Statistically inferred neuronal connections in subsampled neural networks strongly correlate with spike train covariance

Tong Liang[1, 2] and Braden A. W. Brinkman[2, *]

[1] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, 11794, USA*

[2] *Department of Neurobiology and Behavior, Stony Brook University, Stony Brook, NY, 11794, USA*

(Dated: February 1, 2023)

Statistically inferring neuronal connections from observed spike train data is a standard procedure for understanding the structure of the underlying neural circuits. However, the inferred connections seldom reflect true synaptic connections, being skewed by factors such as model mismatch, unobserved neurons, and limited data. On the other hand, spike train covariances, sometimes referred to as "functional connections," make no assumption of the underlying neuron models and provide a straightforward way to quantify the statistical relationships between pairs of neurons. The main drawback of functional connections compared to statistically inferred connections is that the former are not causal, whereas statistically inferred connections are often constrained to be. However, we show in this work that the inferred connections in spontaneously active networks modeled by generalized linear point process models strongly reflect covariances between neurons, not causal information. We investigate this relationship between the neuronal connections inferred with model-matched maximum likelihood inference and the corresponding spike train covariance in a nonlinear spiking neural network model. Strong correlations between inferred neuronal connections and spike train covariances are observed when many neurons are unobserved or when neurons are weakly coupled. This phenomenon occurs across different network structures, including random networks and balanced excitatory-inhibitory networks. A theoretical analysis of maximum likelihood solutions in analytically tractable cases elucidates how the inferred filters relate to ground-truth covariances of the neurons, and opens the door for future investigations.

INTRODUCTION

Identifying the strength and timescales of synaptic transmission between neuron pairs offers enormous opportunities to study how the computational properties of a network are shaped by its structure, which is of great interest not only to neuroscience and network science in general. Therefore, statistical methods to infer interactions between neuron pairs in simultaneously recorded spike train data become extremely valuable for understanding the encoding and decoding properties of many biological neural networks [1], although the inferred connections are oftentimes called "functional" or "effective" interactions to distinguish it from the true synaptic interactions. In the past decade, "shotgun" and "perturbation" based paradigms have been proposed with the hope that causal connection can be measured or inferred [2, 3], to overcome the fact that experimentally determining the true underlying dynamics of neuron pairs in large networks is still challenging and computationally expensive.

While it is desirable to map out the relationship between network properties and the true underlying network structure, it is generally difficult, if not impossible, to recover the true neuronal connections from recorded spike trains alone. This difficulty is exacerbated by the fact that only a fraction of neurons in a circuit can be recorded simultaneously in practice [4, 5]. Spike train covariance, on the other hand, offers an easy and straightforward way to quantify how neuron activities co-vary and makes no assumption of the underlying neuron model. While previous work has established the analytical relationships between the hidden neuron connections and the effective neuronal connections for observed neurons [4], how the *statistically inferred* effective connections from common inference procedures relate to the true causal connections and the spike train covariances is largely unknown.

To better understand how the synaptic interactions between neurons inferred in a statistical relate to the ground truth connections in the underlying generative model, we build a generalized linear point process model (GLM) to simulate spike trains in a 64 neuron network, and infer the effective neuronal connections between pairs of neurons using maximum likelihood estimation (MLE). We focus on the effect that subsampling neurons in the network has on the inferred connections, minimizing other possible artifacts by using the GLM as both our generative model and the model we use for inference.

This paper is organized as follows: in Results we first study how using different numbers of observed neurons and different amounts of spike train data in MLE affects the inferred filters, as compared to the ground-truth coupling filters, in MLE inferred filters vary as the number of observed neurons and data volume vary. We then show that the inferred filters strongly correlate with the empirically estimated covariances, with the correlation growing stronger as fewer neurons are observed or when synaptic connections are weak, in Spike train covariances strongly correlate with MLE inferred filters in sub-sampled networks. This finding is shown to hold for both random networks and networks with a more realistic balanced excitatory-inhibitory (EI) structure. As detailed in Analytic analysis of maximum likelihood inference using a Gaussian process approximation, for the specific case of a GLM with exponential nonlinearity, we are able to show analytically that the synaptic filters only have access to information about the statistical moments of the spike train process, which are non-causal, as opposed to response functions that encode causal information about network responses. We explicitly solve the maximum likelihood equations for the filter of a single observed neuron in some example networks, and contrast the result with the filters predicted by marginalizing out unobserved neurons [4]. We conclude in Discussion by addressing the interpretation of our results in the context of neuronal connection inference (Interpretation of the results) and comparing our results to those of related studies (Comparison to other work), as well as discussing the limitations of our approach and future directions for extensions of this work (Limitations of the study and future directions).

RESULTS

We build a generalized linear point process model (GLM) to simulate the spiking activities in an Erdős-Réyni (random) network of 64 neurons. This GLM can be interpreted as a model for the spike trains of a stochastic leaky integrate-and-fire model [6]. We use this as our generative model because this family of GLMs has been used extensively to fit neuron spike train data [1, 7]. In this model, the number of spikes a neuron i fires within a small window $[t, t+dt]$, $\dot{n}_i(t)dt$, follows a Poisson process conditioned on its past history,

$$\dot{n}_i(t)dt \sim \text{Poiss}\left[\Phi\left(\mu_i + \sum_j J_{ij} * \dot{n}_j\right)dt\right], \tag{1}$$

where $\Phi(x)$ is a nonlinear activation function, μ_i is the baseline drive for neuron i that sets the baseline firing rate, $J_{ij}(t)$ is the interaction or coupling filter from neuron j to neuron i, and $\Phi\left(\mu_i + \sum_j J_{ij} * \dot{n}_j\right)$ gives the instantaneous firing rate of neuron i at time t, with $J_{ij} * \dot{n}_j = \int_{-\infty}^{t} dt'\ J_{ij}(t-t')\dot{n}_j(t')$. We choose an exponential nonlinearity $\Phi(x) \propto \exp(x)$, both for the mathematical simplifications it offers and following previous work fitting neural spike trains [1].

There are few situations in which all neurons in a circuit can be recorded all at once; in most cases, such as *in vivo* recordings, only a subset of the neurons is observed. For example, in Fig. 1A, 3 neurons are observed out of a 64-neuron network, and the observed spike trains are displayed in Fig 1B. Thus, with the recorded spike train data we will only be able to statistically infer the neuronal connections between observed neuron pairs (though see [8–10] for attempts to infer unobserved units). To do so, we partition neurons into "observed" and "hidden" groups, and fit the model above only for the neurons in the observed group:

$$\dot{n}_i(t)dt \sim \text{Poiss}\left[\Phi\left(\hat{\mu}_i + \sum_{j \in obs} \hat{J}_{ij} * \dot{n}_j\right)dt\right],$$

where neuron indices i, j are only in the group of observed neurons and $\hat{J}_{ij}(t)$ and $\hat{\mu}_i$ are to be inferred. We numerically infer these unknowns with our simulated spike train data by maximum likelihood estimation (MLE). The procedures of simulating the spike trains and inferring the neuronal connection with MLE are discussed in depth in Spike train simulation with a linear-nonlinear Poisson cascade model and Neuronal connection inference with maximum likelihood estimation.

In order to infer the synaptic filters $\hat{J}_{ij}(t)$, one must parametrize the function, either by inferring the value of the filter at each time point (requiring as many parameters as the number time-bins used to represent the filter) or by representing them as weighted sums of basis functions and inferring the unknown weights. The basis function approach reduces the number of unknowns to the number of basis functions used, which requires less data than inferring each time point. However, the families of filter shapes that can be inferred are constrained by one's choice of basis functions, whereas inferring each time-point can represent any function given enough temporal resolution and data. In practice, the basis function representation is preferred, but for our analyses the time-point inference will reveal interesting relationships between the inferred filters and ground truth properties of the network.

In Fig. 1C, we show an example of the inferred coupling filters of the 3 observed neurons, both with and without using basis functions, in red solid lines and dots, respectively. As expected, the two approaches yield similar results, although the filters inferred without basis functions tend to be noisier. Notably, the inferred filters differ from the ground truth filters because only 3 out of 64 neurons in the network are observed, and interactions from the rest of the 61 hidden neurons to these 3 observed neurons are generally significant. We also show the normalized spike train covariances in grey bars along with the inferred filters. Surprisingly, the spike train covariances follow the filters inferred without basis functions fairly closely. This observation leads to our study of when and how those two seemingly different quantities correlate to each other.

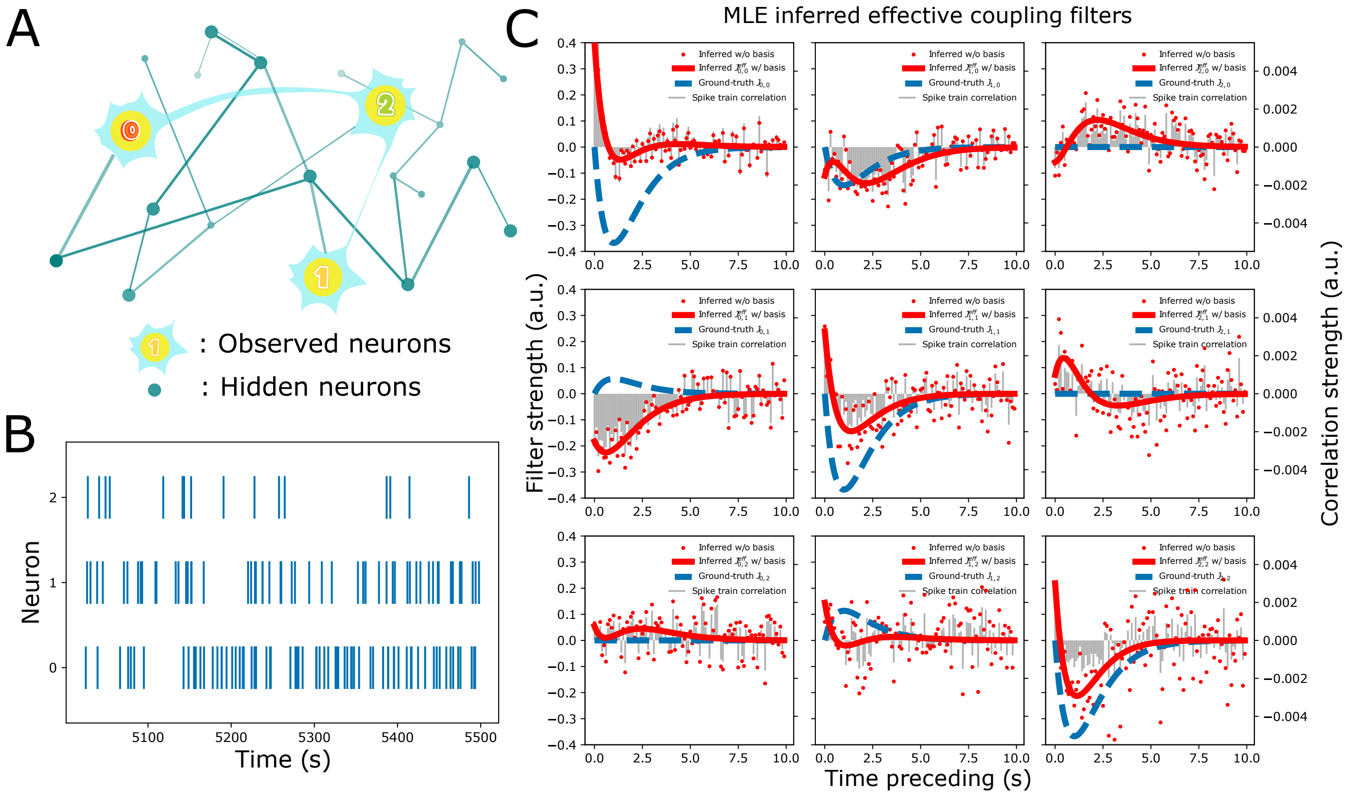

FIG. 1. **Schematics of the hidden neuron problem and effective neuronal connection inference**. **A.** A network of interconnected neurons with three observed and others hidden. **B.** Neuron spike trains recorded from three observed neurons, shown only for 500 s window. **C.** Neuronal connection/coupling filter inference with maximum likelihood estimation based on recorded spike trains. Filters inferred with and without using basis functions are shown, which are compared to the ground-truth coupling filters in blue dashed lines. Normalized spike train covariance is shown in gray bars with the scale on the right. The 0 lag correlation is suppressed for visualization purposes. The spike train correlations closely match the filters inferred without using basis functions.

MLE inferred filters vary as the number of observed neurons and data volume vary

It is expected that the inferred filters will not accurately recover the ground-truth filters when the inference model differs from the true underlying neuron models or if there exist unobserved neurons [4, 5]. In this work, we focus on model-matched inference where the inference model is the same as the ground-truth neuron model. Fig. 2A shows how the inferred filters change when different numbers of neurons are observed in the network, ranging from only 2 neurons to 64, the fully observed case. In this example, inferred filters in the fully observed case match the ground truth, while the inferred filters with fewer observed neurons differ more and more from ground-truth. It is worth noting that neither model-matched inference nor observing all neurons in a biological neural network is easily achievable in real experiments, thus understanding what role the subsampling plays in statistically inferred filters is important.

In addition to the number of observed neurons, the inference procedure is also constrained by the amount of spike train data available in the inference. In Fig. 2B, we inferred the 4 pairwise coupling filters for the first 2 neurons in the network using different amounts of spike train data. Our results confirm the intuition that using less data for inference leads to noisier inferred filters. However, it is also important to note that a coupled GLM is unlikely to be identifiable. This means that it is not guaranteed that the ground truth synaptic interactions are recovered in the fully observed network in the limit of infinite data.

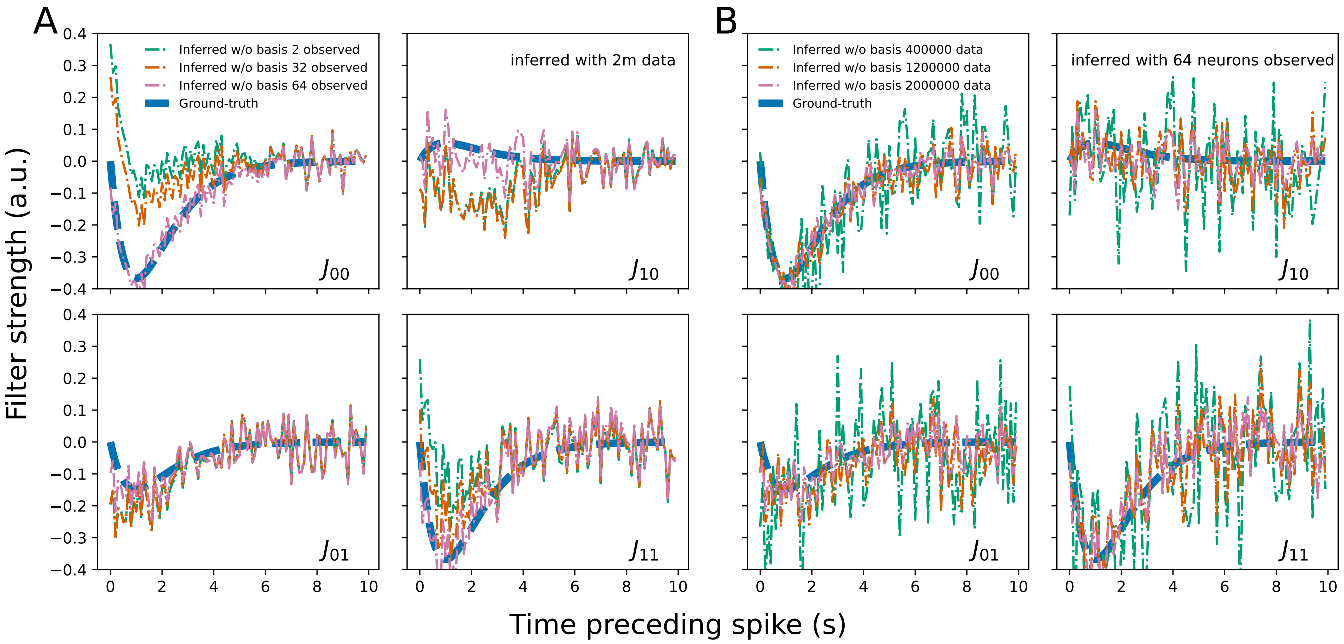

FIG. 2. **Maximum-Likelihood estimation (MLE) of coupling filters change as observed neurons and spike train data volume change**. **A.** The MLE inferred coupling filters for 2 neurons out of the 64 neurons network with 2 million spike train data, when different amounts of neurons are observed. MLE inferred coupling filters without using basis functions approach the ground truth filters in the fully observed case (64 neurons observed). However, when hidden neurons exist, the MLE inferred coupling filters differ from the ground truth as expected, as it differs more with fewer observed neurons. **B.** MLE inferred filters in the fully observed case, with different amounts of spike train data used in MLE inference. When more spike data is used, the inferred filters become less noisy, and vice versa.

Spike train covariances strongly correlate with MLE inferred filters in sub-sampled networks

As shown in Fig. 1C, the inferred filters strongly correlate with their corresponding spike train covariances. We use Pearson correlation to quantify how close these two quantities correlate and how their correlation change in different sampling and network conditions.

Since the MLE inferred filters and spike train covariances were estimated using two independent methods, it is surprising to observe such a strong correlation between them. This correlation persisted across many cases, varying the number of observed neurons, data volume, synaptic coupling strength, and even network architecture, as shown in Fig. 3A-E. We find the correlations are weaker in fully observed or nearly fully observed cases with large amounts of data, but otherwise our results did not seem to be dependent on, e.g., a weak coupling assumption, as the strong correlations persist for synaptic strengths close to the values for which the network would become unstable. Fig. 3F summarizes the median Pearson correlation between the MLE inferred filters and the corresponding spike train covariances, which clearly show the transition of correlation from high to low as more neurons in the network being observed or as more spike train data volume is used (however the latter happens consistently only when over 25% of neurons are observed). We focus here on the self-coupling filters and spike train auto-covariances in Fig. 3, but we show in Appendix Fig. 1 that the correlations of randomly sampled cross-coupling filters and their corresponding cross-covariances are also strong.

In the GLM neuron model used to generate spike trains in this work, the weight matrix coefficient J_0 controls the coupling strength of all of the neuron pairs. The previous results were derived with $J_0 = 3$, close to the largest possible coupling strength we can set to get a stable network for our choice of parameters. As we decreased J_0, the network entered a weaker coupling regime. Fig. 4A shows the change of the mean firing rates of 64 neurons in the random network as the weight matrix coefficient J_0 changes. Smaller J_0 confined the network to a noise-driven regime, where each neuron's firing rate is dominated by the same baseline drive set in the generative model, and a higher J_0 led the network into a strong coupling regime with more variable firing rates across neurons. As shown in Fig. 4B, in the weak coupling limit the Pearson correlations between the MLE inferred filters and the spike train covariances are high even if all the neurons in the network are observed.

To demonstrate that our results are not a quirk of random networks, we also consider balanced networks of excitatory and inhibitory (EI) populations, which are generally considered to be a more realistic network model [11, 12] because the synaptic strengths of excitatory neurons are all positive and the strengths of inhibitory neurons are all negative, as typically observed in real tissue. In Fig. 4D-F we show the strong correlations between inferred filters and spike-train covariance holds for these balanced excitatory-inhibitory networks. Here we use a 64-neuron EI network with 20% inhibitory neurons and 80% excitatory neurons and the details of the network generation process are discussed in Spike train simulation with a linear-nonlinear Poisson cascade model. We tune the weight matrix coefficient J_0 of the EI network from 1 to 7, beyond which the network becomes unstable. Our results also confirm that when the EI network is tuned to a weak coupling regime, as J_0 decreases, the correlations between the MLE inferred filters and the spike train covariances grow in strength, qualitatively similar to the behaviors observed in random networks in Fig. 4A-C.

Analytic analysis of maximum likelihood inference using a Gaussian process approximation

Our simulation results demonstrate a high degree of correlation between the inferred synaptic filters (inferred without using basis functions) and the empirically estimated spike-train covariances. As these quantities are computed by independent methods, some property of the network statistics or maximum likelihood inference procedure must give rise to these strong correlations when the network is subsampled. To better understand what may be going on here, we model the maximum likelihood estimation approach analytically, focusing on the effect that subsampling the network has. We will first derive the MLE equations for the spiking network model with an exponential nonlinearity, and then approximate the spike trains as a Gaussian process to analytically solve the MLE equations for some simple networks, which elucidates how the inferred filters are related to the spike train covariances.

Using the log-likelihood of the generative GLM model described in Eq. 13, we can analytically derive equations the maximum likelihood estimates must satisfy in the limit of infinite data. In the limit of infinitely many trials, the log-likelihood of the model is given by

$$\mathcal{L}(\{\mu\},\{J\}) = \left\langle \sum_{i=1}^{N_{\rm obs}} \int_{-\infty}^{\infty} dt \left\{ \dot{n}_i(t)\ln\Phi_i(t) - \Phi_i(t) - (\dot{n}_i(t)dt)! \right\} \right\rangle, \tag{2}$$

where the angled brackets $\langle \dots \rangle$ indicate expectation over the *true* spike-train process produced by the generative model; this expectation arises as the limit of an average over trials. The parameters to be inferred are $\hat{\mu}_i$ and $\hat{J}_{ij}(t)$, which occur inside the rate nonlinearity,

$$\Phi_i(t) = \lambda_0 \exp\left(\hat{\mu}_i + \sum_j \int dt'\ \hat{J}_{ij}(t-t')\dot{n}_j(t')\right),$$

where λ_0 is a constant and was set to 1 in all simulations. The maximum likelihood equations are obtained by taking derivatives of the log-likelihood with respect to $\hat{\mu}_i$ and $\hat{J}_{ij}(t)$. (We abuse notation and write the log-likelihood in continuous time here, but in practice we take the continuous time limit after deriving the MLE equations).

The resulting equations comprise a system of integral equations for the synaptic filters and baselines. We first derive the general relationship between spike trains and the unknown neuronal coupling filters in this model as follows:

$$\langle \dot{n}_i(t)\rangle = \langle \Phi_i(t)\rangle \tag{3}$$

$$\langle \dot{n}_i(t)\dot{n}_j(t-\tau)\rangle = \langle \Phi_i(t)\dot{n}_j(t-\tau)\rangle, \tag{4}$$

which take advantage of the nonlinearity $\Phi_i(t)$ being an exponential function. In deriving these equations we have assumed the network will be in a stationary steady-state, which means that Eq. 3 will be independent of time t and Eq. 4 will depend only on τ.

In the case of exponential nonlinearity, the expectations over the nonlinearity can be related to the moment gener-

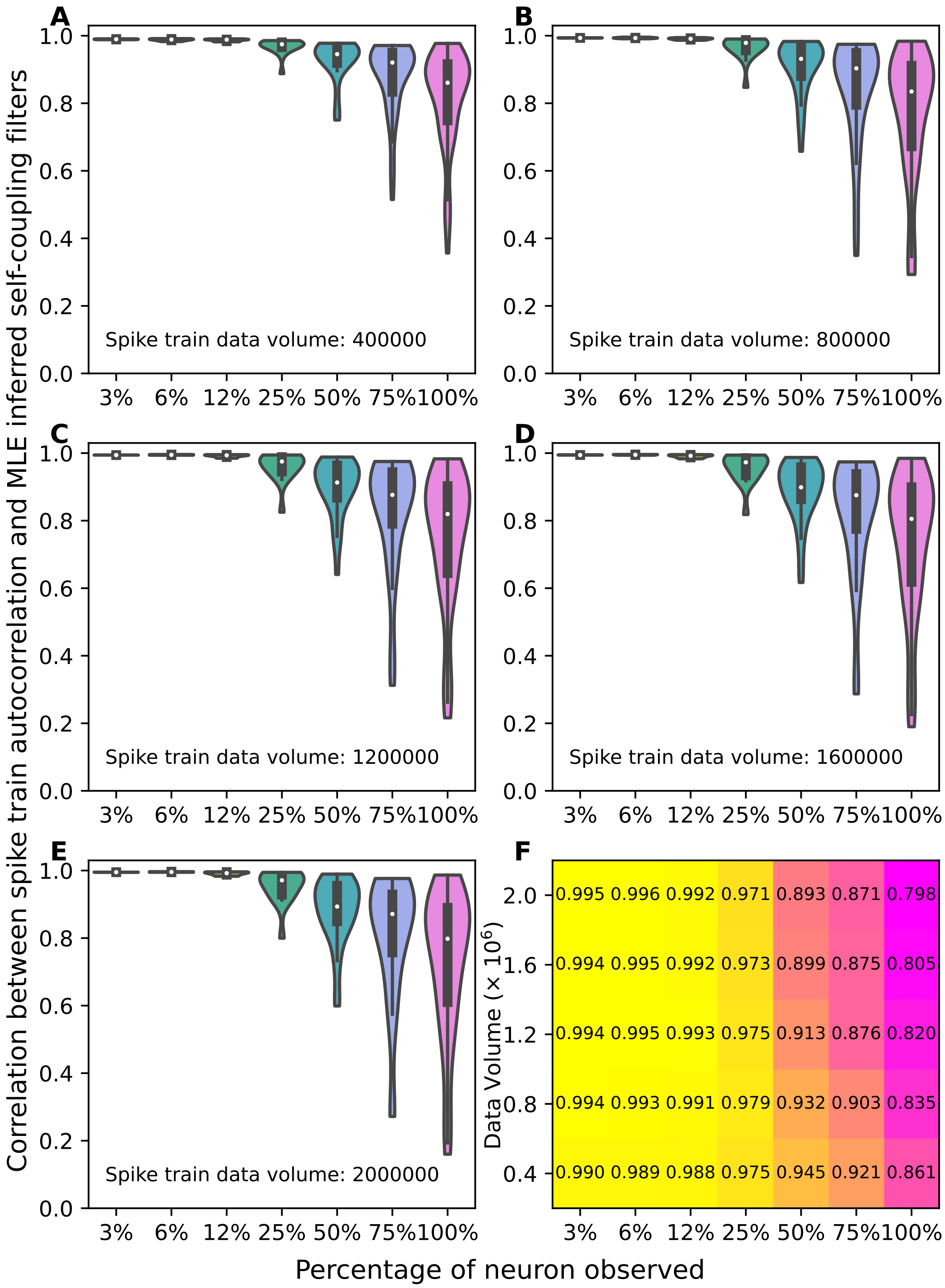

FIG. 3. **Pearson correlation between the spike train correlation and MLE inferred self-coupling filters in a strongly coupled random network of** 64 **neurons**. **A-E.** Violin plots show how the correlations change when the different number of neurons are observed and different amount of spike train data is used in inference. 3%, 6%, 12%, 25%, 50%, 75%, and 100% percentage of observed neurons in this 64 neuron network corresponds to 2, 4, 8, 16, 32, 48, and 64 neurons being observed. The spike train correlation functions strongly correlate with the MLE inferred filters when less neurons are observed and the correlation decreases as more neurons are observed. **F.** Median correlation values summarized from panel A-E showing a transition of the correlation from high to low as more neurons in the network are observed or as more spike train data is used in the MLE inference.

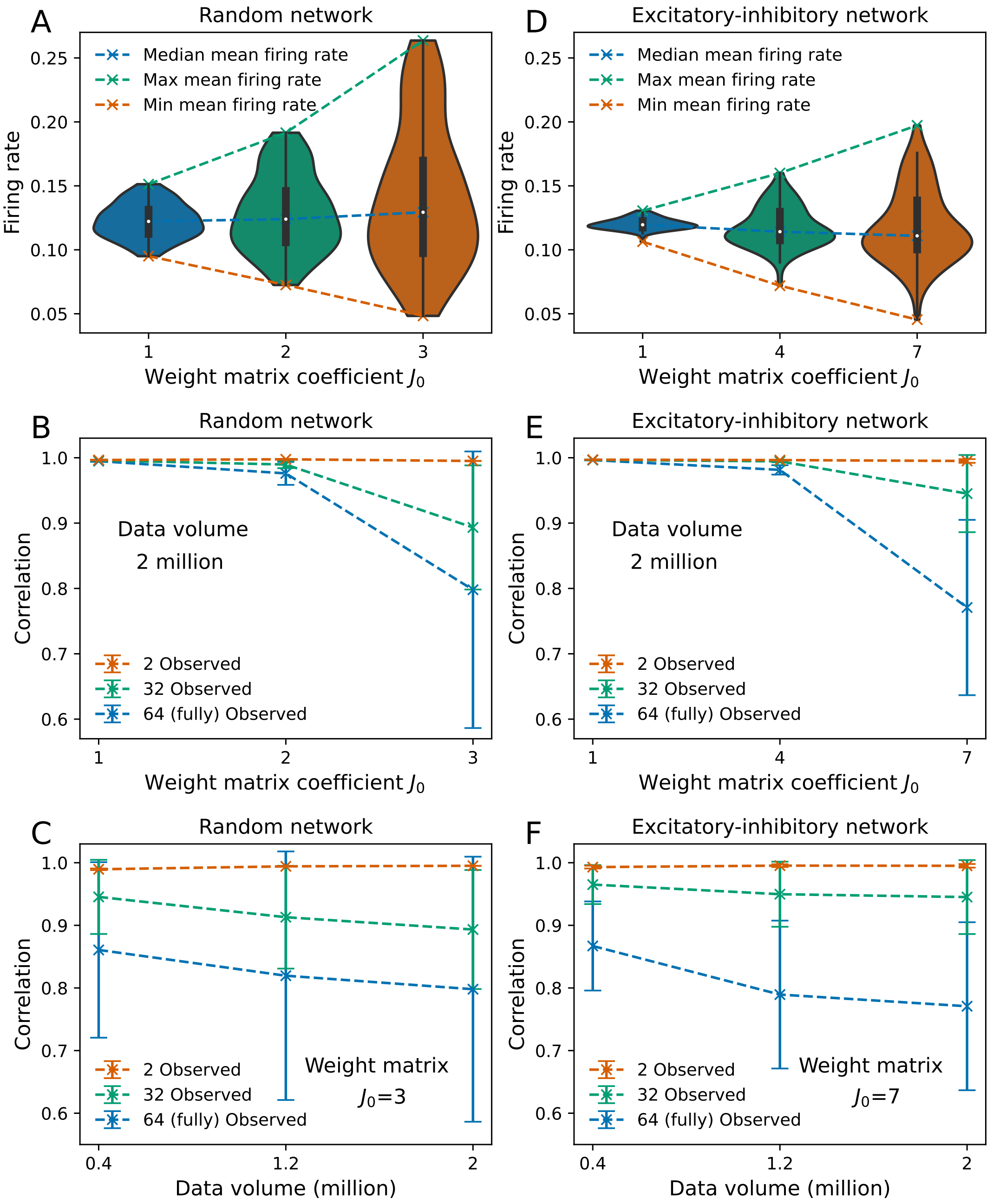

FIG. 4. **Correlations are high for both random and balanced excitatory-inhibitory networks in weak coupling regimes**. **A-C.** Random network. **A.** When the weight matrix coefficient J_0 decreased from 3 to 1, the network transitioned into a noise-driven regime, and the mean firing rates of the neurons varied less and were driven by the same baseline drive in the generative model. **B.** In the weak coupling regime, such as when $J_0 = 1$, high correlations between the MLE inferred filters and spike train covariances were observed even when all the neurons in the network are observed, as compared to the network in a strong coupling regime where the high correlation between the MLE inferred filters and spike train covariances only happen in the sub-sampled network.**C.** For a random network in a strong coupling regime ($J_0 = 3$), strong Pearson correlation coefficients were observed when the MLE was done for a sub-sampled network. Using less spike train data in the inference led to slightly higher correlations. **D-F.** Similar phenomenon holds for a balanced excitatory-inhibitory network, where 20% of the neurons are inhibitory and 80% of them are excitatory. The weight matrix coefficient J_0 was tuned from 1 to 7, with $J_0 = 7$ the largest possible integer value for which the spike train process is stable.

ating functional of the spike train process. For the right hand side of Eq. 3,

$$\langle \Phi_i(t) \rangle = \langle e^{\hat{\mu}_i + \sum_j \hat{J}_{ij} * (\langle \dot{n}_j \rangle + \delta \dot{n}_j)} \rangle = e^{\hat{\mu}_i + \sum_j \hat{J}_{ij} * \langle \dot{n}_j \rangle} Z[\tilde{j}_i(t') = \hat{J}_{ij}(t - t')], \tag{5}$$

where $\dot{n}_j = \langle \dot{n}_j \rangle + \delta \dot{n}_j$ and Z is the moment generating functional of the mean-subtracted spike train process, defined for an arbitrary "source" variable $\tilde{j}_i(t)$ as $Z[\tilde{j}] \equiv \langle \exp(\sum_i \int dt \; \tilde{j}_i(t) \delta \dot{n}_i(t)) \rangle$ (see the full derivation in MLE solution from the path integral formalism of spike train process). Similarly, for the right hand side of Eq. 4 we have

$$\langle \Phi_i(t) \dot{n}_i(t - \tau) \rangle = e^{\hat{\mu}_i + \sum_j \hat{J}_{ij} * \langle \dot{n}_j \rangle} \left. \frac{\delta Z[\tilde{j}]}{\delta \tilde{j}_j(t')} \right|_{\tilde{j}_j(t') = \hat{J}_{ij}(t - \tau)}. \tag{6}$$

The important implication of Eqs. 5 and 6 is that the moment generating functional contains only information about the non-causal statistical *moments* of the spike train process; they do not directly contain any information about *response functions*, which are causal. In a path integral formulation of this stochastic process, one can formulate a more general moment generating functional that contains information about both the statistical moments and the causal response functions of the process. Crucially, however, the information about the response functions drops out of the expectations we have computed, meaning that the MLE equations do not directly contain any information about the response functions. In fully observed systems relationships between the statistical moments and response functions can often be derived [13]; however, in the absence of a fully observed network it may not be possible to recover the response functions from the subset of observed moments. This result suggests that the synaptic filters $\hat{J}_{ij}(t)$, even when restricted to be causal, ultimately reflect information about the non-causal statistical moments, rather than the actual response functions of the network. In the limit of a fully observed network it may be possible to extract this causal information, but in a sub-sampled network our results imply that the causal nature of the inferred connections may not reflect any actual causal response properties of the network.

Eqs. 3 and 4 are too nonlinear to solve in general, but we may glean some insight by approximating the network as a Gaussian process, which amounts to neglecting the contribution of all statistical moments beyond pairwise statistics. We use Eq. 5 to eliminate the dependence on $\hat{\mu}_i$ in Eq. 6, leaving a system of Wiener-Hopf integral equations to solve for the filters $\hat{J}_{ij}(t)$:

$$C_{ij}(t - t') = r_i \sum_k \int dt'' C_{jk}(t' - t'') \hat{J}_{ik}(t - t''). \tag{7}$$

Here, $r_i = \langle \dot{n}_i \rangle$ is the mean firing rate of neuron i and $C_{ij}(t, t') = \langle \dot{n}_i(t) \dot{n}_j(t') \rangle - \langle \dot{n}_i(t) \rangle \langle \dot{n}_j(t') \rangle$ is the spike-train covariance. In practice these quantities would be estimated from data and hence known, but in this analysis we will estimate them using a mean-field analysis of the spiking network model.

Although Eq. 7 looks like it can be solved using Fourier transform methods, the left-hand side is only valid for $t - t' > 0$. If this restriction is neglected, the solution $\hat{J}_{ij}(t)$ will be non-causal in general. Solving this system of equations while imposing causality is difficult in general and still an area of active research, but the procedure is tractable for the case of a single observed neuron, which we highlight here.

To simplify the analytic calculations, we consider a network of all-to-all coupled neurons, $J_{ij}(t) = Jt \exp(-t/\tau)\Theta(t)/\tau^2$, including the self couplings $i = j$, and homogeneous baselines $\mu_i = \mu$. A mean-field analysis yields the following estimate of the mean firing rates,

$$r = \lambda_0 \Phi \left(\mu + NJr \right),$$

which is the same for every neuron due to the homogeneity of the network. For an exponential nonlinearity this equation can be solved in terms of the Lambert W function, $r = -(NJ)^{-1} W_{-1} \left(-NJ\lambda_0 e^{\mu} \right)$, defined as the solution of the transcendental equation $x = W_{-1}(x) \exp(W_{-1}(x))$ for which $-1/e < x < 0$. This restriction defines the branch of the Lambert W function that we must use for excitatory $J > 0$. It follows that $NJ < \exp(-(1 + \mu))/\lambda_0$ in order for the network to be stable.

The covariance of the neurons in this network is estimated to be

$$C_{ij}(t - t') = r \left[\delta_{ij} \delta(t - t') + \frac{(a_-^2 - b_+^2)(b_+^2 - a_+^2)}{(b_+ - b_-)(b_+ + b_-)} \frac{e^{-b_+ |t - t'|/\tau}}{2b_+ \tau} - \frac{(a_-^2 - b_-^2)(b_-^2 - a_+^2)}{(b_+ - b_-)(b_+ + b_-)} \frac{e^{-b_- |t - t'|/\tau}}{2b_- \tau} \right], \tag{8}$$

where $a_\pm = \sqrt{1+(N-1)Jr \pm \sqrt{(N-1)Jr(4-Jr)}}$ and $b_\pm = 1 \pm \sqrt{NJr}$. With this result, we can use the Wiener-Hopf procedure [14] to solve Eq. 7 for the self-history filter of a single observed neuron (see General solution of the integral equation for details). We find

$$\hat{J}(t) = \frac{1}{r}\left[\frac{(a_+ - b_-)(a_+ - b_+)}{a_+ - a_-}e^{-a_+ t/\tau} - \frac{(a_- - b_-)(a_- - b_+)}{a_+ - a_-}e^{-a_- t/\tau}\right]\frac{\Theta(t)}{\tau}. \tag{9}$$

In Fig. 5A, we plot the predicted self-coupling filter in Eq. 9 and the predicted spike train autocovariance for a single observed neuron in a 64-neuron network and compare it with the MLE inferred filter from 2 million spike train data.

A few remarks are in order. First, in the limit that $N \to 1$ this filter indeed recovers the ground-truth filter $J(t) = Jt\exp(-t/\tau)\Theta(t)/\tau^2$. Second, the resulting filter *does not agree* with the predictions of the effective filters of [4], which are obtained by marginalizing out the unobserved neurons. The reason for this discrepancy is subtle: when we make the Gaussian process approximation to close the MLE equations 5 and 6, this turns out to be tantamount to assuming the spike train fluctuations $\delta\dot{n}_i(t)$ are driven by independent Gaussian noise of variance r_i:

$$\delta\dot{n}_i(t) \approx \sum_j \int dt'\ \hat{J}_{ij}(t-t')\delta\dot{n}_j(t') + \xi(t),$$

where $\langle \xi_i(t)\xi_j(t')\rangle = r_i\delta_{ij}\delta(t-t')$. On the other hand, marginalizing out the hidden neurons and applying the Gaussian approximation is tantamount to assuming the neurons are driven by correlated noise $\langle \xi_i(t)\xi_j(t)\rangle = \Sigma_{ij}(t-t') \neq \delta_{ij}\delta(t-t')$. The two calculations are consistent, however, because both approaches produce the *exact same* set of spike-train covariances within the Gaussian approximation. One can transform between any two sets of filter-noise covariance pairs:

$$C_{ij}(\omega) = \sum_k \left(\mathbb{I} - r\mathbf{J}(\omega)\right)^{-1}_{ik}(\omega)\left(\mathbb{I} - r\mathbf{J}(-\omega)\right)\Sigma_{k\ell}(\omega). \tag{10}$$

If the noise covariance matrix can be factored as $\Sigma_{ij}(\omega) = \sum_{mn} S_{im}(\omega)S_{jn}(-\omega)\tilde{\Sigma}_{mn}(\omega)$ for some transformation matrix S and other covariance matrix $\tilde{\Sigma}$, then we can transform from one set of filters to another by $r\tilde{J}_{ij}(\omega) = \delta_{ij} - \sum_k (S^{-1})_{ik}(\omega)(\delta_{kj} - rJ_{kj}(\omega))$. (We give these expressions for our specific homogeneous all-to-all network, but the general expressions of the Gaussian approximation for any network J_{ij} are minor modifications of this result). The MLE equations we have derived here select the solution corresponding to independent driving, while the marginalization calculation of [4] corresponds to correlated driving noise. This implies that, at least for Gaussian processes, one cannot independently infer both the linear filters and the noise covariances without additional constraints or modeling assumptions.

With our approximate solution for the all-to-all network, we can analytically calculate the overlap coefficient between the inferred filter $\hat{J}(t)$ and the auto-covariance $C(t)$, excluding the delta-function peak at $t = 0$:

$$\rho \equiv \frac{\int_{0^+}^{\infty} dt\ \hat{J}(t)C(t)}{\sqrt{\int_0^{\infty} dt\ \hat{J}(t)^2 \int_{0^+}^{\infty} dt\ C(t)^2}}. \tag{11}$$

The overlap is equivalent to the Pearson correlation coefficient estimated from infinite time-points. The general expression for the alpha function filter is rather unwieldy, so we do not write it here. In Fig. 5B, we plot the overlap ρ as a function of $x = NJ\lambda_0 e^{1+\mu} \leq 1$, for various values of N. We observe that the overlap is generally very high away from the edge of stability of the network (at $x = 1$), but drops to 0 as the edge of stability is approached. This drop becomes increasingly rapid as N increases, approaching zero as

$$\rho \simeq 2^{5/4}N^{1/4}\left(1 - NJ\lambda_0 e^{1+\mu}\right)^{1/4} \tag{12}$$

for large N as $NJ\lambda_0 e^{1+\mu} \to 1$. As our single observed neuron represents a smaller and smaller fraction of the observed network when N increases, the window over which the correlation between $\hat{J}(t)$ and $C(t)$ drops to 0 is a range of order $1/N$. Thus, even when the synaptic strength of the network is quite strong, in a heavily subsampled network one must tune extremely close to the edge of stability to see a drop in the correlation.

While this is a simplified test case, it gives us valuable insight into what is occurring in our simulations. The inferred

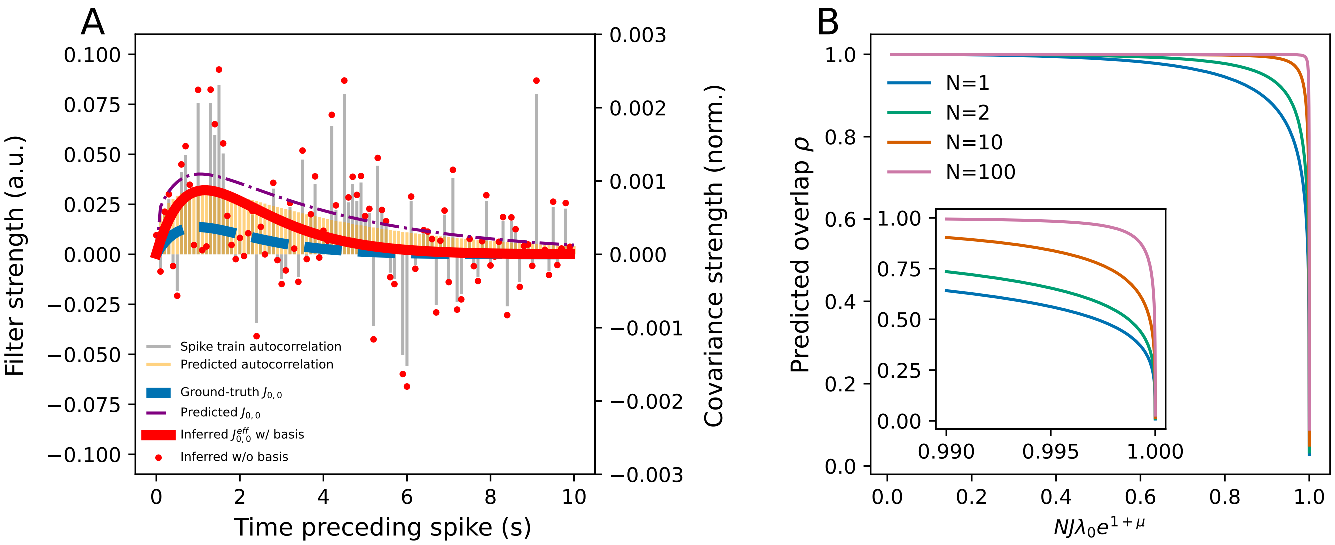

FIG. 5. **Demonstration of the predictions in a homogeneous all-to-all network**. **A.** A homogenous all-to-all network with $J = 0.037$, $\mu = -2$, $\lambda_0 = 1$ is simulated up to 2 million observation windows. The MLE inferred J_{00} (in red solid line and red dots) and estimated spike train covariances (in grey bars) are compared with the predicted J_{00} (in purple dotted line) and the predicted spike train autocorrelation (in orange bars). The predictions are similar in shape and magnitude to the ones derived from the simulated spike trains. **B.** The predicted overlap, equivalent to the Pearson correlation, drops as $NJ\lambda_0 e^{1+\mu}$ increases, with varying speed depending on the network size N. For a fixed N, the increase of $NJ\lambda_0 e^{1+\mu}$ amounts to the increase of the network coupling strength J. The inset shows an enlarged view of the drop.

filters obtained by maximum likelihood estimation are shaped by the spike-train covariances, not any causal response functions of the ground truth network. In a fully observed network, it may be possible to extract information about the response functions from the observed correlation functions, at least at the level of the Gaussian approximation (e.g., by solving Eq. 10 for $\delta_{ij} + r_i J_{ij}(\omega)$, given the ground-truth noise). However, when the network is subsampled it is likely not possible in general to recover these response functions, and the inferred filters increasingly reflect the spike train covariances, despite the imposition of causality on the inferred filters.

DISCUSSION

Interpretation of the results

In this work, we built a Poisson point process Generalized Linear Model (GLM) to study how well the commonly used maximum likelihood estimation (MLE) does at inferring neuronal connections in subsampled spiking neuron networks. Surprisingly, a strong correlation between the MLE inferred coupling filters and the corresponding spike train covariances with Pearson correlation coefficients above 0.99 was observed when a) there exist unobserved neurons in the network, i.e., subsampling the network, or b) the network is in a weak coupling regime dominated by spontaneous activities, as shown in Fig. 4. In the strongest coupling regime possible for these network models, a median Pearson correlation coefficient around 0.8 was still observed even when all the neurons in the network were observed, suggesting a strong tie between the MLE inferred coupling filters and corresponding spike train covariances. This phenomenon is robust both in random networks and in more realistic balanced excitatory-inhibitory networks. Furthermore, using less spike train data (decreasing the spike train data volume) in the inference also led to slightly stronger correlations between those two quantities. The observed strong correlation is puzzling because the inferred filters and the spike train covariances are derived from two seemingly independent routes, even if both are based on the same spike train data. Notably, inferring neuronal coupling filters with MLE requires a neuron model to calculate the likelihood function, while computing the spike train covariances is "model-free."

To gain insights into these phenomena, we turned to the path integral formulation of the spike train process of the GLM neuron model [4, 15, 16]. The standard Poisson GLM choice of an exponential inverse link function (the firing rate nonlinearity) enabled us to analytically derive equations that the maximum likelihood estimation must satisfy in Eq. 5 and Eq. 6, which reduced to Eq. 7 under a Gaussian process approximation of the spiking process. These equations reveal that the MLE inferred filters can only access information on the non-causal moments, not the causal response functions. It is also worth noting that although our analytic calculations for the MLE inferred filters only consider pairwise covariances, the general MLE equations Eq. 5 and Eq. 6 involve all higher order moments from the moment generating functionals. Thus while the MLE inferred filters in practice contain more information than pairwise correlations, they are still not causal. We further demonstrated our results by solving the MLE inferred self-coupling filter for a single observed neuron in a simplified homogeneous all-to-all network and calculated the expected MLE inferred filters, spike train covariances, and Pearson correlation coefficient between those two, as shown in Fig. 5. These demonstrations explained the observed strong correlations in different subsampling and network coupling regimes in Fig. 3 and Fig. 4.

Comparison to other work

Mapping out the network structure and inferring connections between neuron pairs from the recorded spike train data are challenging tasks, and understanding the results one obtains requires careful consideration of the assumptions underlying the statistical models fit to data. It is well-known that when subsampling neurons in a network, the inferred neuronal connections are *effective* or *functional* and may not reflect the casual ground-truth connections [17], although recent attention has shifted towards developing methods to infer *causal* connections in neural circuitry, either through Granger causality, information-theoretic measures, or novel sampling paradigms [2, 3, 18–21]. In particular, Kim *et al.* [18] applied Granger causality to neural spike trains to identify causal connections between recorded neurons by performing hypothesis testing based on the likelihood ratio test. Lepperød *et al.* [3] proposed to combine concepts of instrumental variable and difference in differences from econometrics to perform causal inference. Soudry *et al.* [2], on the other hand, proposed a shotgun sampling method that randomly samples overlapping subsets of the network over a period of time so that the reconstruction of the entire network could in principle be accomplished.

Furthermore, the covariances of neuron activity have historically been considered a different measure of how neuron pairs relate to each other, dubbed "functional connectivity," popular in fields like fMRI [22]. Pernice and Rotter [23], as well as Schiefer *et al.* [24], showed that it is possible to reconstruct neuronal connections from spike train covariances in certain sparse networks. Kobayashi *et al.* [25] proposed a method to fit spike train covariance functions with GLM in order to perform hypothesis testing in determining the existence of connections between neurons. However, the exact relationship between the spike train covariances and the ground-truth coupling filters is generally unknown and can be arbitrarily complex, involving not only the pairwise but also higher order moments [15, 26, 27].

Meanwhile, efforts have been made to infer neuronal connections among observed and hidden neurons through sophisticated statistical methods such as expectation-maximization and latent variable models [8–10, 28]. However, these methods either allow acausal connections between the observed and the hidden neurons [8] or perform well only when the number of hidden neurons is less than the observed neurons [9, 10] or require careful modeling of the hidden neuron populations [28]. More recently, Brinkman *et al.* [4] derived an analytical relationship between the ground-truth neuronal connections and the effective connections that unobserved neurons generate between subsets of observed neurons, although left the inference problem open, which this work seeks to address. Building upon previous studies, our work investigates the question of how the commonly used statistical inference procedure produces effective connections that may differ from or relate to the causal ground-truth connections as well as the spike train covariances in a GLM-based nonlinear spiking network model with different network structures, opening the door for investigating the role spike train covariance function plays in the inferred neuronal connections, in particular in causal connection inference. We found that the maximum likelihood procedure does not infer the effective interactions predicted by [4], which correspond to approximating fluctuations of the spiking process around their mean values as a Gaussian process driven by correlated noise, whereas the MLE solution is equivalent to assuming the fluctuations as Gaussian processes driven by independent noise. The two cases give the same spike-train correlations, however, and so they yield equivalent observable statistics.

Limitations of the study and future directions

While the findings in this work are very relevant to the ongoing study of inferring causal neuronal connections from recorded spike trains, our results are based on simulating the spike train process as a point process Poisson GLM and using MLE to perform a model-matched inference. The choice of the particular neuron model and inference method could potentially have an impact on the conclusions one can draw. Inferring neuronal connections from real neural data usually performs model-mismatched inference [5], as the underlying biological neuron model would always remain unknown. Our choice to focus on a statistical model that matches the form of the generative model (up to different numbers of neurons), likely improves the inference results compared to using a mismatched model. In the appendix we present a short calculation of how fitting a model with exponential nonlinearity to a generative model with a sigmoidal nonlinearity affects the inferred results, finding that the MLE inference will generically estimate weaker coupling strengths than ground truth (Ground-truth networks with non-exponential firing rate nonlinearity $\phi(V)$). It will be interesting to test the conclusion of this work with other neuron models and other inference methods.

Our analytical solutions leveraged the exponential nonlinearity in the inferred GLM, and focused on the steady state of the spike train process, which make the derivation of analytically tractable equations for MLE possible. While in principle a similar analysis could be performed using a inference model with a different nonlinearity, the MLE equations will be much more complicated.

We further simplified our MLE equations for the exponential nonlinearity by approximating the spike train process as a Gaussian process, limiting the statistical moments to second-order. One can go beyond the Gaussian level of approximation of the spike train process to estimate how the higher-order moments have an impact on the statistically inferred neuronal connections, though this would not change the result that the MLE fit only has access to non-causal moments.

Finally, one of the ultimate goals of this line of work is to identify the causal connections between neuron pairs, not just "effective" connections. Establishing the gold-standard causal connections in a network requires perturbation experiments and perhaps novel statistical inference methods that have the potential to make this inference possible in large-scale networks [3]. Such perturbative drives would manifest in our analytic formulas as introducing terms related to the causal response functions of the network. Building on our methodology by including such perturbative inputs will enable investigation of the origin of the inferred neuronal connections: how the non-causal covariances and causal response functions get mixed in, and under what conditions one dominates the other in the inference process.

METHODS

Spike train simulation with a linear-nonlinear Poisson cascade model

In this work, we use a generalized linear model (GLM) to simulate the neuron spike trains. In the GLM generative model, neurons emit spikes probabilistically following a Poisson process, with the rate given by $\Phi\left(\mu_i + \sum_j J_{ij} * \dot{n}_j\right)$. Discrete spikes are generated for a small observation window $dt = 0.1$. A first-order alpha function $J_{ij}(t) = W_{ij}\ t\ e^{-t/\tau}/\tau\Theta(t)$ is used as the ground-truth interaction filters that govern the interaction of neuron j's spike train history on neuron i's instantaneous firing rate. Causality is imposed through the Heaviside step function $\Theta(t) = 1$ if $t > 0$ and 0 otherwise. The weight matrix $\mathbb{W}$ with entry W_{ij} sets the interaction strength of the filters and $\tau = 1$ sets the typical timescale of the decay of the response. The spike train is simulated by solving a second-order differential equation with a 4th-order Runge-Kutta method to conveniently track the spike train history, following the method used in previous works [4, 15, 29]. While simulating the spike train data, we run the simulation up to 2 million observation windows, and use different amounts of the spike train data in inferring the coupling filters, e.g. in Fig. 3. The term "data volume" in the main text and figures refers to the spike train data up to that number of observation windows in the simulation.

Random network weight matrix generation. Following previous works [4, 15], we generated a 64-neuron random network weight matrix with a sparsity $p = 50\%$, so that only half of the connections are non-zero. The non-zero synaptic weight strengths were drawn independently from a normal distribution with zero mean and standard deviation $J_0/\sqrt{pN}$, where J_0 is the weight matrix coefficient and N is the number of neurons in the network. The weight matrix coefficient J_0 took three values 1, 2, 3 while we set the baseline drive $\mu_i = -2$ throughout the study. $J_0 = 3$ was the largest integer value we can set to still have a stable spike train process in the simulation, thus the network is considered to be in a high coupling regime in that case. Importantly, the diagonal entries of the weight matrix were always set to -1 for different J_0 to simulate a soft refractory period for the neurons from their own spike history.

EI network weight matrix generation. Following previous works [11, 12], we generated a 64-neuron excitatory-inhibitory (EI) network weight matrix with 20% (13) inhibitory neurons and 80% (51) excitatory neurons. Excitatory neurons make connections to excitatory neurons with a probability 20% and other all other neuron pairs (E-I and I-I) make connections with a probability 50%. The weight matrix coefficient J_0 was set to be a multiplication factor of the weight matrix. The weight of the excitatory connections was set to 0.04125 and the weight of the inhibitory connections was set to -0.16625 when $J_0 = 1$. Thus for the largest possible integer weight matrix coefficient $J_0 = 7$ in Fig. 4D-F, the excitatory and inhibitory weights were 0.28875 and -1.16375, respectively. The diagonal entries were always set to -1 for different J_0 to simulate a soft refractory period for the neurons from their own spike history.

Neuronal connection inference with maximum likelihood estimation

We infer the neuronal connections based on a generalized linear model with an exponential inverse-link function, which amounts to a model-matched inference given the same model used in generating the spike trains. The observed neuron spike train $\{\dot{n}_i(t)\}$ are assumed to follow $\dot{n}_i(t)dt \sim \text{Poiss}[\Phi(\hat{\mu}_i + \sum_{j\in obs} \hat{J}_{ij} * \dot{n}_j)dt]$, where $\hat{\mu}_i$ and $\hat{J}_{ij}$ are the inferred baseline drive and interaction filters to be determined and the observation window dt is set to 0.1. The likelihood function is thus,

$$L_i(\hat{\mu}, \hat{J}) = \text{Prob}(\{\dot{n}_i(t)\}|\hat{\mu}_i, \hat{J}_{ij}) = \prod_t \frac{(\Phi_i(t)dt)^{\dot{n}_i(t)dt}}{(\dot{n}_i(t)dt)!} e^{-\Phi_i(t)dt}. \tag{13}$$

We use the Tweedie regressor with power 1 and log link function in the scikit-learn (v.0.24.2) package to do the inference [30], which under the hood minimizes the unit deviance and can be shown to be equivalent to maximizing the likelihood function in Eq. 13. Note that no regularization penalty is added for all the inference procedures used in this work.

For the inference of the filters with basis functions as shown in Fig. 1C, alpha basis functions

$$\alpha_n(t) = t^n \exp(-t/\tau)\Theta(t)/\tau^n$$

of orders $n = 0$, 1, and 2 were used. In this scenario, the number of unknowns for inferring the filters decreases to 3, the same as the number of basis functions. The inferred neuronal connections are truncated at 100 observation windows, corresponding to 10 s for the chosen time window $dt = 0.1$ s.

For the inference of the filters without using basis functions, we use the same number of 100 observation windows, and thus 100 unknowns must be inferred to determine the coupling filters at each time point preceding the spikes.

MLE solution from the path integral formalism of spike train process

Maximizing the likelihood function in Eq. 13 amounts to solve for the zero points of its derivatives with respect to the unknowns J_{ij} and $\hat{\mu}_i$ in $\Phi_i(t) = \lambda_0 \exp(\hat{\mu}_i + \sum_j \hat{J}_{ij}\dot{n}_j)$. For mathematical simplicity, we take the logarithm of the likelihood to get the log-likelihood function $\mathcal{L}_i = \log(L_i)$ and maximize the log-likelihood,

$$\frac{\partial \mathcal{L}_i}{\partial \hat{\mu}_i} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} dt \left(\frac{\dot{n}_i(t)}{\Phi_i(t)} - 1\right) \partial_{\hat{\mu}_i}\Phi_i(t) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} dt \left(\frac{\dot{n}_i(t)}{\Phi_i(t)} - 1\right) \Phi_i(t) = 0, \tag{14}$$

where we note that $\partial_{\hat{\mu}_i}\Phi_i(t) = \Phi_i(t)$ for the choice of exponential nonlinearity. For a stationary system the time average will tend to the expected value due to ergodicity, and is equivalent to forming the log-likelihood using a large number of independent trials, which limit to the expected value for infinitely many trials. Thus, Eq. 14 can be simplified to

$$\langle \dot{n}_i(t) \rangle = \langle \Phi_i(t) \rangle . \tag{15}$$

Similarly,

$$\frac{\delta \mathcal{L}_i}{\delta \hat{J}_{ij}(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} dt' \left(\frac{\dot{n}_i(t')}{\Phi_i(t')} - 1\right) \partial_{\hat{J}_{ij}(t)}\Phi_i(t') = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} dt' \left(\frac{\dot{n}_i(t')}{\Phi_i(t')} - 1\right) \Phi_i(t')\dot{n}_j(t' - \tau) = 0, \tag{16}$$

where we note that $\partial_{\hat{J}_{ij}(t)}\Phi_i(t') = \Phi_i(t')\dot{n}_j(t' - t)$ for the choice of exponential nonlinearity and thus the equation can be reduced to Eq. 4.

As explained in the main text, for an exponential nonlinearity we can relate the expectations over $\Phi_i(t)$ to the moment-generating functional of the spiking process and set $\lambda_0 = 1$,

$$\langle \Phi_i(t) \rangle = e^{\hat{\mu}_i + \sum_j \hat{J}_{ij} * \langle \dot{n}_j \rangle} Z[\tilde{j}_i(t') = \hat{J}_{ij}(t - t')],$$

$$\langle \Phi_i(t) \dot{n}_i(t-\tau) \rangle = e^{\hat{\mu}_i + \sum_j \hat{J}_{ij} * \langle \dot{n}_j \rangle} \left. \frac{\delta Z[\tilde{j}]}{\delta \tilde{j}_j(t')} \right|_{\tilde{j}_j(t') = \hat{J}_{ij}(t-\tau)},$$

(Eqs. 5 and 6 in the main text), where $\dot{n}_j = \langle \dot{n}_j \rangle + \delta\dot{n}_j$ and $Z[\tilde{j}] \equiv \langle \exp(\sum_i \int dt\, \tilde{j}_i(t) \delta\dot{n}_i(t)) \rangle$. The moment generating functional cannot generally be solved in closed form, so to make use of these equations we will need an approximation. We will use mean-field theory with Gaussian fluctuation corrections to approximate the spike trains as a Gaussian process, for which the moment generating functional is known in closed form.

Our calculation of the moment generating functional of the spike train process makes use of the path-integral formalism based on the spiking network model [4, 15, 16, 31]. Following Ocker *et al.* [15], we introduce an auxiliary variable $\tilde{n}$, called the "response variable," then and the action of the spike train process under our GLM neuron model becomes

$$S[\tilde{n}, \dot{n}] = \sum_i \int dt \left[\tilde{n}_i(t)\dot{n}_i(t) - (e^{\tilde{n}_i(t)} - 1)\Phi\left(\mu_i + \sum_j (J_{ij} * \dot{n}_j)(t) \right) \right], \quad (17)$$

such that the joint probability distribution of the spike train and auxiliary variable follows,

$$\text{Prob}[\tilde{n}, \dot{n}] \propto e^{-S[\tilde{n}, \dot{n}]}. \quad (18)$$

Going forward, we make a change of variables $\dot{n}_i = r_i + \delta n_i$ where $r_i = \langle \dot{n}_i \rangle$ is the mean firing rate of neuron i, so that the expansion below is around the first moment of the spike train process. Eq. 17 can be split into free and interacting actions. We expand the action in powers of $\delta\dot{n}_i(t)$ and $\tilde{n}_i(t)$, keeping only terms to quadratic order, which amounts to the Gaussian process approximation,

$$S[\tilde{n}, \delta\dot{n}] \approx \sum_{ij} \int dt dt' \left\{ \tilde{n}_i (\Delta^{-1})_{ij}(t, t') \delta\dot{n}_j - \frac{1}{2} \tilde{n}_i(t)^2 \Phi_i \right\}, \quad (19)$$

where $(\Delta^{-1})_{ij}(t, t') = \delta_{ij}\delta(t - t') - \Phi_i^{(1)} J_{ij}(t - t')$ is the inverse of the linear response function and $\Phi_i^{(n)} = \frac{d^n \Phi(x)}{dx^n}|_{x = \mu_i + \sum_j J_{ij} * r_j}$ is the n^{th} derivative of the nonlinear activation function evaluated at the mean firing rate r_i. The linear order terms have been eliminating by imposing that r_i satisfies

$$r_i = \Phi_i \left(\mu_i + \sum_j J_{ij} * r_j \right), \quad (20)$$

and assuming a time-independent solution.

The quadratic action 19 corresponds to a Gaussian distribution for the fluctuations $\delta\dot{n}_i(t)$, which have zero mean and covariance

$$C_{ij}(t', t'') = \sum_k \int dt \Delta_{ik}(t', t) \Phi_k \Delta_{jk}(t'', t). \quad (21)$$

For a Gaussian process the moment generating functional of the spike train can be derived [31], giving

$$Z[\tilde{j}] = \int \mathcal{D}\tilde{n}(t) \mathcal{D}\delta\dot{n}(t) e^{\sum_i \int dt \tilde{j}_i(t) \delta\dot{n}_i(t)}\ e^{-S[\tilde{n}, \delta\dot{n}]} = \exp\left(\frac{1}{2} \sum_{ij} \int dt' dt'' \tilde{j}_i(t') C_{ij}(t', t'') \tilde{j}_j(t'') \right). \quad (22)$$

Combined with Eq. 5 and Eq. 6, the derived moment generating functional can be used to solve the MLE equations in Eq. 3 and Eq. 4, leading to the closed maximum likelihood estimation equations,

$$C_{ij}(t,t') = r_i \sum_k \int dt'' C_{jk}(t',t'')\hat{J}_{ik}(t-t''),$$

(Eq. 7 in the main text), which establishes the relationship between the spike train correlation function C_{ij} and the MLE inferred filters $\hat{J}_{ij}$ under Gaussian process approximation.

For the specific case of an all-to-all network with $J_{ij}(t) = Jte^{-t/\tau}\Theta(t)/\tau$, the mean field equation reduces to a single rate equation (due to homogeneity of the network),

$$r = \lambda_0 \exp\left(\mu + NJr\right),$$

where $\sum_j \int dt'\ J_{ij}(t-t')r_j$ integrates to NJr for constant r. By manipulating this into the form of the Lambert transcendental equation, $W(z)e^{W(z)} = z$ we can write the solution in terms of the Lambert W function,

$$r = -\frac{W_{-1}\left(NJ\lambda_0 e^{\mu}\right)}{NJ}.$$

Using Eq. 21 we can evaluate the covariance for this model, and insert it into Eq. 7 to solve for the inferred filter $\hat{J}(t)$ for a single neuron. Eq. 7 is a Wiener-Hopf integral equation that is difficult to solve in the multivariate case, but is tractable in the scalar case, which corresponds to a single observed neuron in our context. The expressions for the covariance and the Weiner-Hopf equation for the Gaussian approximation of the spiking network process are minor modifications of the equations one obtains for maximum likelihood inferences of a linear Gaussian stochastic process. We give the details of the calculations for the linear Gaussian process in the Appendix (Analysis of linear Gaussian networks), and modify the results to apply to the spiking network model here. Specifically, the results for the spiking network can be obtained from the linear Gaussian process on an all-to-all network by the replacements $J \to Jr$, $\hat{J}(t) \to r\hat{J}(t)$, $C(t) \to C(t)/r$, where r is the mean-field firing rate.

ACKNOWLEDGMENTS

We thank Il Memming Park, Yuan Zhao, Matthew Dowling for helpful discussions, and Siddharth Paliwal for sharing Python code to generate balanced E-I network adjacency matrices. TL thanks Dou Liang for creating Fig. 1A.

* braden.brinkman@stonybrook.edu

[1] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli, Nature **454**, 995 (2008).
[2] D. Soudry, S. Keshri, P. Stinson, M.-h. Oh, G. Iyengar, and L. Paninski, PLoS computational biology **11**, e1004464 (2015).
[3] M. E. Lepperød, T. Stöber, T. Hafting, M. Fyhn, and K. P. Kording, bioRxiv , 463760 (2022).
[4] B. A. Brinkman, F. Rieke, E. Shea-Brown, and M. A. Buice, PLoS computational biology **14**, e1006490 (2018).
[5] A. Das and I. R. Fiete, Nature Neuroscience **23**, 1286 (2020).
[6] S. Ostojic and N. Brunel, PLoS computational biology **7**, e1001056 (2011).
[7] I. M. Park, M. L. Meister, A. C. Huk, and J. W. Pillow, Nature neuroscience **17**, 1395 (2014).
[8] J. Pillow and P. Latham, Advances in Neural Information Processing Systems **20** (2007).
[9] B. Dunn and Y. Roudi, Physical Review E **87**, 022127 (2013).
[10] J. Tyrcha and J. Hertz, Mathematical Biosciences and Engineering **11**, 149 (2014).
[11] M. T. Schaub, Y. N. Billeh, C. A. Anastassiou, C. Koch, and M. Barahona, PLoS computational biology **11**, e1004196 (2015).
[12] T. Rost, M. Deger, and M. P. Nawrot, Biological cybernetics **112**, 81 (2018).
[13] R. Kubo, Reports on progress in physics **29**, 255 (1966).
[14] A. V. Kisil, I. D. Abrahams, G. Mishuris, and S. V. Rogosin, Proceedings of the Royal Society A **477**, 20210533 (2021).
[15] G. K. Ocker, K. Josić, E. Shea-Brown, and M. A. Buice, PLoS computational biology **13**, e1005583 (2017).
[16] M. Kordovan and S. Rotter, arXiv preprint arXiv:2001.05057 (2020).
[17] I. H. Stevenson, J. M. Rebesco, L. E. Miller, and K. P. Körding, Current Opinion in Neurobiology **18**, 582 (2008).
[18] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown, PLoS computational biology **7**, e1001110 (2011).
[19] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, Journal of computational neuroscience **30**, 17 (2011).
[20] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, PloS one **9**, e87636 (2014).
[21] R. Biswas and E. Shlizerman, Frontiers in Systems Neuroscience **16**, 817962 (2022).

[22] R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran, Scientific reports **10**, 1 (2020).

[23] V. Pernice and S. Rotter, Journal of Statistical Mechanics: Theory and Experiment **2013**, P03008 (2013).

[24] J. Schiefer, A. Niederbühl, V. Pernice, C. Lennartz, J. Hennig, P. LeVan, and S. Rotter, PLoS computational biology **14**, e1006056 (2018).

[25] R. Kobayashi, S. Kurita, A. Kurth, K. Kitano, K. Mizuseki, M. Diesmann, B. J. Richmond, and S. Shinomoto, Nature communications **10**, 4468 (2019).

[26] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter, PLoS computational biology **7**, e1002059 (2011).

[27] G. K. Ocker, Y. Hu, M. A. Buice, B. Doiron, K. Josić, R. Rosenbaum, and E. Shea-Brown, Current opinion in neurobiology **46**, 109 (2017).

[28] S. Wang, V. Schmutz, G. Bellec, and W. Gerstner, arXiv preprint arXiv:2205.13493 (2022).

[29] G. B. Ermentrout and D. H. Terman, in *Mathematical Foundations of Neuroscience* (Springer, 2010) pp. 157–170.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).

[31] C. C. Chow and M. A. Buice, The Journal of Mathematical Neuroscience (JMN) **5**, 1 (2015).

[32] "Inverse of constant matrix plus diagonal matrix," StackExchange; accessed January 03 2023.

APPENDIX

Appendix Fig. 1: While we focused on the correlation between self-coupling filters and their spike trains covariances in the main text, here we show that the correlations of the randomly sampled cross-coupling filters and the cross-covariances are also strong.

ANALYSIS OF LINEAR GAUSSIAN NETWORKS

It is useful to analyze the maximum likelihood equations for a simple linear Gaussian model, which will turn out to be formally similar to our equations for the Gaussian approximation of the spiking network model. Consider the Gaussian process $x_i(t)$ defined by

$$x_i(t) = \sum_{j=1}^{N} \int_{-\infty}^{\infty} dt' \; J_{ij}(t-t') x_j(t') + \xi_i(t) \tag{23}$$

where $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t) \xi_j(t') \rangle = \Sigma_{ij}(t-t')$. It follows that $x_i(t)$ is also a Gaussian process with mean zero.

The linear filters $J_{ij}(t-t')$ are assumed to be causal, such that $J_{ij}(t-t') = 0$ for $t - t' < 0$. In the calculations below we will work in continuous time, so we need to be careful with limits as the argument of the filters tends to zero, as $J_{ij}(0)$ is ambiguous. We will therefore introduce a small quantity ϵ such that when we need to make the causal behavior of $J_{ij}(t-t')$ explicit under integrals we will write

$$\int_{-\infty}^{\infty} dt' \; J_{ij}(t-t') = \int_{-\infty}^{t-\epsilon} dt' \; J_{ij}(t-t'),$$

such that the argument of $J_{ij}(t-t')$ is always $\geq \epsilon$. This convention will be important in the next section.

To calculate the covariance we first need to calculate $Q_{ij}(t,t') \equiv \langle x_i(t) \xi_j(t') \rangle$. We multiply Eq. 23 by $\xi_j(t')$ and average over the noise, giving

$$\begin{aligned}\langle x_i(t) \xi_j(t') \rangle &= \sum_{k=1}^{N} \int_{-\infty}^{\infty} dt_1 \; J_{ik}(t-t_1) \langle x_k(t_1) \xi_j(t') \rangle + \langle \xi_i(t) \xi_j(t') \rangle \\ \Rightarrow Q_{ij}(t,t') &= \sum_{k=1}^{N} \int_{-\infty}^{\infty} dt_1 \; J_{ik}(t-t_1) Q_{kj}(t_1,t') + \Sigma_{ij}(t-t').\end{aligned}$$

Thus, $Q_{ij}(t,t')$ is the solution to the equation

$$\sum_{k=1}^{N} \int_{-\infty}^{\infty} dt_1 \; \left[\delta_{ik} \delta(t-t_1) - J_{ik}(t-t_1) \right] Q_{kj}(t_1,t') = \Sigma_{ij}(t-t'); \tag{24}$$

To solve this equation, we define the linear response function $\Delta_{ij}(t,t')$ by:

$$\sum_{k=1}^{N} \int_{-\infty}^{\infty} dt \; \Delta_{\ell i}(t_2,t) \left[\delta_{ik} \delta(t-t_1) - J_{ik}(t-t_1) \right] = \delta_{\ell k} \delta(t_2 - t_1); \tag{25}$$

By multiplying on the right by some right-inverse $\Delta^R_{\ell p}(t_1 - t_3)$, we can show that $\Delta_{\ell p}(t_2 - t_3) = \Delta^R_{\ell p}(t_2 - t_3)$; i.e., $\Delta_{ij}(t-t')$ is both the left and right inverse of $\delta_{ij}\delta(t-t') - J_{ij}(t-t')$.

Using $\Delta_{kj}(t_1,t')$, we can solve for

$$Q_{\ell j}(t_2,t') = \sum_i \int dt \; \Delta_{\ell i}(t_2,t) \Sigma_{ij}(t-t') \tag{26}$$

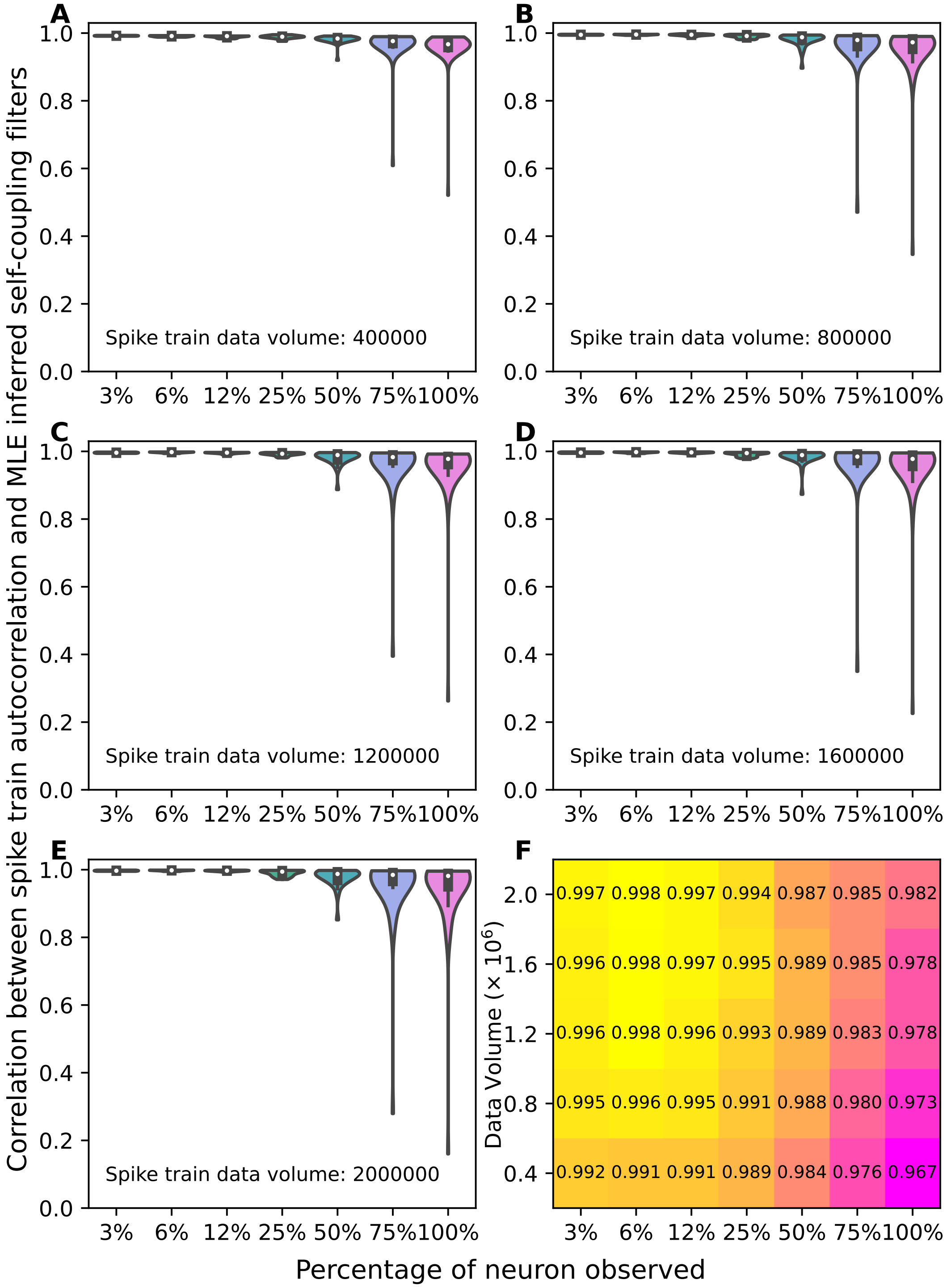

FIG. 1. **Peasron correlations are strong also for randomly sampled cross-coupling filters and the corresponding cross-covariances**. **A-E.** Violin plots show how the correlations change when the different number of neurons are observed and different amount of spike train data is used in inference. 3%, 6%, 12%, 25%, 50%, 75%, and 100% percentage of observed neurons in this 64 neuron network corresponds to 2, 4, 8, 16, 32, 48, and 64 neurons being observed. The normalized spike train covariance functions strongly correlate with the MLE inferred filters when in all conditions. **F.** Median correlation values are summarized from panel A-E showing a transition of the correlation coefficients.

We can now calculate the covariance by multiplying Eq. 23 by $x_j(t')$ and averaging:

$$\langle x_i(t)x_j(t')\rangle = \sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ J_{ik}(t-t_1)\langle x_k(t_1)x_j(t')\rangle + \langle \xi_i(t)x_j(t')\rangle$$
$$\Rightarrow C_{ij}(t,t') = \sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ J_{ik}(t-t_1)C_{kj}(t_1,t') + Q_{ji}(t',t)$$

Rearranging,

$$\sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ \left[\delta_{ik}\delta(t-t_1) - J_{ik}(t-t_1)\right] C_{kj}(t_1,t') = Q_{ji}(t',t)$$
$$\Rightarrow \sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ \left[\delta_{ik}\delta(t-t_1) - J_{ik}(t-t_1)\right] C_{kj}(t_1,t') = Q_{ji}(t',t)$$

Multiplying on the left by $\Delta_{\ell i}(t_2,t)$,

$$\sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ \sum_{i=1}^{N}\int_{-\infty}^{\infty} dt\ \Delta_{\ell i}(t_2,t)\left[\delta_{ik}\delta(t-t_1) - J_{ik}(t-t_1)\right] C_{ij}(t,t') = \sum_{i=1}^{N}\int_{-\infty}^{\infty} dt\ \Delta_{\ell i}(t_2,t)Q_{ji}(t',t)$$
$$\sum_{k=1}^{N}\int_{-\infty}^{\infty} dt_1\ \delta_{\ell k}\delta(t_2-t_1)C_{ij}(t,t') = \sum_{i=1}^{N}\int_{-\infty}^{\infty} dt\ \Delta_{\ell i}(t_2,t)Q_{ji}(t',t).$$

Inserting $Q_{ji}(t',t) = \sum_k \int dt''\ \Delta_{jk}(t'-t'')\Sigma_{ki}(t''-t)$ gives the final expression,

$$C_{\ell j}(t_2-t') = \sum_{i=1}^{N}\sum_{k=1}^{N}\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt''\ \Delta_{\ell i}(t_2-t)\Delta_{jk}(t'-t'')\Sigma_{ki}(t''-t). \tag{27}$$

If we define $(\Delta^T)_{kj}(t''-t') = \Delta_{jk}(t'-t'')$, then we can write this result in a matrix-like notation as $\mathbf{C} = \mathbf{\Delta\Sigma\Delta}^T$.

Maximum likelihood estimation of the linear Gaussian model

We now derive a system of equations for the filters $\hat{J}$ in this linear Gaussian model. For simplicity, we will assume our statistical model matches the form of the generative model. i.e., we want to fit a model of the form

$$x_i(t) = \sum_{j=1}^{N}\int_{-\infty}^{\infty} dt'\ \hat{J}_{ij}(t-t')x_j(t') + \xi_i(t)$$

to our data and infer the filters $\hat{J}_{ij}(t-t')$. We may assume the noise is correlated, it will not affect the equations for the filters. The log-likelihood of the model is

$$\mathcal{L}[\hat{J}] = -\frac{1}{2}\left\langle \sum_{i,j}\int_{-\infty}^{\infty} dtdt'\ \left(x_i(t) - \sum_k \int_{-\infty}^{\infty} dt_1\ \hat{J}_{ik}(t-t_1)x_k(t_1)\right)(\Sigma^{-1})_{ij}(t,t')\left(x_j(t') - \sum_{\ell=1}^{N}\int_{-\infty}^{\infty} dt_2\ \hat{J}_{j\ell}(t'-t_2)x_\ell(t_2)\right)\right\rangle \tag{28}$$
$$= -\frac{1}{2}\sum_{ijk\ell}\int dtdt'dt_1dt_2\ (\Delta^{-1})_{ik}(t,t_1)(\Sigma^{-1})_{ij}(t,t')(\Delta^{-1})_{j\ell}(t',t_2)C_{k\ell}(t_1,t_2) \tag{29}$$

where the angled brackets are averages over the true process (23) in the infinite trial limit and we assume $\hat{J}_{ij}(t-t') = 0$ for $t' \geq t$.

Taking a functional derivative of this with respect to $\hat{J}_{ij}(\tau)$, we obtain

$$\begin{aligned}\frac{\delta\mathcal{L}[\hat{J}]}{\delta\hat{J}_{ij}(\tau)} &= -\sum_{i'j'k\ell}\int dtdt'dt_1dt_2\ \frac{(\Delta^{-1})_{i'k}(t,t_1)}{\delta\hat{J}_{ij}(\tau)}(\Sigma^{-1})_{i'j'}(t,t')(\Delta^{-1})_{j'\ell}(t',t_2)C_{k\ell}(t_1,t_2)\\ &= -\sum_{i'j'k\ell}\int dtdt'dt_1dt_2\ \left(-\delta_{ii'}\delta_{kj}\delta(t-t_1-\tau)\right)(\Sigma^{-1})_{i'j'}(t,t')(\Delta^{-1})_{j'\ell}(t',t_2)C_{k\ell}(t_1,t_2)\\ &= \sum_{j'\ell}\int dtdt'dt_2\ (\Sigma^{-1})_{ij'}(t,t')(\Delta^{-1})_{j'\ell}(t',t_2)C_{j\ell}(t-\tau,t_2)\\ &= \sum_{j'}\int dtdt'\ (\Sigma^{-1})_{ij'}(t,t')\left[\sum_{\ell}\int dt_2(\Delta^{-1})_{j'\ell}(t',t_2)C_{j\ell}(t-\tau,t_2)\right].\end{aligned}$$

In the first line we performed a relabeling of indices to combine the two product rule terms into one.

We now impose that this derivative must be equal to zero. The inverse covariance can be eliminated by using the infinite integrals of integration to shift $t\to t+\tau$, moving the free variable τ into Σ^{-1}:

$$0=\sum_{j'}\int dtdt'\ (\Sigma^{-1})_{ij'}(t+\tau-t')\left[\sum_{\ell}\int dt_2(\Delta^{-1})_{j'\ell}(t'-t_2)C_{j\ell}(t-t_2)\right];$$

we have also used the fact that the covariances and filters are of the convolutional form. Now, we multiply both sides of the equation by $\Sigma_{ki}(\tau'-\tau)$ and sum over i and integrate over τ, employing the inverse relation

$$\sum_i\int d\tau\ \Sigma_{ki}(\tau'-\tau)(\Sigma^{-1})_{ij'}(\tau-(t'-t))=\delta_{kj'}\delta(\tau'-(t'-t)).$$

The noise covariance drops out of the equation, leaving

$$0=\sum_{\ell}\int dtdt_2\ (\Delta^{-1})_{k\ell}(t+\tau'-t_2)C_{j\ell}(t-t_2).$$

We again use the infinite range of integration to shift $t_2\to t_2+t$, eliminating t from the integrand. This leaves an overall integral over t, which would be infinite, but we can impose that the integral must be zero to for all t to avoid this issue. Thus,

$$\begin{aligned}0&=\sum_{\ell}\int dt_2\ (\Delta^{-1})_{k\ell}(\tau'-t_2)C_{j\ell}(-t_2)\\ &=\sum_{\ell}\int dt_2\left[\delta_{k\ell}\delta(\tau'-t_2)-J_{k\ell}(\tau'-t_2)\right]C_{j\ell}(t-\tau-t_2)\\ \Rightarrow C_{jk}(-\tau')&=\sum_{\ell}\int_{-\infty}^{\infty}dt_2\ \hat{J}_{k\ell}(\tau'-t_2)C_{j\ell}(-t_2).\end{aligned}$$

We can perform another shift $t_2\to\tau'-t_2$ and use the fact that $C_{ij}(\tau)=C_{ji}(-\tau)$ to write our equation as

$$C_{kj}(\tau)=\sum_{\ell}\int_{-\infty}^{\infty}dt_2\ \hat{J}_{k\ell}(t_2)C_{j\ell}(t_2-\tau)$$

where we also drop the prime on the variable τ. Using our constraint that the filter is causal, we may write, after some addition variable relabeling,

$$C_{ij}(t)=\sum_{\ell=1}^{N}\int_{+\epsilon}^{\infty}dt'\ \hat{J}_{i\ell}(t')C_{j\ell}(t'-t).\tag{30}$$

We can replace $\epsilon\to0^+$, where it is understood that contributions from generalized functions like $\delta(t')$ will not contribute to the integral.

General solution of the integral equation

Eq. 30 is an integral equation for the unknown filters $\hat{J}_{i\ell}(t')$. One might hope to be able to extend the limits of integration to the entire real line and take a Fourier transform to obtain a matrix system of equations that can be solved, but without explicitly imposing the causality constraint this procedure will generally yield a non-causal solution. To use the Fourier method, one first needs to generalize the equation to

$$G_{ij}(t) = \sum_{\ell=1}^{N} \int_{-\infty}^{\infty} dt'' \; \hat{J}_{i\ell}(t'') C_{j\ell}(t'' - t), \tag{31}$$

where

$$G_{ij}(t) = \begin{cases} C_{ij}(t), & t > 0 \\ G^{-}_{ij}(t), & t \leq 0 \end{cases} \tag{32}$$

for some unknown functions $G^{-}_{ij}(t)$ that must be determined as part of our solution. Although we have introduced an extra set of unknowns, once we solve for $\hat{J}_{i\ell}(t'')$ the $G^{-}_{ij}(t)$'s will be determined. This extra set of functions enables causal solutions for the filter by absorbing any non-causal pieces into them. We apply the Fourier transform to obtain

$$G^{+}_{ij}(\omega) + G^{-}_{ij}(\omega) = \sum_{\ell=1}^{\infty} \hat{J}^{+}_{i\ell}(\omega) C_{\ell j}(\omega),$$

where we use $C_{j\ell}(t) = C_{\ell j}(-t)$ and defined the transforms

$$\begin{aligned} f^{+}(\omega) &= \int_{0+}^{\infty} dt \; e^{-i\omega t} f(t), \\ f^{-}(\omega) &= \int_{-\infty}^{0^{+}} dt \; e^{-i\omega t} f(t), \\ f(\omega) &= \int_{-\infty}^{\infty} dt \; e^{-i\omega t} f(t). \end{aligned}$$

We can write the equation to solve in matrix form,

$$\mathbf{G}^{+}(\omega) + \mathbf{G}^{-}(\omega) = \hat{\mathbf{J}}^{+}(\omega)\mathbf{C}(\omega).$$

Next, we assume we can decompose $\mathbf{C}(\omega) = \mathbf{S}_{+}(\omega)\mathbf{S}_{-}(\omega)$, where $\mathbf{S}_{+}(\omega)$ is analytic and non-vanishing in the upper half plane and $\mathbf{S}_{-}(\omega)$ is analytic and non-vanishing in the lower half plane.

Continuing, we assume $\mathbf{S}_{-}(\omega)$ has an inverse, such that we may write

$$\mathbf{G}^{+}(\omega) \left[\mathbf{S}_{-}(\omega)\right]^{-1} + \mathbf{G}^{-}(\omega) \left[\mathbf{S}_{-}(\omega)\right]^{-1} = \hat{\mathbf{J}}^{+}(\omega)\mathbf{S}_{+}(\omega).$$

Next, we split $\mathbf{G}^{+}(\omega) \left[\mathbf{S}_{-}(\omega)\right]^{-1} = (\mathcal{F}^{-1}[\mathbf{G}^{+} \left[\mathbf{S}_{-}\right]^{-1})^{+}(\omega) + (\mathcal{F}^{-1}[\mathbf{G}^{+} \left[\mathbf{S}_{-}\right]^{-1}])^{-}(\omega)$, where the two terms are defined by first taking the inverse Fourier transform of the left-hand side and then splitting the Fourier transform up into the ± components. We can then rearrange our equation as

$$(\mathcal{F}^{-1}[\mathbf{G}^{+} \left[\mathbf{S}_{-}\right]^{-1}])^{-}(\omega) + \mathbf{G}^{-}(\omega) \left[\mathbf{S}_{-}(\omega)\right]^{-1} = \hat{\mathbf{J}}^{+}(\omega)\mathbf{S}_{+}(\omega) - (\mathcal{F}^{-1}[\mathbf{G}^{+} \left[\mathbf{S}_{-}\right]^{-1})^{+}(\omega),$$

where by construction the left-hand-side has all of its poles in the lower half plane and the right hand side has all of its poles in the upper half plane. Because the two sides are analytic on different half-planes, the only possibility is that they are both equal to the same function, which must be polynomial of degree n if we require the growth at $|\omega| \to \infty$ to be less than $\mathcal{O}(\omega^{n})$ [14]. If we demand that the filters decay as $|\omega| \to \infty$ (which excludes a δ-function component), then the only option is that the two sides must be equal to zero, and hence we arrive at the formal solution

$$\hat{\mathbf{J}}^{+}(\omega) = (\mathcal{F}^{-1}[\mathbf{G}^{+}\,[\mathbf{S}_{-}]^{-1})^{+}(\omega)[\mathbf{S}_{+}(\omega)]^{-1}. \tag{33}$$

In practice, the primary obstacles in performing this procedure are finding a spectral decomposition of the kernel $\mathbf{C}(\omega)$ and then splitting up $\mathbf{G}^{+}(\omega)[\mathbf{S}_{-}(\omega)]^{-1}$ into its separate additive factors that are analytic on different half planes. For a one-dimensional system there is a general procedure for performing both of these steps, but for a system of equations the non-commutativity of matrices prohibits the use of the scalar method.

To this end, in this work we will focus our analytic investigations on the case of a single observed neuron, in order to glean at least some analytic insights into the maximum likelihood inference procedure. This is best illustrated with some concrete examples, which we work through in the next section.

EXAMPLE CASE: ALL-TO-ALL COUPLED NETWORK DRIVEN BY INDEPENDENT NOISE

To evaluate an explicit example, we consider an all-to-all coupled network with $J_{ij}(t-t') = Jg(t-t')$, for some temporal profile $g(t)$; we evaluate the solutions explicitly for an exponential filter, which has simpler analytic expressions, and then give the corresponding results for alpha function filters. We will assume the driving noise $\xi_i(t)$ to be independent white noise of unit variance, $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t')$. In this case we can solve for the response function Δ by first Fourier-transforming Eq. (23) and then performing the matrix inversion:

$$\sum_{k=1}^{N}[\delta_{ik} - Jg(\omega)] = \delta_{ij},$$

where $g(\omega)$ is the Fourier transform of $g(t)$. If we denote $\mathbb{I}$ as the identity and $\mathbf{P}$ as a matrix of all 1's, then the inverse

$$[a\mathbb{I} + b\mathbf{P}]^{-1} = \frac{1}{a}\mathbb{I} - \frac{b}{a(Nb+a)}\mathbf{P}$$

[32]. In our case we have $a = 1$, $b = -Jg(\omega)$, giving

$$\Delta_{ij}(\omega) = \delta_{ij} + \frac{Jg(\omega)}{1 - NJg(\omega)}.$$

Let's now assume an exponential filter $g(t) = \exp(-t/\tau)\Theta(t)/\tau$, which has Fourier transform $g(\omega) = 1/(1+i\omega\tau)$ using the convention $g(\omega) = \int_{-\infty}^{\infty} dt e^{-i\omega t} g(t)$. Thus, in the time-domain $\Delta_{ij}(t)$ is given by

$$\begin{aligned}\Delta_{ij}(t) &= \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}e^{i\omega t}\left(\delta_{ij} + \frac{J}{g(\omega)^{-1} - NJ}\right)\\ &= \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}e^{i\omega t}\left(\delta_{ij} + \frac{J}{i\omega\tau + 1 - NJ}\right)\\ &= \delta_{ij}\delta(t) + J\exp(-(1-NJ)t/\tau)\Theta(t)/\tau,\end{aligned}$$

where to evaluate the second term we used the residue theorem: factoring out a $i\tau$ from the denominator, we observe a pole at $\omega = i(1-NJ)$ in the upper-half plane when $1 > NJ$. This restriction requires either $0 < J < 1/N$ or $J < 0$ for the process to be stable. For $t < 0$, $i\omega t = -iR|t|(\cos\theta + i\sin\theta)$ on a contour of radius R, and the real part of this, $+R|t|\sin\theta$ is only negative in the lower-half plane, so we must close the contour there and the integral evaluates to zero because there are no poles contained in the contour. For $t > 0$ the real part of the arc is $-R|t|\sin\theta$, and we must close the arc in the upper half plane, obtaining the contribution from the pole.

It is important to note that while $\Delta_{ij}(t-t')$ happens to be symmetric in the indices, it is *not* symmetric in time, as it contains a causal piece. (This makes sense, because $x_i(t)$ is causally dependent on the noise $\xi_i(t)$). In a more complicated network, Δ would also not be symmetric in the indices, because $\langle x_i(t)\xi_j(t')\rangle$ is not invariant under an exchange of indices.

We can now calculate the covariance $C_{ij}(t)$. In the Fourier domain we have

$$\begin{aligned}
C_{ij}(\omega) &= \sum_{k=1}^{N} \Delta_{ik}(\omega)\Delta_{jk}(-\omega) \\
&= \sum_{k=1}^{N} \left(\delta_{ik} + \frac{J}{i\omega\tau + 1 - NJ}\right)\left(\delta_{jk} + \frac{J}{-i\omega\tau + 1 - NJ}\right) \\
&= \sum_{k=1}^{N} \left(\delta_{ik}\delta_{jk} + \frac{J\delta_{jk}}{i\omega\tau + 1 - NJ} + \frac{J\delta_{ik}}{-i\omega\tau + 1 - NJ} + \frac{J^2}{|i\omega\tau + 1 - NJ|^2}\right) \\
&= \delta_{ij} + 2J\mathrm{Re}\left[\frac{1}{i\omega\tau + 1 - NJ}\right] + \frac{NJ^2}{(1-NJ)^2 + (\omega\tau)^2} \\
&= \delta_{ij} + \frac{2J(1-NJ) + NJ^2}{(i\omega\tau + 1 - NJ)(-i\omega\tau + 1 - NJ)} \\
&= \delta_{ij} + \frac{J(2-NJ)}{(i\omega\tau + 1 - NJ)(-i\omega\tau + 1 - NJ)}
\end{aligned}$$

We used the fact that the noise covariance is $\Sigma_{ij}(\omega) = \delta_{ij}$. We again evaluate the inverse Fourier transform by using the Residue theorem. There are now two symmetric poles at $\omega = \pm i(1 - NJ)/\tau$, so we get a contribution from both planes, as expected for a covariance. The result is

$$C_{ij}(t) = \delta_{ij}\delta(t) + \frac{J(2-NJ)}{2(1-NJ)}\frac{\exp(-(1-NJ)|t|/\tau)}{\tau}.$$

Solution for a single observed unit

Solving for $\hat{\mathbf{J}}^+(t)$ analytically is in general difficult, but there is a general prescription for the scalar case, meaning we should be able to obtain an exact solution in the case of a single observed unit. We calculate this here for unit $i = 1$ in the all-to-all network. The equation to solve is

$$G^+(\omega) + G^-(\omega) = \hat{J}^+(\omega)C(\omega),$$

where

$$\begin{aligned}
G^+(\omega) &= \int_{0^+}^{\infty} d\tau\ e^{-i\omega\tau}\left(\delta(\tau) + \frac{a}{2b\tau}e^{-b|t|/\tau}\right) \\
&= \frac{a}{2b}\frac{1}{i\omega\tau + b},
\end{aligned}$$

where we introduce $a = J(2 - NJ)$ and $b = 1 - NJ$ to simplify the upcoming formulas. The full Fourier transform of $C(\omega)$

$$C(\omega) = 1 + \frac{a}{(i\omega\tau + b)(-i\omega\tau + b)}.$$

Therefore, we need to solve the equation

$$\begin{aligned}\frac{a}{2b}\frac{1}{i\omega\tau+b}+G^-(\omega)&=\hat{J}^+(\omega)\left(1+\frac{a}{(i\omega\tau+b)(-i\omega\tau+b)}\right)\\&=\hat{J}^+(\omega)\left(\frac{(i\omega\tau+b)(-i\omega\tau+b)+a}{(i\omega\tau+b)(-i\omega\tau+b)}\right)\\&=\hat{J}^+(\omega)\left(\frac{-(i\omega\tau)^2+b^2+a}{(i\omega\tau+b)(-i\omega\tau+b)}\right)\\&=\hat{J}^+(\omega)\left(\frac{(i\omega\tau+\sqrt{b^2+a})(-i\omega\tau+\sqrt{b^2+a})}{(i\omega\tau+b)(-i\omega\tau+b)}\right)\end{aligned}$$

We now separate the factors that are analytic and non-vanishing on the lower-half-plane and the upper half-planes. We have

$$\frac{a}{2b}\frac{1}{i\omega\tau+b}\frac{-i\omega\tau+b}{-i\omega\tau+\sqrt{b^2+a}}+G^-(\omega)\frac{-i\omega\tau+b}{-i\omega\tau+\sqrt{b^2+a}}=\hat{J}^+(\omega)\left(\frac{i\omega\tau+\sqrt{b^2+a}}{i\omega\tau+b}\right).$$

We use partial fractions on the left-hand-side to write

$$\frac{a}{2b}\frac{1}{i\omega\tau+b}\frac{-i\omega\tau+b}{-i\omega\tau+\sqrt{b^2+a}}=\frac{A}{i\omega\tau+b}+\frac{B}{-i\omega\tau+\sqrt{b^2+a}},$$

where

$$A=\frac{a}{b+\sqrt{b^2+a}},\quad B=-\frac{a}{2b}\frac{\sqrt{b^2+a}-b}{\sqrt{b^2+a}+b}.$$

Here we will only care about the filter $\hat{J}^+(\omega)$, so we only need the A term. After separating the terms analytic in the upper versus lower half planes, demanding that the filters decay at infinite ω means we must have

$$\begin{aligned}\hat{J}^+(\omega)\left(\frac{i\omega\tau+\sqrt{b^2+a}}{i\omega\tau+b}\right)&=\frac{a}{b+\sqrt{b^2+a}}\frac{1}{i\omega\tau+b}\\\Rightarrow\hat{J}^+(\omega)&=\frac{a}{b+\sqrt{b^2+a}}\frac{1}{i\omega\tau+\sqrt{b^2+a}}\end{aligned}$$

Because $\hat{J}^+(\omega)$ only has poles in the lower half plane, as desired, we know it will be causal and we can use the regular Fourier transform to recover it in the time domain (as $\hat{J}^+(\omega)=\hat{J}(\omega)$). The result is

$$\hat{J}(t)=\frac{a}{b+\sqrt{b^2+a}}\frac{e^{-\sqrt{b^2+a}\ t/\tau}}{\tau}\Theta(t).$$

Restoring $a=J(2-NJ)$ and $b=1-NJ$ gives

$$\hat{J}(t)=\frac{J(2-NJ)}{1-NJ+\sqrt{(1-NJ)^2+J(2-NJ)}}\frac{e^{-\sqrt{(1-NJ)^2+J(2-NJ)}\ t/\tau}}{\tau}\Theta(t). \tag{34}$$

We can check that we recover the true filter when $N=1$: $(1-J)^2+J(2-J)=1-2J+J^2+2J-J^2=1$, and verify in Mathematica that this solution does satisfy the original integral equation.

Now that we have $\hat{J}(t)$ and $C(t)$ we can evaluate the normalized overlap between them,

$$\rho=\frac{\int_0^\infty dt\ \hat{J}(t)C(t)}{\sqrt{\int_0^\infty dt\ \hat{J}(t)^2\int_0^\infty dt\ C(t)^2}}.$$

Using Mathematica, this works out to

$$\rho = \frac{2\sqrt{1-NJ}\sqrt[4]{(1-NJ)^2 + J(2-NJ)}}{1-NJ+\sqrt{(1-NJ)^2+J(2-NJ)}}.$$

Plotting this as a function of $x = NJ \in [0,1)$ for fixed N, we see that for small x $\rho \approx 1$, and rapidly approaches 0 as $x \to 1$ from below. As N increases the fraction of the range of x for which $\rho \approx 1$ increases.

Alpha function filter

The manipulations work similarly for an alpha function filter $g(t) = te^{-t/\tau}\Theta(t)/\tau^2$, which has Fourier transform $g(\omega) = 1/(i\omega\tau + 1)^2$. This introduces more poles to deal with when using the residue theorem and partial fraction decomposition, but the calculations are tractable for the most part. For the linear Gaussian network model we find the covariance of the units to be

$$C_{ij}(t-t') = \delta_{ij}\delta(t-t') + \frac{(a_-^2 - b_+^2)(b_+^2 - a_+^2)}{(b_+ - b_-)(b_+ + b_-)}\frac{e^{-b_+|t-t'|/\tau}}{2b_+\tau} - \frac{(a_-^2 - b_-^2)(b_-^2 - a_+^2)}{(b_+ - b_-)(b_+ + b_-)}\frac{e^{-b_-|t-t'|/\tau}}{2b_-\tau}, \tag{35}$$

where $a_\pm = \sqrt{1 + (N-1)J \pm \sqrt{(N-1)J(4-J)}}$ and $b_\pm = 1 \pm \sqrt{NJ}$, and the effective self-history filter to be

$$\hat{J}(t) = \frac{(a_+ - b_-)(a_+ - b_+)}{a_+ - a_-}e^{-a_+ t/\tau} - \frac{(a_- - b_-)(a_- - b_+)}{a_+ - a_-}e^{-a_- t/\tau}\frac{\Theta(t)}{\tau}. \tag{36}$$

The reader will notice these expressions are similar to those given in the main text for the Gaussian approximation of the spiking network model. Our results for the spiking network model can be obtained from these results by rescaling $J \to Jr$, $\hat{J}(t) \to r\hat{J}(t)$, and $C(t) \to C(t)/r$.

SUBSAMPLING THE LINEAR GAUSSIAN MODEL

In this section we derive the effective action for the linear Gaussian model and marginalize out unobserved units to derive an action for the subsampled network. In doing so we will show explicitly that the effective filters calculated by this procedure, akin to the method of [4] for the spiking network model, do not match the filters predicted by maximum likelihood inference.

We may write the probability distribution as a path integral

$$P[x(t)] = \int \mathcal{D}\tilde{x}\; e^{-S[\tilde{x},x]},$$

where $S[\tilde{x},x]$ is the action

$$S[\tilde{x},x] = \sum_{i=1}^{N}\int_{-\infty}^{\infty} dt\, \left\{\tilde{x}_i(t)\left[x_i(t) - \sum_{j=1}^{N}\int_{-\infty}^{t^-} dt'\; J_{ij}(t-t')x_j(t')\right] - \frac{1}{2}\tilde{x}_i(t)^2\right\} \tag{37}$$

To subsample the network we divide the sums into recorded and hidden neurons:

$$S[\tilde{x},x] = S_{\rm rec}[\tilde{x},x] + S_{\rm hid}[\tilde{x},x] + \sum_{r,h}\int dtdt'\;\left\{\tilde{x}_r(t)J_{rh}(t-t')x_h(t') + \tilde{x}_h(t)J_{hr}(t-t')x_r(t')\right\},$$

where $S_{\rm rec}[\tilde{x},x]$ and $S_{\rm hid}[\tilde{x},x]$ have the same form as Eq. (37) but the sums only extend over the recorded and hidden neuron subsets, respectively. We marginalize over the hidden units, which requires that we evaluate the expectation

$$\left\langle e^{\sum_{r,h}\int dtdt'\;\left\{\tilde{x}_r(t)J_{rh}(t-t')x_h(t') + \tilde{x}_h(t)J_{hr}(t-t')x_r(t')\right\}}\right\rangle;$$

this can be recognized at the definition of the moment generating function with sources $\tilde{j}_h(t) = \sum_r \int dt\; \tilde{x}_r(t)J_{rh}(t-t')$

and $j_h(t) = \sum_r \int dt'\ J_{hr}(t-t')x_r(t')$. For a zero-mean Gaussian process the result is [31]

$$Z[j,\tilde{j}] = e^{\sum_{hh'} \int dtdt'\ \left\{\frac{1}{2}\tilde{j}_h(t)C_{hh'}(t-t')\tilde{j}_{h'}(t')+\tilde{j}_h(t)\Delta_{hh'}(t-t')j_{h'}(t')\right\}},$$

where $\Delta_{hh'}(t-t')$ is the linear response function of the network, including only hidden neurons, and $C_{hh'}(t-t') = \sum_{h''} \int dt''\ \Delta_{hh''}(t-t'')\Delta_{h'h''}(t'-t'')$ is the covariance function, again including only hidden units. The effective action for the recorded units is therefore

$$\begin{aligned} S_{\text{eff}}[\tilde{x},x] = S_{\text{rec}}[\tilde{x},x] - \sum_{rr'} \int dt_1 dt_2 \Bigg\{ \frac{1}{2}\tilde{x}_r(t_1) \left(\sum_{hh'} \int dtdt'\ J_{rh}(t_1-t)J_{r'h'}(t_2-t')C_{hh'}(t-t') \right) \tilde{x}_{r'}(t_2) \\ + \tilde{x}_r(t_1) \left(\sum_{hh'} \int dtdt'\ J_{rh}(t_1-t)\Delta_{hh'}(t-t')J_{hr'}(t'-t_2) \right) x_{r'}(t_2) \Bigg\} \\ = \sum_r \int_{-\infty}^{\infty} dt\ \left\{ \tilde{x}_r(t) \left[x_r(t) - \sum_{r'} \int_{-\infty}^{t^-} dt'\ J^{\text{eff}}_{rr'}(t-t')x_j(t') \right] \right\} - \frac{1}{2}\sum_{rr'} \int dtdt'\ \tilde{x}_r(t)\Sigma^{\text{eff}}_{rr'}(t-t')\tilde{x}_{r'}(t'), \end{aligned}$$

where the effective filter is

$$J^{\text{eff}}_{rr'}(t-t') = J_{rr'}(t-t') + \sum_{hh'} \int dt_1 dt_2\ J_{rh}(t-t_1)\Delta_{hh'}(t_1-t_2)J_{hr'}(t_2-t') \tag{38}$$

and the effective noise covariance is

$$\Sigma^{\text{eff}}_{rr'}(t-t') = \delta_{rr'}\delta(t-t') + \sum_{hh'} \int dt_1 dt_2\ J_{rh}(t-t_1)J_{r'h'}(t'-t_2)C_{hh'}(t_1-t_2). \tag{39}$$

This result shows that the subsampled network has modified filters and an effective noise that is no longer delta-correlated.

Effective filters for subsampled all-to-all network

We now calculate the effective filter for a single unit from an all-to-all network with exponential filters. In the Fourier domain we have

$$\begin{aligned} J^{\text{eff}}(\omega) &= J(\omega) + \sum_{hh'} J_{1h}(\omega)\Delta_{hh'}(\omega)J_{h1}(\omega) \\ &= \frac{J}{i\omega\tau+1} + \sum_{hh'} \frac{J}{i\omega\tau+1}\left(\delta_{hh'} + \frac{J}{i\omega\tau+b}\right)\frac{J}{i\omega\tau+1} \\ &= \frac{J}{i\omega\tau+1} + \frac{J}{i\omega\tau+1}\left(N-1+\frac{J(N-1)^2}{i\omega\tau+b}\right)\frac{J}{i\omega\tau+1} \\ &= \frac{J}{i\omega\tau+1}\left[\frac{(i\omega\tau+1)(i\omega\tau+b)+J(N-1)(i\omega\tau+b)+J^2(N-1)^2}{(i\omega\tau+1)(i\omega\tau+b)}\right], \end{aligned}$$

which does not match the inferred filter expression.

The effective noise covariance is

$$\begin{aligned} \Sigma^{\text{eff}}(\omega) &= 1 + \sum_{hh'} J_{rh}(\omega)J_{r'h'}(-\omega)C_{hh'}(\omega) \\ &= 1 + \sum_{hh'} \frac{J}{i\omega\tau+1}\frac{J}{-i\omega\tau+1}\left(\delta_{hh'} + \frac{a}{(i\omega\tau+b)(-i\omega\tau+b)}\right) \\ &= 1 + \frac{J^2}{(i\omega\tau+1)(-i\omega\tau+1)}\left(N-1+\frac{a(N-1)^2}{(i\omega\tau+b)(-i\omega\tau+b)}\right) \end{aligned}$$

Using our result for the auto-covariance of a unit driven by correlation noise, the auto-covariance of the observed neuron should be equal to

$$\begin{aligned}
C(\omega) &= \Delta(\omega)\Delta(-\omega)\Sigma^{\mathrm{eff}}(\omega)\\
&= \left|1 - \frac{J}{i\omega\tau+1}\left[\frac{(i\omega\tau+1)(i\omega\tau+b) + J(N-1)(i\omega\tau+b) + J^2(N-1)^2}{(i\omega\tau+1)(i\omega\tau+b)}\right]\right|^{-2}\\
&\qquad\qquad \times\left(1 + \frac{J^2}{(i\omega\tau+1)(-i\omega\tau+1)}\left(N-1+\frac{a(N-1)^2}{(i\omega\tau+b)(-i\omega\tau+b)}\right)\right)\\
&= 1 + \frac{J(2-NJ)}{(i\omega\tau+1-NJ)(-i\omega\tau+1-NJ)},
\end{aligned}$$

which is the expected result (using $a = J(2-(N-1)J)$ and $b = 1-(N-1)J$, accounting for the fact that the observed neuron is removed when calculating $C_{hh'}(\omega)$). We used Mathematica to simplify the expression in going to the last line. This result demonstrates that two different models (one with delta-correlated noise and one with non-trivial correlations) can produce exactly the same covariance function.

GROUND-TRUTH NETWORKS WITH NON-EXPONENTIAL FIRING RATE NONLINEARITY $\phi(V)$

In this work we focus on the effects that unobserved neurons have on inference of their synaptic interactions, but as mentioned in the main text, model mismatch is also a possible source of deviation between inferred filters and ground truth. We can treat a simple case of model mismatch within the context of our analytic analysis: when the ground truth nonlinearity $\phi(V)$ is not exponential, but some other form such as sigmoidal,

$$\phi_i(t) = \lambda_{\mathrm{sig}}\left(1+\exp\left(-\left(\mu_i + \sum_j \int dt'\; J_{ij}(t-t')\dot{n}_j(t')\right)\right)\right)^{-1}.$$

A sigmoidal nonlinearity prevents the generative model's firing rates from diverging, and it remains stable even when the excitatory synaptic couplings between neurons are strong, though the neurons will tend to fire close to the maximum firing rate λ_{sig}. The sigmoidal nonlinearity is used for model inference less frequently than the exponential nonlinearity because it renders the log-likelihood of the model non-convex, opening the door for local minima.

If one fits a GLM with exponential nonlinearity to spike trains generated by the process with non-exponential nonlinearity, our general maximum likelihood equations are not altered, but our Gaussian approximation changes slightly because the factor of $g_i \equiv \phi'(\mu + \sum_j J_{ij} * r_j)$ that appears in the linear response function $\Delta_{ij}(t-t')$ does not reduce to the firing rate r_i. In our all-to-all network example the solution for the inferred self-history filter of a single neuron would be

$$\hat{J}(t) = \frac{1}{r}\frac{Jg(2-NJg)}{1-NJg+\sqrt{(1-NJg)^2+Jg(2-NJg)}}\frac{e^{-\sqrt{(1-NJg)^2+Jg(2-NJg)}\;t/\tau}}{\tau}\Theta(t), \tag{40}$$

where the ground-truth filter is $J(t) = Je^{-t/\tau}\Theta(t)/\tau$ and $g = \phi'(\mu + NJr)$. In the limit $N \to 1$ the inferred filter reduces to

$$\hat{J}(t) = \frac{Jg}{r}\frac{e^{-t/\tau}}{\tau}\Theta(t); \tag{41}$$

compared to the ground truth coupling strength J, the inferred coupling strength is modified by the ratio g/r. For the case of a sigmoidal nonlinearity, if J is large enough that the network is spiking close to its maximum firing rate λ_{sig}, then the gain g will be small compared to the firing rate r, and so the inferred coupling will be much weaker than the true coupling.