# Responses to reviewer comments

We thank the reviewers for their careful reading of the manuscript and a number of valuable suggestions. To address these comments, we performed additional analysis and made several changes to the manuscript as summarized below:

## Reviewer #1.

***Comment 1.*** *I think several parts of the paper could be written more succinctly. For example the presentation of K-fold cross-validation is fairly standard and could be omitted or at least moved in the methods section. The abstract could also be compressed into a single more compact paragraph.  Overall the main message of the paper could be delivered more briefly.*

In the revised manuscript we followed the suggestion of the reviewer and removed the equations defining excess loss, cross-validation loss, and relative cross-validation loss. We also removed the detailed description of the cross-validation procedure from the Results section. This improved the manuscript in making the main points more succinct. All technical details about the cross-validation procedure are now included in the Methods section as recommended by the reviewer.

***Comment 2.*** *For readers unfamiliar with the relationship between partial correlations, precision matrices, and covariance matrices, it might be worth expanding the text around eqs (3-4) a bit.*

In the revised manuscript, we added a paragraph between Eq. 3 and Eq. 4 defining pairwise partial correlations and referencing a comprehensive text on graphical modeling (Whittaker, 1990). This paragraph begins with:

The partial correlation between a pair of variables is the Pearson correlation coefficient of the residuals of the linear least-squares predictor of their activity based on all the other variables. …

Chapter 5 of (Whittaker, 1990) provides detailed derivations that establish the relationship between the inverse covariance matrix and pairwise partial correlations.

***Comment 3.*** *I didn't find Fig 4G-I very interpretable or helpful.  The authors should try to improve these panels, or drop them.*

Figure 4G-I succinctly summarizes the main aim of the project: an inference of functional interactions in dense groups of cortical neurons with known orientation tunings and physical positions. These panels are not meant to lead to quantitative conclusions from this figure (more quantitative analysis is presented in Figures 5 and 6). Rather, the figure is meant to give a sense of the overall connectivity. Even with only 7% connectivity, the network *appears* densely connected, but some cells are not part of the network. Importantly, panels H and I illustrate what panel F shows more abstractly: connectivity is sensitive to the method of its inference (sparsified partial correlations vs. thresholded correlations).

***Comment 4.*** *The authors select the best sparsity and rank hyper parameters by cross-validation.  It would be nice to see how sharply peaked the corresponding objective function is.  For example, Fig 5 is hard to interpret if we don't know how sharply the obtained degree and latent rank are constrained by the observed data.*

We thank the reviewer for this comment. We have now added a supplementary figure (Fig. S1) showing the validation loss as a function of the hyperparameters of the sparse+latent estimator. Note that normal validation loss is proportional to multivariate normal cross-validation log-likelihood per cell per bin. With 292 cells and approximately 10,000 bins used in this analysis, the differences between the adjacent contour lines amount to large likelihood ratios: the step of 0.0005 in the validation loss equates to the likelihood ratio of over 10300 under the assumption that bins are independent.

***Comment 5.*** *Fig 6: the definition/interpretation of "normalized connectivity" was a bit hard to understand.  Could the authors clarify this a bit?*

In the revised document, the analysis in the main text does not rely on “normalized correlations” and “normalized connectivity.” These quantities are only used to produce the plots in Fig. 6. Their definition is now given in the caption of Fig. 6:

Here “connectivity” refers to the fraction of the non-zero elements in the sparse component S of the Csparse+latent estimator, which describes inferred direct interactions between specific pairs of the recorded neurons. Positive and negative connectivities refer to the fractions of the positive and negative partial correlations computed from S, respectively. “Normalized positive connectivity” is the ratio of the positive connectivity for pairs meeting a given condition, e.g. similar tuning with ∆ori < 15◦, to the average connectivity over the entire site. Normalized negative connectivity is computed similarly for the negative connectivity. The average connectivity across sites is shown in Fig. 5 B with only the five most connected sites included in the analysis. The normalization made the effects of tuning and distance more comparable across sites.

***Comment 6.*** *The authors argue in the discussion that inference of conditional dependencies between two neurons requires the observation of all the interacting neurons. This is intuitive but I was wondering to what extent the latent + sparse model can also account for unobserved units as suggested in [49] for gaussian graphical models. This could also be interesting to test either in the simulated or the real data by deliberately leaving some neurons out.*

This is an interesting question. Indeed the convex optimization algorithm proposed by Chandrasekaran, Parrilo, and Wilsky (2010) suggests the ability to discover true latent factors in some graphical models. However, when the number of latent variables is not known *a priori*, the cross-validation approach has limitations that prevent the discovery of specific latent factors in many realistic scenarios. In our simulations, a single unit with average connectivity did not get discovered by the cross-validation algorithm when excluded from the simulated recording (data not shown). To be discovered, the true latent factor must strongly interact with a large fraction of the observed population. In the simulated models used in our study (Fig. 1), the latent units interacted with all observed units. In this idealized case, the algorithm correctly discovered the right number of latent factors. In neural data, the actual number of latent variables must be very large, by far exceeding the number of observed variables. Therefore, the latent factors are likely to represent only common network fluctuations and not individual neurons. Indeed intentionally omitting some neurons from the recording decreased rather than increased the number of inferred latent units (data not shown). Indeed, the more units were recorded, the more latent variables were inferred (Fig. 5A in revised manuscript).

In our manuscript, we did not make specific claims about the identity of the latent factors. The fact that the sparse+latent estimator outperformed other estimators led us to propose that the remaining sparse interactions in such estimates may be more informative about the anatomical connectivity than sample correlations. The latent factors were used to explain the network effects that contribute to pairwise correlations but are not well explained by pairwise effects. Hence, the role of the latent factors was to isolate more significant pairwise effects, which then became the main focus of our. We left the properties of the latent units for future analyses.

***Comment 7.*** *It would also be interesting to repeat the analysis but conditioning on a specific stimulus each time. In this case, I'd expect that the latent part will be smaller for each stimulus (since the common input will be restricted to a specific stimulus). But it would be interesting whether the sparse part will remain more or less unchanged, something that will be more indicative of actual anatomical connectivity.*

This indeed would be a very interesting analysis. We did try some preliminary analysis and found interesting stimulus dependencies of the partial correlation structure. Unfortunately, to achieve statistical significance of these effects we will need more data than included in this manuscript. For this reason, we excluded this analysis from this paper.

***Comment 8.*** *In the papers where the low rank + sparse matrix decomposition method is discussed [49,50], a convex framework is proposed where the l\_1 and the nuclear norm of the matrix are penalized. The authors do cross-validation over all possible values instead of using the nuclear norm, which can reduce shrinkage but leads to a non-convex problem. Is there a specific reason for doing so? (In fact the optimization here would also be computationally more efficient since the low rank part is always a positive semidefinite matrix and thus its nuclear norm would be equal to just the trace.) Similarly, for l\_1 regularization the authors first determine the non-zero pattern with l\_1 regularization and then do an unregularized regression restricted to that set to find the correlation values. Why is that so?*

We followed the suggestion of the reviewer and used the convex framework for both the sparse and the low-rank components of the inverse covariance. The main conclusions of the paper were unaffected.

***Comment 9:*** *The link to the matlab code appears to be broken.*

We kept the repository private until publication but we are happy to provide access to the reviewers via the journal if requested.

## Reviewer #2

***Comment 1.*** *The correlation models differ in the number of parameters, and therefore the merits of different models in predicting the correlation structure of the data might need to be weighted by them. However, while comparing the models presented, the number of parameters is not taken into account. I think that this is quite important, and therefore, using a method such as Akaike information criterion or similar might add some generality to the interpretation of the results of the present manuscript.*

Many approaches have been devised for variable selection. We used nested cross-validation to choose the optimal values of the hyperparameters, which, in turn, determine the dimensionality of the estimates. We agree that the Akaike information criterion or other selection criteria could be used as well. We note, however, that cross-validation is more general and makes fewer assumptions about the data generating process (for review see Arlot (2010), *A survey of cross-validation procedures for model selection*). To address the reviewer’s question, we added the following paragraph under “Model selection” in the Discussion:

Various model selection criteria have been devised to select between families of models and the optimal subset of variables in a given model family based on observed data. Despite its high computational requirements, cross validation is among the most popular model selection methods due to its minimal assumptions about the data generating process \cite{Arlot:2010}

***Comment 2****. In the first part of the paper, data are generated through simulations using a multivariate Gaussian model. However, this model cannot fully represent the structure of the mouse visual cortex data that is modeled later. I think it would be interesting to reproduce Fig. 1 (rows 5 and 6) with a spiking generative model, such a correlated Poisson population, with different correlation structures. For instance, as a factor model, one could consider a Poisson model of independent neurons plus one or more mother spike trains that are summed up to the uncorrelated Poisson spike trains. This kind of analysis would be interesting to understand whether the reported results are robust against non-negativity of the firing rates and against large departures from Gaussianity (e.g. occurring at low firing rates).*

The correspondence between conditional independence and partial correlations in finite samples is a surprisingly non-trivial question sensitive to numerous assumptions and caveats. To properly test this correspondence in a simulated network of neurons, we would need to simulate multiple networks with various types of connectivity and spike generation mechanisms. While this is worthwhile, it goes beyond the scope of the present study. In our study, we present the empirical finding that the sparse+latent estimator of correlations is more efficient than others and hypothesized that the functional connectivity represented by the regularized estimate more closely corresponded to the anatomical connectivity than sample correlations. We propose to test this hypothesis empirically with neural data from ongoing and future experiments.

***Comment 3.*** *Finally, although the authors make predictions about the functional connectivity that is present in mouse visual cortex, a comparison with what it is known anatomically in cortex is not present.*

Following the reviewer’s suggestion we now include a discussion of the relationship between the functional and anatomical connectivity in the “Physiological interpretation and future direction” paragraph of the Discussion. Below is the relevant excerpt:

The distinct positive and negative connectivity patterns may reflect geometric and graphical features of local excitatory and inhibitory networks. Indeed, the relationships between patterns of positive and negative connectivities inferred by the estimator mirrored some regularities in studies of excitatory and inhibitory synaptic connectivities with respect to distance, cortical layers, and feature tuning [23,77,91–96]. For example, while excitatory neurons form synapses within highly specific local cliques [77], inhibitory interneurons form synapses with nearly all excitatory cells within local microcircuits [23,94,97]. To further investigate the link between synaptic connectivity and inferred functional connectivity, in future experiments, we will use molecular markers for various cell types with follow-up multiple whole-cell in vitro experiments [23,28] to directly compare the inferred functional connectivity graphs to the underlying anatomical circuitry.

This paragraph contains 10 references to prior studies of synaptic connectivity with respect to the distances between neurons’ somas, their cortical layers, and their cell types. Due to the great diversity of studies of anatomical connectivity in the cortex, we did not produce a direct comparison of the inferred functional connectivity to the density of anatomical connections described in earlier studies. Instead, we proposed to combine our analysis with tests for anatomical connectivity using whole-cell patching in slices *in vitro* following their functional characterization *in vivo.* This work is left to a future study.

# Summary of other changes

While responding to the reviewers’ comments, we identified and addressed a few other issues. Addressing these issues did not change the major findings of the original submission.

**Estimation of the means.** In the original submission, the mean responses were calculated and subtracted from the traces before the data were divided for cross-validation. In the revised study, the means are estimated in each of the training datasets and applied to both the training and the testing data duration evaluation. The latter approach is more principled since all the parameters necessary for evaluating the correlation matrix are estimated entirely from the training sample.

**Conditioning of variances.**  In the original submission, a common covariance matrix was estimated across all stimulus conditions. In contrast, in published studies of correlations under diverse stimuli, response variances are conditioned on the stimulus. This change required a somewhat more sophisticated cross-validation strategy, which we described in the revised manuscript under “Cross-validation with conditioned variances” in the Methods section.

In this approach, the condition-specific variances are estimated in the training dataset and applied to both the training and testing data during evaluation. This approach avoids potential biases due to the assumption of uniform variances across conditions when variances are strongly influenced by the stimulus.