Separating Spike Count Correlation from Firing Rate Correlation

Giuseppe Vinci
gvinci@andrew.cmu.edu
Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Valérie Ventura
vventura@stat.cmu.edu
Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.,
and Center for the Neural Basis of Cognition, Pittsburgh, PA 15213, U.S.A.

Matthew A. Smith
smithma@pitt.edu
Department of Ophthalmology, University of Pittsburgh, Pittsburgh, PA 15260,
U.S.A., Center for the Neural Basis of Cognition, Pittsburgh, PA 15213, U.S.A.

Robert E. Kass
kass@stat.cmu.edu
Department of Statistics and Department of Machine Learning, Carnegie Mellon
University, Pittsburgh, PA 15213, U.S.A., and Center for the Neural Basis
of Cognition, Pittsburgh, PA 15213, U.S.A.

Populations of cortical neurons exhibit shared fluctuations in spiking activity over time. When measured for a pair of neurons over multiple repetitions of an identical stimulus, this phenomenon emerges as correlated trial-to-trial response variability via spike count correlation (SCC). However, spike counts can be viewed as noisy versions of firing rates, which can vary from trial to trial. From this perspective, the SCC for a pair of neurons becomes a noisy version of the corresponding firing rate correlation (FRC). Furthermore, the magnitude of the SCC is generally smaller than that of the FRC and is likely to be less sensitive to experimental manipulation. We provide statistical methods for disambiguating time-averaged drive from within-trial noise, thereby separating FRC from SCC. We study these methods to document their reliability, and we apply them to neurons recorded in vivo from area V4 in an alert animal. We show how the various effects we describe are reflected in the data: within-trial effects are largely negligible, while attenuation due to trial-to-trial variation dominates and frequently produces comparisons in SCC that, because of noise, do not accurately reflect those based on the underlying FRC.

1 Introduction

Analysis of spike count correlation has provided important neurophysiological insights, as summarized in Cohen and Kohn (2011), Gu et al. (2011), Harris and Thiele (2011), Jeanne, Sharpee, and Gentner (2013), Mitchell, Sundberg, and Reynolds (2009), Smith and Kohn (2008), and Smith and Sommer (2013). However, as documented in these references, a striking feature of the interaction among neurons is that it can occur at multiple timescales. A neuron's spike count collected across a time interval of length T, such as $T = 1000$ ms, reflects both the time-averaged drive to that neuron and additional stochastic fluctuations, which produce irregular spiking at timescales smaller than T, usually described as point process variation (Shadlen & Newsome, 1998; Churchland et al., 2011). That is, the spike count involves both a relatively slowly varying input drive, at the timescale T, and faster point-process fluctuations that determine when the neuron fires. Across an interval of length T, this point-process variation determines the number of times each neuron fires for a given level of time-averaged input drive. Correspondingly, for a given pair of neurons, the spike count correlation (SCC) may reflect (A) the firing rate correlation (FRC), that is, the correlation in the time-averaged drive to these neurons, and (B) point process noise (Kass & Ventura, 2006; Staude, Rotter, & Grún, 2008; Goris, Movshon, & Simoncelli, 2014). It also could involve (C) more precisely timed within-trial correlation effects that might arise, for instance, from (lagged) synchronous spiking (Kelly & Kass, 2012; Harrison, Amarasingham, & Kass, 2013), which has been observed in some cortical areas (macaque V4: Smith & Kohn, 2008; Smith & Sommer, 2013) and may be modulated by state transitions that are more prominent in some anesthetized preparations (Kelly, Smith, Kass, & Lee, 2010; Ecker et al., 2014). Longer timescale effects, but still shorter than trial length, such as correlated up and down states (Steriade & Buzsaki, 1990; Cowan & Wilson, 1994; Timofeev, Grenier, & Steriade, 2001; Steriade, Timofeev, & Grenier, 2001), can also produce within-trial spike count covariability.

In this letter, we describe ways in which the SCC may not accurately measure the FRC in component A due to distortion caused by component B, and we present statistical methods for disambiguating component A from component B. Other studies have developed related techniques: Churchland et al. (2011), Goris et al. (2014), and Staude et al. (2008). Here we go further, using both parametric and nonparametric approaches, in order to provide a comprehensive set of tools that, as we show, perform well under conditions similar to those found with neural data. When component C is negligible, statistical inference and interpretation are simplified. Therefore, we focus especially on methods for evaluating the relevance of that component to the analysis of spike counts, including both statistical tests and estimates of magnitude. We then use the tools we have developed to analyze data
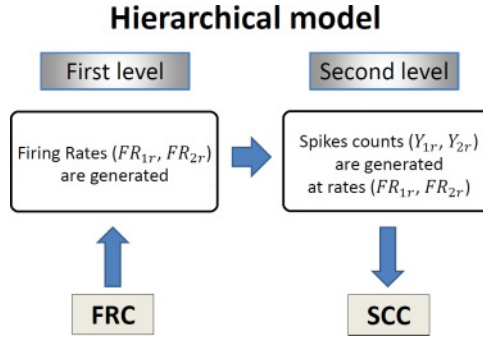
Hierarchical model

First level

Second level

Firing Rates (FR_{1r}, FR_{2r}) are generated

Spikes counts (Y_{1r}, Y_{2r}) are generated at rates (FR_{1r}, FR_{2r})

FRC

SCC

Figure 1: Hierarchical model for correlation sources. First level: the firing rates (FR_1, FR_2) of two neurons have correlation FRC. Second level: the spike counts (Y_{1r}, Y_{2r}) are jointly generated with rates (FR_{1r}, FR_{2r}) and have correlation SCC. The first-level correlation (which is unobserved) describes the relationship between the two neurons, but it is contaminated in the second level with Poisson-like noise.

recorded from area V4 while a monkey performed a fixation task, and we demonstrate that FRC behaves differently than SCC does in this setting.

We begin by assuming that the data are observed across multiple trials of total elapsed time T and that for trials $r = 1, \ldots, n$, each pair of observed spike counts (Y_{1r}, Y_{2r}) for two neurons comes from the doubly stochastic correlated point process depicted in Figure 1. This process consists of two levels. The first level consists of a pair of latent (unobserved) variables (X_{1r}, X_{2r}) generated with some across-trial correlation (trial-to-trial correlation), where, on trial r, X_{ir} is the average firing rate FR_{ir} for neuron i multiplied by T, $X_{ir} = FR_{ir}T$. The variable X_{ir} may be interpreted as an average input drive to neuron i. The second level consists of observed spike counts (Y_{1r}, Y_{2r}), generated jointly with expectation (X_{1r}, X_{2r}). In addition, there may be some within-trial point-process correlation. Thus, the conditional expectation of Y_{ir}, conditionally on (X_{1r}, X_{2r}), is

$$E(Y_{ir} \mid X_{1r}, X_{2r}) = X_{ir}. \tag{1.1}$$

We further assume that the conditional variance is

$$\mathrm{Var}(Y_{ir} \mid X_{1r}, X_{2r}) = \phi_i X_{ir}, \tag{1.2}$$

where $\phi_i \in \mathbb{R}^+$ controls the dispersion of Y_{ir} conditionally on X_{ir} (Kass & Ventura, 2006; Churchland et al., 2011; Goris et al., 2014). The case $\phi_i = 1$ holds when Y_{ir} follows a Poisson distribution, conditionally on X_{ir}.

A fundamental decomposition of variability is given by

$$\text{Cov}(Y_{1r}, Y_{2r}) = \text{E}[\text{Cov}(Y_{1r}, Y_{2r} \mid X_{1r}, X_{2r})]$$
$$+ \text{Cov}(\text{E}[Y_{1r} \mid X_{1r}, X_{2r}], \text{E}[Y_{2r} \mid X_{1r}, X_{2r}]). \qquad (1.3)$$

The observed trial-to-trial spike count correlation (SCC) is the Pearson correlation of the counts Y_{1r} and Y_{2r} across trials, that is the normalized version of the left-hand side of equation 1.3, while the unobserved trial-to-trial firing rate correlation (FRC) is the Pearson correlation of FR_{1r} and FR_{2r} or, equivalently, of X_{1r} and X_{2r} across trials, that is, the normalized version of the second term on the right-hand side. In section 2, we show that on a sequence of independent trials, the SCC can be seen as a perturbed version of FRC. We characterize this phenomenon by the simple formula,

$$\text{SCC} = \text{FRC} \times \text{ATT} + \Gamma, \qquad (1.4)$$

where we identify an attenuation component (ATT), mainly due to trial-to-trial variability, and a deviation component (Γ), which summarizes the impact of within-trial point process correlation (component C in the first paragraph of this letter on SCC).

It is not possible to estimate the components in equation 1.4 separately, based on spike counts alone, without additional modeling assumptions (Amarasingham, Geman, & Harrison, 2015). In conjunction with equation 1.3, Churchland et al. (2011) imposed constraints on the variance of the spike counts, constraints estimated from the data, to extract method of moments estimates of correlations between single-neuron firing rates at different epochs of a trial. By assuming univariate Poisson-gamma models for single-neuron spike counts and then also assuming a sampling distribution for the values of SCC across stimuli, Goris et al. (2014) obtained estimates for FRC and for a quantity related to Γ. Because they used univariate models for the spike counts, they were only able to obtain a single estimate of FRC for all stimuli. However, as illustrated in Figure 3A, the FRC of a neuron pair can (and often does) vary across stimuli. The approach we take allows for this. We build on previous work of Ventura, Cai, and Kass (2005) and Kass and Ventura (2006), who proved that spike count correlation necessarily increases with T in the presence of multiplicative trial-to-trial variation in time-averaged input drive, and Behseta, Berdyyeva, Olson, and Kass (2009), who used a bivariate hierarchical model as in Figure 1 to provide a correction for attenuation of correlation in the presence of point process noise. In section 3 we derive nonparametric estimators of the components of equation 1.4 from spike trains (as opposed to from spike counts alone) by showing that certain estimators of intermediate quantities are unbiased. We also derive a nonparametric estimator of ϕ_i in equation 1.2 from spike trains, in contrast to Goris et al. (2014), who assume $\phi_i = 1$, and Churchland et al.
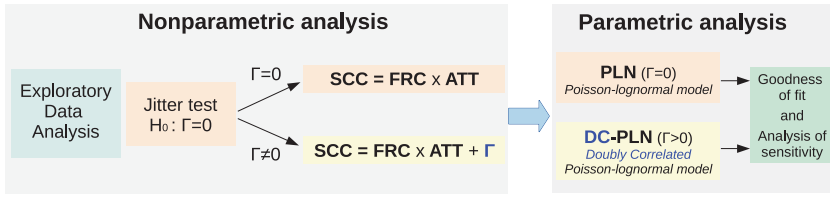
Figure 2: Analysis road map. We first use nonparametric estimators (section 3.2) to explore the data and implement hypothesis tests to determine which pairs of neurons have $\Gamma \neq 0$ (jitter test, section 3.3). We then fit appropriate bivariate parametric models (we impose $\Gamma = 0$ if $\Gamma \leq 0$ was obtained from the jitter test; section 3.1) and check their goodness of fit and sensitivity (sections 3.1.1 and 3.1.2).

(2011), who use ϕ_i like a tuning parameter and choose it such that variances and correlations lie in the required domains. We provide details on bivariate hierarchical models for spike counts that take the trial-to-trial distribution of (X_{1r}, X_{2r}) to be log normal and the distribution of Y_{ir} to be Poisson with mean X_{ir}, and we suggest bootstrap methods to assess the fit of these models. Because analysis and interpretation are more straightforward when $\Gamma = 0$ in equation 1.4, we propose a jitter test of this hypothesis. We show that the test is sufficiently powerful to find substantial departures from $\Gamma = 0$ under realistic scenarios, and because one typically must examine many pairs of neurons, we use methods that control the false discovery rate (FDR) and the false nondiscovery rate (FNR). The latter assesses the likelihood of error when a decision is made to assume $\Gamma = 0$ based on the jitter test. We also introduce generalized forms of tuning curve and tuning curve correlation (GTC and GTCC; see appendix H) that account for the trial-to-trial variability of the firing rates. In section 4, we analyze the V4 data and find that the models fit well and that they support the effects discussed in section 2. In section 5 we discuss implications of our findings.

Figure 2 summarizes the main features of our approach, in which we begin with nonparametric estimators for exploratory purposes (e.g., making informative plots) and to formulate the jitter test. We then move on to model-based inference and, finally, assess goodness-of-fit and sensitivity.

2 Spike Count Correlation as an Attenuated Firing Rate Correlation

From equations 1.1 and 1.2, we have (see appendix A)

$$E[Y_{ir}] = E[X_{ir}] \text{ and } \mathrm{Var}(Y_{ir}) = \phi_i E[X_{ir}] + \mathrm{Var}(X_{ir}), \tag{2.1}$$

which implies that the spike counts are overdispersed ($\mathrm{Var}(Y_{ir}) > E[Y_{ir}]$) if $\phi_i > 1 - F_i$, where $F_i = \mathrm{Var}(X_{ir})/E[X_{ir}] \geq 0$ is the Fano factor of X_{ir}.
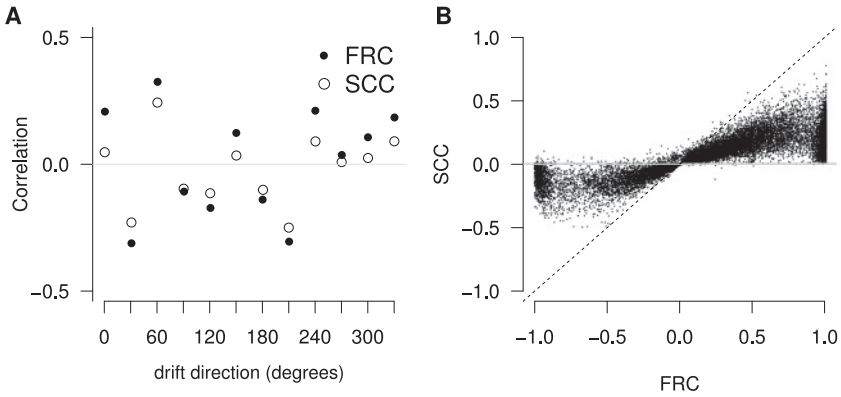
Figure 3: SCC and FRC at the 12 experimental stimuli (drift direction degrees) for a neuron pair (macaque V4 area; section 4). (B) SCC versus FRC for the macaque V4 data analyzed in section 4. The SCC is an attenuated version of the FRC; both generally have the same sign.

Combining equations 1.1, 1.3, and 2.1, SCC = $Cor(Y_{1r}, Y_{2r})$ can be decomposed as

$$SCC = FRC \times ATT + \Gamma \tag{2.2}$$

(see appendix A), where

- FRC = $Cor(E(Y_{1r} \mid X_{1r}), E(Y_{2r} \mid X_{2r})) = Cor(X_{1r}, X_{2r})$ is the trial-to-trial firing rate correlation.
- ATT = $\prod_{i=1,2}(1 + \phi_i/F_i)^{-1/2}$ is the *attenuation coefficient*, ATT $\in$ (0, 1].
- $\Gamma = \gamma / \prod_{i=1,2}(Var(Y_{ir}))^{1/2}$, where $\gamma = E[Cov(Y_{1r}, Y_{2r} \mid X_{1r}, X_{2r})]$ is the trial-averaged within-trial covariance of the spike counts.

The attenuation effect is defined as the case when SCC and FRC have the same sign and SCC has smaller magnitude (|SCC| < |FRC|). Figure 3A shows an example neuron pair (macaque V4 area; section 4) with different attenuation strengths across stimuli, and Figure 3B shows that SCC is an attenuated version of FRC for the majority of neuron pairs in our data set. Because ATT $\in$ (0, 1], attenuation occurs whenever Γ is small, specifically whenever Γ is between FRC $\times$ (1 − ATT) and −FRC $\times$ ATT. The special case $\Gamma = 0$ is of particular interest because equation 2.2 simplifies to SCC = FRC $\times$ ATT, so that the attenuation effect is completely specified by ATT (and the results in Kass & Ventura, 2006, which assume $\Gamma = 0$, hold). In section 3.3, we develop a jitter method to test the hypothesis $\Gamma = 0$, and

in section 4 we show that this hypothesis is reasonable for the majority of neuron pairs in the in vivo macaque V4 data set we analyze.

3 Estimating Sources of Correlations

Our aim here is to estimate the population quantities in equations 2.1 and 2.2 that characterize the hierarchical structure defined in section 2 and Figure 1. In section 3.1 we propose bivariate parametric models for the spike counts of neuron pairs, which imply parametric estimates for the components of equations 2.1 and 2.2. These models assume $\phi_i = 1$, allowing for the overdispersed spike counts that are typical of the visual cortex V4 data we analyze in section 4 (Tomko & Crapper, 1974; Tolhurst, Movshon, & Thompson, 1981; Cohen & Maunsell, 2009; Mitchell et al., 2009; Goris et al., 2014). We also provide methods to evaluate the fit of the models to data and assess their sensitivity to the assumptions. In section 3.2, we propose nonparametric estimators of the population quantities in equations 2.1 and 2.2 based on binned spike counts, which only make assumptions about the timescale of within-trial dependencies between neurons. Because the interpretation of equation 2.2 is more straightforward when $\Gamma = 0$, we develop a model-free jitter test of this hypothesis in section 3.3.

3.1 Parametric Estimation. Goris et al. (2014) used a univariate Poisson-gamma model – $Y_{ir} \mid X_{ir} \sim \text{Poisson}(X_{ir})$ where $X_{ir} \sim \Gamma(\alpha_i, \beta_i)$ – to model single neuron spike counts, but because it has no natural bivariate extension, we extend instead the univariate Poisson-lognormal model of Kass and Ventura (2006):

$$Y_{ir} \mid X_{ir} \sim \text{Poisson}(X_{ir}), \quad i = 1, 2 \tag{3.1}$$

$$\begin{pmatrix} \log X_{1r} \\ \log X_{2r} \end{pmatrix} \sim N\left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right).$$

This bivariate Poisson-lognormal model (PLN) allows for overdispersed spike counts since $\phi_i = 1$ (see section 2 and appendix C) but assumes that $Y_{1r} \mid X_{1r}$ and $Y_{2r} \mid X_{2r}$ are independent, which implies $\gamma = \Gamma = 0$ in equation 2.2. We therefore generalize the model to allow $\gamma \neq 0$. We assume $\gamma \geq 0$ and let

$$Y_{ir} = R_{0r} + R_{ir} \tag{3.2}$$

$$R_{0r} \sim \text{Poisson}(\gamma)$$

$$R_{ir} \mid W_{ir} \sim \text{Poisson}(W_{ir}), \quad i = 1, 2$$

$$\begin{pmatrix} \log W_{1r} \\ \log W_{2r} \end{pmatrix} \sim N(\mu, \Sigma),$$

where R_{0r} and $R_{ir} \mid W_{ir}$, $i = 1, 2$, are mutually independent, and the firing rates are $X_{ir} = \gamma + W_{ir}$. We call this model a bivariate doubly correlated Poisson-lognormal model (DC-PLN) because the spike counts (Y_{1r}, Y_{2r}) given the rates have a correlated Poisson distribution, and the rates (W_{1r}, W_{2r}) are also correlated. Figure 8 shows tolerance regions of the joint distribution of (X_{1r}, X_{2r}) fitted by maximum likelihood (ML; see appendix C) to a neuron pair in macaque area V4 (see section 4), which we also call a generalized tuning curve (GTC; see appendix H). Model 3.2 is appropriate when the spike trains of two neurons share common spikes R_{0r} in addition to, and independent from, (R_{1r}, R_{2r})—for example, if they receive common input from an adjacent third neuron that causes positive conditional within-trial covariance γ. This model does not accommodate the case $\gamma < 0$, which is acceptable for the data set analyzed in section 4, because over 95% of neurons-conditions pairs have $\gamma \geq 0$. When $\gamma = 0$ (equivalently $\Gamma = 0$), model 3.2 reduces to the PLN model in equation 3.1. Section 3.3 details a test of this hypothesis.

Under this model, the spike count correlation formula, equation 2.2, reduces to

$$
\text{SCC} = \underbrace{(e^{\sigma_{12}} - 1) \prod_{i=1,2} (e^{\sigma_i^2} - 1)^{-\frac{1}{2}}}_{\text{FRC}}
$$

$$
\times \underbrace{\prod_{i=1,2} \left(1 + a_i + \left(e^{\mu_i + \sigma_i^2/2}(e^{\sigma_i^2} - 1)\right)^{-1}\right)^{-\frac{1}{2}}}_{\text{ATT}} + \Gamma, \tag{3.3}
$$

where $a_i = \gamma \, (e^{\mu_i + \sigma_i^2/2})^{-2} (e^{\sigma_i^2} - 1)^{-1}$ and $\Gamma = \gamma \prod_{i=1,2}[\gamma + e^{\mu_i + \sigma_i^2/2} + (e^{\mu_i + \sigma_i^2/2})^2 (e^{\sigma_i^2} - 1)]^{-\frac{1}{2}}$. Replacing the parameters by their ML estimates (see appendix C) yields ML estimates of the quantities of interest: $\widehat{\Gamma}_{mle}$, $\widehat{\text{FRC}}_{mle}$, $\widehat{\text{ATT}}_{mle}$, and $\widehat{\text{SCC}}_{mle}$.

Alternatively, the parameters of model 3.2 could be estimated via Bayesian methods (Bernardo & Smith, 2009; Kass, Eden, & Brown, 2014), which require choices of priors for μ, Σ, γ. Assuming squared error loss, FRC would be estimated as the posterior mean from the marginal posterior of Σ.

3.1.1 Assessing the Fit of the Parametric Model. We assess the global goodness-of-fit (GOF) of a parametric model to data using a discrepancy measure between a nonparametric quantity and its parametric analog under the fitted model. A large discrepancy suggests that the model is not a good representation of the true spike count generation process. We consider two discrepancies: the Kullback-Leibler (KL) divergence of the fitted model (univariate or bivariate) from the empirical distribution of the data and, in

the bivariate case, the divergence between the sample SCC and the SCC predicted by the fitted model.

Kullback-Leibler Divergence. Let $Z_r = Y_{ir}$ be the spike count of a neuron in trial r in the univariate case and $Z_r = (Y_{1r}, Y_{2r})$ in the bivariate case. Let $P_n(z) = n^{-1} \sum_{r=1}^{n} I(\|Z_r - z\|_1 = 0)$ be their empirical probability mass function (p.m.f.), where $\|x\|_1 = \sum_j |x_j|$ is the L_1-norm, and $Q_{\widehat{\theta}}$ be a parametric model fitted to $(Z_1, \ldots, Z_n)$. The Kullback-Leibler (KL) divergence between P_n and $Q_{\widehat{\theta}}$ is

$$KL(P_n\|Q_{\widehat{\theta}}) = \sum_z \log(P_n(z)/Q_{\widehat{\theta}}(z))P_n(z) = n^{-1} \sum_{r=1}^{n} \log(P_n(Z_r)/Q_{\widehat{\theta}}(Z_r))$$

$$= n^{-1} \sum_{r=1}^{n} \log P_n(Z_r) - n^{-1} \sum_{r=1}^{n} \log Q_{\widehat{\theta}}(Z_r) \geq 0,$$

where $\sum_{r=1}^{n} \log Q_{\widehat{\theta}}(Z_r) = \max_\theta \log L_Q(\theta; Z_1, \ldots, Z_n)$ is the maximum of the log likelihood of Q_θ. Hence, the KL divergence can be written as

$$KL(P_n\|Q_{\widehat{\theta}}) = n^{-1}(EML - PML),$$

where *EML* stands for "empirical maximum log likelihood" and *PML* for "parametric maximum log likelihood." We use $KL(P_n\|Q_{\widehat{\theta}})$ to test the hypothesis $Z_r \sim Q_\theta$, and approximate its null distribution, F_{KL}, via a parametric bootstrap (Efron & Tibshirani, 1994). We reject the hypothesis if $KL(P_n\|Q_{\widehat{\theta}}) > F_{KL}^{-1}(1 - \alpha)$ for some significance level α.

SCC Divergence. Let Q_θ be a bivariate parametric model for (Y_{1r}, Y_{2r}). Define

$$D = f_Z(\widehat{SCC}_{samp}) - f_Z(\widehat{SCC}_{mle}), \tag{3.4}$$

where $\widehat{SCC}_{samp}$ is the sample SCC of (Y_{1r}, Y_{2r}) (see equation 3.12), $\widehat{SCC}_{mle}$ is the estimate derived from model Q_θ fitted to the data (see equation 3.3), and $f_Z(r) = 1/2 \cdot \log((1 + r)/(1 - r))$ is the Fisher transformation. We reject the hypothesis $Z_r \sim Q_\theta$ if $D < F_D^{-1}(\alpha/2)$ or $D > F_D^{-1}(1 - \alpha/2)$, where F_D is the null distribution of D that we approximate using a parametric bootstrap.

3.1.2 Sensitivity of FRC Estimates to Model Assumptions. Goodness-of-fit tests do not always have high power to detect deviations between the data and the assumed parametric model, so it is useful to assess how sensitive the parametric estimate of FRC is to mild deviations from the DC-PLN model. We do that by simulating spike counts from the more general model specified in equation 3.5, estimating FRC assuming that the data are DC-PLN distributed, and quantifying the deviations of the estimate from its true value.

The generalized bivariate spike count model is defined as

$$Y_{ir} = R_{0r} + R_{ir}, \quad i = 1, 2, \tag{3.5}$$

$$R_{0r} \sim P_{\gamma, \phi},$$

$$R_{ir} \mid W_{ir} \sim P_{W_{ir}, \phi}, \quad i = 1, 2,$$

$$(W_{1r}, W_{2r})^T \sim f_{b, \mu, \Sigma},$$

which has the same structure as model 3.2 but with the Poisson and lognormal distributions replaced by $P_{\lambda, \phi}$ and $f_{b, \mu, \Sigma}$. It includes as a special case the DC-PLN model, equation 3.2, when $\phi = 1$, $b = 0^+$, $\gamma \geq 0$, and the PLN model, equation 3.1, when $\phi = 1$, $b = 0^+$, and $\gamma = 0$. The distributions $P_{\lambda, \phi}$ and $f_{b, \mu, \Sigma}$ are generalizations of the Poisson and lognormal distributions that allow different spread and skewness. They are defined as follows:

1. Under model 3.2, spike counts Y conditional on their rates X are Poisson, which implies $\phi = \text{Var}(Y|X)/\text{E}(Y|X) = 1$ (see section 2). Here we consider a generative model permitting $\phi \neq 1$:

$$P_{\lambda, \phi} =$$

$$\begin{cases} \text{Binomial} \left(N = \lfloor \lambda(1 - \phi)^{-1} \rceil, \; p = 1 - \phi \right) & , \; \phi \in (0, 1) \\ \text{Poisson}(\lambda) & , \; \phi = 1 \\ \text{Negative Binomial} \left(r = \lambda(1 - \phi)^{-1}, \; p = (\phi - 1)\phi^{-1} \right) & , \; \phi > 1, \end{cases}$$

$$\tag{3.6}$$

where $\lfloor x \rceil$ is the closest integer to x. This model has $\text{E}(Y|\lambda) \approx \lambda$ and $\text{Var}(Y|\lambda) \approx \lambda\phi$ (see appendix D). Figure 4A illustrates the deviations of $P_{\lambda, \phi}$ with $\phi = 0.8$ and 1.2 from the Poisson distribution—$P_{\lambda, \phi}$ with $\phi = 1$.

2. Under model 3.2, (W_{1r}, W_{2r}) are bivariate lognormal. Here we generate them as follows: set $b > 0$ and (1) generate $(Z_{1r}, Z_{2r}) \sim N(\mu, \Sigma)$, (2) evaluate $W_{ir} = (1 + Z_{ir}b)^{1/b}$, $i = 1, 2$, and (3) repeat steps 1 and 2 until $W_{1r}, W_{2r} \geq 0$. When $b \to 0^+$, it is easy to show that (W_{1r}, W_{2r}) are bivariate lognormal. Figure 4B shows $f_{b, \mu, \Sigma}$ for $b = 0$ (lognormal) and $b = 1$, with μ, Σ chosen to match typical values in the V4 data analyzed in section 4.

Details of simulations and sensitivity results for the V4 data are in section 4.2.

3.2 Nonparametric Estimation. Let $S_i = \{S_{ir}\}_{r=1}^{n}$ be n independent spike trains of duration $[0, T]$ for neuron i. We divide $[0, T]$ into m bins of equal length, and we let Y_{irj} denote the spike count in trial r and bin j, so that $\sum_{j=1}^{m} Y_{irj} = Y_{ir}$ is the total spike count for neuron i on trial r. We assume
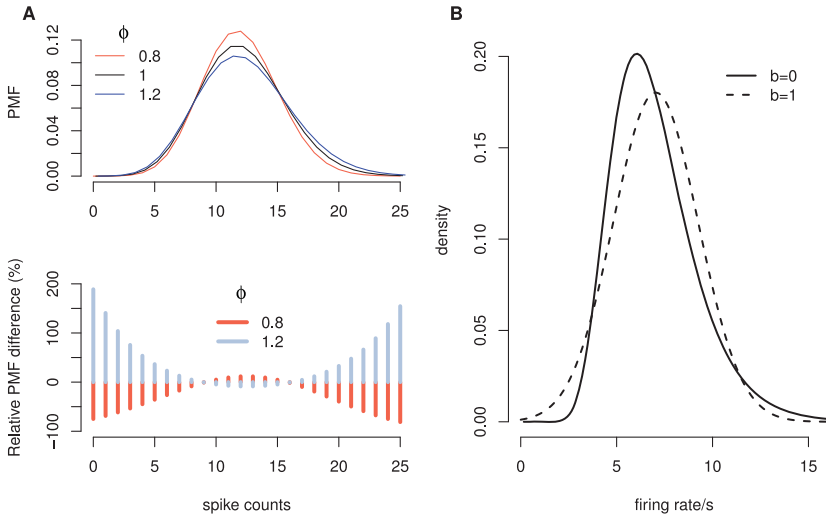
Figure 4: (A) Deviations of the spike counts from Poisson. Top: p.m.f. of model 3.6 for $\phi = 0.8, 1, 1.2$; $\phi = 1$ corresponds to the Poisson p.m.f. Bottom: relative % difference between model 3.6 for $\phi = 0.8$ and 1.2 and the Poisson p.m.f. (B) Deviations of the firing rates from lognormality.

$E[Y_{irj}|X_{1r}, X_{2r}, Y_{1r}, Y_{2r}] = p_{ij}Y_{ir}$, where p_{ij} is the probability that given that neuron i spikes, the spike occurs in bin j, independent of trial r, $\sum_j p_{ij} = 1$; in other words, the expected spike count in bin j in trial r is a portion p_{ij} of the total spike count Y_{ir} in trial r, with p_{ij} constant across trials. We let $0 \leq K < m - 1$ and define[1]

$$G_{m,K}(S_1, S_2) = \frac{\sum_{r=1}^{n} \sum_{|j-h|\leq K}(Y_{1rj} - p_{1j}Y_{1r})(Y_{2rh} - p_{2h}Y_{2r})}{n\left(1 - \sum_{|j-h|\leq K} p_{1j}p_{2h}\right)}. \tag{3.7}$$

Proposition 1. *If the spikes of neurons 1 and 2 given their rates are uncorrelated whenever they are separated by more than K bins, that is, they satisfy $Cov(Y_{irj}, Y_{krh}|X_{ir}, X_{kr}) = 0$ for $1 \leq i \leq k \leq 2$ whenever bins j and h are such that $K < |j - h| \leq m - 1$, then $G_{m,K}(S_1, S_2)$ and $G_{m,K}(S_i, S_i)$ are unbiased estimators of γ and $\phi_i E[X_{ir}]$, respectively.*

---

[1]A sufficient condition for the denominator of equation 3.7 to be strictly positive is $p_{ij} > 0$ for all $i = 1, 2$, since $K < m - 1$ implies $\sum_{|j-h|\leq K} p_{1j}p_{2h} < \sum_{j,h} p_{1j}p_{2h}$, where $\sum_{j,h} p_{1j}p_{2h} = 1$. The p_{ij} could be smoothed to achieve $p_{ij} > 0$ for low firing rate neurons.

The proof is in appendix E. The probabilities p_{ij} are typically unknown but can be estimated by the corresponding observed proportions $\widehat{p}_{ij} = \sum_{r=1}^{n} Y_{irj} / \sum_{r=1}^{n} Y_{ir}$, or smooth versions. In section 4 we estimate p_{ij} based on 50 ms bin widths peristimulus time histograms. The unbiased property depends on K: given the trial length T and number of bins m, we need K such that $K/m \geq \delta/T$, where δ is such that the spikes of the two neurons are correlated only if they are within δ. If the spikes are still correlated when they are farther than T apart, we cannot obtain unbiased estimates of γ and $\phi_i E[X_{ir}]$. Thus, proposition 1 yields unbiased estimates only if within-trial covariabilities occur on timescales δ smaller than the trial duration T, and m is sufficiently large, specifically, $m > T/(T - \delta)$ (see appendix E). Within these constraints, we use the smallest value of K, that is, $K/m \approx \delta/T$, because the variance of $G_{m,K}$ increases with K. For the data in section 4, Smith and Sommer (2013) found that based on jitter-corrected cross-correlograms, short time-scale correlation happened on a 5 to 10 ms half-width, so we use $\delta = 20$ ms to be conservative. The variance of the estimators decreases with the number of bins m because:

1. The trial length T is finite so the cross-terms of the extreme indices in the summand of equation 3.7 are not available, so that the ends of the spike trains do not contribute to estimating γ. Using small bins (m large) therefore means using more data, which reduces the variance of the estimators.
2. $(\delta/T) < 1$ and m large makes it possible for $K = \delta/(T/m)$ to be an integer. Otherwise we must round K up to the next integer so we avoid biasing the estimators, with the consequence that the cross-terms for $|j - k|$ large in equation 3.7 may contain less signal and more noise, and thereby increase the variance of the estimators.
3. Small bins better localize the beginnings and ends of within-trial effects, which reduces the variance of equation 3.7 for the same reason as in item 2.

In section 4 we used $m = 100$ bins, and thus $K = 2$. It would be statistically more efficient to use the largest m possible up to recording accuracy (in our case, this would be $m = 1000$, since $T = 1000$ ms and spike time accuracy is in ms), but the computational burden to perform the within-trial covariability jitter tests (see section 3.3) would be prohibitive (see appendix G).

3.2.1 Plug-In Estimators. We use proposition 1 to define plug-in estimators for the components of equations 2.1 and 2.2:

$$\widehat{\mathrm{ATT}}_{plug} = \prod_{i=1,2} \left(1 + \frac{G_{m,K}(S_i, S_i)}{\widehat{\mathrm{Var}}(Y_{ir}) - G_{m,K}(S_i, S_i)} \right)^{-1/2}, \tag{3.8}$$

$$\widehat{\Gamma}_{plug} = \frac{G_{m,K}(S_1, S_2)}{\sqrt{\widehat{\text{Var}}(Y_{1r})\widehat{\text{Var}}(Y_{2r})}}, \tag{3.9}$$

$$\widehat{\text{FRC}}_{plug} = \frac{\widehat{\text{SCC}}_{samp} - \widehat{\Gamma}_{plug}}{\widehat{\text{ATT}}_{plug}}, \tag{3.10}$$

$$\widehat{\phi}_{i\,plug} = \frac{G_{m,K}(S_i, S_i)}{\bar{Y}_i}, \tag{3.11}$$

where

$$\widehat{\text{Cov}}(Y_{1r}, Y_{2r}) = (n-1)^{-1} \sum_{r=1}^{n} (Y_{1r} - \bar{Y}_{1.})(Y_{2r} - \bar{Y}_{2.}), \quad \bar{Y}_{i.} = n^{-1} \sum_{r=1}^{n} Y_{ir},$$

$$\widehat{\text{Var}}(Y_{ir}) = \widehat{\text{Cov}}(Y_{ir}, Y_{ir}),$$

and

$$\widehat{\text{SCC}}_{samp} = \frac{\widehat{\text{Cov}}(Y_{1r}, Y_{2r})}{\sqrt{\widehat{\text{Var}}(Y_{1r})\widehat{\text{Var}}(Y_{2r})}} \tag{3.12}$$

are the sample estimates of the corresponding population quantities. When $\gamma = \Gamma = 0$, we set $\widehat{\Gamma}_{plug} = 0$ (see section 3.3). Note that the plug-in estimators do not always lie in the required domains (Churchland et al., 2011, encountered similar problems); for example, $\widehat{\text{FRC}}_{plug}$ could lie outside $[-1, 1]$. They are nevertheless useful estimates for exploratory data analysis.

We are able to estimate the components of equations 2.1 and 2.2 without specific assumptions on ϕ_i, in contrast to the assumption that $\phi_i = 1$ in section 3.1 and in Goris et al. (2014), and to the approach of Churchland et al. (2011), who use ϕ_i like a tuning parameter and choose it such that variances and correlations lie in the required domains. Here we derive a nonparametric estimate of ϕ_i, by using the information carried by spike counts at timescales smaller (i.e., spike counts in smaller bins) than the timescale of SCC and FRC.

3.3 Jitter Test for Within-Trial Correlation. In section 2 we presented a hierarchical model where spike counts (Y_{1r}, Y_{2r}) are trial-to-trial correlated through the rates (X_{1r}, X_{2r}), but also, conditionally on (X_{1r}, X_{2r}), they might share within-trial covariability, measured by γ or Γ. The case $\gamma = 0$ (equivalently $\Gamma = 0$) is of particular interest because then the attenuation of FRC is fully explained by the attenuation factor ATT, equation 2.2.

A parametric likelihood ratio test (LRT) of $H_0 : \gamma = 0$ based on model 3.2 is problematic because it is limited to the one-sided alternative $H_a : \gamma > 0$, whereas γ could be negative, and the asymptotic chi-square null distribution of the LRT does not hold because $\gamma = 0$ is at the boundary of the parameter space. A bootstrap would be needed to obtain the null distribution. The parametric estimate of γ also turned out to be less efficient than the nonparametric estimate, equation 3.7, in a simulation study, where we repeatedly generated synthetic data sets similar to ours and estimated γ parametrically and nonparametrically. Both estimators were unbiased, but the nonparametric estimator had smaller variance, which translates into more power in a testing context. The difference in efficiency happens because the two estimators are obtained from different reductions of the observed spike trains: the parametric estimate of γ is based on spike counts in bins of $T = 1000$ ms and the nonparametric estimate on spike counts in 10 ms bins. The spike trains contain more information about γ than spike counts (data processing inequality; Cover & Thomas, 2012), and presumably spike counts in smaller bins contain more information than spike counts in larger bins, since they are closer to the original data.

We therefore prefer a model-free jitter test that accommodates the two-sided alternative $\gamma \neq 0$, does not depend on parametric assumptions, and is likely to be more powerful. We use the nonparametric estimator of γ in section 3.2, $\widetilde{\gamma} = G_{m,K}$ ($m = 100$ and $K = 2$ are appropriate for the data in section 4), and we reject $H_0 : \gamma = 0$ if $\widetilde{\gamma} < F^{-1}(\alpha/2)$ or $\widetilde{\gamma} > F^{-1}(1 - \alpha/2)$, where F is the null distribution of $\widetilde{\gamma}$ obtained using a jitter method (Harrison et al., 2013), as follows. Let $(Y_{ir1}, \ldots, Y_{irm})$ denote the spike counts of neuron i in trial r and time bin $j = 1, \ldots, m$, such that $\sum_{j=1}^{m} Y_{irj} = Y_{ir}$ is the total spike count of neuron i in trial r. Let

$$(Y^*_{ir1}, \ldots, Y^*_{irm}) \mid Y_{ir} \sim \text{Multinomial}(Y_{ir}; p_{i1}, \ldots, p_{im}), \ r = 1, \ldots, n$$

be jittered spike counts for neuron i for each of the n trials, where p_{ij} is the probability that, given that neuron i spikes, the spike occurs in bin j (the same p_{ij}'s appear in section 3.1). Let $\widetilde{\gamma}^*$ be the nonparametric estimator of γ calculated from these jittered spike counts. We repeat this B times to obtain B values of $\widetilde{\gamma}^*$: their empirical distribution approximates the null distribution F of $\widetilde{\gamma}$. To obtain accurate p-values requires a simulation size B in the thousands. However, a normal probability plot revealed that the $\widetilde{\gamma}^*$ were approximately normally distributed (this is not surprising given that equation 3.7 is an average), so a faster approximation of F is the normal distribution with mean 0 and variance equal to the sample variance of B values of $\widetilde{\gamma}^*$, where B can be much smaller, say $B = 100$ (see Canty, Davison, Hinkley, & Ventura, 2006). The jitter test is a conditional parametric bootstrap test that is exact when the p_{ij}'s are known. In our application, we must estimate the p_{ij}'s (see section 3.2), so the test is no longer exact. Figure 5
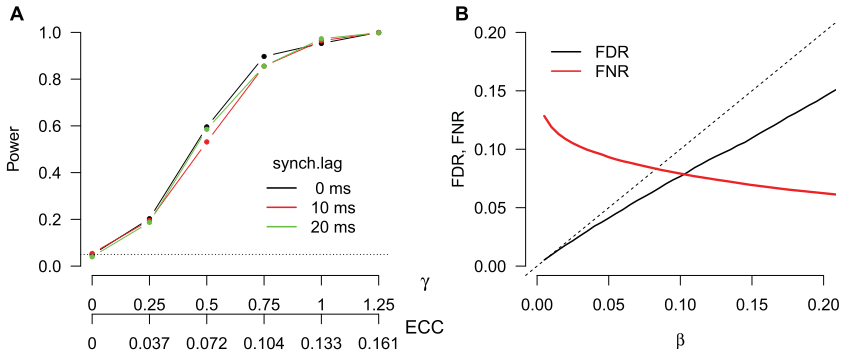
Figure 5: (A) Power of the jitter test to detect lagged synchrony within-trial effects. The test has full power when $\gamma > 1$, approximately. (B) FDR-FNR control: The actual FDR is below the nominal FDR β so the FDR procedure of Benjamini and Hochberg (1995) is reliable. The FNR is below 10% for all $\beta > 5\%$. We expect that the FDR testing procedure will fail to detect less than 10% of existing within-trial effects in data like the V4 data of section 4.

shows that its spurious detection rate is nevertheless close to the significance level α, so the test is reliable.

3.3.1 Statistical Power. The power of a test is the probability of detecting deviations from the null hypothesis when they exist. It is difficult to calculate analytically the power of the jitter test, so we estimated it by simulation. For values of $\gamma \in (0, 1.25)$, we simulated 1000 data sets of 60 trials of pairs of independent spike trains satisfying model 3.1 and induced within-trial covariability by adding an independent homogeneous Poisson spike train with rate γ to each spike train pairs, lagged by 0, 10, or 20 ms. The resulting pairs of spike trains satisfy model 3.2 with parameters $\mu_i = 1.9$, $\sigma_i = 0.31$, $\rho = 0.51$. These parameter values match the mean estimated values in the data in section 4. Then for each γ, we estimated the power as the proportion of times the jitter test rejected the hypothesis $\gamma = 0$ out of the 1000 data sets ($\delta = 20$ ms, $m = 100$, $K = 2$), and plotted this estimate in Figure 5A against γ and a standardized version of γ that takes value in $[-1, 1]$, the expected conditional spike count correlation:

$$\text{ECC} = \text{E}\left[\frac{\text{Cov}(Y_{1r}, Y_{2r} \mid X_{1r}, X_{2r})}{\prod_{i=1,2} \sqrt{\text{Var}(Y_{ir} \mid X_{ir})}} \right]. \qquad (3.13)$$

Appendix F provides nonparametric and parametric estimates of ECC. The power of the jitter test increases with the amount of within-trial covariability γ and has very high power when $\gamma > 0.75$, ECC > 0.1.

3.3.2 False Discovery Rate Control. When many tests are performed, it is useful to cap the expected proportion of false rejections among all rejections—the false discovery rate (FDR)—below some desired β. Benjamini and Hochberg (1995) proceed as follows: given M independent tests with null hypotheses $H_{01}, \ldots, H_{0M}$ and corresponding p-values $P_1, \ldots, P_M$, they reject $H_{0(1)}, H_{0(2)}, \ldots, H_{0(K)}$ where $K = \max\{i$ such that $P_{(i)} \leq i\beta/M\}$ and $P_{(j)}$ is the jth largest p-value. Then, in expectation, fewer than a proportion β of the K rejected tests correspond to true null hypotheses. Benjamini and Yekutieli (2001) show that this procedure also controls the FDR in the case of positive regression-dependent tests, but advocate a more conservative FDR control for tests with general dependencies, which consists of replacing β with $\beta / \sum_{i=1}^{M} i^{-1}$.

The jitter tests are dependent since they are applied to all neuron pairs, but they are not necessarily positive regression dependent, so we investigated the performances of both procedures for data like ours. We simulated $n = 60$ independent trials of spike trains from an N-variate Poisson-lognormal model with $N = 30$ (we extended model 3.1 from $N = 2$ to 30), where μ and Σ were assigned values that match the V4 data in section 4, such that half of the neuron pairs had FRC below 0 and half had FRC above 0. Neurons were randomly assigned to three groups of size $N/3 = 10$. For each trial r, the spike trains of all neurons in a group were contaminated by the same noise spike train with rate γ, and the noise spike trains were independent between groups. We used $\gamma = 0.5, 1$, and 2 for the three groups, respectively, to match approximately the values calculated in the data. The proportion of tests with H_0 false was therefore one-third. We performed the $\binom{N}{2} = 435$ jitter tests for all neuron pairs and recorded the observed FDR and false nondiscovery rate (FNR), that is, the expected proportion of false nulls among the nonrejections, for a range of nominal FDR β. We repeated the simulation 1000 times to obtain accurate mean estimates of the observed FDR and FNR. In Figure 5B, we show that the Benjamini and Hochberg (1995) procedure provides reliable FDR control for data like ours (the observed FDR is below β), perhaps because the tests were only weakly dependent, as suggested by the Brownian covariance test of Székely and Rizzo (2009). We also observe FNR $\leq 10\%$ whenever $\beta \geq 5\%$, which means that most cases with $\Gamma \neq 0$ are detected by the jitter test with FDR control. The Benjamini and Yekutieli (2001) procedure for tests with general dependencies provided too conservative a control (smaller FDR and larger FNR; results not shown), so we did not use it.

4 Analysis of V4 Data

We analyze the rhesus macaque monkey (*Macaca mulatta*) visual V4 data collected by Smith and Sommer (2013) during the following task (see Figure 6). The animal fixated a small blue dot in the center of a computer

A

Fixation
(100 ms)   Stimulus
(1000 ms)   Target
(300 ms)   Saccade

Time

B

Stimulus Set (12 conditions)

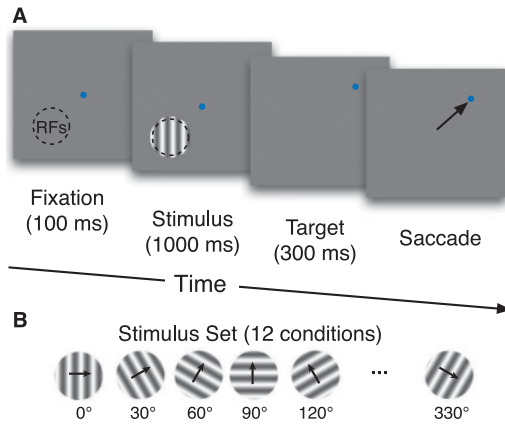0°   30°   60°   90°   120°   330°

Figure 6: (A) The experimental task. The dashed circle indicates the aggregate receptive field (RF) of all the V4 neurons recorded from the 100-electrode array. After the animal fixated on the central blue dot for 100 ms, a drifting grating stimulus (one of 12 directions, block randomized) was shown for 1000 ms. The central dot and grating were then extinguished, and a blue target dot appeared at a random location positioned along a circle 8 degrees from fixation. The animal was rewarded for making a successful saccade to this target dot within 300 ms. (B) The experimental stimulus consisted of a full-contrast drifting sinusoidal grating with 12 equally spaced directions of drift (30 degrees). The spatial frequency (1.3 cpd) and temporal frequency (6.25 Hz) were fixed, and the grating was positioned and sized to cover the aggregate RF of the neurons.

screen while a sinusoidal grating covered the aggregate receptive field area of the recorded neurons. The grating's size and spatial and temporal frequencies were fixed, and the direction of drift was one of 12 equally spaced directions for each trial; $n = 61$ independent trials of duration $T = 1000$ ms were recorded for each of the 12 stimuli (directions). The neural data were acquired from a Utah array (100 electrodes, Blackrock Microsystems) implanted in V4. The electrode voltages were amplified and bandpass filtered (250–7.5 kHz) prior to spike sorting with custom software (Kelly et al., 2007). All 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.

For our statistical analysis, we retained only the 53 well-isolated single units with minimum signal-to-noise ratio 2.75 (Kelly et al., 2007) and 1 Hz minimum trial-averaged firing rate, and studied the relationships between all $\binom{53}{2} \times 12 = 16{,}536$ neuron pairs-stimuli combinations. We proceeded as outlined in Figure 2. We first applied the jitter test (see section 3.3) with a false discovery rate (FDR) of 10% to all neuron pairs-stimuli combinations
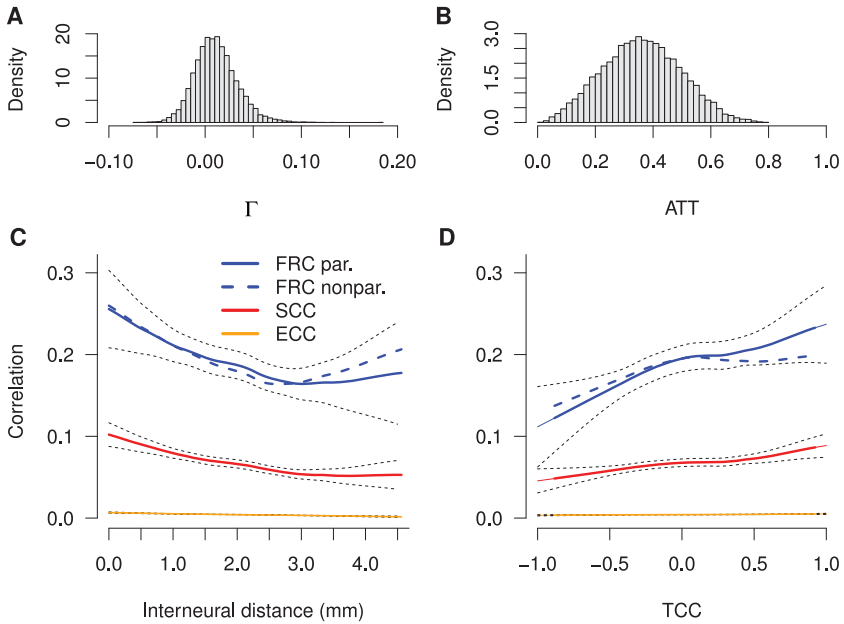
Figure 7: (A) The estimates of $\widehat{\Gamma}_{plug}$, equation 3.9, concentrate around zero: within-trial correlations have a weak effect on SCC. (B) The estimates of $\widehat{ATT}_{plug}$, equation 3.8, concentrate around 0.35: the magnitude of SCC is about 35% that of FRC on average. (CD) SCC, FRC, ECC versus interneural distance and TCC obtained by local linear regressions of the sample estimates of SCC and ECC (see equations 3.12, 3.13, and F.1), and parametric and nonparametric estimates of FRC (see equations 3.3 and 3.10); solid and dashed curves, respectively), along with 95% simultaneous confidence bands. All correlations decrease with interneural distance and increase with TCC. The SCC and FRC curves are similar in shape, but SCC is attenuated compared to FRC. ECC is close to zero, which suggests that within-trial correlations have a small impact on spike counts in 1 s bins.

to determine which pairs have $\Gamma \neq 0$. Only a minority of the V4 neuron pairs exhibited significant within-trial covariance (6.04% using a two-sided test, and 10.07% using a one-sided test with alternative $\gamma > 0$), which suggests that within-trial correlations have a weak impact on SCC, and therefore that the attenuation effect is mainly characterized by ATT in equation 2.2. Figure 7B shows the distribution of the nonparametric estimates $\widehat{ATT}_{plug}$ (see equation 3.8); it is concentrated around 0.35, which means that the magnitude of observed SCCs is on average only 35% of the magnitude of FRC.
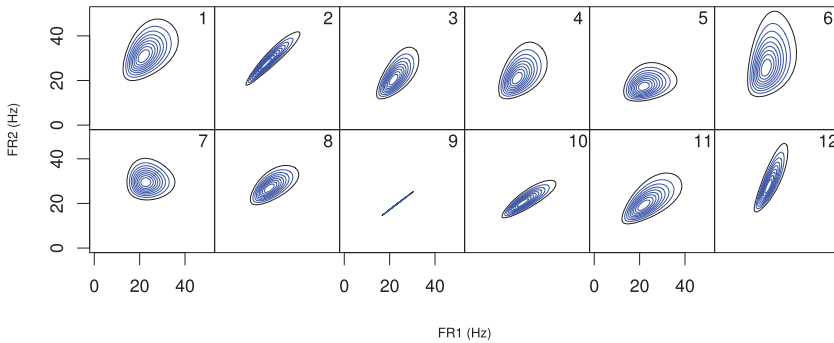
Figure 8: Joint generalized tuning curve (GTC; see appendix H). Tolerance regions (10% to 90% levels) of the estimated joint bivariate lognormal distributions for the rates (X_{1r}, X_{2r}) in model 3.2 for an example pair of neurons at 12 stimuli. The stimulus number is written in the top-right corner of each panel. This plot shows not only how the neurons' average firing rates vary with the stimulus, but also how differently the two rates are correlated for different stimuli.

Next we fitted the parametric models in section 3.1 consistently with the jitter test results (that is, we fitted the DC-PLN model, equation 3.2, if $\Gamma > 0$, and the PLN model, equation 3.1 otherwise) using the maximum likelihood algorithm 2 in appendix C, and obtained estimates of the components of equation 2.2. A component of the model fitted to a neuron pair is shown in Figure 8; goodness-of-fit diagnostics and sensitivity to parametric assumptions are summarized in sections 4.1 and 4.2. Figure 3B shows the sample estimates of SCC in equation 3.12 versus the model-based estimates of FRC in equation 3.3 for all 16,536 neuron pairs-stimuli combinations. The SCC is substantially attenuated compared to the FRC, which is consistent with the nonparametric estimates of ATT being well below one in Figure 7B. Figures 7C and 7D further show that on average, SCC and FRC decrease with pairwise interneural distance (distance between electrodes) and increase with tuning curve correlation (TCC; see equation H.1), which is a measure of similarity between the neural activity of two neurons. Smith and Sommer (2013) had noted these relationships for SCC but not FRC. The sample estimate of ECC, equation F.1, also decreases with interneural distance and increases with TCC, although it is hard to see by the naked eye because ECC is close to zero.

In appendix H, we also display SCC, FRC, and ECC against the generalized tuning curve correlation (GTCC), a combined measure of tuning curve similarity and firing rate covariability between a neuron pair, which is motivated in Figure 8. This figure shows the estimated joint generalized tuning curve (GTC) of a neuron pair, that is, the estimated joint distribution of their firing rates for all stimuli, and shows not only how the neurons'
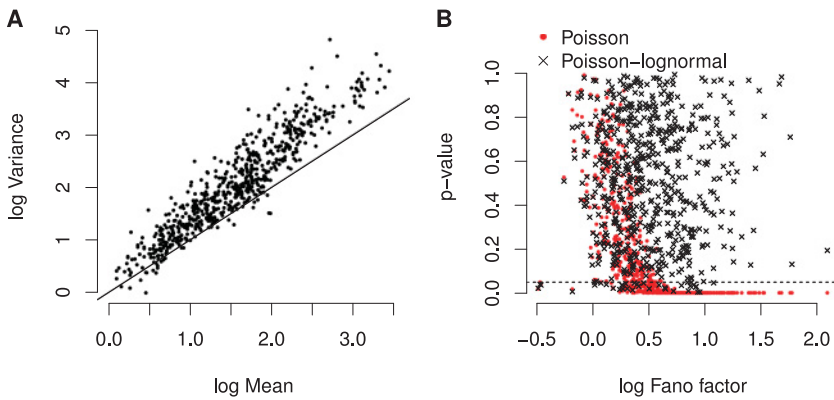
Figure 9: (A) Spike count log variance versus log mean. The spike counts are overdispersed. (B) Kullback-Leibler divergence test p-values. Using $\alpha = 5\%$, the Poisson model is rejected 52% of the times, and the Poisson–lognormal model 5% of the times, which matches the chosen spurious detection rate: the second model provides a better fit to the spike counts.

average firing rates vary jointly with the stimulus—the TCC provides a one number summary of that covariability across stimuli—but also how differently these rates are correlated across trials for different stimuli. For the neuron pair shown in Figure 8, these rates are virtually uncorrelated when stimulus 7 is presented and very strongly positively correlated for stimuli 2, 9, and 12. The GTCC is a one-number summary of both tuning curve similarity and firing rate covariability between a neuron pair.

4.1 Parametric Model Goodness of Fit. Figures 3B, 7C, and 7D display the parametric estimates of FRC. We justify here that the parametric models fit the data.

Figure 9A shows the sample variance versus the sample mean of the spike counts of the 53 neurons for all 12 stimuli: the spike counts are overdispersed, which is a feature of the Poisson-lognormal model. The KL divergence test (see section 3.1) confirms that the univariate Poisson-lognormal model (i.e., the univariate version of equation 3.1) fits single neuron spike counts well; only 5% of the tests were rejected, which matches the spurious detection rate α. To provide a contrasting result, we also performed KL divergence tests assuming Poisson spike counts. Their p-values are plotted in Figure 9B against the spike counts' Fano factors: 52% of the Poisson tests were rejected, including all the tests applied to data with $\log_e$ Fano factors above 0.6. The p-values of the univariate Poisson-lognormal KL divergence tests are also plotted and exhibit no relationship with the Fano factor.

Now we assess the fits of the PLN and DC-PLN models in equations 3.1 and 3.2 to pairs of spike counts. We randomly selected 2500 cases ($\approx$ 15% of the 16,536 neuron pairs-stimuli combinations) to which we applied the KL and SCC divergence tests (see section 3.1). Only about 4% of the neuron pairs did not pass the SCC divergence test, and less than 1.5% did not pass the KL test, at the 5% significance level, which suggests that the bivariate parametric models fit pairs of spike counts well.

4.2 Sensitivity of FRC Estimates to Model Assumptions. Figures 3B, 7C, and 7D display the parametric estimates of FRC where in about 90% of cases, we imposed $\gamma = \Gamma = 0$ according to the results of the jitter test. This test has a low false nondiscovery rate (FNR, Figure 5B) and high power to detect deviations from $\gamma = 0$ when $\gamma > 0.75$ or ECC > 0.1 (see Figure 5A) so we are confident that we fitted the appropriate model, the PLN model, in such cases. However, we would like to know how the parametric estimate of FRC is affected by assuming $\gamma = 0$ when the jitter test fails to detect that γ is small and positive, and how sensitive it is to mild deviations from the Poisson and lognormal assumptions.

We applied the methods in section 3.1.2. We generated $n = 60$ trials of duration $T = 1000$ ms of spike count pairs (Y_{1r}, Y_{2r}) from the generalized bivariate spike count model, equation 3.5, for (ϕ, b, γ) in $(0.8, 1.2) \times (0, 1) \times (0, 1)$, where $\mu(\phi, b, \gamma)$ and $\Sigma(\phi, b, \gamma)$ were calibrated for each (ϕ, b, γ) so that spike counts had means 7, variances 12, and FRC $= 0.5$, to match typical values in the V4 data. We used these data to estimate FRC under model 3.1, and repeated the simulation 10,000 times. Figure 10A shows the mean of the repeat values of $\widehat{\text{FRC}}_{mle}$ and Figure 10B the standard deviation of $\widehat{\text{FRC}}_{mle}$ relative to what it should be if the spike counts were generated from the estimation model, equation 3.1, as functions of the generating model parameters, ϕ, b, and γ (the estimating model has $\phi = 1$, $b = 0^+$, and $\gamma = 0$). The deviations from lognormality ($b > 0$) have practically no effect on $\widehat{\text{FRC}}_{mle}$, and the deviations from $\phi = 1$ moderate effects, that is, $\widehat{\text{FRC}}_{mle}$ is fairly robust to mild deviations from the Poisson-lognormal model. However, the deviations from $\gamma = 0$ increase the bias of $\widehat{\text{FRC}}_{mle}$ substantially and reduce its variance,[2] so inferences are not reliable (for example, a confidence interval for the true FRC would be centered on the wrong value and have lower coverage than expected). The possibility of this happening in our data is small because the FNR is less than 10% (see Figure 5B).

---

[2]The FRC is overestimated likely because the positive within-trial covariance ($\gamma > 0$) causes an increase in SCC (see equation 2.2), which is erroneously attributed to the firing rate's covariance by the estimation model, equation 3.1, which assumes $\text{Cov}(Y_{1r}, Y_{2r}) \equiv \text{Cov}(X_{1r}, X_{2r})$.
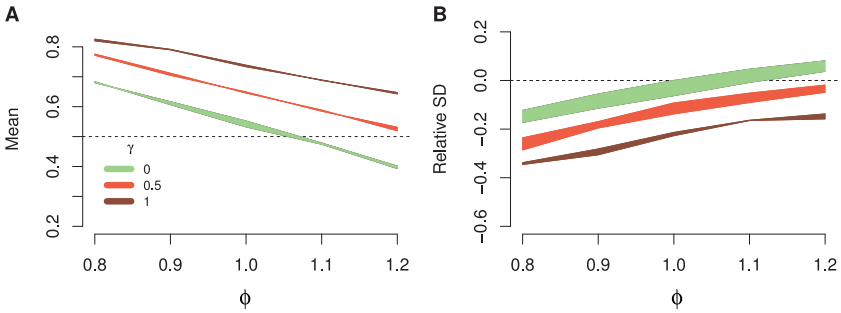
Figure 10: Sensitivity of $\widehat{\text{FRC}}_{mle}$ to deviations of the spike counts from the PLN model, equation 3.1, that is, to deviations from $\phi = 1$ (horizontal axis), $\gamma = 0$ (different colors), and $b = 0$ (thickness of the curves) in model 3.5, in terms of mean and relative standard deviation (SD). The true value of FRC is 0.5. The mean and SD of $\widehat{\text{FRC}}_{mle}$ are not sensitive to changes in $b \in (0, 1)$, somewhat sensitive to changes in dispersion of the spike counts $\phi \in (0.8, 1.2)$, and very sensitive to changes in within-trial covariability γ.

5 Discussion

We have provided a collection of methods for eliminating Poisson-like noise from the SCC to produce the FRC. As shown in Figure 3B, FRC tends to be much larger in magnitude than SCC and should be much more sensitive to experimental manipulations. A relatively simple bivariate hierarchical model may be introduced when the within-trial correlation is sufficiently small that its impact on FRC is negligible, and we proposed a jitter test of this assumption. We showed that the majority of pairs of V4 neurons failed to reject the null hypothesis of no within-trial correlation. Careful analysis requires an evaluation of the power of this test and the sensitivity of the proposed parametric model to mild departures from assumptions, which we supplied, along with an evaluation of the expected false nondiscovery rate (FNR) of the test, which we determined to be low. Together, these analyses strongly suggest that in our data, within-trial correlation rarely induces distortions in the estimate of FRC when the jitter test does not reject the null hypothesis. For pairs of neurons where the hypothesis of no within-trial correlation was rejected, we used a nonparametric estimate of within-trial correlation and then applied a correlated Poisson hierarchical model that allows for positive within-trial correlation. A more complete accounting of within-trial variation, which would include negative within-trial correlation, could be obtained from point process models (Snyder & Miller, 1991; Kass et al., 2014) and is the subject of future research.

Our finding that only a few V4 neuron pairs exhibited within-trial correlations differs from the result of Goris et al. (2014) in the macaque visual area

V1. First, their SCCs are larger than in our V4 data; a similar discrepancy is also documented by the V1 analysis of Smith and Kohn (2008) and the V4 studies of Smith and Sommer (2013), Cohen and Maunsell (2009), and Mitchell et al. (2009). Second, the attenuation effect in Goris et al. (2014) is much weaker than ours, so that SCC is only weakly attenuated compared to FRC (in that paper, FRC is called "gain correlation"), while their point process covariability is much higher (ours is measured by Γ or ECC). These differences may be due to the different area under study. The different anesthetics used in the two experiments (our animal was alert) might also have induced different levels of up and down states (Steriade & Buzsaki, 1990; Cowan & Wilson, 1994; Timofeev et al., 2001; Steriade et al., 2001), which in turn could have affected within-trial covariabilities.

Our overall goal was to provide a useful suite of methods for examining the trial-to-trial correlated behavior of neurons at trial-length timescales. To the extent that neural implementations of maximum likelihood and Bayesian inference are possible (e.g., Ma, Beck, Latham, & Pouget, 2006), downstream neural networks should be able to strip away noise from spike counts and use FRC to advantage, just as we have been able to do statistically. Whether they do so remains an open question.

Appendix A: Spike Count Correlation Formula

Using the laws of total expectation and total variance, equations 1.1 and 1.2 imply

$$E[Y_{ir}] = E\{E[Y_{ir}|X_{ir}]\} = E[X_{ir}]$$

and

$$Var(Y_{ir}) = E\{Var(Y_{ir}|X_{ir})\} + Var(E[Y_{ir}|X_{ir}]) = \phi_i E[X_{ir}] + Var(X_{ir}).$$

Thus,

$$Cor(Y_{1r}, Y_{2r}) = \frac{\gamma + Cov(X_{1r}, X_{2r})}{\prod_{i=1,2} \sqrt{Var(Y_{ir})}} \text{ with } \gamma = E[Cov(Y_{1r}, Y_{2r}|X_{1r}, X_{2r})]$$

$$= \frac{Cov(X_{1r}, X_{2r})}{\prod_{i=1,2} \sqrt{(\phi_i E[X_{ir}] + Var(X_{ir}))}} + \frac{\gamma}{\prod_{i=1,2} \sqrt{Var(Y_{ir})}}$$

$$= \frac{Cov(X_{1r}, X_{2r})}{\prod_{i=1,2} \sqrt{Var(X_{ir})(\phi_i/F_i + 1)}} + \Gamma$$

$$= Cor(X_{1r}, X_{2r}) \prod_{i=1,2} (1 + \phi_i/F_i)^{-1/2} + \Gamma.$$

Appendix B: Lognormal Distribution

Let $(W_{1r}, W_{2r}) = (\log X_{1r}, \log X_{2r}) \sim N(\mu, \Sigma)$. Then:

$$
\begin{aligned}
\mathrm{E}[X_{ir}] &= \mathrm{E}[e^{W_{ir}}] = e^{\mu_i + \sigma_i^2/2} \\
\mathrm{Var}(X_{ir}) &= \mathrm{E}[e^{2W_{ir}}] - (\mathrm{E}[e^{W_{ir}}])^2 \\
&= e^{2\mu_i + 2\sigma_i^2} - e^{2\mu_i + \sigma_i^2} \\
&= (e^{\mu_i + \sigma_i^2/2})^2 (e^{\sigma_i^2} - 1), \\
\mathrm{Cov}(X_{1r}, X_{2r}) &= \mathrm{E}[e^{W_{1r} + W_{2r}}] - \mathrm{E}[e^{W_{1r}}]\mathrm{E}[e^{W_{2r}}] \\
&= e^{\mu_1 + \mu_2 + (\sigma_1^2 + \sigma_2^2)/2 + \sigma_{12}} - e^{\mu_1 + \mu_2 + (\sigma_1^2 + \sigma_2^2)/2} \\
&= e^{\mu_1 + \sigma_1^2/2} e^{\mu_2 + \sigma_2^2/2} (e^{\sigma_{12}} - 1)
\end{aligned}
$$

such that $F_i = e^{\mu_i + \sigma_i^2/2}(e^{\sigma_i^2} - 1)$ and $\mathrm{Cor}(X_{1r}, X_{2r}) = (e^{\rho \sigma_1 \sigma_2} - 1) \prod_{j=1,2}(e^{\sigma_j^2} - 1)^{-1/2}$.

Appendix C: Doubly Correlated Poisson-Lognormal Distribution

Let $p_{\lambda_i}(y)$ be the p.m.f. of Poisson(λ_i). Let $R_{ir} \sim$ Poisson(λ_i), $i = 0, 1, 2$, be independent, and define $Y_{ir} = R_{0r} + R_{ir}$, $i = 1, 2$. The joint distribution of (Y_{1r}, Y_{2r}) is

$$
\begin{aligned}
\mathrm{P}(Y_{1r} = y_1, Y_{2r} = y_2) &= \mathrm{P}(R_{0r} + R_{1r} = y_1, R_{0r} + R_{2r} = y_2) \\
&= \mathrm{P}(R_{1r} = y_1 - R_{0r}, R_{2r} = y_2 - R_{0r}) \\
&= \sum_{z=0}^{\min\{y_1, y_2\}} \mathrm{P}(R_{1r} = y_1 - R_{0r}, R_{2r} = y_2 - R_{0r}, R_{0r} = z) \\
&= \sum_{z=0}^{\min\{y_1, y_2\}} p_{\lambda_1}(y_1 - z) p_{\lambda_2}(y_2 - z) p_{\lambda_0}(z).
\end{aligned}
$$

By integrating with respect to $(\log \lambda_1, \log \lambda_2)^T \sim N(\mu, \Sigma)$ we obtain the p.m.f. of the DC-PLN model, equation 3.2:

$$
p_{\mu, \Sigma, \gamma}(y_1, y_2) = \sum_{z=0}^{\min\{y_1, y_2\}} q_{\mu, \Sigma}(y_1 - z, y_2 - z) p_{\lambda_0}(z) \tag{C.1}
$$

Algorithm 1: *ML Estimator of DC-PLN (Spike Counts).*

---

1. Obtain the profile MOM estimates of μ, Σ, given γ:

$$\widehat{\sigma}_i^2(\gamma) = \log\left(1 + \frac{S_{Y_i}^2 - \bar{Y}_i}{(\bar{Y}_i - \gamma)^2}\right); \widehat{\mu}_i(\gamma) = \log\left(\bar{Y}_i - \gamma\right) - \frac{\widehat{\sigma}_i^2}{2}; \widehat{\sigma}_{12}(\gamma) = \log\left(1 + \frac{S_{Y_1, Y_2} - \gamma}{(\bar{Y}_1 - \gamma)(\bar{Y}_2 - \gamma)}\right).$$

2. Find $\widetilde{\gamma} = \arg\max_{\gamma} \; L\left(\widehat{\mu}(\gamma), \widehat{\Sigma}(\gamma), \gamma; y_{11}, ..., y_{2n}\right).$

3. Find $(\mu^*, \Sigma^*) = \arg\max_{\mu, \Sigma} \; L(\mu, \Sigma, \widetilde{\gamma}; y_{11}, ..., y_{2n}).$

4. Update $\widetilde{\gamma} = \arg\max_{\gamma} \; L(\mu^*, \Sigma^*, \gamma; y_{11}, ..., y_{2n}).$

5. Iterate steps 3–4.

Algorithm 2: *ML Estimator of DC-PLN (Spike Trains).*

---

1. Estimate $\widetilde{\gamma}$ nonparametrically using equation 3.7.

2. Find $(\mu^*, \Sigma^*) = \arg\max_{\mu, \Sigma} \; L(\mu, \Sigma, \widetilde{\gamma}; y_{11}, ..., y_{2n}).$

---

where

$$q_{\mu, \Sigma}(y_1, y_2) = \int_{\mathbb{R}^2} p_{e^{x_1}}(y_1) p_{e^{x_2}}(y_2) g_{\mu, \Sigma}(x_1, x_2) dx_1 dx_2 \qquad (C.2)$$

is the p.m.f. of the PLN model, equation 3.1, and $g_{\mu, \Sigma}$ is the gaussian p.d.f. We can compute the ML estimators of μ, Σ, γ numerically using algorithm 1, when only the spike counts are available, or algorithm 2, when the spike trains are available. In data simulated as in section 4.2, we found that algorithm 2 typically yielded lower-variance estimates of FRC. We therefore recommend algorithm 2, keeping in mind that algorithm 1 could provide better estimates of other quantities. Note that if we assume $\gamma = 0$, that is, the PLN model, then we just use step 2 of algorithm 2 with $\gamma = 0$.

Appendix D: Generalized Count Model

Let $Y|\lambda$ have the distribution in equation 3.6. Thus,

- If $\phi \in (0, 1)$, then $E[Y|\lambda] = \lfloor\frac{\lambda}{1-\phi}\rfloor(1 - \phi) = (\frac{\lambda}{1-\phi} + \epsilon)(1 - \phi) = \lambda + \epsilon(1 - \phi) \approx \lambda$, and $Var(Y|\lambda) = E[Y|\lambda]\phi = \lambda\phi + \epsilon(1 - \phi)\phi \approx \lambda\phi$, where $\epsilon \in (-0.5, 0.5)$, such that $Var(Y|\lambda)/E[Y|\lambda] \approx \phi$.
- If $\phi = 1$, then $E[Y|\lambda] = \lambda$ and $Var(Y|\lambda) = \lambda$.

- If $\phi > 1$, then $\mathrm{E}[Y|\lambda] = rp(1-p)^{-1} = \lambda$ and $\mathrm{Var}(Y|\lambda) = rp(1-p)^{-2} = \lambda\phi$.

Appendix E: Nonparametric Estimators of γ and ϕ_i ——————

Proof of Proposition 1. Let $X_r = (X_{1r}, X_{2r})$ and $Y_r = (Y_{1r}, Y_{2r})$. Exploiting the covariance decomposition as in equation 1.3 and the fact that $\mathrm{E}[Y_{irj}|X_r, Y_r] = p_{ij}Y_{ir}$, we get

$$\mathrm{Cov}(Y_{1rj}, Y_{2rh}|X_r)$$
$$= \mathrm{E}[\mathrm{Cov}(Y_{1rj}, Y_{2rh}|X_r, Y_r)|X_r] + \mathrm{Cov}(\mathrm{E}[Y_{1rj}|X_r, Y_r], \mathrm{E}[Y_{2rh}|X_r, Y_r]|X_r)$$
$$= \mathrm{E}\{\mathrm{E}[(Y_{1rj} - p_{1j}Y_{1r})(Y_{2rh} - p_{2h}Y_{2r})|X_r, Y_r] \mid X_r\} + \mathrm{Cov}(p_{1j}Y_{1r}, p_{2h}Y_{2r}|X_r)$$
$$= \mathrm{E}[(Y_{1rj} - p_{1j}Y_{1r})(Y_{2rh} - p_{2h}Y_{2r})|X_r] + p_{1j}p_{2h}\mathrm{Cov}(Y_{1r}, Y_{2r}|X_r),$$

which implies

$$\mathrm{E}[(Y_{1rj} - p_{1j}Y_{1r})(Y_{2rh} - p_{2h}Y_{2r})|X_r]$$
$$= \mathrm{Cov}(Y_{1rj}, Y_{2rh}|X_r) - p_{1j}p_{2h}\mathrm{Cov}(Y_{1r}, Y_{2r}|X_r). \tag{E.1}$$

Moreover, zero conditional covariance across time bins j, h s.t. $|j - h| > K$ implies

$$\mathrm{Cov}(Y_{1r}, Y_{2r}|X_r) = \sum_{|j-h|\leq K} \mathrm{Cov}(Y_{1rj}, Y_{2rh}|X_r). \tag{E.2}$$

Therefore, combining equations E.1 and E.2, we get

$$\sum_{|j-h|\leq K} \mathrm{E}[(Y_{1rj} - p_{1j}Y_{1r})(Y_{2rh} - p_{2h}Y_{2r})|X_r]$$

$$= \mathrm{Cov}(Y_{1r}, Y_{2r}|X_r)\left(1 - \sum_{|j-h|\leq K} p_{1j}p_{2h}\right),$$

implying that $\mathrm{E}[G_{m,K}(S_1, S_2)] = \mathrm{E}[\mathrm{Cov}(Y_{1r}, Y_{2r}|X_{1r}, X_{2r})]$.

The proof that $G_{m,K}(S_i, S_i)$ is an unbiased estimator of $\phi_i\mathrm{E}[X_{ir}]$ is similar; all we need is to replace (Y_{1r}, Y_{2r}) and (X_{1r}, X_{2r}) by (Y_{ir}, Y_{ir}) and (X_{ir}, X_{ir}).

Table 1: Time (ms) to Compute $G_{m,K}$ (equation 3.7) for a Neuron Pair for Different Values of n and m ($K = \lceil m\delta/T \rceil$).

	δ/T	m	Number of Trials (n)		
			60	80	100
$G_{m,K}$	0.01	50	0.3 (207)	0.3 (275)	0.4 (346)
		100	0.8 (322)	0.9 (426)	1.1 (526)
		200	2.1 (603)	2.5 (867)	3.4 (976)
	0.02	50	0.6 (235)	0.6 (289)	0.8 (357)
		100	1.2 (354)	1.3 (492)	1.8 (581)
		200	3.3 (744)	4.3 (914)	5.5 (1202)

Note: The implementation time of the jitter test based on $G_{m,K}$ (section 3.3) with 100 bootstraps and 20-bin psth is indicated in parentheses.

Note that combining the conditions $K < m - 1$ and $K/m \geq \delta/T$ of section 3.2 yields $\frac{\delta}{T} \leq \frac{K}{m} < \frac{m-1}{m} < 1$, which implies $\frac{\delta}{T} < 1$ and $\frac{m-1}{m} > \frac{\delta}{T} \Rightarrow m > T/(T - \delta)$. These conditions are necessary to obtain unbiased estimators.

Appendix F: Expected Conditional Spike-Count Correlation

The expected conditional spike-count correlation (ECC) in equation 3.13 can be estimated nonparametrically analogous to equation 3.7:

$$\widehat{ECC}_{samp} = \frac{1}{|\mathcal{I}|} \sum_{r \in \mathcal{I}} \frac{\sum_{|j-h| \leq K} (Y_{1rj} - \widehat{p}_{1j} Y_{1r})(Y_{2rh} - \widehat{p}_{2h} Y_{2r})/\widehat{Q}_{12}(K)}{\prod_{i=1,2} \sqrt{\sum_{|j-h| \leq K} (Y_{irj} - \widehat{p}_{ij} Y_{ir})(Y_{irh} - \widehat{p}_{ih} Y_{ir})/\widehat{Q}_{ii}(K)}},$$

(F.1)

where $Q_{il}(K) = 1 - \sum_{|j-h| \leq K} p_{ij} p_{lh}$, and $\widehat{p}_{ij}$'s are as discussed in section 3.2, and $\mathcal{I} = \{r \in \{1, \ldots, n\} : Y_{1r}, Y_{2r} > 0\}$, that is, $\mathcal{I}$ is the set of trials where both neurons spiked at least once. Notice that if $\text{Cov}(Y_{1r}, Y_{2r}|X_{1r}, X_{2r})$ is constant across trials, then ECC $= \gamma \times \text{E}[1/\prod_{i=1,2} \sqrt{\text{Var}(Y_{ir}|X_{ir})}]$, so that ECC $= 0 \Leftrightarrow \gamma = 0$. Otherwise the equivalence is not necessarily true. For instance, under model 3.2, we have ECC $= \gamma \times \text{E}[1/\sqrt{(\gamma + W_{1r})(\gamma + W_{2r})}]$.

Appendix G: Computational Time

We implemented our estimation and testing procedures in R, by using personal codes and modifications of the package poilog to estimate the DC-PLN model, equation 3.2. We used a laptop with an Intel Core™ 2 Duo CPU T6600 (2 × 2.20 GHz), RAM 4GB (DDR 800 MHz); and OS Ubuntu 15.04. Table 1 displays the time to compute $G_{m,K}$ (see equation 3.7) and perform

Table 2: Computational Time (ms) to Estimate the Parameters of the DC-PLN Parametric Model, Equation 3.2, for a Neuron Pair.

	Number of Trials (n)		
	60	80	100
Estimate μ, Σ assuming $\gamma = 0$	240	293	383
Estimate μ, Σ, γ (algorithm 1)	4516	6364	7494
Estimate μ, Σ, γ (algorithm 2, step 2)	2257	2769	3721

Note: Time for step 1 of algorithm 2 can be found in table 1.

the jitter test on a neuron pair, and Table 2 to estimate the parametric model. The times required to process a data set with N neurons and S stimuli can be obtained by multiplying the values in the tables by $\binom{N}{2} \times S$. For example, in section 4, the multiple jitter tests required about $\binom{53}{2} \times 12 \times 354$ ms ≈ 97 minutes.

G.1 Choosing m of $G_{m,K}$. In section 3.1 we argued that the number of bins m should be large. However, the computational burden may be too large when we perform the jitter test (see table 1): the number of elements in the summand in equation 3.7 is $n \times M$, where $M = \text{card}(\{j, h : 1 < j, h \leq m, |j - h| \leq K\}) = m(2K + 1) - K(K + 1)$, so that for a fixed ratio $\pi = K/m = \delta/T$, the computational time of equation 3.7 is proportional to $n \times M = O(n \times m^2)$. To set an upper bound D on M given π, we use $m^*(\pi, D) = \frac{1-\pi+\sqrt{(1-\pi)^2+4(2\pi-\pi^2)D}}{2(2\pi-\pi^2)}$ and $K = \lceil \pi m^* \rceil$.

Appendix H: Generalized Tuning Curve Correlation

The tuning curve of a neuron i is generally defined as the spike count averaged across trials, $f_i(s) = \text{E}[Y_{ir}(s)]$, which varies with the stimulus s when s modulates the neuron (Butts & Goldman, 2006). Under the assumptions of section 2, the tuning curve reduces to $f_i(s) = \text{E}[X_{ir}(s)]$, which contains information about the firing rate averaged across trials but none about the trial-to-trial variability of $X_{ir}(s)$. Instead of reducing the tuning properties to the mean firing rate, we may retain in addition the full distribution of $X_{ir}(s)$, say F_{is}, for each stimulus s: this ensemble is what we call the generalized tuning curve (GTC). Figure 11A displays the GTC for two example V4 neurons, assuming the Poisson-lognormal model, equation 3.1, for their spike counts. Figure 8 shows the joint GTC for a neuron pair, which further displays their shared trial-to-trial covariability.

The tuning curve correlation (TCC) is widely considered an important measure of similarity of the neural activity between two neurons across stimuli. It can be defined as
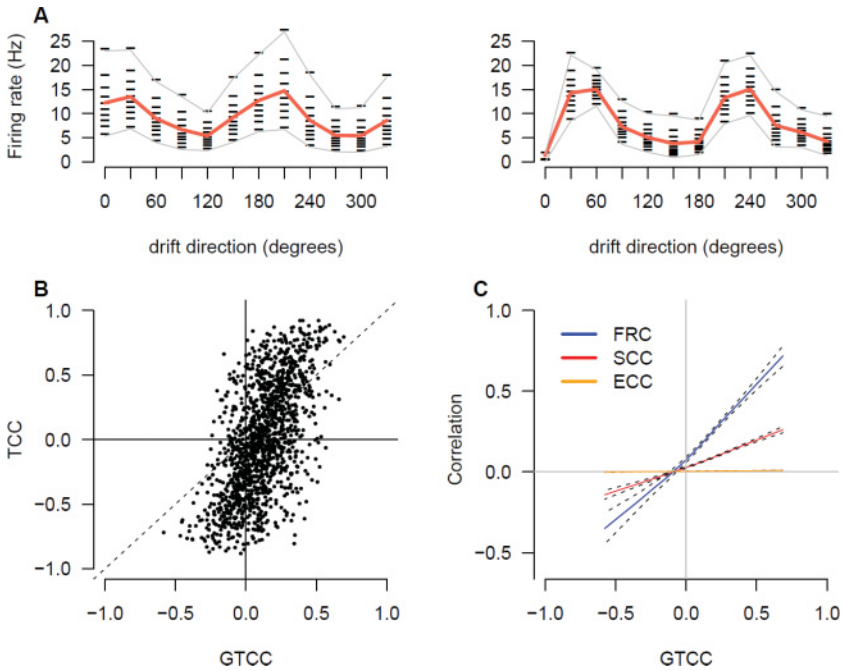
Figure 11: (A) Generalized tuning curves of two V4 neurons with $0.05, 0.15,$ $\dots, 0.95$ quantiles, estimated assuming the Poisson-lognormal model, equation 3.1. The red curves are the standard tuning curves. (B) GTCC versus TCC: They can have different signs, and GTCC appears more attenuated than TCC. (C) SCC, FRC, ECC versus GTCC (local linear regression with 95% simultaneous confidence bands): all correlations increase with GTCC; the relationships are more pronounced than when plotted versus TCC in Figure 7.

$$TCC = Cor(f_1(S), f_2(S)), \tag{H.1}$$

where the covariance is with respect to the random variable S, which codes for the stimulus; for example, when all stimuli are presented with equal frequency, S is uniform over the stimuli. Consider the toy example where, for a stimulus s, $(X_{1r}(s), X_{2r}(s))$ are (trial-to-trial) random variables with generative model:

$$\begin{cases} X_{1r}(s) \sim F_{1s}, \\ X_{2r}(s) = (1 + \beta) f_1(s) - \beta X_{1r}(s), \end{cases}$$

where we choose $\beta > 0$ to achieve $P(X_{2r}(s) \geq 0) = 1$. This model implies $E[X_{1r}(s)] = E[X_{2r}(s)]$ for all s and, in turn, $TCC = 1$, from which one might

conclude that the two neurons have proportional tuning curves. But the model also implies $\text{FRC}_s = \text{Cor}(X_{1r}(s), X_{2r}(s)) = -1$ for all s, which reveals that TCC only summarizes the similarity of the expected firing rates of the two neurons across stimuli but does not convey the sign and magnitude of their shared trial-to-trial variability. This motivates us to define the generalized tuning curve correlation:

$$\text{GTCC} = \text{Cor}(X_{1r}(S), X_{2r}(S)), \qquad\qquad (H.2)$$

where the correlation is with respect to the stimulus random variables S, as in equation H.1, but also to the trial-to-trial random firing rates $X_{ir}(s)$. Hence, the TCC only captures the covariability of the expected firing rates across stimuli, and the GTCC also accounts for changes in their joint distribution, such as trial-to-trial firing rate correlation (FRC). That is, the GTCC summarizes the strength of the relationship between two neurons across both trials and stimuli.

Note that when the experiment involves only one stimulus, so that $S = s$ is not a random variable since it always takes the same value, then GTCC = FRC; when the firing rates have no trial-to-trial variability, that is, when $X_{ir}(s) = \text{E}(X_{ir}(s))$ for all r, then GTCC = TCC since $f_i(s) = X_{ir}(s)$ for all r. When both the stimulus and the firing rates are random, the GTCC can be viewed as an approximate average of FRC across stimuli or as an approximate average of TCC across trials. More precisely, we can rewrite

$$
\begin{aligned}
\text{GTCC} &= \frac{\text{E}[\text{Cov}(X_{1r}(S), X_{2r}(S)|S)] + \text{Cov}(f_1(S), f_2(S))}{\prod_{i=1,2} \sqrt{\text{E}[\text{Var}(X_{ir}(S)|S)] + \text{Var}(f_i(S))}} \\
&= \text{TCC} \times \frac{\text{E}[\text{Cov}(X_{1r}(S), X_{2r}(S)|S)]/\text{Cov}(f_1(S), f_2(S)) + 1}{\prod_{i=1,2} \sqrt{\text{E}[\text{Var}(X_{ir}(S)|S)]/\text{Var}(f_i(S)) + 1}} \\
&= \text{TCC} \times \Omega, \qquad\qquad (H.3)
\end{aligned}
$$

where Ω depends on the joint distributions of the firing rates at different stimuli (Figure 8 shows such a distribution), so that GTCC can be viewed as a modified version of TCC. Figure 11B shows GTCC versus TCC for all the neuron pairs in the V4 data: GTCC has smaller magnitude, and GTCC and TCC sometimes have different signs. Figure 11C shows SCC, FRC, and ECC as functions of GTCC, which resembles the plot of the same quantities versus TCC in Figure 7B, although the relationships appear stronger. This happens because GTCC is approximately an average of FRC across stimuli, as justified above, whereas SCC is an attenuated version of FRC.

Acknowledgments

References

Amarasingham, A., Geman, S., & Harrison, M. T. (2015). Ambiguity and nonidentifiability in the statistical analysis of neural codes. *Proceedings of the National Academy of Sciences*, *112*, 6455–6460.

Behseta, S., Berdyyeva, T., Olson, C. R., & Kass, R. E. (2009). Bayesian correction for attenuation of correlation in multi-trial spike count data. *Journal of Neurophysiology*, *101*, 2186–2193.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*, 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*, 1165–1188.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory*. Hoboken, NJ: Wiley.

Butts, D. A., & Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biology*, *4*(4), e92.

Canty, A. J., Davison, A. C., Hinkley, D. V., & Ventura, V. (2006). Bootstrap diagnostics and remedies. *Canadian Journal of Statistics*, *34*, 5–27.

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X. J., Pouget, A., & Shadlen, (2011). Variance as a signature of neural computations during decision making. *Neuron*, *69*, 818–831.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*, 811–819.

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*, 1594–1600.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. Hoboken, NJ: Wiley.

Cowan, R. L., & Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *Journal of Neurophysiology*, *71*, 17–32.

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., . . . Tolias, A. S. (2014) State dependence of noise correlations in macaque primary visual cortex. *Neuron*, *82*, 235–248.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, *17*, 858–865.

Gu, Y., Liu, S., Fetsch, C. R., Yang, Y., Fok, S., Sunkara, A., . . . Angelaki, D. E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron*, *71*, 750–761.

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, *12*, 509–523.

Harrison, M. T., Amarasingham, A., & Kass, R. E. (2013). Statistical identification of synchronous spiking. In P. D. Lorenzo & J. Victor (Eds.), *Spike timing: Mechanisms and function*. London: Taylor and Francis.

Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2013). Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, *78*, 352–363.

Kass, R. E., Eden, U. T., & Brown, E. N. (2014). *Analysis of neural data*. New York: Springer.

Kass, R. E., & Ventura, V. (2006). Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Computation*, *18*, 2583–2591.

Kelly, R. C., & Kass, R. E. (2012). A framework for evaluating pairwise and multiway synchrony among stimulus-driven neurons. *Neural Computation*, *24*, 2007–2032.

Kelly, R. C., Smith, M. A., Kass, R. E., & Lee, T. S. (2010). Local field potentials indicate network state and account for neuronal response variability. *Journal of Computational Neuroscience*, *29*, 567–579.

Kelly, R. C., Smith, M. A., Samonds, J. M., Kohn, A., Bonds, A. B., Movshon, J. A., & Lee, T. S. (2007). Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *Journal of Neuroscience*, *27*, 261–264.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron*, *63*, 879–888.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, *18*, 3870–3896.

Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, *28*, 12591–12603.

Smith, M. A., & Sommer, M. A. (2013). Spatial and temporal scales of neuronal correlation in visual area v4. *Journal of Neuroscience*, *33*, 5422–5432.

Snyder, D. L., & Miller, M. I. (1991). *Random point processes in time and space*. New York: Springer.

Staude, B., Rotter, S., & Grün, S. (2008). Can spike coordination be differentiated from rate covariation? *Neural Computation*, *20*, 1973–1999.

Steriade, M., & Buzsaki, G. (1990). Parallel activation of thalamic and cortical neurons by brainstem and basal forebrain cholinergic systems. In M. Steriade & D. Biesold (Eds.), *Brain cholinergic systems* (pp. 3–62). New York: Oxford University Press.

Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *Journal of Neurophysiology*, *85*, 1969–1985.

Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *Annals of Applied Statistics*, *3*, 1236–1265.

Timofeev, I., Grenier, F., & Steriade, M. (2001). Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: An intracellular study. *Proceedings of the National Academy of Sciences*, *98*, 1924–1929.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*, 414–419.

Tomko, G. J., & Crapper, D. R. (1974). Neuronal variability: Non-stationary responses to identical visual stimuli. *Brain Research*, *79*, 405–418.

Ventura, V., Cai, C., & Kass, R. E. (2005). Trial-to-trial variability and its effect on time-varying dependency between two neurons. *Journal of Neurophysiology*, *94*, 2928–2939.