

ADJUSTED REGULARIZATION IN LATENT GRAPHICAL MODELS: APPLICATION TO MULTIPLE-NEURON SPIKE COUNT DATA

BY GIUSEPPE VINCI^{*,1}, VALÉRIE VENTURA^{†,§,2}
MATTHEW A. SMITH^{‡,§,3} AND ROBERT E. KASS^{†,§,2}

Rice University^{*}, *Carnegie Mellon University*[†],
University of Pittsburgh[‡] and *Center for the Neural Basis of Cognition*[§]

A major challenge in contemporary neuroscience is to analyze data from large numbers of neurons recorded simultaneously across many experimental replications (trials), where the data are counts of neural firing events, and one of the basic problems is to characterize the dependence structure among such multivariate counts. Methods of estimating high-dimensional covariation based on ℓ_1 -regularization are most appropriate when there are a small number of relatively large partial correlations, but in neural data there are often large numbers of relatively small partial correlations. Furthermore, the variation across trials is often confounded by Poisson-like variation within trials. To overcome these problems we introduce a comprehensive methodology that imbeds a Gaussian graphical model into a hierarchical structure: the counts are assumed Poisson, conditionally on latent variables that follow a Gaussian graphical model, and the graphical model parameters, in turn, are assumed to depend on physiologically-motivated covariates, which can greatly improve correct detection of interactions (nonzero partial correlations). We develop a Bayesian approach to fitting this covariate-adjusted generalized graphical model and we demonstrate its success in simulation studies. We then apply it to data from an experiment on visual attention, where we assess functional interactions between neurons recorded from two brain areas.

1. Introduction. Neurons communicate primarily through series of rapid electrical discharges known as action potentials or spikes, which take place over approximately 1 millisecond (ms); see Figure 1. Sequences of spikes (or spike trains), rather than individual spikes, encode how an animal responds to a stimulus or executes a movement. Spike trains tend to be irregular, so it is natural to model

Received December 2017; revised June 2018.

¹Supported by a NIH Blueprint grant and by the Rice Academy Postdoctoral Fellowship.

²Supported by NIH Grant R01 MH064537.

³Supported by NIH Grant R01 EY022928, NIH CORE Grant P30 EY008098, NSF NCS BCS 1734901, the Eye and Ear Foundation of Pittsburgh, and by an unrestricted grant from Research to Prevent Blindness, New York, NY.

Key words and phrases. Bayesian inference, Gaussian graphical models, Gaussian scale mixture, high dimensionality, lasso, latent variable models, macaque prefrontal cortex, macaque visual cortex, Poisson-lognormal, sparsity, spike-counts.

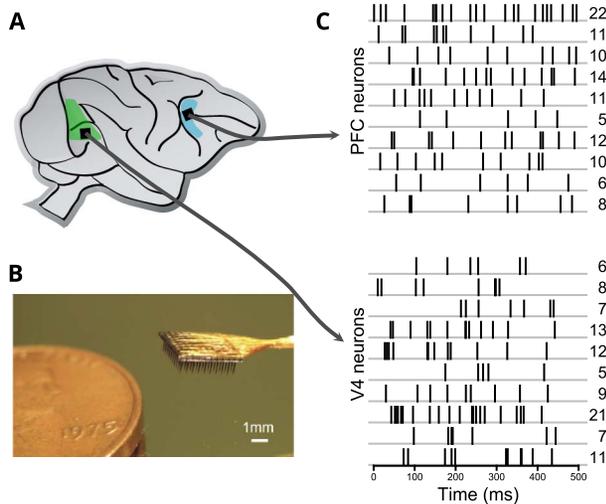


FIG. 1. (A) Sketch of primate brain with the approximate recorded regions of prefrontal cortex (PFC) and visual area V4 indicated in blue and green, respectively. (B) Utah array with 10×10 recording electrodes shown next to a penny to compare sizes. Utah arrays implanted in each of PFC and V4 of a nonhuman primate recorded the spikes of neurons. (C) In each trial of the experiment, spike trains of V4 and PFC neurons were recorded simultaneously. Each row in the diagram is the spike train for one neuron in a particular trial, with spike times being indicated as vertical tick marks. The spike counts over the 500 ms are shown on the right.

them as arising from a point process, for example a Poisson process, or to model the number of spikes in time bins as random variables that follow, for example, a Poisson distribution. In fact, modeling of spike trains was a major motivation for development of point process theory and methods [Cox and Lewis (1972), Perkel, Gerstein and Moore (1967a, 1967b)]. While a faculty member at the University of Chicago, Steve Fienberg interacted with one of the major figures in mathematical modeling of interacting neurons, Jack Cowan [Wilson and Cowan (1972)], and this led Steve to work on a foundational question: What simple principles can be used to generate point process models of neurons? After taking a crack at this, and publishing a short paper, Steve summarized the field in a very nice review article [Fienberg (1974)], which provides a great snapshot of the state of data analysis at that time. Back then the focus was almost entirely on analytically tractable methods, so the subject of Steve’s review was diffusion models, and their extensions, applied to spontaneously active neurons (that is, neurons recorded *in vitro*, having been removed from a living animal). The main results involved characterizing the stochastic process from simple assumptions and determining the resulting waiting-time distribution for boundary crossings, where, for neurons, the waiting times become times between successive spikes. Steve ended his article with a discussion of model fitting, and said, “Little in the way of statistical methodology is available to help with the problem of discriminating among intrinsically differ-

ent models providing adequate fits to a set of data.” When Steve was writing this, one of the most important new ideas in statistics, generalized linear models, had just been developed and, as computer power increased and desktop computers became standard equipment in statistical research environments, resulting techniques became practical and could be applied to solve many of the problems Steve had discussed [Brillinger (1988)]. It took a long time until such approaches began to be applicable to a variety of animal experiments, and to penetrate the literature in neurophysiology [Kass, Ventura and Brown (2005), Kass et al. (2018)].

As recording technologies have progressed, neurophysiology has continued to generate interesting statistical problems. Even though many studies rely on results involving individual neurons, much of the emphasis has shifted to multivariate effects in multiple-neuron data. In this paper, we analyze data from multiple neurons recorded during an experiment on the neurophysiology of attention. We all know that, when faced with a task, we are likely to perform better if we pay attention to what we are trying to do. But what are the mechanisms by which the brain allocates attention? This question has been posed with considerable experimental precision in studying visual processing by nonhuman primates and there is a substantial literature on the neurophysiological basis of visual attention [Maunsell (2015)]. The data we consider here come from two key regions of the brain that are involved in visual processing. The first, prefrontal cortex (PFC) is generally considered to be involved with aspects of control and, in particular, control of attention. The second, area V4, is a mid-level visual region, meaning that it receives inputs from earlier visual-processing regions and sends outputs to later regions, and there is good evidence of changes in activity among neurons in V4 when a subject pays attention to a visual stimulus. What is not known are the ways the neural networks in these areas cooperate, and the extent to which there are changes in patterns of activity, both within and across areas, when the subject pays attention. Thus, the statistical challenge was to construct a framework for identifying patterns of covariation in the activity of many neurons recorded, simultaneously, from PFC and V4. As we will next explain, the methods we have developed should be useful in various contexts, but especially for studies involving spike trains among cortical neurons, meaning neurons residing within the cerebral cortex, the part of the brain most closely associated with higher-level processing.

A small subset of the data is shown in Figure 1. As illustrated there, each experimental trial consisted of multiple spike trains recorded over 500 ms: each row in the diagram is the spike train for one neuron, with the whole set of neural spike trains being recorded simultaneously on one particular experimental trial. The total number of spikes for each neural spike train, known as that neuron’s spike count, is shown on the right. Here we follow a standard simplification by analyzing the spike counts instead of the spike trains. Although the approach we report here could be extended to point processes, we chose to tackle this simpler problem first because, as we will indicate, describing multivariate interactions among the counts was itself a substantial project. In this context, typical questions of scientific interest

would be: How strong is the covariation among neurons within PFC and V4, and how strong is the covariation across these two regions? How does this covariation change when the animal pays attention? We assess covariation in terms of partial correlation within Gaussian graphical models (GGMs). Partial correlation serves to identify the unique association of pairs of neurons, based on input signals that are not common to other recorded neurons. In a GGM, the d -dimensional data vector follows a multivariate Gaussian distribution having a precision matrix (inverse covariance matrix) Ω , and edges connect nodes wherever there are nonzero partial correlations. Because our data set, like many others, involves moderately large numbers of nodes (here, the neurons), that is, d is moderately large, estimation of Ω requires some form of regularization, and it would be natural to consider the graphical LASSO and its variants, which use ℓ_1 regularization [Chandrasekaran, Parrilo and Willsky (2012), Ravikumar et al. (2011), Yuan and Lin (2007)]. However, two major complications arise. First, the correlation across repeated trials between the spike counts of two neurons is corrupted by Poisson-like variation within the trials of each neuron, and this attenuates the estimates of across-trial correlations [Behseta et al. (2009), Vinci et al. (2016)]. Second, methods based on ℓ_1 regularization are better suited to estimate sparse networks [Rothman et al. (2008), Ravikumar et al. (2011)] but in the context of spike count data, networks are not typically very sparse. In previous work we handled the first problem for pairs of neurons using bivariate hierarchical models [Vinci et al. (2016)], where pairs of spike counts were assumed to be Poisson with bivariate log-normally distributed latent means; correlation between the latent means can then be thought of as a Poisson denoised version of spike count correlation. We also handled the second problem, while ignoring the first, by allowing the LASSO penalty to vary with the pair of neurons based on informative covariates [Vinci et al. (2018a)], which improved the detection of correct edges and the accuracy of partial correlation estimates. In this paper we combine those two previous strategies to provide a more comprehensive solution. The fundamental idea is that the observations (here, the spike counts) are considered to depend on latent variables (here, the logs of their Poisson means) that follow a Gaussian graphical model where partial correlations are zero with probabilities determined by covariates.

The key to our approach is the availability of informative covariates, which rests on well-documented neurophysiology concerning correlation among pairs of cortical neurons. In several studies, Pearson correlation has been shown to depend strongly on both inter-neuron distance (physical distance between the neurons) and tuning curve correlation (similarity of the two neurons' average activities over different stimuli): the dependence is typically weaker at larger inter-neuron distance, and stronger for larger tuning-curve correlation [Goris, Movshon and Simoncelli (2014), Smith and Sommer (2013), Yatsenko et al. (2015), Vinci et al. (2016)]. These covariates are available, but their effects may vary across brain areas, neuron types, and recording techniques. In Vinci et al. (2018a) we introduced informative covariate vectors W_{ij} composed of inter-neuron distances and tuning curve

correlations to estimate a GGM using a Bayesian ℓ_1 regularization framework (the graphical LASSO with adjusted regularization, or GAR), where partial correlations between neurons i and j can be penalized differently according to a *data driven* function of W_{ij} . Through an extensive simulation study we showed that this physiologically-motivated procedure performs substantially better (in terms of correct edge detection and accuracy of partial correlation estimates) than off-the-shelf generic tools, including not only graphical LASSO but also several of its variants that have appeared in the literature, the adaptive graphical LASSO [Fan, Feng and Wu (2009)], the latent variable graphical model [Chandrasekaran, Parrilo and Willsky (2012)], and Bayesian graphical LASSO [Wang (2012)]. Improvements in functional connectivity estimation provided by auxiliary variables had previously been observed in analyses of other kinds of neural data by either using a weighted ℓ_1 penalty on the precision matrix, with weights taken to be inverse functions of metrics of the strength of anatomical connections [MEG data, Pineda-Pardo et al. (2014); fMRI and dMRI data, Ng et al. (2012); where the strength of anatomical connections were taken to be the fiber density and fiber count between regions, respectively], or by constraining the support of the precision matrix to reflect structural connections [fMRI data, Hinne et al. (2014); the zeros were imposed in the precision matrix where structural connections were not found]. However, those studies differ from Vinci et al. (2018a) in that the dependence of the regularization on the covariates W_{ij} was pre-specified rather than learned from the data. In Vinci et al. (2018a), GAR used thresholding of estimated partial correlations to set edges to zero. Here, we instead use spike-and-slab priors where the probability of zero for the (i, j) partial correlation depends on W_{ij} , and we let the dependence on W_{ij} be data driven. We also include a Poisson observation model as in Vinci et al. (2016) to avoid attenuation of correlation.

In Section 2 we provide details about our latent covariate-adjusted GGM within a Poisson-lognormal multivariate hierarchical model, and we develop Bayesian methods for estimation in Section 3. Our approach can also treat other exponential families, and thus provides a methodology for covariate-adjusted regularization in generalized graphical models. Section 3 also provides methods to infer graphs with signed edges by controlling the Bayesian false discovery and false non-discovery rates of edge detection. In Section 4 we apply our methodology to the spike count data from areas PFC and V4, and we use inter-neuron distance as the auxiliary quantity W_{ij} . In Sections 4.1 and 4.2, and in Appendix A [Vinci et al. (2018b)], we demonstrate the benefit of our methodology compared to competing methods, and in Section 5 we add a few remarks, especially concerning opportunities for future research.

2. A Bayesian hierarchical model to include auxiliary variables. Assume that $X^{(1)}, \dots, X^{(n)} \stackrel{\text{i.i.d.}}{\sim} N(\mu, \Omega^{-1})$ are n independent and identically distributed (iid) d -dimensional Gaussian random vectors with mean vector μ and covariance

matrix Ω^{-1} . In our context, $X^{(1)}, \dots, X^{(n)}$ are the latent log means of the observed spike counts $Z^{(1)}, \dots, Z^{(n)}$ of d neurons in n identical experimental trials. Even though trials are identical, the neurons' mean spike counts, also called firing rates, can vary from trial to trial due to several sources of variability, including attention, arousal, and adaptation [Churchland et al. (2011), Goris, Movshon and Simoncelli (2014), Vinci et al. (2016)]. We can therefore think of Ω^{-1} as containing denoised versions of spike count covariances. Goris, Movshon and Simoncelli (2014) and Vinci et al. (2016) showed that Gamma or lognormal distributions provide adequate fits to latent firing rates, hence our choice of a multivariate normal distribution to model the joint vector of log firing rates.

To analyze the neural data in Section 4, we further assume that the d -dimensional vectors of spike counts $Z^{(r)} = (Z_1^{(r)}, \dots, Z_d^{(r)})'$, $r = 1, \dots, n$, are distributed according to a d -dimensional Poisson-lognormal distribution:

$$(2.1) \quad \begin{aligned} Z_i^{(r)} | X_i^{(r)} &\sim \text{Poisson}(e^{X_i^{(r)}}), \quad i = 1, \dots, d \\ X^{(r)} = (X_1^{(r)}, \dots, X_d^{(r)})' &\sim N(\mu, \Omega^{-1}), \end{aligned}$$

where the dependence structure of the observed spike counts $Z^{(r)}$ is governed by the dependence structure of their latent log firing rates $X^{(r)}$. Vinci et al. (2016) showed that the hierarchical model in equation (2.1) provides a good fit to the spike counts of pairs of neurons ($d = 2$).

In Bayesian GGMs, a prior h is assumed for the precision matrix Ω , and it is often taken to be a product of Laplace distributions over the entries ω_{ij} of Ω [Wang (2012)], that is

$$(2.2) \quad h(\Omega | (\lambda_{ij})_{i \leq j}) \propto \prod_{i \leq j} l_{ij}(\omega_{ij}) I(\Omega \succ 0),$$

where $l_{ij}(\omega) \propto e^{-\lambda_{ij}|\omega_{ij}|}$, $I(\Omega \succ 0)$ constrains the support of the distribution h on the positive definite cone of $d \times d$ symmetric matrices, thereby ensuring $\omega_{ii} > 0$ for all $i = 1, \dots, d$, and the parameter λ_{ij} determines the amount of shrinkage towards zero of ω_{ij} . In this framework, the maximum a posteriori estimator of Ω corresponds to the graphical LASSO [Yuan and Lin (2007)]. Here we use a more general spike and slab type prior [George and McCulloch (1997)], replacing the Laplace distribution l_{ij} in equation (2.2) by a two-component mixture composed of a point-mass at zero with mixing proportion $1 - \pi_{ij}$, and a Gaussian scale mixture distribution p_{ij} centered at m_{ij} . The point-mass component allows exact zero entries in Ω (i.e., Ω can be sparse), which might induce better concentration around zero of the smaller partial correlation estimates when the graph is indeed locally sparse. Banerjee and Ghosal (2015) and Wang (2015) had previously used a spike and slab prior to estimate graphs, but they assumed the same parameters across all entries of Ω and took p_{ij} to be a Laplace distribution centered at zero. However, allowing for nonzero mean parameters m_{ij} , as we do here, can be useful to better

regularize locally dense precision matrices Ω , where many entries are nonzero, as might happen with neural data. The analytic expression for our spike and slab prior is

$$(2.3) \quad h(\Omega \mid (\lambda_{ij}, m_{ij}, \pi_{ij})_{i \leq j}) \propto \prod_{i \leq j} \{\pi_{ij} p_{ij}(\omega_{ij}) + (1 - \pi_{ij})\delta(\omega_{ij})\} I(\Omega > 0),$$

where $\delta(0) = \infty$ and $\delta(x) = 0$ for all $x \neq 0$ is the Dirac delta function, $\pi_{ij} \in [0, 1]$, $\pi_{ii} = 1$ for all i ,

$$(2.4) \quad p_{ij}(\omega) = \int \frac{\lambda_{ij}}{v} \phi\left(\frac{\lambda_{ij}}{v}(\omega - m_{ij})\right) dG_{ij}(v),$$

$\phi(z)$ is the standard Gaussian density function, and $G_{ij} : \mathbb{R} \rightarrow [0, 1]$ is a cumulative distribution function with $G_{ij}(0) = 0$ and $\int_0^\infty v^{-1} dG_{ij}(v) < \infty$ (these conditions are necessary and sufficient for the existence of p_{ij} [Andrews and Mallows (1974)]; see Lemma 1 in Vinci et al. (2018b)). Without loss of generality, we assume $\int_0^\infty v^2 dG_{ij}(v) = 1$, which implies $\int (\omega - m_{ij})^2 p_{ij}(\omega) d\omega = 1/\lambda_{ij}^2$ and thus allows λ_{ij} to be interpreted as the rate parameter of p_{ij} [see Lemma 1 in Vinci et al. (2018b)]. Equation (2.3) includes common modeling choices as special cases, where typically the prior parameters are the same for all (i, j) , and $m_{ij} = 0$: for example, Gaussian/ridge [Hoerl and Kennard (1981)], Laplace/LASSO [Friedman, Hastie and Tibshirani (2008), Rothman et al. (2008), Tibshirani (1996), Wang (2012), Yuan and Lin (2007), Mazumder and Hastie (2012)], spike-and-slab with Laplace density [Banerjee and Ghosal (2015), Wang (2015)], Elastic net [Zou and Hastie (2005)], the latent variable graphical model [Chandrasekaran, Parrilo and Willsky (2012), Yuan (2012), Giraud and Tsybakov (2012)], as well as less explored choices; see Appendix A [Vinci et al. (2018b)].

The prior distribution for Ω in equation (2.3) has a large number of parameters, but when auxiliary variables are available, it makes sense to let them all depend on those variables and thus reduce greatly the number of effective parameters, as we now describe. Assume that q -dimensional auxiliary variables $W = \{W_{ij}\}_{i < j}$ that carry information about the network of neurons' firing rates are available. We incorporate that information in the prior distribution for Ω to model the neurophysiological mechanism involving W that gives rise to the neuronal correlation structure in Ω . For example, if W is inter-neuron distance, then h might produce covariance structures where two neurons are less dependent if they are farther apart. An example of a 4-dimensional dependence graph is given in Figure 2. We thus let the location, rate, and weight parameters in equations (2.3) and (2.4) depend on W according to

$$(2.5) \quad m_{ij} = (\alpha_i \alpha_j)^{-1} f(W_{ij}),$$

$$(2.6) \quad \lambda_{ij} = \alpha_i \alpha_j g(W_{ij}),$$

$$(2.7) \quad \pi_{ij} = \eta(W_{ij}),$$

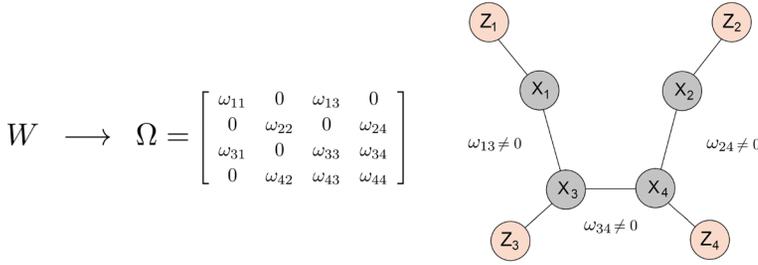


FIG. 2. Example of a d -dimensional dependence graph specified by the hierarchical model in equation (2.1), where $d = 4$ and W is inter-neuron distance, represented here as physical distance. Neuron pairs (1, 2), (1, 4), and (2, 3) are farther apart than the other pairs, and we assume they are not functionally connected. The precision matrix Ω arises from a generative process that reflects the dependence between W and functional connectivity, and here, accordingly, the entries for the three neuron pairs that are far apart are zero. The Gaussian random vector $X \sim N(\mu, \Omega^{-1})$ has dependence structure specified by Ω . The missing edges correspond to zero partial correlations. (Z_1, Z_2, Z_3, Z_4) are the observed spike counts and (X_1, X_2, X_3, X_4) their latent log means (firing rates).

for $i \neq j$, where $f : \mathbb{R}^q \rightarrow \mathbb{R}$, $g : \mathbb{R}^q \rightarrow \mathbb{R}_+$, and $\eta : \mathbb{R}^q \rightarrow [0, 1]$ are flexible functions to be estimated from the data (see Figure 5), and $\alpha = (\alpha_1, \dots, \alpha_d)$ are parameters that ensure that f and g are on the scale of the partial correlations, $\rho_{ij} = -\omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}}$, rather than on the scale of ω_{ij} ; more details are in Appendix C.1.5 [Vinci et al. (2018b)]. Letting π_{ij} , m_{ij} , and λ_{ij} depend on W_{ij} models the information carried by W about Ω , which, when combined with the observed data $(Z^{(1)}, \dots, Z^{(n)})$, can provide better inferences for Ω if W indeed contains relevant information. If no auxiliary variables are available, the model can still be used by letting f , g , and η be constants.

This framework generalizes the graphical LASSO with adjusted regularization (GAR) of Vinci et al. (2018a). GAR assumes that the observation model is the second line in equation (2.1), with no Poisson hierarchy, and that the prior distribution for Ω is a product of Laplace distributions [equation (2.2)] with rate parameters λ_{ij} nonparametrically modulated by W_{ij} , as specified in equation (2.6). Vinci et al. (2018a) applied GAR to square-rooted spike count data from macaque areas V1 and V4 with a two-dimensional W_{ij} composed of inter-neuron distances and tuning curve correlations; the square-root transformation was used to improve the fit to a Gaussian distribution [Georgopoulos and Ashe (2000), Kass, Eden and Brown (2014), Yu et al. (2009)]. The estimated rate function g in equation (2.6) was nonlinearly increasing with inter-neuron distance and decreasing with tuning curve correlation.

Next, we present Bayesian algorithms to estimate the model described above, and we fit it to multiple neuron spike data in Section 4. Estimating the Poisson-lognormal model [equation (2.1)] in high dimensions can be heavily corrupted by noise [Inouye et al. (2017)], and we expect incorporation of auxiliary quantities W to improve estimation.

3. Estimating the precision matrix and the graph. In this section we describe the Bayesian inference of Ω in the model described in Section 2. We start with full and empirical Bayes algorithms for the case when the Gaussian component X is observed (Section 3.1), because they can be useful for Gaussian data, in neuroscience or other fields. We can also extend these algorithms easily when X is latent in an exponential family observation model; Section 3.2 treats the particular Poisson lognormal case in equation (2.1). The full Bayes algorithm uses the Gibbs sampler to approximate the joint distribution of θ and Ω given Gaussian vectors $\mathbb{X}_n = \{X^{(1)}, \dots, X^{(n)}\}$, where $\theta = \{f, g, \eta, \alpha, m, \lambda\}$, and $m = (m_{11}, \dots, m_{dd})$ and $\lambda = (\lambda_{11}, \dots, \lambda_{dd})$ are the means and rates of the p.d.f. in equation (2.4) for $i = j$; f, g, η , and α are defined in equations (2.5), (2.6), and (2.7). We assume that f, g , and η are step-functions; more flexible nonparametric functions are too hard to implement at present. The empirical Bayes algorithm computes first a point estimate $\hat{\theta}$ as the maximum of its marginal posterior, using an EM algorithm that allows f, g , and η to be of any parametric or nonparametric forms, and then estimate Ω through the posterior distribution $h(\Omega | \mathbb{X}_n, W, \hat{\theta})$.

We assign a prior density to θ of the form

$$(3.1) \quad h(\theta | W) \propto h^*(\theta | W) A(\theta, W)^{-1},$$

where $A(\theta, W)$ is the intractable but bounded normalizing constant of equation (2.3) [see Lemma 2 in Vinci et al. (2018b)], and $h^*(\theta | W)$ is a density on θ . Note that conditioning on W in equation (3.1) means that the prior on θ depends on the auxiliary quantity W , but it is θ and W together that specify the parameters in equation (2.3). We combine equation (3.1) with the generative model $h(\Omega | \theta, W)$ [equations (2.3) to (2.7)], a prior distribution $h(\mu)$ on μ [e.g., $h(\mu) \propto 1$; see Appendix C.1.3 in Vinci et al. (2018b)], and the Gaussian likelihood function $L(\mu, \Omega; \mathbb{X}_n)$, to obtain the full joint posterior distribution

$$(3.2) \quad h(\mu, \Omega, \theta | \mathbb{X}_n, W) \propto h(\mu)h(\Omega | \theta, W)h(\theta | W)L(\mu, \Omega; \mathbb{X}_n),$$

which yields posterior means or maximum a posteriori estimates for θ, μ , and Ω .

3.1. *Posterior computation and estimation for observed Gaussian data.* The posterior distribution in equation (3.2) is not analytically tractable because of the positive definiteness constraint on Ω and the point-mass components in equation (2.3) (when $\pi_{ij} < 1$). Instead, we use posterior simulations from the augmented quantity $(\mu, \Omega, \theta, Y, V)$, with posterior distribution

$$(3.3) \quad h_a(\mu, \Omega, \theta, Y, V | \mathbb{X}_n, W) \propto h(\mu)h_a(\Omega, \theta, Y, V | W)L(\mu, \Omega; \mathbb{X}_n),$$

where $V = [V_{ij}]$ and $Y = [Y_{ij}]$ are nuisance parameters such that

$$(3.4) \quad h(\mu, \Omega, \theta | \mathbb{X}_n, W) = \lim_{a \rightarrow 0^+} \int \sum_{Y \in \mathcal{Y}} h_a(\mu, \Omega, \theta, Y, V | \mathbb{X}_n, W) dV$$

and \mathcal{Y} is the set of all 0–1 symmetric matrices with positive diagonal; in practice we take a small a , for example, $a^2 = 10^{-8}$. Basically, we replace the point mass component in equation (2.3) by a Gaussian distribution with a very small variance a^2 , so that it is highly peaked around zero.

The second term on the right-hand side of equation (3.3) is given by

$$(3.5) \quad \begin{aligned} &h_a(\Omega, \theta, Y, V \mid W) \\ &\propto h_a(\Omega \mid \theta, W, Y, V)h_a(V \mid \theta, W, Y)h(Y \mid \theta, W)h_a(\theta \mid W), \end{aligned}$$

where

$$(3.6) \quad \begin{aligned} &h_a(\Omega \mid \theta, W, Y, V) \\ &= \prod_{i \leq j} \tau_{ij} \phi(\tau_{ij}(\omega_{ij} - Y_{ij}m_{ij}))I(\Omega \succ 0)B(\theta, W, Y, V, a), \end{aligned}$$

with $\tau_{ij} = Y_{ij}\lambda_{ij}/V_{ij} + (1 - Y_{ij})/a$ and $Y_{ii} \equiv 1, \forall i$,

$$(3.7) \quad h_a(V \mid \theta, W, Y) = B(\theta, W, Y, V, a)^{-1} \prod_{i \leq j} g_{ij}(V_{ij})A(\theta, W, Y, a),$$

$$(3.8) \quad h(Y \mid \theta, W) = \prod_{i < j} \pi_{ij}^{Y_{ij}} (1 - \pi_{ij})^{1 - Y_{ij}},$$

$$(3.9) \quad h_a(\theta \mid W) \propto h^*(\theta \mid W)A(\theta, W, Y, a)^{-1},$$

$B(\theta, W, Y, V, a)$ is the normalizing constant of $h_a(\Omega \mid \theta, W, Y, V)$, $A(\theta, W, Y, a)$ is the normalizing constant of $h_a(V \mid \theta, W, Y)$, h^* is as in equation (3.1), and g_{ij} is the p.d.f. of G_{ij} in equation (2.4). In Lemma 2 [Vinci et al. (2018b)] we show that the normalizing constants are finite and away from zero, that is $1 < A(\theta, W, Y, a) < \infty$ and $1 < B(\theta, W, Y, V, a) < \infty$, so that equations (3.6), (3.7), and (3.9) are well defined distributions. Lemma 3 [Vinci et al. (2018b)] shows that the quantity (μ, Ω, θ) of a sample $(\mu, \Omega, \theta, Y, V)$ drawn from $h_a(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W)$ in equation (3.3) gets closer to a genuine sample from $h(\mu, \Omega, \theta \mid \mathbb{X}_n, W)$ in equation (3.2) as $a \rightarrow 0^+$.

The representation prescribed by equations (3.3), (3.6), and (3.7) is computationally convenient because it circumvents approximating intractable normalizing constants [mainly due to the positive definiteness constraint $\Omega \succ 0$ in equation (2.3)] and provides a framework where sampling ($\Omega \mid \text{rest}$) does not depend on the mixing distributions G_{ij} 's. Using a specific Gaussian mixture in equation (2.4) only requires a different sampling of $(V_{ij} \mid \text{rest})$ [equation (C.1), Appendix C.1 in Vinci et al. (2018b)], which permits many different regularizations of the precision matrix Ω ; more details and examples are in Appendix A [Vinci et al. (2018b)].

Full Bayes estimation. In Appendix C.1 [Vinci et al. (2018b)] we describe a Gibbs sampler algorithm to draw samples from the posterior distribution of

$(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W)$ in equation (3.3), assuming that the model components f , g , and η in equations (2.5), (2.6), and (2.7) are the step functions

$$\begin{aligned}
 f(w) &= \sum_{k=1}^{K_f} \gamma_k I_{A_k}(w), \\
 g(w) &= \sum_{k=1}^{K_g} \beta_k I_{B_k}(w), \\
 \eta(w) &= \sum_{k=1}^{K_\eta} \chi_k I_{C_k}(w),
 \end{aligned}
 \tag{3.10}$$

where $\{A_1, \dots, A_{K_f}\}$, $\{B_1, \dots, B_{K_g}\}$, and $\{C_1, \dots, C_{K_\eta}\}$ are partitions of the domain of the auxiliary variables W , and $I_D(w) = I(w \in D)$ is the indicator function. Hence, sampling f , g , and η reduces to sampling their coefficients $\gamma = (\gamma_1, \dots, \gamma_{K_f})' \in \mathbb{R}^{K_f}$, $\beta = (\beta_1, \dots, \beta_{K_g})' \in \mathbb{R}_+^{K_g}$, and $\chi = (\chi_1, \dots, \chi_{K_\eta})' \in [0, 1]^{K_\eta}$. The numbers of steps K_g , K_f , and K_η and their locations can be chosen so that the sample partial correlations plotted against W are approximately constant within each step. We use step functions because they are flexible, if not smooth, and they allow the Gibbs sampler to run. Other functions are too difficult to implement at present. In the future we may be able to implement continuous nonparametric functions (e.g. a kernel smoother with a few degrees of freedom) as in the empirical Bayes approach.

We also assume that m_{ii} and V_{ii} in equation (2.4) are such that $m_{ii} \leq 0$ and $V_{ii}^2 \sim \Gamma(1, 1/2)$ for all $i = 1, \dots, d$. The Bayesian counterparts of several penalized MLE methods, such as the graphical LASSO [Yuan and Lin (2007)] and the sparse–low rank model [Chandrasekaran, Parrilo and Willsky (2012)], require the same conditions.

Empirical Bayes estimation. Instead of jointly sampling θ and Ω given the data, we estimate $\theta = \{f, g, \eta, \alpha, m, \lambda\}$ as the maximizer of the marginal posterior distribution

$$\begin{aligned}
 \hat{\theta} &= \arg \max_{\theta} h(\theta \mid \mathbb{X}_n, W) \\
 &= \arg \max_{\theta} \int \int \int \sum_{Y \in \mathcal{Y}} h_a(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W) dV d\mu d\Omega,
 \end{aligned}
 \tag{3.11}$$

and the precision matrix Ω is then inferred from the posterior distribution $h(\Omega \mid \mathbb{X}_n, W, \hat{\theta})$. We solve equation (3.11) using the Expectation–Maximization algorithm described in Appendix C.2 [Vinci et al. (2018b)]. This simpler setting allows f , g , and η to be of any functional form, either parametric or nonparametric.

3.2. *Posterior computation and estimation for latent Gaussian data.* We extend our graph estimation approach to the inference of Ω under the hierarchical model in equation (2.1) by supplementing the previous algorithms with a mechanism to sample X given Z and the rest of the parameters and variables. Full and empirical Bayes estimation algorithms are given in Appendix C.3 [Vinci et al. (2018b)] to obtain estimates of θ and Ω from the posterior distribution

$$(3.12) \quad \begin{aligned} &h_a(\mu, \Omega, \theta, Y, V, \mathbb{X}_n \mid \mathbb{Z}_n, W) \\ &\propto h_a(\Omega, \theta, Y, V \mid W)L(\mu, \Omega; \mathbb{X}_n)p(\mathbb{Z}_n \mid \mathbb{X}_n), \end{aligned}$$

where the first term on the right-hand side is defined in equation (3.5) and

$$(3.13) \quad p(\mathbb{Z}_n \mid \mathbb{X}_n) = \prod_{r=1}^n p(Z^{(r)} \mid X^{(r)})$$

is the likelihood of the observed independent random variables $Z^{(1)}, \dots, Z^{(n)}$ conditioned on $X^{(1)}, \dots, X^{(n)}$. Appendix C.3 [Vinci et al. (2018b)] provides algorithms to fit the Poisson-lognormal model [equation (2.1)], where $X \mid \text{rest}$ is sampled using rejection sampling. In Section 4 we apply this model to spike count data from macaque visual cortex using inter-neuron distance as an auxiliary variable W . Different observational distributions can be fitted in place of Poisson by just modifying the sampling step of $X \mid \text{rest}$. For instance, if $p(z \mid x)$ is binomial or negative-binomial, then the sampling of $X \mid \text{rest}$ may be performed by using rejection sampling or the data augmentation strategy based on the Pólya-Gamma distribution [Polson, Scott and Windle (2013)].

3.3. *Graph estimation.* Once the posterior distribution of $(\Omega \mid \mathbb{X}_n, W)$ is computed, we use it to estimate the signed graph of the network, $E(\delta) = [E_{ij}(\delta)]$, where the edge between nodes i and j is defined as

$$(3.14) \quad E_{ij}(\delta) = \begin{cases} 1, & \rho_{ij} > \delta, \\ -1, & \rho_{ij} < -\delta, \\ 0, & |\rho_{ij}| \leq \delta \end{cases}$$

for some threshold $\delta \in [0, 1)$. Several estimates are listed below.

DEFINITION 1. A plug-in signed δ -graph is the signed matrix $\hat{E}(\delta) = [\hat{E}_{ij}(\delta)]$ where

$$(3.15) \quad \hat{E}_{ij}(\delta) = \begin{cases} 1, & \hat{\rho}_{ij} > \delta, \\ -1, & \hat{\rho}_{ij} < -\delta, \\ 0, & |\hat{\rho}_{ij}| \leq \delta, \end{cases}$$

and $\hat{\rho}_{ij}$ is an estimate of ρ_{ij} .

DEFINITION 2. A posterior signed δ -graph performs a Bayesian test of the hypotheses $H_0 : |\rho_{ij}| \leq \delta$, $H_1 : \rho_{ij} > \delta$, and $H_2 : \rho_{ij} < -\delta$ and can be obtained as the signed matrix $\hat{E}(\delta) = [\hat{E}_{ij}(\delta)]$ where

$$(3.16) \quad \hat{E}_{ij}(\delta) = \arg \max_{s \in \{-1, 0, 1\}} \mathbb{P}_{ij}^\delta(s),$$

and $\mathbb{P}_{ij}^\delta(s) = \mathbb{P}(E_{ij}(\delta) = s \mid \mathbb{X}_n, W)$ is the posterior probability that $E_{ij}(\delta) = s$.

DEFINITION 3. A posterior signed (p, δ) -graph is the signed matrix $\hat{E}(p, \delta) = [\hat{E}_{ij}(p, \delta)]$ where

$$(3.17) \quad \hat{E}_{ij}(p, \delta) = \begin{cases} 1, & \mathbb{P}_{ij}^\delta(0) < 1 - p \quad \text{and} \quad \mathbb{P}_{ij}^\delta(1) \geq \mathbb{P}_{ij}^\delta(-1), \\ -1, & \mathbb{P}_{ij}^\delta(0) < 1 - p \quad \text{and} \quad \mathbb{P}_{ij}^\delta(1) < \mathbb{P}_{ij}^\delta(-1), \\ 0, & \mathbb{P}_{ij}^\delta(0) \geq 1 - p. \end{cases}$$

3.4. *Edge error detection control.* We choose the parameters δ and p of the graph estimators by controlling edge discovery errors. Let \hat{E} be an estimator of $E(\delta)$. Define the false discovery proportion

$$(3.18) \quad \text{FDP} = \frac{\sum_{i < j} |\hat{E}_{ij}| (1 - |E_{ij}(\delta)|)}{\sum_{i < j} |\hat{E}_{ij}|},$$

and the false non-discovery proportion

$$(3.19) \quad \text{FNP} = \frac{\sum_{i < j} (1 - |\hat{E}_{ij}|) |E_{ij}(\delta)|}{\sum_{i < j} (1 - |\hat{E}_{ij}|)},$$

which are, respectively, the proportion of incorrectly detected edges out of all detections, and the proportion of missed true edges out of all non-detections. Further define the signed false discovery proportion

$$(3.20) \quad \text{FDP}^* = \frac{\sum_{i < j} \{I(\hat{E}_{ij} > 0)I(E_{ij}(\delta) \leq 0) + I(\hat{E}_{ij} < 0)I(E_{ij}(\delta) \geq 0)\}}{\sum_{i < j} |\hat{E}_{ij}|},$$

which takes into account sign errors of the detected edges. Lemma 5 [Vinci et al. (2018b)] shows that $\text{FDP} \leq \text{FDP}^*$, so that controlling FDP^* is a more stringent criterion because it also takes into account the sign of the edges. Taking the expectations of FDP , FNP , and FDP^* with respect to the Gaussian data \mathbb{X}_n conditionally on the true parameters μ and Ω gives the frequentist false discovery rate

$$(3.21) \quad \text{FDR} = \mathbb{E}[\text{FDP} \mid \mu, \Omega],$$

false non-discovery rate

$$(3.22) \quad \text{FNR} = \mathbb{E}[\text{FNP} \mid \mu, \Omega],$$

and signed false discovery rate

$$(3.23) \quad \text{FDR}^* = \mathbb{E}[\text{FDP}^* \mid \mu, \Omega].$$

Bounding either of these edge error metrics is a way to choose the parameters δ and p of the graph estimators given in Definitions 1, 2, and 3. These metrics are easy to compute in simulation studies, where the true graph is known, but not otherwise. For experimental data, we use instead their Bayesian counterparts, obtained by taking the expectations of FDP, FNP, and FDP* conditionally on the Gaussian data \mathbb{X}_n and the auxiliary quantity W , yielding

$$(3.24) \quad \text{FDR}_\Pi = \frac{\sum_{i < j} |\hat{E}_{ij}| P(E_{ij}(\delta) = 0 \mid \mathbb{X}_n, W)}{\sum_{i < j} |\hat{E}_{ij}|},$$

$$(3.25) \quad \text{FNR}_\Pi = \frac{\sum_{i < j} (1 - |\hat{E}_{ij}|) P(E_{ij}(\delta) \neq 0 \mid \mathbb{X}_n, W)}{\sum_{i < j} (1 - |\hat{E}_{ij}|)},$$

and

$$(3.26) \quad \text{FDR}^*_\Pi = \frac{\sum_{i < j} \{I(\hat{E}_{ij} > 0)P(E_{ij}(\delta) \leq 0 \mid \mathbb{X}_n, W) + I(\hat{E}_{ij} < 0)P(E_{ij}(\delta) \geq 0 \mid \mathbb{X}_n, W)\}}{\sum_{i < j} |\hat{E}_{ij}|}.$$

The posterior probability $P(E_{ij}(\delta) = 0 \mid \mathbb{X}_n, W)$ in equation (3.24) can be interpreted as a local false discovery rate, so that the FDR_Π can be viewed as the global false discovery rate of the set of putative signals (here edges) in a multiple hypothesis testing problem [Efron et al. (2001), Scott et al. (2015)]. In simulations, FDR_Π , FNR_Π , and FDR^*_Π appear to control their frequentist counterparts [Figure 3; see also Vinci et al. (2018a)].

3.5. *ROC analysis.* We use Bayesian edge discovery error rates to build the graph estimates, as described above, and we evaluate their quality using specificity and sensitivity:

$$(3.27) \quad \text{SPEC} = \frac{\sum_{i < j} I(\hat{E}_{ij} = 0)I(E_{ij}(\delta) = 0)}{\sum_{i < j} I(E_{ij}(\delta) = 0)},$$

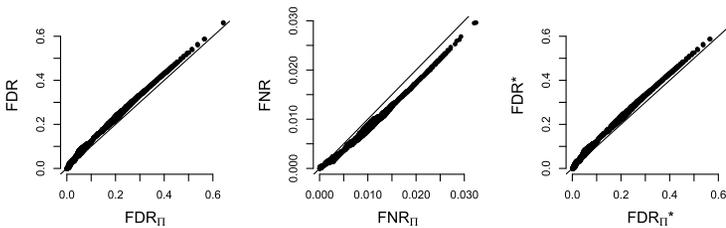


FIG. 3. Bayesian FDR_Π , FNR_Π , and FDR^*_Π plotted against their frequentist counterparts. Results obtained from simulated d -dimensional Gaussian vectors, $d = 150$, with 5580 edges and sample size $n = 5000$. The error rates were computed from (p, δ) -graph estimates (Definition 3).

$$(3.28) \quad \text{SENS} = \frac{\sum_{i < j} I(\hat{E}_{ij} \neq 0) I(E_{ij}(\delta) \neq 0)}{\sum_{i < j} I(E_{ij}(\delta) \neq 0)},$$

which give the proportions of true edges correctly identified and of missing edges correctly omitted, respectively. To account for sign detection performance we further define the positive and negative sensitivities:

$$(3.29) \quad \text{SENS}^+ = \frac{\sum_{i < j} I(\hat{E}_{ij} > 0) I(E_{ij}(\delta) > 0)}{\sum_{i < j} I(E_{ij}(\delta) > 0)},$$

$$(3.30) \quad \text{SENS}^- = \frac{\sum_{i < j} I(\hat{E}_{ij} < 0) I(E_{ij}(\delta) < 0)}{\sum_{i < j} I(E_{ij}(\delta) < 0)},$$

which give the proportions of true positive and true negative edges correctly identified, respectively. The Receiver Operating Characteristic (ROC) curve displays $(1 - \text{SPEC})$ versus SENS for all possible settings (e.g., values of tuning parameters) of a graph estimator. An Area Under the ROC Curve (AUC) close to one indicates good edge estimation performance, but does not account for the signs of the connections. A three-class extension of the ROC curve that accounts for connection sign is the ROC surface given by the graph of SPEC as a bivariate function of SENS^+ and SENS^- over the set $[0, 1]^2$. The Volume Under the ROC Surface (VUS) is a three-class extension of the AUC that takes into account sign detection performance. A graph with signed edges assigned at random has VUS $1/6$ and AUC $1/2$. We thus rescale VUS to ease its interpretation, according to $1/2 + (\text{VUS} - 1/6)6/10$.

4. Estimation of functional connectivity in macaque areas V4 and PFC.

Neural spike train data were recorded simultaneously from visual area V4 and PFC in the same hemisphere of a macaque monkey (*Macaca mulatta*) using two 100-electrode Utah arrays (Figure 1), while the animal was performing an orientation change detection task⁴ (Figure 4). The visual stimuli consisted of drifting gratings, that is, alternating light and dark bars that move in a fixed direction within a circular aperture displayed on a screen in front of the animal. Neurons in area V4 fire more rapidly when the drifting gratings are placed in a particular region of visual space, known as that neuron's receptive field (RF). In the task, one stimulus was placed in the location of the aggregate RFs of the V4 neurons, which was in the lower left corner of the visual space, indicated by the dashed circles in Figure 4, and the other stimulus in the opposite hemifield, in the lower right corner. The animal's task (successful completion of which resulted in a juice reward) was to

⁴All procedures were approved by the Institutional Animal Care and Use Committee of the University of Pittsburgh, and were in compliance with the guidelines set forth in the National Institutes of Health's Guide for the Care and Use of Laboratory Animals.

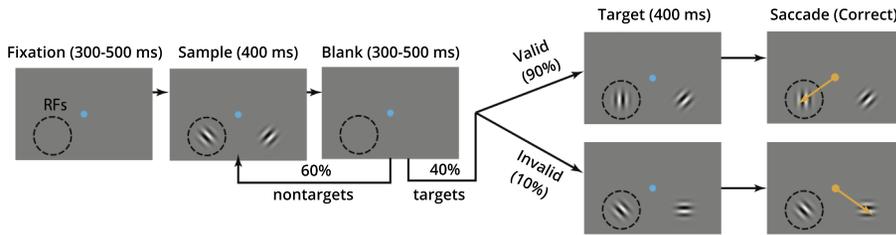


FIG. 4. The experiment. A trial proceeded as follows: after an initial period of 300 to 500 ms of fixation on the blue central spot in the center of the visual space, a sample stimulus was presented, which consisted of two drifting gratings with orthogonal orientations appearing in the aggregate receptive field (RF) of the V4 neurons being recorded, indicated by the dashed circle in the visual space, and its opposite hemifield. The stimulus lasted for 400 ms and was followed by a blank period of 300 to 500 ms. After this blank period, there was a 40% chance that a target (a changed orientation in either the RF (“valid side”) or its opposite hemifield (“invalid side”) appeared, and a 60% chance that the sample stimulus repeated, called a “nontarget”. The animal’s task was to make an eye movement (a saccade, marked in orange) from the central blue fixation spot to the location where the grating changed orientation. The animal was cued in blocks of trials to attend to (pay attention to) either the aggregate V4 RF (“attend in”) or the location in the opposite hemifield (“attend out”), while maintaining fixation on the central blue spot. We analyze the sample period, when the animal presumably prepared to detect a change in grating orientation. The animal’s attention was validated by behavioral metrics, including a decrease in response time and increase in hit rate when the target appeared at the cued location as compared to the uncued location.

make an eye movement (marked in orange) from the central blue fixation spot to the location where the grating changed orientation. The drifting gratings changed in a subtle way (a shift in orientation) that signaled the animal to respond with an eye movement, and the animal was cued in blocks of trials to attend to one of the two stimuli, either in the receptive fields of the V4 neurons being recorded (“attend in”) or the location in the opposite hemifield (“attend out”). The four experimental conditions were therefore defined by the combination of two variables:

1. ORIENTATION ϑ : a drifting grating at orientation degree $\vartheta = 135$ or 45 was presented in the RF, and in the orthogonal orientation in the opposite hemifield.
2. ATTENTION ζ : the animal was cued to either attend towards ($\zeta = 1$, attend in) or away from ($\zeta = 0$, attend out) the receptive field.

This experiment was novel in that it directly investigated functional interactions among large populations of neurons in two brain areas thought to play critical roles in visual attention. A key feature of this experiment (like many others in neurophysiology) involves the substantial number of repeated observations, or trials, under identical experimental conditions. The trial-to-trial variation can provide a means of assessing cooperative activity among neurons. The questions we seek to answer are: How strongly connected, in the sense of covariation of firing rates, are the neurons within V4 and within PFC? How strongly connected are the neurons in PFC to those in V4? How does this connectivity change with attention? How does connectivity change with the orientation of the grating?

A total of 666, 759, 954, and 1077 identical repeat trials of length 500 ms, starting at the stimulus onset of epoch “Sample (400 ms)” in Figure 4, were recorded under the four task combinations $(\vartheta, \zeta) \in \{(135, 1), (45, 1), (135, 0), (45, 0)\}$, respectively, and 264 candidate neuronal units were identified across the two multi-electrode arrays in one recording session. For our statistical analysis we retained only the 148 neurons (71 in V4 and 77 in PFC) that had average firing rates greater than 4 spikes per second in all four conditions. Their firing rates had maximum 67.46 spikes/s, mean 18.96, and 2.5th and 97.5th percentiles 4.78 and 56.80.

To estimate the functional connectivity between and within areas V4 and PFC we fitted the 148-variate Poisson-lognormal (PLN) model [equation (2.1)] in each of the four conditions using the full Bayes algorithm [Algorithm 1 in Appendix C3 in Vinci et al. (2018b)], which provides estimates of the conditional dependence graphs of the latent log firing rates X , with edges representing functional connections. Specifically, we let the prior distribution for Ω be equation (2.3), with the Gaussian mixture distribution p_{ij} [equation (2.4)] taken to be normal with mean m_{ij} and rate λ_{ij} modulated by functions $f(W_{ij})$ and $g(W_{ij})$ [equations (2.5) and (2.6)], respectively, and weight π_{ij} modulated by $\eta(W_{ij})$ [equation (2.7)]. Here, the auxiliary quantity W_{ij} is inter-neuron distance, measured by the distance between electrodes. Neuronal dependences have been observed to weaken when neurons are physically further apart in several cortical areas [Goris, Movshon and Simoncelli (2014), Smith and Kohn (2008), Smith and Sommer (2013), Vinci et al. (2016, 2018a), Yatsenko et al. (2015)], so we expect this quantity to be useful to improve graph estimation. We let the regularizing functions f , g , and η be step functions⁵ with five steps each, the last step modeling distances W greater than 5mm, which corresponds to pairs of neurons on different arrays; the first four steps are taken at the quartiles of the interneuron distances that are less than 5 mm, corresponding to pairs of neurons on the same array. We did not consider including a low-rank covariance structure because the optimal number of factors in a factor model fitted to the square-rooted spike counts of all 148 neurons (using ten-fold cross validation; see Section 4.1 for details) was about 22 factors in each condition, suggesting that the neurons’ covariation structure was not low dimensional. Here, the animal was awake, while the very low-dimensional structure found by Ecker et al. (2014) was in data recorded under anesthesia.

Figure 5 shows the fitted functions f , g , and η in condition $(\vartheta, \zeta) = (45, 0)$; fitted models in the other conditions were similar: $f(W)$ is close to zero, which suggests that the partial correlations are centered around zero for all values of W , and $\eta(W)$ is close to one, likely because the shrinkage induced by $g(W)$ is strong enough and/or there isn’t a large proportion of exactly zero Ω entries. However,

⁵The full Bayes algorithm is currently implemented with f , g , and η in equations (2.5), (2.6), and (2.7) being step functions. The empirical Bayes estimator is more attractive because f , g , and η can be any parametric or nonparametric functions, but the fitting algorithm is very computationally expensive.

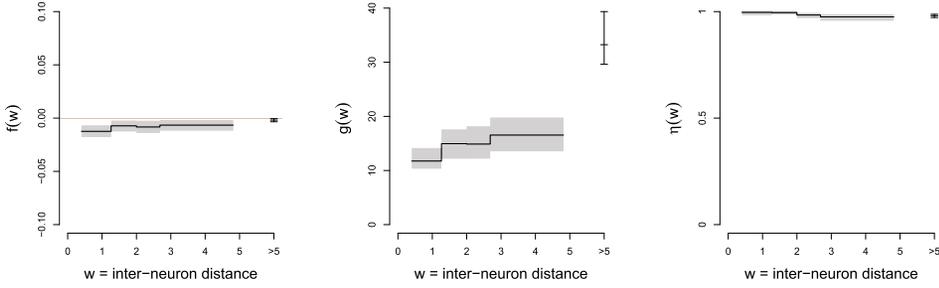


FIG. 5. Step functions $f(W)$, $g(W)$, and $\eta(W)$ fitted to the data in condition $(\vartheta, \zeta) = (45, 0)$: $f \approx 0$ suggests that the partial correlations are centered around zero for all values of W , and $\eta \approx 1$, that the shrinkage induced by g is strong enough and/or there is not a large proportion of exactly zero Ω entries; g increases with W , which means that the partial correlations are penalized more for neurons that are farther apart.

g increases with W , which suggests that incorporating W in the model helped estimate the network [Vinci et al. (2018a); see also Figure 8]. We drew similar conclusions by inspecting the plot of unregularized square-rooted sample spike count partial correlations versus inter-neuron distance [Figure 11, Appendix D, Vinci et al. (2018b)]: although they are noisy versions of the log firing rate partial correlations, their values appear to be centered at zero and their spread to decrease with inter-neuron distance. Therefore, to avoid the excess variability that typically results from fitting nonexistent or weak effects, we refitted the multivariate PLN model to the data, assuming $f = 0$ and $\eta = 1$, that is, $m_{ij} = 0$ and $\pi_{ij} = 1$ for all (i, j) in equations (2.3) and (2.4). Appendix D [Vinci et al. (2018b)] contains some diagnostic plots that suggest a good model fit. Figure 6(C) shows the re-estimated penalty function $g(W)$ in the four task combinations: just as in Figure 5, g increases with inter-neuron distance W . Note that $W > 5$ includes all the neuron pairs that are in different brain areas, so that the penalty across areas is separate from that within areas. The fitted penalty across areas is very strong, indicating much greater sparsity (much less connectivity) across areas than within areas.

Figure 6(A) displays several versions of the graphs estimated by applying 5% Bayesian FDR* control in the four experimental conditions. For ease of visual comparison, for both attend in and attend out conditions we have displayed edges that appear only at orientation 135 degrees but not at 45 degrees, only at orientation 45 degrees but not at 135 degrees, and edges that appear for both orientations (i.e., intersection of the two graphs under the two orientations). Table 1 contains the numbers of edges in these graphs, and the ratio ($\#$ intersection edges/ $\#$ union edges) $\in [0, 1]$, to quantify the similarity of functional connectivity induced by the two stimulus orientations (45 and 135 degrees) in each of the two attention conditions. There is substantial functional connectivity in pairs of neurons across regions, as well as within regions; there is somewhat diminished connectivity (in terms of number of edges) for the attend in conditions as

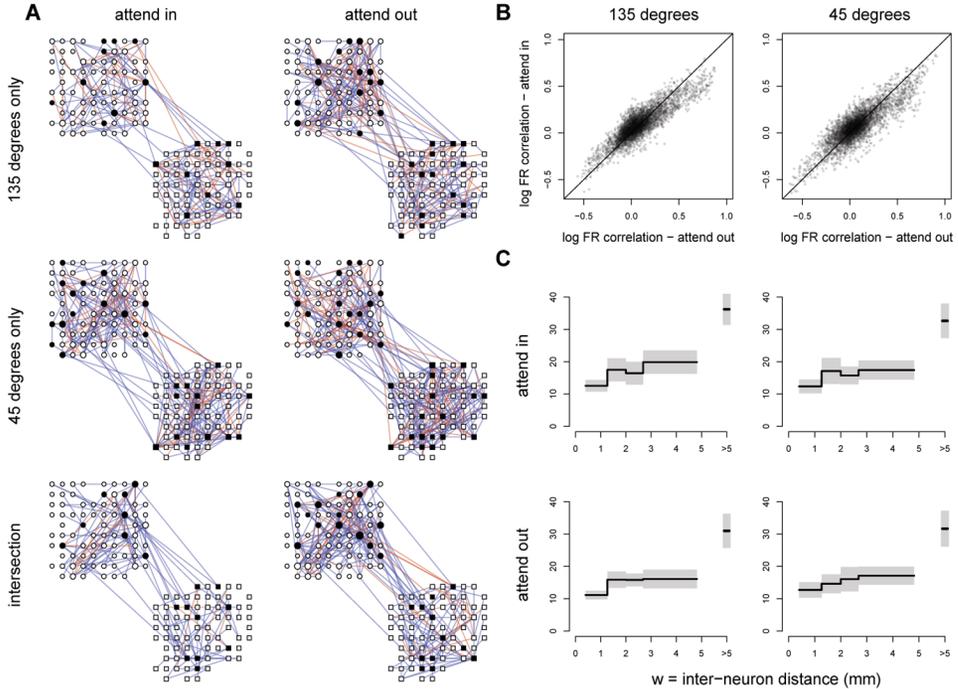


FIG. 6. (A) Functional connectivity graphs of log firing rates of neurons in PFC (round nodes) and V4 (squared nodes); 5% Bayesian FDR* [equation (3.26), Section 3.3; (p, δ) -graphs where δ was taken to be equal to the largest of the 20th percentiles of the magnitudes of the partial correlation estimates across the four conditions]. We show edges that appear only at orientation 135 degrees but not at 45 degrees, only at orientation 45 degrees but not at 135 degrees, and edges that appear for both orientations. Blue and red edges are positive and negative connections, respectively; we blackened the nodes that connect the two areas. (B) Correlations of log firing rates of the fitted PLN models under the attend in and attend out conditions, for both orientations, among pairs of neurons for which one neuron is in V4 and the other is in PFC. The magnitudes of correlations tend to be diminished for the attend in condition (95% confidence intervals for the linear regression slopes of attend in on attend out correlations are [0.74, 0.76] and [0.75, 0.77], respectively). (C) The estimated function g [equation (2.5)] with 95% posterior intervals. The function increases with inter-neuron distance W in all conditions, which implies a stronger penalization on partial correlations for pairs of neurons that are further apart.

compared with the attend out, and somewhat greater connectivity for orientation 45 degrees than for 135 degrees.

We also observed diminished connectivity within V4 and PFC for the attend in conditions as compared with the attend out in terms of edge correlation magnitudes [not shown but similar to Figure 6(B)]. The tendency for diminished magnitude of correlation for the attend in condition is similar to previous findings of reduced trial-to-trial pairwise correlation with attention found within single visual areas [Cohen and Maunsell (2009), Mitchell, Sundberg and Reynolds (2009), Snyder, Morais and Smith (2016)]. Another report [Ruff and Cohen (2016)], how-

TABLE 1

Number of edges and similarity of the graphs (135 degrees versus 45 degrees in the case of attend in or attend out) displayed in Figure 6(A)

	Attend in	Attend out
135° only	208	284
45° only	277	325
Intersection	147	251
Graph similarity	23%	29%

ever, found that spike count correlations between visual cortical areas V1 and MT increased when spatial attention was directed to the joint receptive fields of the recorded neurons. One possible explanation lies in the role of these regions in spatial attention: V1 is thought to provide feedforward input to MT neurons, while PFC has more often been implicated in sending feedback connections to visual areas, including V4. It is also of interest to examine pairwise correlation structure across areas, which we can do here because neurons were recorded simultaneously in V4 and PFC. Figure 6(B) displays a plot of the fitted correlations (the correlations among the latent Poisson log means, representing the log firing rates) under the attend in and attend out conditions, for both orientations, among pairs of neurons for which one neuron is in V4 and the other is in PFC. Once again we find diminished magnitude of correlation in the attend in condition.

Inverting the estimated precision matrix gives the covariance matrix, from which we may compute R^2 using a standard formula from multivariate analysis. In particular, we may choose a single neuron and ask how well its response (its log firing rate) may be predicted based on all the other neurons within its area, or based on all the neurons in the other area. Figure 7 displays plots of R^2 for all neurons in V4 and PFC, where the x -axis is R^2 using neurons within the same brain area, and the y -axis is R^2 using the neurons in the other area. These plots show two things. First, the prediction of a given neuron's firing rate tends to be better when predicting from neurons within the same brain area than from neurons in the other area. This is not surprising. More interesting is the strong relationship between the two, and the relatively small discrepancy between these predictions. This is especially true for area V4 in the attend out condition: the blue dots in the lower left panel cluster strongly around a line that is close to the diagonal; that is, a V4 neuron whose activity is highly predictable (or much less predictable) from other V4 neurons is also highly predictable (or much less predictable) from PFC neurons. This is a striking result, suggesting that major sources of trial-to-trial covariation are shared across these two areas, especially in the attend out condition.

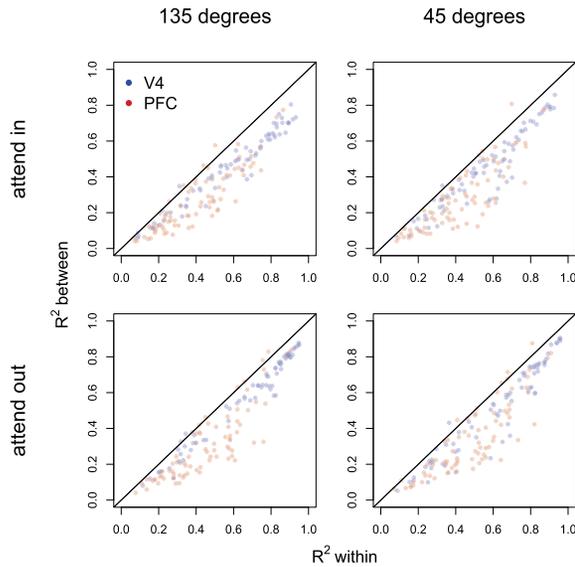


FIG. 7. Coefficients of determinations R^2 of linear regressions of the log firing rate of each neuron in V4 and PFC on the log firing rates of all other neurons within the same brain area (x-axis) and in the other area (y-axis). The values of R^2 were determined from the fitted precision matrix.

4.1. *Analyses with simpler graph estimators.* For comparison's sake, we repeated our data analyses with related methods applied to square-rooted spike-counts (the square root transformation improves the fit of a GGM to count data). The simplest estimator of the covariance matrix is the sample covariance matrix (SAM) whose inverse can be used to obtain sample partial correlations that can be thresholded to obtain a graph. Specifically we applied the Fisher-transformation to z -score the partial correlations and then applied classical 5% FDR control [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)] on the p -values. The graphical LASSO, Glasso [Yuan and Lin (2007)], and its adaptive variant AGlasso [Fan, Feng and Wu (2009)] are graph estimators based on ℓ_1 penalization that favor zeros in the precision matrix; graph structure can be selected via ten-fold cross-validation [CV; Fan, Feng and Wu (2009)]. We do not show the graphs estimated using these three methods, and instead summarize them in Table 2 in Appendix D.3 [Vinci et al. (2018b)], analogously to Table 1. Although the numbers of estimated edges are different than in Table 1, the inferences drawn from these numbers are qualitatively similar to the PLN results, with the exception that Glasso suggests an increased connectivity (in terms of number of edges) for the attend in conditions as compared with the attend out, at the 45 degrees orientation.

For the R^2 analysis reported in Figure 7, we further considered factor analysis (FA): we let the data vector S be a noisy inflation of a latent lower dimensional vector $F \sim N(0, I_q)$, according to $S = AF + \epsilon$, where $\epsilon \sim N(0, D)$, $A \in \mathbb{R}^{d \times q}$, and

D is a positive diagonal matrix; this in turn implies $\Omega^{-1} = AA' + D$. If $q < d/2$ then the number of parameters involved ($d(q + 1)$) is smaller than the full dimensionality of a $d \times d$ symmetric matrix ($d(d + 1)/2$). For each of the four experimental conditions, we computed the maximum likelihood estimate of Ω based on the $d = 148$ neurons assuming a factor structure with q selected via ten-fold CV. We found that about 22 factors minimized the cross-validated error in each condition. In Appendix D.3 [Vinci et al. (2018b)] Figure 13 we reproduce Figure 7 with SAM, Glasso, AGlasso, and FA. Results from the first three methods appear to be noisy versions of those in Figure 7. However, because FA allows the data variables to be close to multicollinear, it produces an uninterpretable plot with all values of R^2 essentially equal to 1.

In the end, we have more faith in the inferences from our covariate-adjusted PLN model, because in Vinci et al. (2018a) we showed via a large simulation study that the simpler case of a covariate-adjusted GGM (GAR) applied to square rooted spike counts with no hierarchical structure in equation (2.1) improved graph estimation, covariance structure inference, and thereby the accuracy of the results of scientific investigations, compared to related methods; and in Section 4.2 we provide additional results that show that covariate-adjusted PLN is better than GAR to extract latent firing rate network information, based on simulated data similar to the experimental data analyzed here.

4.2. Properties of the multivariate PLN model with auxiliary regularization.

We simulate spike count data from a ground truth model similar to the PLN model fitted to the experimental data (Figure 6), estimate the network of their latent firing rates using the PLN model, with and without auxiliary information, also estimate the network of observed spike counts using a GGM applied to their square-roots (GGM-sq), with and without auxiliary information, compare the fits to the ground truth using ROC curves and surfaces (Section 3.5), repeat the simulation 20 times, and report average ROC curves, AUC values, and VUS values.

For the ground truth, we set f to zero and η to one, and we let g be an increasing step function with values 10, 15, 15, 20, 30 over respective bins⁶ (0, 1], (1, 2], (2, 3], (3, 4], (4, 5], with $W \sim \text{Uniform}(0, 5)$, and generate a precision matrix Ω according to equation (2.3). Partial correlations are then set to zero for $W > 4$ to mimic the experimental data, since only a few edges exist between V4 and PFC in Figure 6(A). We simulate $n = 500$ samples from a $d = 50$ dimensional random vector following the multivariate PLN distribution in equation (2.1) with the generated Ω , and μ containing values observed in the experimental data. We fit the GGM-sq and PLN models with $f \equiv 0$, $g(w)$ a step function, and $\eta \equiv 1$, denoted by (g) in the legend of Figure 8, $f(w)$ and $g(w)$ step functions, and

⁶The binning choice is equivalent to setting steps at quintiles of W_{ij} , to obtain five steps as we did with the experimental data.

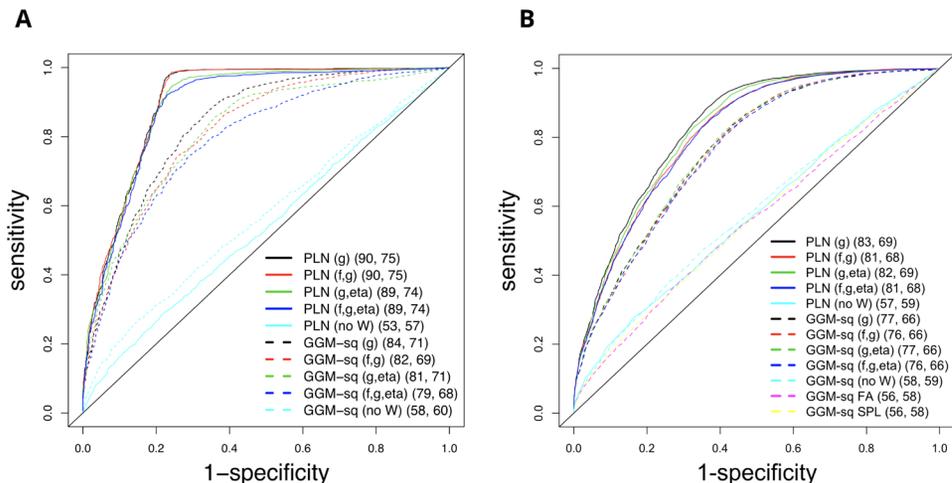


FIG. 8. ROC curves for firing-rate graph estimation based on Poisson-lognormal (PLN) and GGM applied to square rooted spike counts (GGM-sq), averaged across repeat simulations, in the case of a fully recorded network of 50 neurons (A) and a network with 50 recorded and 10 latent neurons (B). The first set of parentheses indicates which functions of W were fitted in the adjusted regularization; the second set contains the area under the curve (AUC) and rescaled volume under the surface (VUS, Section 3.5), averaged across repeat simulations.

$\eta \equiv 1$, denoted by (f, g) , et cetera. The (no W) fit corresponds to using a flat regularization such as the graphical LASSO [Wang (2012), Yuan and Lin (2007)], with $f \equiv 0$, g constant, and $\eta \equiv 1$.

Figure 8(A) shows that the PLN fit retrieves the firing rate network better than the GGM-sq fit, since all the PLN ROC curves are well above their GGM counterparts, and that without auxiliary information, neither PLN nor GGM-sq provides as good a fit as when W is incorporated in the estimation. Note that fitting f , g , and η as functions of W even if they do not all depend on W (only g actually depends on W in this simulation) can degrade the fit somewhat but it remains far superior to the fits that do not account for W . We provide more simulation results in Appendix A [Vinci et al. (2018b)].

In any statistical investigation about dependence among variables, including graphs or multivariate regression, confounding variables are always there lurking behind the scenes. One way to capture their effects is to include as many available variables as possible in the modeling, with the hope to capture the most relevant ones. Alternatively, we may assume that a low-dimensional latent quantity is present. In the context of covariance estimation, factor analysis [FA; Ecker et al. (2014), Rabinowitz et al. (2015)] and the sparse-low rank model [SPL; Chandrasekaran, Parrilo and Willsky (2012), Yatsenko et al. (2015), Yuan (2012)] have been used to model the effects of such low-dimensional latent components. The simulated data here have been generated in a way that could reflect the real data we analyze, and in Figure 8(A) we emulated the case where all neurons in the

network were recorded. We wanted then to consider also the situation where some neurons were latent, to assess how differently FA and SPL would perform compared to GGM-sq (no W) and the methods that use the auxiliary information of W . We thus repeated the simulation, but generated count vectors for 60 neurons, and dropped 10 of them. Results are in Figure 8(B). The performances of FA and SPL (with optimal oracle tuning parameter choices) are similar to the simpler GGM-sq (no W), while our approach maintains performance robustly. We also applied FA and SPL to the case of fully recorded neurons, and conclusions about their performance were the same. However, it is possible that a covariate adjusted version of FA or SPL might produce better results; their implementation is part of our future research.

5. Discussion. We have developed and studied a framework for estimating generalized graphical models in the presence of auxiliary variables that carry information about the graph. The graph is defined by the precision matrix for a multivariate normal distribution of a latent random vector. Here, the components of the latent random vector are logs of Poisson mean parameters, which in the context of spike counts represent log firing rates. The research we report here greatly extends previous work by combining two previously separate aspects of the model: the Poisson log-normal hierarchical aspect (which was previously considered only in the bivariate case), and the idea of using covariates to control the penalty in a sparse Gaussian graphical model. Graphs are summaries of correlation structure, and, together, Figures 6 and 7 indicate the kinds of physiological conclusions that can be drawn from the methodology we have developed: we can either examine the graph itself, as in Figure 6(A), or we can look at summaries of the covariance matrix, as in Figures 6(B) and 7.

The approach we took here was Bayesian, and relied on a flexible and data-driven covariate-adjusted spike and slab model that includes several existing types of network regularization models as special cases, for example, ridge, LASSO, and sparse-low rank decomposition, and that can readily accommodate other types of less explored network regularizations, for example, the Elastic Net, as described in Appendix A [Vinci et al. (2018b)]. The fit of that model to the experimental neuron data of Section 4 suggested that ridge or LASSO-type penalties were adequate for the data we examined. Appendix A contains a toy example where the general spike and slab prior performs better than the currently available priors. Additionally, while our method and estimation algorithms were motivated by neuron spike count data, they are generally applicable to different types of data whenever auxiliary variables are available to help determine latent graphical structure; for example, in equation (2.1), the Poisson distribution may be replaced by others, such as binomial and negative-binomial, or continuous distributions. Because it allows for many precision matrix regularizations and data types, it may be considered a unifying framework for estimating covariate-adjusted networks. In

the future, covariate-adjusted regularization could also be extended to multivariate time series and doubly-stochastic point processes.

The neurophysiological variables used as auxiliary quantities in this and previous research [Vinci et al. (2018a)] were pairwise, that is, one quantity per neuron pair (e.g., inter-neuron distance and tuning curve correlation), so it was a natural choice to introduce such information through an entrywise regularization scheme targeting the entries of the precision matrix, rather than other features (e.g., eigenvalues, factors) that do not map easily onto entrywise auxiliary information. We have provided evidence that this approach can do a good job with estimation of neuronal covariance. Furthermore, graphs provide nice interpretations of the covariance matrix. An alternative approach is to introduce latent factors, which can provide useful dimensionality reduction.

To account for the effect of unrecorded neurons and other latent factors on the dependence structure of the recorded neurons, our framework can be adapted in at least two ways. One approach is analogous in construction to the latent variable graphical model [Chandrasekaran, Parrilo and Willsky (2012)], discussed in Appendix A [Vinci et al. (2018b)]. Alternatively, one could assume $X | F \sim N(\mu + AF, \Omega^{-1})$, where $F \sim N(0, I_q)$ are latent factors, $q < d/2$, $A \in \mathbb{R}^{d \times q}$, and Ω is sparse, so that the covariance matrix of X is $AA' + \Omega^{-1}$, where AA' has rank at most q . By further assuming a prior distribution on Ω as in equation (2.3), with continuous Laplace p.d.f. components in equation (2.4), and a prior distribution on A that constrains the model to be identifiable, we can estimate AA' and Ω by implementing our full or empirical Bayes algorithms [Appendix C, Vinci et al. (2018b)] with additional sampling steps for $F | \text{rest}$ and $A | \text{rest}$. In either case, the sparse component of the precision matrix would represent the dependence structure of the observed variables conditionally on both observed and latent variables. Implementing these approaches was beyond the scope of the work reported in this paper, but we plan to investigate them in the future.

Acknowledgments. We are grateful to Adam Snyder for data collection and assistance with data analysis.

SUPPLEMENTARY MATERIAL

Supplement to “Adjusted regularization in latent graphical models: Application to multiple-neuron spike count data.” (DOI: [10.1214/18-AOAS1190 SUPP](https://doi.org/10.1214/18-AOAS1190SUPP); .pdf). Appendix containing: Appendix A additional details about adjusted regularization and simulations, Appendix B lemmas with proofs, Appendix C algorithms, and Appendix D additional data analyses.

REFERENCES

ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B* **36** 99–102. [MR0359122](https://doi.org/10.2307/2344638)

- BANERJEE, S. and GHOSAL, S. (2015). Bayesian structure learning in graphical models. *J. Multivariate Anal.* **136** 147–162. [MR3321485](#)
- BEHSETA, S., BERDYJEVA, T., OLSON, C. R. and KASS, R. E. (2009). Bayesian correction for attenuation of correlation in multi-trial spike count data. *J. Neurophysiol.* **101** 2186–2193.
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245](#)
- BRILLINGER, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybernet.* **59** 189–200.
- CHANDRASEKARAN, V., PARRILO, P. A. and WILLSKY, A. S. (2012). Latent variable graphical model selection via convex optimization. *Ann. Statist.* **40** 1935–1967. [MR3059067](#)
- CHURCHLAND, A. K., KIANI, R., CHAUDHURI, R., WANG, X. J., POUGET, A. and SHADLEN, M. N. (2011). Variance as a signature of neural computations during decision making. *Neuron* **69** 818–831.
- COHEN, M. R. and MAUNSELL, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12** 1594–1600.
- COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. 401–448. [MR0413254](#)
- ECKER, A. S., BERENS, P., COTTON, R. J., SUBRAMANIAN, M., DENFIELD, G. H., CADWELL, C. R., SMIRNAKIS, S. M., BETHGE, M. and TOLIAS, A. S. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82** 235–248.
- EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. [MR1946571](#)
- FAN, J., FENG, Y. and WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* **3** 521–541. [MR2750671](#)
- FIENBERG, S. E. (1974). Stochastic models for single neuron firing trains: A survey. *Biometrics* **30** 399–427. [MR0359082](#)
- FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- GEORGE, E. I. and MCCULLOCH, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7** 339–373.
- GEORGOPOULOS, A. P. and ASHE, J. (2000). One motor cortex, two different views. *Nat. Neurosci.* **3** 964–963; author reply 965.
- GIRAUD, C. and TSYBAKOV, A. (2012). Discussion: Latent variable graphical model selection via convex optimization [MR3059067]. *Ann. Statist.* **40** 1984–1988. [MR3059071](#)
- GORIS, R. L., MOVSHON, J. A. and SIMONCELLI, E. P. (2014). Partitioning neuronal variability. *Nat. Neurosci.* **17** 858–865.
- HINNE, M., AMBROGIONI, L., JANSSEN, R. J., HESKES, T. and VAN GERVEN, M. A. J. (2014). Structurally-informed Bayesian functional connectivity analysis. *NeuroImage* **86** 294–305.
- HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.
- INOUE, D. I., YANG, E., ALLEN, G. I. and RAVIKUMAR, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdiscip. Rev.: Comput. Stat.* **9** e1398, 25. [MR3648601](#)
- KASS, R. E., EDEN, U. T. and BROWN, E. N. (2014). *Analysis of Neural Data*. Springer, New York. [MR3244261](#)
- KASS, R. E., VENTURA, V. and BROWN, E. N. (2005). Statistical issues in the analysis of neuronal data. *J. Neurophysiol.* **94** 8–25.
- KASS, R. E., AMARI, S. I., ARAI, K., BROWN, E. N., DIEKMAN, C. O., DIESMANN, M. . . . and FUKAI, T. (2018). Computational neuroscience: Mathematical and statistical perspectives. *Ann. Rev. Statist. Appl.* **5** 183–214.

- MAUNSELL, J. H. (2015). Neuronal mechanisms of visual attention. *Ann. Rev. Vision Sci.* **1** 373–391.
- MAZUMDER, R. and HASTIE, T. (2012). The graphical lasso: New insights and alternatives. *Electron. J. Stat.* **6** 2125–2149. [MR3020259](#)
- MITCHELL, J. F., SUNDBERG, K. A. and REYNOLDS, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63** 879–888.
- NG, B., VAROQUAUX, G., POLINE, J. B. and THIRION, B. (2012). A novel sparse graphical approach for multimodal brain connectivity inference. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 707–714. Springer, Berlin, Heidelberg.
- PERKEL, D. H., GERSTEIN, G. L. and MOORE, G. P. (1967a). Neuronal spike trains and stochastic point processes: I. The single spike train. *Biophys. J.* **7** 391–418.
- PERKEL, D. H., GERSTEIN, G. L. and MOORE, G. P. (1967b). Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophys. J.* **7** 419–440.
- PINEDA-PARDO, J. A., BRUÑA, R., WOOLRICH, M., MARCOS, A., NOBRE, A. C., MAESTÚ, F. and VIDAURRE, D. (2014). Guiding functional connectivity estimation by structural connectivity in MEG: An application to discrimination of conditions of mild cognitive impairment. *NeuroImage* **101** 765–777.
- POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. [MR3174712](#)
- RABINOWITZ, N. C., GORIS, R. L., COHEN, M. and SIMONCELLI, E. P. (2015). Attention stabilizes the shared gain of V4 populations. *eLife* **4** e08998.
- RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. [MR2836766](#)
- ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. [MR2417391](#)
- RUFF, D. A. and COHEN, M. R. (2016). Stimulus dependence of correlated variability across cortical areas. *J. Neurosci.* **36** 7546–7556.
- SCOTT, J. G., KELLY, R. C., SMITH, M. A., ZHOU, P. and KASS, R. E. (2015). False discovery rate regression: An application to neural synchrony detection in primary visual cortex. *J. Amer. Statist. Assoc.* **110** 459–471. [MR3367240](#)
- SMITH, M. A. and KOHN, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28** 12591–12603.
- SMITH, M. A. and SOMMER, M. A. (2013). Spatial and temporal scales of neuronal correlation in visual area V4. *J. Neurosci.* **33** 5422–5432.
- SNYDER, A. C., MORAIS, M. J. and SMITH, M. A. (2016). Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *J. Neurophysiol.* **116** 1807–1820.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- VINCI, G., VENTURA, V., SMITH, M. A. and KASS, R. E. (2016). Separating spike count correlation from firing rate correlation. *Neural Comput.* **28** 849–881.
- VINCI, G., VENTURA, V., SMITH, M. A. and KASS, R. E. (2018a). Adjusted regularization of cortical covariance. *J. Comput. Neurosci.* To appear.
- VINCI, G., VENTURA, V., SMITH, M. A. and KASS, R. E. (2018b). Supplement to “Adjusted regularization in latent graphical models: Application to multiple-neuron spike count data.” DOI:10.1214/18-AOAS1190SUPP.
- WANG, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Anal.* **7** 867–886. [MR3000017](#)
- WANG, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Anal.* **10** 351–377. [MR3420886](#)
- WILSON, H. R. and COWAN, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12** 1–24.

- YATSENKO, D., JOSIĆ, K., ECKER, A. S., FROUDARAKIS, E., COTTON, R. J. and TOLIAS, A. S. (2015). Improved estimation and interpretation of correlations in neural circuits. *PLoS Comput. Biol.* **11** e1004083.
- YU, B. M., CUNNINGHAM, J. P., SANTHANAM, G., RYU, S. I., SHENOY, K. V. and SAHANI, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in Neural Information Processing Systems* 1881–1888.
- YUAN, M. (2012). Discussion: Latent variable graphical model selection via convex optimization [MR3059067]. *Ann. Statist.* **40** 1968–1972. [MR3059068](#)
- YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. [MR2367824](#)
- ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. [MR2137327](#)

G. VINCI
DEPARTMENT OF STATISTICS
RICE UNIVERSITY
DUNCAN HALL
6100 MAIN ST.
HOUSTON, TEXAS 77005
USA
E-MAIL: gv9@rice.edu

V. VENTURA
R. E. KASS
DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
BAKER HALL 132
5000 FORBES AVENUE
PITTSBURGH, PENNSYLVANIA 15203
USA
E-MAIL: vventura@stat.cmu.edu
kass@stat.cmu.edu

M. A. SMITH
DEPARTMENT OF OPHTHALMOLOGY
EYE AND EAR INSTITUTE
UNIVERSITY OF PITTSBURGH
ROOM 914
203 LOTHROP ST.
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: smithma@pitt.edu