Differential Covariance: A New Class of Methods to Estimate Sparse Connectivity from Neural Recordings

Tiger W. Lin[1, 2], **Anup Das**[1, 5], **Giri P. Krishnan**[4], **Maxim Bazhenov**[4], **Terrence J. Sejnowski**[1, 3]

[1]Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037.

[2]Neurosciences Graduate Program, University of California San Diego, La Jolla, CA 92092.

[3]Institute for Neural Computation, University of California San Diego, La Jolla, CA 92092.

[4]Department of Medicine, University of California San Diego, La Jolla, CA 92092.

[5]Jacobs School of Engineering, University of California San Diego, La Jolla, CA 92092.

Keywords: Functional connectivity, correlation estimation, spiking neural network, sparse connectivity, neural recordings, local field potential, calcium imaging

Abstract

With our ability to record more neurons simultaneously, making sense of these data is a challenge. Functional connectivity is one popular way to study the relationship between multiple neural signals. Correlation-based methods are a set of currently well-used techniques for functional connectivity estimation. However, due to explaining away and unobserved common inputs [Stevenson et al., 2008], they produce spurious connections. The general linear model (GLM), which models spikes trains as Poisson processes [Okatan et al., 2005, Truccolo et al., 2005, Pillow et al., 2008], avoids these confounds. We develop here a new class of methods by using differential signals based on intracellular voltage recordings. We then expand the results to local field potential (LFP) recordings and calcium imaging. The differential covariance-based methods achieved better or similar performance to the GLM method and required less data samples. This new class of methods provides alternative ways to analyze neural signals.

1 Introduction

Simultaneous recording of large population of neurons is an inexorable trend in current neuroscience research [Kandel et al., 2013]. Over the last five decades, the number of

simultaneously recorded neurons doubles approximately every 7 years [Stevenson and Kording, 2011]. One way to make sense of this big data is to measure the functional connectivity between neurons [Friston, 2011], and link the function of the neural circuit to behavior. As previously reviewed [Stevenson et al., 2008], correlation-based methods have been used to estimate functional connectivity for a long time. However, they are suffering from the problem of explaining away and unobserved common inputs [Stevenson et al., 2008], which make it difficult to interpret the link between the estimated correlation and the physiological network. More recently, Okatan et al. [2005], Truccolo et al. [2005], Pillow et al. [2008] applied the generalized linear model to spike train data and showed that the method avoids these confounds.

To overcome these issues, we developed a new class of methods that achieve better performance to the GLM method but are free from the Poisson process model and require less data samples. They provide directionality information about sources and sinks in a network, which is important to determine the hierarchical structure of a neural circuit.

To validate our new methods, we first generated data using a simple passive neuron model, and provide theoretical proof for why the new methods perform better. Then, we used a more realistic Hodgkin-Huxley (HH) based thalamocortical model to simulate intracellular recordings and local field potential data. This model can successfully generate sleep patterns such as spindles and slow waves [Bazhenov et al., 2002, Chen et al., 2012, Bonjean et al., 2011]. Since the model has a cortical layer and a thalamic layer, we further assume that the neural signals in the cortical layer are visible by the recording instruments while those from the thalamic layer are not. This is a reasonable assumption for many experiments. Since, thalamus is a deep brain structure, most experiments involve only measurements from cortex.

Next, we simulated 1000 Hodgkin-Huxley neurons networks with 80% excitatory neurons and 20% inhibitory neurons sparsely connected. As in real experiments, We recorded simulated calcium signals from only a small percentage of the network (50 neurons) and compared the performance of different methods. In all situations, our differential covariance-based methods achieve better or similar performance to the GLM method. And in the LFP and calcium imaging dataset, they achieve the same performance with less data samples.

The paper is organized as follow: In section 2, we introduce our new methods. In section 3, we show the performance of all methods and explain why our methods perform better. In section 4, we discuss the advantage and generalizability of our methods. We also propose a few improvements that can be done in the future.

2 Methods

2.1 Simulation models used to benchmark the methods

2.1.1 Passive neuron model

To validate and test our new methods, we first developed a passive neuron model. Because of its simplicity, we can provide some theoretical proof for why our new class of methods are better. Every neuron in this model has a passive cell body with capacitance

C and a leakage channel with conductance g_l. Neurons are connected with a directional synaptic conductance g_{syn}; For example, neuron i receiving inputs from neurons $i-4$ and $i-3$:

$$C\frac{dV_i}{dt} = g_{syn}V_{i-4} + g_{syn}V_{i-3} + g_l V_i + \mathcal{N}_i \tag{1}$$

Here, we let $C = 1, g_{syn} = 3, g_l = -5$, and $\mathcal{N}_i$ is a Gaussian noise with standard deviation of 1. The connection pattern is shown in Fig.3A. There are 60 neurons in this circuit. The first 50 neurons are connected with a pattern that: neuron i projects to neuron $i+3$ and $i+4$. To make the case more realistic, aside of these 50 neurons that are visible, we added 10 more neurons (neuron 51 to 60 in Fig.3A) that are invisible during our estimations (i.e. only the membrane voltages of the first 50 neurons are used to estimate connectivity). These 10 neurons send latent inputs to the visible neurons and introduce external correlations into the system. Therefore, we update our passive neuron's model as:

$$C\frac{dV_i}{dt} = g_{syn}V_{i-4} + g_{syn}V_{i-3} + g_{latent}V_{latent1} + g_l V_i + \mathcal{N}_i \tag{2}$$

where $g_{latent} = 10$.

We added the latent inputs to the system because unobserved common inputs exist in real-world problems [Stevenson et al., 2008]. For example, one could be using two-photon imaging to record calcium signals from the cortical circuit. The cortical circuit might receive synaptic inputs from deeper layers in the brain, such as the thalamus, which is not visible to the microscopy. Each invisible neuron projects to many visible neurons leading to common synaptic currents to the cortical circuit and cause neurons in the cortical circuit to be highly correlated. Later, we discuss how to remove interference from the latent inputs.

2.1.2 Thalamocortical model

To test and benchmark the differential covariance-based methods in a more realistic model, we simulated neural signals from a Hodgkin-Huxley based spiking neural network. The thalamocortical model used in this study was based on several previous studies, which were used to model spindle and slow wave activity [Bazhenov et al., 2002, Chen et al., 2012, Bonjean et al., 2011]. A schematic of the thalamocortical model in this work is shown in Fig. 1. As shown, the thalamocortical model was structured as a one-dimensional, multi-layer array of cells. The thalamocortical network consisted of 50 cortical pyramidal (PY) neurons, 10 cortical inhibitory (IN) neurons, 10 thalamic relay (TC) neurons and 10 reticular (RE) neurons. The connections between the 50 PY neurons follow the pattern in the passive neuron model and are shown in Fig. 7A. For the rest of the connection types, a neuron connects to all target neurons within the radius listed in Table 1 [Bazhenov et al., 2002]. The network is driven by spontaneous oscillations. Details of the model is explained in appendix C.

For each simulation, we simulated the network for 600 secs. The data were sampled at 1000 Hz.

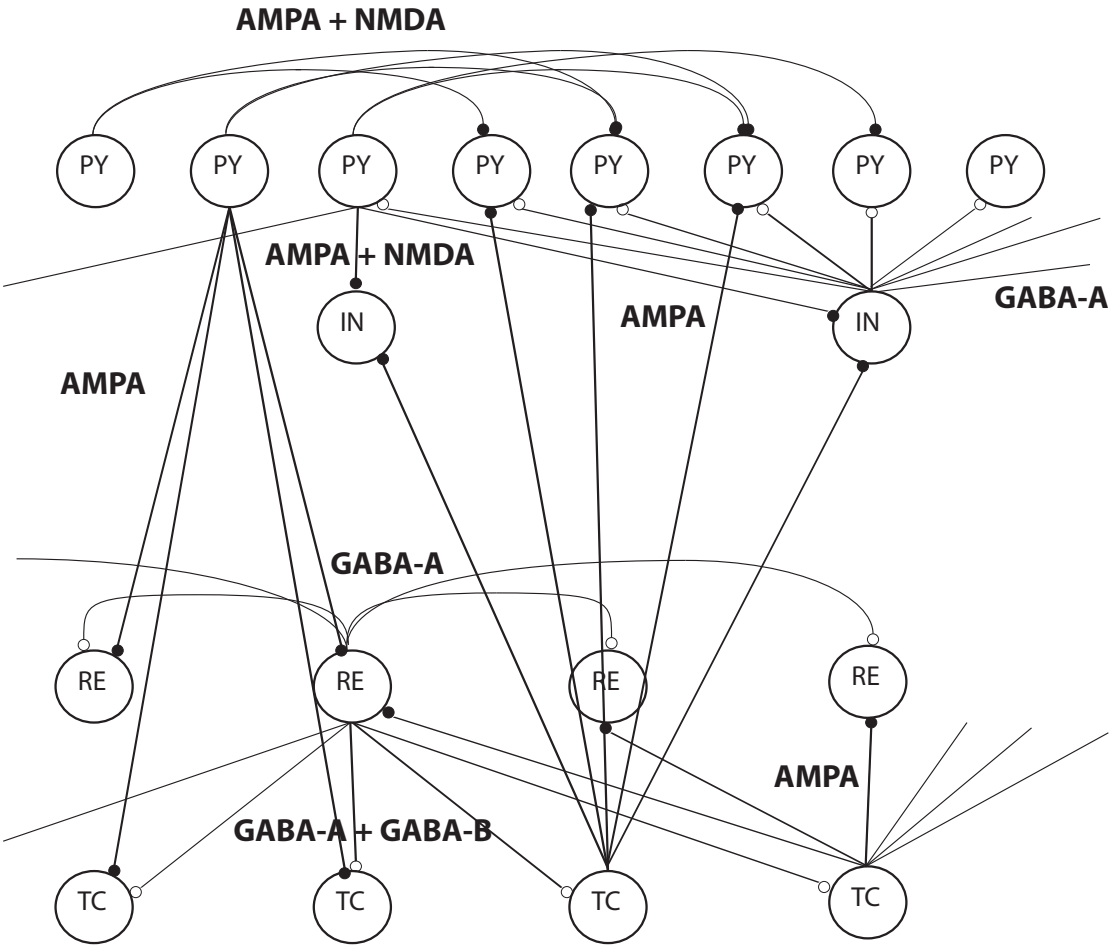

Figure 1: Network model for the thalamocortical interactions, which included four layers of neurons. Thalamocortical (TC), reticular nucleus (RE) of the thalamus, cortical pyramidal (PY) and inhibitory (IN) neurons. Small solid dots indicate excitatory synapses. Small open dots indicate inhibitory synapses.

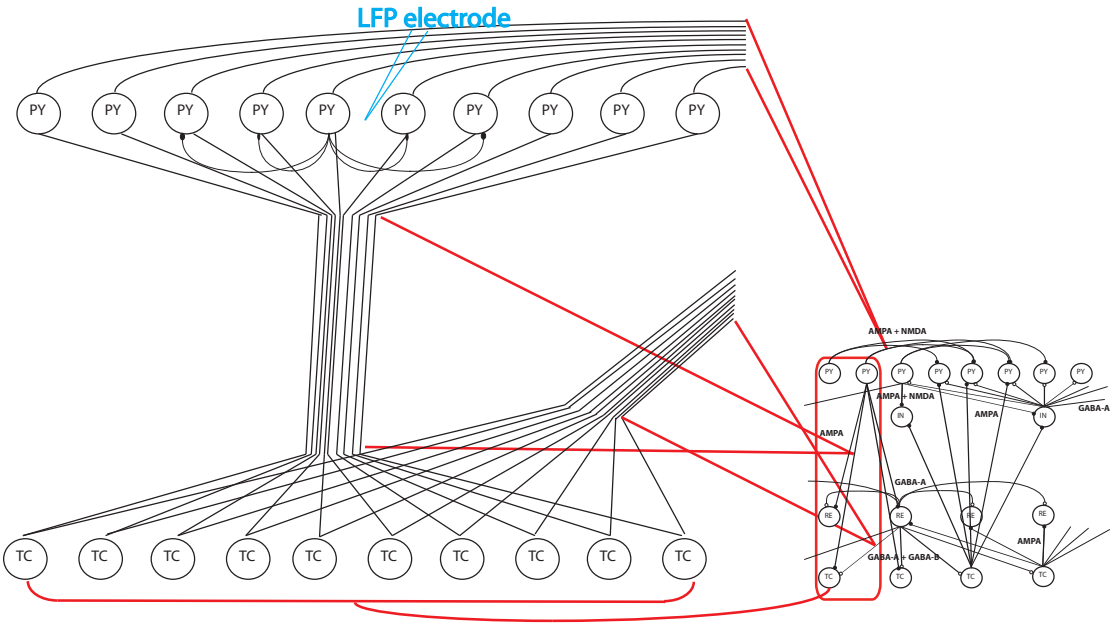

Figure 2: The LFP model was transformed from the thalamocortical model. Shown in this figure, we used the TC $\rightarrow$ PY connection in the red box as an example. Each neuron in the original thalamocortical model was expanded into a sub-network of 10 neurons. Each connection in the original thalamocortical model was transformed into 10 parallel connections between two sub-networks. Moreover, sub-networks transformed from PY neurons have local connections. Each PY neuron in this sub-network connects to its 2 nearest neighbors on each side. We put a LFP electrode at the center of each PY sub-network. The electrode received neurons' signals inversely proportional to the distance.

Table 1: Connectivity properties

	PY→TC	PY→RE	TC→PY	TC→IN	PY→IN	IN→PY	RE→RE	TC→RE	RE→TC
Radius	8	6	15	3	1	5	5	8	8

2.1.3 Local field potential (LFP) model

To simulate local field potential data, we expanded the previous thalamocortical model by 10 times (Fig. 2). Each connection in the previous network (Fig. 1) is transformed into a fiber bundle connecting 10 pairs of neurons in a sub-network. For cortical pyramidal neurons (PY), we connect each neuron to its 4 neighboring neurons inside the sub-network. Other settings for the neurons and the synapses are the same as the previous thalamocortical model. We simulated the network for 600 secs. The data were sampled at 1000 Hz.

We plant a LFP electrode at the center of each of the 50 PY neuron local circuits. The distance between the 10 local neurons is $100\,\mu$m, and the distance between the PY sub-networks is 1 cm. The LFP is calculated according to the "standard model" as previously mentioned in Bonjean et al. [2011], Destexhe [1998], Nunez and Srinivasan [2005]. The LFPs are mainly contributed by elongated dendrites of the cortical pyramidal neurons. In our model, each cortical pyramidal neuron has a 2 mm long dendrite.

For each LFP S_i,

$$S_i = \sum_j \left(\frac{I_{syn}}{r_d^i} - \frac{I_{syn}}{r_s^i}\right)_j \tag{3}$$

Where the sum is taken over all excitatory cortical neurons. I_{syn} is the post-synaptic current of neuron j. r_d is the distance from the electrode to the center of the dendrite of neuron j. r_s is the distance from the electrode to the soma of neuron j.

2.1.4 Calcium imaging model

Vogelstein et al. [2009] proposed a transfer function between spike trains and calcium florescence signals.

$$\begin{aligned} [Ca]_t - [Ca]_{t-1} &= -\frac{\Delta t}{\tau_{Ca}}[Ca]_{t-1} + A_{Ca} n_t \\ F &= \frac{[Ca]}{[Ca]+K_d} + \eta_t \end{aligned} \tag{4}$$

Where $A_{Ca} = 50\,\mu$M is a step influx of calcium molecules at each action potential. n_t is the number of spikes at each time step. $K_d = 300\,\mu$M is the saturation concentration of calcium. η_t is a Gaussian noise with a standard deviation of 0.000003. Since our data is sampled at 1000 Hz, we can resolve every single action potential. So in our data, there is no multiple spikes at one time step. $\tau_{Ca} = 1$ s is the decay constant for calcium molecules. To maintain the information in the differential signal, instead of setting a hard cutoff value and transform the intracellular voltages to binary spike trains, we use a sigmoid function to transform the voltages to the calcium influx activation parameter (n_t).

$$n_t = \frac{1}{1 + e^{-(V(t) - V_{thre})}} \tag{5}$$

Where $V_{thre} = -50$ mV is the threshold potential.

In real experiments, we can only image a small percentage of neurons in the brain. Therefore, we simulated HH-based networks of 1000 neurons and only record from 50 neurons. We used the 4 connection patterns (normal-1 $\sim$ normal-4) provided on-line (`https://www.kaggle.com/c/connectomics`) [Stetter et al., 2012]. Briefly, the networks have 80% excitatory neurons and 20% inhibitory neurons. The connection probability is 0.01, i.e. one neuron connects to about 10 neurons, so it is a sparsely connected network.

Similar to our thalamocortical model, we used AMPA and NMDA synapses for the excitatory synapses, and GABA synapses for the inhibitory synapses. The simulations ran 600secs and were sampled at 1000Hz. The intracellular voltages obtained from the simulations were then transferred to calcium florescence signals and down sampled to 50 Hz. For each of the 4 networks, we conducted 25 recordings. Each recording contains 50 randomly selected neurons' calcium signals.

For accurate estimations, the differential covariance-based methods require the reconstructed action potentials from the calcium imaging. While this is an active research area and many methods have been proposed [Quan et al., 2010, Rahmati et al., 2016], In this study, we simply reversed the transfer function. By assuming the transfer function from action potentials to calcium fluorescence signals is known, we can reconstruct the action potentials as:

Given $\hat{F}$ as the observed fluorescence signal

$$\begin{aligned} [\hat{Ca}] &= \hat{F} * K_d/(1-\hat{F}) \\ \hat{n_t} &= (d[\hat{Ca}] + 1/\tau_{Ca}[Ca]_t)/(A/\Delta t) \\ \hat{V} &= log(1/n_t - 1) \end{aligned} \tag{6}$$

2.2 Differential covariance-based Methods

In this section, we introduce a new class of methods to estimate the functional connectivity of neurons (code is provided on-line at `https://github.com/tigerwlin/diffCov`).

2.2.1 Step1: differential covariance

The input to the method, $S(t)$, is a $N \times T$ neural recording dataset. N is the number of neurons/channels recorded, and T is the number of data samples during recordings. We compute the derivative of each time series with $dS(t) = (S(t+1) - S(t-1))/(2dt)$. Then, the covariance between S(t) and dS(t) is computed and denoted as Δ_C, which is a $N \times N$ matrix defined as the following:

$$\Delta_{C_{i,j}} = cov(dS_i(t), S_j(t)) \tag{7}$$

where $dS_i(t)$ is the differential signal of neuron/channel i, $S_j(t)$ is the signal of neuron/channel j, and $cov()$ is the commonly used covariance function for two time series.

2.2.2 Step 2: applying partial covariance method

As previously mentioned in Stevenson et al. [2008], one problem of the correlation method is the propagation of correlation. Here we designed a customized partial covariance algorithm to reduce this type of error in our methods. We use $\Delta_{P_{i,j}}$ to denote the estimation after applying the partial covariance method.

Using the derivation from appendix B.2, we have:

$$\Delta_{P_{i,j}} = \Delta_{C_{i,j}} - COV_{j,Z} * COV_{Z,Z}^{-1} * \Delta_{C_{i,Z}}^{T} \tag{8}$$

Where Z is a set of all neurons/channels except $\{i, j\}$.

$$Z = \{1, 2, ..., i-1, i+1, ..., j-1, j+1, ...N\}$$

$\Delta_{C_{i,j}}$ and $\Delta_{C_{i,Z}}$ were computed from the previous step, and COV is the covariance matrix.

Note in this case, Δ_P is not equivalent to the precision matrix Δ_C^{-1} explained in appendix B.2. The two are only equivalent for the correlation matrix and when the partial correlation is controlling on all variables in the observed set. In our case, the differential covariance matrix is non-symmetric because it is the covariance between recorded signals and their differential signals. We have signals and differential signals in our observed set, however, we are only controlling on the original signals for the partial covariance algorithm. Due to these differences, we developed this customized partial covariance algorithm (Eq. 8), which performs well for neural signals in the form of Eq. 2.

2.2.3 Step 3: sparse latent regularization

Finally, we applied the sparse latent regularization method to our estimation [Chandrasekaran et al., 2011, Yatsenko et al., 2015]. As explained in appendix B.3, in the sparse latent regularization method, we made the assumption that there are observed neurons and unobserved common inputs in a network. If the connections between the observed neurons are sparse and the number of unobserved common inputs is small, this method can separate the covariance into two parts and the sparse matrix is the intrinsic connections between the observed neurons.

Here we define Δ_S as the sparse result from the method, and L as the low-rank result from the method.

Then by solving:

$$\underset{\Delta_S, L}{\arg\min} \, ||\Delta_S||_1 + \alpha * tr(L) \tag{9}$$

under the constraint that

$$\Delta_P = \Delta_S + L \tag{10}$$

Where, $||\ ||_1$ is the L1-norm of a matrix, and $tr()$ is the trace of a matrix. α is the penalty ratio between the L1-norm of Δ_S and the trace of L. It was set to $1/\sqrt{N}$ for all our estimations. Δ_P is the partial differential covariance computed from section 2.2.2. We receive a sparse estimation, Δ_S, of the connectivity.

2.3 Performance quantification

The performance of each method is judged by 4 quantified values. The first 3 values indicate the method's abilities to reduce the 3 types of false connections (Fig. 4). The last one indicates the method's ability to correctly estimate the true positive connections against all possible interference.

Let's define G as the ground truth connectivity matrix, where:

$$G_{i,j} = \begin{cases} 1, & \text{if neuron i projects to neuron j with excitatory synapse} \\ -1, & \text{if neuron i projects to neuron j with inhibitory synapse} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Then, we can use a 3-dimensional tensor to represent the false connections caused by common inputs. For example, neuron j and neuron k receive common input from neuron i:

$$M_{i,j,k} = \begin{cases} 1, & \text{iff } G_{i,j} = 1 \text{ and } G_{i,k} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Therefore, we can compute a mask that labels all the type 1 false connections:

$$mask_{1_{j,k}} = \sum_{i \in \{\text{observed neurons}\}} M_{i,j,k} \tag{13}$$

For the type 2 false connections (e.g. neuron i projects to neuron k, then neuron k projects to neuron j), the mask is defined as:

$$mask_{2_{i,j}} = \sum_{k \in \{\text{observed neurons}\}} G_{i,k} G_{k,j} \tag{14}$$

or, in simple matrix notation:

$$mask_2 = G \cdot G \tag{15}$$

Similar to $mask_1$, the false connections caused by unobserved common inputs is:

$$mask_{3_{j,k}} = \sum_{i \in \{\text{unobserved neurons}\}} M_{i,j,k} \tag{16}$$

Lastly, $mask_4$ is defined as:

$$mask_{4_{i,j}} = \begin{cases} 1, & \text{if } G_{i,j} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

Given a connectivity matrix estimation result: Est, the 4 values for the performance quantification are computed as the area under the ROC curve for two sets: the true positive set and the false positive set.

$$\begin{aligned} |Est_{i,j}| &\in \{\text{true positive set}\}_k, \text{if } G_{i,j} \neq 0 \text{ and } mask_{k_{i,j}} = 0 \\ |Est_{i,j}| &\in \{\text{false positive set}\}_k, \text{if } mask_{k_{i,j}} \neq 0 \text{ and } G_{i,j} = 0 \end{aligned} \tag{18}$$

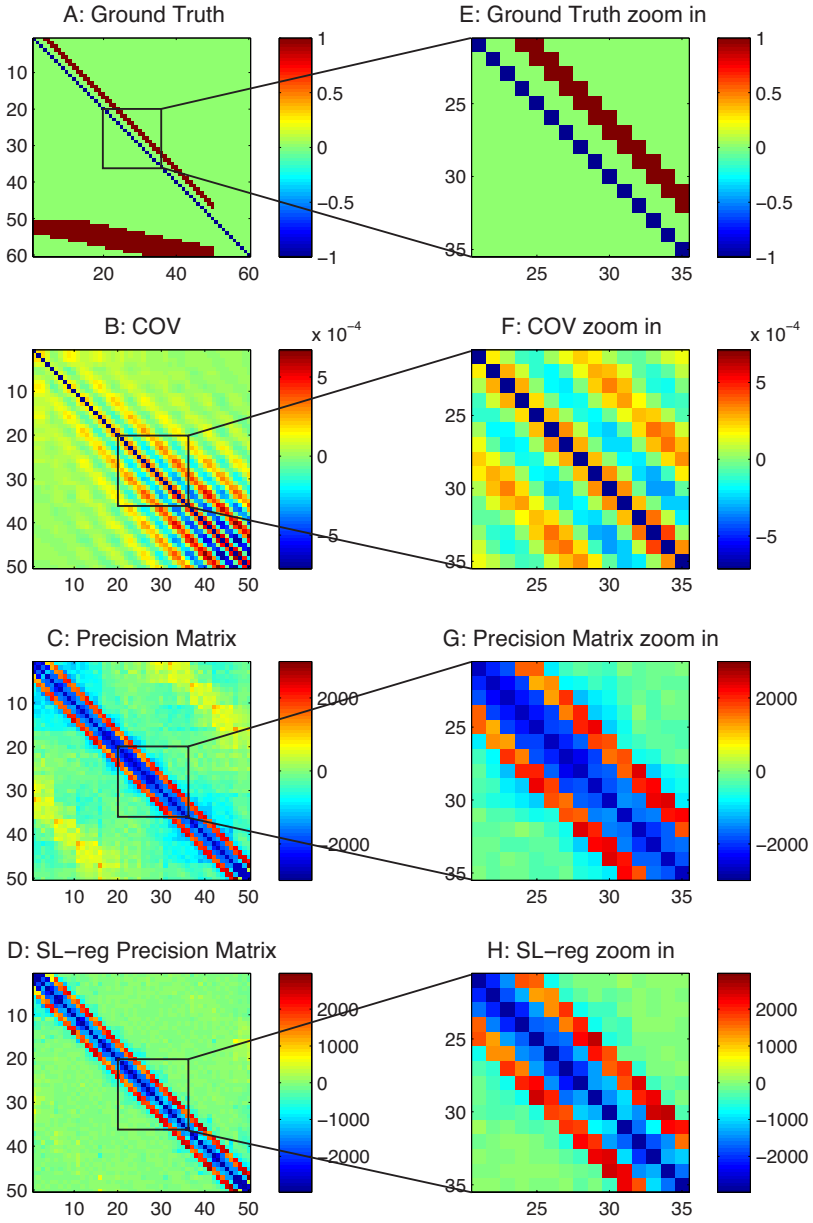

Figure 3: Ground truth connections from a passive neuron model and estimations from correlation-based methods. A) Ground truth connection matrix. neurons 1-50 are observed neurons. neurons 51-60 are unobserved neurons that introduce common inputs. B) Estimation from the correlation method. C) Estimation from the precision matrix. D) Estimation from the sparse+latent regularized precision matrix. E) Zoom in of panel A. F) Zoom in of panel B. G) Zoom in of panel C. H) Zoom in of panel D.

3 Results

3.1 False connections in correlation-based methods

When applied to neural circuits, the commonly used correlation-based methods produce systematic false connections. As shown, Fig. 3 A is the ground truth of the connections in our passive neuron model (Neurons 1-50 are the observed neurons). Fig. 3B is from the correlation method, Fig. 3C is the precision matrix, and Fig. 3D is the sparse+latent regularized precision matrix. As shown, all of these methods produce extra false connections.

For the convenience of explanation, we define the diagonal strip of connections in the ground truth (first 50 neurons in Fig. 3A) as the -3 and -4 diagonal lines, because they are 3 and 4 steps away from the diagonal line of the matrix. As shown in Fig. 3, all these methods produce false connections on the ± 1 diagonal lines. The precision matrix method (Fig. 3C) also has square box shape false connections in the background.

3.1.1 Type 1 false connections

Shown in Fig. 4A, the type 1 false connections are produced because two neurons receive the same input from another neuron. The same synaptic current that passes into the two neurons generates positive correlation between the two neurons. However, there is no physiological connection between these two neurons. In our connection pattern (Fig. 3 A), we notice that two neurons next to each other receive common synaptic inputs, therefore there are false connections on the ± 1 diagonal lines of the correlation-based estimations.

3.1.2 Type 2 false connections

Shown in Fig. 4B, the type 2 false connections are due to the propagation of correlation. Because one neuron V_A connects to another neuron V_B and neuron V_B connects to another neuron V_C, the correlation method presents correlation between V_A and V_C, which do not have a physical connection. This phenomenon is shown in Fig. 3B as the extra diagonal strips. Shown in Fig. 3C, this problem can be greatly reduced by the precision matrix/partial covariance method.

3.1.3 Type 3 false connections

Shown in Fig. 4C, the type 3 false connections are also due to the common currents pass into two neurons. However, in this case, they are from the unobserved neurons. For this particular passive neuron model, it is due to the inputs from the 10 unobserved neurons (Neurons 51-60) as shown on Fig. 3A. Because the latent neurons have broad connections to the observed neurons, they introduce a square box shape correlation pattern into the estimations (Fig. 3C. Fig. 3B also contains this error, but it is hard to see). Since, the latent neurons are not observable, partial covariance method cannot be used to regress out this type of correlation. On the other hand, the sparse latent regularization can be applied if the sparse and low-rank assumption is valid, and the

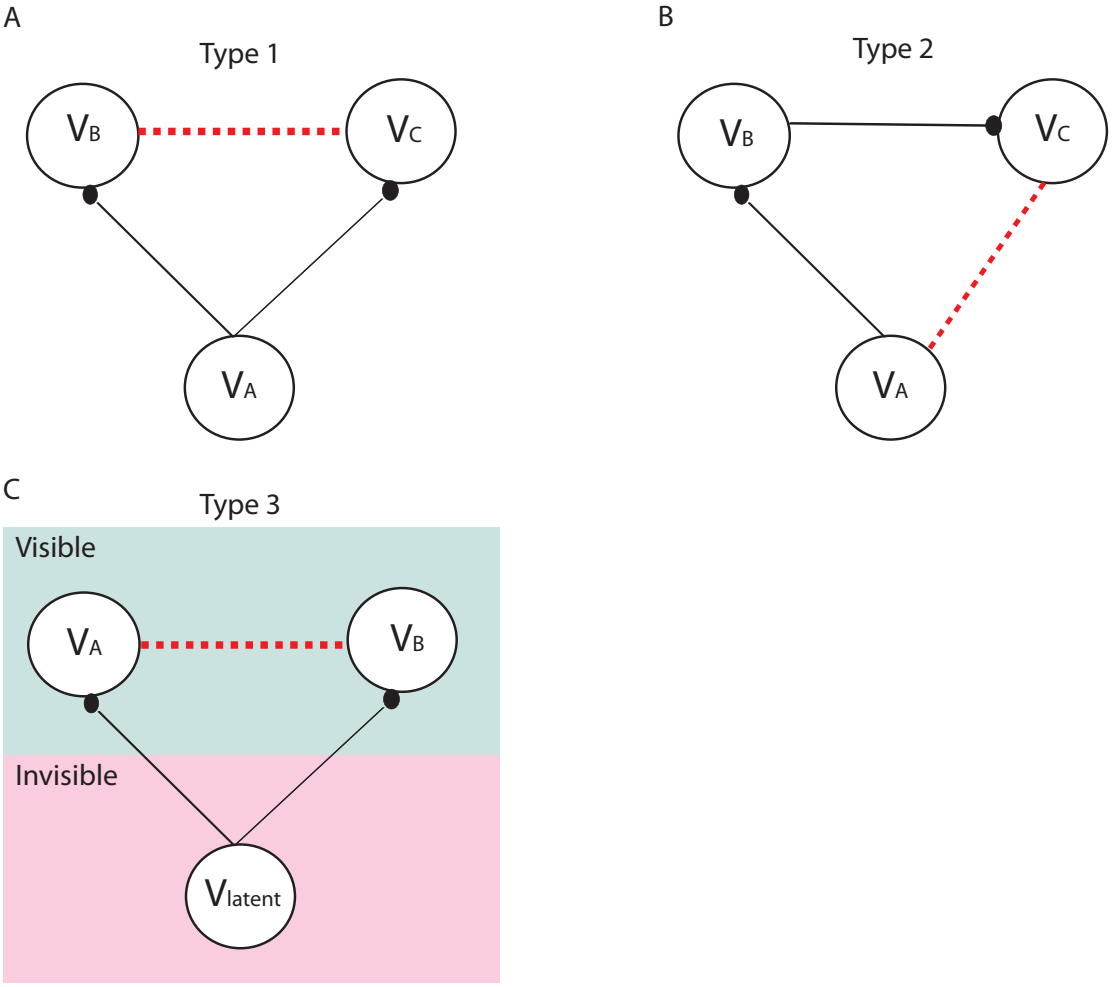

Figure 4: Illustrations of the 3 types of false connections in the correlation-based methods. Solid lines indicate the physical wiring between neurons, and the black circles at the end indicate the synaptic contacts (i.e. the direction of the connections). The dotted lines are the false connections introduced by the correlation-based methods. A) Type 1 false connections, which are due to two neurons receiving the same synaptic inputs. B) Type 2 false connections, which are due to the propagation of correlation. C) Type 3 false connections, which are due to unobserved common inputs.

sparse+latent regularized result is shown in Fig. 3D. However, even after using this regularization, the correlation-based methods still leave false connections in Fig. 3D.

3.2 Estimations from differential covariance-based methods

Table 2: Performance quantification (area under the ROC curve) of different methods with respect to their abilities to reduce the 3 types of false connections and their abilities to estimate the true positive connections under 5 different passive neuron model settings.

	Cov	Precision	Precision+SL-reg	Δ_C	Δ_P	Δ_S
cxcx34 g5						
Error 1	0	0	0	0.6327	0.3469	0.8776
Error 2	0.1469	0.9520	0.9915	0.3757	0.8347	1.0000
Error 3	0.4638	0.9362	0.9797	0.6541	0.8391	0.9986
True Positive	0.7312	1.0000	1.0000	0.9677	0.9946	1.0000
cxcx34 g30						
Error 1	0	0	0	0.0510	0.5816	0.9490
Error 2	0.0056	0.8927	0.9972	0.2881	0.9548	1.0000
Error 3	0.2164	0.9188	0.9942	0.5430	0.9662	1.0000
True Positive	0.5591	1.0000	0.9892	0.6559	1.0000	1.0000
cxcx34 g50						
Error 1	0	0	0	0	0.2041	0.6531
Error 2	0	0.7034	0.9944	0.0523	0.9054	1.0000
Error 3	0.3179	0.8000	0.9894	0.4145	0.9309	1.0000
True Positive	0.9140	1.0000	0.9946	0.9516	0.9785	1.0000
cxcx56789 g5						
Error 1	0	0.0053	0.0053	0.6895	0.6263	0.8526
Error 2	0.1896	0.6229	0.7896	0.5240	0.7896	0.9938
Error 3	0.3573	0.6085	0.7659	0.6957	0.7591	0.9817
True Positive	0.6884	0.9442	0.6674	0.9930	0.9605	0.9837
cxcx56789 g50						
Error 1	0	0	0	0.0263	0.2816	0.6842
Error 2	0.1083	0.5312	0.8240	0.2844	0.6990	0.9979
Error 3	0.4256	0.4927	0.7762	0.5091	0.7116	0.9835
True Positive	0.9256	0.9116	0.6698	0.9395	0.9279	0.9419

Comparing the ground truth connections in Fig. 5A with our final estimation in Fig. 5D, we see that our methods essentially transformed the connections in the ground truth into a map of sources and sinks in a network. An excitatory connection, $i \to j$, in our estimations have negative value for $\Delta_{S_{ij}}$ and positive value for $\Delta_{S_{ji}}$, which means the current is coming out of the source i, and goes into the sink j. We note that there is another ambiguous case, an inhibitory connection $j \to i$, which produces the same results in our estimations. Our methods can not differentiate these two cases, instead, they indicate sources and sinks in a network.

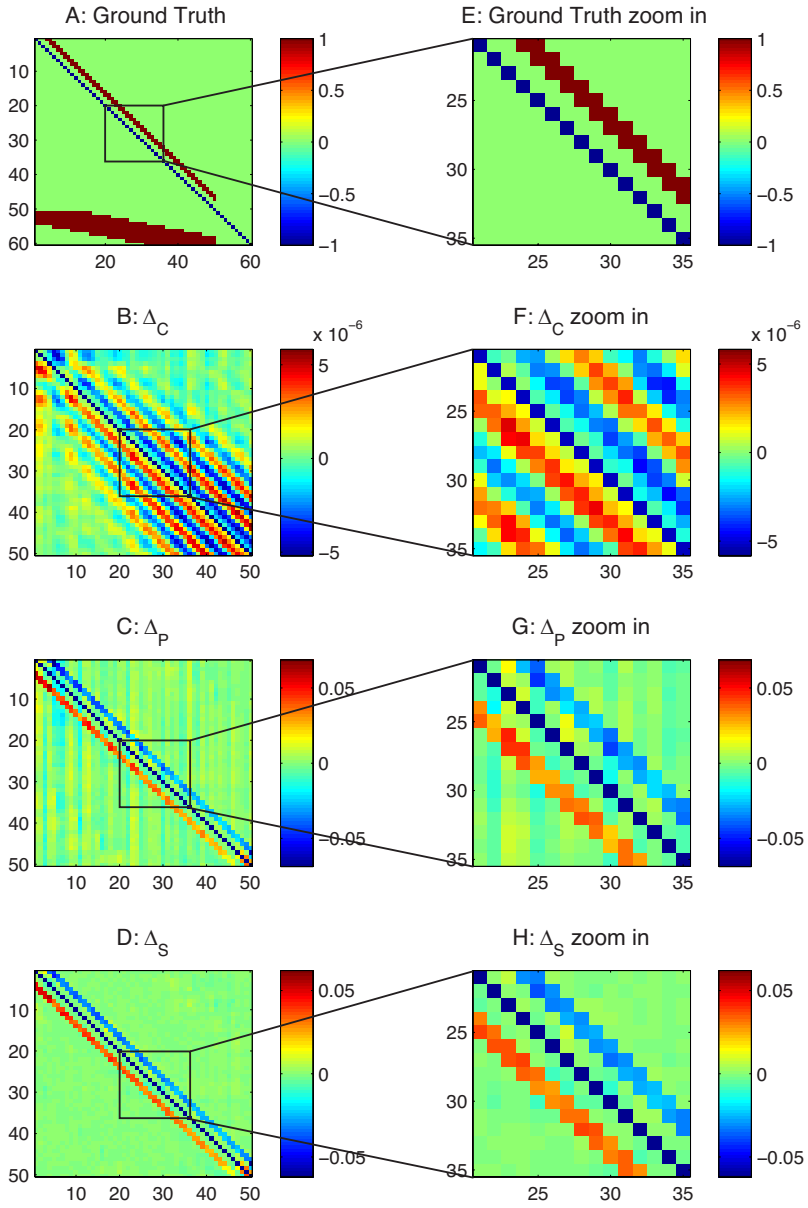

Figure 5: differential covariance analysis of the passive neuron model. The color in B,C,D,F,G,H indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method. E) Zoom in of panel A. F) Zoom in of panel B. G) Zoom in of panel C. H) Zoom in of panel D.

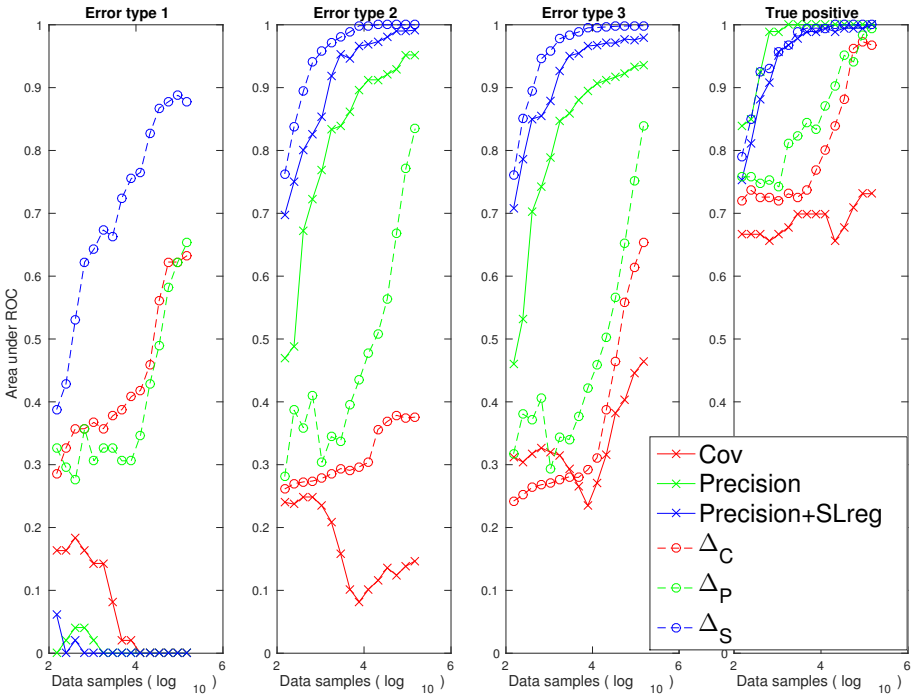

Figure 6: Performance quantification (area under the ROC curve) of different methods with respect to their abilities to reduce the 3 types of false connections and their abilities to estimate the true positive connections using the passive neuron dataset.

3.2.1 The differential covariance method reduces type 1 false connections

By comparing Fig. 3 B with Fig. 5 B, we see that the type 1 false connections on the ± 1 diagonal lines of Fig. 3 B is reduced in Fig. 5 B. This is reflecting the theorems we proved in appendix A, in particular theorem 5, which shows that the strength of the type 1 false connections is reduced in the differential covariance method by a factor of g_l/g_{syn}. Moreover, the differential covariance method's performance on reducing the type 1 false connections is quantified in Fig. 6.

3.2.2 The partial covariance method reduces type 2 false connections

Second, we see that, due to the propagation of correlation, there are extra diagonal strips in Fig. 5B. These are removed in Fig. 5C by applying the partial covariance method. And each estimator's performance for reducing type 2 false connections is quantified in Fig. 6.

3.2.3 The sparse+latent regularization reduces type 3 false connections

Third, we use the sparse+latent regularization to remove the correlation introduced by the latent inputs. As mentioned in the method section, when the observed neurons' connections are sparse and the number of unobserved common inputs is small, the covariance introduced by the unobserved common inputs can be removed. As shown in Fig. 5D, the external covariance in the background of Fig. 5C is removed, while the true diagonal connections and the directionality of the connections are maintained. This regularization is also effective for correlation-based methods, but type 1 false connections maintain in the estimation even after applying this regularization (Fig. 3D). Each estimator's performance for reducing type 3 false connections is quantified in Fig. 6.

3.2.4 The differential covariance-based methods provide directionality information of the connections

Using this passive neuron model, in appendix A.3, we provide a mathematical explanation for why the differential covariance-based methods provide directional information for the connections. Given an excitatory connection $g_{i\to j}$ (neuron i projects to neuron j), from Theorem 6 in appendix A.3, we have:

$$\begin{aligned} \Delta_{C_{j,i}} &> 0 \\ \Delta_{C_{i,j}} &< 0 \end{aligned} \tag{19}$$

However, we wish to note here that, there is another ambiguous setting that provides the same result, which is an inhibitory connection $g_{j\to i}$. Conceptually, the differential covariance indicates the current sources and sinks in a neural circuit, but the exact type of synapse is unknown.

3.2.5 Performance quantification for the passive neuron dataset

In Figure 6, we quantified the performance of the estimators for one example dataset. We see that, with the increase of the sample size, our differential covariance-based

methods reduce the 3 types of false connections, while maintaining high true positive rates. Also as we apply more advanced techniques ($\Delta_C \rightarrow \Delta_P \rightarrow \Delta_S$), the estimator's performance increases in all 4 panels of the quantification indices. Although the precision matrix and the sparse+latent regularization help the correlation method reduce the type 2 and type 3 error, all correlation-based methods handle poorly of the type 1 false connections. We also note that the masks we used to quantify each type of false connections are not mutually exclusive (i.e. there are false connections that belong to more than one type of false connections). Therefore, in Figure 6, it seems like a regularization is reducing multiple types of false connections. For example, the sparse+latent regularization is reducing both type 2 and type 3 false connections.

In Table 2, we provide quantified results (area under the ROC curve) for two different connection patterns (cxcx34, and cxcx56789) and three different conductance settings (g5, g30, and g50). We see that, the key results in Fig. 6 are also generalized here. By applying more advanced techniques to the original differential covariance estimator ($\Delta_C \rightarrow \Delta_P \rightarrow \Delta_S$, the performance increases with respect to the 3 types of error, while the true positive rate is not sacrificed. We also note that, although the precision matrix and the sparse+latent regularization help the correlation method reduce the type 2 and the type 3 error, all correlation-based methods handle poorly of the type 1 error.

3.3 Thalamocortical model results

We further tested the methods in a more realistic Hodgkin-Huxley based model. Because the synaptic conductances in the Hodgkin-Huxley model are no longer constants but become nonlinear dynamic functions, which depend on the pre-synaptic voltages, the derivations above can only be considered as a first-order approximation.

Shown in Fig. 7A is the ground truth connections between the cortical neurons. These cortical neurons also receive latent inputs from and sending feedback currents to inhibitory neurons in the cortex (IN) and thalamic neurons (TC). For clarity of representation, these latent connections are not shown here, but the detailed connections are described in the Method section.

Similar to the passive neuron model, in Fig. 7B, the correlation method still suffers from those 3 types of false connections. As shown, the latent inputs generate false correlations in the background. And the ± 1 diagonal line false connections, which are due to the common currents, exist in all correlation-based methods (see Fig. 7B,C,D). Comparing Fig. 7C, D, because the type 1 false connections are strong in the Hodgkin-Huxley based model, the sparse+latent regularization removed the true connections but kept these false connections in its final estimation.

As shown in Fig. 8B, differential covariance method reduces the type 1 false connections. Then in Fig. 8C, the partial differential covariance method reduces type 2 false connections in Fig. 8B (yellow color connections around the red strip in Fig. 8B). Finally, in Fig. 8D, the sparse latent regularization removes the external covariance in the background of Fig. 8C. The current sources (positive value, red color) and current sinks (negative value, blue color) in the network are also indicated on our estimators.

In Fig. 9, each estimator's performance on each type of false connections is quantified. In this example, our differential covariance-based methods achieve similar perfor-

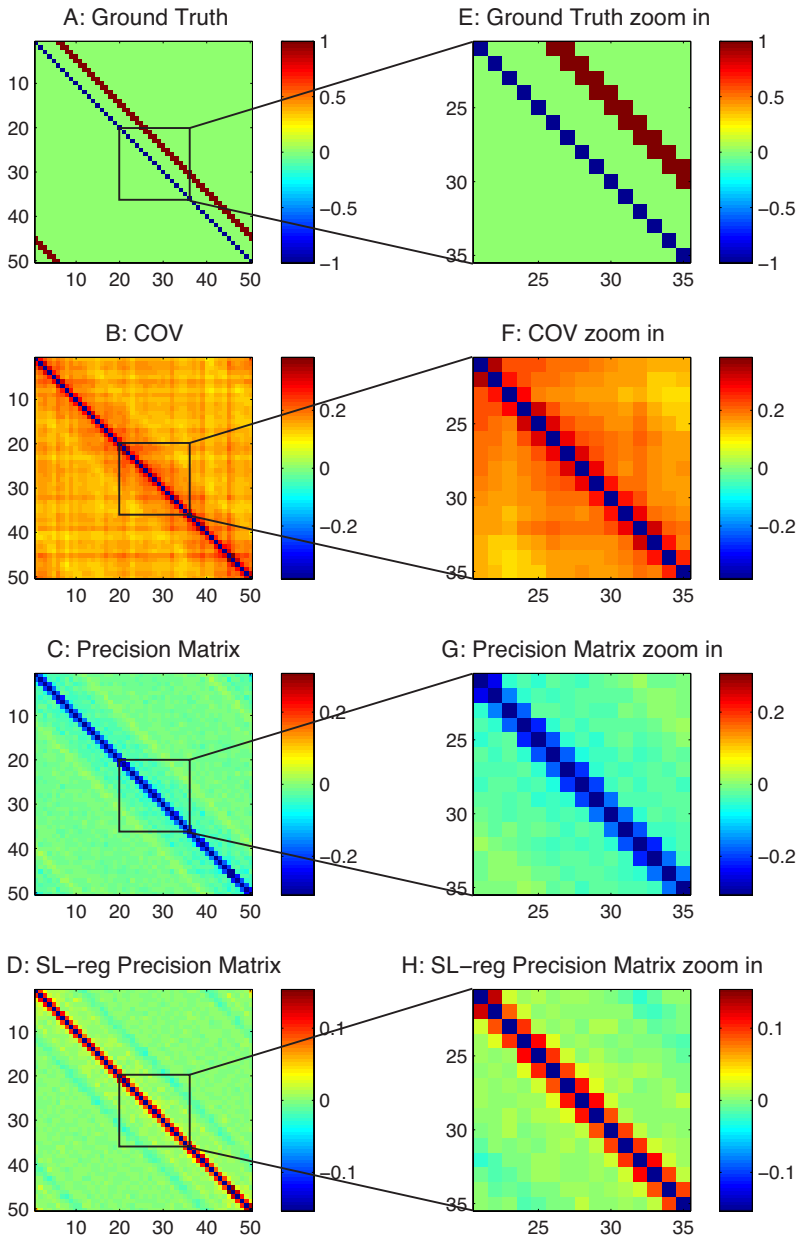

Figure 7: Analysis of the thalamocortical model with correlation-based methods. A) Ground truth connections of the PY neurons in the thalamocortical model. B) Estimation from the correlation method. C) Estimation from the precision matrix method. D) Estimation from the sparse+latent regularized precision matrix method. E) Zoom in of panel A. F) Zoom in of panel B. G) Zoom in of panel C. H) Zoom in of panel D.

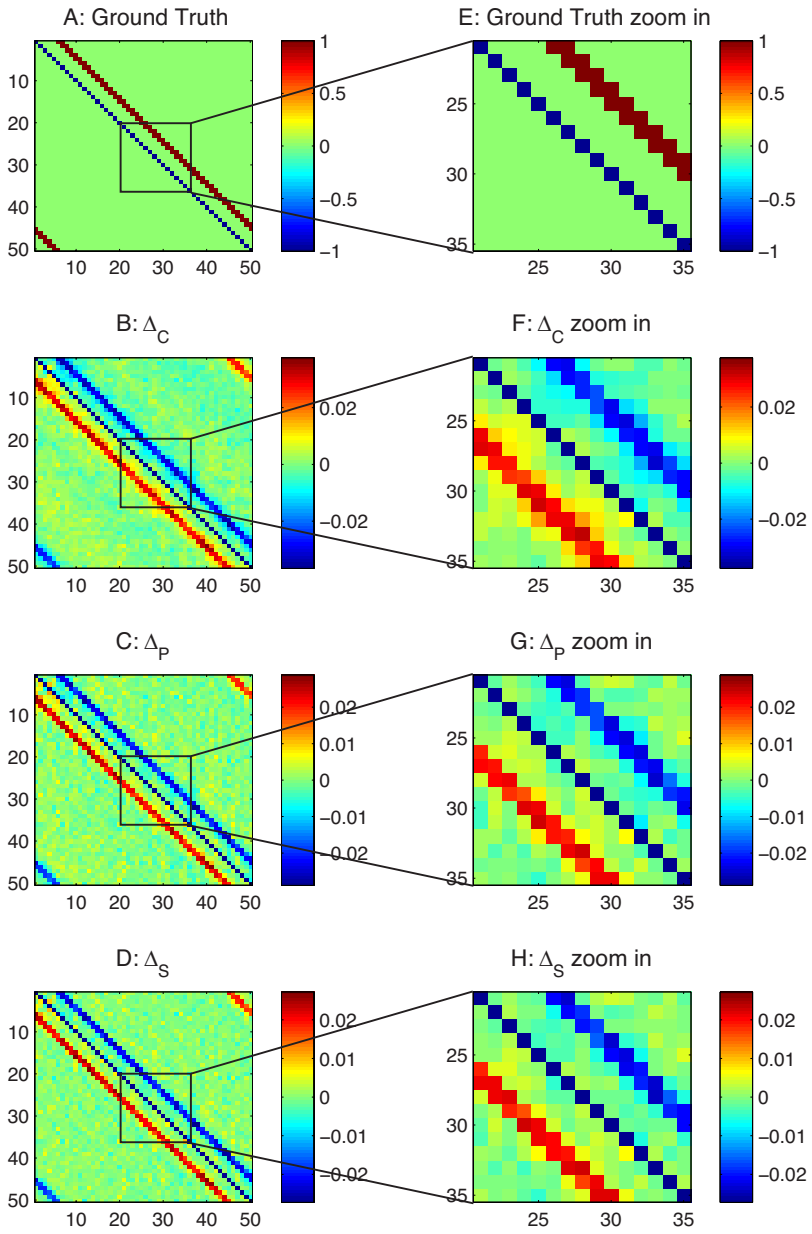

Figure 8: Analysis of the thalamocortical model with differential covariance-based methods. The color in B,C,D,F,G,H indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method. E) Zoom in of panel A. F) Zoom in of panel B. G) Zoom in of panel C. H) Zoom in of panel D.

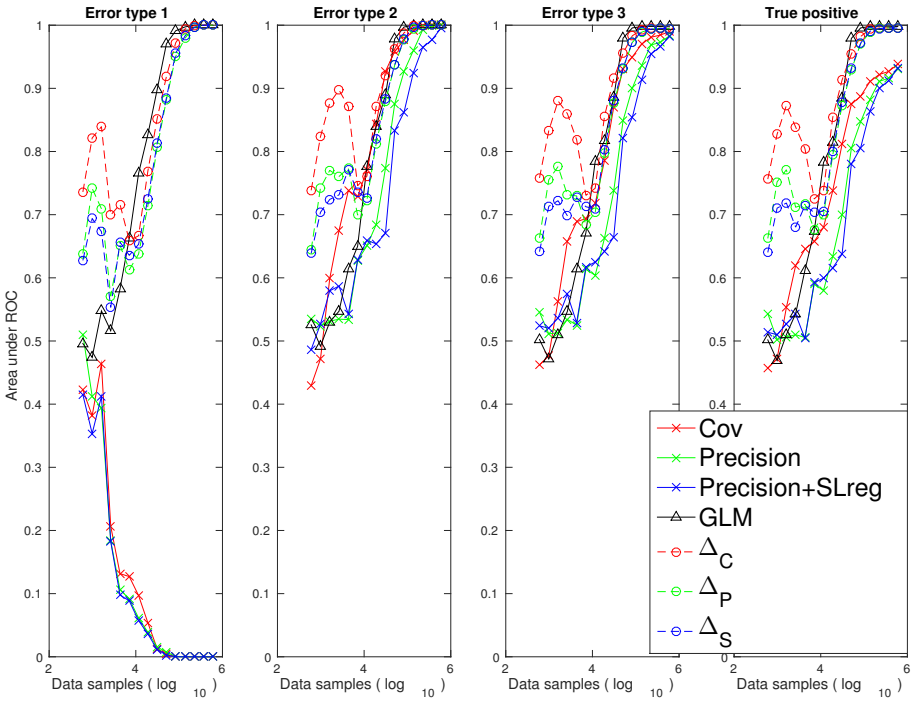

Figure 9: Performance quantification (area under the ROC curve) of different methods with respect to their abilities to reduce the 3 types of false connections and their abilities to estimate the true positive connections using the thalamocortical dataset.

mance to the GLM method.

3.4 LFP results

For population recordings, our methods have similar performance to the thalamocortical model example. While the correlation-based methods still suffering from the problem of type 1 false connections (Fig. 10), our differential covariance-based methods can reduce all 3 types of false connections (Fig. 11). In Fig. 12, each estimator's performance on LFP data is quantified. In this example, with sufficient data samples, our differential covariance-based methods achieve similar performance to the GLM method. However, for smaller sample sizes, our new methods perform better than the GLM method.

3.5 Calcium imaging results

Lastly, because current techniques only allow recording of a small percentage of neurons in the brain, we tested our methods on a calcium imaging dataset of 50 neurons recorded from 1000 neurons networks. In this example, our differential covariance-based methods (Fig. 14) match better with the ground truth than the correlation-based methods (Fig. 13).

In Fig. 15, We performed 25 sets of recordings with 50 neurons randomly selected in each of the 4 large networks and quantified the results. The markers on the plots are the average area under the ROC curve values across the 100 sets, and the error bars indicate the standard deviations across these 100 sets of recordings. Our differential covariance-based methods perform better than the GLM method, and the performance differences seem to be greater in situations with less data samples.

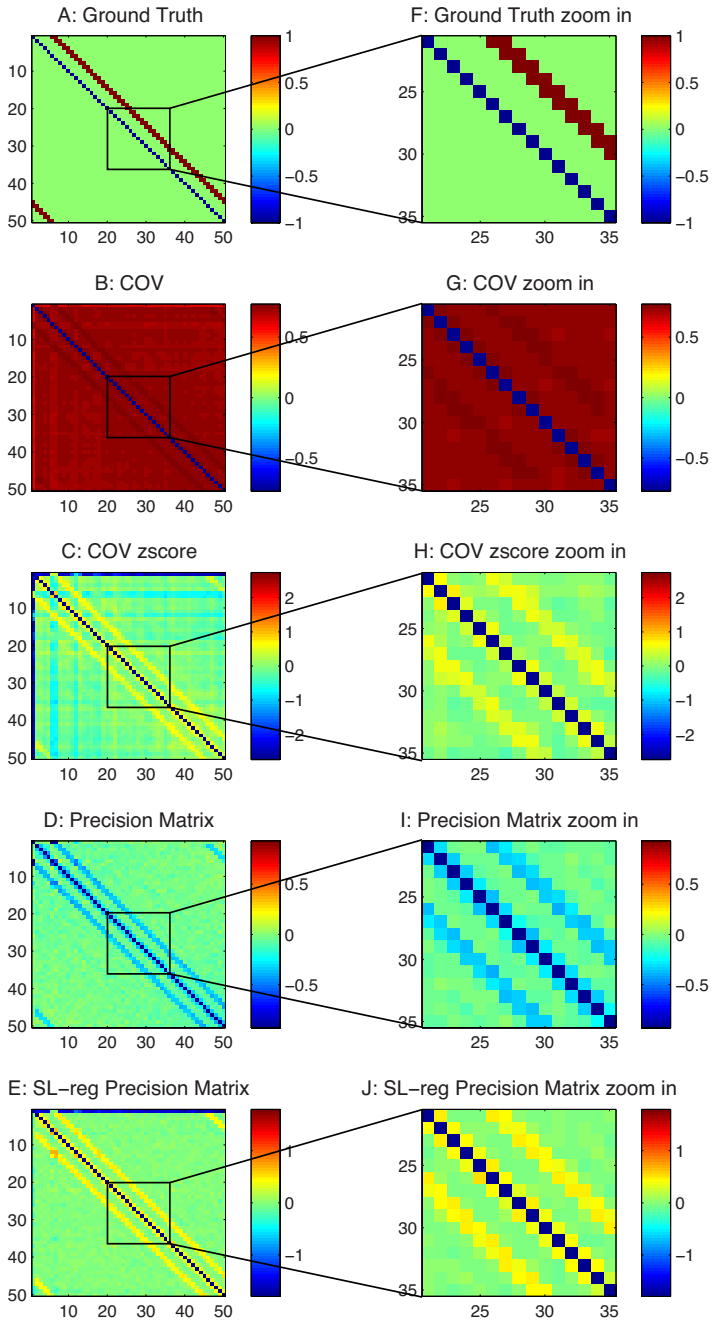

Figure 10: Analysis of the LFP model with correlation-based methods. A) Ground truth connection matrix B) Estimation from the correlation method. C) z-score of the correlation matrix. D) Estimation from the precision matrix method. E) Estimation from the sparse+latent regularized precision matrix method. F) Zoom in of panel A. G) Zoom in of panel B. H) Zoom in of panel C. I) Zoom in of panel D. J) Zoom in of panel E.

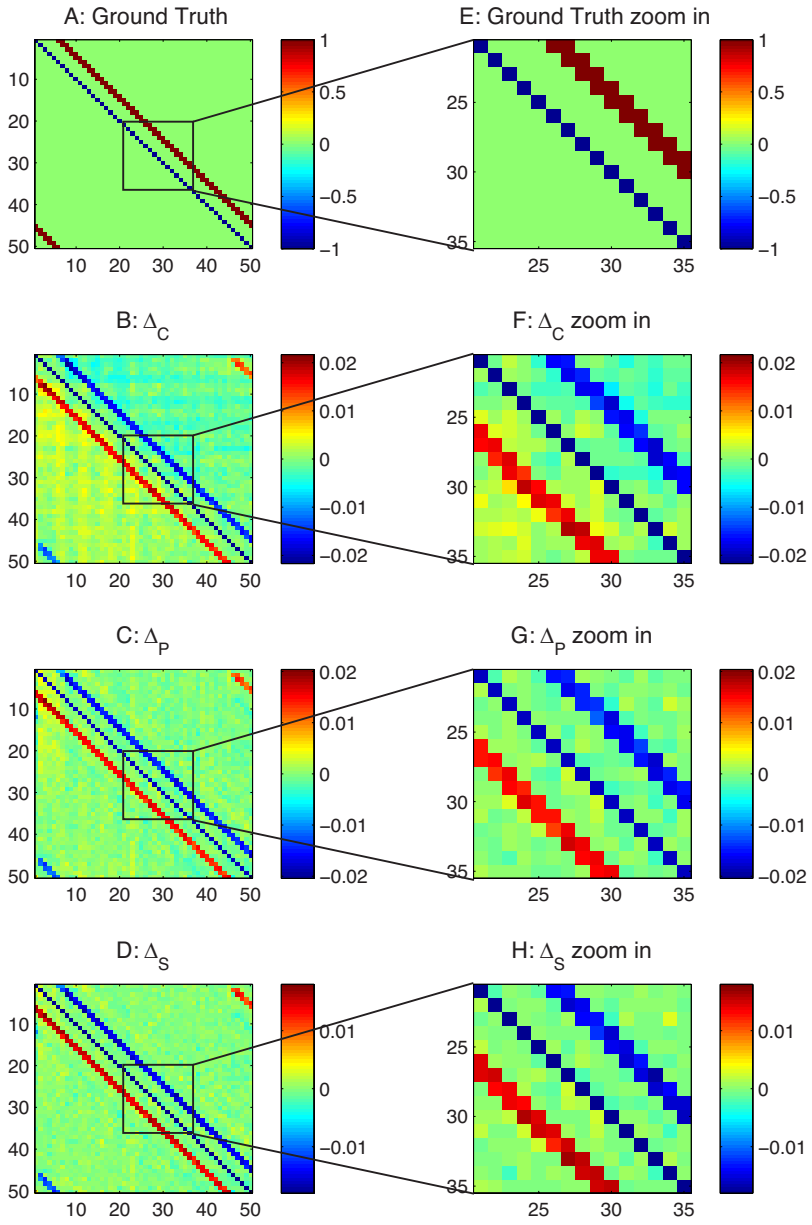

Figure 11: Analysis of the LFP model with differential covariance-based methods. The color in B,C,D,F,G,H indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method. E) Zoom in of panel A. F) Zoom in of panel B. G) Zoom in of panel C. H) Zoom in of panel D.

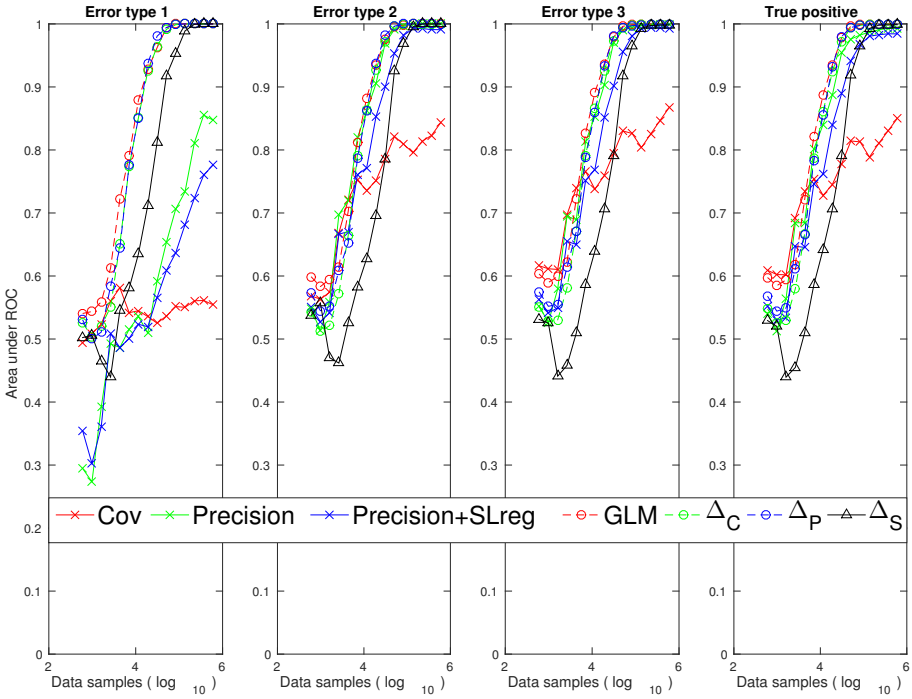

Figure 12: Performance quantification (area under the ROC curve) of different methods with respect to their abilities to reduce the 3 types of false connections and their abilities to estimate the true positive connections using the LFP dataset.

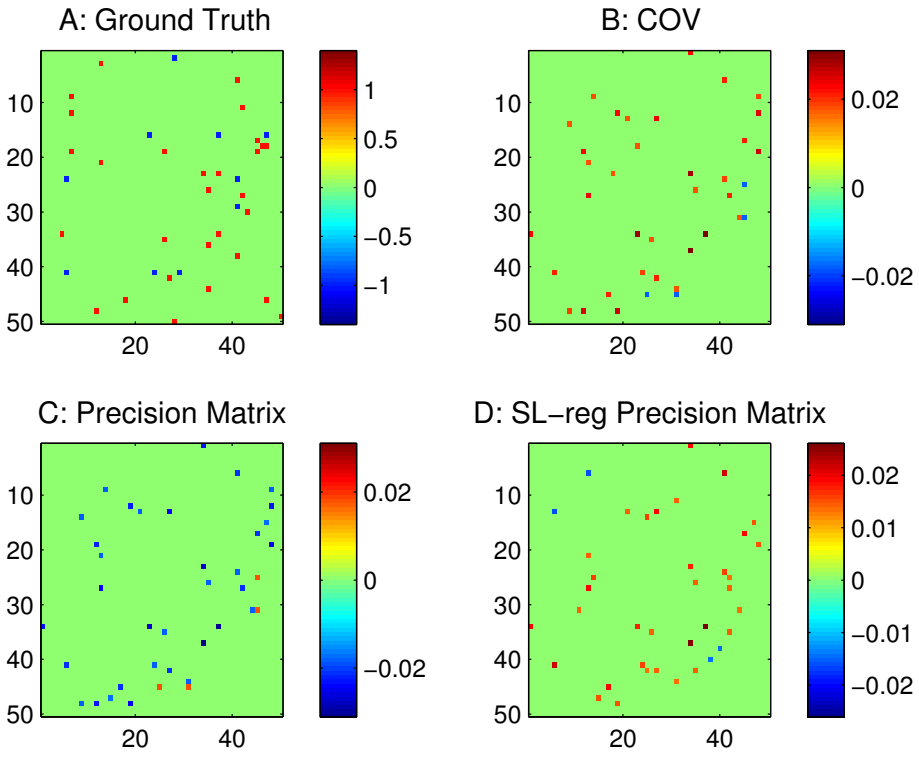

Figure 13: Analysis of the calcium imaging dataset with correlation-based methods. A) Ground truth connection matrix B) Estimation from the correlation method. C) Estimation from the precision matrix method. D) Estimation from the sparse+latent regularized precision matrix method. For clarity purpose, panel B,C,D are thresholded to show only the most strong connections, so one can compare it with the ground truth.

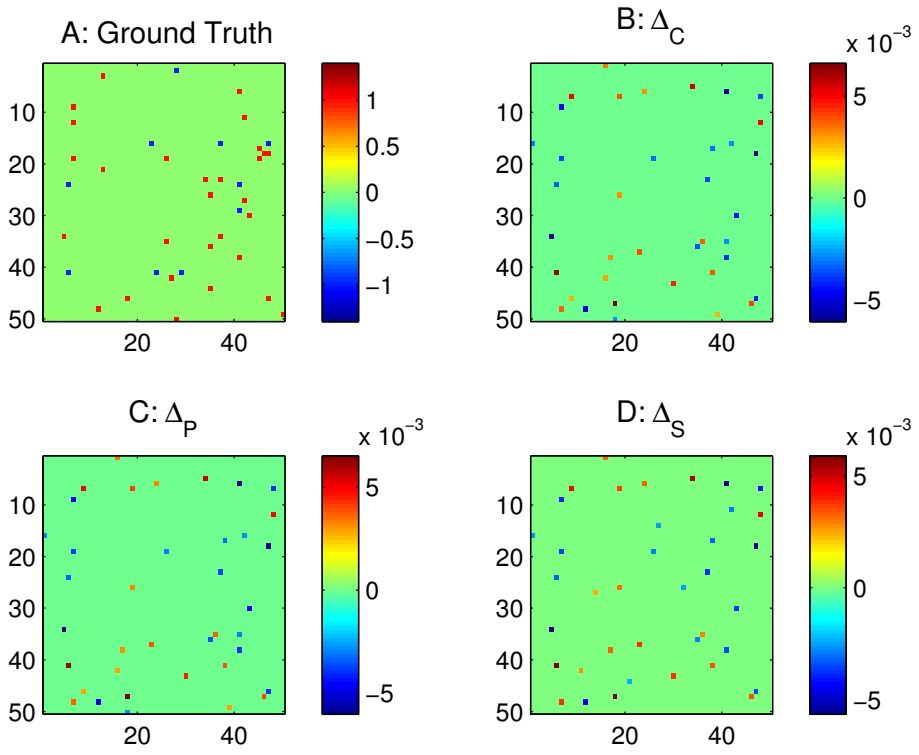

Figure 14: Analysis of the calcium imaging dataset with differential covariance-based methods. The color in B,C,D indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method. For clarity purpose, panel B,C,D are thresholded to show only the most strong connections, so one can compare it with the ground truth.

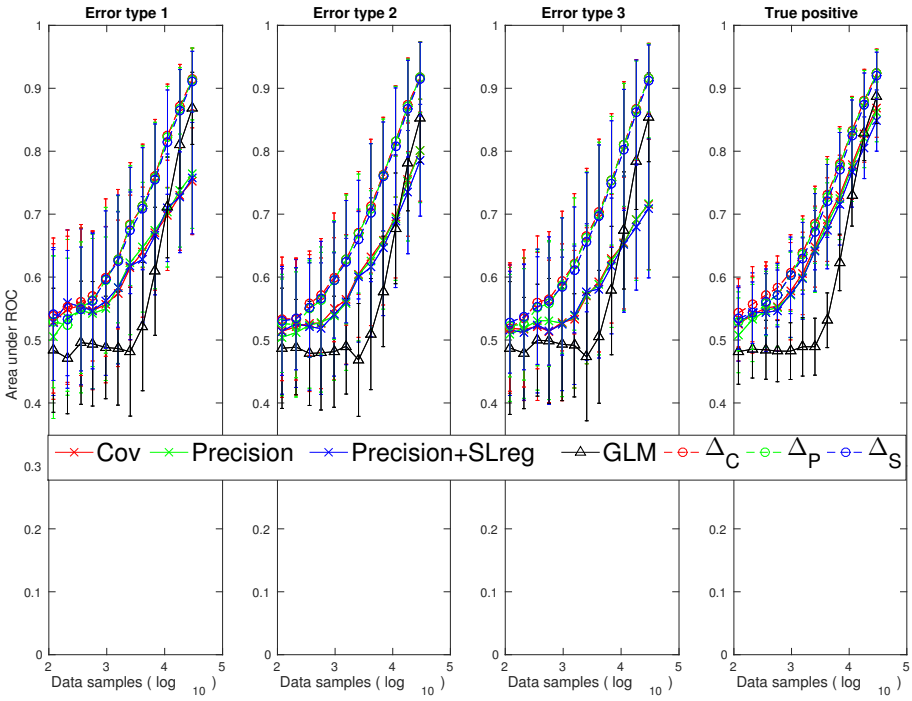

Figure 15: Performance quantification (area under the ROC curve) of different methods with respect to their abilities to reduce the 3 types of false connections and their abilities to estimate the true positive connections using the calcium imaging dataset. Error bar is the standard deviation across 100 sets of experiments. Each experiment randomly recorded 50 neurons in a large network. The markers on the plots indicate the average area under the ROC curve values across the 100 sets of experiments.

4 Discussion

4.1 Generalizability and applicability of the differential covariance-based methods to real experimental data

Many methods have been proposed to solve the problem of reconstructing the connectivity of a neural network. Winterhalder et al. [2005] reviewed the non-parametric methods and Granger causality based methods. Many progress have been made recently using Ising model and Hopfield network [Huang, 2013, Dunn and Roudi, 2013, Battistin et al., 2015, Capone et al., 2015, Roudi and Hertz, 2011] with sparsity regularization [Pernice and Rotter, 2013]. The GLM method [Okatan et al., 2005, Truccolo et al., 2005, Pillow et al., 2008] and the maximum entropy method [Schneidman et al., 2006] are two popular classes of methods, which are the main modern approaches for modeling multi-unit recordings [Roudi et al., 2015].

In the trend of current research, people are recording more and more neurons and looking for new data analysis techniques to handle bigger data with higher dimensionality. The field is in favor of algorithms that require less samples and scale well with dimensionality, but at the same time not sacrificing the accuracy. Also, an algorithm that is model free or make minimum assumptions about the hidden structure of the data has the potential to be applied to multiple types of neural recordings.

The key difference between our methods and other methods is that we use the relationship between a neuron's differential voltage and a neuron's voltage rather than finding the relationship between voltages. This provides better performance because the differential voltage is a proxy of a neuron's synaptic current. And the relationship between a neuron's synaptic current and its input voltages is more linear, which is suitable for data analysis techniques like the covariance method. While this linear relationship hold only for our passive neuron model, we still see similar or better performance of our methods in our Hodgkin-Huxley model based examples, where we relaxed this assumption and allowed loops in the networks. This implies that this class of methods are still applicable even when the ion channels' conductances vary non-linearly with the voltages, which makes the linear relationship only weakly holds.

4.2 Caveats and future directions

One open question for the differential covariance-based methods is how to improve the way they handle neural signals that are non-linearly transformed from the intracellular voltages. Currently, to achieve good performance in the calcium imaging example, we need to assume knowing the exact transfer function and reversely reconstruct the action potentials. We find this reverse transform method to be prone to additive Gaussian noise. Further study is needed to find better way to preprocess calcium imaging data for the differential covariance-based methods.

Our Hodgkin-Huxley simulations did not include axonal or synaptic delays, which is a critical feature of a real neural circuit. Unfortunately, it is non-trivial to add this feature to our Hodgkin-Huxley model. Nevertheless, we tested our methods with the passive neuron model using the same connection patterns but with random synaptic delays between neurons. In appendix D, we show that for up to a 10 ms uniformly

distributed synaptic delay pattern, our methods still outperform the correlation-based methods.

Acknowledgement

We would like to thank Dr. Thomas Liu, and all members of the Computational Neurobiology Lab for providing helpful feedback. This research is supported by ONR MURI (N000141310672), Swartz Foundation and Howard Hughes Medical Institute.

Appendix

A Differential covariance derivations

In this section, we first build a simple 3-neuron network to demonstrate that our differential covariance-based methods can reduce the type 1 false connections. Then we develop a generalized theory, which shows that the type 1 false connections' strength is always lower in our differential covariance-based methods than the original correlation-based methods.

A.1 A 3-neuron network

Let us assume a network of 3 neurons, where neuron A projects to neuron B and C:

$$\begin{aligned} I_A &= dV_A/dt = g_l V_A + \mathcal{N}_A \\ I_B &= dV_B/dt = g_1 V_A + g_l V_B + \mathcal{N}_B \\ I_C &= dV_C/dt = g_2 V_A + g_l V_C + \mathcal{N}_C \end{aligned} \tag{20}$$

Here, the cell conductance is g_l, neuron A's synaptic connection strength to neuron B is g_1, and neuron A's synaptic connection strength to neuron C is g_2. $\mathcal{N}_A, \mathcal{N}_B, \mathcal{N}_C$ are independent white Gaussian noises.

From Eq.18 of Fan et al. [2011], we can derive the covariance matrix of this network:

$$vec(COV) = -(G \otimes I_n + I_n \otimes G)^{-1}(D \otimes D)vec(I_m) \tag{21}$$

Where,

$$G = \begin{bmatrix} g_l & g_1 & g_2 \\ 0 & g_l & 0 \\ 0 & 0 & g_l \end{bmatrix}^T \tag{22}$$

is the transpose of the ground truth connection of the network. And,

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{23}$$

since each neuron receives independent noise. I_n is an identity matrix of the size of G and I_m is an identity matrix of the size of D. $\otimes$ is the Kronecker product and $vec()$ is the column vectorization function.

Therefore, we have the covariance matrix of the network as:

$$COV = \begin{bmatrix} -1/(2*g_l) & g_1/(4*g_l^2) & g_2/(4*g_l^2) \\ g_1/(4*g_l^2) & -(g_1^2+2*g_l^2)/(4*g_l^3) & -(g_1*g_2)/(4*g_l^3) \\ g_2/(4*g_l^2) & -(g_1*g_2)/(4*g_l^3) & -(g_2^2+2*g_l^2)/(4*g_l^3) \end{bmatrix} \quad (24)$$

When computing the differential covariance, we plug in Eq. 20. For example:

$$COV(I_C, V_B) = g_2 COV(V_A, V_B) + g_l COV(V_C, V_B) \quad (25)$$

Therefore, from Eq. 24, we can compute the differential covariance as:

$$\Delta_P = \begin{bmatrix} -1/2 & g_1/(4*g_l) & g_2/(4*g_l) \\ -g_1/(4*g_l) & -1/2 & 0 \\ -g_2/(4*g_l) & 0 & -1/2 \end{bmatrix} \quad (26)$$

Notice that, because the ratio between $COV(V_A, V_B)$ and $COV(V_C, V_B)$ is $-g_l/g_2$, in differential covariance, the type 1 false connection $COV(I_C, V_B)$ has value 0.

A.2 Type 1 false connection's strength is reduced in differential covariance

In this section, we propose a theory. Given a network, which consists of passive neurons in the following form:

$$I_i(t) = C\frac{dV_i(t)}{dt} = \sum_{k \in \{pre_i\}} g_{k \to i} V_k(t) + g_l V_i(t) + dB_i(t) \quad (27)$$

Where $\{pre_i\}$ is the set of neurons that project to neuron i, $g_{k \to i}$ is the synaptic conductance for the projection from neuron k to neuron i. $B_i(t)$ is a Brownian motion.

And further assume that:

- All neurons' leakage conductance g_l and membrane capacitance C are constants and the same.
- There is no loop in the network.
- $g_{syn} << g_l$, where g_{syn} is the maximum of $|g_{i \to j}|$, for $\forall i, j$

Then, we prove below that:

- For two neurons that have physical connection, their covariance is $O(\frac{g_{syn}}{g_l^2})$.
- For two neurons that do not have physical connection, their covariance is $O(\frac{g_{syn}^2}{g_l^3})$.
- For two neurons that have physical connection, their differential covariance is $O(\frac{g_{syn}}{g_l})$.
- For two neurons that do not have physical connection, their differential covariance is $O(\frac{g_{syn}^3}{g_l^3})$.
- The type 1 false connection's strength is reduced in differential covariance.

Lemma 1 The asymptotic auto-covariance of a neuron is

$$Cov[V_i, V_i] = -(1 + 2 \sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_i])/2g_l \tag{28}$$

Proof From Eq.9 of Fan et al. [2011], we have,

$$d(V_i - E[V_i]) = \sum_{k \in \{pre_i\}} g_{k \to i}(V_k - E[V_k])dt + g_l(V_i - E[V_i]) + dB_i(t) \tag{29}$$

Where $E[]$ is the expectation operation. (t) is dropped from $V_i(t)$ when the meaning is unambiguous.

From Theorem 2 of Fan et al. [2011], integrating by parts using Itô calculus gives

$$\begin{array}{l} d((V_i - E[V_i])(V_i - E[V_i])) = d(V_i - E[V_i]) \cdot (V_i - E[V_i]) \\ +(V_i - E[V_i]) \cdot d(V_i - E[V_i]) + d[V_i - E[V_i], V_i - E[V_i]] \end{array} \tag{30}$$

Taking the expectation of both sides with Eq. 29 gives

$$dCov[V_i, V_i] = 2(\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_i] + g_l Cov[V_i, V_i])dt + dE[[B_i(t), B_i(t)]] \tag{31}$$

when $t \to +\infty$, Eq. 31 becomes

$$\begin{array}{l} 0 = 2(\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k(+\infty), V_i(+\infty)] + g_l Cov[V_i(+\infty), V_i(+\infty)]) + 1 \\ Cov[V_i(+\infty), V_i(+\infty)] = -(1 + 2 \sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k(+\infty), V_i(+\infty)])/2g_l \end{array} \tag{32}$$

Lemma 2 The asymptotic covariance between two neurons is

$$Cov[V_i, V_j] = -(\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_j] + \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_i])/2g_l \tag{33}$$

Proof From Eq.9 of Fan et al. [2011], we have,

$$d(V_i - E[V_i]) = \sum_{k \in \{pre_i\}} g_{k \to i}(V_k - E[V_k])dt + g_l(V_i - E[V_i]) + dB_i(t) \tag{34}$$

From Theorem 2 of Fan et al. [2011], integrating by parts using Itô calculus gives

$$\begin{array}{l} d((V_i - E[V_i])(V_j - E[V_j])) = d(V_i - E[V_i]) \cdot (V_j - E[V_j]) \\ +(V_j - E[V_j]) \cdot d(V_i - E[V_i]) + d[V_i - E[V_i], V_j - E[V_j]] \end{array} \tag{35}$$

Taking the expectation of both sides with Eq. 34 gives

$$\begin{array}{l} dCov[V_i, V_i] = (\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_i] + \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_j] \\ +2g_l Cov[V_i, V_j])dt + dE[[B_i(t), B_j(t)]] \end{array} \tag{36}$$

when $t \to +\infty$, Eq. 36 becomes

$$\begin{aligned} 0 = & \sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k(+\infty), V_i(+\infty)] \\ & + \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k(+\infty), V_j(+\infty)] \\ & + 2 g_l Cov[V_i(+\infty), V_j(+\infty)] \\ Cov[V_i(+\infty), V_j(+\infty)] = & -(\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k(+\infty), V_i(+\infty) \\ & + \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k(+\infty), V_j(+\infty)])/2 g_l \end{aligned} \tag{37}$$

Theorem 1 The auto-covariance of a neuron is $O(\frac{1}{g_l})$. The covariance of two different neurons with or without physical connection is $O(\frac{g_{syn}}{g_l^2})$.

Proof We prove this by induction.

The basis: The base case contains two neurons:

$$\begin{aligned} C \frac{dV_1}{dt} &= g_l V_1 + \mathcal{N}_1 \\ C \frac{dV_2}{dt} &= g_{1 \to 2} V_1 + g_l V_2 + \mathcal{N}_2 \end{aligned} \tag{38}$$

From Lemma 1, we have:

$$Cov[V_1, V_1] = -1/2 g_l \tag{39}$$

Then, from Lemma 2, we have:

$$\begin{aligned} Cov[V_1, V_2] &= -g_{1 \to 2} Cov[V_1, V_1]/2 g_l \\ Cov[V_1, V_2] &= g_{1 \to 2}/4 g_l^2 \end{aligned} \tag{40}$$

And, from Lemma 1, we have:

$$\begin{aligned} Cov[V_2, V_2] &= -(1 + 2 g_{1 \to 2} Cov[V_1, V_2])/2 g_l \\ Cov[V_1, V_2] &= -1/2 g_l - g_{1 \to 2}^2/4 g_l^3 \end{aligned} \tag{41}$$

So the statement holds.

The inductive step: If the statement holds for a network of $n - 1$ neurons, we add one more neuron to it.

Part 1: First, let's prove the covariance of any neuron with n is also $O(\frac{g_{syn}}{g_l^2})$.

From Lemma 2, we have:

$$Cov[V_i, V_n] = -(\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_n] + \sum_{k \in \{pre_n\}} g_{k \to n} Cov[V_k, V_i])/2 g_l \tag{42}$$

where, $\{pre_i\}$ are the neurons projecting to neuron i, $\{pre_n\}$ are the neurons projecting to neuron n.

Note that, because $\{pre_n\}$ are neurons from the old network, $Cov[V_k, V_i], k \in \{pre_n\}$ is at most $O(\frac{1}{g_l})$, and it is $O(\frac{1}{g_l})$ only when $k = i$.

Now, we need to prove that $Cov[V_k, V_n], k \in \{pre_i\}$ is also $O(\frac{1}{g_l})$. We prove this by contradiction. Let's suppose that $Cov[V_k, V_n], k \in \{pre_i\}$ is larger than $O(\frac{1}{g_l})$. Then similar to Eq. 42, we have:

For $p \in \{pre_i\}$

$$Cov[V_p, V_n] = -(\sum_{k \in \{pre_p\}} g_{k \to p} Cov[V_k, V_n] + \sum_{k \in \{pre_n\}} g_{k \to n} Cov[V_k, V_p])/2g_l \quad (43)$$

Here we separate the problem into two situations,

Case 1: neuron i projects to neuron n Since there is no loop in the network, $n \notin \{pre_i\}$. Therefore, $Cov[V_k, V_p], k \in \{pre_n\}, p \in \{pre_i\}$ is the covariance of two neurons from the old network and is $O(\frac{1}{g_l})$. $Cov[V_k, V_n], k \in \{pre_p\}$ must be larger than $O(\frac{1}{g_l})$, such that $Cov[V_p, V_n], p \in \{pre_i\}$ is larger than $O(\frac{1}{g_l})$.

Therefore, if a neuron's covariance with neuron n is larger than $O(\frac{1}{g_l})$, one of its antecedents' covariance with neuron n is also larger than $O(\frac{1}{g_l})$. Since we assume there is no loop in this network, there must be at least one antecedent (say, neuron m) whose covariance with neuron n is larger than $O(\frac{1}{g_l})$ and it has no antecedent.

However, from Lemma 2:

$$Cov[V_m, V_n] = -(\sum_{k \in \{pre_n\}} g_{k \to n} Cov[V_k, V_m])/2g_l \quad (44)$$

Since $Cov[V_m, V_k]), k \in \{pre_n\}$ is $O(\frac{1}{g_l})$, $Cov[V_m, V_n]$ is $O(\frac{g_{syn}}{g_l^2})$, which is smaller than $O(\frac{1}{g_l})$. This is a contradiction. So $Cov[V_k, V_n], k \in \{pre_i\}$ is no larger than $O(\frac{1}{g_l})$. Therefore, $Cov[V_i, V_n]$ is $O(\frac{g_{syn}}{g_l^2})$.

Case 2: neuron i does not project to neuron n Now, in Eq. 43, it is possible that $n \in \{pre_i\}$. However, in case 1, we just proved that the covariance of any neuron that projects to neuron n is $O(\frac{g_{syn}}{g_l^2})$. Therefore, $Cov[V_k, V_p], k \in \{pre_n\}, p \in \{pre_i\}$ is $O(\frac{g_{syn}}{g_l^2})$ regardless of whether $p = n$.

Then, similar to case 1, there must be an antecedent of neuron i (say neuron m), whose covariance with neuron n is larger than $O(\frac{1}{g_l})$ and it has no antecedent. Then from Eq. 44 we know this is a contradiction. So, $Cov[V_i, V_n]$ is $O(\frac{g_{syn}}{g_l^2})$.

Part 2: Then, for the auto-covariance of neuron n:

$$Cov[V_n, V_n] = -(1 + 2 \sum_{k \in \{pre_n\}} g_{k \to n} Cov[V_k, V_n])/2g_l \quad (45)$$

As we already proved that any neuron's covariance with neuron n is $O(\frac{g_{syn}}{g_l^2})$, the dominant term in Eq. 45 is $-1/2g_l$. Therefore, the auto-covariance of neuron n is also $O(\frac{1}{g_l})$.

End of proof.

Theorem 2 The covariance of two neurons that are physically connected is $O(\frac{g_{syn}}{g_l^2})$. The covariance of two neurons that are not physically connected is $O(\frac{g_{syn}^2}{g_l^3})$.

Proof From Lemma 2, we have:

$$Cov[V_i, V_j] = -(\sum_{k\in\{pre_i\}} g_{k\to i} Cov[V_k, V_j] + \sum_{k\in\{pre_j\}} g_{k\to j} Cov[V_k, V_i])/2g_l \tag{46}$$

If neuron i and neuron j are physically connected, let's say $i \to j$, then $i \in \{pre_j\}$. Thus one of the V_k for $k \in \{pre_j\}$ is V_i. Therefore, $Cov[V_k, V_i]$ is $O(\frac{1}{g_l})$. Since there is no loop in the network, $j \notin \{pre_i\}$, so $Cov[V_k, V_j]$ is $O(\frac{g_{syn}}{g_l^2})$. Therefore, $Cov[V_i, V_j]$ is $O(\frac{g_{syn}}{g_l^2})$.

If neuron i and neuron j are not physically connected, we have $i \notin \{pre_j\}$, so $Cov[V_k, V_i])$ is $O(\frac{g_{syn}}{g_l^2})$. And $j \notin \{pre_i\}$, so $Cov[V_k, V_j])$ is $O(\frac{g_{syn}}{g_l^2})$. Therefore, $Cov[V_i, V_j]$ is $O(\frac{g_{syn}^2}{g_l^3})$.

End of proof.

Lemma 3 The differential covariance of two neurons,

$$Cov[\frac{dV_i}{dt}, V_j] + Cov[V_i, \frac{dV_j}{dt}] = 0 \tag{47}$$

Proof From Lemma 2, we have,

$$\begin{aligned} &2g_l Cov[V_i, V_j] + \sum_{k\in\{pre_i\}} g_{k\to i} Cov[V_k, V_j] + \sum_{k\in\{pre_j\}} g_{k\to j} Cov[V_k, V_i] = 0 \\ &Cov[\sum_{k\in\{pre_i\}} g_{k\to i} V_k + g_l V_i, V_j] + Cov[\sum_{k\in\{pre_j\}} g_{k\to j} V_k + g_l V_j, V_i] = 0 \\ &Cov[\tfrac{dV_i}{dt}, V_j] + Cov[V_i, \tfrac{dV_j}{dt}] = 0 \end{aligned} \tag{48}$$

End of proof.

Theorem 3 The differential covariance of two neurons that are physically connected is $O(\frac{g_{syn}}{g_l})$.

Proof Assume two neurons have physical connection as $i \to j$. The differential covariance of them is:

$$\begin{aligned} &Cov[\tfrac{dV_i}{dt}, V_j] = Cov[\sum_{k\in\{pre_i\}} g_{k\to i} V_k + g_l V_i, V_j] \\ &= \sum_{k\in\{pre_i\}} g_{k\to i} Cov[V_k, V_j] + g_l Cov[V_i, V_j] \end{aligned} \tag{49}$$

From Theorem 2, we know $Cov[V_i, V_j]$ is $O(\frac{g_{syn}}{g_l^2})$, and

- If V_k is projecting to V_j, $Cov[V_k, V_j]$ is $O(\frac{g_{syn}}{g_l^2})$.
- If V_k is not projecting to V_j, $Cov[V_k, V_j]$ is $O(\frac{g_{syn}^2}{g_l^3})$.

Therefore, the dominant term is $Cov[V_i, V_j]$. And $Cov[\frac{dV_i}{dt}, V_j]$ is $O(\frac{g_{syn}}{g_l})$.

From Lemma 3,

$$Cov[V_i, \frac{dV_j}{dt}] = -Cov[\frac{dV_i}{dt}, V_j] \tag{50}$$

Therefore, $Cov[V_i, \frac{dV_j}{dt}]$ is $O(\frac{g_{syn}}{g_l})$.

End of proof.

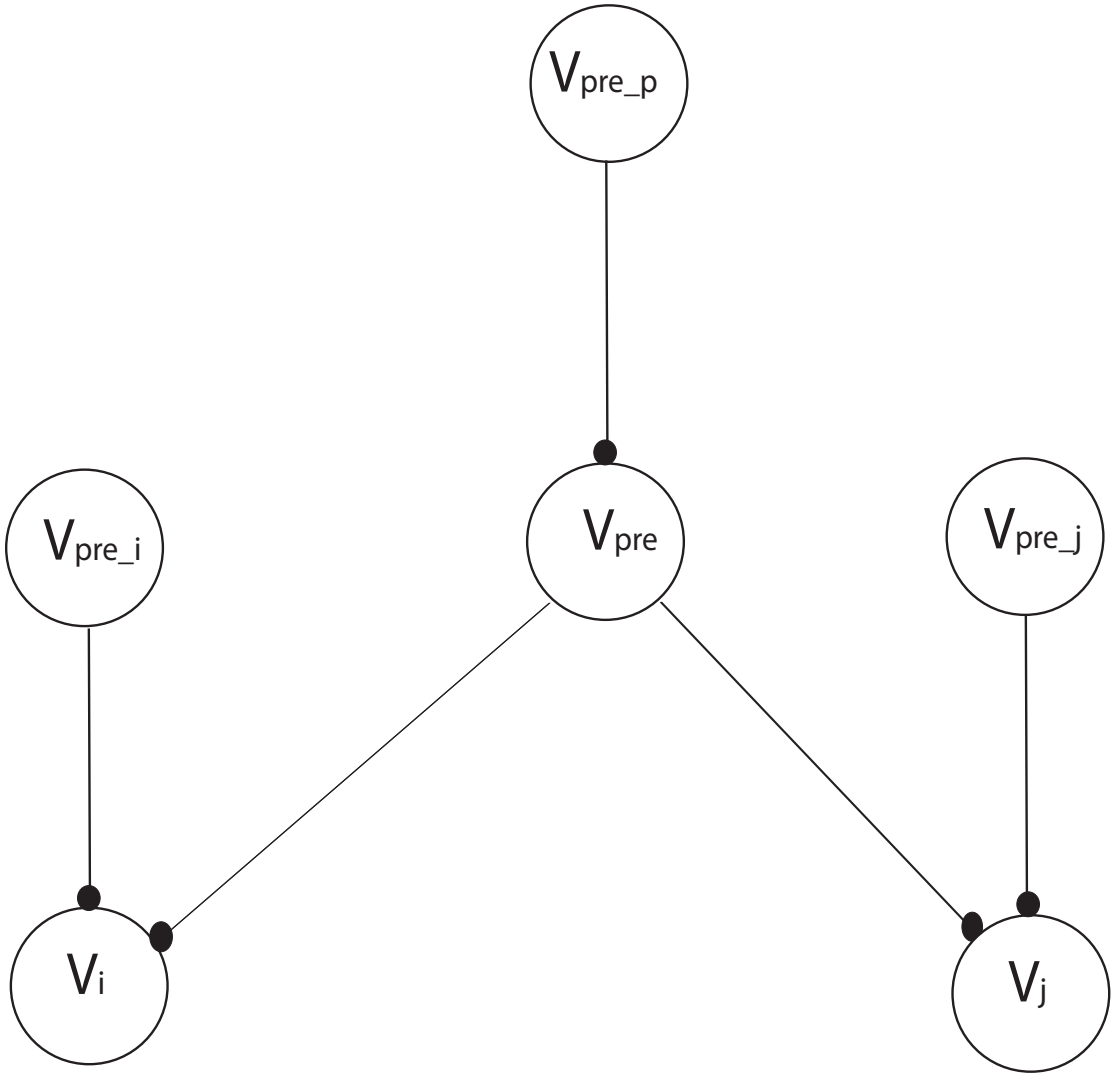

Figure 16: The network used in Theorem 4's proof.

Theorem 4 The differential covariance of two neurons that are not physically connected is $O(\frac{g_{syn}^3}{g_l^3})$.

Proof First, let's define the antecedents of neuron i and neuron j. Shown in Fig. 16, $\{pre\}$ is the set of common antecedents of neuron i and neuron j. $\{pre_p\}$ is the set of antecedents of $p \in \{pre\}$. $\{pre_i\}$ is the set of exclusive antecedents of neuron i. $\{pre_j\}$ is the set of exclusive antecedents of neuron j.

From Lemma 2, we have, for any neuron $p \in \{pre\}$

$$\begin{aligned} Cov[V_i, V_p] = -(&\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_p] + \sum_{k \in \{pre\}} g_{k \to i} Cov[V_k, V_p] \\ &+ \sum_{k \in \{pre_p\}} g_{k \to p} Cov[V_k, V_i])/2g_l \end{aligned} \tag{51}$$

$$\begin{aligned} Cov[V_j, V_p] = -(&\sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_p] + \sum_{k \in \{pre\}} g_{k \to j} Cov[V_k, V_p] \\ &+ \sum_{k \in \{pre_p\}} g_{k \to p} Cov[V_k, V_j])/2g_l \end{aligned} \tag{52}$$

For simplicity, we define

$$\begin{aligned} \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_p] &= C_p \\ \sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_p] &= D_p \\ \sum_{k \in \{pre_p\}} g_{k \to p} Cov[V_k, V_j] &= E_p \\ \sum_{k \in \{pre_p\}} g_{k \to p} Cov[V_k, V_i] &= F_p \end{aligned} \tag{53}$$

From Lemma 2, we also have,

$$\begin{aligned} Cov[V_i, V_j] = -(&\sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_j] + \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_i] \\ &+ \sum_{k \in \{pre\}} g_{k \to i} Cov[V_k, V_j] + \sum_{k \in \{pre\}} g_{k \to j} Cov[V_k, V_i])/2g_l \end{aligned} \tag{54}$$

For simplicity, we define

$$\begin{aligned} \sum_{k \in \{pre_i\}} g_{k \to i} Cov[V_k, V_j] &= A \\ \sum_{k \in \{pre_j\}} g_{k \to j} Cov[V_k, V_i] &= B \end{aligned} \tag{55}$$

Plug in Eq. 51, Eq. 52 to Eq. 54, we have,

$$\begin{aligned} Cov[V_i, V_j] = &\sum_{p \in \{pre\}} \frac{g_{p \to i}}{4g_l^2} (\sum_{k \in \{pre\}} g_{k \to j} Cov[V_k, V_p] + C_p + E_p) \\ &+ \sum_{p \in \{pre\}} \frac{g_{p \to j}}{4g_l^2} (\sum_{k \in \{pre\}} g_{k \to i} Cov[V_k, V_p] + D_p + F_p) \\ &- \frac{A}{2g_l} - \frac{B}{2g_l} \end{aligned} \tag{56}$$

Now, we look at the differential covariance between neuron i and neuron j,

$$Cov[\frac{dV_i}{dt}, V_j] = \sum_{k\in\{pre_i\}} g_{k\to i} Cov[V_k, V_j] + \sum_{k\in\{pre\}} g_{k\to i} Cov[V_k, V_j] + g_l Cov[V_i, V_j] \tag{57}$$

Plug in Eq. 52, Eq. 56, we have,

$$\begin{aligned} Cov[\frac{dV_i}{dt}, V_j] = A &+ \sum_{p\in\{pre\}} \frac{-g_{p\to i}}{2g_l}(\sum_{k\in\{pre\}} g_{k\to j} Cov[V_k, V_p] + C_p + E_p) \\ &+ g_l[\sum_{p\in\{pre\}} \frac{g_{p\to i}}{4g_l^2}(\sum_{k\in\{pre\}} g_{k\to j} Cov[V_k, V_p] + C_p + E_p) \\ &+ \sum_{p\in\{pre\}} \frac{g_{p\to j}}{4g_l^2}(\sum_{k\in\{pre\}} g_{k\to i} Cov[V_k, V_p] + D_p + F_p) \\ &- \frac{A}{2g_l} - \frac{B}{2g_l}] \\ = \frac{A}{2} &- \frac{B}{2} + \sum_{p\in\{pre\}} \frac{g_{p\to j}}{4g_l}(D_p + F_p) - \sum_{p\in\{pre\}} \frac{g_{p\to i}}{4g_l}(C_p + E_p) \end{aligned} \tag{58}$$

Note,

- There is no physical connection between a neuron in $\{pre_i\}$ and neuron j, otherwise, this neuron belongs to $\{pre\}$. Therefore, from Theorem 2, A is $O(\frac{g_{syn}^2}{g_l^3}) * g_{k\to i} = O(\frac{g_{syn}^3}{g_l^3})$.
- There is no physical connection between a neuron in $\{pre_j\}$ and neuron i, otherwise, this neuron belongs to $\{pre\}$. Therefore, from Theorem 2, B is $O(\frac{g_{syn}^2}{g_l^3}) * g_{k\to j} = O(\frac{g_{syn}^3}{g_l^3})$.
- There could be physical connections between neurons in $\{pre_p\}$ and $\{pre_j\}$, so C_p is $O(\frac{g_{syn}}{g_l^2}) * g_{k\to j} = O(\frac{g_{syn}^2}{g_l^2})$.
- There could be physical connections between neurons in $\{pre_p\}$ and $\{pre_i\}$, so D_p is $O(\frac{g_{syn}}{g_l^2}) * g_{k\to i} = O(\frac{g_{syn}^2}{g_l^2})$.
- There could be physical connections between neurons in $\{pre_p\}$ and neuron j, so E_p is $O(\frac{g_{syn}}{g_l^2}) * g_{k\to p} = O(\frac{g_{syn}^2}{g_l^2})$.
- There could be physical connections between neurons in $\{pre_p\}$ and neuron i, so F_p is $O(\frac{g_{syn}}{g_l^2}) * g_{k\to p} = O(\frac{g_{syn}^2}{g_l^2})$.

Therefore,

$$
\begin{aligned}
Cov[\frac{dV_i}{dt}, V_j] &= O(\frac{g_{syn}^3}{g_l^3}) - O(\frac{g_{syn}^3}{g_l^3}) + \sum_{p\in\{pre\}}\sum_{k\in\{pre\}} \frac{g_{p\to j}}{4g_l}(O(\frac{g_{syn}^2}{g_l^2}) + O(\frac{g_{syn}^2}{g_l^2})) \\
&\quad - \sum_{p\in\{pre\}}\sum_{k\in\{pre\}} \frac{g_{p\to i}}{4g_l}(O(\frac{g_{syn}^2}{g_l^2}) + O(\frac{g_{syn}^2}{g_l^2})) \\
&= O(\frac{g_{syn}^3}{g_l^3})
\end{aligned}
\tag{59}
$$

From Lemma 3, we know, $Cov[V_i, \frac{dV_j}{dt}]$ is also $O(\frac{g_{syn}^3}{g_l^3})$.
End of proof.

Theorem 5 The type 1 false connection's strength is reduced in differential covariance

Proof From theorem 1 and theorem 2, we know that, in the correlation method, the strength of a non-physical connection ($O(\frac{g_{syn}}{g_l^2})$) is $\frac{g_{syn}}{g_l}$ times that of a physical connection ($O(\frac{g_{syn}^2}{g_l^3})$).

From theorem 3 and theorem 4, we know that, in the differential covariance method, the strength of a non-physical connection ($O(\frac{g_{syn}}{g_l})$) is $\frac{g_{syn}^2}{g_l^2}$ times that of a physical connection ($O(\frac{g_{syn}^3}{g_l^3})$).

Because $g_{syn} << g_l$, the relative strength of the non-physical connections is reduced in the differential covariance method.

End of proof.

A.3 Directionality information in differential covariance

Theorem 6 If neuron i projects to neuron j with an excitatory connection, $Cov[\frac{dV_i}{dt}, V_j] > 0$ and $Cov[V_i, \frac{dV_j}{dt}] < 0$

Proof Given the model above, similar to Theorem 4, from Lemma 2, we have, for any neuron $p \in \{pre\}$

$$
\begin{aligned}
Cov[V_i, V_p] = -(&\sum_{k\in\{pre_i\}} g_{k\to i} Cov[V_k, V_p] + \sum_{k\in\{pre\}} g_{k\to i} Cov[V_k, V_p] \\
&+ \sum_{k\in\{pre_p\}} g_{k\to p} Cov[V_k, V_i])/2g_l
\end{aligned}
\tag{60}
$$

$$
\begin{aligned}
Cov[V_j, V_p] = -(&\sum_{k\in\{pre_j\}} g_{k\to j} Cov[V_k, V_p] + \sum_{k\in\{pre\}} g_{k\to j} Cov[V_k, V_p] \\
&+ \sum_{k\in\{pre_p\}} g_{k\to p} Cov[V_k, V_j] + g_{i\to j} Cov[V_i, V_p])/2g_l
\end{aligned}
\tag{61}
$$

For simplicity, we define

$$\begin{aligned}\sum_{k\in\{pre_j\}} g_{k\to j}Cov[V_k,V_p] &= C_p\\ \sum_{k\in\{pre_i\}} g_{k\to i}Cov[V_k,V_p] &= D_p\\ \sum_{k\in\{pre_p\}} g_{k\to p}Cov[V_k,V_j] &= E_p\\ \sum_{k\in\{pre_p\}} g_{k\to p}Cov[V_k,V_i] &= F_p\end{aligned} \tag{62}$$

From Lemma 2, we also have,

$$\begin{aligned}Cov[V_i,V_j] = -(&\sum_{k\in\{pre_i\}} g_{k\to i}Cov[V_k,V_j] + g_{i\to j}Cov[V_i,V_i] + \sum_{k\in\{pre_j\}} g_{k\to j}Cov[V_k,V_i]\\ &+ \sum_{k\in\{pre\}} g_{k\to i}Cov[V_k,V_j] + \sum_{k\in\{pre\}} g_{k\to j}Cov[V_k,V_i])/2g_l\end{aligned} \tag{63}$$

For simplicity, we define

$$\begin{aligned}\sum_{k\in\{pre_i\}} g_{k\to i}Cov[V_k,V_j] &= A\\ \sum_{k\in\{pre_j\}} g_{k\to j}Cov[V_k,V_i] &= B\end{aligned} \tag{64}$$

Plug in, we have,

$$\begin{aligned}Cov[V_i,V_j] = &\sum_{p\in\{pre\}} \frac{g_{p\to i}}{4g_l^2}(\sum_{k\in\{pre\}} g_{k\to j}Cov[V_k,V_p] + C_p + E_p + g_{i\to j}Cov[V_i,V_p])\\ &+ \sum_{p\in\{pre\}} \frac{g_{p\to j}}{4g_l^2}(\sum_{k\in\{pre\}} g_{k\to i}Cov[V_k,V_p] + D_p + F_p)\\ &- \frac{A}{2g_l} - \frac{B}{2g_l} - \frac{g_{i\to j}Cov[V_i,V_i]}{2g_l}\end{aligned} \tag{65}$$

Now, we look at the differential covariance between neuron i and neuron j,

$$\begin{aligned}Cov[\frac{dV_i}{dt},V_j] = &\frac{A}{2} - \frac{B}{2} + \sum_{p\in\{pre\}} \frac{g_{p\to j}}{4g_l}(D_p + F_p)\\ &- \sum_{p\in\{pre\}} \frac{g_{p\to i}}{4g_l}(C_p + E_p + + g_{i\to j}Cov[V_i,V_p]) - \frac{g_{i\to j}Cov[V_i,V_i]}{2}\end{aligned} \tag{66}$$

Note, in Theorem 4, we already proved the scale of A, B, C_p, D_p, E_p, F_p. We also have,

- There are physical connections between neurons in $\{pre\}$ and neuron i, so $g_{i\to j}Cov[V_i,V_p]$ is $O(\frac{g_{syn}}{g_l^2}) * g_{k\to p} = O(\frac{g_{syn}^2}{g_l^2})$.

- From Lemma 1, the auto-covariance of neuron i is $O(\frac{1}{g_l})$. So $g_{i\to j}Cov[V_i, V_i]$ is $O(\frac{1}{g_l}) * g_{i\to j} = O(\frac{g_{syn}}{g_l})$.

Therefore, $g_{i\to j}Cov[V_i, V_i]$ is the dominant term in $Cov[\frac{dV_i}{dt}, V_j]$. Since $Cov[V_i, V_i] > 0$, for excitatory connection $g_{i\to j} > 0$, $Cov[\frac{dV_i}{dt}, V_j] < 0$.

From Lemma 3,

$$Cov[V_i, \frac{dV_j}{dt}] = -Cov[\frac{dV_i}{dt}, V_j] \tag{67}$$

Therefore, $Cov[V_i, \frac{dV_j}{dt}] > 0$.

End of proof.

Corollary 1 If neuron i projects to neuron j with an inhibitory connection, $Cov[\frac{dV_i}{dt}, V_j] > 0$ and $Cov[V_i, \frac{dV_j}{dt}] < 0$

Proof The proof is similar to Theorem 6. Again, we know $g_{i\to j}Cov[V_i, V_i]$ is the dominant term in $Cov[\frac{dV_i}{dt}, V_j]$. Since $Cov[V_i, V_i] > 0$, now for an inhibitory connection $g_{i\to j} < 0$, $Cov[\frac{dV_i}{dt}, V_j] > 0$.

From Lemma 3,

$$Cov[V_i, \frac{dV_j}{dt}] = -Cov[\frac{dV_i}{dt}, V_j] \tag{68}$$

Therefore, $Cov[V_i, \frac{dV_j}{dt}] < 0$.

End of proof.

B Benchmarked methods

We compared our methods to a few popular methods.

B.1 Correlation method

The correlation matrix is defined as:

$$COV_{x,y} = \frac{E[(x - \bar{x})(y - \bar{y})^T]}{\sqrt{E[(x - \bar{x})(x - \bar{x})^T]E[(y - \bar{y})(y - \bar{y})^T]}} \tag{69}$$

Where x(t) is the membrane voltage recording of neuron x, and y(t) is the membrane voltage recording of neuron y. $\bar{x}$ is the mean voltage of neuron x, and $\bar{y}$ is the mean voltage of neuron y. E[] is the expectation operation.

B.2 Precision matrix

The precision matrix is the inverse of the correlation matrix:

$$P = COV^{-1}, \tag{70}$$

It is equivalent to the partial correlation. Here we briefly review this derivation, because we use it to develop our new method. The derivation here is based on and adapted from Cox and Wermuth [1996].

We begin by considering a pair of variables (x, y), and remove the correlation in them introduced from a control variable z.

First, we define the correlation matrix as:

$$COV_{xyz} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{71}$$

By solving the linear regression problem:

$$\begin{aligned} w_x &= \arg\min_w E(x - w * z)^2 \\ w_y &= \arg\min_w E(y - w * z)^2 \end{aligned} \tag{72}$$

we have:

$$\begin{aligned} w_x &= \sigma_{xz}\sigma_{zz}^{-1} \\ w_y &= \sigma_{yz}\sigma_{zz}^{-1} \end{aligned} \tag{73}$$

then, we define the residual of x, y as,

$$\begin{aligned} r_x &= x - w_x * z \\ r_y &= y - w_y * z \end{aligned} \tag{74}$$

Therefore, the correlation of r_x, r_y is:

$$COV_{r_x,r_y} = \sigma_{xy} - \sigma_{xz} * \sigma_{zz}^{-1} * \sigma_{yz} \tag{75}$$

On the other hand, if we define the precision matrix as:

$$P_{xyz} = \begin{bmatrix} p_{xx} & p_{xy} & p_{xz} \\ p_{yx} & p_{yy} & p_{yz} \\ p_{zx} & p_{zy} & p_{zz} \end{bmatrix} \tag{76}$$

Using Cramer's rule, we have:

$$p_{xy} = \frac{-\begin{vmatrix} \sigma_{xy} & \sigma_{xz} \\ \sigma_{zy} & \sigma_{zz} \end{vmatrix}}{|COV_{xyz}|} \tag{77}$$

Therefore,

$$p_{xy} = \frac{-\sigma_{zz}}{|COV_{xyz}|}(\sigma_{xy} - \sigma_{xz} * \sigma_{zz}^{-1} * \sigma_{yz}) \tag{78}$$

Since the diagonal terms of the correlation matrix is 1, p_{xy} and COV_{r_x,r_y} are differed by a ratio of $-|COV_{xyz}|$. So the precision matrix and the partial correlation are equivalent.

B.3 Sparse latent regularization

Prior studies[Banerjee et al., 2006, Friedman et al., 2008] have shown that regularizations can provide better estimation if the ground truth connection matrix has a known structure (e.g. sparse). For all data tested in this paper, the sparse latent regularization[Yatsenko et al., 2015] worked best. For a fair comparison, we applied the sparse latent regularization to both the precision matrix method and our differential covariance method.

In the original sparse latent regularization method, people made the assumption that a larger precision matrix S is the joint distribution of the p observed neurons and d latent neurons [Yatsenko et al., 2015]. i.e.

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

Where S_{11} corresponds to the observable neurons. If we can only measure the observable neurons, the partial correlation computed from the observed neural signals is,

$$C_{ob}^{-1} = S_{ob} = S_{11} - S_{12} * S_{22}^{-1} * S_{21} \tag{79}$$

because the invisible latent neurons as shown in Eq. 2 introduce correlations into the measurable system. We denote this correlation introduced from the latent inputs as

$$L = S_{12} * S_{22}^{-1} * S_{21} \tag{80}$$

If we can make the assumption that the connection between the visible neurons are sparse, i.e. S_{11} is sparse and the number of latent neurons is much smaller than the number of visible neurons, i.e. $d << p$. Then, prior works [Chandrasekaran et al., 2011] have shown that if S_{ob} is known, S_{11} is sparse enough and L's rank is low enough (within the bound defined in Chandrasekaran et al. [2011]), then the solution of

$$S_{11} - L = S_{ob} \tag{81}$$

is uniquely defined and can be solved by the following convex optimization problem

$$\arg\min_{S_{11},L} ||S_{11}||_1 + \alpha * tr(L) \tag{82}$$

under the constraint that

$$S_{ob} = S_{11} - L \tag{83}$$

Here, $||\ ||_1$ is the L1-norm of a matrix, and $tr()$ is the trace of a matrix. α is the penalty ratio between the L1-norm of S_{11} and the trace of L and is set to $1/\sqrt{N}$ for all our estimations.

However, the above method is used to regularize precision matrix. For our differential covariance estimation, we need to make small changes to the derivation. Note that if we assume the neural signals of the latent neurons are known, and let l be the indexes of these latent neurons, then from our previous section (section 2.2.2),

$$\Delta_{S_{i,j}} = \Delta_{P_{i,j}} - COV_{j,l} * COV_{l,l}^{-1} * \Delta_{C_{i,l}}^{T} \tag{84}$$

removes the V_{latent} terms in Eq. 2.

Even if l is unknown,

$$COV_{j,l} * COV_{l,l}^{-1} * \Delta_{C_{i,l}}^{T}$$

is low-rank, because it is bounded by the dimensionality of $COV_{l,l}$, which is d. And Δ_S is the internal connections between the visible neurons, which should be a sparse matrix. Therefore, letting

$$\begin{aligned} S_{ob} &= \Delta_P \\ S_{11} &= \Delta_S \\ L &= -COV_{j,l} * COV_{l,l}^{-1} * \Delta_{C_{i,l}}^{T} \end{aligned} \tag{85}$$

we can use the original sparse+latent method to solve for Δ_S. In this paper, we used the inexact robust PCA algorithm (`http://perception.csl.illinois.edu/matrix-rank/sample_code.html`) to solve this problem [Lin et al., 2011].

B.4 the generalized linear model method

As summarized by Roudi et al. [2015], GLMs assume that every neuron spikes at a time-varying rate which depends on earlier spikes (both those of other neurons and its own) and on external covariates (such as a stimulus or other quantities measured in the experiment). As they explained, the influence of earlier spikes on the firing probability at a given time is assumed to depend on the time since they occurred. For each pre-postsynaptic pair i, j, it is described by a function $J_{ij}(\tau)$ of this time lag [Roudi et al., 2015]. In this paper, we average this temporal dependent function $J_{ij}(\tau)$ over time to obtain the functional connectivity estimation of this method.

The Spike trains used for the GLM method were computed using the spike detection algorithm from Quiroga et al. [2004]. The GLM code was obtained from Pillow et al. [2008] (`http://pillowlab.princeton.edu/code_GLM.html`).

C Details about the thalamocortical model

C.1 Intrinsic currents

For the thalamocortical model, a conductance-based formulation was used for all neurons. The cortical neuron consisted of two compartments: dendritic and axo-somatic compartments, similar to previous studies [Bazhenov et al., 2002, Chen et al., 2012, Bonjean et al., 2011] and is described by the following equations,

$$\begin{aligned} C_m \tfrac{dV_D}{dt} &= -I_d^{K-leak} - I_d^{leak} - I_d^{Na} - I_d^{Nap} - I_d^{Ca} - I_d^{Km} - I^{syn} \\ g_c^s(V_S - V_D) &= -I_S^{Na} - I_S^{K} - I_S^{Nap} \end{aligned} \tag{86}$$

where the subscripts s and d correspond to axo-somatic and dendritic compartment, I^{leak} is the Cl^- leak currents, I^{Na} is fast Na^+ channels, I^{Nap} is persistent sodium current, I^K is fast delayed rectifier K^+ current, I^{Km} is slow voltage-dependent non-inactivating K^+ current, I^{KCa} is slow Ca^{2+} dependent K^+ current, I^{Ca} is high-threshold Ca^{2+} current, I^h is hyperpolarization-activated depolarizing current and I^{syn} is the sum of synaptic

currents to the neuron. All intrinsic currents were of the form: $g(V-E)$, where g is the conductance, V is the voltage of the corresponding compartment and E is the reversal potential. The detailed descriptions of individual currents are provided in previous publications [Bazhenov et al., 2002, Chen et al., 2012]. The conductance of the leak currents were 0.007 mS/cm^2 for I_d^{K-leak} and 0.023 mS/cm^2 for I_d^{leak}. The maximal conductance for different currents were, I_d^{Nap}: 2.0 mS/cm^2, I_d^{Na}: 0.8 mS/cm^2, I_d^{Km}: 0.012 mS/cm^2, I_d^{KCa}: 0.015 mS/cm^2 , I_d^{Km}: 0.012 mS/cm^2, I_s^{Na}: 3000mS/cm^2, I_s^{K}: 200 mS/cm^2 and I_s^{Nap}: 15 mS/cm^2. C_m was $0.075\,\mu$F/cm^2.

The following describes the IN neurons:

$$\begin{array}{rcl} C_m\frac{dV_D}{dt} & = & -I_d^{K-leak}-I_d^{leak}-I_d^{Na}-I_d^{Ca}-I_d^{KCa}-I_d^{Km}-I^{syn} \\ g_c^s(V_S-V_D) & = & -I_S^{Na}-I_S^{K} \end{array} \tag{87}$$

The conductance for leak currents for IN neurons were 0.034 mS/cm^2 for I_d^{K-leak} and 0.006 mS/cm^2 for I_d^{leak}. Maximal conductance for other currents were, I_d^{Na} : 0.8 mS/cm^2, I_d^{Ca}: 0.012 mS/cm^2, I_d^{KCa}: 0.015 mS/cm^2 , I_d^{Km}: 0.012 mS/cm^2, I_s^{Na}: 2500 mS/cm^2 and I_s^{K}: 200 mS/cm^2 .

The TC neurons consisted of only single compartment and was described as follows,

$$\frac{dV_D}{dt} = -I^{K-leak}-I^{leak}-I^{Na}-I^{K}-I^{LCa}-I^{h}-I^{syn} \tag{88}$$

The conductance of leak currents were, I^{leak}: 0.01 mS/cm^2, I^{K-leak}: 0.007 mS/cm^2. The maximal conductance for other currents were, fast Na^+ (I^{Na}) current: 90 mS/cm^2, fast K^+ (I^{k}) current: 10 mS/cm^2, low threshold Ca^{2+}(I^{LCa}) current: 2.5 mS/cm^2, hyperpolarization-activated depolarizing current (I^{h}): 0.015 mS/cm^2.

The RE cells were also modeled as a single compartment neuron as follows,

$$\frac{dV_D}{dt} = -I^{K-leak}-I^{leak}-I^{Na}-I^{K}-I^{LCa}-I^{h}-I^{syn} \tag{89}$$

The conductance for leak currents were, I^{leak}: 0.05 mS/cm^2, I^{K-leak}: 0.016 mS/cm^2. The maximal conductance for other currents were, fast Na^+ (I^{Na}) current: 100 mS/cm^2, fast K^+ (I^{K}) current: 10 mS/cm^2, low threshold Ca^{2+}(I^{LCa}) current: 2.2 mS/cm^2.

C.2 Synaptic currents

GABA-A, NMDA and AMPA synaptic currents were described by first-order activation schemes [Timofeev et al., 2000]. The equations for all synaptic currents used in this model are given in our previous publications [Bazhenov et al., 2002, Chen et al., 2012]. Briefly, below we mention only the relevant equations.

$$\begin{array}{r} I_{syn}^{AMPA}=g_{syn}[O](V-E_{AMPA}) \\ I_{syn}^{NMDA}=g_{syn}[O](V-E_{NMDA}) \\ I_{syn}^{GABA}=g_{syn}[O](V-E_{GABA}) \end{array} \tag{90}$$

D Supplementary figures

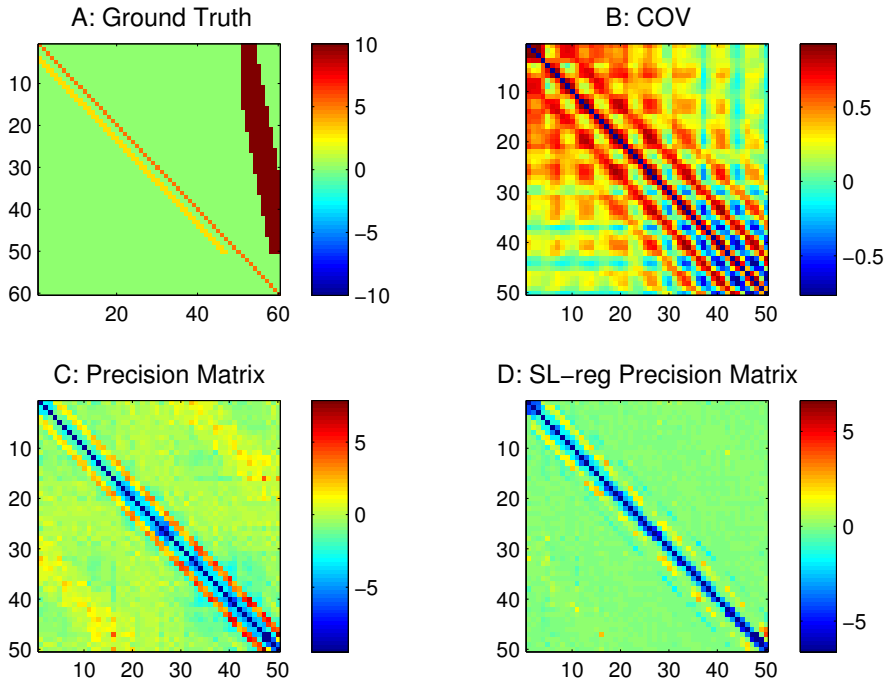

Figure 17: Supplementary1, passive neuron model with 5 ms fixed synaptic delay. Results from correlation-based methods. A) Ground truth connection matrix. neurons 1-50 are visible neurons. neurons 51-60 are invisible neurons. B) Estimation from the correlation method. C) Estimation from the precision matrix. D) Sparse+latent regularized precision matrix.

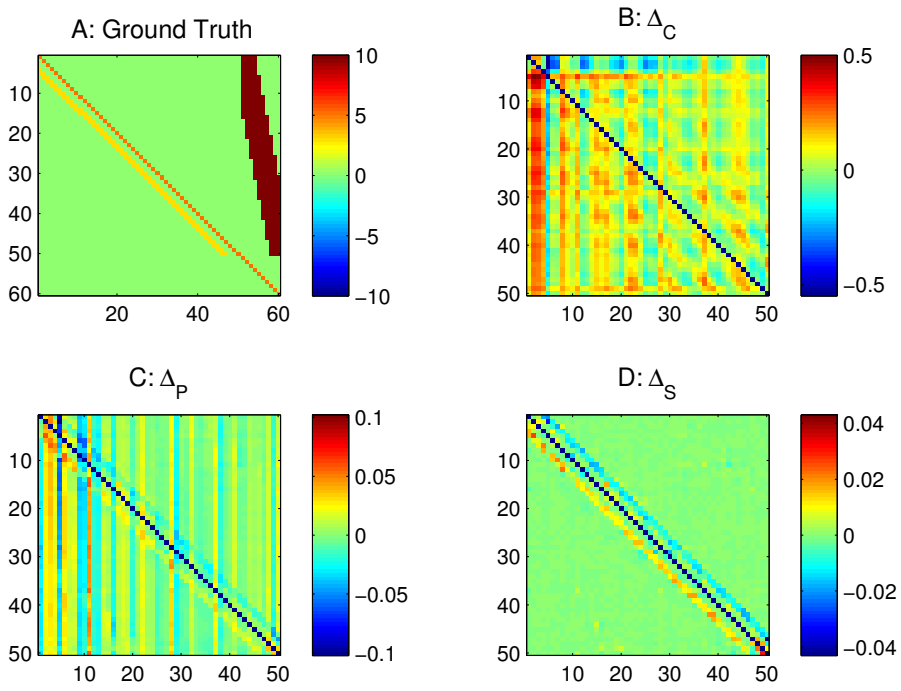

Figure 18: Supplementary2, differential covariance analysis of the passive neuron model with 5 ms fixed synaptic delay. The color in B,C,D indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method.

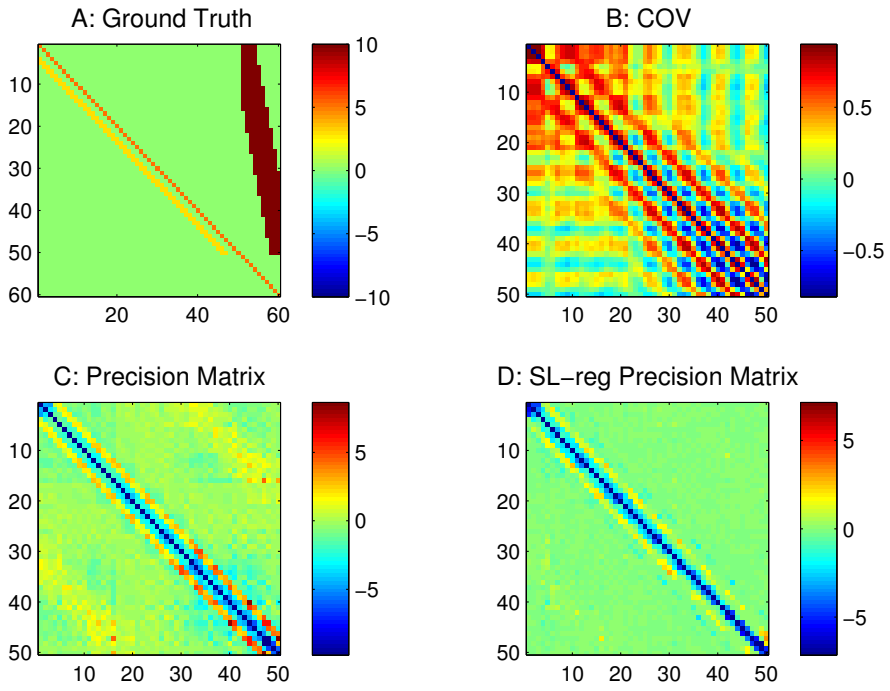

Figure 19: Supplementary3, passive neuron model with 0-10 ms uniformly distributed synaptic delay. Results from correlation-based methods. A) Ground truth connection matrix. neurons 1-50 are visible neurons. neurons 51-60 are invisible neurons. B) Estimation from the correlation method. C) Estimation from the precision matrix. D) Sparse+latent regularized precision matrix.

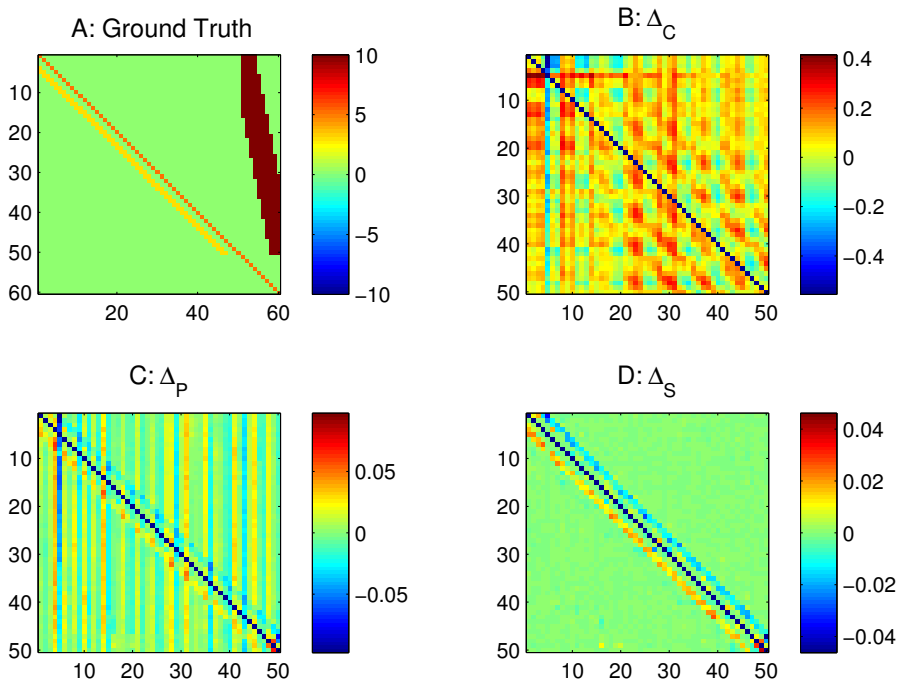

Figure 20: Supplementary4, differential covariance analysis of the passive neuron model with 0-10 ms uniformly distributed synaptic delay. The color in B,C,D indicates direction of the connections. For element A_{ij}, warm color indicates i is the sink, j is the source, i.e. $i \leftarrow j$, and cool color indicates j is the sink, i is the source, i.e. $i \rightarrow j$. A) Ground truth connection matrix. B) Estimation from the differential covariance method. C) Estimation from the partial differential covariance method. D) Estimation from the sparse+latent regularized partial differential covariance method.

References

Onureena Banerjee, Laurent El Ghaoui, Alexandre d'Aspremont, and Georges Natsoulis. Convex optimization techniques for fitting sparse gaussian graphical models. In *Proceedings of the 23rd international conference on Machine learning*, pages 89–96. ACM, 2006.

Claudia Battistin, John Hertz, Joanna Tyrcha, and Yasser Roudi. Belief propagation and replicas for inference and learning in a kinetic ising model with hidden spins. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(5):P05021, 2015.

Maxim Bazhenov, Igor Timofeev, Mircea Steriade, and Terrence J Sejnowski. Model of thalamocortical slow-wave sleep oscillations and transitions to activated states. *The Journal of Neuroscience*, 22(19):8691–8704, 2002.

Maxime Bonjean, Tanya Baker, Maxime Lemieux, Igor Timofeev, Terrence Sejnowski, and Maxim Bazhenov. Corticothalamic feedback controls sleep spindle duration in vivo. *The Journal of Neuroscience*, 31(25):9124–9134, 2011.

Cristiano Capone, Carla Filosa, Guido Gigante, Federico Ricci-Tersenghi, and Paolo Del Giudice. Inferring synaptic structure in presence of neural interaction time scales. *PloS one*, 10(3):e0118412, 2015.

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21 (2):572–596, 2011.

Jen-Yung Chen, Sylvain Chauvette, Steven Skorheim, Igor Timofeev, and Maxim Bazhenov. Interneuron-mediated inhibition synchronizes neuronal activity during slow oscillation. *The Journal of physiology*, 590(16):3987–4010, 2012.

David Roxbee Cox and Nanny Wermuth. *Multivariate dependencies: Models, analysis and interpretation*, volume 67. CRC Press, 1996.

Alain Destexhe. Spike-and-wave oscillations based on the properties of gabab receptors. *The Journal of neuroscience*, 18(21):9099–9111, 1998.

Benjamin Dunn and Yasser Roudi. Learning and inference in a nonequilibrium ising model with hidden nodes. *Physical Review E*, 87(2):022127, 2013.

Hua Fan, Xiuming Shan, Jian Yuan, and Yong Ren. Covariances of linear stochastic differential equations for analyzing computer networks. *Tsinghua Science & Technology*, 16(3):264–271, 2011.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1 (1):13–36, 2011.

Haiping Huang. Sparse hopfield network reconstruction with 1 regularization. *The European Physical Journal B*, 86(11):1–7, 2013.

Eric R Kandel, Henry Markram, Paul M Matthews, Rafael Yuste, and Christof Koch. Neuroscience thinks big (and collaboratively). *Nature Reviews Neuroscience*, 14(9): 659–664, 2013.

Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems*, pages 612–620, 2011.

PL Nunez and R Srinivasan. Electric fields of the brain oxford university press. *New York*, 2005.

Murat Okatan, Matthew A Wilson, and Emery N Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural computation*, 17(9):1927–1961, 2005.

Volker Pernice and Stefan Rotter. Reconstruction of sparse connectivity in neural networks from spike train covariances. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03008, 2013.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

Tingwei Quan, Xiuli Liu, Xiaohua Lv, Wei R Chen, and Shaoqun Zeng. Method to reconstruct neuronal action potential train from two-photon calcium imaging. *Journal of biomedical optics*, 15(6):066002–066002, 2010.

R Quian Quiroga, Zoltan Nadasdy, and Yoram Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*, 16(8):1661–1687, 2004.

Vahid Rahmati, Knut Kirmse, Dimitrije Marković, Knut Holthoff, and Stefan J Kiebel. Inferring neuronal dynamics from calcium imaging data using biophysical models and bayesian inference. *PLoS Comput Biol*, 12(2):e1004736, 2016.

Yasser Roudi and John Hertz. Mean field theory for nonequilibrium network reconstruction. *Physical review letters*, 106(4):048702, 2011.

Yasser Roudi, Benjamin Dunn, and John Hertz. Multi-neuronal activity and functional connectivity in cell assemblies. *Current opinion in neurobiology*, 32:38–44, 2015.

Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

Olav Stetter, Demian Battaglia, Jordi Soriano, and Theo Geisel. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Comput Biol*, 8(8):e1002653, 2012.

Ian H Stevenson and Konrad P Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–142, 2011.

Ian H Stevenson, James M Rebesco, Lee E Miller, and Konrad P Körding. Inferring functional connections between neurons. *Current opinion in neurobiology*, 18(6): 582–588, 2008.

I Timofeev, F Grenier, M Bazhenov, TJ Sejnowski, and Mircea Steriade. Origin of slow cortical oscillations in deafferented cortical slabs. *Cerebral Cortex*, 10(12): 1185–1199, 2000.

Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089, 2005.

Joshua T Vogelstein, Brendon O Watson, Adam M Packer, Rafael Yuste, Bruno Jedynak, and Liam Paninski. Spike inference from calcium imaging using sequential monte carlo methods. *Biophysical journal*, 97(2):636–655, 2009.

Matthias Winterhalder, Björn Schelter, Wolfram Hesse, Karin Schwab, Lutz Leistritz, Daniel Klan, Reinhard Bauer, Jens Timmer, and Herbert Witte. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal processing*, 85(11):2137–2160, 2005.

Dimitri Yatsenko, Krešimir Josić, Alexander S Ecker, Emmanouil Froudarakis, R James Cotton, et al. Improved estimation and interpretation of correlations in neural circuits. *PLoS Computational Biology*, 2:199–207, 2015.