Direct Extraction of Signal and Noise Correlations from Two-Photon Calcium Imaging of Ensemble Neuronal Activity

Anuththara Rupasinghe[1,2], **Nikolas A Francis**[3], **Ji Liu**[3], **Zac Bowen**[3], **Patrick O Kanold**[3,4], **Behtash Babadi**[1,2*]

***For correspondence:**
behtash@umd.edu (BB)

[1]Department of Electrical & Computer Engineering, University of Maryland, College Park, MD, USA; [2]The Institute for Systems Research, University of Maryland, College Park, MD, USA; [3]Department of Biology, University of Maryland, College Park, MD, USA; [4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

Abstract Neuronal activity correlations are key to understanding how populations of neurons collectively encode information. While two-photon calcium imaging has created a unique opportunity to record the activity of large populations of neurons, existing methods for inferring correlations from these data face several challenges. First, the observations of spiking activity produced by two-photon imaging are temporally blurred and noisy. Secondly, even if the spiking data were perfectly recovered via deconvolution, inferring network-level features from binary spiking data is a challenging task due to the non-linear relation of neuronal spiking to endogenous and exogenous inputs. In this work, we propose a methodology to explicitly model and directly estimate signal and noise correlations from two-photon fluorescence observations, without requiring intermediate spike deconvolution. We provide theoretical guarantees on the performance of the proposed estimator and demonstrate its utility through applications to simulated and experimentally recorded data from the mouse auditory cortex.

Introduction

Neuronal activity correlations are essential in understanding how populations of neurons encode information. Correlations provide insights into the functional architecture and computations carried out by neuronal networks (***Abbott and Dayan, 1999***; ***Averbeck et al., 2006***; ***Cohen and Kohn, 2011***; ***Hansen et al., 2012***; ***Kohn et al., 2016***; ***Kohn and Smith, 2005***; ***Lyamzin et al., 2015***; ***Montijn et al., 2014***; ***Smith and Sommer, 2013***; ***Sompolinsky et al., 2001***; ***Yatsenko et al., 2015***). Neuronal activity correlations are often categorized in two groups: *signal* correlations and *noise* correlations (***Cohen and Kohn, 2011***; ***Cohen and Maunsell, 2009***; ***Gawne and Richmond, 1993***; ***Josić et al., 2009***; ***Lyamzin et al., 2015***; ***Vinci et al., 2016***). Given two neurons, signal correlation quantifies the similarity of neural responses that are time-locked to a repeated stimulus across trials, whereas noise correlation quantifies the stimulus-independent trial-to-trial variability shared by neural responses that are believed to arise from common latent inputs.

Two-photon calcium imaging has become increasingly popular in recent years to record *in vivo* neural activity simultaneously from hundreds of neurons (***Ahrens et al., 2013***; ***Romano et al., 2017***; ***Stosiek et al., 2003***; ***Svoboda and Yasuda, 2006***). This technology takes advantage of intracellular

calcium flux mostly arising from spiking activity and captures calcium signaling in neurons in living animals using fluorescence microscopy. The observed fluorescence traces of calcium concentrations, however, are indirectly related to neuronal spiking activity. Extracting spiking activity from fluorescence traces is a challenging signal deconvolution problem, and has been the focus of active research (***Deneux et al., 2016***; ***Friedrich et al., 2017***; ***Grewe et al., 2010***; ***Jewell et al., 2020***; ***Jewell and Witten, 2018***; ***Kazemipour et al., 2018***; ***Pachitariu et al., 2018***; ***Pnevmatikakis et al., 2016***; ***Stringer and Pachitariu, 2019***; ***Theis et al., 2016***; ***Vogelstein et al., 2010***, ***2009***).

The most commonly used approach to infer signal and noise correlations from two-photon data is to directly apply the classical definitions of correlations for firing rates (***Lyamzin et al., 2015***), to fluorescence traces (***De Vico Fallani et al., 2015***; ***Francis et al., 2018***; ***Rothschild et al., 2010***; ***Winkowski and Kanold, 2013***). However, it is well known that fluorescence observations are noisy and blurred surrogates of spiking activity, because of dependence on observation noise, calcium dynamics and the temporal properties of calcium indicators. Due to temporal blurring, the resulting signal and noise correlation estimates are highly biased. An alternative approach is to carry out the inference in a two-stage fashion: first, infer spikes using a deconvolution technique, and then compute firing rates and evaluate the correlations (***Kerlin et al., 2019***; ***Najafi et al., 2020***; ***Ramesh et al., 2018***; ***Soudry et al., 2015***; ***Yatsenko et al., 2015***). These two-stage estimates are highly sensitive to the accuracy of spike deconvolution, and require high temporal resolution and signal-to-noise ratios (***Lütcke et al., 2013***; ***Pachitariu et al., 2018***). Furthermore, these deconvolution techniques are biased towards obtaining accurate first-order statistics (i.e., spike timings) via spatiotemporal priors, which may be detrimental to recovering second-order statistics (i.e., correlations). Finally, both approaches also undermine the non-linear dynamics of spiking activity as governed by stimuli, past activity and other latent processes (***Truccolo et al., 2005***). There are a few existing studies that aim at improving estimation of neuronal correlations, but they either do not consider signal correlations (***Rupasinghe and Babadi, 2020***; ***Yatsenko et al., 2015***), or aim at estimating surrogates of correlations from spikes such as the connectivity/coupling matrix (***Aitchison et al., 2017***; ***Mishchenko et al., 2011***; ***Soudry et al., 2015***; ***Keeley et al., 2020***).

Here, we propose a methodology to *directly* estimate both signal and noise correlations from two-photon imaging observations, without requiring an intermediate step of spike deconvolution. We pose the problem under the commonly used experimental paradigm in which neuronal activity is recorded during trials of a repeated stimulus. We avoid the need to perform spike deconvolution by integrating techniques from point processes and state-space modeling that explicitly relate the signal and noise correlations to the observed fluorescence traces in a multi-tier model. Thus, we cast signal and noise correlations within a parameter estimation setting. To solve the resulting estimation problem in an efficient fashion, we develop a solution method based on variational inference (***Jordan et al., 1999***; ***Blei et al., 2017***), by combining techniques from Pólya-Gamma augmentation (***Polson et al., 2013***) and compressible state-space estimation (***Rauch et al., 1965***; ***Kazemipour et al., 2018***; ***Ba et al., 2014***). We also provide theoretical guarantees on the bias and variance performance of the resulting estimator.

We demonstrate the utility of our proposed estimation framework through application to simulated and real data from the mouse auditory cortex during presentations of tones and acoustic noise. Our results corroborate existing hypotheses regarding the invariance of the noise correlation structure under spontaneous activity and stimulus-driven conditions, and its distinction from the signal correlation structure in the stimulus-driven condition (***Keeley et al., 2020***; ***Rumyantsev et al., 2020***; ***Bartolo et al., 2020***). Furthermore, while application of our proposed method to spatial analysis of signal and noise correlations in the mouse auditory cortex is consistent with existing work (***Winkowski and Kanold, 2013***), it reveals novel and distinct spatial trends in the correlation structure of layers 2/3 and 4. In summary, our method improves on existing work by: 1) joint estimation of signal and noise correlations directly from two-photon fluorescence observations without requiring intermediate spike deconvolution, 2) providing theoretical guarantees on the performance of the proposed estimator, and 3) gaining access to closed-form posterior approximations, with low-

complexity and iterative update rules and minimal dependence on training data. Our proposed method can thus be used as a robust and scalable alternative to existing approaches for extracting signal and noise correlations from two-photon imaging data.

Results

In this section we first demonstrate the utility of our proposed estimation framework through simulation studies as well as applications on experimentally-recorded data from the mouse auditory cortex. Then, we present theoretical performance bounds on the proposed estimator. Before presenting the results, we will give an overview of the proposed signal and noise correlation inference framework, and outline our contributions and their relationship to existing work. For the ease of reproducibility, we have archived a MATLAB implementation of our proposed method in GitHub (***Rupasinghe, 2020***), and have deposited the data used in this work in the Digital Repository at the University of Maryland (***Rupasinghe et al., 2021***).

Signal and Noise correlations

We consider a canonical experimental setting in which the same external stimulus, denoted by $\mathbf{s}_t$, is repeatedly presented across L independent trials and the spiking activity of a population of N neurons are indirectly measured using two-photon calcium fluorescence imaging. ***Figure 1*** (forward arrow) shows the generative model that is used to quantify this procedure. The fluorescence observation in the l^{th} trial from the j^{th} neuron at time frame t, denoted by $y_{t,l}^{(j)}$, is a noisy surrogate of the intracellular calcium concentrations. The calcium concentrations in turn are temporally blurred surrogates of the underlying spiking activity $n_{t,l}^{(j)}$, as shown in ***Figure 1***.

In modeling the spiking activity, we consider two main contributions: 1) the common known stimulus $\mathbf{s}_t$ affects the activity of the j^{th} neuron via an unknown kernel $\mathbf{d}_j$, akin to the receptive field; 2) the trial-to-trial variability and other intrinsic/extrinsic neural covariates that are not time-locked to the stimulus $\mathbf{s}_t$ are captured by a trial-dependent latent process $x_{t,l}^{(j)}$. Then, we use a Generalized Linear Model to link these underlying neural covariates to spiking activity (***Truccolo et al., 2005***).

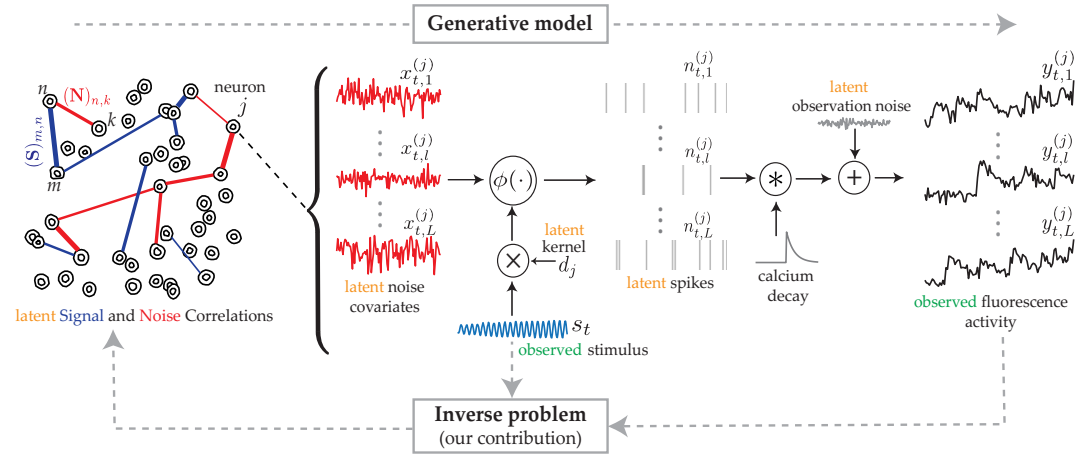


Figure 1. The proposed generative model and inverse problem. Observed (green) and latent (orange) variables pertinent to the j^{th} neuron are indicated, according to the proposed model for estimating the signal (blue) and noise (red) correlations from two-photon calcium fluorescence observations. Calcium fluorescence traces $\left(y_{t,l}^{(j)}\right)$ of L trials are observed, in which the repeated external stimulus $\left(\mathbf{s}_t\right)$ is known. The underlying spiking activity $\left(n_{t,l}^{(j)}\right)$, trial-to-trial variability and other intrinsic/extrinsic neural covariates that are not time-locked with the external stimulus $\left(x_{t,l}^{(j)}\right)$, and the stimulus kernel $\left(\mathbf{d}_j\right)$ are latent. Our main contribution is to solve the inverse problem: recovering the underlying latent signal $(\mathbf{S})$ and noise $(\mathbf{N})$ correlations directly from the fluorescence observations, without requiring intermediate spike deconvolution.

More specifically, we model spiking activity as a Bernoulli process:

$$n_{t,l}^{(j)} \sim \text{Bernoulli}\left(\phi\left(x_{t,l}^{(j)}, \mathbf{d}_j{}^{\top}\mathbf{s}_t\right)\right),$$

where $\phi(\cdot)$ is a mapping function, which could in general be non-linear.

The *signal* correlations aim to measure the correlations in the temporal response that is time-locked to the repeated stimulus, $\mathbf{s}_t$. On the other hand, *noise* correlations in our setting quantify connectivity arising from covariates that are unrelated to the stimulus, including the trial-to-trial variability (***Keeley et al., 2020***). Based on the foregoing model, we propose to formulate the signal $\left((\mathbf{\Sigma}_s)_{i,j}\right)$ and noise $\left((\mathbf{\Sigma}_x)_{i,j}\right)$ covariance between the i^{th} neuron and j^{th} neuron as:

$$(\mathbf{\Sigma}_s)_{i,j} := \mathbf{d}_i^{\top}\,\text{cov}\left(\mathbf{s}_t, \mathbf{s}_t\right)\mathbf{d}_j, \quad (\mathbf{\Sigma}_x)_{i,j} := \text{cov}\left(x_{t,l}^{(i)}, x_{t,l}^{(j)}\right), \tag{1}$$

where $\text{cov}(\cdot)$ is the empirical covariance function defined as $\text{cov}\left(\mathbf{u}_t, \mathbf{v}_t\right) := \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{u}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\right)\left(\mathbf{v}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_t\right)^{\top}$, for a total observation duration of T time frames.

Our main contribution is to provide an efficient solution for the so-called inverse problem: direct estimation of $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_x$ from the fluorescence observations, without requiring intermediate spike deconvolution (***Figure 1***, backward arrow). The signal and noise correlation matrices, denoted by $\mathbf{S}$ and $\mathbf{N}$, can then be obtained by standard normalization of $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_x$:

$$(\mathbf{S})_{i,j} := \frac{(\mathbf{\Sigma}_s)_{i,j}}{\sqrt{(\mathbf{\Sigma}_s)_{i,i}.(\mathbf{\Sigma}_s)_{j,j}}}, \quad (\mathbf{N})_{i,j} := \frac{(\mathbf{\Sigma}_x)_{i,j}}{\sqrt{(\mathbf{\Sigma}_x)_{i,i}.(\mathbf{\Sigma}_x)_{j,j}}}, \quad \forall i, j = 1, 2, \cdots, N. \tag{2}$$

We note that when spiking activity is directly observed using electrophysiology recordings, the conventional signal $\left((\mathbf{\Sigma}_s^{\text{con}})_{i,j}\right)$ and noise $\left((\mathbf{\Sigma}_x^{\text{con}})_{i,j}\right)$ covariances of spiking activity between the i^{th} and j^{th} neuron are defined as (***Lyamzin et al., 2015***):

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} := \text{cov}\left(\frac{1}{L}\sum_{l=1}^{L} n_{t,l}^{(i)},\ \frac{1}{L}\sum_{l=1}^{L} n_{t,l}^{(j)}\right), \quad (\mathbf{\Sigma}_x^{\text{con}})_{i,j} := \frac{1}{L}\sum_{l=1}^{L}\text{cov}\left(n_{t,l}^{(i)} - \frac{1}{L}\sum_{l=1}^{L} n_{t,l}^{(i)},\ n_{t,l}^{(j)} - \frac{1}{L}\sum_{l=1}^{L} n_{t,l}^{(j)}\right), \tag{3}$$

which after standard normalization in ***Equation 2*** give the conventional signal $\left((\mathbf{S}^{\text{con}})_{i,j}\right)$ and noise $\left((\mathbf{N}^{\text{con}})_{i,j}\right)$ correlations. While at first glance our definitions of signal and noise covariances in ***Equation 1*** seem to be a far departure from the conventional ones in ***Equation 3***, we show that the conventional notions of correlation indeed approximate the same quantities as in our definitions:

$$\mathbf{S}^{\text{con}} \approx \mathbf{S} \quad \text{and} \quad \mathbf{N}^{\text{con}} \approx \mathbf{N},$$

under asymptotic conditions (i.e., T and L sufficiently large). We prove this assertion of asymptotic equivalence in ***Appendix 1***, which highlights another facet of our contributions: our proposed estimators are designed to robustly operate in the regime of finite (and typically small) T and L, aiming for the very same quantities that the conventional estimators could only recover accurately under ideal asymptotic conditions.

Existing methods used for performance comparison

In order to compare the performance of our proposed method with existing work, we consider three widely available methods for extracting neuronal correlations. In simulation studies, we additionally benchmark these estimates with respect to the known ground truth. The existing methods considered are the following:

Pearson Correlations from the Two-Photon Data

In this method, fluorescence observations are assumed to be the direct measurements of spiking activity, and thus empirical Pearson correlations of the two-photon data are used to compute the signal and noise correlations (***Rothschild et al., 2010***; ***Winkowski and Kanold, 2013***; ***Francis et al., 2018***; ***Bowen et al., 2020***). Explicitly, these estimates are obtained by simply replacing $n_{t,l}^{(j)}$ in ***Equation 3*** by $y_{t,l}^{(j)}$, without performing spike deconvolution.

Two-stage Pearson Estimation

Unlike the previous method, in this case spikes are first inferred using a deconvolution technique. Then, following temporal smoothing via a narrow Gaussian kernel the Pearson correlations are computed using the conventional definitions of ***Equation 3***. For spike deconvolution, we primarily used the FCSS algorithm (***Kazemipour et al., 2018***). In order to also demonstrate the sensitivity of these estimates to the deconvolution technique that is used, we provide a comparison with the f-oopsi deconvolution algorithm (***Pnevmatikakis et al., 2016***) in ***Figure 2–Figure Supplement 1***.

Two-stage GPFA Estimation

Similar to the previous method, spikes are first inferred using a deconvolution technique. Then, a latent variable model called Gaussian Process Factor Analysis (GPFA) (***Yu et al., 2009***) is applied to the inferred spikes in order to estimate the latent covariates and receptive fields. Based on those estimates, the signal and residual noise correlations are derived through a formulation similar to ***Equation 1*** and ***Equation 2*** (***Ecker et al., 2014***).

Simulation study 1: Neuronal ensemble driven by external stimulus

We simulated calcium fluorescence observations according to the proposed generative model given in Methods and Materials, from an ensemble of $N = 8$ neurons for a duration of $T = 5000$ time frames. We considered $L = 20$ repeated trials driven by the same external stimulus, which we modeled by an autoregressive process (see Methods and Materials for details). ***Figure 2*** shows the corresponding estimation results.

The first column of ***Figure 2***-A shows the ground truth noise (top) and signal (bottom) correlations (diagonal elements are all equal to 1 and omitted for visual convenience). The second column shows estimates of the noise and signal correlations using our proposed method, which closely match the ground truth. The third, fourth and fifth columns, respectively, show the results of the Pearson correlations from the two-photon data, two-stage Pearson, and two-stage GPFA estimation methods. Through a qualitative visual inspection, it is evident that these methods incur high false alarms and mis-detections of the ground truth correlations.

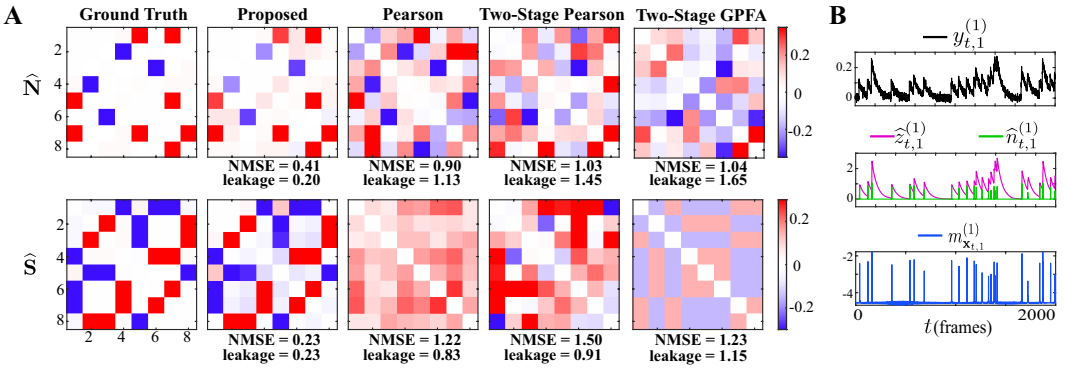


Figure 2. Results of simulation study 1. A) Estimated noise and signal correlation matrices from different methods. Rows from left to right: ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson estimates and two-stage GPFA estimates. The normalized mean squared error (NMSE) of each estimate with respect to the ground truth and the leakage effect quantified by the ratio between out-of-network and in-network power (leakage) are indicated below each panel. B) Simulated fluorescence observations (black), estimated calcium concentrations (purple), putative spikes (green) and estimated mean of the latent state (blue) by the proposed method, for the first trial of neuron 1.

Figure 2–Figure supplement 1. Sensitivity of two-stage estimates to the choice of the underlying spike deconvolution technique.

Figure 2–Figure supplement 2. Performance of two-stage estimates based on ground truth spikes.

Figure 2–Figure supplement 3. Proposed estimates based on simulated data with model mismatch and at lower SNR.

To quantify these comparisons, the normalized mean square error (NMSE) of different estimates with respect to the ground truth are shown below each of the subplots (***Figure 2***-A). Our proposed method achieves the lowest NMSE compared to the others. Furthermore, we observed a significant mixing between signal and noise correlations in these other estimates. To quantify this leakage effect, we first classified each of the correlation entries as in-network or out-of-network, based on being non-zero or zero in the ground truth, respectively (see Methods and Materials). We then computed the ratio between the power of out-of-network components and the power of in-network components as a measure of leakage. The leakage ratios are also reported in ***Figure 2***-A. The leakage of our proposed estimates is the lowest of all four techniques, in estimating both the signal and noise correlations. In order to further probe the performance of our proposed method, the simulated observations $y_{t,1}^{(1)}$, estimated calcium concentration $\hat{z}_{t,1}^{(1)}$, the putative spikes $\hat{n}_{t,1}^{(1)} := \hat{z}_{t,1}^{(1)} - \alpha\hat{z}_{t-1,1}^{(1)}$, and the estimated mean of the latent state $m_{\mathbf{x}_{t,1}}^{(1)}$, for the first trial of the first neuron are shown in ***Figure 2***-B. These results demonstrate the ability of the proposed estimation framework in accurately identifying the latent processes, which in turn leads to an accurate estimation of the signal and noise correlations as shown in ***Figure 2***-B.

The main sources of the observed performance gap between our proposed method and the existing ones are the bias incurred by treating the fluorescence traces as spikes, low spiking rates, non-linearity of spike generation with respect to intrinsic and external covariates, and sensitivity to spike deconvolution. For the latter, we demonstrated the sensitivity of the two-stage Pearson estimates to the choice of the deconvolution technique in ***Figure 2–Figure Supplement 1***. Furthermore, in order to isolate the effect of said non-linearities on the estimation performance, we applied the two-stage methods to ground truth spikes in ***Figure 2–Figure Supplement 2***. Our analysis showed that both two-stage estimates incur significant estimation errors even if the spikes were recovered perfectly, mainly due to the limited number of trials ($L = 20$ here). In accordance with our theoretical analysis of the asymptotic behavior of the conventional signal and noise correlation estimates given in ***Appendix 1***, we also showed in ***Figure 2–Figure Supplement 2*** that the performance of the two-stage Pearson estimates based on ground truth spikes, but using $L = 1000$ trials, dramatically improves. Our proposed method, however, was capable of producing reliable estimates with the number of trials as low as $L = 20$, which is typical in two-photon imaging experiments.

Finally, since real data does not necessarily follow the proposed generative model, to test the robustness of the proposed algorithm and modeling framework (with first-order autoregressive calcium dynamics assumption as outlined in Methods and Materials), we applied our method on simulated data generated based on a mismatched model (second-order autoregressive calcium dynamics), and at a lower signal-to-noise ratio (SNR) compared to the setting of ***Figure 2***. ***Figure 2–Figure Supplement 3*** shows the corresponding noise and signal correlations estimated by the proposed method under these conditions. Even though the performance slightly degrades (in terms of NMSE and leakage), our method is able to recover the underlying correlations faithfully under model mismatch and low SNR.

Simulation study 2: Spontaneous activity

Next, we present the results of a simulation study in the absence of external stimuli (i.e. $\mathbf{s}_t = \mathbf{0}$), pertaining to the spontaneous activity condition. It is noteworthy that the proposed method can readily be applied to estimate noise correlations during spontaneous activity, by simply setting the external stimulus $\mathbf{s}_t$ and the receptive field $\mathbf{d}_j$ to zero in the update rules (see Methods and Materials for details). We simulated the ensemble spiking activity based on a Poisson process (***Smith and Brown, 2003***) using a discrete time-rescaling procedure (***Brown et al., 2002***; ***Smith and Brown, 2003***), so that the data are generated using a different model than that used in our inference framework (i.e., Bernoulli process with a logistic link as outlined in Methods and Materials). As such, we eliminated potential performance biases in favor of our proposed method by introducing the aforementioned model mismatch. We simulated $L = 20$ independent trials of spontaneous activity of $N = 30$ neurons, observed for a time duration of $T = 5000$ time frames. The number of neurons

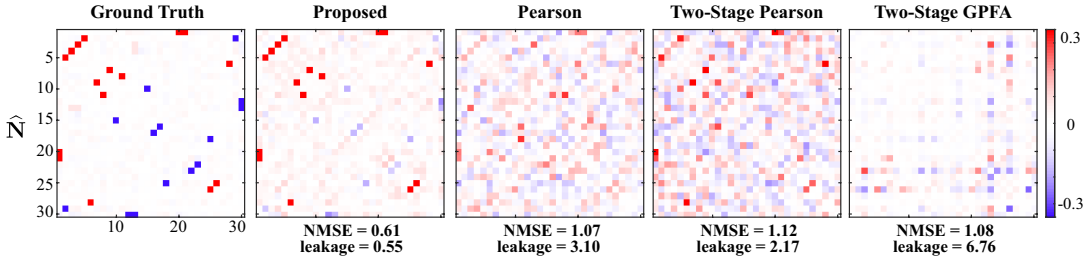


Figure 3. Results of simulation study 2. Estimated noise correlation matrices using different methods based from spontaneous activity data. Rows from left to right: ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson and two-stage GPFA estimates. The normalized mean squared error (NMSE) of each estimate with respect to the ground truth and the ratio between out-of-network power and in-network power (leakage) are shown below each panel.

in this study is notably larger than that used in the previous one, to examine the scalability of our proposed approach with respect to the ensemble size.

Figure 3 shows the comparison of the noise correlation matrices estimated by our proposed method, Pearson correlations from two-photon recordings, two-stage Pearson, and two-stage GPFA estimates, with respect to the ground truth. The Pearson and the two-stage estimates are highly variable and result in excessive false detections. Our proposed estimate, however, closely follows the ground truth, which is also reflected by the comparatively lower NMSE and leakage ratios, in spite of the mismatch between the models used for data generation and inference. It is noteworthy that the proposed method exhibits favorable scaling with respect to the ensemble size, thanks to the underlying low-complexity variational updates (see Methods and Materials).

Real data study 1: Mouse auditory cortex under random tone presentation

We next applied our proposed method to experimentally recorded two-photon observations from the mouse primary auditory cortex (A1). The dataset consisted of recordings from 371 excitatory neurons in layer 2/3 A1, from which we selected $J = 16$ neurons which exhibited the highest level of activity. A random sequence of four tones was presented to the mouse, with the same sequence being repeated for $L = 10$ trials. Each trial consisted of $T = 3600$ time frames, and each tone was two seconds long followed by a four-second silent period (see Methods and Materials for details). The comparison of the noise and signal correlation estimates obtained by our proposed method, Pearson correlations from two-photon recordings, two-stage Pearson and two-stage GPFA methods is shown in ***Figure 4***-A. The spatial map of the 16 neurons considered in the analysis in the field of view is shown in ***Figure 4***-B. ***Figure 4***-C shows the stimulus tone sequence s_t, two-photon observations $y_{t,1}^{(1)}$, estimated calcium concentration $\hat{z}_{t,1}^{(1)}$, putative spikes $\hat{n}_{t,1}^{(1)} := \hat{z}_{t,1}^{(1)} - \alpha \hat{z}_{t-1,1}^{(1)}$ and the estimated mean of the latent state $m_{\mathbf{x}_{t,1}}^{(1)}$, for the first trial of the first neuron.

We estimated the Best Frequency (BF) of each neuron as the tone that resulted in the highest level of fluorescence activity. The results in ***Figure 4***-A are organized such that the neurons with the same BF are neighboring, with the BF increasing along the diagonal. Thus, expectedly (***Bowen et al., 2020***) our proposed method as well as the Pearson and two-stage Pearson estimates show high signal correlations along the diagonal. However, the two-stage GPFA estimates do not reveal such a structure. By visual inspection, as also observed in the simulation studies, the Pearson correlations from two-photon recordings, two-stage Pearson and two-stage GPFA estimates have significant leakage between the signal and noise correlations, whereas our proposed signal and noise correlation estimates in ***Figure 4***-A suggest distinct spatial structures.

To quantify this visual comparison, we used a statistic based on the Tanimoto similarity metric (***Lipkus, 1999***), denoted by $T_s(\mathbf{X}, \mathbf{Y})$ for two matrices $\mathbf{X}$ and $\mathbf{Y}$. As a measure of dissimilarity, we used $T_d(\mathbf{X}, \mathbf{Y}) := 1 - T_s(\mathbf{X}, \mathbf{Y})$ (see Methods and Materials). The comparison of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ for the four estimates is presented in the second column of ***Table 1***. To assess statistical significance, for each

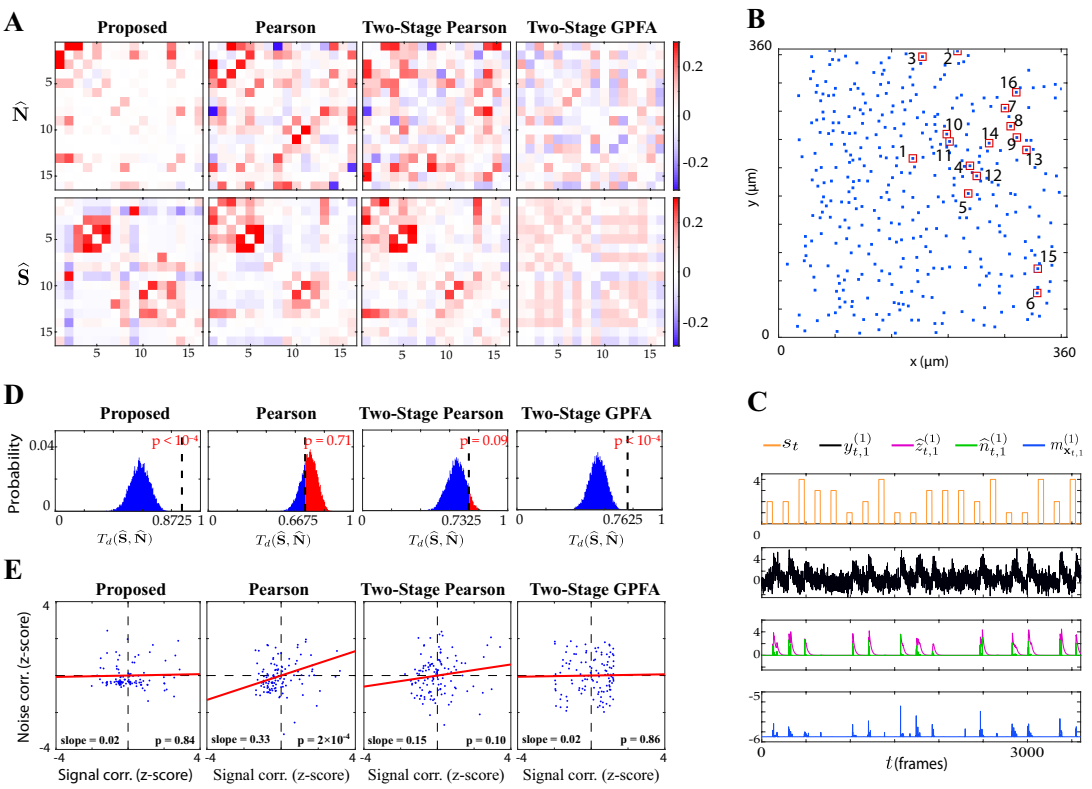


Figure 4. Application to experimentally-recorded data from the mouse A1. A) Estimated noise (top) and signal (bottom) correlation matrices using different methods. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson and two-stage GPFA estimates. B) Location of the selected neurons with the highest activity in the field of view. C) Presented tone sequence (orange), observations (black), estimated calcium concentrations (purple), putative spikes (green) and estimated mean latent state (blue) in the first trial of the first neuron. D) Null distributions of chance occurrence of dissimilarities between signal and noise correlation estimates using different methods. The observed test statistic in each case is indicated by a dashed vertical line. E) Scatter plots of signal vs. noise correlations for individual cell pairs (blue dots) corresponding to each method. Data were normalized for comparison by computing z-scores. For each case, the linear regression model fit is shown in red, and the slope and p-value of the t-test are indicated as insets.

Table 1. Dissimilarity metric statistics for the estimates in ***Figure 4***-A (also illustrated in ***Figure 4***-D), linear regression statistics of the comparison between signal and noise correlations in ***Figure 4***-E, and the average NMSE across 50 trials used in the shuffling procedure illustrated in ***Figure 5***-A.

Estimate	Dissimilarity $T_d(\widehat{\mathbf{S}}, \widehat{\mathbf{N}})$ (***Figure 4***-D)	Regression statistics (***Figure 4***-E)		Shuffling test (***Figure 5***)	
		slope (p-value)	R^2 value	NMSE in $\widehat{\mathbf{N}}$	NMSE in $\widehat{\mathbf{S}}$
Proposed	$\mathbf{0.8725}$ $(p < 10^{-4})$	0.02 $(p = 0.84)$	4×10^{-4}	$\mathbf{1.07 \pm 0.16}$	$\mathbf{1.32 \pm 0.19}$
Pearson	0.6675 $(p = 0.71)$	0.33 $(p = 2 \times 10^{-4})$	0.11	0	0
Two-Stage Pearson	0.7325 $(p = 0.09)$	0.15 $(p = 0.10)$	0.02	1.84 ± 0.34	0.55 ± 0.12
Two-Stage GPFA	0.7625 $(p < 10^{-4})$	0.02 $(p = 0.86)$	3×10^{-4}	2.32 ± 0.52	2.26 ± 0.51

comparison we obtained null distributions corresponding to chance occurrence of dissimilarities using a shuffling procedure as shown in ***Figure 4***-D, and then computed one-tailed p-values from those distributions (see Methods and Materials for details). ***Table 1*** and ***Figure 4***-D includes these p-values, which show that the proposed estimates (boldface numbers in ***Table 1***, second column) indeed have the highest dissimilarity between signal and noise correlations. The higher leakage effect in the other three estimates is also reflected in their smaller $T_d(\widehat{\mathbf{S}}, \widehat{\mathbf{N}})$ values. To further

investigate this effect, we have depicted the scatter plots of signal vs. noise correlations estimated by each method in ***Figure 4***-E. To examine the possibility of the leakage effect on a pairwise basis, we performed linear regression in each case. The slope of the model fit, the p-value for the corresponding t-test, and the R^2 values are reported in the third and fourth columns of ***Table 1*** (the slope and p-values are also shown as insets in ***Figure 4***-E). Consistent with the results of ***Winkowski and Kanold*** (***2013***), the Pearson estimates suggest a significant correlation between the signal and noise correlation pairs (as indicated by the higher slope in ***Figure 4***-E). However, it is noteworthy that none of the other estimates (including the proposed estimates) in ***Figure 4***-E register a significant trend between signal and noise correlations. This further corroborates our assessment of the high leakage between signal and noise correlations in Pearson estimates, since such a leakage effect could result in overestimation of the trend between the signal and noise correlation pairs. It is noteworthy that the signal and noise correlations estimated by our proposed method show no pairwise trend, suggesting distinct patterns of stimulus-dependent and stimulus-independent functional connectivity.

Given that the ground truth correlations are not available for a direct comparison, we instead performed a test of specificity that reveals another key limitation of existing methods. Fluorescence observations exhibit structured dynamics due to the exponential intracellular calcium concentration decay (as shown in ***Figure 4***-C, for example), which are in turn related to the underlying spikes that are driven non-linearly by intrinsic/extrinsic stimuli as well as the properties of the indicator used. As such, an accurate inference method is expected to be specific to this temporal structure. To test this, we randomly shuffled the T time frames consistently in the same order in all trials, in order to fully break the temporal structure governing calcium decay dynamics, and then estimated correlations from these shuffled data using the different methods. The resulting estimates of noise correlations are shown in ***Figure 5***-A for one instance of such shuffled data. The average NMSE for a total of 50 shuffled samples with respect to the original un-shuffled estimates (in ***Figure 4***-A) are tabulated in the fifth and sixth columns of ***Table 1***, and are also indicated below each panel in ***Figure 5***-A.

A visual inspection of ***Figure 5***-A shows that the Pearson correlations from two-photon recordings expectedly remain unchanged. Since this method treats each time frame to be independent, temporal shuffling does not impact the correlations in anyway. On the other extreme, both of the

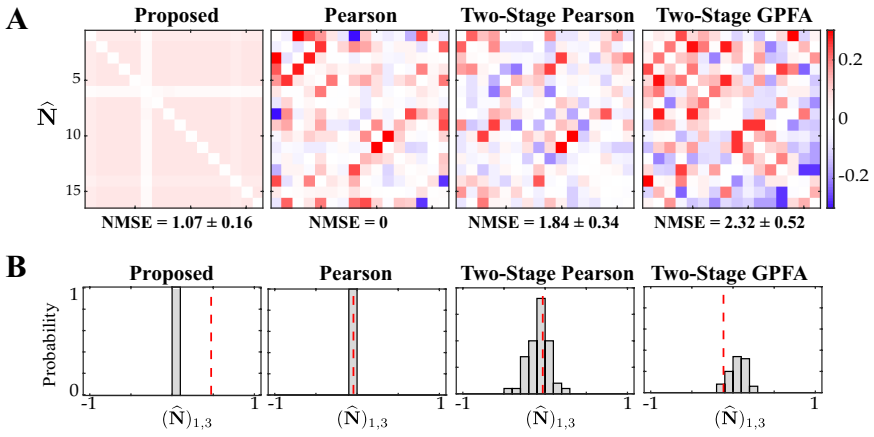


Figure 5. Assessing the specificity of different estimation results shown in ***Figure 4***. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson and two-stage GPFA estimates. A) The estimated noise correlations using different methods after random temporal shuffling of the observations. The mean and standard deviation of the NMSE across 50 trials are indicated below each panel. B) Histograms of the noise correlation estimates between the first and third neurons over the 50 temporal shuffling trials. The estimate based on the original (un-shuffled) data in each case is indicated by a dashed vertical line.

two-stage estimates seem to detect highly variable and large correlation values, despite operating on data that lacks any relevant temporal structure. Our proposed method, however, remarkably produces negligible correlation estimates. Although both the two-stage and proposed estimates show variability with respect to the shuffled data (***Table 1***), the standard deviation of the NMSE values of our proposed method are considerably smaller than those of the two-stage methods (***Table 1***, fifth column). For further inspection, the histograms of a single element ($(\widehat{\mathbf{N}})_{1,3}$) of the estimated correlation matrices across the 50 shuffling trials are shown in ***Figure 5***-B. The original un-shuffled estimates are marked by the dashed vertical lines in each case. The proposed estimate in ***Figure 5***-B is highly concentrated around zero, even though the un-shuffled estimate is non-zero. However, the two-stage estimates produce correlations that are widely variable across the shuffling trials. This analysis demonstrates that our proposed method is highly specific to the temporal structure of fluorescence observations, whereas the Pearson correlations from two-photon recordings, two-stage Pearson and two-stage GPFA methods fail to be specific.

Real data study 2: Spontaneous vs. stimulus-driven activity in the mouse A1

To further validate the utility of our proposed methodology, we applied it to another experimentally-recorded dataset from the mouse layer 2/3 A1. This experiment pertained to trials of presenting a sequence of short white noise stimuli, randomly interleaved with silent trials of the same duration. The two-photon recordings thus contained episodes of stimulus-driven and spontaneous activity (see Methods and Materials for details). Under these experimental conditions, it is expected that the noise correlations are invariant across the spontaneous and stimulus-driven conditions, and that the signal and noise correlation patterns are distinct (***Kohn et al., 2016***; ***Montijn et al., 2014***; ***Rothschild et al., 2010***; ***Keeley et al., 2020***). Each trial consisted of $T = 765$ frames. We selected $N = 10$ neurons with the highest level of activity for the analysis, each with $L = 10$ trials.

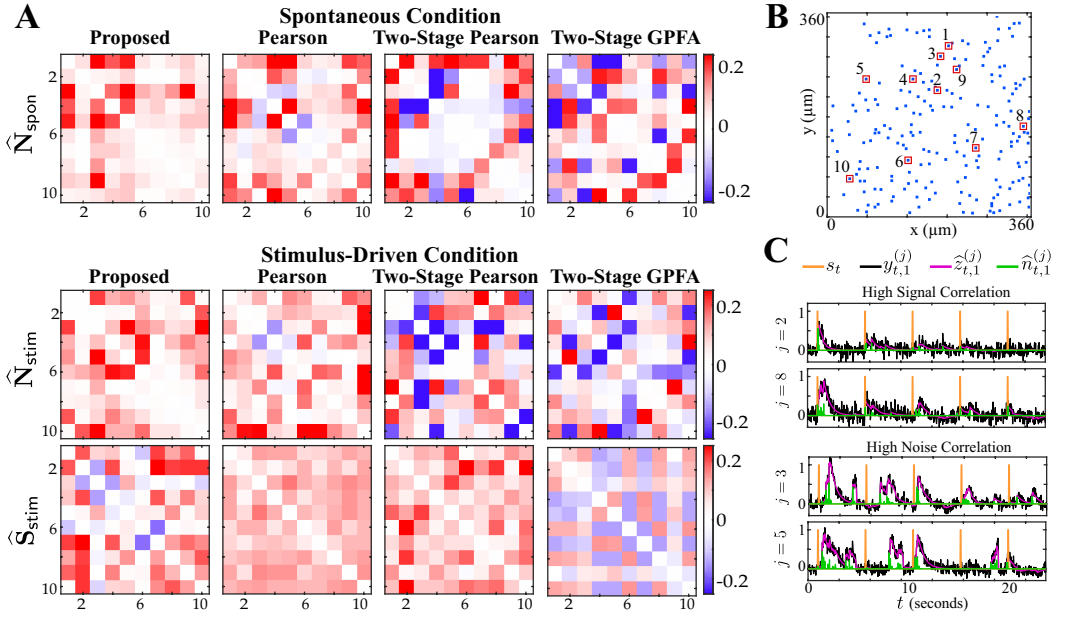


Figure 6. Comparison of spontaneous and stimulus-driven activity in the mouse A1. A) Estimated noise and signal correlation matrices under spontaneous (top) and stimulus-driven (bottom) conditions. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson and two-stage GPFA estimates. B) Location of the selected neurons with highest activity in the field of view. C) Stimulus onsets (orange), observations (black), estimated calcium concentrations (purple) and putative spikes (green) for the first trial from two pairs of neurons with high signal correlation (top) and high noise correlation (bottom), as identified by the proposed estimates.

Figure 6–Figure supplement 1. Histograms of the similarity/dissimilarity metrics under the shuffling procedure.

Table 2. Similarity/dissimilarity metric statistics for the estimates in ***Figure 6***.

Estimation Method	$T_s(\widehat{\mathbf{N}}_{\text{spon}}, \widehat{\mathbf{N}}_{\text{stim}})$	$T_d(\widehat{\mathbf{S}}_{\text{stim}}, \widehat{\mathbf{N}}_{\text{stim}})$
Proposed	**0.5518** ($p = 0.01$)	**0.7625** ($p = 0.01$)
Pearson	0.3031 ($p = 0.61$)	0.5025 ($p = 0.92$)
Two-Stage Pearson	0.2790 ($p = 0.05$)	0.7925 ($p = 0.39$)
Two-Stage GPFA	0.2008 ($p = 0.50$)	0.7825 ($p = 0.22$)

Figure 6-A shows the resulting noise and signal correlation estimates under the spontaneous ($\widehat{\mathbf{N}}_{\text{spon}}$, top) and stimulus-driven ($\widehat{\mathbf{N}}_{\text{stim}}$ and $\widehat{\mathbf{S}}_{\text{stim}}$, bottom) conditions. ***Figure 6***-B shows the spatial map of the 10 neurons considered in the analysis in the field of view. A visual inspection of the first column of ***Figure 6***-A indeed suggests that $\widehat{\mathbf{N}}_{\text{spon}}$ and $\widehat{\mathbf{N}}_{\text{stim}}$ are saliently similar, and distinct from $\widehat{\mathbf{S}}_{\text{stim}}$. The Pearson correlations obtained from two-photon data (second column) and the two-stage Pearson and GPFA estimates (third and fourth columns, respectively), however, evidently lack this structure. As in the previous study, we quantified this visual comparison using the similarity metric $T_s(\mathbf{X}, \mathbf{Y})$ and the dissimilarity metric $T_d(\mathbf{X}, \mathbf{Y})$ (see Methods and Materials for details). These statistics are reported in ***Table 2*** along with the p-values (null distributions are shown in ***Figure 6–Figure Supplement 1***), which show that the only significant outcomes (boldface numbers) are those of our proposed method.

Furthermore, ***Figure 6***-C shows the time course of the stimulus, observations, estimated calcium concentrations and putative spikes for the first trial from two pairs of neurons with high signal correlation ($j = 2, 8$, top) and high noise correlation ($j = 3, 5$, bottom). As expected, the putative spiking activity of the neurons with high signal correlation (top) are closely time-locked to the stimulus onsets. The activity of the two neurons with high noise correlation (bottom), however, is not time-locked to the stimulus onsets, even though the two neurons exhibit highly correlated activity. The correlations estimated via the proposed method thus encode substantial information about the inter-dependencies of the spiking activity of the neuronal ensemble.

Real data study 3: Spatial analysis of signal and noise correlations in the mouse A1

Lastly, we applied our proposed method to examine the spatial distribution of signal and noise correlations in the mouse A1 layers 2/3 and 4 (data from ***Bowen et al.*** (***2020***)). The dataset included fluorescence activity recorded during multiple experiments of presenting sinusoidal amplitude-modulated tones, with each stimulus being repeated across several trials (see Methods and Materials and ***Bowen et al.*** (***2020***) for experimental details). In each experiment, we selected around 20 neurons with highest spiking rates for the subsequent analysis. For brevity, we compare the estimates of signal and noise correlations using our proposed method only with those obtained by Pearson correlations from the two-photon data. The latter method was also used in previous analyses of data from this experimental paradigm (***Winkowski and Kanold, 2013***).

In parallel to the results reported in ***Winkowski and Kanold*** (***2013***), ***Figure 7***-A and ***Figure 7***-B illustrate the correlation between the signal and noise correlations in layers 2/3 and 4, respectively. Consistent with the results of ***Winkowski and Kanold*** (***2013***), the signal and noise correlations exhibit positive correlation in both layers, regardless of the method used. However, the correlation coefficients (i.e., slopes in the insets) identified by our proposed method are notably smaller than those obtained from Pearson correlations, in both layer 2/3 (***Figure 7***-A) and layer 4 (***Figure 7***-B). Comparing this result with our simulation studies suggests that the stronger linear trend between the signal and noise correlations observed using the Pearson correlation estimates is likely due to the mixing between the estimates of signal and noise correlations. As such, our method suggests that the signal and noise correlations may not be as highly correlated with one another as indicated in previous studies of layer 2/3 and 4 in mouse A1.

Next, to evaluate the spatial distribution of signal and noise correlations, we plotted the cor-

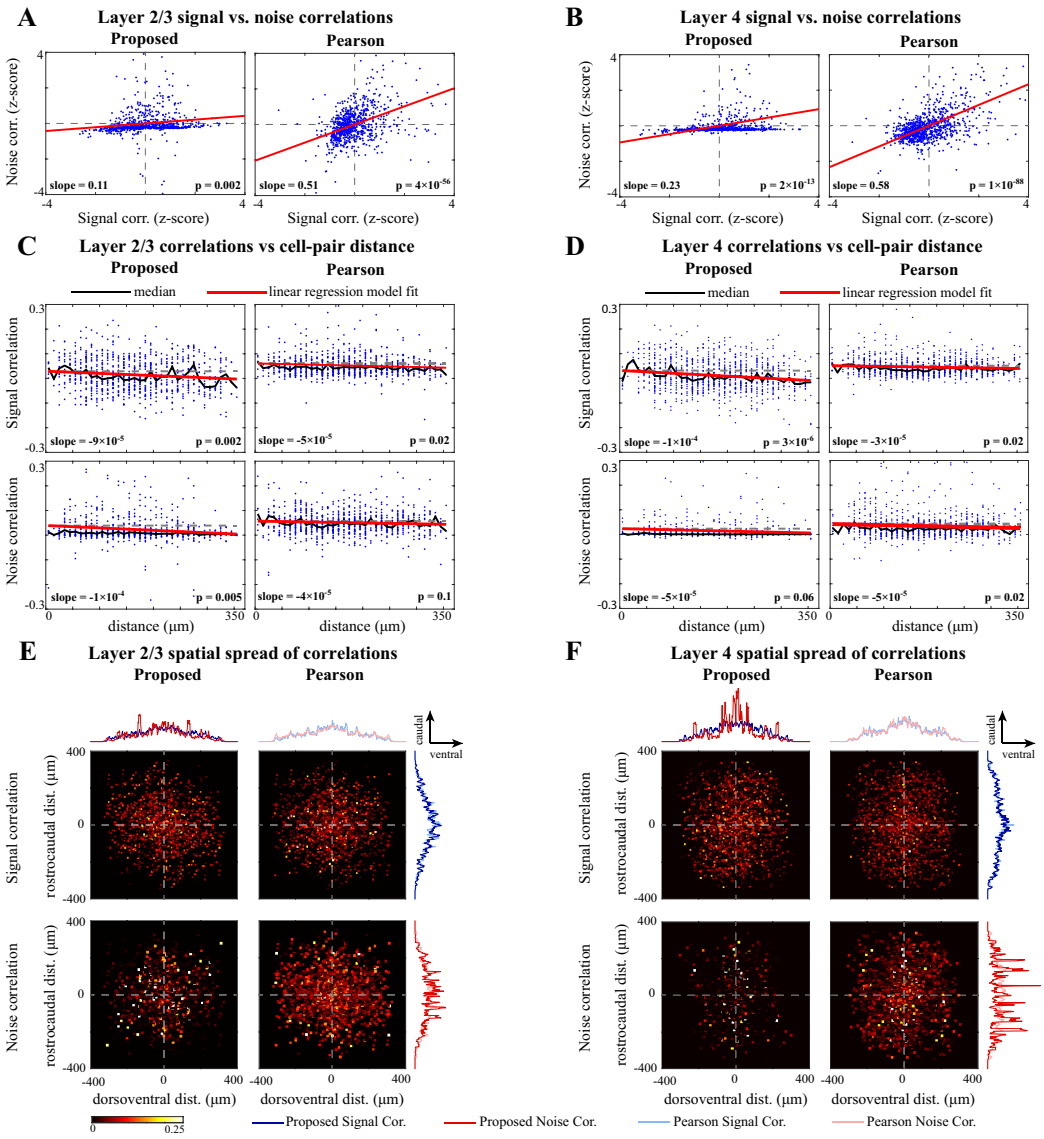


Figure 7. Comparison of signal and noise correlations across layers 2/3 and 4. A) Scatter-plot of noise vs. signal correlations (blue) for individual cell-pairs in layer 2/3, based on the proposed (left) and Pearson estimates (right). Data were normalized for comparison by computing z-scores. The linear model fits are shown in red, and the slope and p-value of the t-tests are indicated as insets. Panel B corresponds to layer 4 in the same organization as panel A. C) Signal (top) and noise (bottom) correlations vs. cell-pair distance in layer 2/3, based on the proposed (left) and Pearson estimates (right). Distances were binned to $10\ \mu m$ intervals. The median of the distributions (black) and the linear model fit (red) are shown in each panel. The slope of the linear model fit, and the p-value of the t-test are also indicated as insets. Dashed horizontal lines indicate the zero-slope line for ease of visual comparison. Panel D corresponds to layer 4 in the same organization as panel C. E) Spatial spread of signal (top) and noise (bottom) correlations in layer 2/3, based on the proposed (left) and Pearson estimates (right). The horizontal and vertical axes in each panel respectively represent the relative dorsoventral and rostrocaudal distances between each cell-pair, and the heat-map indicates the magnitude of correlations. Marginal distributions of the signal (blue) and noise (red) correlations along the dorsoventral and rostrocaudal axes for the proposed method (darker colors) and Pearson method (lighter colors) are shown at the top and right side of the sub-panels. Panel F corresponds to layer 4 in the same organization as panel E.

Figure 7–Figure supplement 1. Comparing the marginal distributions of signal and noise correlations along the dorsoventral and rostrocaudal axes.

Figure 7–Figure supplement 2. Marginal angular distributions of signal and noise correlations.

Table 3. Linear regression statistics for the analysis of correlations vs. cell-pair distance

Correlations	Statistics of layer 2/3 correlations		Statistics of layer 4 correlations	
	slope (p-value)	R^2 value	slope (p-value)	R^2 value
Proposed Signal Corr.	$\mathbf{-9 \times 10^{-5}}$ ($p = 0.002$)	0.012	$\mathbf{-1 \times 10^{-4}}$ ($p = 3 \times 10^{-6}$)	0.023
Pearson Signal Corr.	-5×10^{-5} ($p = 0.02$)	0.007	-3×10^{-5} ($p = 0.02$)	0.005
Proposed Noise Corr.	$\mathbf{-1 \times 10^{-4}}$ ($p = 0.005$)	0.010	$\mathbf{-5 \times 10^{-5}}$ ($p = 0.06$)	0.004
Pearson Noise Corr.	-4×10^{-5} ($p = 0.1$)	0.003	-5×10^{-5} ($p = 0.02$)	0.005

relation values for pairs of neurons as a function of their distance for layer 2/3 (***Figure 7***-C) and layer 4 (***Figure 7***-D). The distances were discretized using bins of length $10\ \mu m$. The scatter of the correlations along with their median at each bin are shown in all panels. Then, to examine the spatial trend of the correlations, we performed linear regression in each case. The slope of the model fit, the p-value for the corresponding t-test, and the R^2 values are reported in ***Table 3*** (the slope and p-values are also shown as insets in ***Figure 7***-C & D).

From ***Table 3*** and ***Figure 7***-C & D (upper panels), it is evident that the signal correlations show a significant negative trend with respect to distance, using both methods and in both layers. However, the slope of these negative trends identified by our method (boldface numbers in ***Table 3***) is notably steeper than those identified by Pearson correlations. On the other hand, the trends of the noise correlations with distance (bottom panels) are different between our proposed method and Pearson correlations: our proposed method shows a significant negative trend in layer 2/3, but not in layer 4, whereas the Pearson correlations of the two-photon data suggest a significant negative trend in layer 4, but not in layer 2/3. In addition, the slopes of these negative trends identified by our method (boldface numbers in ***Table 3***) are steeper than or equal to those identified by Pearson correlations.

Our proposed estimates indicate that noise correlations are sparser and less widespread in layer 4 (***Figure 7***-D) than in layer 2/3 (***Figure 7***-C). To further investigate this observation, we depicted the two-dimensional spatial spread of signal and noise correlations in both layers and for both methods in ***Figure 7***-E & F, by centering each neuron at the origin and overlaying the individual spatial spreads. The horizontal and vertical axes in each panel respectively represent the relative dorsoventral and rostrocaudal distances, and the heat-maps represent the magnitude of correlations. Comparing the proposed noise correlation spread in ***Figure 7***-E with the corresponding spread in ***Figure 7***-F, we observe that the noise correlations in layer 2/3 are indeed more widespread and abundant than in layer 4.

It is notable that the spatial spreads of signal and noise correlations based on the Pearson estimates are remarkably similar in both layers (***Figure 7***-E & F, right panels), whereas they are saliently different for our proposed estimates (***Figure 7***-E & F, left panels). This further corroborates our hypothesis on the possibility of high mixing between the signal and noise correlation estimates obtained by the Pearson correlation of two-photon data. To further examine the differences between the signal and noise correlations, the marginal distributions along the dorsoventral and rostrocaudal axes are shown in ***Figure 7***-E & F, selectively overlaid for ease of visual comparison. To quantify the differences between the spatial distributions of signal and noise correlations estimated by each method, we performed Kolmogorov–Smirnov (KS) tests on each pair of marginal distributions, which are summarized in ***Figure 7–Figure Supplement 1***. Although the marginal distributions of signal and noise correlations are significantly different in all cases from both methods, the effect sizes of their difference (KS statistics) are notably higher for our proposed estimates compared to those of the Pearson estimates.

Finally, it is noteworthy that the spatial spreads of correlations for either method and in each layer suggest non-uniform angular distributions with possibly directional bias. To test this effect, we computed the angular marginal distributions and performed KS tests for non-uniformity, which are

reported in ***Figure 7–Figure Supplement 2***. These tests indicate that all distributions are significantly non-uniform. In addition, the angular distributions of both signal and noise correlations in layer 4 exhibit salient modes in the rostrocaudal direction, whereas they are less directionally selective in layer 2/3 (***Figure 7–Figure Supplement 2***).

In summary, the spatial trends identified by our proposed method are consistent with empirical observations of spatially heterogeneous pure-tone frequency tuning by individual neurons in auditory cortex (***Winkowski and Kanold, 2013***). The improved correspondence of our proposed method compared to results obtained using Pearson correlations could be the result of the demixing of signal and noise correlations in our method. As a result of the demixing, our proposed method also suggests that noise correlations have a negative trend with distance in layer 2/3, but are much sparser and spatially flat in layer 4. In addition, the spatial spread patterns of signal and noise correlations are more structured and remarkably more distinct for our proposed method than those obtained by the Pearson estimates.

Theoretical analysis of the bias and variance of the proposed estimators

Finally, we present a theoretical analysis of the bias and variance of the proposed estimator. Note that our proposed estimation method has been developed as a scalable alternative to the intractable maximum likelihood (ML) estimation of the signal and noise covariances (see Methods and Materials). In order to benchmark our estimates, we thus need to evaluate the quality of said ML estimates. To this end, we derived bounds on the bias and variance of the ML estimators of the kernel $\mathbf{d}_j$ for $j = 1, \cdots, N$ and the noise covariance $\mathbf{\Sigma}_x$. In order to simplify the treatment, we posit the following mild assumptions:

Assumption (1). We assume a scalar time-varying external stimulus (i.e. $\mathbf{s}_t = s_t$, and hence $\mathbf{d}_j = d_j, \mathbf{d} = [d_1, d_2, \cdots, d_N]^\top$). Furthermore, we set the observation noise covariance to be $\mathbf{\Sigma}_w = \sigma_w^2 \mathbf{I}$, for notational convenience.

Assumption (2). We derive the performance bounds in the regime where T and L are large, and thus do not impose any prior distribution on the correlations, which are otherwise needed to mitigate overfitting (see Methods and Materials).

Assumption (3). We assume the latent noise process and stimulus to be slowly varying signals, and thus adopt a piece-wise constant model in which these processes are constant within consecutive windows of length W (i.e., $\mathbf{x}_{t,l} = \mathbf{x}_{W_k,l}$ and $s_t = s_{W_k}$, for $(k-1)W + 1 \leq t < kW$ and $k = 1, \cdots, K$ with $W_k = (k-1)W + 1$ and $KW = T$) for our theoretical analysis, as is usually done in spike count calculations for conventional noise correlation estimates.

Our main theoretical result is as follows:

Theorem 1 (Performance Bounds)**.** *Let $q > \frac{1}{64}$, $0 < \epsilon < 1/2$, and $0 < \eta \leq 1/2$ be fixed constants, $\sigma_m^2 := \max_i (\mathbf{\Sigma}_x)_{i,i}$ and $\sigma_s^2 := \frac{1}{K}\sum_{k=1}^K s_{W_k}^2$. Then, under Assumptions (1) - (3), the bias and variance of the maximum likelihood estimators $\widehat{\mathbf{d}}$ and $\widehat{\mathbf{\Sigma}}_x$, conditioned on an event $\mathscr{A}_W$ with $\mathbb{P}\left(\mathscr{A}_W\right) \geq 1 - \eta$ satisfy:*

$$\left|\text{bias}_{\mathscr{A}_W}\left(\widehat{d}_j\right)\right| \leq \frac{1}{\sqrt{W^{1-2\epsilon}}} C_1 \left(2\sigma_w\sqrt{1+\alpha^2}+1\right) + \tau_j,$$

$$\sqrt{\text{Var}_{\mathscr{A}_W}\left(\widehat{d}_j\right)} \leq \sqrt{\frac{(\mathbf{\Sigma}_x)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{1}{\sqrt{W^{1-2\epsilon}}} C_2 \left(2\sigma_w\sqrt{1+\alpha^2}+1\right) + \widetilde{\tau}_j,$$

$$\left|\text{bias}_{\mathscr{A}_W}\left((\widehat{\mathbf{\Sigma}}_x)_{i,j}\right)\right| \leq \frac{\left|(\mathbf{\Sigma}_x)_{i,j}\right|}{KL(1-\eta)} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_3 \left(14\sigma_w\sqrt{1+\alpha^2}+3\right) + \xi_{i,j},$$

$$\sqrt{\text{Var}_{\mathscr{A}_W}\left((\widehat{\mathbf{\Sigma}}_x)_{i,j}\right)} \leq \sqrt{\frac{(KL-1)\left((\mathbf{\Sigma}_x)_{i,j}^2 + (\mathbf{\Sigma}_x)_{i,i}(\mathbf{\Sigma}_x)_{j,j}\right)}{K^2L^2(1-\eta)}} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_4 \left(2\sigma_w\sqrt{1+\alpha^2}+1\right) + \widetilde{\xi}_{i,j},$$

for all $i, j = 1, 2, \cdots, N$, if $\log W \geq \max\left\{\frac{\log(8KLN/\eta)}{q}, \frac{32\sigma_m^2 q}{\epsilon^2}, \frac{2\log(64q)}{1-2\epsilon}, \frac{\max\{6.25,\, 4\left(\|\boldsymbol{\mu}_x\|_\infty + \max_{k,j}\{\,|s_{W_k} d_j|\,\}\right)^2\}}{8q\sigma_m^2}, \log 2\right\}$,

where τ_j and $\widetilde{\tau}_j$ denote bounded terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O}\left(\frac{1}{W}\right)$, $\xi_{i,j}$ and $\widetilde{\xi}_{i,j}$ denote bounded terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right)$ and C_1, C_2, C_3 and C_4 are bounded constants given in **Appendix 2**.

Proof. The proof of Theorem ***1*** is provided in ***Appendix 2***. □

In order to discuss the implications of this theoretical result, several remarks are in order:

Remark 1: Achieving near oracle performance

A common benchmark in estimation theory is the performance of the idealistic *oracle* estimator, in which an oracle directly observes the true latent process $\mathbf{x}_{t,l}$ and the true kernel d_j and forms the correlation estimates. In this case, the oracle would incur zero bias and variance of order $\mathcal{O}(1/KL)$ in estimating d_j, and outputs an estimate of $\mathbf{\Sigma}_x$ with bias and variance in the order of $\mathcal{O}(1/KL)$. Theorem 1 indeed states that for sufficiently large W and small σ_w, the bias and variance of the ML estimators are arbitrarily close to those of the oracle estimator. Recall that our variational inference framework is in fact a solution technique for the regularized ML problem. Hence, the bounds in Theorem 1 provide a benchmark for the expected performance of the proposed estimators, by quantifying the excess bias and variance over the performance of the oracle estimator.

Remark 2: Effect of the observation noise and observation duration

As the assumed window of stationarity $W \to \infty$ (and hence the observation duration $T \to \infty$), the loss of performance of the proposed estimators only depends on σ_w^2, the variance of the observation noise. As a result, at a given observation noise variace σ_w^2, these bounds provide a sufficient upper bound on the time duration of the observations required for attaining a desired level of estimation accuracy. It is noteworthy that σ_w^2 is typically small in practice, as it pertains to the effective observation noise and is significantly diminished by pixel averaging of the fluorescence traces following cell segmentation.

Remark 3: Effect of the number of trials

Finally, note that the bounds in Theorem 1 have terms that also drop as the number of trials L grows. These terms in fact pertain to the performance of the oracle estimator. As the number of trials grows ($L \to \infty$), the oracle estimates become arbitrarily close to the true parameters $\mathbf{\Sigma}_x$ and $\mathbf{d}_j$. Thus, our theoretical performance bounds also provide a sufficient upper bound on the number of trials L required for the oracle estimator to attain a desired level of estimation accuracy.

Discussion

We developed a novel approach for the joint estimation of signal and noise correlations of neuronal activities directly from two-photon calcium imaging observations and tested our method with experimental data. Existing widely used methods either take the fluorescence traces as surrogates of spiking activity, or first recover the unobserved spikes using deconvolution techniques, both followed by computing Pearson correlations or connectivity matrices. As such, they typically result in estimates that are highly biased and are heavily dependent on the choice of the spike deconvolution technique. We addressed these issues by using data with multiple repeated trials and explicitly relating the signal and noise covariances to the observed two-photon data via a multi-tier Bayesian model that accounts for the observation process and non-linearities involved in spiking activity. We developed an efficient estimation framework by integrating techniques from variational inference and state-space estimation. We established performance bounds on the bias and variance of the proposed estimators, which revealed favorable scaling with respect to the observation noise and trial length.

We demonstrated the utility of our proposed estimation framework on both simulated and experimentally-recorded data from the mouse auditory cortex. Our analysis showed that, unlike the aforementioned methods, our estimates provide noise correlation structures that are highly

invariant across spontaneous and stimulus-driven conditions, while producing signal correlation structures that are largely distinct from those given by the noise correlation. These results provide evidence for the involvement of distinct functional neuronal network structures in encoding the stimulus-dependent and stimulus-independent information.

Our analysis of the relationship between the signal and noise correlations in layers 2/3 and 4 in mouse A1 indicates a smaller correlation between signal and noise correlations than previously reported (***Winkowski and Kanold, 2013***). Thus, our proposed method suggests that the signal and noise correlations reflect distinct circuit mechanisms of sound processing in layers 2/3 vs 4. The spatial distribution of signal correlations obtained by our method was consistent with previous work showing significant negative trends with distance (***Winkowski and Kanold, 2013***). However, in addition, our proposed method revealed a significant negative trend of noise correlations with distance in layer 2/3, but not in layer 4, in contrast to the outcome of Pearson correlation analysis. The lack of a negative trend in layer 4 could be attributed to the sparse nature of the noise correlation spread in layer 4, as revealed by our analysis of two-dimensional spatial spreads. The latter analysis indeed revealed that the noise correlations in layer 2/3 are more widespread than those in layer 4, consistent with existing work based on whole-cell patch recordings (***Meng et al., 2017a***,b). In addition, the two-dimensional spatial spreads of signal and noise correlations obtained by our method are more distinct than those obtained by Pearson correlations. The spatial spreads also allude to directionality of the functional connectivity patterns, with a notable rostrocaudal preference in layer 4.

It is noteworthy that the proposed method can scale up favorably to larger populations of neurons, thanks to the underlying low-complexity variational updates in the inference procedure. Due to its minimal dependence on training data, our estimation framework is also applicable to single-session analysis of two-photon data with limited number of trials and duration. Another useful byproduct of the proposed framework is gaining access to approximate posterior densities in closed-form, which allows further statistical analyses such as construction of confidence intervals. Our proposed methodology can thus be used as a robust and scalable alternative to existing approaches for extracting neuronal correlations from two-photon calcium imaging data.

A potential limitation of our proposed model is the assumption that there is at most one spiking event per time frame for each neuron, in light of the fact that typical two-photon imaging frame durations are in the range of 30–100 ms. Average spike rates of excitatory neurons in mouse A1 layers 2/3 and 4 are of the order of < 10 Hz (***Forli et al., 2018***) and thus our assumption is reasonable for the current study, although it might not be optimal during bursting activity. Even though this assumption can be made more precise by adopting a Poisson model, that would render closed-form variational density updates intractable. Furthermore, in the regime of extremely low spiking rate and high observation noise, the proposed forward model may fail to capture the underlying correlations faithfully, and the performance is expected to degrade to those of existing methods based on Pearson correlations. Nevertheless, our method addresses key limitations of conventional signal and noise correlation estimators that persist in high spiking rate and high SNR conditions.

Our proposed estimation framework can be used as groundwork for incorporating other notions of correlation such as the connected correlation function (***Martin et al., 2020***), and to account for non-Gaussian and higher-order structures arising from spatiotemporal interactions (***Kadirvelu et al., 2017***; ***Yu et al., 2011***). Other possible extensions of this work include leveraging variational inference beyond the mean-field regime ***Wang and Blei*** (***2013***), extension to time-varying correlations that underlie rapid task-dependent dynamics, and extension to non-linear models such as those parameterized by neural networks (***Aitchison et al., 2017***). In the spirit of easing reproducibility, a MATLAB implementation of our proposed method as well as the data used in this work are made publicly available (***Rupasinghe, 2020***; ***Rupasinghe et al., 2021***).

Methods and Materials

Proposed forward model

Suppose we observe fluorescence traces of N neurons, for a total duration of T discrete-time frames, corresponding to L independent trials of repeated stimulus. Let $\mathbf{y}_{t,l} := [y_{t,l}^{(1)}, y_{t,l}^{(2)}, \cdots, y_{t,l}^{(N)}]^\top$, $\mathbf{z}_{t,l} := [z_{t,l}^{(1)}, z_{t,l}^{(2)}, \cdots, z_{t,l}^{(N)}]^\top$, and $\mathbf{n}_{t,l} := [n_{t,l}^{(1)}, n_{t,l}^{(2)}, \cdots, n_{t,l}^{(N)}]^\top$ be the vectors of noisy observations, intracellular calcium concentrations, and ensemble spiking activities, respectively, at trial l and frame t. We capture the dynamics of $\mathbf{y}_{t,l}$ by the following state-space model:

$$\mathbf{y}_{t,l} = \mathbf{A}\,\mathbf{z}_{t,l} + \mathbf{w}_{t,l}, \qquad \mathbf{z}_{t,l} = \alpha\,\mathbf{z}_{t-1,l} + \mathbf{n}_{t,l},$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ represents the scaling of the observations, $\mathbf{w}_{t,l}$ is zero-mean i.i.d. Gaussian noise with covariance $\mathbf{\Sigma}_w$, and $0 \le \alpha < 1$ is the state transition parameter capturing the calcium dynamics through a first order model. Note that this state-space is non-Gaussian due to the binary nature of the spiking activity, i.e., $n_{t,l}^{(j)} \in \{0, 1\}$. We model the spiking data as a point process or Generalized Linear Model with Bernoulli statistics (***Eden et al., 2004***; ***Paninski, 2004***; ***Smith and Brown, 2003***; ***Truccolo et al., 2005***):

$$n_{t,l}^{(j)} \sim \text{Bernoulli}\left(\lambda_{t,l}^{(j)}\right), \qquad \lambda_{t,l}^{(j)} = \phi\left(x_{t,l}^{(j)}, \mathbf{d}_j{}^\top \mathbf{s}_t\right)$$

where $\lambda_{t,l}^{(j)}$ is the conditional intensity function (***Truccolo et al., 2005***), which we model as a non-linear function of the known external stimulus $\mathbf{s}_t$ and the other latent intrinsic and extrinsic trial-dependent covariates, $\mathbf{x}_{t,l} := [x_{t,l}^{(1)}, x_{t,l}^{(2)}, \cdots, x_{t,l}^{(N)}]^\top$. While we assume the stimulus $\mathbf{s}_t \in \mathbb{R}^M$ to be common to all neurons, we model the distinct effect of this stimulus on the j^{th} neuron via an unknown kernel $\mathbf{d}_j \in \mathbb{R}^M$, akin to the receptive field.

The non-linear mapping of our choice is the logistic link, which is also the canonical link for a Bernoulli process in the point process and Generalized Linear Model frameworks (***Truccolo et al., 2005***). Thus, we assume:

$$\phi\left(x_{t,l}^{(j)}, \mathbf{d}_j{}^\top \mathbf{s}_t\right) = \frac{\exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)}{1 + \exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)}.$$

Finally, we assume the latent trial dependent covariates to be a Gaussian process $\mathbf{x}_{t,l} \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{\Sigma}_x)$, with mean $\boldsymbol{\mu}_x := [\mu_x^{(1)}, \mu_x^{(2)}, \cdots, \mu_x^{(N)}]^\top$ and covariance $\mathbf{\Sigma}_x$.

The probabilistic graphical model in ***Figure 8*** summarizes the main components of the aforementioned forward model. According to this forward model, the underlying noise covariance matrix that captures trial-to-trial variability can be identified as $\mathbf{\Sigma}_x$. The signal covariance matrix, representing the covariance of the neural activity arising from the repeated application of the stimulus $\mathbf{s}_t$, is given by $\mathbf{\Sigma}_s := \mathbf{D}^\top \text{cov}(\mathbf{s}_t, \mathbf{s}_t)\ \mathbf{D}$, where $\mathbf{D} := [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N] \in \mathbb{R}^{M\times N}$. The signal and noise correlation matrices, denoted by $\mathbf{S}$ and $\mathbf{N}$, can then be obtained by standard normalization of $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_x$:

$$(\mathbf{S})_{i,j} := \frac{(\mathbf{\Sigma}_s)_{i,j}}{\sqrt{(\mathbf{\Sigma}_s)_{i,i}.(\mathbf{\Sigma}_s)_{j,j}}}, \quad (\mathbf{N})_{i,j} := \frac{(\mathbf{\Sigma}_x)_{i,j}}{\sqrt{(\mathbf{\Sigma}_x)_{i,i}.(\mathbf{\Sigma}_x)_{j,j}}}, \quad \forall i, j = 1, 2, \cdots, N.$$

The main problem is thus to estimate $\{\mathbf{\Sigma}_x, \mathbf{D}\}$ from the noisy and temporally blurred data $\{\mathbf{y}_{t,l}\}_{t=1,l=1}^{T,L}$.

Overview of the proposed estimation method

First, given a limited number of trials L from an ensemble with typically low spiking rates, we need to incorporate suitable prior assumptions to avoid overfitting. Thus, we impose a prior $p_{\text{pr}}(\mathbf{\Sigma}_x)$ on the noise covariance, to compensate sparsity of data. A natural estimation method to estimate $\{\mathbf{\Sigma}_x, \mathbf{D}\}$ in a Bayesian framework is to maximize the observed data likelihood $p(\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L} | \mathbf{\Sigma}_x, \mathbf{D})$, i.e., maximum likelihood (ML). Thus, we consider the joint likelihood of the observed data and latent

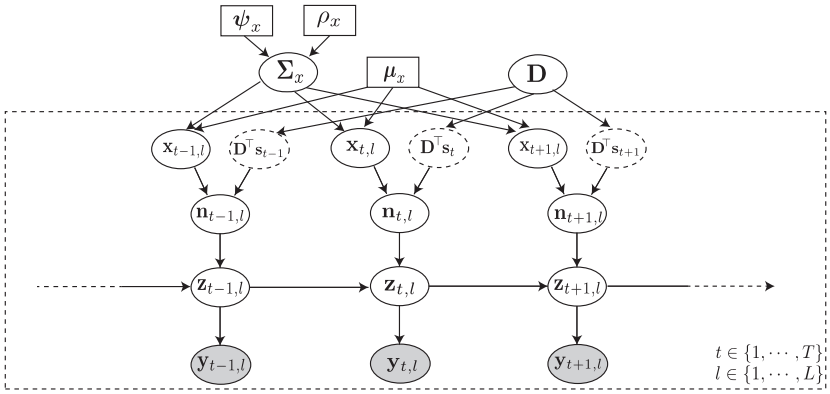


Figure 8. Probabilistic graphical model of the proposed forward model. The fluorescence observations at the t^{th} time frame and l^{th} trial: $\mathbf{y}_{t,l}$, are noisy surrogates of the intracellular calcium concentrations: $\mathbf{z}_{t,l}$. The calcium concentration at time t is a function of the spiking activity $\mathbf{n}_{t,l}$, and the calcium activity at the previous time point $\mathbf{z}_{t-1,l}$. The spiking activity is driven by two independent mechanisms: latent trial-dependent covariates $\mathbf{x}_{t,l}$, and contributions from the known external stimulus $\mathbf{s}_t$, which we model by $\mathbf{D}^\top \mathbf{s}_t$ (in which the receptive field $\mathbf{D}$ is unknown). Then, we model $\mathbf{x}_{t,l}$ as a Gaussian process with constant mean $\boldsymbol{\mu}_x$, and unknown covariance $\boldsymbol{\Sigma}_x$. Finally, we assume the covariance $\boldsymbol{\Sigma}_x$ to have an inverse Wishart prior distribution with hyper-parameters $\boldsymbol{\psi}_x$ and ρ_x. Based on this forward model, the inverse problem amounts to recovering the signal and noise correlations by directly estimating $\boldsymbol{\Sigma}_x$ and $\mathbf{D}$ (top layer) from the fluorescence observations $\{\mathbf{y}_{t,l}\}_{t=1,l=1}^{T,L}$ (bottom layer).

processes to perform Maximum a Posteriori (MAP) estimation:

$$p(\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\Sigma}_x | \mathbf{D})$$

$$= p_{\text{pr}}(\boldsymbol{\Sigma}_x) \prod_{t,l=1}^{T,L} \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}_w|}} \exp\left(-\frac{1}{2}(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l})^\top \boldsymbol{\Sigma}_w^{-1} (\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l})\right) \prod_{t,l,j=1}^{T,L,N} \frac{\left(\exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)\right)^{z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}}}{1 + \exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)} \prod_{t,l=1}^{T,L} \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}_x|}} \exp\left((\mathbf{x}_{t,l} - \boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_x^{-1} (\mathbf{x}_{t,l} - \boldsymbol{\mu}_x)\right) \tag{4}$$

Inspecting this MAP problem soon reveals that estimating $\boldsymbol{\Sigma}_x$ and $\mathbf{D}$ is a challenging task: 1) standard approaches such as Expectation-Maximization (EM) (***Shumway and Stoffer, 1982***) are intractable due to the complexity of the model, arising from the hierarchy of latent processes and the non-linearities involved in their mappings, and 2) the temporal coupling of the likelihood in the calcium concentrations makes any potential direct solver scale poorly with T.

Thus, we propose an alternative solution based on Variational Inference (VI) (***Beal, 2003***; ***Blei et al., 2017***; ***Jordan et al., 1999***). VI is a method widely used in Bayesian statistics to approximate unwieldy posterior densities using optimization techniques, as a low-complexity alternative strategy to Markov Chain Monte Carlo sampling (***Hastings, 1970***) or empirical Bayes techniques such as EM. To this end, we treat $\{\mathbf{x}_{t,l}\}_{t,l=1}^{T,L}$ and $\boldsymbol{\Sigma}_x$ as latent variables and $\{\mathbf{z}_{t,l}\}_{t,l=1}^{T,L}$ and $\mathbf{D}$ as unknown parameters to be estimated. We introduce a framework to update the latent variables and parameters sequentially, with straightforward update rules. We will describe the main ingredients of the proposed framework in the following subsections. Hereafter, we use the shorthand notations $\mathbf{y} := \{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$, $\mathbf{z} := \{\mathbf{z}_{t,l}\}_{t,l=1}^{T,L}$, and $\mathbf{x} := \{\mathbf{x}_{t,l}\}_{t,l=1}^{T,L}$.

Preliminary assumptions

For the sake of simplicity, we assume that the constants α, $\mathbf{A}$, $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\mu}_x$ are either known or can be consistently estimated from pilot trials. For example, we assume $\mathbf{A}$ to be diagonal and estimate the diagonal elements based on the magnitudes of spiking events. Further, we estimate $\boldsymbol{\mu}_x$ from the average firing rate and $\boldsymbol{\Sigma}_w$ using the background fluorescence in the absence of spiking events. Next, we take $p_{\text{pr}}(\boldsymbol{\Sigma}_x)$ to be an Inverse Wishart density:

$$\boldsymbol{\Sigma}_x \sim \text{InvWish}_N(\boldsymbol{\psi}_x, \rho_x),$$

which turns out to be the conjugate prior in our model. Thus, $\boldsymbol{\psi}_x$ and ρ_x will be the hyper-parameters of our model. It is noteworthy that although we fix α, it can be updated similar to the other two hyper-parameters for better accuracy. The details of the procedure followed for hyper-parameter tuning are given in the subsection Hyper-parameter tuning.

Decoupling via Pólya-Gamma augmentation

Direct application of VI to problems containing both discrete and continuous random variables results in intractable densities. Specifically, finding a variational distribution for $\mathbf{x}_{t,l}$ in our model with a standard distribution is not straightforward, due to the complicated posterior arising from co-dependent Bernoulli and Gaussian random variables. In order to overcome this difficulty, we employ Pólya-Gamma (PG) latent variables (***Pillow and Scott, 2012***; ***Polson et al., 2013***; ***Linderman et al., 2016***). We observe from ***Equation 4*** that the posterior density, $p(\mathbf{x}|\mathbf{z}, \mathbf{D}, \boldsymbol{\Sigma}_x)$ is conditionally independent in t, l with:

$$p(\mathbf{x}_{t,l}|\mathbf{z}, \mathbf{D}, \boldsymbol{\Sigma}_x) \propto p(\mathbf{x}_{t,l}|\boldsymbol{\Sigma}_x) \prod_{j=1}^{N} \frac{\left(\exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)\right)^{z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}}}{1 + \exp\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)}.$$

Thus, upon careful inspection we see that this density has the desired form for the PG augmentation scheme (***Polson et al., 2013***). Accordingly, we introduce a set of auxiliary PG-distributed i.i.d. latent random variables $\boldsymbol{\omega}_{t,l} := [\omega_{t,l}^{(1)}, \omega_{t,l}^{(2)}, \cdots, \omega_{t,l}^{(N)}]^\top$, $\omega_{t,l}^{(j)} \sim \mathrm{PG}(1, 0)$ for $1 \le j \le N$, $1 \le t \le T$ and $1 \le l \le L$, to derive the complete data log-likelihood:

$$\begin{aligned}
&\log p(\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x|\mathbf{D}) \\
&= -\frac{TL}{2} \log|\boldsymbol{\Sigma}_x| + \log p_{\mathsf{pr}}\left(\boldsymbol{\Sigma}_x\right) + \sum_{t,l=1}^{T,L} \Bigg\{ -\frac{1}{2}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right)^\top \boldsymbol{\Sigma}_w^{-1}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right) - \frac{1}{2}\left(\mathbf{x}_{t,l} - \boldsymbol{\mu}_x\right)^\top \boldsymbol{\Sigma}_x^{-1}\left(\mathbf{x}_{t,l} - \boldsymbol{\mu}_x\right) \\
&\quad + \sum_{j=1}^{N} \left\{ \left(z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} - \frac{1}{2}\right)\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right) - \frac{1}{2}\omega_{t,l}^{(j)}\left(x_{t,l}^{(j)} + \mathbf{d}_j{}^\top \mathbf{s}_t\right)^2 + \log p_{\mathrm{PG}(1,0)}\left(\omega_{t,l}^{(j)}\right) \right\} \Bigg\} + C, \qquad (5)
\end{aligned}$$

where $\boldsymbol{\omega} := \left\{\boldsymbol{\omega}_{t,l}\right\}_{t,l=1}^{T,L}$ and C accounts for terms not depending on $\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x$ and $\mathbf{D}$. The complete data log-likelihood is notably *quadratic* in $\mathbf{z}_{t,l}$, which as we show later admits efficient estimation procedures with favorable scaling in T.

Deriving the optimal variational densities

In this section, we will outline the procedure of applying VI to the latent variables $\mathbf{x} = \left\{\mathbf{x}_{t,l}\right\}_{t,l=1}^{T,L}$, $\boldsymbol{\omega} = \left\{\boldsymbol{\omega}_{t,l}\right\}_{t,l=1}^{T,L}$ and $\boldsymbol{\Sigma}_x$, assuming that the parameter estimates $\hat{\mathbf{z}}$ and $\hat{\mathbf{D}}$ of the previous iteration are available. The methods that we propose to update the parameters $\hat{\mathbf{z}}$ and $\hat{\mathbf{D}}$ subsequently, will be discussed in the next section.

The objective of variational inference is to posit a family of approximate densities $\mathcal{Q}$ over the latent variables, and to find the member of that family that minimizes the Kullback-Leibler (KL) divergence to the exact posterior:

$$q^*(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x|\hat{\mathbf{z}}, \hat{\mathbf{D}}) = \underset{q \in \mathcal{Q}}{\operatorname{argmin}}\ \mathrm{KL}\left(q(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x|\hat{\mathbf{z}}, \hat{\mathbf{D}}) \,\middle\|\, p(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x|\mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{D}})\right).$$

However, evaluating the KL divergence is intractable, and it has been shown (***Blei et al., 2017***) that an equivalent result to this minimization can be obtained by maximizing the alternative objective function, called the evidence lower bound (ELBO):

$$\mathrm{ELBO}\,(q) = \mathbb{E}[\log p(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x, \mathbf{y}|\hat{\mathbf{z}}, \hat{\mathbf{D}})] - \mathbb{E}[\log q(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x|\hat{\mathbf{z}}, \hat{\mathbf{D}})].$$

Further, we assume $\mathcal{Q}$ to be a mean-field variational family (***Blei et al., 2017***), resulting in the overall variational density of the form:

$$q\left(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x\right) = q\left(\boldsymbol{\Sigma}_x\right) \prod_{t,l=1}^{T,L} \left(q\left(\mathbf{x}_{t,l}\right) \prod_{j=1}^{N} q\left(\omega_{t,l}^{(j)}\right) \right). \qquad (6)$$

Under the mean field assumptions, the maximization of the ELBO can be derived using the optimization algorithm 'Coordinate Ascent Variational Inference' (CAVI) (***Bishop, 2006***; ***Blei et al., 2017***). Accordingly, we see that the optimal variational densities in ***Equation 6*** take the forms:

$$\begin{aligned}\log\, q^*\left(\mathbf{x}_{t,l}\right) &\propto \mathbb{E}_{q^*(\boldsymbol{\Sigma}_x)q^*(\boldsymbol{\omega}_{t,l})}\left[\log p\left(\mathbf{x}_{t,l}|\boldsymbol{\omega}_{t,l},\boldsymbol{\Sigma}_x,\mathbf{y},\hat{\mathbf{z}},\widehat{\mathbf{D}}\right)\right]\\ \log\, q^*\left(\omega_{t,l}^{(j)}\right) &\propto \mathbb{E}_{q^*(\mathbf{x}_{t,l})}\left[\log p\left(\omega_{t,l}^{(j)}|\mathbf{x}_{t,l},\boldsymbol{\Sigma}_x,\mathbf{y},\hat{\mathbf{z}},\widehat{\mathbf{D}}\right)\right]\\ \log\, q^*\left(\boldsymbol{\Sigma}_x\right) &\propto \mathbb{E}_{q^*(\mathbf{x})}\left[\log p\left(\boldsymbol{\Sigma}_x|\mathbf{x},\mathbf{y},\hat{\mathbf{z}},\widehat{\mathbf{D}}\right)\right]\end{aligned}$$

Upon evaluation of these expectations, we derive the optimal variational distributions as:

$$q^*(\mathbf{x}_{t,l}) \sim \mathcal{N}(\mathbf{m}_{\mathbf{x}_{t,l}}, \mathbf{Q}_{\mathbf{x}_{t,l}}), \quad q^*\left(\omega_{t,l}^{(j)}\right) \sim \mathrm{PG}\left(1, c_{t,l}^{(j)}\right), \quad q^*(\boldsymbol{\Sigma}_x) \sim \mathrm{InvWish}_N(\mathbf{P}_x, \gamma_x).$$

whose parameters $\mathbf{m}_{\mathbf{x}_{t,l}} := [m_{\mathbf{x}_{t,l}}^{(1)}, m_{\mathbf{x}_{t,l}}^{(2)}, \cdots, m_{\mathbf{x}_{t,l}}^{(N)}]^\top$, $\mathbf{Q}_{\mathbf{x}_{t,l}}$, $c_{t,l}^{(j)}$, $\mathbf{P}_x$, and γ_x can be updated given parameter estimates $\widehat{\mathbf{D}}$ and $\hat{\mathbf{z}}$:

$$\mathbf{Q}_{\mathbf{x}_{t,l}} = (\widetilde{\boldsymbol{\Omega}}_{t,l} + \gamma_x \mathbf{P}_x^{-1})^{-1}, \quad \mathbf{m}_{\mathbf{x}_{t,l}} = \mathbf{Q}_{\mathbf{x}_{t,l}}\left(\hat{\mathbf{z}}_{t,l} - \alpha\hat{\mathbf{z}}_{t-1,l} - \frac{1}{2}\mathbf{1} - \tilde{\boldsymbol{\Omega}}_{t,l}\,\widehat{\mathbf{D}}^\top\mathbf{s}_t + \gamma_x\mathbf{P}_x^{-1}\boldsymbol{\mu}_x\right),$$

$$\mathbf{P}_x := \boldsymbol{\psi}_x + \sum_{t,l=1}^{T,L}\left\{\mathbf{Q}_{\mathbf{x}_{t,l}} + \mathbf{m}_{\mathbf{x}_{t,l}}\mathbf{m}_{\mathbf{x}_{t,l}}^\top - \boldsymbol{\mu}_x\mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mathbf{m}_{\mathbf{x}_{t,l}}\boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\right\}, \quad c_{t,l}^{(j)} = \sqrt{\left(\mathbf{Q}_{\mathbf{x}_{t,l}}\right)_{j,j} + \left(m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t\right)^2},$$

and $\gamma_x := \rho_x + TL$, with $\widetilde{\boldsymbol{\Omega}}_{t,l} \in \mathbb{R}^{N\times N}$ denoting a diagonal matrix with entries $(\widetilde{\boldsymbol{\Omega}}_{t,l})_{j,j} := \frac{1}{2c_{t,l}^{(j)}}\tanh\left(\frac{c_{t,l}^{(j)}}{2}\right)$ and $\mathbf{1} \in \mathbb{R}^N$ denoting the vector of all ones.

Low-complexity parameter updates

Note that even though $\mathbf{z}$ is composed of the latent processes $\mathbf{z}_{t,l}$, we do not use VI for its inference, and instead consider it as an unknown parameter. This choice is due to the temporal dependencies arising from the underlying state-space model in ***Equation 4***, which hinders a proper assignment of variational densities under the mean field assumption. We thus seek to estimate both $\mathbf{z}$ and $\mathbf{D}$ using the updated variational density $q^*(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x)$.

First, note that the log-likelihood in ***Equation 5*** is decoupled in l, which admits independent updates to $\{\mathbf{z}_{t,l}\}_{t=1}^T$, for $l = 1, \cdots, L$. As such, given an estimate $\widehat{\mathbf{D}}$, we propose to estimate $\{\mathbf{z}_{t,l}\}_{t=1}^T$ as:

$$\begin{aligned}\{\hat{\mathbf{z}}_{t,l}\}_{t=1}^T &= \underset{\{\mathbf{z}_{t,l}\}_{t=1}^T}{\mathrm{argmax}} \quad \mathbb{E}_{q^*(\mathbf{x},\boldsymbol{\omega},\boldsymbol{\Sigma}_x)}\left[\log p\left(\mathbf{y},\mathbf{z},\mathbf{x},\boldsymbol{\omega},\boldsymbol{\Sigma}_x|\widehat{\mathbf{D}}\right)\right]\\ &= \underset{\{\mathbf{z}_{t,l}\}_{t=1}^T}{\mathrm{argmin}} \quad \sum_{t=1}^T\left\{\frac{1}{2}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right)^\top\boldsymbol{\Sigma}_w^{-1}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right) - \sum_{j=1}^N\left(m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t\right)\left(z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}\right)\right\},\end{aligned}$$

under the constraints $0 \le z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} \le 1$, for $t = 1, \cdots, T$ and $j = 1, \cdots, N$. These constraints are a direct consequence of $n_{t,l}^{(j)} = z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}$ being a Bernoulli random variable with $\mathbb{E}\left[n_{t,l}^{(j)}\right] \in [0, 1]$. While this problem is a quadratic program and can be solved using standard techniques, it is not readily decoupled in t, and thus standard solvers would not scale favorably in T.

Instead, we consider an alternative solution that admits a low-complexity recursive solution by relaxing the constraints. To this end, we relax the constraint $\mathbf{z}_{t,l} - \alpha\mathbf{z}_{t-1,l} \le \mathbf{1}$ and replace the constraint $\mathbf{z}_{t,l} - \alpha\mathbf{z}_{t-1,l} \ge \mathbf{0}$ by penalty terms proportional to $\left|z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}\right|$. The resulting relaxed problem is thus given by:

$$\min_{\{\mathbf{z}_{t,l}\}_{t=1}^T} \quad \sum_{t=1}^T\left\{\frac{1}{2}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right)^\top\boldsymbol{\Sigma}_w^{-1}\left(\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}\right) + \sum_{j=1}^N\nu_{t,l}^{(j)}\left|z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}\right|\right\}, \tag{7}$$

where $\nu_{t,l}^{(j)} := \beta\,|m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t|$ with $\beta \ge 1$ being a hyper-parameter. Given that the typical spiking rates are quite low in practice, $m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t$ is expected to be a negative number. Thus, we have assumed that $-m_{\mathbf{x}_{t,l}}^{(j)} - \widehat{\mathbf{d}}_j^\top\mathbf{s}_t = |m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t|$.

The problem of ***Equation 7*** pertains to *compressible* state-space estimation, for which fast recursive solvers are available (***Kazemipour et al., 2018***). The solver utilizes the Iteratively Re-weighted Least Squares (IRLS) (***Ba et al., 2014***) framework to transform the absolute value in the second term of the cost function into a quadratic form in $\mathbf{z}_{t,l}$, followed by Fixed Interval Smoothing (FIS) (***Rauch et al., 1965***) to find the minimizer. At iteration k, given a current estimate $\mathbf{z}^{[k-1]}$, the problem reduces to a Gaussian state-space estimation of the form:

$$\mathbf{y}_{t,l} = \mathbf{A}\mathbf{z}_{t,l} + \mathbf{w}_{t,l}, \quad \mathbf{z}_{t,l} = \alpha \mathbf{z}_{t-1,l} + \mathbf{v}_{t,l}, \tag{8}$$

with $\mathbf{w}_{t,l} \sim \mathcal{N}(0, \mathbf{\Sigma}_w)$ and $\mathbf{v}_{t,l} \sim \mathcal{N}(0, \mathbf{\Sigma}^{[k]}_{\mathbf{v}_{t,l}})$, where $\mathbf{\Sigma}^{[k]}_{\mathbf{v}_{t,l}} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $\left(\mathbf{\Sigma}^{[k]}_{\mathbf{v}_{t,l}}\right)_{j,j} := \sqrt{\left(\hat{z}^{(j)[k-1]}_{t,l} - \alpha \hat{z}^{(j)[k-1]}_{t-1,l}\right)^2 + \varepsilon^2} \Big/ \nu^{(j)}_{t,l}$, for some small constant $\varepsilon > 0$. This problem can be efficiently solved using FIS, and the iterations proceed for a total of K times or until a standard convergence criterion is met (***Kazemipour et al., 2018***). It is noteworthy that our proposed estimator of the calcium concentration $\mathbf{z}_{t,l}$ can be thought of as *soft* spike deconvolution, which naturally arises from our variational framework, as opposed to the *hard* spike deconvolution step used in two-stage estimators.

Finally, given $q^*(\mathbf{x}, \boldsymbol{\omega}, \mathbf{\Sigma}_x)$ and the updated $\hat{\mathbf{z}}$, the estimate of $\mathbf{d}_j$ for $j = 1, 2, \cdots, N$ can be updated in closed-form by maximizing the expected complete log-likelihood $\mathbb{E}_{q^*(\mathbf{x},\boldsymbol{\omega},\mathbf{\Sigma}_x)}\left[\log p(\mathbf{y}, \hat{\mathbf{z}}, \mathbf{x}, \boldsymbol{\omega}, \mathbf{\Sigma}_x | \mathbf{D})\right]$:

$$\hat{\mathbf{d}}_j = \left(\sum_{t,l=1}^{T,L} \left((\widetilde{\mathbf{\Omega}}_{t,l})_{j,j} \mathbf{s}_t \mathbf{s}_t^\top\right)\right)^{-1} \left(\sum_{t,l=1}^{T,L} \left\{ \left(\hat{z}^{(j)}_{t,l} - \alpha \hat{z}^{(j)}_{t-1,l} - \frac{1}{2}\right) \mathbf{s}_t - (\widetilde{\mathbf{\Omega}}_{t,l})_{j,j} m^{(j)}_{\mathbf{x}_{t,l}} \mathbf{s}_t \right\}\right).$$

The VI procedure iterates between updating the variational densities and parameters until convergence, upon which we estimate the noise and signal covariances as:

$$\hat{\mathbf{\Sigma}}_x := \text{mode}\{q^*(\mathbf{\Sigma}_x)\} = \frac{\mathbf{P}_x}{\gamma_x + N + 1}, \qquad \hat{\mathbf{\Sigma}}_s := \hat{\mathbf{D}}^\top \mathbb{E}[\mathbf{s}_t \mathbf{s}_t^\top] \hat{\mathbf{D}}.$$

The overall combined iterative procedure is outlined in Algorithm 1. Furthermore, a MATLAB implementation of this algorithm is publicly available in ***Rupasinghe*** (***2020***).

Model parameter settings

Simulation study 1

In the first simulation study, we set $\alpha = 0.98$, $\beta = 8$, $\mathbf{A} = 0.1\mathbf{I}$, $\boldsymbol{\mu}_x = -4.5\mathbf{1}$ and $\mathbf{\Sigma}_w = 2 \times 10^{-4}\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{8\times 8}$ represents the identity matrix and $\mathbf{1} \in \mathbb{R}^8$ represents the vector of all ones), so that the SNR of simulated data was in the same range as that of experimentally-recorded data. We used a 6^{th} order autoregressive process with a mean of -1 as the stimulus (s_t), and considered $M = 2$ lags of the stimulus (i.e., $\mathbf{s}_t = [s_t, s_{t-1}]^\top$) in the subsequent analysis.

Simulation study 2

In the second simulation study, we set $\alpha = 0.98$, $\mathbf{A} = 0.1\mathbf{I}$, $\boldsymbol{\mu}_x = -4.5\mathbf{1}$ and $\mathbf{\Sigma}_w = 10^{-4}\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{30\times 30}$ represents the identity matrix and $\mathbf{1} \in \mathbb{R}^{30}$ represents the vector of all ones) when generating the fluorescence traces $\left\{\mathbf{y}_{t,l}\right\}_{t,l=1}^{T,L}$, so that the SNR of the simulated data was in the same range as of real calcium imaging observations. Furthermore, we simulated the spike trains based on a Poisson process (***Smith and Brown, 2003***) using the discrete time re-scaling procedure (***Brown et al., 2002***; ***Smith and Brown, 2003***). Following the assumptions in ***Brown et al.*** (***2002***), we used an exponential link to simulate the observations:

$$n^{(j)}_{t,l} \sim \text{Poisson}\left(\lambda^{(j)}_{t,l}\right), \quad \lambda^{(j)}_{t,l} = \exp\left(x^{(j)}_{t,l}\right).$$

as opposed to the Bernoulli-logistic assumption in our recognition model. Then, we estimated the noise covariance $\hat{\mathbf{\Sigma}}_x$ using the Algorithm 1, with a slight modification. Since there are no external stimuli, we set $\mathbf{s}_t = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$. Accordingly, in Algorithm 1, we initialized $\hat{\mathbf{D}} = \mathbf{0}$ and did not perform the update on $\hat{\mathbf{D}}$ in the subsequent iterations.

Algorithm 1 Estimation of $\boldsymbol{\Sigma}_x$ and $\mathbf{D}$ through the proposed iterative procedure

Inputs: Ensemble of fluorescence measurements $\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$, constants $\alpha, \mathbf{A}, \boldsymbol{\Sigma}_w$ and $\boldsymbol{\mu}_x$, hyper-parameters $\boldsymbol{\psi}_x$, ρ_x, β and ϵ, tolerance at convergence δ and the external stimulus $\mathbf{s}_t$

Outputs: $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$

Initialization: Initial choice of $\boldsymbol{\Sigma}_{\mathbf{v}_{t,l}}$, $\tilde{\boldsymbol{\Omega}}_t$, $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$, residual $= 10\,\delta$, $\gamma_x = \rho_x + LT$

while residual $\geq \delta$ **do**

Estimate calcium concentrations using Fixed Interval Smoothing

  for $l = 1, \cdots, L$ **do**

  Forward filter:

    for $t = 1, \cdots, T$ **do**

$$\mathbf{z}_{(t|t-1),l} = \alpha\, \mathbf{z}_{(t-1|t-1),l}$$

$$\mathbf{P}_{(t|t-1),l} = \alpha^2\, \mathbf{P}_{(t-1|t-1),l} + \boldsymbol{\Sigma}_{\mathbf{v}_{t,l}}$$

$$\mathbf{B}_{t,l} = \mathbf{P}_{(t|t-1),l}\mathbf{A}^\top(\mathbf{A}\ \mathbf{P}_{(t|t-1),l}\ \mathbf{A}^\top + \boldsymbol{\Sigma}_w)^{-1}$$

$$\mathbf{z}_{(t|t),l} = \mathbf{z}_{(t|t-1),l} + \mathbf{B}_{t,l}(\mathbf{y}_{t,l} - \mathbf{A}\,\mathbf{z}_{(t|t-1),l})$$

$$\mathbf{P}_{(t|t),l} = (\mathbf{I} - \mathbf{B}_{t,l}\,\mathbf{A})\mathbf{P}_{(t|t-1),l}$$

    end for

  Backward smoother:

    for $t = T-1, \cdots, 1$ **do**

$$\hat{\mathbf{z}}_{t,l} = \mathbf{z}_{(t|t),l} + \alpha\mathbf{P}_{(t|t),l}\mathbf{P}^{-1}_{(t+1|t),l}(\hat{\mathbf{z}}_{t+1,l} - \mathbf{z}_{(t+1|t),l})$$

    end for

  end for

Update variational parameters

  for $t = 1, \cdots, T$ and $l = 1, \cdots, L$ **do**

$$\mathbf{Q}_{\mathbf{x}_{t,l}} = (\widetilde{\boldsymbol{\Omega}}_{t,l} + \gamma_x\mathbf{P}_x^{-1})^{-1}$$

$$\mathbf{m}_{\mathbf{x}_{t,l}} = \mathbf{Q}_{\mathbf{x}_{t,l}}\left(\hat{\mathbf{z}}_{t,l} - \alpha\hat{\mathbf{z}}_{t-1,l} - \tfrac{1}{2}\mathbf{1} - \tilde{\boldsymbol{\Omega}}_{t,l}\,\widehat{\mathbf{D}}^\top\mathbf{s}_t + \gamma_x\mathbf{P}_x^{-1}\boldsymbol{\mu}_x\right)$$

$$\nu_{t,l}^{(j)} := \beta\,|m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t|$$

    for $j = 1, \cdots, N$ **do**

$$c_{t,l}^{(j)} = \sqrt{\left(\mathbf{Q}_{\mathbf{x}_{t,l}}\right)_{j,j} + \left(m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top\mathbf{s}_t\right)^2}$$

$$(\widetilde{\boldsymbol{\Omega}}_{t,l})_{j,j} := \frac{1}{2c_{t,l}^{(j)}}\tanh\left(\frac{c_{t,l}^{(j)}}{2}\right)$$

    end for

  end for

$$\mathbf{P}_x := \boldsymbol{\psi}_x + \sum_{t,l=1}^{T,L}\{\mathbf{Q}_{\mathbf{x}_{t,l}} + \mathbf{m}_{\mathbf{x}_{t,l}}\mathbf{m}_{\mathbf{x}_{t,l}}^\top - \boldsymbol{\mu}_x\mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mathbf{m}_{\mathbf{x}_{t,l}}\boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\}$$

Update IRLS covariance approximation

  for $l = 1, \cdots, L$, $t = 1, \cdots, T$ and $j = 1, \cdots, N$ **do**

$$\left(\boldsymbol{\Sigma}_{\mathbf{v}_{t,l}}\right)_{j,j} := \sqrt{\left(\hat{z}_{t,l}^{(j)} - \alpha\hat{z}_{t-1,l}^{(j)}\right)^2 + \epsilon^2}\Big/\nu_{t,l}^{(j)}$$

  end for

Update outputs and the convergence criterion

  for $j = 1, \cdots, N$ **do**

$$\widehat{\mathbf{d}}_j = \left(\sum_{t,l=1}^{T,L}\left((\widetilde{\boldsymbol{\Omega}}_{t,l})_{j,j}\mathbf{s}_t\mathbf{s}_t^\top\right)\right)^{-1}\left(\sum_{t,l=1}^{T,L}\left\{\left(\hat{z}_{t,l}^{(j)} - \alpha\hat{z}_{t-1,l}^{(j)} - \frac{1}{2}\right)\mathbf{s}_t - (\widetilde{\boldsymbol{\Omega}}_{t,l})_{j,j}m_{\mathbf{x}_{t,l}}^{(j)}\mathbf{s}_t\right\}\right)$$

  end for

$$(\widehat{\mathbf{D}})_{\text{prev}} = \widehat{\mathbf{D}},\ \widehat{\mathbf{D}} = \left[\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \cdots, \hat{\mathbf{d}}_N\right]$$

$$(\widehat{\boldsymbol{\Sigma}}_x)_{\text{prev}} = \widehat{\boldsymbol{\Sigma}}_x,\ \widehat{\boldsymbol{\Sigma}}_x = \frac{\mathbf{P}_x}{\gamma_x + N + 1}$$

$$\text{residual} = \|(\widehat{\boldsymbol{\Sigma}}_x)_{\text{prev}} - \widehat{\boldsymbol{\Sigma}}_x\|_2/\|(\widehat{\boldsymbol{\Sigma}}_x)_{\text{prev}}\|_2 + \|(\widehat{\mathbf{D}})_{\text{prev}} - \widehat{\mathbf{D}}\|_2/\|(\widehat{\mathbf{D}})_{\text{prev}}\|_2$$

end while

Return $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$

Real data study 1

The dataset consisted of recordings from 371 excitatory neurons, from which we selected $N = 16$ neurons with high level of activity for the analysis. Each trial consisted of $T = 3600$ time frames (the sampling frequency was 30 Hz, and each trial had a duration of 120 seconds), with the presentation of a random sequence of four tones. The spiking events were very sparse and infrequent, and hence this dataset fits our model with at most one spiking event in a time frame.

We encoded the stimulus in this experiment based on the tone onsets of the four tones. Suppose that the tone onset of the p^{th} tone ($p = 1, \cdots, P$, where $P = 4$) is given by the binary sequence $\mathbf{f}_t^{(p)} \in \mathbb{R}^T$. We assumed that the response at each time t depends only on the R most recent time lags of the stimulus. For each time t, we formulated the effective stimulus corresponding to the tone p: $\mathbf{s}_t^{(p)} \in \mathbb{R}^R$, using the R recent lags of the tone onset sequence $\mathbf{f}_t^{(p)}$ starting at t. Likewise, we encoded all P tones, and then formulated the overall effective stimulus at the t^{th} time frame, $\mathbf{s}_t := [{\mathbf{s}_t^{(1)}}^\top, \cdots, {\mathbf{s}_t^{(P)}}^\top]^\top_{RP\times 1}$. Note that the weight vector $\mathbf{d}_j$ would be $M = R \times P$ dimensional under this setting. Further, based on the duration of the tones and silent periods, we considered $R = 25$ time lags in this analysis.

We set $\alpha = 0.95$ and $\mathbf{A} = \mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{16\times 16}$ represents the identity matrix), after considering the magnitude of the spiking events in observations. Further, we estimated $\boldsymbol{\mu}_x$ by a linear function of average fluorescence activity. Finally, we assumed that the observation noise covariance $\boldsymbol{\Sigma}_w$ is diagonal, and estimated the diagonal elements using the background fluorescence in the absence of spiking events.

Real data study 2

Each trial consisted of $T = 765$ frames (25.5 seconds) at a sampling frequency of 30 Hz. The layer 2/3 auditory neurons, are known to exhibit spiking rates < 5 Hz (e.g., Fig. 2-F in ***Petrus et al.*** (***2014***)), which makes the Bernoulli spiking assumption plausible for the dataset considered. Further, the auditory neurons studied here had notably low response rates (in both time and space), with only ~ 10 neurons exhibiting meaningful response. Thus, we selected $N = 10$ neurons with the highest level of activity and $L = 10$ trials for the analysis, and chose $M = 40$ lags of the stimulus in the model for the stimulus-driven condition.

We set $\alpha = 0.95$ and $\mathbf{A} = 0.75\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{10\times 10}$ represents the identity matrix), considering the magnitude of the spiking events in fluorescence observations. We used the same methods as in the first real data study to determine the optimal settings of $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_w$.

Real data study 3

Each experiment consisted of $L = 5$ trials of $P = 9$ different tone frequencies repeated at 4 different amplitude levels, resulting in each concatenated trial being ~ 180 second long (see ***Bowen et al.*** (***2020***) for more details). We used the same procedure as in the first real data study to encode this stimulus, setting the number of time lags to be $R = 25$. For each layer, we analyzed fluorescence observations from six experiments. In each experiment, we selected the most responsive $N \sim 20$ neurons for the subsequent analysis.

We set $\alpha = 0.95$, $\mathbf{A} = \mathbf{I}$ and used the same methods as in the previous two studies to determine the optimal settings of $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_w$ in each experiment.

Performance evaluation

Simulation studies

Since the ground truth is known in simulations, we directly compared the performance of each signal and noise correlation estimate with the ground truth signal and noise correlations, respectively. Suppose the ground truth correlations are given by the matrix $\mathbf{X}$ and the estimated correlations are given by the matrix $\hat{\mathbf{X}}$. To quantify the similarity between $\mathbf{X}$ and $\hat{\mathbf{X}}$, we defined the following two metrics:

Normalized Mean Squared Error (NMSE): The NMSE computes the mean squared error of $\hat{\mathbf{X}}$ with respect to $\mathbf{X}$ using the Frobenius Norm:

$$\text{NMSE} := \frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2}{\|\mathbf{X}\|_F^2}.$$

Ratio between out-of-network power and in-network power (leakage): First, we identified the in-network and out-of-network components from the ground truth correlation matrix $\mathbf{X}$. Suppose that if the true correlation between the i^{th} neuron and the j^{th} neuron is non-zero, then $\left|(\mathbf{X})_{i,j}\right| > \delta_x$, for some $\delta_x > 0$. Thus, we formed a matrix $\mathbf{X}^{\text{in}}$ that masks the in-network components, by setting $\left(\mathbf{X}^{\text{in}}\right)_{i,j} = 1$ if $\left|(\mathbf{X})_{i,j}\right| > \delta_x$ and $\left(\mathbf{X}^{\text{in}}\right)_{i,j} = 0$ if $\left|(\mathbf{X})_{i,j}\right| \leq \delta_x$. Likewise, we also formed a matrix $\mathbf{X}^{\text{out}}$ that masks the out-of-network components, by setting $\left(\mathbf{X}^{\text{out}}\right)_{i,j} = 1$ if $\left|(\mathbf{X})_{i,j}\right| \leq \delta_x$ and $\left(\mathbf{X}^{\text{out}}\right)_{i,j} = 0$ if $\left|(\mathbf{X})_{i,j}\right| > \delta_x$. Then, using these two matrices we quantified the leakage effect of $\hat{\mathbf{X}}$ comparative to $\mathbf{X}$ by:

$$\text{leakage} := \frac{\|\hat{\mathbf{X}} \cdot \mathbf{X}^{\text{out}}\|_F^2}{\|\hat{\mathbf{X}} \cdot \mathbf{X}^{\text{in}}\|_F^2},$$

where $(\cdot)$ denotes element-wise multiplication.

Real data studies

To quantify the similarity and dissimilarity between signal and noise correlation estimates, we used a statistic based on the Tanimoto similarity metric (***Lipkus, 1999***), denoted by $T_s(\mathbf{X}, \mathbf{Y})$ for two matrices $\mathbf{X}$ and $\mathbf{Y}$. For two vectors $\mathbf{a}$ and $\mathbf{b}$ with *non-negative* entries, the Tanimoto coefficient (***Lipkus, 1999***) is defined as:

$$T(\mathbf{a}, \mathbf{b}) := \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a} + \mathbf{b}^\top \mathbf{b} - \mathbf{a}^\top \mathbf{b}}.$$

The Tanimoto similarly metric between two matrices can be defined in a similar manner, by vectorizing the matrices. Thus, we formulated a similarity metric between two correlation matrices $\mathbf{X}$ and $\mathbf{Y}$ as follows. Let $\mathbf{X}^+ := \max\{\mathbf{X}, 0\mathbf{I}\}$ and $\mathbf{X}^- := \max\{-\mathbf{X}, 0\mathbf{I}\}$, with the $\max\{\cdot,\cdot\}$ operator interpreted element-wise. Note that $\mathbf{X} = \mathbf{X}^+ - \mathbf{X}^-$, and $\mathbf{X}^+, \mathbf{X}^-$ have non-negative entries. We then defined the similarity matrix by combining those of the positive and negative parts as follows:

$$T_s(\mathbf{X}, \mathbf{Y}) := \varepsilon\, T(\mathbf{X}^+, \mathbf{Y}^+) + (1 - \varepsilon)\, T(\mathbf{X}^-, \mathbf{Y}^-)$$

where $\varepsilon \in [0, 1]$ denotes the percentage of positive entries in $\mathbf{X}$ and $\mathbf{Y}$. As a measure of dissimilarity, we used $T_d(\mathbf{X}, \mathbf{Y}) := 1 - T_s(\mathbf{X}, \mathbf{Y})$. The values of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ in ***Table 1*** and $T_s(\hat{\mathbf{N}}_{\text{spon}}, \hat{\mathbf{N}}_{\text{stim}})$ and $T_d(\hat{\mathbf{S}}_{\text{stim}}, \hat{\mathbf{N}}_{\text{stim}})$ reported in ***Table 2*** were obtained based on the foregoing definitions.

To further assess the statistical significance of these results, we performed following randomized tests. To test the significance of $T_s(\hat{\mathbf{N}}_{\text{spon}}, \hat{\mathbf{N}}_{\text{stim}})$, for each comparison and each algorithm, we fixed the first matrix (i.e. $\hat{\mathbf{N}}_{\text{spon}}$) and randomly shuffled the entries of the second one ($\hat{\mathbf{N}}_{\text{stim}}$ in both cases) while respecting symmetry. We repeated this procedure for 10000 trials, to derive the null distributions that represented the probabilities of chance occurrence of similarities between two random groups of neurons.

To test the significance of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ and $T_d(\hat{\mathbf{S}}_{\text{stim}}, \hat{\mathbf{N}}_{\text{stim}})$, for each comparison and each algorithm, again we fixed the first matrix (i.e. signal correlations). Then, we formed the elements of the second matrix (akin to noise correlations) as follows. For each element of the second matrix, we assigned either the same element as the signal correlations (in order to model the leakage effect) or a random noise (with same variance as the elements in the noise correlation matrix) with equal probability. As before, we repeated this procedure for 10000 trials, to derive the null distributions that represent the probabilities of chance occurrence of dissimilarities between two matrices that have some leakage between them.

Hyper-parameter tuning

The hyper-parameters that directly affect the proposed estimation are the inverse Wishart prior hyper-parameters: $\boldsymbol{\psi}_x$ and ρ_x. Given that ρ_x appears in the form of $\gamma_x := TL + \rho_x$, we will consider $\boldsymbol{\psi}_x$ and γ_x as the main hyper-parameters for simplicity. Here, we propose a criterion for choosing these two hyper-parameters in a data-driven fashion, which will then be used to construct the estimates of the noise covariance matrix $\widehat{\boldsymbol{\Sigma}}_x$ and weight matrix $\widehat{\mathbf{D}}$. Due to the hierarchy of hidden layers in our model, an empirical Bayes approach for hyper-parameter selection using a likelihood-based performance metric is not straightforward. Hence, we propose an alternative empirical method for hyper-parameter selection as follows.

For a given choice of $\boldsymbol{\psi}_x$ and γ_x, we estimate $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$ following the proposed method. Then, based on the generative model in Proposed forward model, and using the estimated values of $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$, we sample an ensemble of simulated fluorescence traces $\hat{\mathbf{y}} = \left\{\hat{\mathbf{y}}_t^{(l)}\right\}_{t,l=1}^{T,L}$, and compute the metric $d\left(\boldsymbol{\psi}_x, \gamma_x\right)$:

$$d\left(\boldsymbol{\psi}_x, \gamma_x\right) := D_{\text{frob}}\left(\text{cov}(\hat{\mathbf{y}}, \hat{\mathbf{y}}), \text{cov}(\mathbf{y}, \mathbf{y})\right),$$

where $\text{cov}(\cdot)$ denotes the empirical covariance and $D_{\text{frob}}(\mathbf{X}, \mathbf{Y}) := \|\mathbf{X} - \mathbf{Y}\|_F^2$. Note that $D_{\text{frob}}(\mathbf{X}, \mathbf{Y})$ is strictly convex in $\mathbf{X}$. Thus, minimizing $D_{\text{frob}}(\mathbf{X}, \mathbf{Y})$ over $\mathbf{X}$ for a given $\mathbf{Y}$ has a unique solution. Accordingly, we observe that $d\left(\boldsymbol{\psi}_x, \gamma_x\right)$ is minimized when $\text{cov}(\hat{\mathbf{y}}, \hat{\mathbf{y}})$ is nearest to $\text{cov}(\mathbf{y}, \mathbf{y})$. Therefore, the corresponding estimates $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$ that generated $\hat{\mathbf{y}}$, best match the second-order statistics of $\mathbf{y}$ that was generated by the true parameters $\boldsymbol{\Sigma}_x$ and $\mathbf{D}$.

The typically low spiking rate of sensory neurons observed in practice may render the estimation problem ill-posed. It is thus important to have an accurate choice of the scale matrix $\boldsymbol{\psi}_x$ in the prior distribution. However, an exhaustive search for optimal tuning of $\boldsymbol{\psi}_x$ is not computationally feasible, given that it has $N(N+1)/2$ free variables. Thus, the main challenge here is finding the optimal choice of the scale matrix $\boldsymbol{\psi}_{x,\text{opt}}$.

To address this challenge, we propose the following method. First, we fix $\boldsymbol{\psi}_{x,\text{init}} = \tau\mathbf{I}$, where τ is a scalar and $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix. Next, given $\boldsymbol{\psi}_{x,\text{init}}$ we find the optimal choice of γ_x as:

$$\gamma_{x,\text{init}} = \underset{\gamma_x \in \mathcal{S}_\gamma}{\text{argmin}}\; d(\boldsymbol{\psi}_{x,\text{init}}, \gamma_x),$$

where $\mathcal{S}_\gamma$ is a finite set of candidate solutions for $\gamma_x > N-1$. Let $\widehat{\boldsymbol{\Sigma}}_{x,\text{init}}$ denote the noise covariance estimate corresponding to hyper-parameters $\left(\boldsymbol{\psi}_{x,\text{init}}, \gamma_{x,\text{init}}\right)$. We will next use $\widehat{\boldsymbol{\Sigma}}_{x,\text{init}}$ to find a suitable choice of $\boldsymbol{\psi}_x$. To this end, we first fix $\gamma_{x,\text{opt}} := TL + \widetilde{\rho}_x$, for some $N-1 < \widetilde{\rho}_x \ll TL$. Note that by choosing $\widetilde{\rho}_x$ to be much smaller than TL, the final estimates become less sensitive to the choice of γ_x. Then, we construct a candidate set $\mathcal{S}_\psi$ for $\boldsymbol{\psi}_{x,\text{opt}}$ by scaling $\widehat{\boldsymbol{\Sigma}}_{x,\text{init}}$ with a finite set of scalars $\eta \in \mathbb{R}^+$: $\mathcal{S}_\psi = \left\{\eta\widehat{\boldsymbol{\Sigma}}_{x,\text{init}},\ \eta \in \mathbb{R}^+\right\}$. To select $\boldsymbol{\psi}_{x,\text{opt}}$, we match it with the choice of $\gamma_{x,\text{opt}}$ by solving:

$$\boldsymbol{\psi}_{x,\text{opt}} = \underset{\boldsymbol{\psi}_x \in \mathcal{S}_\psi}{\text{argmin}}\; d\left(\boldsymbol{\psi}_x, \gamma_{x,\text{opt}}\right).$$

Finally, we use these hyper-parameters $\left(\boldsymbol{\psi}_{x,\text{opt}}, \gamma_{x,\text{opt}}\right)$ to obtain the estimators $\widehat{\boldsymbol{\Sigma}}_x$ and $\widehat{\mathbf{D}}$ as the output of the algorithm.

Experimental procedures

All procedures were approved by the University of Maryland Institutional Animal Care and Use Committee. Imaging experiments were performed on a P60 (for real data study 1) and P83 (for real data study 2) female F1 offspring of the CBA/CaJ strain (The Jackson Laboratory; stock #000654) crossed with transgenic C57BL/6J-Tg(Thy1-GCaMP6s)GP4.3Dkim/J mice (The Jackson Laboratory; stock #024275) (CBAxThy1), and F1 (CBAxC57). The third real data study was performed on data from P66-P93 and P166-P178 mice (see ***Bowen et al.*** (***2020***) for more details). We used the F1 generation of the crossed mice because they have good hearing into adulthood (***Frisina et al., 2011***).

We performed cranial window implantation and 2-photon imaging as previously described in ***Francis et al.*** (***2018***); ***Liu et al.*** (***2019***); ***Bowen et al.*** (***2019***). Briefly, we implanted a cranial window of 3 mm in diameter over the left auditory cortex. We used a scanning microscope (Bergamo II series, B248, Thorlabs) coupled to Insight X3 laser (Spectra-physics) (study 1) or pulsed femtosecond Ti:Sapphire 2-photon laser with dispersion compensation (Vision S, Coherent) (studies 2 and 3) to image GCaMP6s fluorescence from individual neurons in awake head-fixed mice with an excitation wavelengths of $\lambda = 920$ nm and $\lambda = 940$ nm, respectively. The microscope was controlled by ThorImageLS software. The size of the field of view was 370×370 μm. Imaging frames of 512×512 pixels (pixel size 0.72 μm) were acquired at 30 Hz by bidirectional scanning of an 8 kHz resonant scanner. The imaging depth was around 200 μm below pia. A circular ROI was manually drawn over the cell body to extract fluorescence traces from individual cells.

Stimuli for real data study 1

During imaging experiments, we presented 4 tones (4, 8, 16 and 32 kHz) at 70 dB SPL. The tones were 2 s in duration with an inter-trial silence of 4 s. For the sequence of tones, we first generated a randomized sequence that consisted of 5 repeats for each tone (20 tones in total) and then the same sequence was repeated for 10 trials.

Stimuli for real data study 2

During imaging experiments, we presented 97 repetitions of a 75 dB SPL 100 ms broadband noise (4–48 kHz; 8 s inter-stimulus intervals). Spontaneous neuronal activity was collected from activity during 113 randomly interleaved periods of silence (8.1 s) between 1 s long noise presentations.

Stimuli for real data study 3

During imaging experiments, sounds were played at four sound levels (20, 40, 60, and 80 dB SPL). Auditory stimuli consisted of sinusoidal amplitude-modulated (SAM) tones (20 Hz modulation, cosine phase), ranging from 3–48 kHz. The frequency resolution was 2 tones/octave (0.5 octave spacing) and each of these tonal stimuli was 1 s long, repeated five times with a 4—6 s inter-stimulus interval (see ***Bowen et al.*** (***2020***) for details).

Acknowledgments

This work is supported in part by the National Science Foundation Award No. 1807216 (BB) and the National Institutes of Health Award No. 1U19NS107464 (BB & POK). The authors would like to thank Daniel E. Winkowski for collecting the data in ***Bowen et al.*** (***2019***) that was also used in this work.

References

Abbott LF, Dayan P. The effect of correlated variability on the accuracy of a population code. Neural Computation. 1999; 11(1):91–101. https://doi.org/10.1162/089976699300016827, doi: 10.1162/089976699300016827.

Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nature methods. 2013; 10(5):413. https://doi.org/10.1038/nmeth.2434.

Aitchison L, Russell L, Packer AM, Yan J, Castonguay P, Hausser M, Turaga SC. Model-based Bayesian inference of neural activity and connectivity from all-optical interrogation of a neural circuit. In: *Advances in Neural Information Processing Systems*; 2017. p. 3486–3495. http://papers.nips.cc/paper/6940-model-based-bayesian-inference-of-neural-activity-and-connectivity-from-all-optical-interrogation-of-a-neural-circuit.pdf.

Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. Nature Reviews Neuroscience. 2006; 7(5):358–366. https://doi.org/10.1038/nrn1888.

Ba D, Babadi B, Purdon PL, Brown EN. Convergence and Stability of Iteratively Re-weighted Least Squares Algorithms. IEEE Transactions on Signal Processing. 2014 Jan; 62(1):183–195. doi: 10.1109/TSP.2013.2287685.

Bartolo R, Saunders RC, Mitz AR, Averbeck BB. Information-Limiting Correlations in Large Neural Populations. Journal of Neuroscience. 2020; 40(8):1668–1678. https://www.jneurosci.org/content/40/8/1668, doi: 10.1523/JNEUROSCI.2072-19.2019.

Beal MJ. Variational algorithms for approximate Bayesian inference. PhD thesis, University of London, University College London (United Kingdom); 2003.

Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag; 2006.

Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. Journal of the American Statistical Association. 2017; 112(518):859–877. doi: https://doi.org/10.1080/01621459.2017.1285773.

Boucheron S, Lugosi G, Massart P. Concentration Inequalities: A Nonasymptotic Theory of Independence. OUP Oxford; 2013.

Bowen Z, Winkowski DE, Kanold PO. Functional organization of mouse primary auditory cortex in adult C57BL/6 and F1 (CBAxC57) mice. Scientific Reports. 2020; 10(10905). https://doi.org/10.1038/s41598-020-67819-4.

Bowen Z, Winkowski DE, Seshadri S, Plenz D, Kanold PO. Neuronal Avalanches in Input and Associative Layers of Auditory Cortex. Frontiers in Systems Neuroscience. 2019; 13:45. https://www.frontiersin.org/article/10.3389/fnsys.2019.00045, doi: 10.3389/fnsys.2019.00045.

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The Time-rescaling Theorem and Its Application to Neural Spike Train Data Analysis. Neural Comput. 2002 Feb; 14(2):325–346. https://doi.org/10.1162/08997660252741149.

Cohen M, Kohn A. Measuring and interpreting neuronal correlations. Nature Neuroscience. 2011; 14(7):811–819. https://doi.org/10.1038/nn.2842.

Cohen MR, Maunsell JH. Attention improves performance primarily by reducing interneuronal correlations. Nature Neuroscience. 2009; 12(12):1594. doi: https://doi.org/10.1038/nn.2439.

De Vico Fallani F, Corazzol M, Sternberg JR, Wyart C, Chavez M. Hierarchy of Neural Organization in the Embryonic Spinal Cord: Granger-Causality Graph Analysis of In Vivo Calcium Imaging Data. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2015 May; 23(3):333–341. doi: 10.1109/TNSRE.2014.2341632.

Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, Rózsa B, Vanzetta I. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. Nature Communications. 2016; 7(1):1–17. doi: https://doi.org/10.1038/ncomms12190.

Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, Tolias AS. State Dependence of Noise Correlations in Macaque Primary Visual Cortex. Neuron. 2014; 82(1):235 – 248. http://www.sciencedirect.com/science/article/pii/S0896627314001044, doi: https://doi.org/10.1016/j.neuron.2014.02.006.

Eden U, Frank L, Barbieri R, Solo V, Brown E. Dynamic analysis of neural encoding by point process adaptive filtering. Neural Computation. 2004; 16(5):971–998. https://www.mitpressjournals.org/doi/abs/10.1162/089976604773135069.

Forli A, Vecchia D, Binini N, Succol F, Bovetti S, Moretti C, Nespoli F, Mahn M, Baker CA, Bolton MM, et al. Two-photon bidirectional control and imaging of neuronal excitability with high spatial resolution in vivo. Cell reports. 2018; 22(11):3087–3098. https://doi.org/10.1016/j.celrep.2018.02.063.

Francis NA, Winkowski DE, Sheikhattar A, Armengol K, Babadi B, Kanold PO. Small Networks Encode Decision-Making in Primary Auditory Cortex. Neuron. 2018; 97(4):885 – 897.e6. http://www.sciencedirect.com/science/article/pii/S0896627318300229, doi: https://doi.org/10.1016/j.neuron.2018.01.019.

Friedrich J, Zhou P, Paninski L. Fast online deconvolution of calcium imaging data. PLOS Computational Biology. 2017 03; 13(3):1–26. https://doi.org/10.1371/journal.pcbi.1005423, doi: 10.1371/journal.pcbi.1005423.

Frisina RD, Singh A, Bak M, Bozorg S, Seth R, Zhu X. F1 (CBA× C57) mice show superior hearing in old age relative to their parental strains: Hybrid vigor or a new animal model for "Golden Ears"? Neurobiology of aging. 2011; 32(9):1716–1724. http://www.sciencedirect.com/science/article/pii/S0197458009003157, doi: https://doi.org/10.1016/j.neurobiolaging.2009.09.009.

Gawne T, Richmond B. How independent are the messages carried by adjacent inferior temporal cortical neurons? Journal of Neuroscience. 1993; 13(7):2758–2771. https://www.jneurosci.org/content/13/7/2758, doi: 10.1523/JNEUROSCI.13-07-02758.1993.

Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. Nature Methods. 2010; 7(5):399. doi: https://doi.org/10.1038/nmeth.1453.

Hansen BJ, Chelaru MI, Dragoi V. Correlated variability in laminar cortical circuits. Neuron. 2012; 76(3):590–602. doi: https://doi.org/10.1016/j.neuron.2012.08.029.

Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970; 57(1):97–109. https://www.jstor.org/stable/2334940.

Jewell SW, Hocking TD, Fearnhead P, Witten DM. Fast nonconvex deconvolution of calcium imaging data. Biostatistics. 2020 October; 21. https://doi.org/10.1093/biostatistics/kxy083.

Jewell SW, Witten DM. Exact spike train inference via ℓ_0 optimization. The Annals of Applied Statistics. 2018; 12(4):2457–2482. https://projecteuclid.org/euclid.aoas/1542078052.

Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK. An introduction to variational methods for graphical models. Machine Learning. 1999; 37(2):183–233. doi: https://doi.org/10.1023/A:1007665907178.

Josić K, Shea-Brown E, Doiron B, de la Rocha J. Stimulus-dependent correlations and population codes. Neural Computation. 2009; 21(10):2774–2804. https://doi.org/10.1162/neco.2009.10-08-879, doi: 10.1162/neco.2009.10-08-879, pMID: 19635014.

Kadirvelu B, Hayashi Y, Nasuto SJ. Inferring structural connectivity using Ising couplings in models of neuronal networks. Scientific Reports. 2017 aug; 7(1). doi: https://doi.org/10.1038/s41598-017-05462-2.

Kazemipour A, Liu J, Solarana K, Nagode DA, Kanold PO, Wu M, Babadi B. Fast and Stable Signal Deconvolution via Compressible State-Space Models. IEEE Transactions on Biomedical Engineering. 2018 Jan; 65(1):74–86. doi: https://doi.org/10.1109/TBME.2017.2694339.

Keeley SL, Aoi MC, Yu Y, Smith SL, Pillow JW. Identifying signal and noise structure in neural population activity with Gaussian process factor models. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. *Advances in Neural Information Processing Systems 33* Curran Associates, Inc.; 2020.https://papers.nips.cc/paper/2020/file/9eed867b73ab1eab60583c9d4a789b1b-Paper.pdf.

Kerlin A, Mohar B, Flickinger D, MacLennan BJ, Dean MB, Davis C, Spruston N, Svoboda K. Functional clustering of dendritic activity during decision-making. eLife. 2019 oct; 8:e46966. https://doi.org/10.7554/eLife.46966, doi: 10.7554/eLife.46966.

Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A. Correlations and Neuronal Population Information. Annual Review of Neuroscience. 2016; 39(1):237–256. doi: https://doi.org/10.1146/annurev-neuro-070815-013851, pMID: 27145916.

Kohn A, Smith MA. Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. Journal of Neuroscience. 2005; 25(14):3661–3673. https://www.jneurosci.org/content/25/14/3661, doi: 10.1523/JNEUROSCI.5106-04.2005.

Linderman S, Adams RP, Pillow JW. Bayesian latent structure discovery from multi-neuron recordings. In: *Advances in Neural Information Processing Systems*; 2016. p. 2002–2010. http://papers.nips.cc/paper/6185-bayesian-latent-structure-discovery-from-multi-neuron-recordings.pdf.

Lipkus AH. A proof of the triangle inequality for the Tanimoto distance. Journal of Mathematical Chemistry. 1999 Oct; 26(1):263–265. https://doi.org/10.1023/A:1019154432472.

Liu J, Whiteway MR, Sheikhattar A, Butts DA, Babadi B, Kanold PO. Parallel Processing of Sound Dynamics across Mouse Auditory Cortex via Spatially Patterned Thalamic Inputs and Distinct Areal Intracortical Circuits. Cell Reports. 2019; 27(3):872 – 885.e7. http://www.sciencedirect.com/science/article/pii/S2211124719303997, doi: https://doi.org/10.1016/j.celrep.2019.03.069.

Lyamzin DR, Barnes SJ, Donato R, Garcia-Lazaro JA, Keck T, Lesica NA. Nonlinear Transfer of Signal and Noise Correlations in Cortical Networks. Journal of Neuroscience. 2015; 35(21):8065–8080. https://www.jneurosci.org/content/35/21/8065, doi: 10.1523/JNEUROSCI.4738-14.2015.

Lütcke H, Gerhard F, Zenke F, Gerstner W, Helmchen F. Inference of neuronal network spike dynamics and topology from calcium imaging data. Frontiers in Neural Circuits. 2013; 7:201. https://www.frontiersin.org/article/10.3389/fncir.2013.00201, doi: 10.3389/fncir.2013.00201.

Martin DA, Ribeiro TL, Cannas SA, Grigera TS, Plenz D, Chialvo DR, Box-scaling as a proxy of finite-size correlations; 2020. https://arxiv.org/abs/2007.08236.

Meng X, Kao JP, Lee HK, Kanold PO. Intracortical circuits in thalamorecipient layers of auditory cortex refine after visual deprivation. Eneuro. 2017; 4(2). https://www.eneuro.org/content/4/2/ENEURO.0092-17.2017, doi: 10.1523/ENEURO.0092-17.2017.

Meng X, Winkowski DE, Kao JP, Kanold PO. Sublaminar subdivision of mouse auditory cortex layer 2/3 based on functional translaminar connections. Journal of Neuroscience. 2017; 37(42):10200–10214. https://www.jneurosci.org/content/37/42/10200, doi: 10.1523/JNEUROSCI.1361-17.2017.

Mishchenko Y, Vogelstein J, Paninski L. A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. Annals of Applied Statistics. 2011; 5(2B):1229–1261. https://www.jstor.org/stable/23024850.

Montijn JS, Vinck M, Pennartz CMA. Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. Frontiers in Computational Neuroscience. 2014; 8:58. https://www.frontiersin.org/article/10.3389/fncom.2014.00058, doi: 10.3389/fncom.2014.00058.

Najafi F, Elsayed GF, Cao R, Pnevmatikakis E, Latham PE, Cunningham JP, Churchland AK. Excitatory and Inhibitory Subnetworks Are Equally Selective during Decision-Making and Emerge Simultaneously during Learning. Neuron. 2020; 105(1):165 – 179.e8. http://www.sciencedirect.com/science/article/pii/S0896627319308487, doi: https://doi.org/10.1016/j.neuron.2019.09.045.

Pachitariu M, Stringer C, Harris KD. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. Journal of Neuroscience. 2018; 38(37):7976–7985. https://www.jneurosci.org/content/38/37/7976, doi: 10.1523/JNEUROSCI.3339-17.2018.

Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. Network: Computation in Neural Systems. 2004; 15(4):243–262. https://doi.org/10.1088/0954-898X_15_4_002.

Petrus E, Isaiah A, Jones AP, Li D, Wang H, Lee HK, Kanold PO. Crossmodal Induction of Thalamocortical Potentiation Leads to Enhanced Information Processing in the Auditory Cortex. Neuron. 2014; 81(3):664 – 673. http://www.sciencedirect.com/science/article/pii/S089662731301091X.

Pillow JW, Scott J. Fully Bayesian inference for neural models with negative-binomial spiking. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems 25* Curran Associates, Inc.; 2012.p. 1898–1906. http://papers.nips.cc/paper/4567-fully-bayesian-inference-for-neural-models-with-negative-binomial-spiking.pdf.

Pnevmatikakis E, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W, Ahrens M, Bruno R, Jessell TM, Peterka D, Yuste R, Paninski L. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron. 2016; 89(2):285–299. doi: https://doi.org/10.1016/j.neuron.2015.11.037.

Polson NG, Scott JG, Windle J. Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. Journal of the American Statistical Association. 2013; 108(504):1339–1349. https://www.jstor.org/stable/24247065.

Ramesh RN, Burgess CR, Sugden AU, Gyetvan M, Andermann ML. Intermingled ensembles in visual association cortex encode stimulus identity or predicted outcome. Neuron. 2018; 100(4):900–915. doi: https://doi.org/10.1016/j.neuron.2018.09.024.

Rauch HE, Tung F, Striebel CT. Maximum likelihood estimates of linear dynamic systems. AIAA Journal. 1965; 3(8):1445–1450. https://doi.org/10.2514/3.3166, doi: 10.2514/3.3166.

Romano SA, Pérez-Schuster V, Jouary A, Boulanger-Weill J, Candeo A, Pietri T, Sumbre G. An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics. PLoS Computational Biology. 2017; 13(6):e1005526. doi: https://doi.org/10.1371/journal.pcbi.1005526.

Rothschild G, Nelken I, Mizrahi A. Functional organization and population dynamics in the mouse primary auditory cortex. Nature Neuroscience. 2010; 13(3):353. doi: https://doi.org/10.1038/nn.2484.

Rumyantsev OI, Lecoq JA, Hernandez O, Zhang Y, Savall J, Chrapkiewicz R, Li J, Zeng H, Ganguli S, Schnitzer MJ. Fundamental bounds on the fidelity of sensory cortical coding. Nature. 2020 April; 580(7801):100—105. https://doi.org/10.1038/s41586-020-2130-2.

Rupasinghe A, Babadi B. Robust Inference of Neuronal Correlations from Blurred and Noisy Spiking Observations. In: *2020 54th Annual Conference on Information Sciences and Systems (CISS)*; 2020. p. 1–5. https://doi.org/10.1109/CISS48834.2020.1570617409.

Rupasinghe A, Direct Extraction of Signal and Noise Correlations from Two-Photon Calcium Imaging of Ensemble Neuronal Activity MATLAB Codes. GitHub Repository; 2020. https://github.com/Anuththara-Rupasinghe/Signal-Noise-Correlation.

Rupasinghe A, Francis N, Liu J, Bowen Z, Kanold PO, Babadi B, Experimental Data from 'Direct Extraction of Signal and Noise Correlations from Two-Photon Calcium Imaging of Ensemble Neuronal Activity'. Digital Repository at the University of Maryland (DRUM); 2021. http://hdl.handle.net/1903/26917.

Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. Journal of time series analysis. 1982; 3(4):253–264. https://doi.org/10.1111/j.1467-9892.1982.tb00349.x.

Smith AC, Brown EN. Estimating a State-space Model from Point Process Observations. Neural Comput. 2003 May; 15(5):965–991. https://doi.org/10.1162/089976603765202622.

Smith MA, Sommer MA. Spatial and temporal scales of neuronal correlation in visual area V4. Journal of Neuroscience. 2013; 33(12):5422–5432. https://doi.org/10.1523/JNEUROSCI.2929-08.2008.

Sompolinsky H, Yoon H, Kang K, Shamir M. Population coding in neuronal systems with correlated noise. Phys Rev E. 2001 Oct; 64:051904. https://link.aps.org/doi/10.1103/PhysRevE.64.051904.

Soudry D, Keshri S, Stinson P, Oh Mh, Iyengar G, Paninski L. Efficient "shotgun" inference of neural connectivity from highly sub-sampled activity data. PLoS Computational Biology. 2015; 11(10). https://doi.org/10.1371/journal.pcbi.1004464.

Stosiek C, Garaschuk O, Holthoff K, Konnerth A. In vivo two-photon calcium imaging of neuronal networks. Proceedings of the National Academy of Sciences. 2003; 100(12):7319–7324. https://doi.org/10.1073/pnas.1232232100.

Stringer C, Pachitariu M. Computational processing of neural recordings from calcium imaging data. Current Opinion in Neurobiology. 2019; 55:22–31. https://doi.org/10.1016/j.conb.2018.11.005.

Svoboda K, Yasuda R. Principles of Two-Photon Excitation Microscopy and Its Applications to Neuroscience. Neuron. 2006; 50(6):823 – 839. https://doi.org/10.1016/j.neuron.2006.05.019.

Theis L, Berens P, Froudarakis E, Reimer J, Román Rosón M, Baden T, Euler T, Tolias A, Bethge M. Benchmarking Spike Rate Inference in Population Calcium Imaging. Neuron. 2016; 90(3):471–482. https://doi.org/10.1016/j.neuron.2016.04.014.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of Neurophysiology. 2005; 93(2):1074–1089. https://doi.org/10.1152/jn.00697.2004.

Vinci G, Ventura V, Smith MA, Kass RE. Separating spike count correlation from firing rate correlation. Neural computation. 2016; 28(5):849–881. https://www.mitpressjournals.org/doi/full/10.1162/NECO_a_00831, doi: 10.1162/NECO_a_00831.

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, Paninski L. Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging. Journal of Neurophysiology. 2010; 104(6):3691–3704. https://doi.org/10.1152/jn.01073.2009.

Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. Biophysical Journal. 2009; 97(2):636–655. https://doi.org/10.1016/j.bpj.2008.08.005.

Wang C, Blei DM. Variational Inference in Nonconjugate Models. J Mach Learn Res. 2013 Apr; 14(1):1005–1031. doi: https://dl.acm.org/doi/pdf/10.5555/2567709.2502613.

Winkowski DE, Kanold PO. Laminar Transformation of Frequency Organization in Auditory Cortex. Journal of Neuroscience. 2013; 33(4):1498–1508. https://www.jneurosci.org/content/33/4/1498, doi: 10.1523/JNEUROSCI.3101-12.2013.

Wong R. Asymptotic Approximations of Integrals. Society for Industrial and Applied Mathematics; 2001. https://epubs.siam.org/doi/abs/10.1137/1.9780898719260.

Yatsenko D, Josić K, Ecker AS, Froudarakis E, Cotton RJ, Tolias AS. Improved Estimation and Interpretation of Correlations in Neural Circuits. PLOS Computational Biology. 2015 03; 11(3):1–28. https://doi.org/10.1371/journal.pcbi.1004083.

Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M. Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. Journal of Neurophysiology. 2009; 102(1):614–635. https://doi.org/10.1152/jn.90941.2008.

Yu S, Yang H, Nakahara H, Santos GS, Nikolić D, Plenz D. Higher-Order Interactions Characterized in Cortical Activity. Journal of Neuroscience. 2011; 31(48):17514–17526. https://www.jneurosci.org/content/31/48/17514.

Appendix 1

Relationship to existing definitions of Signal and Noise correlations

Recall that the conventional definitions of signal and noise covariance of spiking activity between the i^{th} and j^{th} neuron are (***Lyamzin et al., 2015***):

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} = \text{cov}\left(\frac{1}{L}\sum_l n_{t,l}^{(i)}, \frac{1}{L}\sum_l n_{t,l}^{(j)}\right),$$

$$(\mathbf{\Sigma}_x^{\text{con}})_{i,j} = \frac{1}{L}\sum_l \text{cov}\left(n_{t,l}^{(i)} - \frac{1}{L}\sum_l n_{t,l}^{(i)}, n_{t,l}^{(j)} - \frac{1}{L}\sum_l n_{t,l}^{(j)}\right),$$

where $\text{cov}\left(\mathbf{u}_t, \mathbf{v}_t\right) := \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{u}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\right)\left(\mathbf{v}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_t\right)^{\top}$, is the empirical covariance. The correlations, are then derived by the standard normalization:

$$(\mathbf{S}^{\text{con}})_{i,j} := \frac{(\mathbf{\Sigma}_s^{\text{con}})_{i,j}}{\sqrt{(\mathbf{\Sigma}_s^{\text{con}})_{i,i}.(\mathbf{\Sigma}_s^{\text{con}})_{j,j}}}, \quad (\mathbf{N}^{\text{con}})_{i,j} := \frac{(\mathbf{\Sigma}_x^{\text{con}})_{i,j}}{\sqrt{(\mathbf{\Sigma}_x^{\text{con}})_{i,i}.(\mathbf{\Sigma}_x^{\text{con}})_{j,j}}}, \quad \forall i,j = 1,2,\cdots,N. \tag{9}$$

Suppose that the spiking events follow the forward model:

$$n_{t,l}^{(j)} \sim \text{Bernoulli}\left(\lambda_{t,l}^{(j)}\right),$$

$$\lambda_{t,l}^{(j)} = \phi\left(x_{t,l}^{(j)}, \mathbf{d}_j{}^{\top}\mathbf{s}_t\right),$$

where $\phi : \mathbb{R}^2 \to [0,1]$ is a differentiable non-linear mapping. We assume $\mathbf{x}_{t,l}$ and $\mathbf{s}_t$ to be independent. Without loss of generality, let $\mathbb{E}\left[\mathbf{s}_t\right] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{x}_{t,l}\right] = \boldsymbol{\mu}_x$. Further, we define the notation $X_t \approx Y_t$ to denote almost sure equivalence, i.e., $X_t \xrightarrow{\text{a.s.}} Z$ and $Y_t \xrightarrow{\text{a.s.}} Z$ for some random variable Z.

First, let us consider $(\mathbf{S}^{\text{con}})_{i,j}$. Noting that $\mathbb{E}\left[n_{t,l}^{(j)}\right] = \mathbb{E}\left[\lambda_{t,l}^{(j)}\right]$ and $\mathbb{E}\left[n_{t,l}^{(i)} n_{t,l}^{(j)}\right] = \mathbb{E}\left[\lambda_{t,l}^{(i)}\lambda_{t,l}^{(j)}\right]$, we conclude as $T \to \infty$:

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov}\left(\frac{1}{L}\sum_l \lambda_{t,l}^{(i)}, \frac{1}{L}\sum_l \lambda_{t,l}^{(j)}\right),$$

from the law of large numbers. Then, if we consider the Taylor series expansion of $\phi\left(x_{t,l}^{(j)}, \mathbf{d}_j{}^{\top}\mathbf{s}_t\right)$ around the mean $(\mu_x^{(j)}, 0)$, we get:

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov}\Bigg(\phi\left(\mu_x^{(i)}, 0\right) + \frac{1}{L}\sum_l \left(x_{t,l}^{(i)} - \mu_x^{(i)}\right)\phi_{\left(x_{t,l}^{(i)}\right)}\left(\mu_x^{(i)}, 0\right) + \left(\mathbf{d}_i{}^{\top}\mathbf{s}_t\right)\phi_{\left(\mathbf{d}_i{}^{\top}\mathbf{s}_t\right)}\left(\mu_x^{(i)}, 0\right) + \epsilon_{t,l}^{(i)},$$

$$\phi\left(\mu_x^{(j)}, 0\right) + \frac{1}{L}\sum_l \left(x_{t,l}^{(j)} - \mu_x^{(j)}\right)\phi_{\left(x_{t,l}^{(j)}\right)}\left(\mu_x^{(j)}, 0\right) + \left(\mathbf{d}_j{}^{\top}\mathbf{s}_t\right)\phi_{\left(\mathbf{d}_j{}^{\top}\mathbf{s}_t\right)}\left(\mu_x^{(j)}, 0\right) + \epsilon_{t,l}^{(j)}\Bigg),$$

where $\epsilon_{t,l}^{(i)}$ and $\epsilon_{t,l}^{(j)}$ represent the higher order terms. Then, as $L \to \infty$, we get:

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov}\left(\left(\mathbf{d}_i{}^{\top}\mathbf{s}_t\right)\phi_{\left(\mathbf{d}_i{}^{\top}\mathbf{s}_t\right)}\left(\mu_x^{(i)}, 0\right), \left(\mathbf{d}_j{}^{\top}\mathbf{s}_t\right)\phi_{\left(\mathbf{d}_j{}^{\top}\mathbf{s}_t\right)}\left(\mu_x^{(j)}, 0\right)\right) + \epsilon_{t,l},$$

since $\lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L}\left(x_{t,l}^{(j)}\right) = \mu_x^{(j)}$ by the Law of Large numbers. Thus, we see that:

$$(\mathbf{\Sigma}_s^{\text{con}})_{i,j} \approx C_S \mathbf{d}_i{}^{\top}\,\text{cov}\left(\mathbf{s}_t, \mathbf{s}_t\right)\mathbf{d}_j{}^{\top} + \epsilon_{t,l}$$
$$= C_S(\mathbf{\Sigma}_s)_{i,j} + \epsilon_{t,l},$$

where C_S is a constant and $\epsilon_{t,l}$ is typically small if the latent process $\mathbf{x}_{t,l}$ and the stimulus $\mathbf{s}_t$ are concentrated around their means. Then, the signal correlations are obtained by

normalization of the signal covariance as in ***Equation 9***, through which the scaling factor C_S cancels and we get:

$$(\mathbf{S}^{\text{con}})_{i,j} \quad \approx \quad (\mathbf{S})_{i,j}.$$

Thus, as $T, L \to \infty$, we see that $\mathbf{S}$ is indeed the signal correlation matrix that is aimed to be approximated by the conventional definitions.

Next, let us consider $(\mathbf{N}^{\text{con}})_{i,j}$. Similar to foregoing analysis of the signal covariance, as $T \to \infty$ we get:

$$(\mathbf{\Sigma}_x^{\text{con}})_{i,j} \approx \frac{1}{L}\sum_l \text{cov}\left(\lambda_{t,l}^{(i)} - \frac{1}{L}\sum_l \lambda_{t,l}^{(i)}, \lambda_{t,l}^{(j)} - \frac{1}{L}\sum_l \lambda_{t,l}^{(j)}\right).$$

Then, from a Taylor series expansion, we get:

$$(\mathbf{\Sigma}_x^{\text{con}})_{i,j} \approx \frac{1}{L}\sum_l \text{cov}\Bigg(x_{t,l}^{(i)}\phi_{\left(x_{t,l}^{(i)}\right)}\left(\mu_x^{(i)},0\right) - \frac{1}{L}\sum_l x_{t,l}^{(i)}\phi_{\left(x_{t,l}^{(i)}\right)}\left(\mu_x^{(i)},0\right) + \xi_{t,l}^{(i)},$$
$$x_{t,l}^{(j)}\phi_{\left(x_{t,l}^{(j)}\right)}\left(\mu_x^{(j)},0\right) - \frac{1}{L}\sum_l x_{t,l}^{(j)}\phi_{\left(x_{t,l}^{(j)}\right)}\left(\mu_x^{(j)},0\right) + \xi_{t,l}^{(j)}\Bigg),$$

where $\xi_{t,l}^{(i)}$ and $\xi_{t,l}^{(j)}$ represent the higher order terms. Then, as $L \to \infty$:

$$(\mathbf{\Sigma}_x^{\text{con}})_{i,j} \approx \frac{1}{L}\sum_l \text{cov}\left(\left(x_{t,l}^{(i)} - \mu_x^{(i)}\right)\phi_{\left(x_{t,l}^{(i)}\right)}\left(\mu_x^{(i)},0\right), \left(x_{t,l}^{(j)} - \mu_x^{(j)}\right)\phi_{\left(x_{t,l}^{(j)}\right)}\left(\mu_x^{(j)},0\right)\right) + \xi_{t,l},$$

from the law of large numbers. Accordingly, we see that:

$$(\mathbf{\Sigma}_x^{\text{con}})_{i,j} \approx C_N \frac{1}{L}\sum_l \text{cov}\left(x_{t,l}^{(i)} - \mu_x^{(i)}, x_{t,l}^{(j)} - \mu_x^{(j)}\right) + \xi_{t,l}$$
$$= C_N\,(\mathbf{\Sigma}_x)_{i,j} + \xi_{t,l},$$

where C_N is a constant and $\xi_{t,l}$ is typically small if the latent process $\mathbf{x}_{t,l}$ and the stimulus $\mathbf{s}_t$ are concentrated around their means. Then, the noise correlations are derived by normalization of the noise covariance given in ***Equation 9***. This cancels out the scaling factor C_N, and we get:

$$(\mathbf{N}^{\text{con}})_{i,j} \quad \approx \quad (\mathbf{N})_{i,j}.$$

Thus, we similarly conclude that as $T, L \to \infty$, the conventional definition of noise correlation $\mathbf{N}^{\text{con}}$ indeed aims to approximate $\mathbf{N}$.

As a numerical illustration, we demonstrated in ***Figure 2–Figure Supplement 2*** that the conventional definitions of the correlations indeed approximate our proposed definitions, but require much larger number of trials to be accurate. More specifically, in order to achieve comparable performance to our method using $L = 20$ trials, the conventional correlation estimates require $L = 1000$ trials.

Appendix 2

Proof of Theorem 1

In what follows, we present a comprehensive proof of Theorem 1. Recall the following key assumptions:

Assumption (1). We assume a scalar time-varying external stimulus (i.e. $\mathbf{s}_t = s_t$, and hence $\mathbf{d}_j = d_j, \mathbf{d} = [d_1, d_2, \cdots, d_N]^\top$). Furthermore, we set the observation noise covariance to be $\mathbf{\Sigma}_w = \sigma_w^2 \mathbf{I}$, for notational convenience.

Assumption (2). We derive the performance bounds in the regime where T and L are large, and thus do not impose any prior distribution on the correlations (i.e., $p_{\text{pr}}(\mathbf{\Sigma}_x) \propto 1$), which are otherwise needed to mitigate overfitting (see Methods and Materials).

Assumption (3). We assume the latent noise process and stimulus to be slowly varying signals, and thus adopt a piece-wise constant model in which these processes are constant within consecutive windows of length W (i.e., $\mathbf{x}_{t,l} = \mathbf{x}_{W_k,l}$ and $s_t = s_{W_k}$, for $(k-1)W+1 \le t < kW$ and $k = 1, \cdots, K$ with $W_k = (k-1)W+1$ and $KW = T$) for our theoretical analysis, as is usually done in spike count calculations for conventional noise correlation estimates.

Proof of Theorem 1. First, recall the proposed forward model (see Methods and Materials) under Assumption (1)–(3):

$$\begin{aligned}
\mathbf{y}_{t,l} &= \mathbf{A}\mathbf{z}_{t,l} + \mathbf{w}_{t,l}, \\
\mathbf{z}_{t,l} &= \alpha\, \mathbf{z}_{t-1,l} + \mathbf{n}_{t,l}, \\
n_{t,l}^{(j)} &\sim \text{Bernoulli}\left(\phi\left(x_{W_k,l}^{(j)}\right)\right), \\
\mathbf{x}_{W_k,l} &\sim \mathcal{N}\left(\boldsymbol{\mu}_x + s_{W_k}\mathbf{d},\, \mathbf{\Sigma}_x\right),
\end{aligned}$$

where $\phi(\cdot) := \frac{\exp(\cdot)}{1+\exp(\cdot)}$, is the logistic function. Note that we have re-defined the latent process $\mathbf{x}_{t,l}$ by absorbing the stimulus activity $s_t\mathbf{d}$ to the mean of $\mathbf{x}_{t,l}$ for notational convenience, without loss of generality. Hereafter, we also assume that $\mathbf{A} = \mathbf{I}$ without loss of generality. For a truncation level B (to be specified later), consider the event $\mathscr{A}_W = \left\{ \left|x_{W_k,l}^{(j)}\right| \le B \text{ and } \frac{1}{2(1+\exp(B))} \le \overline{n}_{W_k,l}^{(j)} \le 1 - \frac{1}{2(1+\exp(B))} \text{ for } j = 1, \cdots, N, k = 1, \cdots, K \text{ and } l = 1, \cdots, L \right\}$, such that $\overline{\mathbf{n}}_{W_k,l} = \left[\overline{n}_{W_k,l}^{(1)}, \overline{n}_{W_k,l}^{(2)}, \cdots, \overline{n}_{W_k,l}^{(N)}\right]^\top := \frac{1}{W}\sum_{w=1}^{W} \mathbf{n}_{(k-1)W+w,l}$. First, we derive convenient forms of the maximum likelihood estimators via the Laplace's approximations and asymptotic expansions (***Wong, 2001***) through the following lemma:

Lemma 1. *Conditioned on event $\mathscr{A}_W$, the maximum likelihood estimators of the stimulus kernel of the j^{th} neuron and the noise covariance between the i^{th} and j^{th} neurons take the forms:*

$$\begin{aligned}
\widehat{d}_j &= \widetilde{d}_j \left(1 + \mathcal{O}\left(\sigma_w^2\right)\right)\left(1 + \mathcal{O}\left(\frac{1}{W}\right)\right) \quad \textit{and} \\
(\widehat{\mathbf{\Sigma}}_x)_{i,j} &= (\widetilde{\mathbf{\Sigma}}_x)_{i,j} \left(1 + \mathcal{O}\left(\sigma_w^2\right)\right)\left(1 + \mathcal{O}\left(\frac{1}{W}\right)\right),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{d}_j &= \frac{1}{L\sum_{k=1}^{K} s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k}\left(\phi^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)}\right) \quad \textit{and} \\
(\widetilde{\mathbf{\Sigma}}_x)_{i,j} &= \frac{1}{KL}\sum_{k,l=1}^{K,L}\left(\phi^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)} - s_{W_k}\widetilde{d}_i\right)\left(\phi^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)} - s_{W_k}\widetilde{d}_j\right),
\end{aligned}$$

with $\widetilde{\mathbf{n}}_{W_k,l} = \left[\widetilde{n}_{W_k,l}^{(1)}, \widetilde{n}_{W_k,l}^{(2)}, \cdots, \widetilde{n}_{W_k,l}^{(N)}\right]^\top := \frac{1}{W}\sum_{w=1}^{W}\left(\mathbf{y}_{(k-1)W+w,l} - \alpha\mathbf{y}_{(k-1)W+w-1,l}\right)$ *and* $\phi^{-1}(z) := \ln(z/(1-z))$.

Proof of Lemma 1. First, maximizing the data likelihood, we derive the estimators:

$$\widehat{d}_j = \underset{d_j}{\operatorname{argmax}}\, p(\mathbf{y}|\boldsymbol{\Sigma}_x, \mathbf{d}) = \frac{\int \left(\frac{1}{L\sum_{k=1}^{K} s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right) p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\boldsymbol{\Sigma}_x, \mathbf{d})\, d\mathbf{n} d\mathbf{x}}{\int p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\boldsymbol{\Sigma}_x, \mathbf{d})\, d\mathbf{n} d\mathbf{x}}, \tag{10}$$

and

$$(\widehat{\boldsymbol{\Sigma}}_x)_{i,j} = \underset{(\boldsymbol{\Sigma}_x)_{i,j}}{\operatorname{argmax}}\, p(\mathbf{y}|\boldsymbol{\Sigma}_x, \mathbf{d}) = \frac{\int \left(\frac{1}{KL} \sum_{k,l=1}^{K,L} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} \widehat{d}_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} \widehat{d}_j \right) \right) p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\boldsymbol{\Sigma}_x, \mathbf{d})\, d\mathbf{n} d\mathbf{x}}{\int p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\boldsymbol{\Sigma}_x, \mathbf{d})\, d\mathbf{n} d\mathbf{x}}. \tag{11}$$

where $W_k = (k-1)W + 1$. Then, we simplify these integrals based on the saddle point method of asymptotic expansions (***Wong, 2001***). To that end, first consider the numerator of ***Equation 10*** denoted by $I_{\text{num}}^{(1)}$. First, we evaluate the integration in $I_{\text{num}}^{(1)}$ with respect to the variable $\mathbf{n}$. To that end, note:

$$I_{\text{num}}^{(1)} = \int h_{\text{num}}^{(1)}(\mathbf{n}) \exp\left(A_1 f_1(\mathbf{n}) \right) d\mathbf{n},$$

where $h_{\text{num}}^{(1)}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^{TNL}\sigma_w^{2TNL}}} \int \left(\frac{1}{L\sum_{k=1}^{K} s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\boldsymbol{\Sigma}_x, \mathbf{d})\, d\mathbf{x}$, $A_1 = \frac{1}{\sigma_w^2}$, $f_1(\mathbf{n}) = -\frac{1}{2} \sum_{t,l,j} \left(y_{t,l}^{(j)} - \sum_{k=1}^{t} \alpha^{t-k} n_{t,l}^{(j)} \right)^2$ and $d\mathbf{n}$ is shorthand notation for the product measure of the discrete random vector $\mathbf{n}$. Observing that $\nabla f_1(\widehat{\mathbf{n}}) = \mathbf{0}$ for $\widehat{\mathbf{n}} := \left\{ \widehat{\mathbf{n}}_{t,l} = \mathbf{y}_{t,l} - \alpha \mathbf{y}_{t-1,l} \right\}_{t,l=1}^{T,L}$, using the method of asymptotic expansions, $I_{\text{num}}^{(1)}$ can be evaluated as:

$$I_{\text{num}}^{(1)} = h_{\text{num}}^{(1)}(\widehat{\mathbf{n}}) \times \exp\left(A_1 f_1(\widehat{\mathbf{n}}) \right) \sqrt{\frac{(2\pi)^{TLN}}{-A_1 \left| H(f_1) \right|}} \left(1 + \mathcal{O}\left(\frac{1}{A_1} \right) \right), \tag{12}$$

where the determinant of the Hessian matrix $\left| H(f_1) \right|$, is a negative function of α. Note that the covariance of this Gaussian integral $\left(-(H(f_1))^{-1} \right)$ is a function of $\alpha \in (0,1)$, and hence is bounded. Thus, all higher order error terms in ***Equation 12*** are also bounded, as higher order moments of Gaussian distributions are functions of the covariance.

Next, we simplify the integral $h_{\text{num}}^{(1)}(\widehat{\mathbf{n}})$ in ***Equation 12*** using a similar procedure. We have:

$$h_{\text{num}}^{(1)}(\widehat{\mathbf{n}}) = \int r_{\text{num}}^{(1)}(\mathbf{x}) \exp\left(A_2 f_2(\mathbf{x}) \right) d\mathbf{x}, \tag{13}$$

where $f_2(\mathbf{x}) = \sum_{k,l,j} \left(\widetilde{n}_{W_k,l}^{(j)} x_{W_k,l}^{(j)} - \log\left(1 + \exp\left(x_{W_k,l}^{(j)} \right) \right) \right)$ with $\widetilde{\mathbf{n}}_{W_k,l} = \left[\widetilde{n}_{W_k,l}^{(1)}, \widetilde{n}_{W_k,l}^{(2)}, \cdots, \widetilde{n}_{W_k,l}^{(N)} \right]^\top := \frac{1}{W} \sum_{w=1}^{W} \widehat{\mathbf{n}}_{(k-1)W+w,l}$, $r_{\text{num}}^{(1)}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{(W+1)KLN} \sigma_w^{2TNL} |\Sigma_x|^{KL}}} \exp\left(-\frac{1}{2} \sum_l \left(\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d} \right)^\top \Sigma_x^{-1} \left(\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d} \right) \right) \times \left(\frac{1}{L\sum_{k=1}^{K} s_{W_k}^2} \sum_{k=1}^{K} s_{W_k,l} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right)$ and $A_2 = W$. Then, we note that the gradient of f_2, $\nabla f_2(\widehat{\mathbf{x}}) = \mathbf{0}$ for $\widehat{\mathbf{x}} := \left\{ \widehat{x}_{W_k,l}^{(j)} = \phi^{-1}\left(\widetilde{n}_{W_k,l}^{(j)} \right) \right\}_{k,l,j=1}^{K,L,N}$, where $\phi^{-1}(z) := \operatorname{logit}(z) = \ln(z/(1-z))$. Accordingly, by re-applying the saddle point method of asymptotic expansions, we evaluate the integral in ***Equation 13*** as:

$$h_{\text{num}}^{(1)}(\widehat{\mathbf{n}}) = r_{\text{num}}^{(1)}(\widehat{\mathbf{x}}) \times \exp\left(A_2 f_2(\widehat{\mathbf{x}}) \right) \sqrt{\frac{(2\pi)^{KLN}}{-A_2 |H(f_2(\widehat{\mathbf{x}}))|}} \left(1 + \mathcal{O}\left(\frac{1}{A_2} \right) \right), \tag{14}$$

where the determinant of the Hessian, $|H(f_2(\widehat{\mathbf{x}}))| = -\prod_{k,l,j} \widetilde{n}_{W_k,l}^{(j)} \left(1 - \widetilde{n}_{W_k,l}^{(j)} \right) < 0$ when conditioned on event $\mathscr{A}_W$. The higher order terms in ***Equation 14*** will be bounded if the covariance of the saddle point approximation $\left(-(H(f_2(\widehat{\mathbf{x}})))^{-1} \right)$ is bounded, which we ensure by conditioning on event $\mathscr{A}_W$. This completes the evaluation of $I_{\text{num}}^{(1)}$.

Following the same sequence of arguments, we evaluate the denominator of ***Equation 10*** denoted by $I^{(1)}_{\text{den}}$. Accordingly, we derive:

$$I^{(1)}_{\text{den}} = h^{(1)}_{\text{den}}(\widehat{\mathbf{n}}) \times \exp\left(A_1 f_1(\widehat{\mathbf{n}})\right)\sqrt{\frac{(2\pi)^{TLN}}{-A_1|H(f_1)|}}\left(1+\mathcal{O}\left(\frac{1}{A_1}\right)\right),$$
$$h^{(1)}_{\text{den}}(\widehat{\mathbf{n}}) = r^{(1)}_{\text{den}}(\widehat{\mathbf{x}}) \times \exp\left(A_2 f_2(\widehat{\mathbf{x}})\right)\sqrt{\frac{(2\pi)^{KLN}}{-A_2|H(f_2(\widehat{\mathbf{x}}))|}}\left(1+\mathcal{O}\left(\frac{1}{A_2}\right)\right), \tag{15}$$

where $r^{(1)}_{\text{den}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{(W+1)KLN}\sigma_w^{2TNL}|\Sigma_x|^{KL}}} \exp\left(-\frac{1}{2}\sum_{k,l}\left(\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k}\mathbf{d}\right)^{\top}\Sigma_x^{-1}\left(\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k}\mathbf{d}\right)\right)$.
Finally, by combining ***Equation 12***, ***Equation 14*** and ***Equation 15***, the maximum likelihood estimator in ***Equation 10*** takes the form:

$$\widehat{d}_j = \frac{I^{(1)}_{\text{num}}}{I^{(1)}_{\text{den}}} = \widetilde{d}_j \frac{\left(1+\mathcal{O}\left(\frac{1}{A_1}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{A_2}\right)\right)}{\left(1+\mathcal{O}\left(\frac{1}{A_1}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{A_2}\right)\right)} = \widetilde{d}_j\left(1+\mathcal{O}\left(\sigma_w^2\right)\right)\left(1+\mathcal{O}\left(\frac{1}{W}\right)\right).$$

Further, following the same sequence of reasoning, simplifying the numerator $(I^{(2)}_{\text{num}})$ and denominator $(I^{(2)}_{\text{den}})$ of ***Equation 11*** yields:

$$(\widehat{\boldsymbol{\Sigma}}_x)_{i,j} = \frac{I^{(2)}_{\text{num}}}{I^{(2)}_{\text{den}}} = (\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\frac{\left(1+\mathcal{O}\left(\frac{1}{A_1}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{A_2}\right)\right)}{\left(1+\mathcal{O}\left(\frac{1}{A_1}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{A_2}\right)\right)} = (\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\left(1+\mathcal{O}\left(\sigma_w^2\right)\right)\left(1+\mathcal{O}\left(\frac{1}{W}\right)\right).$$

This concludes the proof of Lemma 1. □

Given that $\phi^{-1}(z)$ is unbounded for $z = 0$ or $z = 1$, we consider another truncation: $\phi^{-1}_{B'}(z) := \min\{\max\{\phi^{-1}(z), -B'\}, B'\}$, where $B' = 2\log(2\exp(B)+1)$. This choice of B' guarantees that over $\mathscr{A}_W$, $\left|\phi^{-1}_{B'}\left(\overline{n}^{(j)}_{W_k,l}\right)\right| < B'$ for all $j = 1,\cdots,N$, $k = 1,\cdots,K$ and $l = 1,\cdots,L$: and thus $\phi^{-1}_{B'}\left(\overline{n}^{(j)}_{W_k,l}\right) = \phi^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)$ on $\mathscr{A}_W$.

From Lemma 1, the bias and variance of the maximum likelihood estimators, $\widehat{d}_j$ and $(\widehat{\boldsymbol{\Sigma}}_x)_{i,j}$ are upper-bounded, if those of $\widetilde{d}_j$ and $(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}$ are bounded:

$$\left|\text{bias}\left(\widehat{d}_j\right)\right| \le \left|\text{bias}\left(\widetilde{d}_j\right)\right| + \zeta_j, \quad \text{Var}\left(\widehat{d}_j\right) \le \text{Var}\left(\widetilde{d}_j\right) + \widetilde{\zeta}_j, \tag{16}$$

and

$$\left|\text{bias}\left((\widehat{\boldsymbol{\Sigma}}_x)_{i,j}\right)\right| \le \left|\text{bias}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right)\right| + \upsilon_{i,j}, \quad \text{Var}\left((\widehat{\boldsymbol{\Sigma}}_x)_{i,j}\right) \le \text{Var}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right) + \widetilde{\upsilon}_{i,j}, \tag{17}$$

where ζ_j, $\widetilde{\zeta}_j$, $\upsilon_{i,j}$ and $\widetilde{\upsilon}_{i,j}$ represent terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O}\left(\frac{1}{W}\right)$. Thus, we seek to derive the performance bounds of $\widetilde{d}_j$ and $(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}$.

Bounding the bias of $\widehat{d}_j$

Let us first consider $\widetilde{d}_j$. Note that:

$$\begin{aligned}\left|\text{bias}\left(\widetilde{d}_j\right)\right| &:= \left|\mathbb{E}\left[\widetilde{d}_j\right] - d_j\right| \\ &\overset{(a)}{=} \left|\mathbb{E}\left[\widetilde{d}_j - \left(d_{\text{Oracle}}\right)_j\right]\right| \\ &\overset{(b)}{\le} \frac{1}{L\sum_{k=1}^{K} s_{W_k}^2}\sum_{k,l=1}^{K,L}\left|s_{W_k}\right|\mathbb{E}\left[\left|\phi^{-1}_{B'}\left(\overline{n}^{(j)}_{W_k,l}\right) - x^{(j)}_{W_k,l}\right|\right]\end{aligned} \tag{18}$$

where (a) holds since the Oracle estimator, $\left(d_{\text{Oracle}}\right)_j = \frac{1}{L\sum_{k=1}^{K} s^2_{W_k}} \sum_{k,l=1}^{K,L} s_{W_k}\left(x^{(j)}_{W_k,l} - \mu_x^{(j)}\right)$ (i.e., observing $\mathbf{x}_{t,l}$ directly) is unbiased and (b) follows through the application of Jensen's inequality and triangle inequality. To simplify this bound, the triangle inequality yields:

$$\mathbb{E}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - x^{(j)}_{W_k,l}\right|\right] \leq \mathbb{E}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|\right] + \mathbb{E}\left[\left|\phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right) - x^{(j)}_{W_k,l}\right|\right]. \tag{19}$$

Then, to bound each of these terms, we establish a piece-wise linear Lipschitz-type bound on $\phi_{B'}^{-1}(z)$. First, consider the first term $\mathbb{E}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|\right]$. We seek to upper-bound this expectation by bounding $\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|$ via the following technical lemma:.

Lemma 2. *Conditioned on event $\mathscr{A}_W$, the following bound holds for all $j = 1, \cdots, N$, $k = 1, \cdots, K$ and $l = 1, \cdots, L$:*

$$\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right) := \left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right| \leq g\left(B\right)\left|\widetilde{n}^{(j)}_{W_k,l} - \overline{n}^{(j)}_{W_k,l}\right|,$$

where

$$g\left(B\right) = \max\left\{4\left(1+\exp\left(B\right)\right)^2, 4\exp\left(-B\right)\log\left(2\exp\left(B\right)+1\right)\left(1+\left(2\exp\left(B\right)+1\right)^2\right)\right\}.$$

Proof of Lemma 2. First, consider the case $\overline{n}^{(j)}_{W_k,l} \leq 0.5$. We bound the function $\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$ in a piece-wise fashion as follows. Note that $\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right)$ is convex for $\widetilde{n}^{(j)}_{W_k,l} \geq 0.5$ and concave for $\widetilde{n}^{(j)}_{W_k,l} \leq 0.5$. Thus, it immediately follows that for $\widetilde{n}^{(j)}_{W_k,l} \leq \overline{n}^{(j)}_{W_k,l}$, $\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$ is convex and hence:

$$\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right) \quad \leq \quad \frac{\left|B' + \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|}{\left|\overline{n}^{(j)}_{W_k,l} - \frac{1}{1+\exp(B')}\right|}\left(\overline{n}^{(j)}_{W_k,l} - \widetilde{n}^{(j)}_{W_k,l}\right). \tag{20}$$

Furthermore, for $\overline{n}^{(j)}_{W_k,l} \leq \widetilde{n}^{(j)}_{W_k,l} \leq 0.5$, $\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$ is concave, and hence is bounded by the tangent at $\overline{n}^{(j)}_{W_k,l}$:

$$\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right) \quad \leq \quad \frac{1}{\overline{n}^{(j)}_{W_k,l}(1-\overline{n}^{(j)}_{W_k,l})}\left(\widetilde{n}^{(j)}_{W_k,l} - \overline{n}^{(j)}_{W_k,l}\right). \tag{21}$$

Finally, for the case of $\widetilde{n}^{(j)}_{W_k,l} \geq 0.5$, consider the line,

$$h\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right) := \frac{\left|B' - \phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|}{\left|\frac{1}{1+\exp(-B')} - \overline{n}^{(j)}_{W_k,l}\right|}\left(\widetilde{n}^{(j)}_{W_k,l} - \overline{n}^{(j)}_{W_k,l}\right). \tag{22}$$

From the convexity of $\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$, $h\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$ upper bounds $\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right)$ for $\widetilde{n}^{(j)}_{W_k,l} \geq 0.5$, since $h\left(0.5, \overline{n}^{(j)}_{W_k,l}\right) \geq \varepsilon\left(0.5, \overline{n}^{(j)}_{W_k,l}\right)$ for $\overline{n}^{(j)}_{W_k,l} \leq 0.5$. Combining the piece-wise bounds in ***Equation 20***, ***Equation 21*** and ***Equation 22***, we conclude that for $\overline{n}^{(j)}_{W_k,l} \leq 0.5$:

$$\varepsilon\left(\widetilde{n}^{(j)}_{W_k,l}, \overline{n}^{(j)}_{W_k,l}\right) \leq \tilde{g}\left(\overline{n}^{(j)}_{W_k,l}, B'\right)\left|\widetilde{n}^{(j)}_{W_k,l} - \overline{n}^{(j)}_{W_k,l}\right|, \tag{23}$$

where

$$\tilde{g}\left(\overline{n}_{W_k,l}^{(j)}, B'\right) = \max\left\{\frac{1}{\overline{n}_{W_k,l}^{(j)}(1-\overline{n}_{W_k,l}^{(j)})}, \frac{\left|B' + \phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)\right|}{\left|\overline{n}_{W_k,l}^{(j)} - \frac{1}{1+\exp(B')}\right|}, \frac{\left|B' - \phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)\right|}{\left|\frac{1}{1+\exp(-B')} - \overline{n}_{W_k,l}^{(j)}\right|}\right\}.$$

Due to the symmetry of $\varepsilon\left(\widetilde{n}_{W_k,l}^{(j)}, \overline{n}_{W_k,l}^{(j)}\right)$, the same bound in ***Equation 23*** can be established for $\overline{n}_{W_k,l}^{(j)} > 0.5$ as well.

Then, using $\left|\phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)\right| \leq B'$ and conditioning on event $\mathscr{A}_W$, we simplify this bound as:

$$\tilde{g}\left(\overline{n}_{W_k,l}^{(j)}, B'\right) \leq \max\left\{4\left(1+\exp\left(B\right)\right)^2, \frac{4B'\left(1+\exp\left(B'\right)\right)\left(1+\exp\left(B\right)\right)}{\exp\left(B'\right)-\left(2\exp\left(B\right)+1\right)}\right\}.$$

Finally, based on the fact that $B' = 2\log\left(2\exp\left(B\right)+1\right)$, the latter is further upper bounded as:

$$\tilde{g}\left(\overline{n}_{W_k,l}^{(j)}, B'\right) \leq g\left(B\right),$$

where

$$g\left(B\right) = \max\left\{4\left(1+\exp\left(B\right)\right)^2, 4\exp\left(-B\right)\log\left(2\exp\left(B\right)+1\right)\left(1+\left(2\exp\left(B\right)+1\right)^2\right)\right\}.$$

This concludes the proof of Lemma 2. □

Following Lemma 2, by conditioning on the event $\mathscr{A}_W$ we have:

$$\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)\right|\right] \leq g\left(B\right)\mathbb{E}_{\mathscr{A}_W}\left[\left|\widetilde{n}_{W_k,l}^{(j)} - \overline{n}_{W_k,l}^{(j)}\right|\right]. \tag{24}$$

Then, we note that:

$$\begin{aligned}\mathbb{E}\left[\left|\widetilde{n}_{W_k,l}^{(j)} - \overline{n}_{W_k,l}^{(j)}\right|\right] &\overset{(c)}{\leq} \sqrt{\mathbb{E}\left[\left|\widetilde{n}_{W_k,l}^{(j)} - \overline{n}_{W_k,l}^{(j)}\right|^2\right]} \\ &\overset{(d)}{=} \frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}},\end{aligned} \tag{25}$$

where in (c) we have used the Cauchy-Schwarz inequality, and in (d) we have used the fact that the observation noise across the W time instances is i.i.d. and white. From the bounds in ***Equation 24*** and ***Equation 25***, we conclude that the first expectation in ***Equation 19***, conditioned on event $\mathscr{A}_W$ is bounded as:

$$\begin{aligned}\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)\right|\right] &\leq g\left(B\right)\mathbb{E}_{\mathscr{A}_W}\left[\left|\widetilde{n}_{W_k,l}^{(j)} - \overline{n}_{W_k,l}^{(j)}\right|\right] \\ &\leq g\left(B\right)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}\,\mathbb{P}\left(\mathscr{A}_W\right)}.\end{aligned} \tag{26}$$

The foregoing sequence of reasoning similarly follows for $\mathbb{E}\left[\left|\phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right) - x_{W_k,l}^{(j)}\right|\right]$, since $\frac{1}{1+\exp\left(B\right)} \leq \phi\left(x_{W_k,l}^{(j)}\right) \leq 1 - \frac{1}{1+\exp\left(B\right)}$ for $k = 1,\cdots,K$, $l = 1,\cdots,L$ and $j = 1,\cdots,N$ (as a consequence of $|x_{W_k,l}^{(j)}| < B$ for $k = 1,\cdots,K$, $l = 1,\cdots,L$ and $j = 1,\cdots,N$, conditioned on $\mathscr{A}_W$). Accordingly,

we derive the upper bound on the second term in ***Equation 19***, conditioned on event $\mathscr{A}_W$:

$$\begin{aligned}\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\overline{n}_{W_k,l}^{(j)}\right)-x_{W_k,l}^{(j)}\right|\right] &\leq g(B)\,\mathbb{E}_{\mathscr{A}_W}\left[\left|\overline{n}_{W_k,l}^{(j)}-\phi\left(x_{W_k,l}^{(j)}\right)\right|\right]\\ &\overset{(e)}{\leq}\frac{g(B)}{W\,\mathbb{P}\left(\mathscr{A}_W\right)}\sqrt{\mathbb{E}\left[\left(\sum_{w=1}^{W}n_{(k-1)W+w,l}^{(j)}-W\,\phi\left(x_{W_k,l}^{(j)}\right)\right)^2\right]}\\ &\overset{(f)}{=}\frac{g(B)}{W\,\mathbb{P}\left(\mathscr{A}_W\right)}\sqrt{\mathbb{E}\left[W\phi\left(x_{W_k,l}^{(j)}\right)\left(1-\phi\left(x_{W_k,l}^{(j)}\right)\right)\right]}\\ &\overset{(g)}{\leq}\frac{g(B)}{2\sqrt{W}\,\mathbb{P}\left(\mathscr{A}_W\right)},\end{aligned}\tag{27}$$

where (e) follows from the application of Jensen's inequality, (f) follows from the formula for the variance of a Binomial random variable, and (g) follows from the inequality $\phi\left(x_{W_k,l}^{(j)}\right)\times\left(1-\phi\left(x_{W_k,l}^{(j)}\right)\right)\leq 1/4$, for $\phi\left(x_{W_k,l}^{(j)}\right)\in[0,1]$. Combining the results in ***Equation 26*** and ***Equation 27***, the overall expectation in ***Equation 19***, conditioned on the event $\mathscr{A}_W$ is upper-bounded by:

$$\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right)-x_{W_k,l}^{(j)}\right|\right]\leq\frac{2g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right),\tag{28}$$

where we have lower bounded the probability of the event $\mathscr{A}_W$ by $1/2$ (that is, $\mathbb{P}\left(\mathscr{A}_W\right)>1/2$). Thus, from ***Equation 18*** and ***Equation 28*** we derive:

$$\begin{aligned}\left|\mathrm{bias}_{\mathscr{A}_W}\left(\widetilde{d}_j\right)\right| &\leq\frac{2g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)\frac{\sum_{k,l=1}^{K,L}\left|s_{W_k}\right|}{L\sum_{k=1}^{K}s_{W_k}^2}\\ &\overset{(h)}{\leq}\frac{2g(B)}{\sigma_s\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right),\end{aligned}$$

where in (h) we have used the Cauchy-Schwarz inequality $\sum_{k=1}^{K}\left|s_{W_k}\right|\leq\sqrt{K}\sqrt{\sum_{k=1}^{K}s_{W_k}^2}$ while defining $\sigma_s^2:=\frac{1}{K}\sum_{k=1}^{K}s_{W_k}^2$.

Then, for $B\geq 2.5$, we have $g(B)=4(1+\exp(B))^2$ and $B'=2\log(2\exp(B)+1)\leq 3B$. Let $B:=\sigma_m\sqrt{8q\log W}$ for some $q>\frac{1}{64}$. Further, for some $\epsilon<1/2$, suppose that:

$$\log W\geq\max\left\{\frac{\log(8KLN/\eta)}{q},\frac{32\sigma_m^2 q}{\epsilon^2},\frac{2\log(64q)}{1-2\epsilon},\frac{\max\left\{6.25,\,4\left(\|\boldsymbol{\mu}_x\|_\infty+\max_{k,j}\left\{\left|s_{W_k}d_j\right|\right\}\right)^2\right\}}{8q\sigma_m^2},\log 2\right\}.\tag{29}$$

Under these conditions,

$$\begin{aligned}g(B)&\leq 4\left(1+\exp(\sigma_m\sqrt{8q\log W})\right)^2\\ &\overset{(i)}{\leq}16\exp\left(2\sigma_m\sqrt{8q\log W}\right)\\ &\leq 16W^{\epsilon},\end{aligned}\tag{30}$$

where in (i) we have used the fact that $e^x\geq 1$ for $x\geq 0$. Thus, under the conditions in ***Equation 29***, we have:

$$\left|\mathrm{bias}_{\mathscr{A}_W}\left(\widetilde{d}_j\right)\right|\leq\frac{32}{\sigma_s\sqrt{W^{1-2\epsilon}}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right).\tag{31}$$

Finally, from ***Equation 16*** and ***Equation 31***, we conclude that:

$$\left|\text{bias}_{\mathscr{A}_W}\left(\widehat{d}_j\right)\right| \leq \frac{1}{\sqrt{W^{1-2\epsilon}}} C_1 \left(2\sigma_w\sqrt{1+\alpha^2}+1\right) + \mathcal{O}\left(\sigma_w^2\right) + \mathcal{O}\left(\frac{1}{W}\right),$$

where $C_1 := \frac{16}{\sigma_s}$.

Bounding the variance of $\widehat{d}_j$

Next, we prove the upper bound on the variance of the maximum likelihood estimator, $\widehat{d}_j$. To that end, we upper-bound the variance of $\widetilde{d}_j$. First, using the Cauchy-Schwarz inequality, we have:

$$\text{Var}\left(\widetilde{d}_j\right) := \mathbb{E}\left[\left|\widetilde{d}_j - \mathbb{E}\left[\widetilde{d}_j\right]\right|^2\right] \leq \left\{\sqrt{\mathbb{E}\left[\left|\widetilde{d}_j - \left(d_{\text{Oracle}}\right)_j\right|^2\right]} + \sqrt{\text{Var}\left(\left(d_{\text{Oracle}}\right)_j\right)}\right\}^2. \tag{32}$$

Then, we upper-bound the conditional second moment of $\left|\widetilde{d}_j - \left(d_{\text{Oracle}}\right)_j\right|$ using the same techniques as we used in bounding the first moment. Accordingly, we get:

$$\begin{aligned}
\mathbb{E}_{\mathscr{A}_W}\left[\left|\widetilde{d}_j - \left(d_{\text{Oracle}}\right)_j\right|^2\right] &= \frac{1}{\left(L\sum_{k=1}^K s_{W_k}^2\right)^2}\mathbb{E}_{\mathscr{A}_W}\left[\left|\sum_{k,l=1}^{K,L} s_{W_k}\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - x_{W_k,l}^{(j)}\right)\right|^2\right] \\
&\overset{(k)}{\leq} \frac{1}{\left(L\sum_{k=1}^K s_{W_k}^2\right)^2}\left\{\sum_{k,l=1}^{K,L}\left|s_{W_k}\right|\sqrt{\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - x_{W_k,l}^{(j)}\right|^2\right]}\right\}^2 \\
&\overset{(l)}{\leq} \left\{\frac{\sqrt{2}g(B)}{\sigma_s\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)\right\}^2
\end{aligned} \tag{33}$$

where in (k) we have used the Cauchy-Schwarz inequality and (l) follows from $\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - x_{W_k,l}^{(j)}\right|^2\right] \leq \frac{2(g(B))^2}{W}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)^2$, which can be proven by the same techniques as before.

Next, we note that the variance of the Oracle estimator $\left(d_{\text{Oracle}}\right)_j$:

$$\begin{aligned}
\text{Var}\left(\left(d_{\text{Oracle}}\right)_j\right) &= \frac{1}{\left(L\sum_{k=1}^K s_{W_k}^2\right)^2}\sum_{k,l=1}^{K,L} s_{W_k}^2\,\text{Var}\left(\left(x_{W_k,l}^{(j)} - \mu_x^{(j)}\right)\right) \\
&= \frac{\left(\boldsymbol{\Sigma}_x\right)_{j,j}}{L\sum_{k=1}^K s_{W_k}^2} = \frac{\left(\boldsymbol{\Sigma}_x\right)_{j,j}}{LK\sigma_s^2}
\end{aligned} \tag{34}$$

Combining ***Equation 32***, ***Equation 33*** and ***Equation 34***, we can upper-bound the conditional variance of $\widetilde{d}_j$ as:, following ***Equation 32***:

$$\sqrt{\text{Var}_{\mathscr{A}_W}\left(\widetilde{d}_j\right)} \leq \sqrt{\frac{\left(\boldsymbol{\Sigma}_x\right)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{\sqrt{2}g(B)}{\sigma_s\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)$$

Then, following ***Equation 16***, under the conditions for W in ***Equation 29***, we conclude the proof of the conditional variance of $\widehat{d}_j$:

$$\sqrt{\text{Var}_{\mathscr{A}_W}\left(\widehat{d}_j\right)} \leq \sqrt{\frac{\left(\boldsymbol{\Sigma}_x\right)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{1}{\sqrt{W^{1-2\epsilon}}} C_2\left(2\sigma_w\sqrt{1+\alpha^2}+1\right) + \mathcal{O}(\sigma_w^2) + \mathcal{O}\left(\frac{1}{W}\right), \tag{35}$$

where $C_2 := \frac{8\sqrt{2}}{\sigma_s}$.

Bounding the bias of $(\widehat{\boldsymbol{\Sigma}}_x)_{i,j}$

Next, following the foregoing techniques, we upper-bound the bias and variance of the noise covariance estimator $(\widehat{\boldsymbol{\Sigma}}_x)_{i,j}$. To that end, we first note:

$$\begin{aligned}\left|\text{bias}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right)\right| &:= \left|\mathbb{E}\left[(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right] - (\boldsymbol{\Sigma}_x)_{i,j}\right| \\ &\overset{(m)}{\leq} \left|\mathbb{E}\left[(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j} - \left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right]\right| + \left|\text{bias}\left(\left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right)\right|\end{aligned} \tag{36}$$

where (m) follows from the triangle inequality, with the Oracle noise covariance estimator (i.e., observing $\mathbf{x}_{t,l}$ directly), being defined as: $\left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j} = \frac{1}{KL}\sum_{k,l=1}^{K,L}\left(x^{(i)}_{W_k,l} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i\right) \times \left(x^{(j)}_{W_k,l} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j\right)$. Then, to simplify the first term in ***Equation 36***, we use similar techniques as before. Accordingly,

$$\begin{aligned}&\left|\mathbb{E}\left[(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j} - \left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right]\right| \\ &= \left|\mathbb{E}\left[\frac{1}{KL}\sum_{k,l=1}^{KL}\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right) - \mu_x^{(i)} - s_{W_k}\widetilde{d}_i\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \mu_x^{(j)} - s_{W_k}\widetilde{d}_j\right) - \frac{1}{T}\sum_{t=1}^{T}\left(x^{(i)}_{W_k,l} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i\right)\left(x^{(j)}_{W_k,l} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j\right)\right]\right| \\ &\overset{(n)}{\leq} \frac{1}{KL}\sum_{k,l=1}^{K,L}\mathbb{E}\left[\left|\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \mu_x^{(j)}\right) - \left(x^{(i)}_{W_k,l} - \mu_x^{(i)}\right)\left(x^{(j)}_{W_k,l} - \mu_x^{(j)}\right)\right|\right] \\ &+ \frac{1}{KL^2\sum_{k=1}^{K}s_{W_k}^2}\sum_{k,l=1}^{K,L}\mathbb{E}\left[\left|\sum_{k,l=1}^{K,L}s_{W_k}\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right) - \mu_x^{(i)}\right)\sum_{k',l'=1}^{K,L}s_{W_{k'}}\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_{k'},l'}\right) - \mu_x^{(j)}\right) - \sum_{k,l=1}^{K,L}s_{W_k}\left(x^{(i)}_{W_k,l} - \mu_x^{(i)}\right)\sum_{k',l'=1}^{K,L}s_{W_{k'}}\left(x^{(j)}_{W_{k'},l'} - \mu_x^{(j)}\right)\right|\right],\end{aligned} \tag{37}$$

where (n) follows through the application of Jensen's inequality and triangle inequality. Next, we have:

$$\begin{aligned}&\mathbb{E}_{\mathscr{A}_W}\left[\left|\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \mu_x^{(j)}\right) - \left(x^{(i)}_{W_k,l} - \mu_x^{(i)}\right)\left(x^{(j)}_{W_k,l} - \mu_x^{(j)}\right)\right|\right] \\ &\leq \mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right)\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - x^{(i)}_{W_k,l}x^{(j)}_{W_k,l}\right|\right] + \mu_x^{(j)}\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right) - x^{(i)}_{W_k,l}\right|\right] + \mu_x^{(i)}\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - x^{(j)}_{W_k,l}\right|\right] \\ &\leq \mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(i)}_{W_k,l}\right)\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - \phi_{B'}^{-1}\left(\overline{n}^{(i)}_{W_k,l}\right)\phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right)\right|\right] + \mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\overline{n}^{(i)}_{W_k,l}\right)\phi_{B'}^{-1}\left(\overline{n}^{(j)}_{W_k,l}\right) - x^{(i)}_{W_k,l}x^{(j)}_{W_k,l}\right|\right] \\ &\quad + 2\mu_m\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}^{(j)}_{W_k,l}\right) - x^{(j)}_{W_k,l}\right|\right] \\ &\leq 2g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}\left(g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}} + 4\log(2\exp(B)+1)\right) + \frac{2g(B)}{\sqrt{W}}\left\{\frac{g(B)}{4\sqrt{W}} + B\right\} \\ &\quad + \frac{4\mu_m g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2} + \frac{1}{2}\right),\end{aligned} \tag{38}$$

where $\mu_m = \|\boldsymbol{\mu}_x\|_\infty$ and we have used $B' = 2\log(2\exp(B)+1)$. Similarly, the second term in ***Equation 37*** can be bounded as:

$$\begin{aligned}
&\mathbb{E}_{\mathscr{A}_W}\left[\left|\sum_{k,l=1}^{K,L} s_{W_k}\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right)-\mu_x^{(i)}\right)\sum_{k',l'=1}^{K,L} s_{W_{k'}}\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(j)}\right)-\mu_x^{(j)}\right)-\sum_{k,l=1}^{K,L} s_{W_k}\left(x_{W_k,l}^{(i)}-\mu_x^{(i)}\right)\sum_{k',l'=1}^{K,L} s_{W_{k'}}\left(x_{W_{k'},l'}^{(j)}-\mu_x^{(j)}\right)\right|\right] \\
&\le \sum_{k,k',l,l'=1}^{K,K,L,L}\left|s_{W_k}s_{W_{k'}}\right|\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right)\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(j)}\right)-x_{W_k,l}^{(i)}x_{W_{k'},l'}^{(j)}\right|\right] \\
&\quad+\mu_x^{(i)}\sum_{k,k',l,l'=1}^{K,K,L,L}\left|s_{W_k}s_{W_{k'}}\right|\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(j)}\right)-x_{W_{k'},l'}^{(j)}\right|\right] \\
&\quad+\mu_x^{(j)}\sum_{k,k',l,l'=1}^{K,K,L,L}\left|s_{W_k}s_{W_{k'}}\right|\mathbb{E}_{\mathscr{A}_W}\left[\left|\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right)-x_{W_k,l}^{(i)}\right|\right] \qquad (39)\\
&\le\left(L\sum_{k=1}^{K}\left|s_{W_k}\right|\right)^2\left\{2g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}\left(g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}+4\log(2\exp(B)+1)\right)\right. \\
&\quad\left.+\frac{2g(B)}{\sqrt{W}}\left\{\frac{g(B)}{4\sqrt{W}}+B\right\}+\frac{4\mu_m g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)\right\}. \qquad (40)
\end{aligned}$$

Then, by combining the bounds in ***Equation 38*** and ***Equation 40*** and using an instance of Cauchy-Schwarz inequality $\left(\sum_{k=1}^{K}\left|s_{W_k}\right|\right)^2 \le K\sum_{k=1}^{K}s_{W_k}^2$, we see that the bound in ***Equation 37*** can be expressed as:

$$\begin{aligned}
&\left|\mathbb{E}_{\mathscr{A}_W}\left[(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}-(\boldsymbol{\Sigma}_{\text{Oracle}})_{i,j}\right]\right| \\
&\le 4g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}\left(g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}+4\log(2\exp(B)+1)+2\mu_m\right)+\frac{4g(B)}{\sqrt{W}}\left\{\frac{g(B)}{4\sqrt{W}}+B+\mu_m\right\}. \qquad (41)
\end{aligned}$$

Next, we see that the oracle estimator follows an Inverse Wishart distribution, that is $KL\boldsymbol{\Sigma}_{\text{Oracle}} \sim \text{InvWish}_N(\boldsymbol{\Sigma}_x, KL-1)$. Therefore, we get:

$$\mathbb{E}\left[\boldsymbol{\Sigma}_{\text{Oracle}}\right] = \frac{(KL-1)}{KL}\boldsymbol{\Sigma}_x.$$

Thus, the bias of the oracle estimator is given by:

$$\left|\text{bias}\left((\boldsymbol{\Sigma}_{\text{Oracle}})_{i,j}\right)\right| = \frac{1}{KL}\left|(\boldsymbol{\Sigma}_x)_{i,j}\right|. \qquad (42)$$

Combining the results in ***Equation 41*** and ***Equation 42***, the bias of $(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}$ can be bounded as:

$$\begin{aligned}
\left|\text{bias}_{\mathscr{A}_W}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right)\right| &\le \frac{\left|(\boldsymbol{\Sigma}_x)_{i,j}\right|}{KL(1-\eta)}+4g(B)\frac{\sigma_w\sqrt{1+\alpha^2}}{\sqrt{W}}\left(4\log(2\exp(B)+1)+2\mu_m\right) \\
&\quad+\frac{4g(B)}{\sqrt{W}}(B+\mu_m)+\mathcal{O}\left(\frac{g(B)^2}{W}\right). \qquad (43)
\end{aligned}$$

Finally, under the conditions for W in ***Equation 29***, the latter inequality simplifies to:

$$\begin{aligned}
\left|\text{bias}_{\mathscr{A}_W}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right)\right| &\overset{(p)}{\le} \frac{\left|(\boldsymbol{\Sigma}_x)_{i,j}\right|}{KL(1-\eta)}+\frac{Bg(B)}{\sqrt{W}}\left(28\sigma_w\sqrt{1+\alpha^2}+6\right)+\mathcal{O}\left(\frac{g(B)^2}{W}\right) \\
&\overset{(q)}{\le} \frac{\left|(\boldsymbol{\Sigma}_x)_{i,j}\right|}{KL(1-\eta)}+64\sigma_m\sqrt{\frac{2q\log W}{W^{1-2\epsilon}}}\left(14\sigma_w\sqrt{1+\alpha^2}+3\right)+\mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right) \qquad (44)
\end{aligned}$$

where in (p) we have used $2\log(2\exp(B)+1) \le 3B$ and $B > 2\mu_m$ and in (q) we have used $Bg(B) \le 16L^\epsilon\sigma_m\sqrt{8q\log L}$, which follows from ***Equation 30***. Thus, following ***Equation 17*** we

derive the bound on the bias of the maximum likelihood estimator:

$$\left|\text{bias}_{\mathscr{A}_W}\left(\left(\widehat{\boldsymbol{\Sigma}}_x\right)_{i,j}\right)\right| \leq \frac{\left|\left(\boldsymbol{\Sigma}_x\right)_{i,j}\right|}{KL(1-\eta)} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}}\, C_3\left(14\sigma_w\sqrt{1+\alpha^2}+3\right) + \mathcal{O}\left(\sigma_w^2\right) + \mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right),$$

where $C_3 := 64\sigma_m\sqrt{2q}$.

Bounding the variance of $(\widehat{\boldsymbol{\Sigma}}_x)_{i,j}$

Next, we establish an upper bound on the variance of the maximum likelihood estimator of the noise covariance. To that end, we upper-bound the variance of $(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}$. First, using the Cauchy-Schwarz inequality, we get:

$$\text{Var}\left(\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j}\right) := \mathbb{E}\left[\left|\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j} - \mathbb{E}\left[\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j}\right]\right|^2\right] \leq \left\{\sqrt{\mathbb{E}\left[\left|\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j} - \left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right|^2\right]} + \sqrt{\text{Var}\left(\left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right)}\right\}^2. \tag{45}$$

Then, we upper-bound the conditional second moment of $\left|\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j} - \left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right|$ using the same techniques used in bounding its first moment. Accordingly, we derive:

$$\begin{aligned}
&\mathbb{E}_{\mathscr{A}_W}\left[\left|\left(\widetilde{\boldsymbol{\Sigma}}_x\right)_{i,j} - \left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right|^2\right] \\
&= \tfrac{1}{K^2L^2}\mathbb{E}_{\mathscr{A}_W}\left[\left\{\textstyle\sum_{k,l=1}^{K,L}\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)} - s_{W_k}\widetilde{d}_i\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)} - s_{W_k}\widetilde{d}_j\right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j\right)\right)\right\}^2\right] \\
&\leq \tfrac{1}{K^2L^2}\left\{\textstyle\sum_{k,l=1}^{K,L}\left\{\mathbb{E}_{\mathscr{A}_W}\left[\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)} - s_{W_k}\widetilde{d}_i\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)} - s_{W_k}\widetilde{d}_j\right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j\right)\right)^2\right]\right\}^{\frac{1}{2}}\right\}^2
\end{aligned} \tag{46}$$

where the last bound follows from the Cauchy-Schwarz inequality. Then, we derive:

$$\begin{aligned}
&\mathbb{E}_{\mathscr{A}_W}\left[\left\{\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)} - s_{W_k}\widetilde{d}_i\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)} - s_{W_k}\widetilde{d}_j\right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j\right)\right\}^2\right] \\
&= \mathbb{E}_{\mathscr{A}_W}\left[\left\{\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)} - s_{W_k}\frac{1}{L\sum_{k'=1}^K {s_{W_{k'}}}^2}\textstyle\sum_{k',l'=1}^{K,L} s_{W_{k'}}\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(i)}\right) - \mu_x^{(i)}\right)\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)} - s_{W_k}\frac{1}{L\sum_{k''=1}^K {s_{W_{k''}}}^2}\textstyle\sum_{k'',l''=1}^{K,L} s_{W_{k''}}\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k''},l''}^{(j)}\right) - \mu_x^{(j)}\right)\right)\right. \\
&\quad \left. - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}\frac{1}{L\sum_{k'=1}^K {s_{W_{k'}}}^2}\textstyle\sum_{k',l'=1}^{K,L} s_{W_{k'}}\left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)}\right)\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}\frac{1}{L\sum_{k''=1}^K {s_{W_{k''}}}^2}\textstyle\sum_{k'',l''=1}^{K,L} s_{W_{k''}}\left(x_{W_{k''},l''}^{(j)} - \mu_x^{(j)}\right)\right)\right\}^2\right] \\
&= \mathbb{E}_{\mathscr{A}_W}\left[\left\{\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)}\right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)}\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)}\right)\right)\right. \\
&\quad - \frac{s_{W_k}}{L\sum_{k''=1}^K {s_{W_{k''}}}^2}\textstyle\sum_{k'',l''=1}^{K,L} s_{W_{k''}}\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(i)}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k''},l''}^{(j)}\right) - \mu_x^{(j)}\right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)}\right)\left(x_{W_{k''},l''}^{(j)} - \mu_x^{(j)}\right)\right) \\
&\quad - \frac{s_{W_k}}{L\sum_{k'=1}^K {s_{W_{k'}}}^2}\textstyle\sum_{k',l'=1}^{K,L} s_{W_{k'}}\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(i)}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_k,l}^{(j)}\right) - \mu_x^{(j)}\right) - \left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)}\right)\left(x_{W_k,l}^{(j)} - \mu_x^{(j)}\right)\right) \\
&\quad \left. + \frac{{s_{W_k}}^2}{L^2\sum_{k'=1}^K {s_{W_{k'}}}^2\sum_{k''=1}^K {s_{W_{k''}}}^2}\textstyle\sum_{k',k'',l',l''=1}^{K,L} s_{W_{k'}}s_{W_{k''}}\left(\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k'},l'}^{(i)}\right) - \mu_x^{(i)}\right)\left(\phi_{B'}^{-1}\left(\widetilde{n}_{W_{k''},l''}^{(j)}\right) - \mu_x^{(j)}\right) - \left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)}\right)\left(x_{W_{k''},l''}^{(j)} - \mu_x^{(j)}\right)\right)\right\}^2\right] \\
&\leq \frac{2\left(g\left(B\right)\right)^2}{W}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)^2\left(\frac{g\left(B\right)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)+2(B+\mu_m)\right)^2 \\
&\quad \times\left(1 + \frac{s_{W_k}L\sum_{k'=1}^K s_{W_{k'}}}{L\sum_{k'=1}^K {s_{W_{k'}}}^2} + \frac{s_{W_k}L\sum_{k''=1}^K s_{W_{k''}}}{L\sum_{k''=1}^K {s_{W_{k''}}}^2} + \frac{s_{W_k}^2L\sum_{k'=1}^K s_{W_{k'}}L\sum_{k''=1}^K s_{W_{k''}}}{L^2\sum_{k'=1}^K {s_{W_{k'}}}^2\sum_{k''=1}^K {s_{W_{k''}}}^2}\right)^2.
\end{aligned} \tag{47}$$

Using the final bound of ***Equation 47*** in ***Equation 46***, we get:

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\mathscr{A}_W}\left[\left|(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}-(\boldsymbol{\Sigma}_{\text{Oracle}})_{i,j}\right|^2\right]}\\
&\leq \frac{\sqrt{2}g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)\left(\frac{g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)+2(B+\mu_m)\right)\\
&\quad\times\frac{1}{KL}\left(KL+\frac{L\sum_{k=1}^K|s_{W_k}|\sum_{k'=1}^K|s_{W_{k'}}|}{\sum_{k'=1}^K {s_{W_{k'}}}^2}+\frac{L\sum_{k=1}^K|s_{W_k}|\sum_{k''=1}^K|s_{W_{k''}}|}{\sum_{k''=1}^K {s_{W_{k''}}}^2}+\frac{L\sum_{k=1}^K s_{W_k}^2\sum_{k'=1}^K|s_{W_{k'}}|\sum_{k''=1}^K|s_{W_{k''}}|}{\sum_{k'=1}^K {s_{W_{k'}}}^2\sum_{k''=1}^K {s_{W_{k''}}}^2}\right)\\
&\leq \frac{4\sqrt{2}g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)\left(\frac{g(B)}{\sqrt{W}}\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)+2(B+\mu_m)\right),
\end{aligned}\tag{48}$$

where the last inequality follows from an instance of the Cauchy-Schwarz inequality, i.e., $\left(\sum_{k=1}^K\left|s_{W_k}\right|\right)^2\leq K\sum_{k=1}^K s_{W_k}^2$.

Then, following the observation $KL\,\boldsymbol{\Sigma}_{\text{Oracle}}\sim\text{InvWish}_{\text{N}}(\boldsymbol{\Sigma}_x,KL-1)$, we derive the variance of $\left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}$:

$$\text{Var}\left(\left(\boldsymbol{\Sigma}_{\text{Oracle}}\right)_{i,j}\right)=\delta_{i,j}^2=\frac{(KL-1)\left((\boldsymbol{\Sigma}_x)_{i,j}^2+(\boldsymbol{\Sigma}_x)_{i,i}(\boldsymbol{\Sigma}_x)_{j,j}\right)}{K^2L^2}.\tag{49}$$

Combining ***Equation 45***, ***Equation 48*** and ***Equation 49***, we express the upper bound on the conditional variance of $(\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}$ as:

$$\sqrt{\text{Var}_{\mathscr{A}_{\text{W}}}\left((\widetilde{\boldsymbol{\Sigma}}_x)_{i,j}\right)}\leq\frac{1}{\sqrt{1-\eta}}\delta_{i,j}+\frac{8\sqrt{2}g(B)}{\sqrt{W}}(B+\mu_m)\left(\sigma_w\sqrt{1+\alpha^2}+\frac{1}{2}\right)+\mathcal{O}\left(\frac{g(B)^2}{W}\right).$$

Then, following ***Equation 17*** and the conditions in ***Equation 29***, we conclude the proof of the upper bound on the conditional variance of $(\widehat{\boldsymbol{\Sigma}}_x)_{i,j}$:

$$\sqrt{\text{Var}_{\mathscr{A}_{\text{W}}}\left((\widehat{\boldsymbol{\Sigma}}_x)_{i,j}\right)}\leq\frac{1}{\sqrt{1-\eta}}\delta_{i,j}+\sqrt{\frac{\log W}{W^{1-2\epsilon}}}C_4\left(2\sigma_w\sqrt{1+\alpha^2}+1\right)+\mathcal{O}\left(\sigma_w^2\right)+\mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right),$$

where $C_4:=384\sigma_m\sqrt{q}$.

Finally, it only remains to prove that the event $\mathscr{A}_W$ occurs with high probability for sufficiently large W:

Lemma 3. *The probability of occurrence of the event* $\mathscr{A}_W=\left\{\left|x_{W_k,l}^{(j)}\right|\leq B \text{ and } \frac{1}{2(1+\exp(B))}\leq\overline{n}_{W_k,l}^{(j)}\leq 1-\frac{1}{2(1+\exp(B))} \text{ for } j=1,\cdots,N,k=1,\cdots,K \text{ and } l=1,\cdots,L\right\}$ *is upper-bounded as follows:*

$$\mathbb{P}\left(\mathscr{A}_W\right)\geq 1-\eta,$$

for some constant $0<\eta\leq 1/2$ *satisfying the conditions of Eq. (29).*

Proof of Lemma 3. First, using the union bound, we have:

$$\mathbb{P}\left(\mathscr{A}_W\right)\geq 1-\sum_{k,l,j=1}^{K,L,N}\left\{\mathbb{P}\left(\left|x_{W_k,l}^{(j)}\right|>B\right)+\mathbb{P}\left(\overline{n}_{W_k,l}^{(j)}<\frac{1}{2(1+\exp(B))}\right)+\mathbb{P}\left(\overline{n}_{W_k,l}^{(j)}>1-\frac{1}{2(1+\exp(B))}\right)\right\}.\tag{50}$$

Next, we bound the probabilities on the right hand side using Chernoff's inequality (***Boucheron et al., 2013***). First, note that:

$$\begin{aligned}\mathbb{P}\left(x^{(j)}_{W_k,l} > B\right) &= \mathbb{P}\left(x^{(j)}_{W_k,l} - \mu^{(j)}_x - s_{W_k} d_j > B - \mu^{(j)}_x - s_{W_k} d_j\right)\\ &\overset{(r)}{\leq} \mathbb{P}\left(x^{(j)}_{W_k,l} - \mu^{(j)}_x - s_{W_k} d_j > \frac{B}{2}\right)\\ &\overset{(s)}{\leq} \exp\left(-\frac{B^2}{8\sigma_m^2}\right),\end{aligned}$$

where (r) follows if $B > 2\left(\|\boldsymbol{\mu}_x\|_\infty + \max_{k,j}\left\{\,|s_{W_k} d_j|\,\right\}\right)$ (which will hold under the conditions in ***Equation 29***) and (s) has been derived by applying the Chernoff's bound on the Gaussian random variable $x^{(j)}_{W_k,l}$. From the same reasoning we see that $\mathbb{P}\left(x^{(j)}_{W_k,l} < -B\right) \leq \exp\left(-\frac{B^2}{8\sigma_m^2}\right)$. Combining these two results, we get the upper bound:

$$\mathbb{P}\left(\left|x^{(j)}_{W_k,l}\right| > B\right) \leq 2\exp\left(-\frac{B^2}{8\sigma_m^2}\right). \tag{51}$$

Next, note that:

$$\begin{aligned}\mathbb{P}\left(\overline{n}^{(j)}_{W_k,l} < \frac{1}{2(1+\exp(B))}\right) &\overset{(u)}{\leq} \mathbb{P}\left(\overline{n}^{(j)}_{W_k,l} - \phi\left(x^{(j)}_{W_k,l}\right) < \frac{-1}{2(1+\exp(B))}\right)\\ &\overset{(v)}{\leq} \exp\left(-\frac{W}{16(1+\exp(B))^2}\right),\end{aligned} \tag{52}$$

where (u) follows from the observation $\frac{1}{1+\exp(B)} < \phi\left(x^{(j)}_{W_k,l}\right)$ (which is a consequence of $|x^{(j)}_{W_k,l}| < B$). Then, we note that the zero-mean random variable $\overline{n}^{(j)}_{W_k,l} - \phi\left(x^{(j)}_{W_k,l}\right)$ is sub-Gaussian with variance factor $\frac{2}{W}$. Thus, using the Chernoff's inequality on sub-Gaussian random variables (***Boucheron et al., 2013***), we derive the upper-bound (v) in ***Equation 52***. In a similar fashion, based on the observation $\phi\left(x^{(j)}_{W_k,l}\right) < 1 - \frac{1}{1+\exp(B)}$, we conclude the bound:

$$\mathbb{P}\left(\overline{n}^{(j)}_{W_k,l} > 1 - \frac{1}{2(1+\exp(B))}\right) \leq \exp\left(-\frac{W}{16(1+\exp(B))^2}\right). \tag{53}$$

By combining the bounds in ***Equation 51***, ***Equation 52*** and ***Equation 53***, the upper bound on $\mathbb{P}\left(\mathscr{A}_W\right)$ in ***Equation 50*** takes the form:

$$\mathbb{P}\left(\mathscr{A}_W\right) \geq 1 - 2KLN\exp\left(-\frac{W}{16(1+\exp(B))^2}\right) - 2KLN\exp\left(-\frac{B^2}{8\sigma_m^2}\right).$$

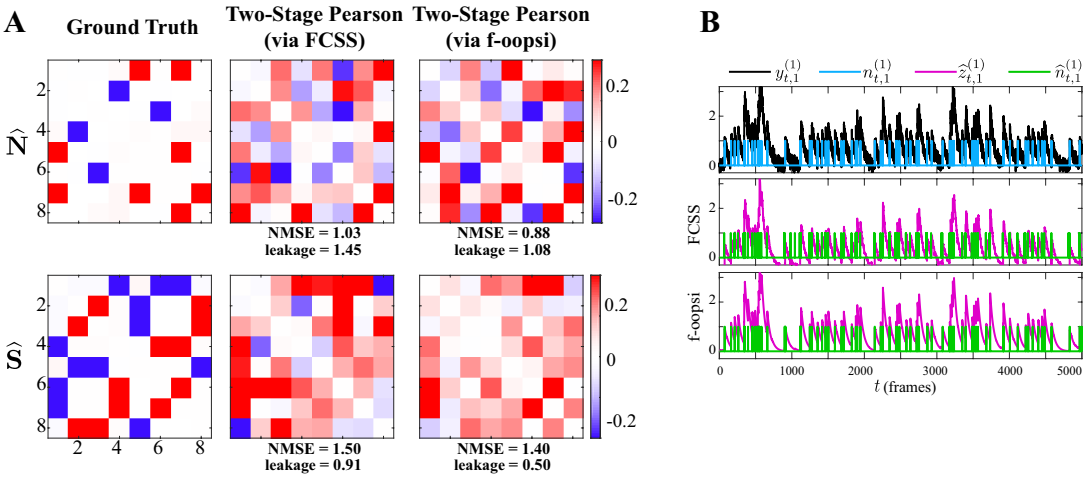
Finally, under the assumptions in ***Equation 29***, we further simplify this bound as:

$$\mathbb{P}\left(\mathscr{A}_W\right) \geq 1 - 2KLN\exp\left(-\frac{W^{1-\epsilon}}{64}\right) - \frac{2KLN}{W^q} \geq 1 - \frac{4KLN}{W^q},$$

where we have used $W \geq 2$ (which gives $\log W \geq 2\log\log W$) and $\log W \geq \frac{2\log(64q)}{1-2\epsilon}$ to show that $\frac{W^{1-\epsilon}}{64} \geq q\log W$. Thus, $\log W \geq \frac{\log(8KLN/\eta)}{q}$ ensures that $\mathbb{P}\left(\mathscr{A}_W\right) \geq 1-\eta$, for $0 < \eta \leq \frac{1}{2}$. □

This concludes the proof of Theorem 1.

□



1523

Figure 2–Figure supplement 1. A) Noise (first row) and signal (second row) correlations corresponding to the ground truth (first column), estimated by the two-stage Pearson method using the FCSS (***Kazemipour et al., 2018***) (second column) and constrained f-oopsi (***Pnevmatikakis et al., 2016***) (third column) spike deconvolution techniques, for the simulation study in **Figure 2**. The NMSE and leakage ratios of the estimates are indicated below each panel. While the correlation estimates based on these two methods are comparable, there exist notable differences between them, as a result of the slight discrepancies in the deconvolved spikes. This demonstrates that the two-stage estimates are notably sensitive to minor differences in the estimated spikes obtained by different deconvolution techniques. In addition, both two-stage Pearson estimates fail to capture the ground truth correlations (as is also evident from the high NMSE and leakage values). B) Simulated observations (black, re-scaled for ease of visual comparison) and ground truth spikes (blue), as well as the estimated calcium concentrations (purple) and putative spikes (green) for the 1st trial of neuron 1 in the simulation study of **Figure 2**, using the FCSS (***Kazemipour et al., 2018***) (second row) and constrained f-oopsi (***Pnevmatikakis et al., 2016***) (third row) spike deconvolution methods.

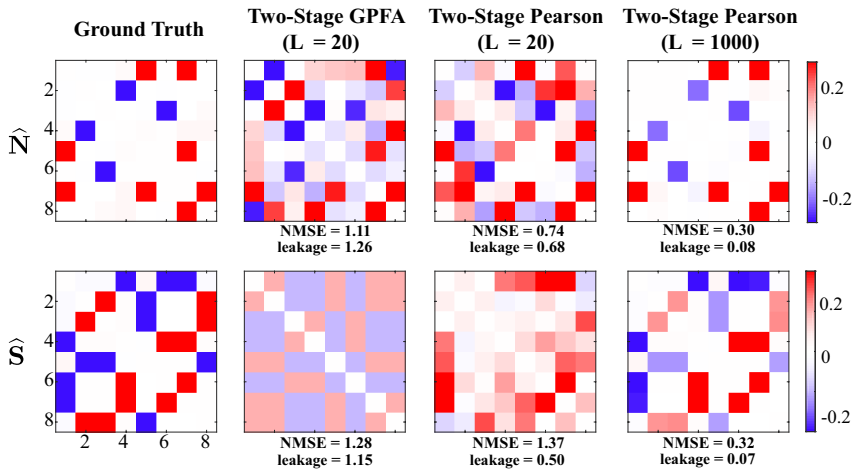


Figure 2–Figure supplement 2. Performance of two stage estimates based on ground truth spikes. Noise (first row) and signal (second row) correlations corresponding to the ground truth (first column) are repeated from **Figure 2**. The second and third columns show the results of two-stage GPFA and two-stage Pearson methods using $L = 20$ trials, respectively. The fourth column shows the results of the two-stage Pearson method using $L = 1000$ trials. All estimates were obtained using the ground truth spikes, as opposed to extracting the spikes via a deconvolution technique. Thus, these results isolate the effect of the non-linearities involved in spike generation on the estimation performance. The NMSE and leakage ratios of the estimates are indicated below each panel. Even though the ground truth spikes are used, the NMSE and leakage ratios indicated in the second and third columns are remarkably high. This further shows that the usage of conventional definitions and GPFA estimates is not optimal for the recovery of signal and noise correlations. In accordance with our theoretical analysis in **Appendix 1**, the performance of the two-stage Pearson method significantly improves as the number of trials is increased to $L = 1000$, a number that is unrealistic in the context of typical two-photon imaging experiments. However, our proposed method shown in **Figure 2** achieves comparable performance with number of trials as low as $L = 20$. In summary, these results suggest that the two-stage methods produce highly biased estimates under limited number of trials, even if the ground truth spikes were ideally deconvolved from the two-photon data.

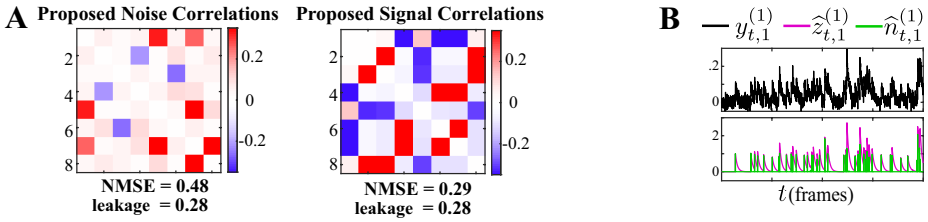


Figure 2–Figure supplement 3. A) Proposed noise and signal correlation estimates for data simulated at lower SNR than the setting of **Figure 2** and model mismatch introduced by using a second-order autoregressive model for the calcium decay. The ground truth correlations are the same as those in **Figure 2**. The NMSE and leakage ratio are given at the bottom. B) putative spikes (green) and estimated calcium concentrations (purple). The model mismatch and lower SNR result in slight performance degradation compared to **Figure 2** (in terms of NMSE and leakage), and our method is capable of recovering the underlying correlations faithfully.

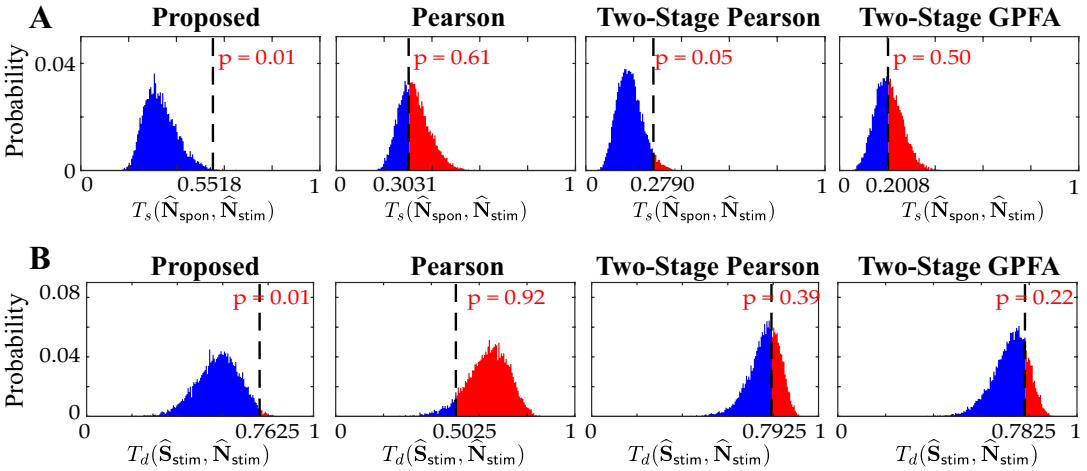


Figure 6–Figure supplement 1. Null distributions of A) the similarities between $\mathbf{N}_{\text{spon}}$ and $\mathbf{N}_{\text{stim}}$ (top: $T_s(\widehat{\mathbf{N}}_{\text{spon}}, \widehat{\mathbf{N}}_{\text{stim}})$) and B) the dissimilarities between $\widehat{\mathbf{S}}_{\text{stim}}$ and $\widehat{\mathbf{N}}_{\text{stim}}$ (bottom: $T_d(\widehat{\mathbf{S}}_{\text{stim}}, \widehat{\mathbf{N}}_{\text{stim}})$), obtained by the shuffling procedure applied to the results of real data study 2 in **Figure 6**. The observed test statistic in each case is indicated by a dashed vertical line. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson correlations and two-stage GPFA estimates. These results show that the only statistically significant outcomes (with $p \leq 0.05$) are the similarities and dissimilarities obtained by our proposed method.

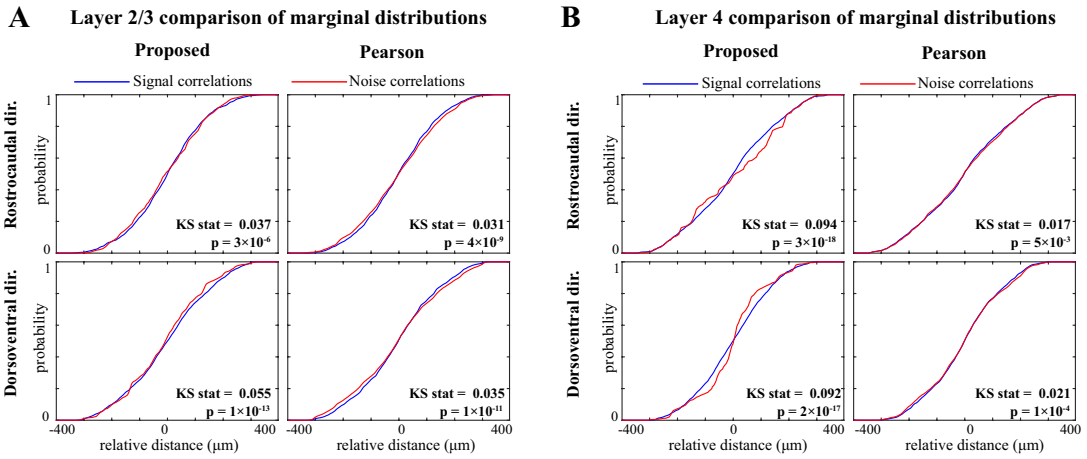


Figure 7–Figure supplement 1. Comparison of marginal distributions of signal and noise correlations. A) Cumulative marginal probability distributions of signal (blue) and noise (red) correlations along the rostrocaudal (top) and dorsoventral (bottom) directions, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right), in layer 2/3 neurons. The Kolmogorov–Smirnov (KS) test statistic along with the corresponding p-values are indicated as insets in each panel. Panel B shows the results for layer 4 in the same organization as panel A. These results show that along both directions and in both layers, the signal correlation distributions are significantly different from the corresponding noise correlation distributions, consistently for both methods. However, the KS statistics (i.e., effect sizes) for the proposed estimate are remarkably larger than those obtained from the Pearson estimates.

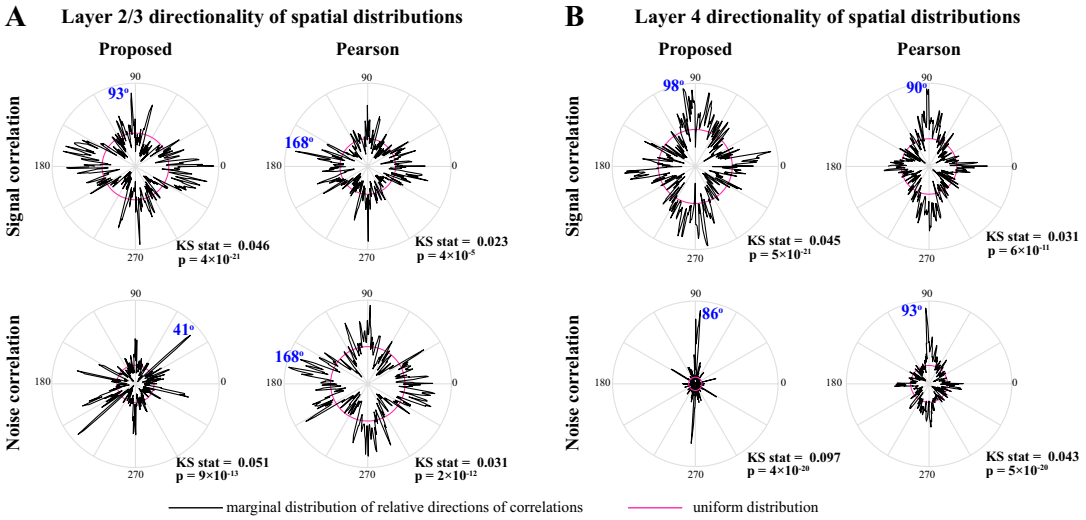


Figure 7–Figure supplement 2. Polar plots of the angular marginal distributions of correlations. A) Polar histograms indicating the distribution of signal (top) and noise (bottom) correlations as a function of relative angle (in the dorsoventral-rostrocaudal coordinate system) between pairs of neurons in layer 2/3, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right). The KS test statistic comparing each polar distribution with a uniform distribution (shown in magenta), along with the corresponding p-values are indicated below each polar plot. The mode of each probability distribution is also indicated in blue fonts. Panel B shows the results for layer 4 in the same organization as panel A. All distributions are significantly non-uniform, and particularly indicate a rostrocaudal directionality in layer 4 (as indicated by the mode angles in panel B).