{sparse} from $C_{\text{sparse+latent}}$. The histograms are based on 31 imaged sites in 24 mice. All medians (red dashed lines) were significantly greater than zero, indicating the dominance of $C_{\text{sparse+latent}}$ over the other estimators. The arrow heads indicate the results for the site shown in Fig. 3 and Fig. 5.
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, $p = 5.5 \times 10^{-6}$.

Correlation structure and circuit architecture Having demonstrated that $C_{\text{sparse+latent}}$ dominated the other estimators, we examined the low-dimensional structure revealed by this estimator for individual imaged sites. By design, the sparse+latent estimator finds the mixture of sparse partial correlations between visible neurons and common fluctuations described as latent units. If this estimate dominates, we can hypothesize that it better describes underlying physiological interactions than other estimators, including the sample covariance. In particular, the sparse component of the partial correlation matrix suggests direct interactions between pairs of neurons, whereas the low-rank component suggest common fluctuations such as those caused by common inputs or other collective synchronized activity.

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

In previous studies of the structure of correlations, a sparse network was produced by thresholding the correlations coefficients at a level deemed significant. A network was formed by connecting neurons whose total correlations exceed this threshold [25, 58]. There was fairly little overlap between the network of interactions revealed by such thresholding and those revealed by the sparse+latent estimator (Fig. 5 F). When thresholded to the same sparsity (82.1%), only 42% of cell pairs connected in one network were connected in the other while the average magnitude of such correlations was much lower in the case of



the regularized estimator (Fig. 5 F, H, and I). In particular, many low sample correlations translated into negative interactions in the regularized estimate. Indeed, the absence of correlations between pairs of cells that both correlate similarly to several of their neighbors should be considered as significant as a high correlation coefficient. Regularized partial correlations reveal such phenomena whereas regular sample correlations cannot.

The average partial correlations revealed by the sparse+latent estimator at all 31 sites were about 5 times lower than the sample correlations and less variable across sites (Fig. 6 A). The average node degree of the sparse component of the partial correlations and the number of inferred latent units varied widely between sites but generally increased with recorded population size (Fig. 6 B and C). However, there was an inverse relationship between the number of latent units and the average node degree (Fig. 6 D). Several sites, even with relatively large population sizes, had fairly few pairwise interactions and were dominated by latent units. These differences have multiple explanations, including differences in brain states and recording quality (neuropil contamination, motion).

We also examined the relationship of differences in orientation preference and physical distances (lateral and by depth) between pairs of cells to the average sample correlations, average regularized partial correlations, and inferred connectivity between them (Fig. 7). The average partial correlations fell more rapidly with difference in preferred orientation (Fig. 7 A and D) and lateral displacements at equal depths ($\pm 25 \mu\text{m}$) (Fig. 7 B and E), and differences in depth at small ($\pm 25 \mu\text{m}$) lateral displacements (Fig. 7 C and F).

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

Discussion

Functional connectivity Effective phenomenological or statistical descriptions of neural activity can serve as a starting point for investigation of underlying physiological mechanisms. For example, the discoveries of receptive fields and tuning properties of single neurons, over the past 60 years, have led to fruitful hypotheses about the underlying patterns of synaptic connectivity and principles of neural coding, and to the discovery of other organizational principles such as various cortical feature maps [59]. The

Figure 5 (preceding page). Example of low-dimensional correlation structure revealed by the sparse+low-rank estimator. **A.** The regularized estimate of the correlation matrix (top-right) closely approximates the sample correlation matrix (bottom left). This close approximation is also demonstrated by the scatter plot of the correlation coefficients produced by the two estimates (**D**). However, the partial correlation matrices from the two estimate show more pronounced differences (**B** and **E**). **C.** Furthermore, the partial correlation matrix of the regularized estimate is decomposed into a sparse component with 82.2% off-diagonal zeros (bottom-left) and low-rank component of rank 15 (top-right). **F.** The sparse component of the regularized partial correlation matrix had little resemblance to the sample correlations. The gray interval indicates the range of correlations containing 82.2% of cell pairs, equal to the fraction of zeros in the sparse partial correlation matrix. This interval contained 58.9% of the partial correlations. **G.** The graphical depiction of the positive (green) and negative (magenta) partial correlations as edges between observed neurons. The line density is proportional to the magnitude of the correlation. **H.** A subset of neurons from the center of the cluster shown in **G** showing the regularized partial correlations. **I.** The same subset with sample correlations thresholded to match the sparsity of the regularized interactions.

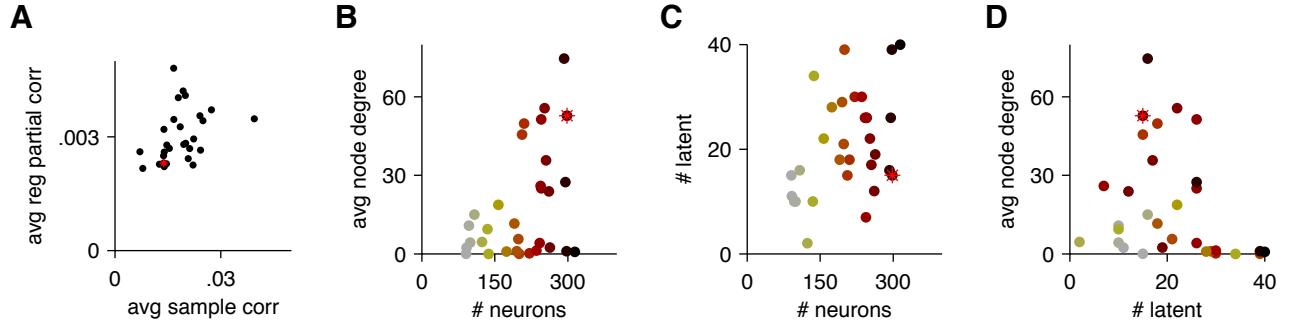


Figure 6. Properties of sparse+low-rank regularized estimates from all imaged sites

A. The average sample correlations vs. average partial correlations for each imaged site. In each plot, the red asterisk indicates the site shown in figures 3 and 5. **B.** The average node degree for sparse partial correlations vs. population size in each imaged site. **C.** The number of inferred latent units vs. population size in each imaged site. **D.** The number latent units vs. average node degree for sparse partial correlations for each site.

search for similarly effective descriptions of observed *multineuronal* activity patterns that are not reducible to the response properties of the individual cells has produced various statistical expressions collectively known as ‘functional connectivity’. Functional connectivity is commonly expressed as the degree of correlation, synchronization, or other forms of statistical association between neurons and is thought to reflect internal local recurrent connectivity, shared interactions with other regions, or endogenous network activity. To remove the confounding effects of stimuli, the average stimulus responses are usually subtracted or otherwise accounted for. Although various measures of functional connectivity are only phenomenological descriptions of population activity, they provide experimental neuroscientists with tools to relate population activity to the anatomical and functional architecture of neural circuits and to generate hypotheses about underlying mechanisms.

Functional connectivity as a network of pairwise interactions The intended purpose of many studies of functional connectivity is to estimate the circuits anatomical connectivity solely from the observed multineuronal spiking activity. For example, characteristic peaks and troughs in the pairwise cross-correlograms of recorded spike trains contain statistical signatures of directional monosynaptic connections or shared synaptic inputs [12, 14, 30, 31, 60]. Such signatures are still ambiguous as they can arise from diverse network effects other than direct synaptic connections [61]. With simultaneous recordings from a larger subset of neurons in the circuit, ambiguities can be resolved by inferring the conditional dependencies between pairs of neurons. Since direct causal interactions between neurons produce statistical dependency even after conditioning on the state of the rest of the network, conditionally independent neurons can be inferred to lack a direct causal influence. Such conditional dependencies can be inferred by fitting a probabilistic model of the joint population activity. For example, generalized linear models (GLMs) have been constructed to include biophysically plausible synaptic integration and membrane kinetics, individual neurons stimulus drive, and a network of synaptic couplings [62]. Another family of models are maximum entropy models that are constrained by the observed pairwise correlations [?, 63–66]. Although not informed by physiology, maximum entropy models are attractive because they account for the observed statistics with fewest other assumptions about the distribution: the investigator need only choose which statistics to match. Under the multivariate normal distribution, the conditional dependencies between pairs of neurons are expressed by the partial correlations between them. Each probabilistic model, fitted to the same data may reveal a different network of ‘interactions’, *i.e.* conditional

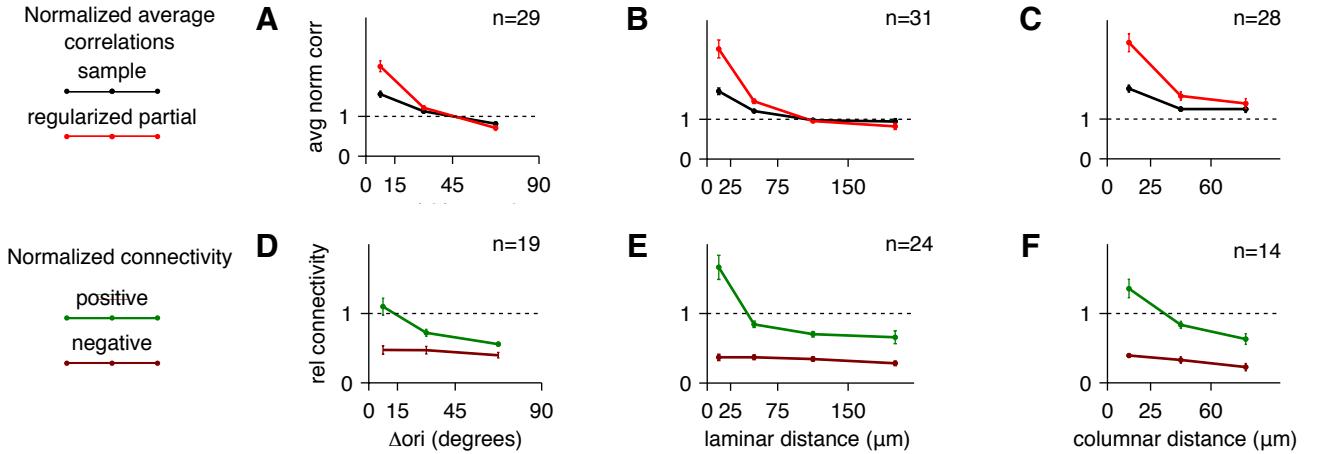


Figure 7. Dependence of correlations and partial correlations on orientation tuning differences and physical distance between cell pairs. A–C. Average partial correlations (red) estimated by $C_{\text{sparse+latent}}$ and average sample correlations (black) averaged across multiple imaged sites. In each site, the correlations were normalized by the respective average correlation shown in Fig. 6 A. The number n of sites that qualified to be included in the analysis is indicated. Sites were included if they had at least 20 pairs of neurons in each of the intervals. The error bars indicate the standard error of the mean for based on n . A. Average correlations between pairs of neurons tuned to orientation with differences in preferred orientation in the intervals of 0–15°, 15–45° and 45–90°. B. Average correlations between pairs of neurons located at the same depth ($\pm 25\mu\text{m}$) separated by lateral distances in the intervals of 0–25 μm , 25–75 μm , 75–150 μm , and 150+ μm . C. Average correlations between pairs of neurons displaced laterally by less than 25 μm separated in depth by distances in the intervals of 0–25 μm , 25–60 μm , and 60+ μm . D–F. Normalized connectivity of positive (green) and negative (dark red) interactions from the sparse component obtained from $C_{\text{sparse+latent}}$. Normalized connectivity was computed as the fraction of pairs connected by interactions of corresponding signs in each interval divided by the fraction of non-zero interactions across the entire site. Site were included in the analysis only if they had at least 20 cell pairs in each interval. The error bars indicate the standard error of the mean based on the number n of sites included in the analysis.

dependencies between pairs of neurons.

The inference of the conditional dependencies also depends on the completeness of the recorded population: To properly condition the interaction between a pair of neurons on the activity of the other neurons under a given probabilistic model, the recording must include all the other neurons with which the pair can potentially interact. Otherwise, interactions with the unobserved portions of the circuit will introduce false conditional dependencies between the observed neurons. For this reason, the most successful applications of statistical models of population activity thus far have been in *in vitro* preparations of the retina or cell cultures where high-quality recordings from the complete populations have been possible [62]. When inserted in cortical tissue, electrode arrays record from a small fraction of cells in a given volume, limiting the validity of the inference of the conditional dependencies between pairs of cells. Perhaps for this reason, until our study, partial correlations have not been used to describe the functional connectivity between cortical neurons. Despite this limitation, many studies have fitted probabilistic models to multielectrode recordings and analyzed the structure of the resulting conditional dependencies [65, 67].

Because the validity of the inference of interactions from probabilistic models depends on both the

correctness of their mathematical form in the model and the completeness of the recorded population, it is far from clear which approach provides best correspondence to the anatomical connectivity in the circuit, with no experimental evidence yet available to address this question. The result is often treated as a graphical representation of the functional connectivity without a statement about its anatomical implications. Whatever these implications may be, investigators have applied graph-theoretical analyses to the inferred networks to make statements about the design principles and efficiency of circuit architectures [58, 65, 68].

Two-photon imaging of population calcium signals presents unique advantages for the estimation of functional connectivity despite some of the disadvantages. The temporal resolution of calcium signals is limited by calcium dye kinetics. Fast imaging techniques combined with advanced spike inference algorithms have been demonstrated to provide millisecond-scale temporal resolution of single action potentials [55, 56, 69]. However, such high temporal precision comes at the cost of the accuracy of the inferred spike rates: better accuracy can be achieved when calcium signals are analyzed on the temporal scales of tens of milliseconds [56]. For what calcium imaging lack in temporal resolution, it make up in spatial resolution and the ability to characterize the spatial arrangements of cells and identify their types. Recently, advanced imaging techniques have allowed recording from nearly every cell in a volume of cortical tissue *in vivo* [55, 56] and even from entire nervous systems [70, 71]. Such completeness may open new opportunities for inferring networks of conditional dependencies between neurons with greater validity than with electrophysiological recordings.

The low temporal resolution of calcium signals limits the use of functional connectivity methods that rely on millisecond-scale binning of signals (cross-correlograms, some GLMs, and maximum entropy models on the binary domain). Most studies of functional connectivity have relied on instantaneous sample correlations [22, 25, 58, 72]. Although some investigators have interpreted such correlations as indicators of (chemical or electrical) synaptic connectivity, most used them as more general indicators of functional connectivity without a statement about underlying mechanisms.

In our study, we sought to express pairwise connectivity networks among neurons in local cortical microcircuits. We hypothesized that partial correlations would correspond more closely to underlying mechanisms than sample correlations. Since neurons form synaptic connections mostly locally and sparsely [73], we a priori favored solutions in with sparse partial correlations. The importance of partial correlations may be justified by the principle of maximum entropy: The maximum entropy distribution on discrete or continuous multivariate domains constrained by the observed mean firing rates, their variances, and correlations is the multivariate normal distribution. The interactions (conditional dependencies) between pairs of neurons inferred by this distributed are expressed by the precision matrix (inverse of the covariance) or, equivalently, by the matrix of the pairwise partial correlations. Therefore, under the aforementioned assumptions that the recorded population is sufficiently complete and that the model correctly represents the nature of interactions, the network of partial correlations can be hypothesized to be a better representation of functional dependencies than sample correlations.

Functional connectivity as coactivations Another approach to describe the functional connectivity of a circuit is to isolate patterns of multineuronal coactivations [36–39, 74, 75]. 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 [74, 75], *i.e.* groups of tightly interconnected groups of cells involved in a common computation. Coactivation patterns could also be explained as shared input from unobserved parts of the circuit. Coactivations could also be explained by global network fluctuations modulating the activity of the local circuit [76].

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

In our study, the analysis of functional connectivity through modes of coactivations was represented by the factor analysis estimator C_{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 [77], high-order correlations [78], or global network fluctuations [79].

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 [50,51]. The resulting estimator, $C_{\text{sparse+latent}}$, effectively decomposed the functional connectivity into a sparse network of pairwise interactions and coactivation mode vectors.

Addressing ill-posedness Inferring the conditional dependencies between variables in a probabilistic model is an ill-posed problem: small variations in the data produce large errors in the inferred network of dependencies. The problem becomes increasingly worse as the size of the recorded population of neurons increases until such models lose their statistical validity [80]. As the recorded neuronal population sizes increased gradually, experimental neuroscientists addressed this problem by extending the recording durations to keep sampling noise in check and verified that existing models are not overfitted and that the stationarity assumption are observed [79]. However, with ambitious projects, such as the BRAIN initiative currently underway [81] aiming to record from significantly larger populations, the problem must be addressed by regularizing the solution. Regularization limits the solution to a smaller subspace to counteract the effect of the sampling noise in the empirical data. However, limiting the solution to an inappropriate subspace does not allow significant improvement.

Several strategies have been developed to limit the model space in order to improve the quality of the estimate. For example, [78] 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 [62].

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

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

Model selection In our study, the covariance matrix estimators were evaluated with respect to the cross-validated normal log likelihood. However, this does not limit the applicability of its conclusions to normal distributions. Indeed the major findings in this paper could be reproduced with respect to other loss functions (compare Fig. 4 and Fig. S2). Other probabilistic models, fitted to the same data, would also serve as estimators of the covariance matrix. If a different model yields better estimation of the covariance matrix than the estimator proposed here, we believe that its structure should deserve consideration as the better representation of the functional connectivity.

The results of model selection must be interpreted with caution. As we demonstrated by simulation, a simple model can produce a better estimate even if it does not correctly represent the real dependencies in the data than a more complex model with correct representation of dependencies. Therefore, showing that a more constrained model has better cross-validated performance than a more complex model does not support the inference that it reveals a better representation of dependencies in the data. This caveat is related to the phenomenon known as *Steins Paradox* [82]: in which the biasing of an estimate toward an arbitrary target is guaranteed to produce *some* improvement.

Physiological interpretation and future directions Although the network of interactions inferred by the optimal estimator, $C_{\text{sparse+latent}}$, alludes synaptic interactions, it is unlikely that the anatomical and functional networks are identical. The improved estimation performance gives the model extra credence over other estimates of the functional connectivity. The inferred interaction network was substantially different from that inferred from by considering the network of most significant sample correlations (Fig. 5 F, H, and I). In particular, sample correlations do not draw attention to the absence of correlation between a pair of neurons that should be strongly correlated based on their common correlations with other neurons. The $C_{\text{sparse+latent}}$ estimator identified many such pairs and inferred negative interactions between them (Fig. 5 F). The resulting network had a larger fraction of negative interactions than the fraction of negative correlations in the sample correlation matrix. The negative interactions had a distinct spatial organization from Reproducing our study with labeled interneurons of various types could indicate whether the negative interactions

The estimator effectively decomposed the correlation structure into two distinct components: pairwise interactions and common coactivation modes denoted as ‘latent units’. Since the distinction led to improved estimation, the two types of components could be further investigated. For example, slow-wave network fluctuations due to anesthesia or slow-wave sleep could be strongly reflected on the common components/latent units but much less on the pairwise interactions. Previous studies found that correlations measured under anesthesia retained little in common to correlations measured in wakefulness and concluded that studies of the anesthetized brain did not inform about its behaviorally relevant functions [72]. When conditioned on the brain state, optimal estimation of covariance matrices in dense cortical networks could reveal that slow-wave activity alters the common components but does not affect the pairwise interactions as strongly. The correlation structure also changes as a function of the stimulus condition [56] on time scales that are much faster than changes in the anatomical connectivity. Descriptions of specific changes in the functional connectivity due to stimulus properties could shed light on how circuits orchestrate their activity to represent diverse stimuli.

The malleability of the joint structure of activity in populations of neurons makes functional connectivity statistics indispensable wholly apart from the ability to infer the anatomical connectivity. Effective statistical descriptions of the functional connectivity will need to be developed to infer the most significant features of the functional organization in large ensembles of neurons from limited data. These approaches will need to rely on techniques exemplified by our study: selection among multiple mathematical representations of dependencies, regularization, and model selection.

Methods

Surgery and two-photon Imaging All procedures were conducted in accordance with the ethical guidelines of the National Institutes of Health and were approved by the Baylor College of Medicine IACUC. The surgical procedures and data acquisition were performed as described in [56]. Briefly, C57BL/6J mice (aged p40–60) were used. Anesthesia was initiated with isoflurane (3%) and the mixture of fentanyl (0.05 mg/kg), midazolam (5 mg/kg), and medetomidine (0.5 mg/kg), with boosts of half the initial dose every 3 hours. A craniotomy was performed over the right primary visual area. Membrane-permeant calcium indicator Oregon Green 488 BAPTA-1 AM (OGB-1, Invitrogen) was loaded by bolus injection. The craniotomy was sealed using a glass coverslip secured with dental cement.

Calcium imaging began 1 hour after dye injection. All imaging was performed using the 3D-RAMP two-photon microscope described in [56]. First, a 3D stack was acquired and cells manually segmented. To collect calcium signals, the system repeatedly hopped between the selected neurons.

Visual stimulus Full-field drifting gratings with 90% contrast, luminance of 10 cd/m², spatial frequency of 0.08 cycles/degree, and temporal frequency of 2 cycles/s. Two sets of stimuli were presented for each imaging site: the first to map directional tuning and the second to estimate noise correlations.

Directional tuning was map using a pseudo-random sequence of drifting gratings at sixteen equally spaced directional of motion changing at 2 Hz for 3 min without blanks. The data for covariance estimation were collected during presentations of full-field drifting gratings with the same parameters as those used in directional tuning except only two directions (in 9 dataset) or five directions (in 22 datasets) were used and the presentations lasted 1 second and, separated by 1-second blanks. Each stimulus condition was presented at least 180 times.

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

The collected fluorescent traces were deconvolved to reconstruct the firing rates for each neuron. First, the first principal component was subtracted from the traces, which reduced common mode noise related to small movement and cardiovascular artifacts. The resulting traces were low-pass filtered below 0.1 Hz and downsampled to 20 Hz. Firing rates were estimated using a fast non-negative deconvolution algorithm [57].

Orientation tuning was computed by fitting the mean firing rates in response to gratings of directions ϕ with two-peaked von Mises tuning functions of the form $f(\phi) = a + b \exp\left[\frac{1}{w}(\cos(\phi - \theta) - 1)\right] + c \exp\left[\frac{1}{w}(\cos(\phi - \theta + \pi) - 1)\right]$ where $b \geq c$ are amplitudes of the two respective peaks, w is the tuning width, and θ is the preferred direction. The significance of the fit was determined by the permutation test: the labels of the direction were randomly permuted 10,000 times. The p -value of the fit was computed as the fraction of permuted datasets for which the R^2 value of the tuning function fit exceeded that of the real data. Cells were considered tuned for p -values not exceeding 0.05.

For covariance estimation, the analysis was limited to the period with 2 or 5 stimulus conditions and lasted between 12 and 20 mins. Only traces whose average deconvolved signal was greater than 1% of the median of the recorded population were included in the analysis.

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

Since each of the regularized estimators had one or two hyperparameters, we used *nested cross-validation*. The outer loop evaluated the performance of the estimators with optimal values of the hyperparameters. The optimization of the hyperparameters was performed within the inner loop in two phases: random search to find a good starting point and pattern search to find the global minimum. The inner cross-validation loop subdivided the training dataset from the outer loop to perform 10-fold cross-validation in order to evaluate each choice of the hyperparameter values. Thus the size of the training dataset within the inner loop comprised 81% of the entire recording.

When cross-validation loss was not required, only the inner loop of cross-validation was used, applied to the entire dataset. This approach was used to compute the covariance matrix estimates and their excess-loss in the simulation study (Fig. 2 rows 3 and 4) and to analyze the partial correlation structure of the sparse+latent estimator (Fig. 5-7).

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

Estimator C_{diag} (Eq. 5) used two hyperparameters: the covariance shrinkage intensity $\lambda \in [0, 1]$ and variance shrinkage intensity $\alpha \in [0, 1]$. The variances (the diagonal of C_0) were shrunk toward (linearly

mixed with) their mean value:

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

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

The `corpcor` package in the programming language R also implements this estimator [83], although its analytical optimization of the shrinkage intensities is based on the mean squared error whereas we optimized them with respect to the loss function in Eq. 9.

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

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

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

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

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

where $\tilde{Z} \succ 0$ denotes the constraint that \tilde{Z} be a positive definite matrix and $\|\tilde{Z}\|_1$ is the element-wise L_1 norm of the matrix \tilde{Z} . This problem formulation is known as *graphical lasso* [41]. To solve this minimization problem, we modified the alternative-direction method of multipliers (ADMM) algorithm developed by [51]. Then, after the zero structure was determined, the remaining coefficients were fitted without penalty:

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

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

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

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

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

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

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

The estimate is found in two steps. First, we use the ADMM algorithm to find the zero structure of S by minimizing the L_1 -penalized loss [50, 51]:

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

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

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

The partial correlation matrix computed from the overall estimate, computed from $C_{\text{sparse+latent}}$ according to Eq. 4, includes the effects of interactions between the visible and latent units. This estimate of the partial correlations was used in analyses where sparsity was not required (Fig. 5 B and E and Fig. 6 A, and Fig. 7 A–C). For analyses that required a sparse network (Fig. 5 F, G, H and Fig. 6 B, D and Fig. 7 D–F) was computed using the extended precision matrix S^* :

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

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

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

Acknowledgments

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

References

1. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366.
2. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143.
3. Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25: 3661–73.
4. Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior. *The journal of Neuroscience* 21: 1676–1697.
5. Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327: 584–587.
6. Abbott L, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural computation* 11: 91–101.
7. Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Physical Review E* 64: 051904.
8. 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. Smith MA, Sommer MA (2013) Spatial and temporal scales of neuronal correlation in visual area v4. *The Journal of Neuroscience* 33: 5422–5432.
22. 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.
23. Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. *Neuron* 76: 590–602.
24. Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. *Journal of neurophysiology* 109: 940–947.
25. 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.
26. 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.

27. Goard M, Dan Y (2009) Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci* 12: 1444-9.
28. Kohn A, Zandvakili A, Smith MA (2009) Correlations and brain states: from electrophysiology to functional imaging. *Curr Opin Neurobiol* 19: 434-8.
29. 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.
30. Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal* 7: 419-440.
31. Moore GP, Segundo JP, Perkel DH, Levitan H (1970) Statistical signs of synaptic interaction in neurons. *Biophysical Journal* 10: 876-900.
32. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870-96.
33. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience* 2: 539-550.
34. 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.
35. Rosenbaum R, Josić K (2011) Mechanisms that modulate the transfer of spiking correlations. *Neural computation* 23: 1261-1305.
36. Chapin JK, Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods* 94: 121-140.
37. 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.
38. Lopes-dos Santos V, Conde-Ocacionez S, Nicolelis MA, Ribeiro ST, Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis. *PloS one* 6: e20996.
39. Lopes-dos Santos V, Ribeiro S, Tort AB (2013) Detecting cell assemblies in large neuronal populations. *Journal of neuroscience methods* .
40. Dempster A (1972) Covariance selection. *Biometrics* : 157-175.
41. Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9: 432-441.
42. Anderson T (2003) An introduction to multivariate statistical analysis. Wiley series in probability and statistics .
43. 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.
44. Peng J, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* 104.

45. 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.
46. 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.
47. Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88: 365–411.
48. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, et al. (2006) Regularization in statistics. *Test* 15: 271–344.
49. 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.
50. 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.
51. Ma S, Xue L, Zou H (2013) Alternating direction methods for latent variable gaussian graphical model selection. *Neural computation* : 1–27.
52. 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.
53. Fan J, Fan Y, Lv J (2008) High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186–197.
54. Reddy GD, Saggau P (2005) Fast three-dimensional laser scanning scheme using acousto-optic deflectors. *J Biomed Opt* 10: 064038.
55. 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* .
56. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS (2013) Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7: 151.
57. 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.
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. Reid RC (2012) From functional architecture to functional connectomics. *Neuron* 75: 209–217.
60. Alonso JM, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature neuroscience* 1: 395–403.
61. Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *Journal of neurophysiology* 61: 900–917.

62. 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.
63. 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.
64. Tkacik G, Schneidman E, Berry I, Michael J, Bialek W (2006) Ising models for networks of real neurons. arXiv preprint q-bio/0611072 .
65. Yu S, Huang D, Singer W, Nikolic D (2008) A small world of neuronal synchrony. *Cereb Cortex* 18: 2891-901.
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. 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.
68. Feldt S, Bonifazi P, Cossart R (2011) Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends in neurosciences* 34: 225–236.
69. 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.
70. Leung LC, Wang GX, Mourrain P (2013) Imaging zebrafish neural circuitry from whole brain to synapse. *Frontiers in neural circuits* 7.
71. 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.
72. 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.
73. 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.
74. Gerstein GL, Bedenbaugh P, Aertsen AM (1989) Neuronal assemblies. *Biomedical Engineering, IEEE Transactions on* 36: 4–14.
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. 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.
77. 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 .
78. 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.
79. 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 .

80. 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.
81. Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, et al. (2013) The brain activity map. Science 339: 1284–1285.
82. Efron B, Morris CN (1977) Stein's paradox in statistics. WH Freeman.
83. Schäfer J, Opgen-Rhein R, Strimmer K (2010) corpcor: Efficient estimation of covariance and (partial) correlation. R package version 15.7.

Supporting Information

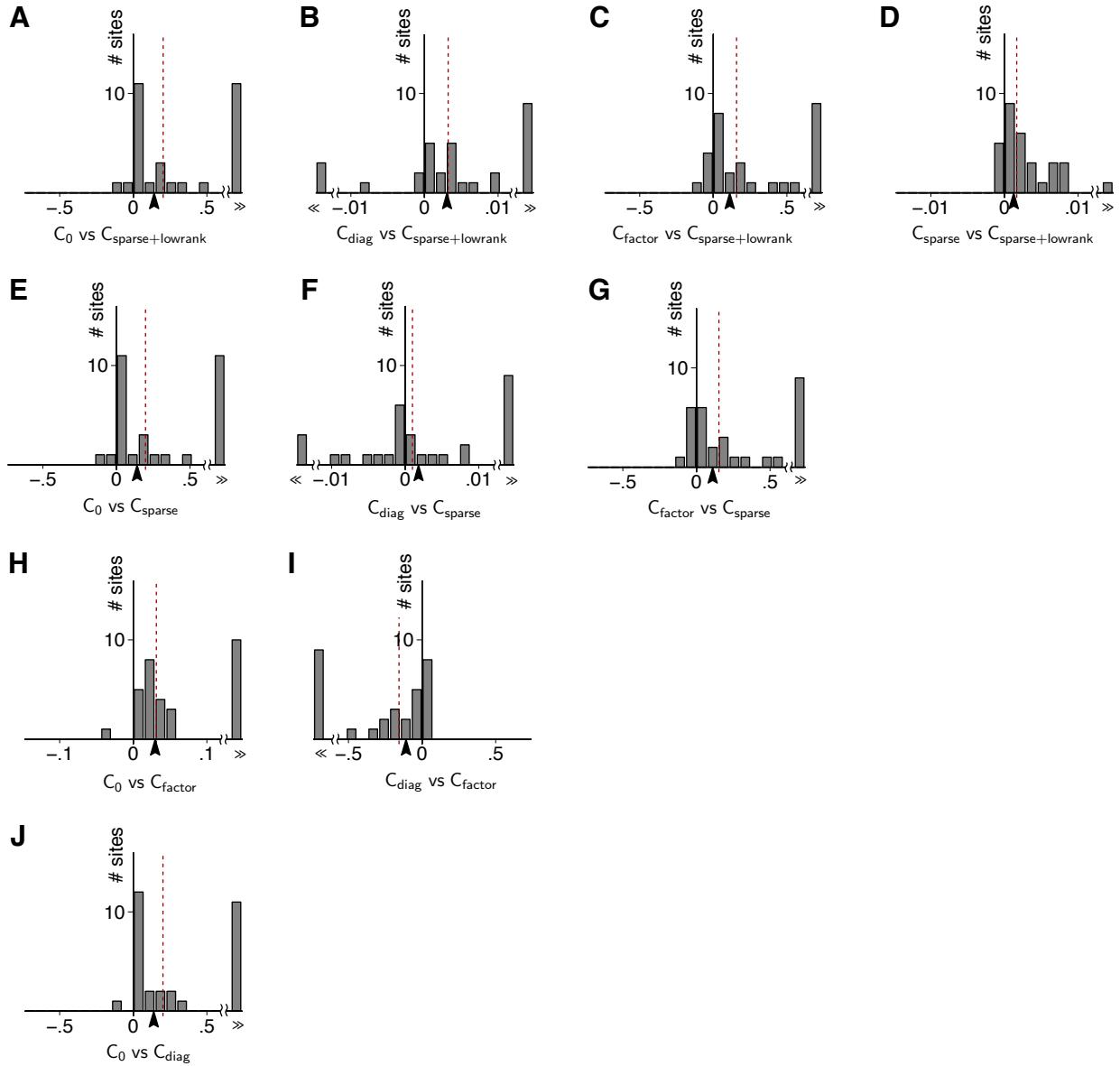


Figure 1. All-to-all performance comparisons of the sample covariance matrix and the four regularized estimators with respect to multivariate normal cross-validation loss.

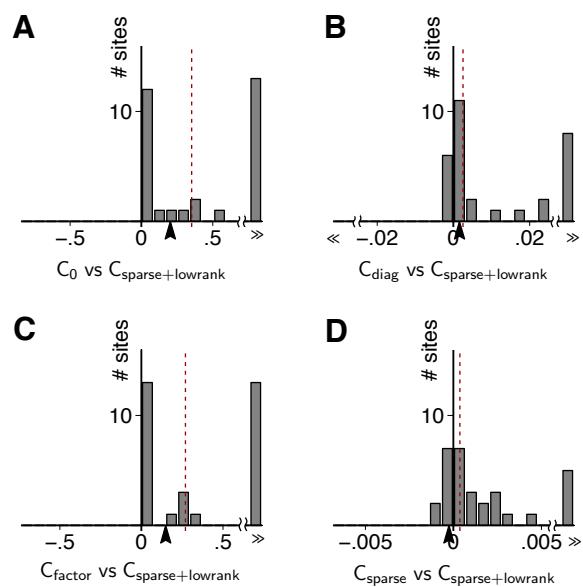


Figure 2. The sparse+latent estimator outperforms the other estimators with respect to quadratic cross-validation loss
 $\mathcal{L}(C, C'_0) = \frac{1}{p^2} \text{tr}(C^{-1}C'_0 - I)^2$.