From Univariate to Multivariate Coupling Between Continuous Signals and Point Processes: A Mathematical Framework

Shervin Safavi
shervin.safavi@tuebingen.mpg.de
MPI for Biological Cybernetics, and IMPRS for Cognitive and Systems Neuroscience, University of Tübingen, 72076 Tübingen, Germany

Nikos K. Logothetis
nikos.logothetis@tuebingen.mpg.de
MPI for Biological Cybernetics, 72076 Tübingen, Germany; International Center for Primate Brain Research, Songjiang, Shanghai 200031, China; and University of Manchester, Manchester M13 9PL, U.K.

Michel Besserve
michel.besserve@tuebingen.mpg.de
MPI for Biological Cybernetics and MPI for Intelligent Systems, 72076 Tübingen, Germany

Time series data sets often contain heterogeneous signals, composed of both continuously changing quantities and discretely occurring events. The coupling between these measurements may provide insights into key underlying mechanisms of the systems under study. To better extract this information, we investigate the asymptotic statistical properties of coupling measures between continuous signals and point processes. We first introduce martingale stochastic integration theory as a mathematical model for a family of statistical quantities that include the phase locking value, a classical coupling measure to characterize complex dynamics. Based on the martingale central limit theorem, we can then derive the asymptotic gaussian distribution of estimates of such coupling measure that can be exploited for statistical testing. Second, based on multivariate extensions of this result and random matrix theory, we establish a principled way to analyze the low-rank coupling between a large number of point processes and continuous signals. For a null hypothesis of no coupling, we establish sufficient conditions for the empirical distribution of squared singular values of the matrix to converge, as the number of measured signals increases, to the well-known Marchenko-Pastur (MP) law, and the largest squared singular value converges to the upper end of the MP support. This justifies a simple thresholding approach to assess the significance of multivariate coupling. Finally, we illustrate with

simulations the relevance of our univariate and multivariate results in the context of neural time series, addressing how to reliably quantify the interplay between multichannel local field potential signals and the spiking activity of a large population of neurons.

1 Introduction

The observation of highly multivariate temporal point processes, corresponding to the activity of a large number of individuals or units, is pervasive in many applications (for example, neurons in brain networks; Johnson, 1996) and members in social networks (Dai, Wang, Trivedi, & Song, 2016; De, Valera, Ganguly, Bhattacharya, & Rodriguez, 2016). As the number of observed events per unit may remain small, inferring the underlying dynamical properties of the studied system from such observations is challenging. However, in many cases, it is possible to observe continuous signals whose coupling with the events can offer key insights.

In neuroscience, this is the case of the extracellular electrical field, which provides information complementary to spiking activity. Local field potentials (LFP) are mesoscopic (Liljenstroem, 2012) signals resulting from the superposition of the electric potentials generated by ionic currents flowing across the membranes of the cells located close to the tip of recording electrodes. The LFP reflects neural cooperation due to the anisotropic cytoarchitecture of most brain regions, allowing the summation of the extracellular currents resulting from the activity of neighboring cells. As such, a number of subthreshold integrative processes (i.e., modifying the neurons' internal state without necessarily triggering spikes) contribute to the LFP signal (Buzsaki, Anastassiou, & Koch, 2012; Buzsaki, Logothetis, & Singer, 2013; Einevoll, Kayser, Logothetis, & Panzeri, 2013; Pesaran et al., 2018; Herreras, 2016).

Reliably quantifying the coupling between activities of individual units (e.g., spikes generated by individual neurons) in a circuit and the aggregated measures (such as the LFP) may provide insights into underlying network mechanisms, as illustrated in the electrophysiology literature. At the single neuron level, the relationship of spiking activity to subthreshold activity has broad implications for the underlying cellular and network mechanisms at play. For instance, it has been suggested that synaptic plasticity triggers changes in the coupling between spikes and LFPs (Grosmark, Mizuseki, Pastalkova, Diba, & Buzsáki, 2012; Grosmark & Buzsáki, 2016). Regarding the putative functional role of such observed couplings, it has been hypothesized to support cognitive functions such as attention. Such coordination by oscillations hypothesis proposes that network oscillations modulate differentially the excitability of several target populations, such that a sender population can emit messages during the window of time for which a selected target is active, while unselected targets are silenced (Fries, 2005, 2015; Womelsdorf et al., 2007).

In the case of two continuous signals, coupling measures such as coherence and phase locking value (PLV) (Rosenblum, Pikovsky, Kurths, Schäfer, & Tass, 2001; Pereda, Quiroga, & Bhattacharya, 2005) are widely used, and their statistical properties have been investigated, in particular in the stationary gaussian case (Brillinger, 1981; Aydore, Pantazis, & Leahy, 2013). In a similar way, PLV (Ashida, Wagner, & Carr, 2010) and spike-field coherence (SFC) (Mitra, 2007) can measure spike-LFP coupling (see among others: Vinck, Battaglia, Womelsdorf, & Pennartz, 2012; Vinck, van Wingerden, Womelsdorf, Fries, & Pennartz, 2010; Jiang, Bahramisharif, van Gerven, & Jensen, 2015; Zarei, Jahed, & Daliri, 2018; Li, Cui, & Li, 2016) and are broadly used to makes sense of the role played by neurons in coordinated network activity (Buzsaki & Schomburg, 2015). There are notable contributions investigating potential biases of those measures when both point processes and continuous signals are involved (Lepage, Kramer, & Eden, 2011; Kovach, 2017). However, two issues relevant for practical applications remain: (1) the effect of the intrinsic variability of spike occurrence on key statistical properties of the estimates, such as the variance, have not yet been thoroughly described, and (2) how to extend the rigorous statistical analysis of spike-field coupling in the context of the highly multivariate signals available with modern recording techniques remains largely unaddressed.

We address these two issues by using continuous time martingale theory (see Liptser & Shiryaev, 2013a), the related concept of stochastic integration (see Protter, 2005) and random matrix theory (Anderson, Guionnet, & Zeitouni, 2010; Capitaine & Donati-Martin, 2016). The martingale central limit theorem (CLT) allows us to derive analytically the asymptotic gaussian distribution of a general family of coupling measures that can be expressed as stochastic integrals. We exploit this general result to show that the classical univariate PLV estimator is also asymptotically normally distributed and provide the analytical expression for its mean and variance. Furthermore, we study potential sources of bias for the commonly used von Mises coupling model (Ashida et al., 2010). We then go beyond univariate coupling measures and analyze the statistical properties of a family of multivariate coupling measures taking the form of a matrix with stochastic integral coefficients. We characterize the jointly gaussian asymptotic distribution of matrix coefficients, and exploit random matrix theory (RMT) principles to show that after appropriate normalization, the spectral distribution of such large matrices under the null hypothesis (of absence of coupling), follows approximately the Marchenko-Pastur (MP) law (Marchenko & Pastur, 1967), while the magnitude of the largest singular value converges to a fixed value whose simple analytic expression depends only on the shape of the matrix. We finally show how this result provides a fast and principled procedure to detect significant singular values of the coupling matrix, reflecting an actual dependency between the underlying signals. In the appendixes, we included detailed proofs and background material on RMT

and stochastic integration, such that nonexpert readers can further apply these tools in neuroscience.

2 Background

2.1 Spike-Field Coupling in Neuroscience. Although our results are relevant to a broad range of applications within and beyond neuroscience, we will use the estimation of spike-LFP coupling introduced above as the guiding example of this letter. Spikes convey information communicated between individual neurons. This information is believed to be encoded in the occurrence times of successive spike events, which are typically modeled with point processes—for example, Poisson (Softky & Koch, 1993) or Hawkes process (Truccolo, 2016; Krumin, Reutsky, & Shoham, 2010).

While oscillatory dynamics is ubiquitous in the brain and instrumental to its coordinated activity (Buzsaki, 2006; Buzsaki et al., 2013; Peterson & Voytek, 2018), it is often challenging to uncover based solely on the sparse spiking activity of recorded neurons. On the other hand, LFPs often exhibit oscillatory components that can be isolated with signal processing tools (typically bandpass filtering or template matching), such that pairing the temporal information from LFPs and spiking activity can help extract reliable markers of neural coordination.

An example of a coupling measure achieving such pairing is the phase locking value (PLV). Given, on the one hand, event (spike) times $\{t_j\}$ where $j \in \{1, 2, \ldots, N\}$ (with N the number of spikes in the spike train), and on the other hand, $\phi(t)$ the time-varying phase of an oscillatory continuous signal, which is typically a bandpassed filtered LFP, phase locking between these two signals is estimated by the complex number

$$\widehat{\text{PLV}} = \frac{1}{N} \sum_{j=1}^{N} e^{i\phi(t_j)}, \text{ with } i^2 = -1. \tag{2.1}$$

We use a "hat" notation to reflect that this quantity is empirical: indeed, even if we assume a fixed ϕ, the PLV depends on the specific values of event times t_j. In this work, we assume these points are drawn from a Poisson process, with a possibly time-varying rate (inhomogeneous Poisson process), such that we can define a population statistics that is a function of the point process population distribution instead of its empirical counterpart. We then address under which conditions the empirical PLV reflects a true coupling between the rate of the underlying point process and ϕ.

2.2 Counting Process Martingales. We use a continuous time framework leading to powerful results based on concise deterministic and stochastic integral expressions, which can trivially be approximated using discrete time signals in practice. A (continuous time) stochastic process
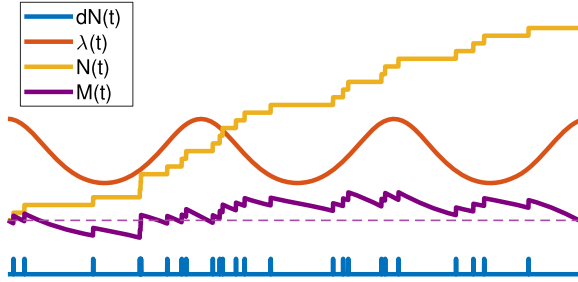
Figure 1: Doob-Meyer decomposition for an example inhomogeneous Poisson process with oscillatory of rate $\lambda(t)$ of frequency $f = 1$ Hz; average firing rate $\lambda_0 = 5$ Hz (dashed line indicates the reference 0). See section 3.3 for details of the simulation.

$M = \{M(t); t \in [0, \tau]\}$ is a zero-mean martingale relative[1] to the filtration $\{\mathcal{F}_t\}$ (which represents the past information accumulated up to time t) if (1) $M(0) = 0$, (2) it is adapted to $\{\mathcal{F}_t\}$ (informally the law of M up to time t "uses" only past information up to t), and (3) it satisfies the martingale property:

$$E\left[M(t)|\mathcal{F}_s\right] = M(s), \quad \text{for all} \quad t > s. \tag{2.2}$$

Consider now a (univariate) counting process $\left\{(N(t), \mathcal{F}_t); t \geq 0\right\}$, counting the number of events that occurred up to time t, adapted to filtration $\{\mathcal{F}_t\}$ (Aalen, Borgan, & Gjessing, 2008, chap. 2). Under mild assumptions, it has a Doob-Meyer decomposition,

$$N(t) = M(t) + \int_0^t \lambda(t)dt, \tag{2.3}$$

where $\lambda(t)$ is a predictable process with respect to $\{\mathcal{F}_t\}$ called the intensity function and $M(t)$ is a martingale, called the compensated counting process. Figure 1 shows an illustration of this decomposition for a Poisson process with sinusoidal intensity.

Consider now an empirical coupling measure c between a (real or complex) predictable process $x(t)$ and $N(t)$ observed during time interval $[0, T]$, which takes the form of the stochastic integral (see Protter, 2005),

$$\widehat{c} = \sum_{t_k < T} x(t_k) = \int_0^T x(t)dN(t), \tag{2.4}$$

---

[1] Any martingale in this paper is zero-mean.

where $\{t_k\}$ denote the jump times of the counting process (note that the PLV defined in equation 2.1 is a normalized version of such coupling). The empirical coupling measure, $\widehat{c}$, can then be decomposed as

$$\widehat{c} = \int_0^T x(t)\lambda(t)dt + \int_0^T x(t)dM(t). \tag{2.5}$$

Interestingly, it can be shown that the second integral on the right-hand side is also a martingale (see Liptser & Shiryaev, 2013b, theorem 18.7).

In order to keep our results concise, we assume the following deterministic setting in the remainder of this letter (see section 5 for potential extensions).

Assumption 1. Assume the intensity function, $\lambda(t) = \lambda(t|\mathcal{F}_t)$ of $N(t)$ and the signal $x(t)$ are deterministic bounded left-continuous and adapted to $\mathcal{F}_t$ over $[0, T]$.

Note this entails that $N(t)$ is a (possibly inhomogeneous) Poisson process (Liptser & Shiryaev, 2013b, theorem 18.10). Under assumption 1, the terms of equation 2.5 separate the deterministic part from the (zero-mean) random fluctuations of the measure that are integrally due to the martingale term. Using martingale properties, the statistics of the coupling measure are[2]

$$c^* \triangleq \mathbb{E}[\widehat{c}] = \int_0^T x(t)\lambda(t)dt \quad \text{and}$$

$$\mathrm{Var}[\widehat{c}] = \mathbb{E}\left[|\widehat{c} - c^*|^2\right] = \int_0^T |x^2(t)|\lambda(t)dt. \tag{2.6}$$

In case $x(t)$ integrates to zero, the expected coupling c^* thus reflects the covariation across time between $x(t)$ and the intensity of the point process up to random fluctuations.

2.3 Random Matrix Theory. As data sets get increasingly high dimensional, it becomes important to replace the above univariate measure $\widehat{c}$ by a quantity that summarizes the coupling between a large number of units and continuous signals. This extension leads to assessing the spectral properties of a coupling matrix $\widehat{\mathbf{C}}$ that gathers all pairwise measurements. However, such task is nontrivial due to the martingale fluctuations affecting $\widehat{\mathbf{C}}$, leading to spurious nonzero coupling coefficients, and can also hide the deterministic structure of the matrix associated with significant coupling.

Random matrix theory allows investigating the spectral properties of some matrices in noisy settings by studying their asymptotic spectral properties as dimensions grow to infinity. Any $(p \times p)$ complex Hermitian or

---

[2]See section B.1.1 for more details.

real symmetric matrix M has a set of p real eigenvalues $\{\ell_k\}$ (where we put several times the same eigenvalue in the set according to its multiplicity). One classically studied quantity is the empirical spectral distribution (ESD) (or empirical eigenvalue distribution, see Mingo & Speicher, 2017 and Anderson et al., 2010) of the set of all eigenvalues $\{\ell_k\}$. ESD indistinctly refers (with a slight abuse of language), to either the probability measure (also called *spectral measure* in our case),

$$\mu_M(t) = \frac{1}{p}(\delta_{\ell_1}(t) + \cdots + \delta_{\ell_p}(t)), \ t \in \mathbb{R},$$

where δ_{ℓ_k} is the dirac measure with unit mass in ℓ_k, or to its associated cumulative distribution:

$$F_M(t) = \int_{-\infty}^{t} d\mu_M(s).$$

Seminal works by Wigner (1955, 1958), Marchenko and Pastur (1967), and many others have established the convergence of the ESD of large random matrix ensembles (see section B.2 for the precise notions of convergence). In particular, for a sequence of matrices $\{X_n\}_{n>0}$ of dimension $p \times n$ such that $\frac{p}{n} \underset{n \to +\infty}{\to} \alpha \leq 1$, with coefficients sampled independently and identically distributed (i.i.d.) from a (possibly complex) standard Normal distribution, the ESD of the Wishart matrix $S_n = \frac{1}{n}X_nX_n^H$ (where $.^H$ indicates the transposed complex conjugate) converges to the Marchenko-Pastur (MP) law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

$$\frac{d\mu_{MP}}{dx}(x) = \begin{cases} \frac{1}{2\pi\alpha x}\sqrt{(b-x)(x-a)}, & a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

with $a = (1 - \sqrt{\alpha})^2$ and $b = (1 + \sqrt{\alpha})^2$. Additionally, the smallest and largest eigenvectors converge to a and b, respectively. Importantly, these convergences also hold in the case $\alpha > 1$, but equation 2.7 is modified to account for the rank deficiency of the Wishart matrix, imposing $p - n$ zero eigenvalues in the spectrum (see section B.3.1 for details).

Notably, recent developments in the field of random matrix theory extend the classic results that were only valid for independent coefficients (uncorrelated Wishart matrices) to various forms of dependencies between coefficients. For instance, El Karoui (2007, 2008) showed that the ESD and the distribution of the largest eigenvalue for a sequence of matrices $\{X_n\}_{n>0}$ with general covariance matrices (not necessarily with identity covariance matrix) follow similar laws and Banna, Merlevède, and Peligrad (2015) investigate the case of symmetric random matrices with correlated entries.

Furthermore, the behavior of high-dimensional autocovariance matrices in the context of discrete time stochastic processes is discussed in Liu, Aue, and Paul (2015) and Bhattacharjee and Bose (2016). Applications of this framework have also been considerably extended including global finance (Namaki et al., 2020) and various aspects of machine learning and signal processing such as shallow (Louart, Liao, & Couillet, 2018) and deep (Pennington & Bahri, 2017; Pennington & Worah, 2019) neural networks, denoising (Bun, Bouchaud, & Potters, 2017) and dimensionality reduction (Johnstone & Onatski, 2020).

In this study, we show that the martingale fluctuations of the coupling matrices also cause spectral convergence to the MP law in the absence of actual coupling between the signals. Recent results on the low-rank perturbation (Capitaine & Donati-Martin, 2016; Loubaton & Vallet, 2011; Benaych-Georges & Nadakuditi, 2012) of random matrices suggest this convergence can be exploited to further assess the significance of the largest eigenvalues of the coupling matrix with respect to the null hypothesis that they only reflect random fluctuations.

3 Assessment of Univariate Coupling

3.1 Mathematical Formulation. We consider the setting of $K \geq 1$ independent trials of measurements on $[0, T]$ available to estimate the coupling statistics by the trial average

$$\widehat{c}_K = \frac{1}{K} \sum_{k=1}^{K} \int_0^T x(t) dN^{(k)}(t),$$

where $\{N^{(k)}\}$ are K independent copies of the process $N(t)$, associated with each trial. As this letter focuses on the statistical properties induced by the intrinsic variability of point process realizations, we assumed above that the continuous signal does not change across trials. However, including some forms of variability across trials, such as random time shifts affecting all processes in the same way, would not affect the results, barring additional technical details.

We exploit a central limit theorem (CLT) for martingales to show the residual variability (difference between the empirically estimated $\widehat{c}_K$ and the expected coupling c^* of equation 2.6) is asymptotically normally distributed. We formally state it in theorem 1.

Theorem 1. *Assume $(\mathcal{F}_t, x(t), \lambda(t))$ satisfy assumption 1, and $x(t)$ real-valued. Then,*

$$\mathbb{E}[\widehat{c}_K] \triangleq c^* = \int_0^T x(t)\lambda(t)dt \quad and \quad \mathrm{Var}[\widehat{c}_K] = \frac{1}{K} \int_0^T x^2(t)\lambda(t)dt.$$

Moreover, as the number of trials increases, fluctuations converge in distribution:

$$\sqrt{K}\,(\hat{c}_K - c^*) \xrightarrow[K \to +\infty]{} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right).$$

Sketch of the proof. We rely on the decomposition of equation 2.5. As described in section B.1.1, the martingale property is preserved by the stochastic integral term and allows us to exploit a martingale CLT to prove convergence to a gaussian distribution. $\square$

The case of $x(t)$ complex-valued can be dealt with by distinguishing the real and imaginary parts of the signal, as is done in the proofs of the following corollaries. We can exploit theorem 1 to derive the asymptotic properties of the PLV introduced in section 2.1. For that, we adapt the empirical estimate of equation 2.1 to the K trials setting introduced above and define

$$\widehat{\text{PLV}}_K = \frac{1}{\sum_{k=1}^K N_k} \sum_{k=1}^K \sum_{j=1}^{N_k} e^{i\phi(t_j^k)}, \tag{3.1}$$

where N_k is the number of events observed during trial k and $\{t_j^k\}$ is the collection of the time stamps of these events. The specificity of this multi-trial estimate is to use a single normalization constant corresponding to the total number of events pooled across trials.[3] For this estimate, we get the following result.

Corollary 1. *Assume* $(\mathcal{F}_t, x(t) = e^{i\phi(t)}, \lambda(t))$ *satisfy assumption 1, where ϕ is real-valued and stands for the phase of the signal x. Then the expectation of the PLV statistics $\widehat{\text{PLV}}_K$ estimated from K trials of measurements on $[0, T]$ tends to the limit*

$$\text{PLV}^* = \int_0^T e^{i\phi(t)}\lambda(t)dt / \Lambda(T), \quad with \quad \Lambda(T) = \int_0^T \lambda(t)dt. \tag{3.2}$$

Moreover, as $K \to +\infty$, the residual,

$$\sqrt{K}\left(\widehat{\text{PLV}}_K - \text{PLV}^*\right), \tag{3.3}$$

---

[3]This allows the normalization factor to converge to a deterministic quantity as $K \to +\infty$ equation 2.1.

converges in distribution to a zero-mean complex gaussian variable Z (i.e., the joint distribution of real and imaginary parts is gaussian), such that

$$\text{Cov}\left(\begin{bmatrix} Re\{Z\} \\ Im\{Z\} \end{bmatrix}\right) = \frac{1}{\Lambda(T)^2} \int_0^T M(t)\lambda(t)dt,$$

$$where\, M(t) = \begin{bmatrix} \cos^2(\phi(t)) & \sin(2\phi(t))/2 \\ \sin(2\phi(t))/2 & \sin^2(\phi(t)) \end{bmatrix}.$$

Sketch of the proof. This relies on applying theorem 1 to the real and imaginary parts of $e^{i\phi(t)}$. In addition, the coupling between both quantities is taken into account by replacing the variance of univariate quantities $\widetilde{V}(t)$ in theorem 1 by a covariance matrix that can be assessed with martingale results given in section B.1.1.   □

Remark 1. For the simple case of a T/k-periodic sinusoidal signal (k integer), such that $\phi(t) = 2\pi kt/T$, and a sinusoidal modulation of the intensity with phase shift φ_0 and modulation amplitude $\varkappa$ such that

$$\lambda(t) = \lambda_0 \left(1 + \varkappa \cos\left(\phi(t) - \varphi_0\right)\right),\, \lambda_0 > 0,\, 0 \le \varkappa \le 1,$$

we get easily with trigonometric identities that PLV* $= \frac{1}{2}\varkappa e^{i\varphi_0}$ and the residual of equation 3.3 converges to an isotropic complex gaussian of total variance[4] $\frac{1}{\lambda_0 T}$ such that the coupling strength $\varkappa$ affects the mean but not the variance of the PLV estimate.

Also, it is easy to see that if $\lambda(t)$ is modulated by a sine wave at a different integer multiple $m \ne k$ of the fundamental frequency $1/T$, such that

$$\lambda(t) = \lambda_0 + \varkappa \cos\left(2\pi mt/T - \varphi_0\right),$$

the PLV* vanishes and the residual's variance remains the same. These properties make PLV straighforward to interpret and test for sinusoidal coupling with a carefully chosen observation duration T. Assumption 3 and corollary 5, in appendix C, provide formal statements of this remark.

We can use corollary 1 to predict the statistics of PLV estimates for other models of phase-locked spike trains. A classical model uses the von Mises distribution (also known as circular normal distribution) with parameter $\kappa \ge 0$ to model the concentration of spiking probability around a specified locking phase ϕ_0 (for more details, see Ashida et al., 2010). The original model uses a purely sinusoidal time series by assuming a linearly

---

[4]The sum of the variances of real and imaginary parts.

increasing phase $\phi(t) = 2\pi f t$, where f is the modulating frequency, to derive the intensity of an inhomogeneous Poisson spike train,

$$\lambda(t) = \lambda_0 \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right). \tag{3.4}$$

resulting in an analytical expression for the asymptotic complex-valued PLV,

$$\text{PLV}^* = e^{\mathbf{i}\varphi_0} \frac{\int_0^\pi \cos(\theta) \exp(\kappa \cos(\theta)) d\theta}{\int_0^\pi \exp(\kappa \cos(\theta)) d\theta} = e^{\mathbf{i}\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)},$$

with the I_k's denoting the modified Bessel functions of the first kind for k integer (see Abramowitz & Stegun, 1972, p. 376):

$$I_k(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta.$$

Compared to the sinusoidal coupling described in remark 1, whose PLV magnitude can reach at most $1/2$, this model can achieve arbitrarily large PLV, which might explain why it is more frequently used in applications.

Corollaries 2 and 3 derive the asymptotic covariance of the variability of the PLV estimate around this theoretical value (which is novel to the best of our knowledge). Furthermore, the results are derived in a more general model setting accounting for "biases"[5] due to nonlinear phase increases $\phi(t)$ and observation intervals that are not multiples of the modulating oscillation period. It should be noted that the mentioned biases are inherent in the estimator's definition. They happen independent of additional biases originating from the phase estimation procedure (e.g., phase extraction via Hilbert transform; see Kovach, 2017).

We thus assume a coupling, parameterized by κ between a possibly nonlinearly increasing phase $\phi(t)$ and a point process with intensity

$$\lambda(t) = \lambda_0 \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) \frac{d\phi}{dt}(t). \tag{3.5}$$

Note that for linearly increasing phases, this coupling amounts to the classical von Mises model of equation 3.4. The additional factor $\frac{d\phi}{dt}(t)$ allows preserving the analytical expression of PLV statistics even for nonlinearly increasing phases, providing a novel generalization of the von Mises model (see corollary 4 in appendix C for a simplified version of corollary 2, assuming a linearly increasing phase $\phi(t) = 2\pi f t$ with frequency f).

---

[5]They are biases in the sense that one would expect a coupling measure to vanish if there is no coupling in the data generating procedure.

Corollary 2. *Under the assumptions of corollary 1, assume additionally that $\phi(t)$ is continuous, strictly increasing, and piece-wise differentiable on $[0, T]$ and the intensity of the point-process is given by equation 3.5 for a given $\kappa \geq 0$, then the expectation of the multitrial PLV estimate converges (for $K \to +\infty$) to*

$$\text{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}. \tag{3.6}$$

If in addition $[0, T]$ corresponds to an integer number of periods of the oscillation,

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_0^\pi \cos(\theta) \exp(\kappa \cos(\theta)) d\theta}{\int_0^\pi \exp(\kappa \cos(\theta)) d\theta} = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{3.7}$$

and the scaled residual $\sqrt{K}\left(\widehat{\text{PLV}}_K - \text{PLV}^*\right)$ converges to a zero mean complex gaussian Z with the following covariance:

$$\text{Cov} \begin{bmatrix} Re\{Ze^{-i\varphi_0}\} \\ Im\{Ze^{-i\varphi_0}\} \end{bmatrix}$$

$$= \frac{1}{2\lambda_0(\phi(T) - \phi(0))I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \tag{3.8}$$

Sketch of the proof. This is based on plugging the intensity function $\lambda(t)$ of equation 3.5 in corollary 1. Using change of variable in the integrals ($\phi(t)$ to θ) and exploiting the symmetries of the functions, the integrals in the analytical expressions of the expectation and covariance turn into modified Bessel functions I_k for k integer. □

The above result has important consequences for the assessment of PLV from data. In particular, it exhibits key experimental requirements for PLV estimates to match the classical Bessel functions expression of equation 3.7: (1) evaluate PLV on an integer number of periods (this is critical for trials with short duration) and (2) take into account the fluctuations of the rate of increase of the phase $\phi(t)$ across the oscillation period. This second point is critical in applications where the phase is inferred from signals (such as LFPs) through the Hilbert transform, as nonlinearities of the underlying phenomena may lead to nonsinusoidal oscillations, with periodic fluctuations of the time derivative of the phase $\phi'(t)$. To further emphasize the consequences of this aspect, we also derive the asymptotic distribution of PLV for a homogeneous Poisson process that corresponds to the special case $\kappa = 0$ of the classical von Mises coupling of equation 3.5. Although there is no actual coupling between events and the continuous signal in such a

case,[6] the nonlinear phase increase leads asymptotically (for K large) to a nonvanishing PLV estimate and to false detection of coupling.

Corollary 3. *Under the assumptions of corollary 1, we assume additionally that the point process is homogeneous Poisson with rate λ_0 and that $\phi(t)$ is strictly increasing (almost everywhere) and differentiable on $[0, T]$. Let $\theta \mapsto \tau(\theta)$ be its inverse function (such that $\tau(\phi(t)) = t$). Then the expectation of $\widehat{PLV}_K$ converges (for $K \to +\infty$) to*

$$\text{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{i\theta} \tau'(\theta) d\theta}{\phi(T) - \phi(0)}, \tag{3.9}$$

and the scaled residual,

$$Z = \sqrt{K}\left(\widehat{PLV}_K - PLV^*\right),$$

converges to a zero mean complex gaussian:

$$\sqrt{K}\left(\widehat{PLV}_K - PLV^*\right) \xrightarrow[K \to +\infty]{} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{Cov}(Z)\right),$$

with the following covariance:

$$\text{Cov}(Z) = \frac{1}{\lambda_0 T^2} \int_{\phi(0)}^{\phi(T)} \begin{bmatrix} \cos^2(\theta) & \sin(2\theta)/2 \\ \sin(2\theta)/2 & \sin^2(\theta) \end{bmatrix} \tau'(\theta) d\theta.$$

Sketch of the proof. The result stems from using the intensity function λ_0 in corollary 1 and then using change of variable in the integrals and exploiting the symmetries of the functions.   □

This corollary will be further illustrated in the next paragraphs.

3.2 Application to Bias Assessment. Corollary 3 predicts scenarios where in the absence of modulation of spiking activity (having a constant intensity function $\lambda(t) = \lambda_0$), the expectation of the PLV estimates remains far from zero even when the number of trials is large, that is, the coupling between a homogeneous point process and a continuous oscillatory signal would appear significant and reflect a form of bias. Corollary 3 allows computing this bias and therefore correcting it.

---

[6]In the sense that we can generate the homogeneous spike train and the oscillation without parametric models that do not share any information.

One such case is when the observation interval is not an integer number of oscillation periods. To demonstrate this analytically, we can start from the PLV expectation with the constant intensity λ_0,

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} = \frac{\lambda_0 \int_0^T e^{i\phi(t)}dt}{\lambda_0 \int_0^T dt} = \frac{1}{T}\int_0^T e^{i\phi(t)}dt. \tag{3.10}$$

Furthermore, we assume $\phi(t)$ has linear phase (assumption 3): $\phi(t) = 2\pi ft$, where f is the frequency of oscillation of the continuous signal. We then get

$$\text{PLV}^* = \frac{1}{T}\int_0^T e^{i2\pi ft}dt = \frac{1}{2\pi\gamma_T i}\left(e^{2\pi\gamma_T i} - 1\right), \tag{3.11}$$

where $\gamma_T = Tf$ is the ratio of the length of the time series (T) to the period of oscillation $\frac{1}{f}$. As is noticeable in equation 3.11, the coupling measure PLV^* is not zero when γ_T is not an integer number. Notably, this bias affects both the magnitude and the phase of the PLV^* estimate.

Furthermore, even using an observation interval covering an integer number of periods, nonlinear increases in phase may lead to a nonvanishing PLV. This can be demonstrated with a simple example. Again, we can start from the original definition of PLV expectation, equation 3.2, but now we do not assume the linearity of the phase. As introduced in corollary 3, let $\theta \mapsto \tau(\theta)$ be the inverse of $\phi(t)$, and let us use equation 3.9 to compute the PLV^*. Taking a sinusoidal modulation over the oscillation period, $\tau(\theta) = \theta + \epsilon\sin(\theta)$ with $|\epsilon| < 1$,[7] we get a nonvanishing asymptotic expected PLV:

$$\text{PLV}^* = \frac{1}{2\pi}\int_0^{2\pi} e^{i\theta}(1 + \epsilon\cos(\theta))d\theta = \epsilon\int_0^\pi e^{i\theta}\cos(\theta)d\theta = \epsilon/2 \neq 0, \text{ if } \epsilon \neq 0.$$

Our theoretical framework can be used for developing methods to correct such biases. In the linear phase setting, bias can be avoided simply by using an integer number of periods for coupling estimation. In the case of a nonlinear phase evolution of the continuous signal, we can use the theoretical phase (if available) or its empirical estimate to evaluate PLV^* under constant spike intensity assumptions with equation 3.9 and subtract this quantity to the estimated PLV. For resolving issues that arise due to the nonlinearity of the estimated phase, specialized methods have been suggested. For instance, Hurtado, Rubchinsky, and Sigvardt (2004) dealt with phase jumps (a particular form of nonlinearity) by interpolating the signal from the available data before and after the sudden change and Cole and

---

[7] To guarantee the phase to be strictly increasing.

Voytek (2019) introduced a cycle-by-cycle method for analyzing oscillatory dynamics. In this method, they consider a linear phase for each detected cycle of oscillation. Therefore, with this linear choice of phase, one can avoid the spurious coupling that can appear due to phase nonlinearities. Based on our framework, theoretically motivated methods that are not relying on the linearization of the phase can be developed.

3.3 Simulations. We demonstrate the outcome of our theoretical results using simulated phase-locked spike trains (similar to what has been introduced in corollaries 2 and 4) and sinusoidal oscillations. For generating phase-locked spike trains, we adopt the method introduced in Ashida et al. (2010). As the model has already been described elsewhere, we restrict ourselves to a brief explanation.

To generate phase-locked or periodic spike trains based on the classical von Mises model with rate $\lambda(t)$ as introduced in equation 3.4, we use a purely sinusoidal continuous signal $x(t)$ with linearly increasing phase $\phi(t) = 2\pi f t$, with $f = 1$ Hz and various coupling strength (κ) (see appendix E for lists of parameters used for each figure). Based on this simulation we perform two numerical experiments to demonstrate the practical relevance of our (asymptotic) theoretical results.

3.3.1 Experiment 1. In order to demonstrate the validity of corollaries 2 and 4, in Figure 2 we show the empirical distribution of the normalized residual of the PLV estimate and compare it to its asymptotic theoretical distribution. We simulate two cases, one with homogeneous Poisson spike trains ($\kappa = 0$) and one with phase-locked spike trains ($\kappa = 0.5$) with Poisson statistics. In both cases, we observe the agreement between theory and simulation, as the joint distribution of real and imaginary part approaches an isotropic gaussian. The slightly non-gaussian shape of the real part histogram for $\kappa = 0.5$ suggests, however, a slower convergence to the normal distribution in the case of coupled signals.

3.3.2 Experiment 2. We demonstrate an application of corollary 3 for bias evaluation with a simple simulation. In section 3.2 we pointed out that using a noninteger fT (T is not a multiple of the oscillation period) can lead to spurious correlation between the point process and the oscillatory continuous signal. By using equation 3.11 we can compute this bias.

We use a simulation similar to the one used in the previous experiment with an oscillatory signal and a homogeneous Poisson spike train ($\kappa = 0$) and investigate the coupling between these two signals. If the length of the continuous signal is not an integer number of the oscillation period, the PLV estimate has a nonzero empirical mean (see Figures 3A and B) while when it is a multiple of number of the oscillation period, the estimate matches the ground truth (see Figure 3C). In Figure 3D we compare the theoretical prediction and the numerical simulation for various length of the signals,
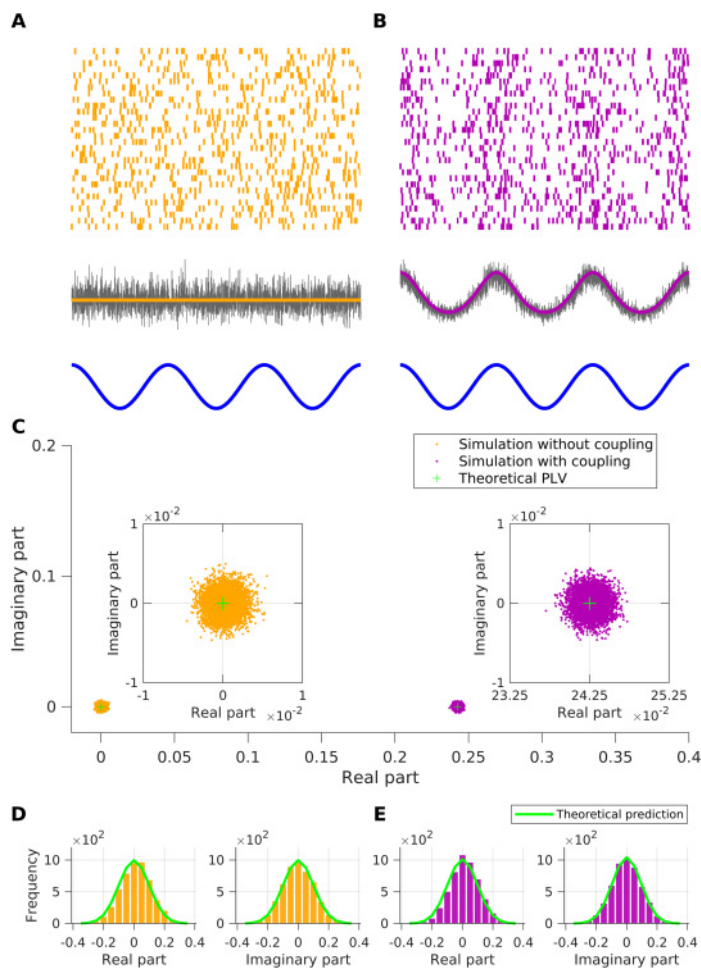
Figure 2: Simulation of (A) homogeneous Poisson spike trains and (B) phase-locked spike train with Poisson statistics (von Mises model with $\kappa = 0.5$). First row: Example raster plot of the spikes. Second row: Empirical firing rate (gray line) and ground truth firing rate (orange and purple traces). Third row: Continuous signal $x(t)$. (C) Scatter plots represent the complex-valued PLVs estimates. Each dot represents one realization of the simulation. Insets depict the zoomed version of both distributions. Green crosses indicate the theoretical complex-valued PLV. (D, E) Histograms of real and imaginary parts of scaled residuals for simulations (D) without coupling and (E) with coupling. Green lines indicate the theoretical predictions of corresponding distributions according to corollaries 2 and 4, and the bars indicate the empirical distributions. Note the subtle difference between real and imaginary parts in panels D versus E. See Table 1 for parameters used for this figure.
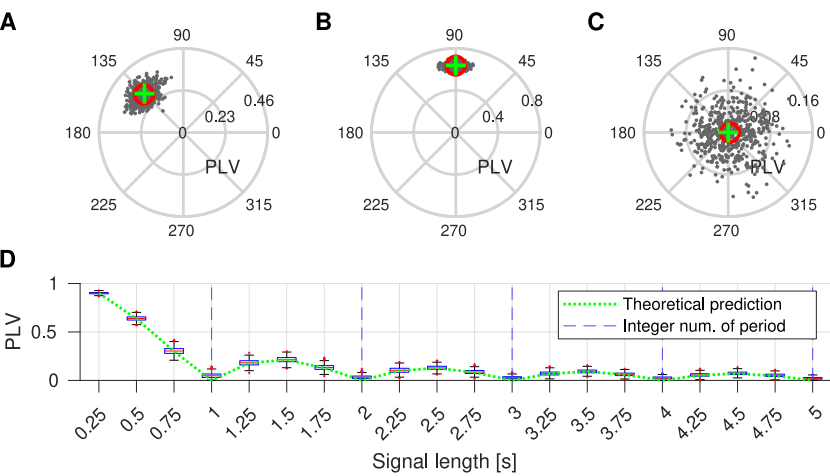
Figure 3: (A–C) Distribution of simulated complex-valued PLVs (gray dots), average of the simulated PLVs (red circle), and theoretical prediction based on equation 3.11 (green crosses) for (A) $\gamma_T = 0.75$, (B) $\gamma_T = 0.5$, and (C) $\gamma_T = 1$. All complex-valued PLVs are represented in the complex plane. Angles indicate the locking phase and the radius the PLV. (D) PLV for different interval lengths T. Box plots represent the simulated PLVs, and the dashed green trace represents theoretical prediction of the expectation based on equation 3.11. Vertical broken blue lines indicate integer number of oscillation periods. See Table 2 for parameters used for this figure.

showing that this effect disappears when the observation window covers a larger number of oscillation periods.

4 Assessment of Multivariate Coupling

High-dimensional data sets have become increasingly important in biology (Bühlmann, Kalisch, & Meier, 2014). More specifically in neuroscience, state-of-the-art multichannel electrophysiology recording systems (Dickey, Suminski, Amit, & Hatsopoulos, 2009; Jun et al., 2017; Juavinett, Bekheet, & Churchland, 2019) allow the simultaneous recording of thousands of sites (Pesaran et al., 2018; Jun et al., 2017; Buzsáki, 2004; Fukushima, Chao, & Fujii, 2015). This growth in dimensionality requires the development of appropriate tools (Stevenson & Kording, 2011; O'Leary, Sutton, & Marder, 2015; Gao & Ganguli, 2015; Williamson, Doiron, Smith, & Yu, 2019) for computing an interpretable summary of the coupling between neurophysiological quantities reflecting the collective dynamics of the underlying neural ensembles (Truccolo, 2016; Safavi et al., 2020). To achieve this aim, deriving low-rank approximations of high-dimensional matrices is supported

by empirical evidence and theoretical predictions of the existence of low-dimensional structures in neural activity (Ermentrout & Kleinfeld, 2001; Ermentrout & Pinto, 2007; Truccolo, Hochberg, & Donoghue, 2010; Gallego, Perich, Miller, & Solla, 2017; Mastrogiuseppe & Ostojic, 2018; Sohn, Narain, Meirhaeghe, & Jazayeri, 2019; Cueva et al., 2020). This section provides statistical results for such approximation in the context of the coupling between point processes and continuous signals.

As a natural extension of the scalar case discussed in the previous section, we now consider the expected coupling matrix C^* between an n-dimensional vector of counting processes N with associated intensity vector $\lambda(t)$ and a multivariate p-dimensional signal $x(t)$, and its estimate based on independent trials $\widehat{C}_K$, respectively defined as

$$C^* = \int_0^T x(t)\lambda(t)^\top dt \quad \text{and} \quad \widehat{C}_K = \frac{1}{K}\sum_{k=1}^K \int_0^T x(t)dN^{(k)}(t)^\top. \tag{4.1}$$

In this multivariate setting, the coupling matrix between the point process and continuous signal can be characterized by the singular value(s) of C^*,

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0,$$

and associated orthonormal singular vectors $\{(u_k, v_k)\}$, such that

$$C^* = \sum_{k=1}^p u_k \sigma_k v_k^H.$$

When the dimension of the coupling matrix gets large, recovering the entire structure of C^* using its estimate $\widehat{C}_K$ becomes unlikely due to the fluctuations of individual coupling coefficients investigated in the previous section. However, the largest singular values may remain reliably estimated because they correspond to low-rank structures of the matrix that stand out from the noise. Random matrix theory provides justifications for this approach by characterizing the spectral properties of "noisy" matrices. Up to a normalization explained later, this will involve indirectly characterizing the behavior of the empirical singular values $\{\widehat{\sigma}_k\}$ of the estimate matrix $\widehat{C}_K$ by analyzing the eigenvalues of the hermitian matrix $\frac{1}{n}\widehat{C}_K\widehat{C}_K^H$ denoted

$$\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p \geq 0.$$

These are related to each other by the relation $\widehat{\sigma}_k = \sqrt{n\ell_k}$ for all k.

4.1 Mathematical Formulation. We now replace assumption 1 to adapt to this multivariate setting. By restricting ourselves to homogeneous

Poisson processes, we investigate a null hypothesis of no coupling between continuous signals and point processes. Let us denote $\bar{x}$ the complex conjugate of x and δ the Kronecker delta symbol:

$$\delta_{lj} = \begin{cases} 1, & \text{if } l = j, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Assumption 2 (Complex Multivariate Case). We consider an infinite sequence $\{x_j(t)\}_{j \geq 1}$ of complex valued left-continuous deterministic functions uniformly bounded on $[0, T]$ and assume

(1) For all $i, j \geq 1$, $\frac{1}{T} \int_0^T \bar{x}_i x_j dt = \delta_{ij}$ and $\int_0^T x_i x_j dt = 0$.
(2) For all $i \geq 1$, $\int_0^T x_i dt = 0$,
(3) There exist $0 < \lambda_{\min} < \lambda_{\max}$ and a sequence of independent homogeneous Poisson processes $\{N_i\}_{i \in \mathbb{N}^*}$'s with associated rates $\{\lambda_i\}_{i \in \mathbb{N}^*}$ in the interval $[\lambda_{\min}, \lambda_{\max}]$.

While the assumptions on $\{x_i(t)\}$ are designed for complex signals, which is the classical case when dealing with PLV-like quantities, the results of this section also hold for real signals by using the assumption $\frac{1}{T} \int_0^T x_i x_j dt = \delta_{ij}$ instead of the above condition 1. Condition 2 is also added to ensure that there is no trivial bias leading to a nonvanishing expectation of the coupling coefficients (as illustrated in section 3.2). Indeed, when the time average of each signal vanishes, based on theorem 1, the expectation of all univariate coupling measures for a homogeneous Poisson process vanishes. We then exploit a multivariate generalization of the martingale CLT to characterize the distribution of the coupling matrix given these assumptions.

Theorem 2. *For given* n, $p \geq 1$ *and all* $K \geq 1$, *we use sequences of signals defined in assumption 2 to build multivariate continuous signal* $\boldsymbol{x}(t) = (x_j)_{j=1\dots p}$ *and* K *independent copies of multivariate Poisson process* $\boldsymbol{N}(t) = (N_i)_{i=1\dots n}$ *with rate vector* $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]^\top$. *Then the normalized coupling matrix* $\sqrt{K}\widehat{\boldsymbol{C}}_K \text{diag}(\sqrt{T\boldsymbol{\lambda}})^{-1}$, *with* $\widehat{\boldsymbol{C}}_K$ *given by equation 4.1, converges in distribution for* $K \to +\infty$ *to a matrix with i.i.d. complex standard normal coefficients.*

Sketch of the proof. This essentially uses a generalization of the CLT to multivariate point processes described in Aalen et al. (2008, appendix B). Based on the statistics of stochastic integrals presented in section B.1.1, assumptions on x entail vanishing correlations between all matrix coefficients and lead to the analytical expression of the covariance matrix. $\square$

This result suggests that for large n and $p = p(n)$, coupling matrices $\widehat{C}_K^n$ of increasing size can be used to build the Wishart-like matrix sequence,

$$S_n \triangleq \frac{K}{n} \widehat{C}_K^n \text{diag}(T\boldsymbol{\lambda})^{-1} (\widehat{C}_K^n)^H, \tag{4.3}$$

whose ESD may converge to the Marchenko-Pastur law. This is, however, not guaranteed by classical results due to the nongaussianity and dependence of the matrix coefficients of $\widehat{C}_K^n$ for fixed n and K. Convergence will thus depend on how much the departure from these assumptions plays a role as n becomes large. We show in the following theorem that increasing the number of trials as a function of the dimension guarantees convergence to the MP law.

Theorem 3. *In addition to assumption 2, assume an increasing, positive integer sequence $\{p(n), K(n)\}_{n \in \mathbb{N}^*}$ such that $\frac{p(n)}{n} \xrightarrow[n \to +\infty]{} \alpha \in (0, +\infty)$, and*

$$\frac{1}{n^2 K(n)^2} \sum_{\Gamma} \left(\int_0^T \bar{x}_j x_l x_{j'} \bar{x}_{l'} dt \right)^2 \to 0, \ \ uniformly \ in \ k \leq n, \tag{4.4}$$

where $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l' \ or \ j = l' \neq j' = l\}$. Consider the sequence $\{\widehat{C}_{K(n)}^n\}_{n \in \mathbb{N}^*}$ built as in theorem 2 for $p = p(n)$; then the corresponding sequence $\{S_n\}$ defined by equation 4.3 has an ESD converging weakly with probability one to the MP law of equation 2.7.

Sketch of the proof. We use theorem 1.1 of Bai and Zhou (2008) addressing the case of matrices with dependence of coefficients within columns. We use Itô's formula (see appendix B) to check the simplified necessary conditions provided in corollary 1.1 of Bai and Zhou (2008). This implies convergence of the Stieltjes transform to the same function as the transform of the MP distribution. By classical results on the Stieltjes transform (Anderson et al., 2010, theorem 2.4.4), this implies weak convergence to the MP measure (i.e., convergence for the weak topology; see appendix B.2). □

Remark 2. Condition in equation 4.4 determines how many trials are needed at most for spectral convergence. Due to the uniform boundedness assumption on signal $x(t)$ and given the number of terms in the sum bounded by n^4, we can already see that $\frac{n}{K(n)} \to 0$, that is, having the number of trials increasing at an even slightly faster rate than the dimension is enough for convergence for any choice of continuous signals respecting orthonormality assumption 2. However, there are cases where even fewer trials than dimensions are required. An important example is the Fourier basis of the $[0, T]$ interval, $x_l(t) = \frac{1}{\sqrt{T}} \exp(i 2\pi l t / T)$. Then all terms in the sum of equation 4.4 vanish, except the ones satisfying $j - j' - l + l' = 0$, such that we are left with a number of bounded terms that scale with n^3. As a consequence, the condition on the number of trials to achieve spectral convergence becomes $\frac{\sqrt{n}}{K(n)} \to 0$, such that we need increasingly fewer trials than dimensions. On the contrary, due to the uniform bound that we impose on the signals, choosing a basis of signals with decreasing support, such as a wavelet basis, typically departs from our condition 1 of assumption 2, as the normalization of condition 1 of assumption 2 imposes unit

norm on each signal, requiring their amplitude to increase as their support decreases, violating the uniform bound assumption. This limitation supports the intuition that statistical regularities exploited by our asymptotic results deteriorate with highly transient signals.

This convergence of the spectral measure to the MP law guarantees eigenvalues do not accumulate in a large proportion above the upper end of the support of the MP law; however, they do not provide rigorous guarantees regarding convergence of individual eigenvalues and, in particular, the largest eigenvalue. Although such convergence is satisfied in classical settings (gaussian i.i.d. coefficients), they typically require stronger assumptions than for the (weak) spectral convergence to the MP law, and still only very few results are available in the non-i.i.d. setting. We could, however, prove such convergence by adding a constraint to our model.

Theorem 4. *In addition to assumption 2, assume all homogeneous rates λ_k are equal. Assume two increasing, positive integer sequences $\{p(n), K(n)\}_{n \in \mathbb{N}^*}$ such that*

$$\frac{p(n)}{n} \to \alpha \in (0, +\infty) \quad and \quad \frac{1}{K(n)} \sum_{1 \leq i, k \leq p(n)} \int_0^T |x_i x_j|^2(t) dt < B, \quad (4.5)$$

for some constant B. Then S_n defined in equation 4.3 has an ESD converging weakly with probability one to the MP law of equation 2.7. Moreover, let ℓ_1 and ℓ_p be the largest and smallest eigenvalues of $\{S_n\}$, respectively. Then in probability

$$\ell_1(n) \to (1 + \sqrt{\alpha})^2 \quad and \quad \ell_p(n) \to (1 - \sqrt{\alpha})^2 \mathbf{1}_{\alpha < 1}.$$

Sketch of the proof. The identical intensities assumption allows us to use the result of Chafaï and Tikhomirov (2018) for matrices with i.i.d. columns. We first checked that their proof holds also for the complex case by replacing symmetric matrices by Hermitian matrices and squared scalar product by an absolute squared Hermitian product. We satisfy their strong tail projection (STP) assumption using Chebyshev's inequality. The necessary fourth-order moment conditions exploit the same stochastic integration results as theorem 3. □

Remark 3. Without additional assumptions, the moment condition of equation 4.5 is satisfied by choosing $K(n) = n^2$ (as there are p^2 bounded moments, scaling as n^2 when n grows). It is likely from the proof that taking into account more information about the moments of the continuous signal sequence $\{x_j\}$, we can achieve convergence with a lower rate of increase for the number of trials. This is left to future work.

This result thus provides the guarantees that under a null hypothesis of no coupling (due to homogeneity of the Poisson processes), the extreme eigenvalues of S will asymptotically cover exactly the full support

of the MP law. This will be used in section 4.2 to assess the significance of the eigenvalues ℓ_k by simply checking whether they are larger than $(1 + \sqrt{\alpha})^2$.

This significance analysis relies as well on understanding what happens to the eigenvalues when the model departs from the null hypothesis. In a practical setting, we hypothesize that the coupling matrix has a deterministic structure superimposed on the martingale noise modeled in the above results. One qualitative justification of this assumption can be found in remark 1, showing that for sinusoidal coupling, a nonvanishing expectation proportional to the coupling is superimposed to martingale noise, whose distribution is unaffected by coupling, such that the noisy part of the matrix satisfies the conditions of the above theorems. As typically done in applications, we are mostly interested in the low-rank structure associated with the largest singular values of the coupling matrix, providing an interpretable summary of the multivariate interactions.

This naturally leads to a modeling departure from the null hypothesis with a low-rank perturbation assumption. In such a case, the eigenvalues related to significant coupling are expected to be reflected in the spectrum of the perturbed matrix, such that they can be isolated from the remaining eigenvalues associated with the martingale noise. This intuition is justified by results in the case of the Wishart ensemble (Loubaton & Vallet, 2011); see also Benaych Georges and Nadakuditi (2012) for a more general result and Capitaine and Donati Martin (2016) for an overview of matrix perturbation results), that we restate here:

Theorem 5 (From Loubaton & Vallet, 2011, Theorem 6). *Let $\boldsymbol{X}_n$ be an $n \times p$ sequence of i.i.d. complex gaussian matrices defined in section 2.3 and A_n be a finite rank perturbation of the null matrix with nonzero eigenvalues θ_i. Let $\boldsymbol{M}_n = (\frac{1}{\sqrt{n}}\boldsymbol{X}_n + A_n)(\frac{1}{\sqrt{n}}\boldsymbol{X}_n + A_n)^H$. Then as $n \to \infty$ and $\frac{p}{n} \to \alpha \in (0, 1)$, almost surely,*

$$
\lambda_i(M_n) \to \begin{cases} \frac{(1+\theta_i)(c+\theta_i)}{\theta_i}, & \text{if } \theta_i > \sqrt{\alpha}, \\ (1 + \sqrt{\alpha})^2, & \text{otherwise.} \end{cases}
$$

A demonstration that this further applies rigorously to our nongaussian, non-i.i.d. case is left to further work (but see Benaych-Georges & Nadakuditi, 2012, for a generalization in this direction). This result shows the upper end of the MP support is indeed the critical threshold for the eigenvalues of A_n to stand out from the noise. Below this threshold, the largest eigenvalue convergence to the upper end of the support of the MP distribution is not informative about θ_i. Above this threshold, the value of θ_i can be recovered and detected by comparing the largest eigenvalue to the upper end of the MP distribution.

We next illustrate the interest of these theoretical predictions in the context of neural time series for reliably quantifying the interplay between multichannel LFP signals and the spiking of multiple neurons. Nevertheless, the results are potentially applicable in other domains as well. In neuroscience, x may represent LFP measurements collected on each recording channel and N the spiking activity of different neurons, called units. The number of recording channels n_c and recorded units n_u correspond to p and n, respectively. These numbers may differ, and as a consequence, the coupling matrix is generally rectangular.

4.2 Application to Significance Assessment. In order to statistically assess the significance of the largest singular value(s) of coupling matrix $\widehat{C}_K^n$, considered as a measure of coupling between point processes and continuous signals, we need a null hypothesis. Hypothesis testing based on the generation of surrogate data is one of the common methods for significant assessment in neuroscience and other fields. Generating appropriate surrogate data can be not only challenging (see Grün, 2009, and Elsayed & Cunningham, 2017, for examples in neuroscience), but also computationally expensive due to the increasingly large dimensions of modern data sets. Exploiting our theoretical results for this setting allows us to perform such statistical assessment in a principled way, without using surrogate data and sparing computational resources.

In order to exploit the results of the theoretical part, it is best to preprocess the $p \times q$ matrix of time-discretized signals $\mathbf{L}$ that correspond to q samples over interval $[0, T]$ with sampling interval $\Delta = T/q$. The chosen signals are driven by the application (in our case, they are preprocessed LFPs, see section 4.3 for a simulation reproducing the context of neurophysiolgy data). We assume the rows of $\mathbf{L}$ sum to zero to match condition 2 of assumption 2 (and avoid bias in the coupling measure similar to what is described in section 3.3). We then need to process further this signal such that condition 1 of assumption 2 is satisfied approximately. In order to achieve this, we perform classical whitening of the signals to generate matrix $\mathbf{X}$, the discrete time approximation of $x(t)$, according to

$$\mathbf{X} = \mathbf{WL}, \quad \text{with} \quad \mathbf{W} = \left(\frac{1}{q}\mathbf{LL}^{\mathrm{H}}\right)^{-\frac{1}{2}}, \tag{4.6}$$

where the power in the expression of the whitening matrix $\mathbf{W}$ describes the inversion of a matrix square root, typically achieved via eigenvalue decomposition, and which may require PCA-like dimensionality reduction in practice to minimize the numerical effects of small eigenvalues. This procedure decorrelates the martingale fluctuations of coefficients within the same column of the coupling matrix (see theorem 2), a key requirement for convergence to the MP law.
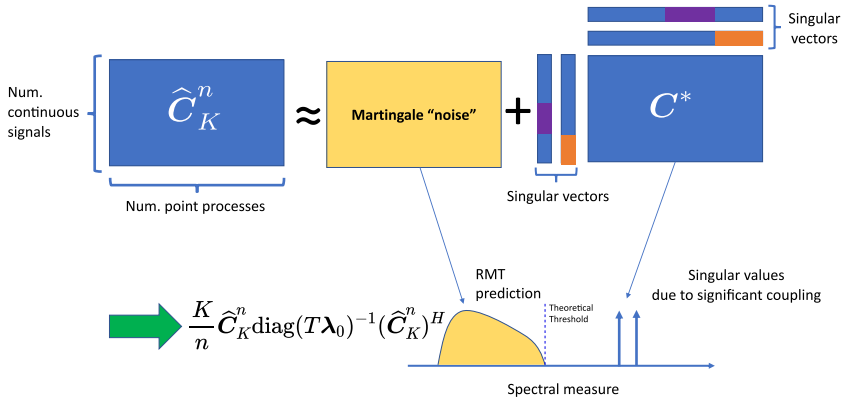
Figure 4: We assume $\widehat{C}_K^n$ is a superposition of martingale noise and a low-rank deterministic matrix C^* reflecting the actual coupling. If the singular values of a normalized version of C^* are large enough (larger than the upper end of the MP law support), theory suggests that they will correspond to the largest eigenvalues of S_n appearing beyond the support of MP distributed eigenvalues reflecting martingale noise. They can thus be detected with a simple thresholding approach (see equation 4.7).

As explained in section 4.1, theoretical results support using $\theta_{DET} = (1 + \sqrt{\alpha})^2$, the upper end of the support of the MP law, as a detection threshold for the significance of the eigenvalues of the Hermitian matrix,

$$S_n = \frac{K}{n}\widehat{C}_K^n \operatorname{diag}(T\boldsymbol{\lambda}_0)^{-1}(\widehat{C}_K^n)^H.$$

The null hypothesis of nonsignificance of the kth largest singular value $\widehat{\sigma}_k$ of the normalized coupling matrix,

$$\sqrt{K}\widehat{C}_K^n \operatorname{diag}(\sqrt{T\boldsymbol{\lambda}_0})^{-1},$$

should thus be rejected if the corresponding kth largest eigenvalue ℓ_k of S_n is superior to the significance threshold, leading to the condition

$$\widehat{\sigma}_k = \sqrt{n\ell_k} > \sqrt{n\theta_{DET}} = \sqrt{n}(1 + \sqrt{\alpha}). \tag{4.7}$$

Therefore, this last condition on the empirical singular values is used to identify those reflecting a significant coupling between the multivariate point process and continuous signal. An illustration of our overall significance assessment approach is shown in Figure 4.

4.3 Simulation. We use a simulation to demonstrate the outcome of our (asymptotic) theoretical results on mutlivariate coupling. Similar to the simulations of section 3.3 for the univariate case, we use simulated phase-locked spike trains with Poisson statistics. The main difference between this simulation and the previous one is in synthesizing the LFP. In order to simulate multichannel oscillatory signals that lead to a low-rank structure for C^*, we use a combination of noisy oscillatory components.

The LFPs contain N_{osc} oscillatory groups of channels; each channel l within the same group contains the same oscillatory component with index $j(l)$, with the time course of all these components being $O_j(t) = e^{2\pi i f_j t}$, $j \in \{1, \ldots, N_{osc}\}$, with all frequencies f_j in the range $[f_{\min}, f_{\max}]$, and all multiples of $1/T$. Due to the necessary time axis discretization, the bracket notation $[t]$ indicates the oscillation is sampled at equispaced discrete times $t = \{k\Delta\}_{k=1,\ldots,q}$. The synthesized discrete time multichannel LFP ($\Psi[t] = \{\psi_l[t]\}_{l=1,\ldots,n_c}$) can be written as

$$\Psi_l[t] = O_{j(l)}[t] \odot \exp(i\eta_l[t]), \tag{4.8}$$

with $\odot$ entrywise product and $\{\eta_l[t])\}$ i.i.d. sampled (white) phase noises contaminating each channel independently (see appendix D for more details).

In this simulation, the frequencies of the oscillatory components range from 11 to 15 Hz. We used 100 LFP channels ($n_c = 100$) and different choices for the number of spiking units (10, 50, and 90). Spiking activities are simulated in different scenarios, with and without coupling to the LFP oscillations. In the latter case, we have two populations of neurons (each consisting of 1/5th of the total number of neurons) that are each coupled to one of the oscillatory groups of LFP channels. Both populations are coupled to their respective oscillation with identical strength ($\kappa = 0.15$) and phase ($\phi_0 = 0$).

To compute the coupling matrix $\widehat{C}_K$, we first preprocess $\Psi[t]$ by applying bandpass filtering in a range covering $[f_{\min}, f_{\max}]$ and convert it to an analytic signal via the discrete time Hilbert transform, leading to data matrix $\mathbf{L}$, following the standards of PLV analysis in neuroscience (Chavez, Besserve, Adam, & Martinerie, 2006).

This signal matrix is then whitened according to equation 4.6 to yield matrix $\mathbf{X}$, the discrete time version of $x(t)$. The coupling matrix $\widehat{C}_K$ is then computed according to equation 4.1 using 10 trials (barring trivial approximation to the closest time sample in $\mathbf{X}$).

Then in order to approximate the normalization $\sqrt{K}\widehat{C}_K^n \text{diag}(\sqrt{T\lambda_0})^{-1}$ based on empirical data, we use the total number of events for unit u occurring across all K trials $N_{tot}^u = \sum_{k=1}^{K} N_k^u$ and multiply each column u of the coupling matrix by
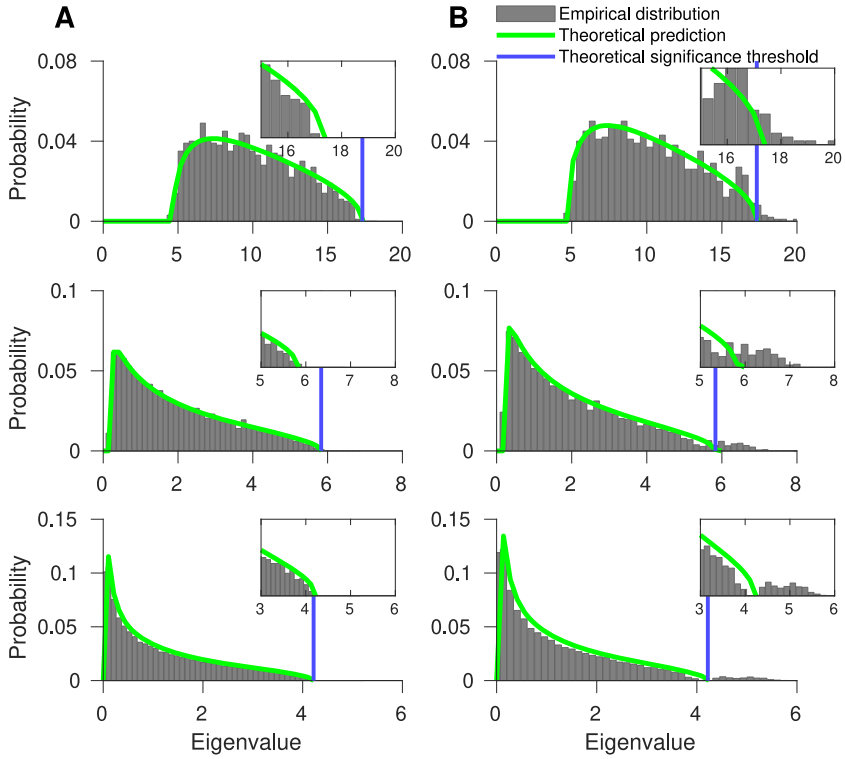
Figure 5: Theoretical Marchenko-Pastur distribution (green lines) and empirical distribution (gray bars) for (A) simulation without coupling ($\kappa = 0$) and (B) with coupling ($\kappa = 0.15$) between multivariate spikes and LFP. Rows represent the spectral distribution of simulations with different number of spiking units—rows 1, 2, and 3, respectively, 10, 50, and 90 (which leads to different α for MP law). Insets zoom the tail of the distributions. Parameters used for this figure are denoted in Table 3.

$$\frac{K}{\sqrt{N_{\text{tot}}^u}} \approx \frac{K}{\sqrt{K \int_0^T \lambda_u(t)dt}}$$

for the corresponding unit u, asymptotically matching the theoretical normalization in the homogeneous Poisson case.

We observe in Figure 5A that in the absence of coupling, the distribution of eigenvalues originating from the random matrix structure is very close to the theoretically predicted MP distribution. In Figure 5B, we have coupling between spike and eigenvalues reflecting the coupling beyond the MP support (blue line in Figure 5), and the eigenvalue bulk below the threshold is

also close to MP distribution. This suggests an easy thresholding approach for significance assessment.

5 Discussion

5.1 Insights for Data Analysis. Our theoretical results provide guarantees for specific coupling models to respect univariate and multivariate asymptotic statistics that can be easily exploited for statistical testing. The required assumptions provide guidelines for practical settings that are likely of interest beyond the strict framework that we imposed to get the rigorous results. For the univariate coupling measure, corollaries and simulations point out the importance of the choice of the observation interval $[0, T]$, which is particularly sensitive when considering short intervals covering only a few oscillation periods. This is the case when doing time-resolved analysis or dealing with experiments with short trial durations. Moreover, the univariate results also emphasize the effect of nonlinear phase increases, highly relevant in neuroscience due to the pervasive effects of nonlinear dynamics in the mesoscopic signals. Our results provide asymptotic bias correction terms that can be used for statistical testing.

In the same way, theoretical results in the multivariate setting may seem to be constrained by our assumptions, but they provide critical guidelines to interpret singular values. First, whitening the continuous signals and normalizing the coupling by the square root of the rate are key preprocessing steps for making the asymptotic behavior of the martingale noise invariant to the specifics of the data at hand. This then reduces to an analytical model, the MP law, dependent on only a single matrix shape parameter. After assessment of the significance of the singular value of the normalized coupling matrix, it is of course possible to revert these preprocessing steps to get a low-rank approximation of the original coupling matrix (nonnormalized, nonwhitened) to summarize the significant coupling structure in an interpretable way. A second insight provided by the multivariate results is the role of fourth-order moments of the continuous signals, represented by the integrals of order four monomials of components of $x(t)$, in the MP convergence results. The magnitude of these moments determines the number of trials asymptotically needed to achieve convergence. Since these moments can be estimated empirically, we can check how they grow with the dimension of the signals in a specific application. With our minimal assumptions on the signals, the number of trials need only to grow at most sublinearly in the dimension for spectral convergence; however, we could only show that convergence of the largest eigenvalues requires at most quadratic increase in the dimension n. This last result might be improved in future work, with extra assumptions, to reach linear growth.

Our theoretical results can be extended in two directions. The first is toward exploiting point processes different from inhomogeneous Poisson

(e.g., Hawkes processes) in order to be able to apply the framework in the context of stochastic intensities. The second direction is toward exploiting recent developments in RMT, in order to develop a probabilistic significance assessment.

5.2 Extension of Signal Assumptions. Our theoretical results assume deterministic continuous signals and point process intensities (see assumption 1). This entails limitations, such as implicitly assuming the considered point processes are (homogeneous or inhomogeneous) Poisson processes. This assumption may be too restrictive in realistic scenarios (for examples in neuroscience, see Deger, Helias, Boucsein, & Rotter, 2012; Reimer, Staude, Ehm, & Rotter, 2012; Nawrot et al., 2008; Shinomoto, Shima, & Tanji, 2003; Maimon & Assad, 2009; Shinomoto et al., 2009). However, the stochastic integration methods that provide the basis of our results allow the treatment of random signals and intensities, provided they are predictable, which encompasses a wide enough class of processes to cover most applications (Protter, 2005). Our results thus have potential for generalizations to the case of random continuous signal, with the difference that the variance of the estimates would increase due to the additional variability induced by the signal fluctuations, and to the case of random intensities, leading to different asymptotic properties of the coupling measures, which may or may not have simple analytical expressions.

As a potential direction for extending this framework, Hawkes processes (Hawkes, 1971) are point processes for which the probability of occurrence of future events can also depend on the sequence of past events. Due to this history dependency, they are also called self-exciting processes. Hawkes processes are used for modeling recurrent interactions in various fields; for instance—in finance it is used to model buy or sell transaction events on stock markets (Embrechts, Liniger, & Lin, 2011) in geology to model the origin times and magnitudes of earthquakes (Ogata, 1988), in online social media to model user actions over time (Rizoiu, Lee, Mishra, & Xie, 2017), and even modeling reliability of information on the web and controlling the spread misinformation (Tabibian et al., 2017; Kim, Tabibian, Oh, Schölkopf, & Gomez-Rodriguez, 2018), and in neuroscience to model spike trains (Krumin et al., 2010). We conjecture that such history dependency can be incorporated in our analytic treatment of the coupling measure, such that our theoretical results can be extended to this model.

5.3 Extension beyond Binary Significance Assessment. We show that the Marchenko-Pastur distribution provides a good approximation of the distribution of eigenvalues in the absence of coupling, and the upper end of its support approximates the largest eigenvalue. This provides us a threshold to assess the significance of empirical singular values. Nevertheless, this hard thresholding approach does not take into account the actual fluctuations of the largest eigenvalue around this asymptotic upper end of the

support and thus does not provide meaningful *p*-value for the statistical test.

It has been shown that the appropriately rescaled and recentered[8] largest eigenvalue of Wishart matrices is asymptotically distributed as the Tracy-Widom distribution—for example, see Johnstone (2001); Tracy and Widom (2002); and El Karoui (2003, 2005, 2007). However, note that in some cases of practical relevance, the normal distribution might be more appropriate (Bai & Yao, 2008). Such asymptotic distribution of the largest eigenvalue can be exploited for reporting a theoretical *p*-value for the significance of the coupling and therefore extending the significance assessment from a binary decision to a probabilistic one. For example, Kritchman and Nadler (2009) exploit this idea (but in a simpler scenario) to determine the number of signal components in noisy data. This extension would allow a precise probabilistic assessment of the significance of weaker couplings leading to eigenvalues in the neighborhood of the asymptotic threshold introduced above.

6 Conclusion

We investigated the statistical properties of coupling measures between continuous signals and point processes. We first used martingale theory to characterize the distributions of univariate coupling measures such as the PLV. Then, based on multivariate extensions of this result and RMT, we established predictions regarding the null distribution of the singular values of coupling matrices between a large number of point processes and continuous signals and a principled way to assess significance of such multivariate coupling. These theoretical results build a solid basis for the statistical assessment of such coupling in applications dealing with high dimensional data.

Appendix: Proofs of Theorems in the Main Text

Proof of Theorem 1. For the first part of the theorem (expectation), we use the martingale $M^{(k)}$ associated with each copied process $N^{(k)}$ to rewrite

$$\widehat{c}_K = \frac{1}{K} \sum_{k=1}^{K} \int_0^T x(t) dM^{(k)}(t) + \frac{1}{K} \sum_{k=1}^{K} \int_0^T x(t) \lambda(t) dt(t). \tag{A.1}$$

Elements of the sum in the first term are then zero mean martingale, and by linearity, so is the whole term. As a consequence (using the zero mean

---

[8]The required recentering and rescaling of the eigenvalues is studied in the literature (Johnstone, 2001; El Karoui, 2003, 2007).

property), the expectation of the first term is zero so only the second term remains:

$$\mathbb{E}\left[\hat{c}_K\right] = \int_0^T x(t)\lambda(t)dt(t).$$

We then exploit a central limit theorem (CLT) for martingales to prove the second part of the theorem (convergence to gaussian distribution). To satisfy the CLT in such a case, it is sufficient to find a particular martingale $\widetilde{M}^{(K)}$ sequence that will satisfy the conditions described in Aalen et al. (2008, p. 63) ($\xrightarrow{P}$ indicate convergence in probability):

1. $\mathrm{Var}(\widetilde{M}^{(K)}(t)) \xrightarrow[K\to+\infty]{P} \widetilde{V}(t)$ for all $t \in [0, T]$, with $\widetilde{V}$ increasing and $\widetilde{V}(0) = 0$.
2. Informally, the size of the jumps of $\widetilde{M}^{(K)}$ tends to zero (see Aalen et al., 2008, p. 63). Formally, for any $\epsilon > 0$, the martingale $\widetilde{M}_\epsilon^{(K)}(t)$ gathering the jumps $> \epsilon$ satisfies $\mathrm{Var}\left(\widetilde{M}_\epsilon^{(K)}(t)\right) \xrightarrow[K\to+\infty]{P} 0$.

Then $\widetilde{M}^{(K)}(t)$ converges in distribution to a gaussian martingale of variance $\widetilde{V}(t)$.

To achieve these conditions, we define $M^{(k)}$, the sequence of i.i.d. zero mean martingales defined on $[0, T]$ canonically associated with the point process of each trial $N^{(k)}$. Then we build martingales $M_x^{(k)}(t) = \int_0^t x(s)dM^{(k)}(s)ds$ and construct $\widetilde{M}^{(K)} = 1/\sqrt{K}\sum_{k=1}^K M_x^{(k)}$.

The variance of this latter martingale (also called its predictable variation processes) can be computed based on the rules provided in section B.1.1. First, due to trial independence,

$$\widetilde{V}(t) = \mathrm{Var}\left(\widetilde{M}^{(K)}(t)\right) = \mathrm{Var}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^K M_x^{(k)}(t)\right) = \sum_{k=1}^K \mathrm{Var}\left(\frac{1}{\sqrt{K}}M_x(t)\right),$$

$$(A.2)$$

and using equation B.4, we get

$$\widetilde{V}(t) = \frac{1}{K}\sum \int_0^t x^2(t)\lambda(t)dt = \int_0^t x^2(t)\lambda(t)dt. \qquad (A.3)$$

Equation A.3 clearly fulfills CLT's condition 1.

For condition 2, due to assumption 1, $x(t)$ is bounded, such that there is a $B > 0$ satisfying $|x(t)| < B$ over $[0, T]$. As a consequence, the size of all jumps is bounded by $B/\sqrt{K}$, and for any ϵ, $\widetilde{M}_\epsilon^{(K)}(t)$ is the constant zero for $K > \frac{B^2}{\epsilon^2}$ and condition 2 is satisfied.

Fulfillment of both conditions leads to convergence in distribution to a gaussian martingale of variance $\widetilde{V}(t)$;

$$\widetilde{M}^{(K)} \underset{K \to +\infty}{\longrightarrow} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right). \tag{A.4}$$

Finally, using equation A.1, we conclude the proof by noticing that the above martingale corresponds exactly to the quantity $\sqrt{K}\,(\widehat{c}_K - c^*)$. Therefore,

$$\sqrt{K}\,(\widehat{c}_K - c^*) \underset{K \to +\infty}{\longrightarrow} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right). \tag{A.5}$$

$\square$

Proof of Corollary 1. We apply theorem 1 to $e^{i\phi(t)}$ (i.e., replacing $x(t)$ with $e^{i\phi(t)}$). As $e^{i\phi(t)}$ is complex valued, we should have a covariance function for its predictable variation process $\widetilde{V}(t)$. The covariance between a martingale's real part,

$$M_{\text{Re}}(t) = \int_0^t \text{Re}(e^{i\phi(s)})dM(s)ds,$$

and imaginary part,

$$M_{\text{Im}}(t) = \int_0^t \text{Im}(e^{i\phi(s)})dM(s)ds,$$

is given by

$$\int_0^t \text{Re}(e^{i\phi(s)})\text{Im}(e^{i\phi(s)})\lambda(s)ds. \tag{A.6}$$

The diagonal elements of the covariance function are the predictable variation process of M_{Re} and M_{Im} that can be computed based on equation B.4, and the off-diagonal elements are the covariance between martingale's real and imaginary part that can be computed based on equation B.5. Therefore, the covariance function for its predictable variation process is

$$\text{Cov}\left(\begin{bmatrix} \text{Re}\{Z\} \\ \text{Im}\{Z\} \end{bmatrix}\right)$$
$$= \begin{bmatrix} \int_0^t \left(\text{Re}(e^{i\phi(s)})\right)^2 \lambda(s)ds & \int_0^t \text{Re}(e^{i\phi(s)})\text{Im}(e^{i\phi(s)})\lambda(s)ds \\ \int_0^t \text{Re}(e^{i\phi(s)})\text{Im}(e^{i\phi(s)})\lambda(s)ds & \int_0^t \left(\text{Im}(e^{i\phi(s)})\right)^2 \lambda(s)ds \end{bmatrix} \tag{A.7}$$

$$= \int_0^t \begin{bmatrix} \cos^2(\phi(s)) & \sin(2\phi(s))/2 \\ \sin(2\phi(s))/2 & \sin^2(\phi(s)) \end{bmatrix} \lambda(s)ds. \tag{A.8}$$

Similar to theorem 1, as $K \to +\infty$, the residuals converge in distribution to a zero-mean complex gaussian variable Z (i.e., the joint distribution of real and imaginary parts is gaussian):

$$\sqrt{K}\,(\widehat{c}_K - c^*) \xrightarrow[K \to +\infty]{} \mathcal{N}\,(0, \text{Cov}(Z))\,.$$

Because theorem 1 guarantees that the $\sqrt{K}(\widehat{c}_K - c^*)$ tends to a gaussian with finite variance, $\widehat{c}_K$ tends to the Dirac measure at c^*.

However, given that we use $x(t) = e^{i\phi(t)}$, $\widehat{c}_K$ is not exactly the multitrial PLV estimate—more precisely,

$$\widehat{c}_K = \frac{1}{K} \sum_{k=1}^{K} \int_0^T e^{i\phi(t)} dN^{(k)}(t) = \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{N_k} e^{i\phi(t_j^k)} = \frac{\left(\sum_{k=1}^{K} N_k\right)}{K} \widehat{\text{PLV}}_K.$$

Thus, we can write $\widehat{\text{PLV}}_K = \nu_K \cdot \widehat{c}_K$, with $\nu_K = \frac{K}{\left(\sum_{k=1}^{K} N_k\right)}$. With the same techniques (using $x(t) = 1$), we can show convergence in the distribution of ν_K to a constant:

$$\frac{1}{\nu_K} = \frac{\left(\sum_{k=1}^{K} N_k\right)}{K} = 1/K \sum_k \int_0^T 1 \cdot dN^{(k)} \xrightarrow[K \to +\infty]{} \int_0^T \lambda(t)dt = \Lambda(T).$$

This leads to

$$\nu_K \xrightarrow[K \to +\infty]{} \frac{1}{\Lambda(T)}.$$

Following a version of Slutsky's theorem (Mittelhammer, 1996, theorem 5.10), since ν_k and $\widehat{c}_K$ tend to a limit in distribution, and one of these limits is a constant, the product tends to the product of the limits such that we get

$$\text{PLV}^* = \lim_{K \to \infty} \nu_K \cdot \widehat{c}_K = \frac{c^*}{\Lambda(T)}$$

and can decompose the PLV residual as follows:

$$\sqrt{K}\left(\widehat{\text{PLV}}_K - \text{PLV}^*\right) = \sqrt{K}\nu_K\,(\widehat{c}_K - c^*) + \sqrt{K}\,(\nu_K c^* - \text{PLV}^*)\,.$$

Taking the limit of the above equation, the second term clearly vanishes (see the above limit of ν_K), and the first term, using again the limit of products, leads to the final result:

$$\sqrt{K}\left(\widehat{\text{PLV}}_K - \text{PLV}^*\right) \xrightarrow[K\to+\infty]{} \mathcal{N}\left(0, \frac{1}{\Lambda(T)^2}\text{Cov}(Z)\right).$$

$\square$

Proof of Corollary 2. We use the intensity function introduced in equation 3.5 in corollary 1. The PLV asymptotic value (PLV*) can be derived from definition introduced in equation 3.2:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{A.9}$$

$$= \frac{r_o \int_0^T e^{i\phi(t)} \exp(\kappa \cos(\phi(t) - \varphi_0))\phi'(t)dt}{r_o \int_0^T \exp(\kappa \cos(\phi(t) - \varphi_0))\phi'(t)dt}. \tag{A.10}$$

We change the integration variable from $\phi(t)$ to θ:

$$\text{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0))d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta - \varphi_0))d\theta}. \tag{A.11}$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ, ($\psi = \theta - \varphi_0$):

$$\text{PLV}^* = \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa \cos(\psi))d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa \cos(\psi))d\psi} \tag{A.12}$$

$$= e^{i\varphi_0} \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i\psi} \exp(\kappa \cos(\psi))d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa \cos(\psi))d\psi}. \tag{A.13}$$

Given that that integrand is a 2π-periodic functions (thus, the integral is invariant to translations of the integration interval), we get

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa \cos(\psi))d\psi}{\int_{-\pi}^{\pi} \exp(\kappa \cos(\psi))d\psi}.$$

Observing that the integrand of the denominator is even, while for the numerator, the imaginary part is odd and the real part is even, we get

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_0^{\pi} \cos(\psi) \exp(\kappa \cos(\psi))d\psi}{\int_0^{\pi} \exp(\kappa \cos(\psi))d\psi}.$$

This proves the first part of the corollary—equation 3.6. By using the integral form of the modified Bessel functions I_k for k integer (see, e.g., Watson, 1995, p. 181):

$$I_k(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta + \frac{\sin(k\pi)}{\pi} \int_0^{+\infty} e^{-\kappa \cosh t - kt} dt \quad \text{(A.14)}$$

$$= \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta, \quad \text{(A.15)}$$

we can derive the compact form:

$$\text{PLV}^* = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}. \quad \text{(A.16)}$$

The covariance matrix of the asymptotic distribution can be easily derived by plugging equation 3.5 as $\lambda(t)$ in corollary 1 and integrating on $[0, T]$:

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \quad \text{(A.17)}$$

Based on the above developments and noticing that the integration intervals correspond to $2\pi\gamma_T$, with γ_T the number of oscillation periods, we have

$$\Lambda(T) = \lambda_0 2\gamma_T \pi I_0(\kappa) = \lambda_0 2 \frac{\phi(T) - \phi(0)}{2\pi} \pi I_0(\kappa),$$

such that

$$(\text{Cov}(Z))_{11} = \frac{1}{\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2}$$

$$\times \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \quad \text{(A.18)}$$

To simplify the rest of the derivations, we transform the complex variable coordinates by using $e^{i\phi(t)} e^{-i\varphi_0}$ instead of $e^{i\phi(t)}$ as predictable with respect to $\{\mathcal{F}_t\}$ (i.e., replacing $x(t)$ with $e^{i\phi(t)} e^{-i\varphi_0}$ in theorem 1). With this change, equation A.18 becomes

$$(\text{Cov}(Z))_{11} = \frac{1}{\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2}$$

$$\times \int_0^T \cos^2(\phi(t) - \varphi_0) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \quad \text{(A.19)}$$

We change the variable of the integral from $\phi(t) - \varphi_0$ to θ and use the following trigonometric identity,

$$\cos^2(\theta) = \frac{1}{2}\left(1 + \cos(2\theta)\right), \tag{A.20}$$

to obtain

$$
\begin{aligned}
(\text{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \int_{\phi(0)}^{\phi(T)} \left(1 + \cos(2\theta)\right) \exp\left(\kappa \cos(\theta)\right) d\theta \\
&= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \\
&\quad \times \int_{\phi(0)}^{\phi(T)} \left(\exp\left(\kappa \cos(\theta)\right) + \cos(2\theta) \exp\left(\kappa \cos(\theta)\right)\right) d\theta.
\end{aligned}
$$

Using again that the integration interval is $2\pi \gamma_T$ with γ_T integer, and integrates 2π-periodic functions (thus, the integral is invariant to translations of the integration interval), we get

$$
\begin{aligned}
(\text{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[\int_0^{2\pi\gamma_T} \exp\left(\kappa \cos(\theta)\right) d\theta \right. \\
&\quad \left. + \int_0^{2\pi\gamma_T} \cos(2\theta) \exp\left(\kappa \cos(\theta)\right) d\theta\right],
\end{aligned}
$$

$$(\text{Cov}(Z))_{11} = \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)\right] \tag{A.21}$$

$$= \frac{2\pi\gamma_T}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa)\right] \tag{A.22}$$

$$= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right) I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa)\right], \tag{A.23}$$

where γ_T is the number of oscillation periods contained in $[0, T]$.

We can have a similar calculation for the imaginary part, $(\text{Cov}(Z))_{22}$, as well, but using the identity $\sin^2(\theta) = \frac{1}{2}\left(1 - \cos(2\theta)\right)$ instead of equation A.20. The off-diagonal elements of the covariance matrix vanish due to symmetries of integrand.

Therefore, we showed that for a given $\kappa \geq 0$, scaled residual

$$Z' = e^{-i\varphi_0} \sqrt{K} \left(\widehat{\text{PLV}}_K - \text{PLV}^*\right)$$

converges to a zero mean complex gaussian with the following covariance:

$$\mathrm{Cov}\begin{bmatrix} \mathrm{Re}\{Z'\} \\ \mathrm{Im}\{Z'\} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\{Ze^{-i\varphi_0}\} \\ \mathrm{Im}\{Ze^{-i\varphi_0}\} \end{bmatrix}$$

$$= \frac{1}{2\lambda_0(\phi(T) - \phi(0))I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}.$$

$\square$

Proof of Corollary 3. Similar to corollary 2, we can derive the asymptotic PLV, equation 3.9, for this case, from the definition in equation 3.2. We apply the intensity function $\lambda = \lambda_0$ in corollary 1. The PLV asymptotic value (PLV*) can be derived simply by changing the integration variable from $\phi(t)$ to θ (and let $\theta \mapsto \tau(\theta)$ be its inverse).

The covariance matrix of the asymptotic distribution, can be derived by the procedure we used for the proof of corollary 2. We plug the rate λ_0 as $\lambda(t)$ in corollary 1 and integrate on $[0, T]$:

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t))dt. \tag{A.24}$$

By changing the variable from $\phi(t)$ to θ, we get

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \tag{A.25}$$

As $\Lambda(T) = \int_0^T \lambda_0 dt = \lambda_0 T$, we have

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta \tag{A.26}$$

$$= \frac{1}{\lambda_0 T^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \tag{A.27}$$

With a similar calculation for other coefficients of the covariance matrix, we get

$$\mathrm{Cov}(Z) = \frac{1}{\lambda_0 T^2} \int_{\phi(0)}^{\phi(T)} \begin{bmatrix} \cos^2(\theta) & \sin(2\theta)/2 \\ \sin(2\theta)/2 & \sin^2(\theta) \end{bmatrix} \tau'(\theta)d\theta.$$

Therefore, we showed that the scaled residual,

$$Z = \sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right),$$

converges to a zero mean complex gaussian:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) \xrightarrow[K\to+\infty]{} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathrm{Cov}(Z)\right).$$

$\square$

Proof of Theorem 2. Similar to the proof of theorem 1, we rely on a CLT, but this time adapted to the case of vector-valued martingales (Aalen et al., 2008, appendix B) to prove this theorem.

We start from the single trial empirical vector-valued coupling measure of equation 4.1:

$$C = \int_0^t x(t)dN(t)^\top. \tag{A.28}$$

As for the univariate case, under mild assumptions, we can associate a martingale with a vector-valued counting process $N(t)$:

$$M(t) = N(t) - \int_0^t \lambda(s)ds. \tag{A.29}$$

As in this theorem, we assume $\lambda(t) = \lambda_0, t \in [0, T]$, we get

$$M(t) = N(t) - \lambda_0 t. \tag{A.30}$$

The $(p \times n)$ matrix-valued martingale for the empirical coupling matrix of equation 4.1, resulting from stochastic integration, is

$$M_x(t) = \int_0^t x(s)dM^\top(s)ds \tag{A.31}$$

and can be decomposed similarly to equation B.3 as

$$M_x(t) = \int_0^t x(s)dN(s)^\top - \int_0^t x(s)\lambda_0 ds. \tag{A.32}$$

By generalizing the steps of theorem 1, we introduced the $(p \times n)$-variate martingale:

$$\widetilde{M}^{(K)}(t) = 1/\sqrt{K}\sum_{k=1}^{K} M_x^{(k)}(t) \tag{A.33}$$

$$= 1/\sqrt{K}\sum_{k=1}^{K} \int_0^t x(s)\left(dM^{(k)}\right)^\top(s)ds. \tag{A.34}$$

We now state the CLT theorem for multivariate stochastic integrals.

Proposition 1 (Multivariate Martingale CLT; Aalen et al., 2008, Section B.3).
Given the (real) matrix valued predictable functions $H^{(K)}(t)$, consider the multivariate stochastic integral of multivariate martingale $M^{(K)}$ with intensity vector $\lambda^{(K)}(t)$:

$$\int_0^t H^{(K)}(u)dM^{(K)}(u).$$

Assume:

1. $\int_0^t H^{(K)}(u)diag\{\lambda^{(K)}(u)\}H^{(K)}(u)^\top du \xrightarrow{P} V(t).$
2. $\sum_{j=1}^k \int_0^t (H^{(K)}(u))^2 \mathbf{1}_{|H^{(K)}(u)|>\epsilon}\lambda_j^{(K)}(u)du \xrightarrow{P} 0,$ *for all $t \in [0.T]$ and $\epsilon > 0$.*

The above stochastic integral converges in distribution to a mean-zero gaussian martingale of covariance $V(t)$.

We notice that when summing across K trials (see equation A.34), deterministic signals x remain identical and point processes are pooled across K-trials. Given that trials are independent, the counting processes derived from the trial-pooled Poisson processes $\sum_{k=1}^K N^{(k)}(t)$ are distributed as multivariate Poisson processes with intensity vector $K\lambda_0$, such that

$$\widetilde{M}^{(K)}(t) = 1/\sqrt{K} \int_0^t x(s)dP^\top(s)ds, \tag{A.35}$$

where P is the martingale associated with the pooled process,

$$P(t) = \left(\sum_{k=1}^K N^{(k)}(t)\right) - \int_0^t K\lambda(s)ds. \tag{A.36}$$

Given that the coupling matrix is matrix valued, we have to vectorize it in order to apply the above CLT. Let Vec{.} be the operator that concatenates the successive columns of a matrix into a larger column vector. $\widetilde{M}^{(K)}(t)$ is a $(p \times n)$-variate matrix-valued process, and its vectorized version, Vec$\{\widetilde{M}^{(K)}(t)\}$, is a $(pn \times 1)$-variate vector process. We can write equation A.35 in vectorized form as

$$\text{Vec}\{\widetilde{M}^{(K)}(t)\} = \int_0^t H(s)dP^\top(s)ds,$$

with the $(pn \times n)$-variate block diagonal matrix:

$$H(s) = \frac{1}{\sqrt{K}} \begin{bmatrix} x(s) & 0 & \cdots & \cdots & 0 \\ 0 & x(s) & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & x(s) \end{bmatrix}. \tag{A.37}$$

The variance of $\text{Vec}\{\widetilde{M}^{(K)}(t)\}$ (a $(pn \times pn)$-variate covariance matrix which is also called predictable variation process) can be written, based on proposition 1, as

$$\widetilde{V}(t) = \int_0^t H(s)\text{diag}\left\{\boldsymbol{\lambda}(s)\right\} H(s)^\top ds. \tag{A.38}$$

Since we assume a constant intensity function, $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 = \{\lambda_k\}_k$ ($(n \times 1)$-variate matrix), we can simplify equation A.38 as follows:

$$\widetilde{V}(t) = \int_0^t H(s)\text{diag}\left\{K\boldsymbol{\lambda}_0\right\} H(s)^\top ds. \tag{A.39}$$

Replacing $H(s)$ with the block diagonal matrix defined in equation A.37 leads us to

$$\widetilde{V}(t) = \frac{1}{K} \begin{bmatrix} \int_0^t K\lambda_1 x(s)x(s)^H ds & 0 & \cdots & \cdots & 0 \\ 0 & \int_0^t K\lambda_2 x(s)x(s)^H ds & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & & \ddots & \ddots & 0 \\ 0 & & \cdots & & \cdots & 0 & \int_0^t K\lambda_n x(s)x(s)^H ds \end{bmatrix} \tag{A.40}$$

$$= \begin{bmatrix} \lambda_1 \int_0^t x(s)x(s)^H ds & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 \int_0^t x(s)x(s)^H ds & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & & \ddots & \ddots & 0 \\ 0 & & \cdots & & \cdots & 0 & \lambda_n \int_0^t x(s)x(s)^H ds \end{bmatrix}. \tag{A.41}$$

This fulfills condition 1 of the CLT for all $t \in [0, T]$. For the second condition, it is enough to see that the coeffcients of H are bounded by a term decreasing in $\frac{1}{\sqrt{K}}$. The CLT is thus satisfied, and we get convergence in the distribution to a zero-mean complex gaussian of covariance $\widetilde{V}(t)$ for each t. Specializing the result for $t = T$, we get, based on assumption 2, a diagonal covariance matrix with block-constant diagonal coefficients,

$$
\widetilde{V}(T) = \begin{bmatrix}
T\lambda_1 I_p & 0 & \cdots & \cdots & 0 \\
0 & T\lambda_2 I_p & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & 0 & T\lambda_n I_p
\end{bmatrix}, \tag{A.42}
$$

where I_p indicates the $(p \times p)$ identity matrix, which provides the covariance matrix of the (vectorized) coefficients of matrix $\sqrt{K}\widehat{C}_K$.

Therefore, for the normalized coupling matrix, $\widehat{C}_K \mathrm{diag}(\sqrt{T\lambda_0})^{-1}$, the column by-column normalization, normalizes each block of the above covariance matrix by a multiplicative term $\frac{1}{T\lambda_k}$ to lead to an identity covariance. This proves convergence of the normalized coupling matrix in distribution for $K \to +\infty$ to a random matrix with i.i.d. unit variance complex gaussian coefficients (because lack of correlations implies independence in the gaussian case):

$$
\sqrt{K}\mathrm{Vec}\{\widehat{C}_K\mathrm{diag}(\sqrt{T\lambda_0})^{-1}\} \xrightarrow[K \to +\infty]{} \mathcal{N}\left(\mathbf{0}_{pn}, \mathrm{I}_{pn}\right). \tag{A.43}
$$

Proof of Theorem 3. Based on proposition 6 in section B.3, we need only to check the four following necessary conditions, using the Kronecker delta notation of equation 4.2:

1. $\mathbb{E}\bar{X}_{jk}X_{lk} = \delta_{lj}$, for all k.
2. $\frac{1}{n}\max_{j \neq l} \mathbb{E}\left|\bar{X}_{jk}X_{lk}\right|^2 \to 0$ uniformly in $k \leq n$.
3. $\frac{1}{n^2}\sum_{\Gamma}\left(\mathbb{E}[(\bar{X}_{jk}X_{lk} - \delta_{lj})(X_{j'k}\bar{X}_{l'k} - \delta_{j'l'})]\right)^2 \to 0$ uniformly in $k \leq n$, where $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l'$ or $j = l' \neq j' = l\}$.
4. $p/n \to \alpha \in (0, \infty)$.

Based on the same developments as theorem 2, we use the auxiliary processes

$$
X_{lk}(t) = \frac{\sqrt{K}}{\sqrt{\lambda_k T}}\frac{1}{K}\int_0^t x_l(s)dP_k(s) = \frac{1}{\sqrt{K\lambda_k T}}\int_0^t x_l(s)dP_k(s) = \int_0^t H_{lk}(s)dP_k(s)
$$

with P_k zero-mean martingale associated with the Poisson process of intensity $K\lambda_k$ (see equation A.36) and

$$H_{lk}(t) = \frac{x_l(t)}{\sqrt{K\lambda_k T}},$$

and will denote $X_{lk} = X_{lk}(T)$—that is, random variables that we are concerned with are the final values (at $t = T$) of those processes.

Condition 1 is a direct application of results from equation A.42 in the proof of theorem 2 because $\mathbb{E}\left[\bar{X}_{jk}X_{lk}\right]$ is the covariance between the coefficients of the normalized coupling matrix.

For condition 2, let us first evaluate

$$\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2.$$

For that, we can use Ito's formula of equation B.10 and derive the expression of $\bar{X}_{jk}X_{lk}$ as a stochastic integral, using the function $F(\bar{X}_{jk}, X_{lk}) = \bar{X}_{jk}X_{lk}$ We obtain

$$
\begin{aligned}
\bar{X}_{jk}X_{lk} = &-\int_0^T \left(X_{lk}\bar{H}_{jk}(s) + \bar{X}_{jk}H_{lk}(s)\right) K\lambda_k ds \\
&+ \int_0^T \left[\left(\bar{X}_{jk}(s_-) + \bar{H}_{jk}(s_-)\right)\left(X_{lk}(s_-) + H_{lk}(s_-)\right)\right. \\
&\left. - \bar{X}_{jk}X_{lk}(s_-)\right] \left(dP_k(s) + K\lambda_k dt\right), \\
= &\int_0^T \left(X_{lk}\bar{H}_{jk}(s_-) + \bar{X}_{jk}H_{lk}(s_-)\right) dP_k(s) \\
&+ \int_0^T \left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]\left(dP_k(s) + K\lambda_k ds\right).
\end{aligned}
\tag{A.44}
$$

The first term is a stochastic integral of a zero mean martingale, while the second term is a stochastic integral of a Poisson counting process, which we can verify (due to assumption 2) that it has mean δ_{ij}. As a consequence, $\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2$ is the variance of the above expression, which is (by stochastic integral formula)

$$
\begin{aligned}
\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2 = &-\int_0^T \mathbb{E}\left[\left(X_{lk}(s_-)\bar{H}_{jk}(s_-) + \bar{X}_{jk}(s_-)H_{lk}(s_-)\right)^2\right] K\lambda_k ds \\
&+ \int_0^T \left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]^2 K\lambda_k ds.
\end{aligned}
\tag{A.45}
$$

Applying again the formula for predictable variation process, we obtain

$$\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2 = -\int_0^T \left[\int_0^s \left(H_{lk}(u)\bar{H}_{jk}(s_-) + \bar{H}_{jk}(u)H_{lk}(s_-)\right)^2 K\lambda_k du\right] K\lambda_k ds$$

$$+ \int_0^T \left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]^2 K\lambda_k ds. \tag{A.46}$$

Due to assumption 2, this expression is bounded uniformly for any values of i, j, n, k, and condition 2 is fulfilled.

For condition 3, we use the auxiliary result presented in proposition 2 to compute the required fourth-order moments:

$$\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk}\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right] = \int_0^T H_{lk}H_{j'k}ds \int_0^T \bar{H}_{jk}\bar{H}_{l'k}ds$$

$$+ \int_0^T H_{lk}\bar{H}_{jk}ds \int_0^T H_{j'k}\bar{H}_{l'k}ds + \int_0^T H_{lk}\bar{H}_{l'k}ds \int_0^T \bar{H}_{jk}H_{j'k}ds$$

$$+ \frac{1}{K\lambda_k}\int_0^T H_{lk}\bar{H}_{jk}H_{j'k}\bar{H}_{l'k}ds$$

$$= \frac{1}{\lambda_k^2 T^2 K^2}\left[\int_0^T x_l x_{j'}ds \int_0^T \bar{x}_j \bar{x}_{l'}ds + \int_0^T x_l \bar{x}_j ds \int_0^T x_{j'}\bar{x}_{l'}ds\right.$$

$$\left.+ \int_0^T x_l \bar{x}_{l'}ds \int_0^T \bar{x}_{jk}x_{j'k}ds\right] + \frac{1}{K^3\lambda_k^3 T^2}\int_0^T x_l \bar{x}_j x_{j'}\bar{x}_{l'}ds.$$

We first consider the term consisting in all products of two integrals, which we call *integral product term*; the last term in this expression will be dealt with independently. Given assumption 2, it is clear that for l, j, j', l', all different from each other, the integral product term is vanishing. If there happen to be only two indices that are equal, the moment also vanishes (at least one term of each product vanishes). For the case $j = l = k' = l'$, the integral product term possibly does not vanish, but is uniformly bounded, and only n terms satisfy this relation, such that it will not affect the limit of the relevant expression for condition 3 (due to the $1/n^2$ factor).

It remains the case in which three indices exactly are identical. In such a case, one among δ_{jl} or $\delta_{j'l'}$ is one while the other is zero. Take $\delta_{jl} = 1$ and $\delta_{j'l'} = 0$ without loss of generality, assuming $j = l = j' \neq l'$. The relevant quantity of condition 3 is

$$\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk} - 1\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right]$$

$$= \frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk}\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right] - \frac{1}{K^2\lambda_k^2}\mathbb{E}\left[X_{j'k}\bar{X}_{l'k}\right]$$

$$= \frac{1}{\lambda_k^2 T^2 K^2} \left[\int_0^T x_l x_{j'} ds \int_0^T \bar{x}_j \bar{x}_{l'} ds + \int_0^T (x_l \bar{x}_j - T) ds \int_0^T x_{j'} \bar{x}_{l'} ds \right.$$

$$\left. + \int_0^T x_l \bar{x}_{l'} ds \int_0^T \bar{x}_{jk} x_{j'k} ds \right] + \frac{1}{K^3 \lambda_k^3 T^2} \int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds,$$

in which, due to assumption 2, the integral product term still vanishes. As a consequence, the asymptotic behavior we are interested in is given by the behavior of the remaining single integral term of the moment: $\frac{1}{K\lambda_k} \int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds$ (the only remaining nonvanishing terms are bounded and intervene only in n terms of the sum), such that

$$\lim \frac{1}{n^2} \sum_\Gamma \left(\mathbb{E} \left[(\bar{X}_{jk} X_{lk} - \delta_{lj}) (X_{j'k} \bar{X}_{l'k} - \delta_{j'l'}) \right] \right)^2$$

$$= \lim \frac{1}{n^2 K^2 \lambda_k^2} \sum_\Gamma \left(\mathbb{E} \left[(\bar{X}_{jk} X_{lk} - \delta_{lj}) (X_{j'k} \bar{X}_{l'k} - \delta_{j'l'}) \right] \right)^2. \tag{A.47}$$

Thus condition 3 is satisfied due to the theorem's assumption.

To sum up, all four necessary conditions for the application of proposition 6 are fulfilled (condition 4 is part of the assumptions), and the convergence to the MP law follows immediately. □

Proof of Theorem 4. Let us use the result of Chafaï and Tikhomirov (2018) adapted to our complex case and adapt the dimension notation ($n \to p(n)$, $m_n \to n$, but we keep the notation X_n). We additionally checked in all proofs and lemmas that the result still holds when we replace symmetric matrices by Hermitian ones and the scalar product of real vectors by Hermitian products of complex vectors, putting an absolute value on the Hermitian product when the original scalar product was squared. We consider $\{X_n\}$, a sequence of isotropic (i.e. identity covariance) zero mean random vectors and consider the empirical covariance matrix estimated from observing n independent copies of X_n,

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n X_n^{(k)} X_n^{(k)H}.$$

We rely on the strong tail projection property (STP) that guarantees convergence of the spectral measure of the empirical covariance to the MP law, and convergence of the extreme eigenvalues to the ends of the MP support.

Definition 1 (Strong Tail Projection Property (STP)). *STP holds when there exist $f : \mathbb{N} \to [0, 1]$, $g : \mathbb{N} \to \mathbb{R}^+$ such that $f(r) \to 0$ and $g(r) \to 0$ as $r \to \infty$, and for every $p \in \mathbb{N}$, for any orthogonal projection $P : \mathbb{C}^p \to \mathbb{C}^p$ of rank $r > 0$, for any real $t > f(r).r$ we have*

$$\mathbb{P} \left(\|P X_n\|^2 - r \geq t \right) \leq \frac{g(r)r}{t^2}.$$

By noting that $\mathbb{E}\left\|P\mathbf{X}_n\right\|^2 = r$, we can use Chebyshev's inequality to satisfy such property. Let σ^2 be the variance of $\|P\mathbf{X}_n\|^2$. The inequality leads to, for any t,

$$\mathbb{P}\left(\|P\mathbf{X}_n\|^2 - r \geq \sigma t\right) \leq \mathbb{P}\left(\left|\|P\mathbf{X}_n\|^2 - r\right| \geq \sigma t\right) \leq \frac{1}{t^2},$$

so we get $\mathbb{P}\left(\|P\mathbf{X}_n\|^2 - r \geq t\right) \leq \sigma^2/t^2$ and just need to find an upper bound of σ^2 of the form $g(r)r$. To limit the complexity of the rank-dependent analysis, we will look for $g(r) = C/r$ for a fixed positive constant C, such that we just need to bound the above variance by a constant. Finer bounds are likely possible but left to future work.

In our specific case, in line with the proof of theorem 3, we use

$$\mathbf{X}_n = \int_0^T \frac{\mathbf{x}(t)}{\sqrt{K\lambda T}} dP(t),$$

with P the compensated Poisson process martingale of rate $K\lambda$. In an orthonormal basis adapted to the othogonal projection P with rank r, we can rewrite

$$\|P\mathbf{X}_n\|^2 = \sum_{k=1}^r |\langle \mathbf{w}_k, \mathbf{X}_n\rangle|^2,$$

where $\{\mathbf{w}_k\}$ are r orthonormal vectors in $\mathbb{C}^p$. Then we have

$$\sigma^2 = \sum_{k,l \leq r} \mathbb{E}\left[|\langle \mathbf{w}_k, \mathbf{X}_n\rangle|^2 \, |\langle \mathbf{w}_l, \mathbf{X}_n\rangle|^2 - 1\right].$$

Using similar fourth-order moment results as in theorem 3 (based on proposition 2) leads to an expansion for which all terms vanish but one per expectation, leading to

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l \leq r} \int_0^T \langle \mathbf{w}_k, \mathbf{x}(t)\rangle\langle \mathbf{x}(t), \mathbf{w}_k\rangle\langle \mathbf{w}_l, \mathbf{x}(t)\rangle\langle \mathbf{x}(t), \mathbf{w}_l\rangle \, dt,$$

which can be rewritten using the Hermitian operator $\mathcal{X}$ acting on the space of $p \times p$ matrices as a positive definite bilinear form,

$$\mathcal{X}(U, V) = \int_0^T \langle V, \mathbf{x}\mathbf{x}^H(t)\rangle\langle \mathbf{x}\mathbf{x}^H(t), U\rangle \, dt,$$

with associated eigenvalues $\xi_1 \geq \ldots \geq \xi_{p^2} \geq 0$ such that

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l \leq r} \mathcal{X}\left(\boldsymbol{w}_k \boldsymbol{w}_l^H, \boldsymbol{w}_k \boldsymbol{w}_l^H\right).$$

This sum is maximized when the r^2 unitary tensor matrices of the sum $\boldsymbol{w}_k \boldsymbol{w}_l$ are eigenvectors associated with the largest eigenvalues of the operator, such that we get

$$\sigma^2 \leq \frac{1}{K\lambda T^2} \sum_{k=1 \leq r^2} \xi_k,$$

which is itself upper bounded by the trace of the operator, leading to

$$\sigma^2 \leq \frac{1}{K\lambda T^2} \sum_{k,l \leq p(n)} \int_0^T |x_k x_l|^2 dt,$$

which is bounded according to the theorem's assumptions, completing the proof.

Appendix B: Additional Background and Useful Results

B.1 Jump Processes. Jump processes exhibit discontinuities related to the occurrence of random events, which are distributed according to the given point process models. In this letter, we are concerned with jump times distributed according to (possibly inhomogeneous) Poisson processes.

B.1.1 Martingales Related to Counting Processes. As introduced in section 2.2 (see equation 2.3), under mild assumptions, we can associate a zero-mean martingale with a counting process $N(t)$:

$$M(t) = N(t) - \int_0^t \lambda(s)ds. \tag{B.1}$$

In addition, in our case (deterministic intensity), the variance of $M(t)$ is given by

$$V(t) = \mathbb{E}\left[M(t)^2\right] = \int_0^t \lambda(s)ds.$$

B.1.2 Stochastic Integrals. Now, if we consider for a deterministic predictable process H (with regard to the same filtration $\mathcal{F}_t$), the stochastic integration

$$M_H(t) = \int_0^t H(s)dM(s)ds. \tag{B.2}$$

Using equation B.1, we can write

$$M_H(t) = \int_0^t H(s)dN(s) - \int_0^t H(s)\lambda(s)ds, \tag{B.3}$$

which is equivalent to equation 2.5, which introduced the separation of the deterministic component of empirical coupling measure from the (zero-mean) random fluctuations of the measure. $M_H(t)$ is also a zero-mean martingale with respect to history $\{\mathcal{F}_t\}$. This trivially entails that $\mathbb{E}[M_H(t)] = 0$ at all times.

A.1.3 Second Order Statistics. In addition, the second-order statistics of such stochastic integrals can be explicitly derived from the original intensities. In particular, for $M_H(t) = \int_0^t H(s)dM(s)ds$, we have the variance

$$V_H(t) = \mathbb{E}[M_H(t)^2] = \int_0^t H(s)^2\lambda(s)ds, \tag{B.4}$$

which corresponds to its predictable variation process (see Aalen et al., 2008, sec. 2.2.6). A similar result applies to covariance as well. Let G and H be deterministic predictable; then

$$V_{H,G}(t) = \mathbb{E}[M_H(t)M_G(t)] = \int_0^t H(s)G(s)\lambda(s)ds. \tag{B.5}$$

Importantly, we note that this nonvanishing covariance reflects the fact that both stochastic integrals are computed from the same realization of $M(t)$. If two stochastic integrals are derived from independent point processes, the resulting covariance between them is zero.

B.1.4 General Jump Stochastic Processes. For the proofs of our results, it is convenient to state some general results for jump processes that combine deterministic and a jump stochastic integral, decomposable as

$$X(t) = X(0) + \int_0^t f(X(s), s)ds + \int_0^t h(X(s), s)dN(s), \tag{B.6}$$

with $N(t)$ a Poisson process with intensity $\lambda(t)$, f and h square integrable. This clearly includes the martingales defined above.

B.1.5 Mean Stochastic Jump Integrals. According to Hanson (2007, theorem 3.20), we can compute the expectation of $X(t)$ defined in equation B.6:

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)] + \int_0^t f(X(s), s)ds + \int_0^t \mathbb{E}\left[h(X(s), s)\right]\lambda(s)ds. \quad \text{(B.7)}$$

This allows retrieval of the zero-mean property of the stochastic integral of martingales.

B.1.6 Itô's Formula. Itô's formula or Itô's lemma is an identity to find the differential of a function of a stochastic process. It is a counterpart of the chain rule used to compute the differential of composed functions. We restrict ourselves to the case of a time-independent scalar function of a jump process, while different formulas exist for other cases.

A generalized chain rule for the time derivative of such processes allows deriving an integral formula for scalar process $Y(t) = F(X(t))$ with F continuously differentiable (see Hanson, 2007, lemma 4.22, rule 4.23):

$$Y(t) = Y(0) + \int_0^t \frac{dF}{dx}(X(s))f(X(s), s)ds$$

$$+ \int_0^t \left[F\left(X(s_-) + h(X(s_-), s)\right) - F(X(s_-))\right]dN(s), \quad \text{(B.8)}$$

where $X(s_-) = \lim_{t \to s_-} X(t)$ indicates the left limit.

For a scalar function of a multivariate process $Y(t) = F(\boldsymbol{X}(t))$ with

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \int_0^t \boldsymbol{f}(\boldsymbol{X}(s), s)ds + \int_0^t \boldsymbol{h}(\boldsymbol{X}(s), s)dN(s), \quad \text{(B.9)}$$

the generalization is straightforward:

$$Y(t) = Y(0) + \int_0^t \sum_k \frac{dF}{dx_k}(\boldsymbol{X}(s))f_k(\boldsymbol{X}(s), s)ds$$

$$+ \int_0^t \left[F\left(\boldsymbol{X}(s_-) + \boldsymbol{h}(\boldsymbol{X}(s_-), s)\right) - F(\boldsymbol{X}(s_-))\right]dN(s). \quad \text{(B.10)}$$

This allows retrieving the expression of martingale second-order statistics presented above, as well as computing higher-order moments required in the proof of theorem 3.

An application of this formula that we will use follows:

Proposition 2. *Assume that* $W(t) = \int_0^t A(s)dM(s)$, $X(t) = \int_0^t B(s)dM(s)$, $Y(t) = \int_0^t C(s)dM(s)$, *and* $Z(t) = \int_0^t D(s)dM(s)$ *are stochastic integrals with respect to the same (possibly inhomogeneous) Poisson process martingale* $M(t) = N(t) - \int_0^t \lambda(s)ds$ *with intensity* $\lambda(t)$. *Then*

$$
\begin{aligned}
\mathbb{E}\left[WXYZ\right](t) = & \int_0^t ABCD(s_-)\lambda(s)ds \\
& + \left(\int_0^t AB(s)\lambda(s)ds\right)\left(\int_0^t CD(s)\lambda(s)ds\right) \\
& + \left(\int_0^t AC(s)\lambda(s)ds\right)\left(\int_0^t BD(s)\lambda(s)ds\right) \\
& + \left(\int_0^t AD\lambda(s)(s)ds\right)\left(\int_0^t BC(s)\lambda(s)ds\right).
\end{aligned}
\tag{B.11}
$$

Proof. We apply the above formula to $F(W, X, Y, Z) = WXYZ$, yielding

$$
\begin{aligned}
WXYZ(t) = & -\int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s))\lambda ds \\
& + \int_0^t [(W(s_-) + A)(X(s_-) + B)(Y(s_-) + C)(Z(s_-) + D) \\
& - WXYZ(s_-)]dN(s).
\end{aligned}
$$

Expanding the second term, we obtain the formula

$$
\begin{aligned}
WXYZ(t) = & \int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s))\, dM(s) \\
& + \int_0^t (ABYZ(s_-) + AXCZ(s_-) + AXYD(s_-) + WBCZ(s_-) \\
& + WBYD(s_-) + WXCD(s_-))dN(s) + \int_0^t ABCD(s_-)dN(s) \\
& + \int_0^t (ABCZ(s_-) + AXCD(s_-) + ABYD(s_-) + WBCD(s_-))dN(s).
\end{aligned}
$$

The first and last integral terms in this formula have vanishing expectation, the first because it is a stochastic integral of zero mean martingale M, the last because each term inside the integral contains only one random

variable, which is itself a stochastic integral of the martingale M (and thus zero mean). Thus, for the expectation, we get

$$\mathbb{E}\left[WXYZ\right](t) = \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t (AB\mathbb{E}YZ(s_-) + AC\mathbb{E}XZ(s_-)$$
$$+ AD\mathbb{E}XY(s_-) + BC\mathbb{E}WZ(s_-) + BD\mathbb{E}WY(s_-)$$
$$+ CD\mathbb{E}WX(s_-))\lambda(s)ds. \tag{B.12}$$

Based on the Itô integral formula, one can easily derive an expression for the expectation of each product of two variables (see equation A.44), leading to, after reordering the terms,

$$\mathbb{E}\left[WXYZ\right](t) = \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t \left(AB(s_-)\int_0^s CD(u_-)\lambda(u)du\right.$$
$$+ CD(s_-)\int_0^s AB(u_-)\lambda(u)du + AC(s_-)\int_0^s BD(u_-)\lambda(u)du$$
$$+ BD(s_-)\int_0^s AC(u_-)\lambda(u)du + AD(s_-)\int_0^s BC(u_-)\lambda(u)du$$
$$+ \left. BC(s_-)\int_0^s AD(u_-)\lambda(u)du\right)\lambda(s)ds. \tag{B.13}$$

We then observe that the terms inside the integral can be paired such that the integral form of the product derivative formula $(\int f \int g = \int (g \int f + f \int f))$ can be applied, leading directly to equation B.11. $\qquad\square$

B.2 Notions of Convergence. In contrast to finite-dimensional vectors, there are different and nonequivalent notions of convergence for functions and random variables. We explain the two types of convergence encountered in this letter. For a random variable X, we consider its probability measure μ_X such that

$$\mu_X(A) = P(X \in A),$$

and its associated cumulative distribution function (CDF),

$$F_X(x) = \mu_X((-\infty, x]) = P(X \leq x).$$

B.2.1 Convergence in Distribution. The classical definition is based on the CDF.

Definition 2 (Convergence in Distribution). *We say that the sequence of random variables $\{X_n\}$ converges in distribution (or in law) to X whenever*

$$F_{X_n}(x) \underset{n \to +\infty}{\longrightarrow} F_X,$$

at all continuity points of F_X. This is then denoted $X_n \overset{D}{\longrightarrow} X$.

An equivalent definition can be formulated in terms of weak convergence:

Proposition 3. *$X_n \overset{D}{\longrightarrow} X$ if and only if, for any bounded continuous function f,*

$$\mathbb{E}\left[f(X_n)\right] = \int f d\mu_{X_n} \to \mathbb{E}\left[f(X)\right] = \int f d\mu_X,$$

that is, in classical topological terms, the measure μ_{X_n} converges weakly to μ_X.

The generalization to multidimensional variables encountered in theorem 2 consists simply in replacing the cumulative distribution by its multivariate version, $F_X(x) = P(X_1 < x_1, \ldots, X_n < x_n)$, in definitions. A simple necessary and sufficient condition for $X \longrightarrow Y$ is that for all vectors t, $t^\top X \longrightarrow t^\top Y$ (this is the Cramér-Wold theorem, see Billingsley, 1995).

B.2.2 Convergence in Probability. This stronger notion of convergence denotes $X_n \overset{P}{\longrightarrow} X$, stating that for any $\epsilon > 0$,

$$P\left(|X_n - X| > \epsilon\right) \underset{n \to +\infty}{\longrightarrow} 0. \tag{B.14}$$

It can be shown that convergence in probability implies convergence in distribution. The converse is true only in special cases:

Proposition 4. *If X converges in distribution to a (deterministic) constant c, then it also converges to it in probability.*

An extension to the multivariate case is obtained in finite vector spaces by replacing the absolute value in equation B.14 by any norm, or simply by requiring the convergence of all components individually.

B.2.3 Convergence of Random Measures. The ESDs are random measures, and as such, random variables, leaving in an infinite-dimensional space of measures. This means that for a fixed realization ω, the random measure $\boldsymbol{\mu}$ takes the deterministic value $\boldsymbol{\mu}(\omega)$.

Several types of convergence can be defined. First, the notion of *convergence weakly in probability* can be seen as a combination of the above definitions. It is known that the weak convergence of deterministic measures (see proposition 3) can be associated with a (nonunique) metric (the topological

space of weak convergence is metrizable). Let us pick such a metric $\rho(\mu, \nu)$ between two deterministic measures; then:

Definition 3 (Convergence Weakly in Probability). *The sequence of random measures $\boldsymbol{\mu}_n$ converges weakly in probability to the deterministic measure ν for any $\epsilon > 0$:*

$$P\left(\rho(\boldsymbol{\mu}_n, \nu) > \epsilon\right) \underset{n \to +\infty}{\longrightarrow} 0. \tag{B.15}$$

Next, we can also define convergence with probability 1 (also called *almost sure convergence*).

Definition 4 (Convergence (Weakly) with Probability One). *The sequence of random measures $\boldsymbol{\mu}_n$ converges weakly with probability one to the deterministic measure ν for any $\epsilon > 0$:*

$$P\left(\rho(\boldsymbol{\mu}_n(\omega), \nu) \underset{n \to +\infty}{\longrightarrow} 0\right) = 1. \tag{B.16}$$

As for the case of scalar random variables, convergence with probability one implies convergence in probability.

B.3 Random Matrix Theory. Random matrix theory resulted from fairly recent developments in high-dimensional statistics. It has various application in physics (Guhr, Müller-Groeling, & Weidenmüller, 1998; Doussal, Majumdar, & Schehr, 2016), machine learning (Pennington & Bahri, 2017; Pennington & Worah, 2017; Louart et al., 2018), and neuroscience (Timme, Geisel, & Wolf, 2006; Veraart et al., 2016; Almog et al., 2019).

B.3.1 Wishart Ensemble. Let X be a $p \times n$ data matrix. Assume that the coefficients of X, x_{ij} are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. $\mathcal{N}_{\mathbb{C}}$ specifies a standard complex normal distribution. By definition, this means that $x_{ij} = x_{ij}^{real} + i x_{ij}^{imag}$, where $x_{ij} = x_{ij}^{real}$ and x_{ij}^{imag} are independent (real) $\mathcal{N}(0, \frac{1}{2})$. This implies that columns of X are i.i.d. $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_p, \mathrm{I}_p\right)$ and, similarly, the real and imaginary parts are $\mathcal{N}\left(\mathbf{0}_p, \mathrm{I}_p/2\right)$.

As n grows and $\frac{p}{n} \underset{n \to +\infty}{\to} \alpha \in (0, +\infty)$, the ESD of the so-called Wishart ensemble, $S_n = \frac{1}{n} X X^H$, converges to the Marchenko-Pastur law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

$$\frac{d\mu_{MP}}{dx}(x) = \frac{1-\alpha}{\alpha} \mathbf{1}_{\alpha > 1} \delta_0 + \frac{1}{2\pi\alpha x} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}, \tag{B.17}$$

with $a = (1 - \sqrt{\alpha})^2$ and $b = (1 + \sqrt{\alpha})^2$ (see examples for Marchenko-Pastur law for different values of α in Figure 6).
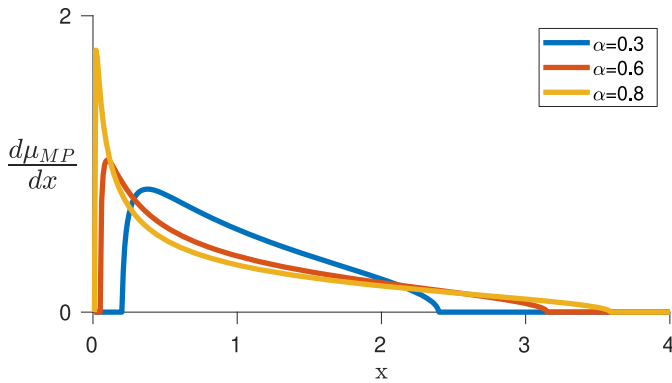
Figure 6: Density of the Marchenko-Pastur law for different values of the aspect ratio of the matrices, α, in equation 2.7.

We wrote here the general formula that holds for all $\alpha > 0$, accounting for zero eigenvalues with a Dirac mass in zero in the rank-deficient case $\alpha > 1$.

B.3.2 Stieltjes Transform of ESD. The Stieltjes transform is a very useful tool to establish the convergence of ESD and determine its limit. The Stieltjes transform of a measure μ is defined as

$$m_\mu(z) = \int \frac{1}{x - z} d\mu(x), z \in \mathbb{C} \setminus \mathbb{R}.$$

A key example for us is the Stieltjes transform of the MP law:

$$m(z) = \frac{1 - c - z + \sqrt{(1 + c - z)^2 - 4c}}{2cz}.$$

Many important results relate measures to their Stieltjes transform. We only need the property that the Stieltjes transform identifies the limit of a sequences of measures, with the following proposition that immediately derives from Anderson et al. (2010, theorem 2.4.4).

Proposition 5. *If two sequences of random measures $\{\mu_k\}$ and $\{\nu_k\}$ converge weakly in probability to a deterministic with identical Stieltjes transform, they converge to the same measure.*

B.3.3 Convergence to MP for Matrices with Dependent Coefficients. Based on the above, we can now write a result that is a combination of results found in Bai and Zhou (2008—mainly theorem 1.1 and corollary 1.1) adapted to

our specific case. We consider a sequence of random matrices $\{X_n\}$ with independent columns and study the ESD of

$$S_n = \frac{1}{n} X_n X_n^H.$$

In the following proposition, we use the Kronecker delta symbold delta (see equation 4.2) and denote by $\bar{X}$ the complex conjugate of X.

Proposition 6. *Let As $n \to \infty$, and assume the following. Let*

1. $\mathbb{E}\bar{X}_{jk}X_{lk} = \delta_{lj}$, for all k.
2. $\frac{1}{n}\max_{j\neq l}\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2 \to 0$ *uniformly in* $k \leq n$.
3. $\frac{1}{n^2}\sum_\Gamma \left(\mathbb{E}\left(\bar{X}_{jk}X_{lk} - \delta_{lj}\right)\left(X_{j'k}\bar{X}_{l'k} - \delta_{j'l'}\right)\right)^2 \to 0$ *uniformly in* $k \leq n$, *where* $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l'$ *or* $j = l' \neq j' = l\}$.
4. $p/n \to \alpha \in (0, \infty)$.

Then, with probability 1, the ESD of S_n tends (weakly) to the MP law.

Sketch of the proof. We use theorem 1.1 from Bai and Zhou (2008) combined with the sufficient condition of corollary 1.1, assuming the identity matrix T_n. These conditions are compatible with the case of the Wishart ensemble, such that the ESD converges to a distribution with the same Stieltjes transform as the MP law.[9] As a consequence of proposition 5, we get that the limit ESD is the MP law.   □

Appendix C: Additional Corollaries

The additional results in this appendix are corollaries based on simplifying assumption 3, where a linear phase is considered instead of the general assumption on phase that was used in corollaries 2 and 3.

Assumption 3. Assume that $\phi(t)$ is a linear function of t on $[0, T]$,

$$\phi(t) = mt, \quad m = 2\pi f = 2\pi/\tau, \tag{C.1}$$

where $f > 0$ (interpretable as the frequency of an oscillation for the continuous signal) and γ_T is the ratio of length (T) of signal to period of oscillation τ:

$$\gamma_T = \frac{T}{\tau} = \frac{\phi(T) - \phi(0)}{2\pi}.$$

---

[9]This requires checking that the self-consistency equation 1.1 in Bai and Zhou (2008) has a unique solution, which they establish by equation 1.2.

Corollary 4. *Under the assumptions of corollary 2, assume additionally assumption 3 is also satisfied, and the intensity of the point-process is given by*

$$\lambda(t) = \lambda_0 \exp(\kappa \cos(\phi(t) - \varphi_0)), \tag{C.2}$$

for a given $\kappa \geq 0$. Then the expectation of the multitrial PLV estimate converges (for $K \to +\infty$) to

$$\text{PLV}^* = \frac{\int_0^T e^{i2\pi ft} \exp(\kappa \cos(2\pi ft - \varphi_0))dt}{\int_0^T \exp(\kappa \cos(2\pi ft - \varphi_0))dt}. \tag{C.3}$$

If, in addition, $[0, T]$ corresponds to an integer number $\gamma_T > 0$ of periods of the oscillation,

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{\phi(0)}^{\phi(T)} \cos(\theta) \exp(\kappa \cos(\theta))d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta))d\theta} = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{C.4}$$

and the scaled residual $\sqrt{K}\left(\widehat{\text{PLV}}_K - \text{PLV}^*\right)$ converges to a zero mean complex gaussian Z with the following covariance:

$$\text{Cov}\begin{bmatrix} Re\{Ze^{-i\varphi_0}\} \\ Im\{Ze^{-i\varphi_0}\} \end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \tag{C.5}$$

Proof. We use the intensity function introduced in equation C.2. The PLV asymptotic value (PLV*) can be derived from definition introduced in equation 3.2 by using assumption 3:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{C.6}$$

$$= \frac{\lambda_0 \int_0^T e^{i\phi(t)} \exp(\kappa \cos(\phi(t) - \varphi_0))dt}{\lambda_0 \int_0^T \exp(\kappa \cos(\phi(t) - \varphi_0))dt} \tag{C.7}$$

$$= \frac{\lambda_0 \int_0^T e^{imt} \exp(\kappa \cos(mt - \varphi_0))dt}{\lambda_0 \int_0^T \exp(\kappa \cos(mt - \varphi_0))dt}. \tag{C.8}$$

We change the integration variable from mt to θ:

$$\text{PLV}^* = \frac{\int_{\theta(0)}^{\theta(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0))d\theta}{\int_{\theta(0)}^{\theta(T)} \exp(\kappa \cos(\theta - \varphi_0))d\theta}. \tag{C.9}$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ, ($\psi = \theta - \varphi_0$),

$$\text{PLV}^* = \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa \cos(\psi)) d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi} \tag{C.10}$$

$$= e^{i\varphi_0} \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi}. \tag{C.11}$$

When $[0, T]$ corresponds to an integer number of periods of the oscillation (i.e., is an integer number), and given that the integration interval is $2\pi \gamma_T$, and integrates 2π-periodic functions (thus the integral is invariant to translations of the integration interval), we have

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{-\pi}^{\pi} \exp(\kappa \cos(\psi)) d\psi}.$$

Observing that the integrand of the denominator is even, while for the numerator the imaginary part is odd and the real part is even, we get

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{0}^{\pi} \cos(\psi) \exp(\kappa \cos(\psi)) d\psi}{\int_{0}^{\pi} \exp(\kappa \cos(\psi)) d\psi}.$$

We prove the first part of the corollary, equation C.3. By using the integral form of the modified Bessel functions I_k for k integer (see Watson, 1995, p. 181)

$$I_k(\kappa) = \frac{1}{\pi} \int_{0}^{\pi} \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta + \frac{\sin(k\pi)}{\pi} \int_{0}^{+\infty} e^{-\kappa \cosh t - kt} dt \tag{C.12}$$

$$= \frac{1}{\pi} \int_{0}^{\pi} \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta, \tag{C.13}$$

we can derive the compact form:

$$\text{PLV}^* = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{C.14}$$

The covariance matrix of the asymptotic distribution can be easily derived by plugging equation C.2 as $\lambda(t)$ in corollary 1 and integrating on $[0, T]$

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t)) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt. \qquad \text{(C.15)}$$

As we have

$$\Lambda(T) = \lambda_0 T I_0(\kappa),$$

we can continue with equation C.15 as,

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t)) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt. \quad \text{(C.16)}$$

To simplify the rest of the derivations, we transform the complex variable coordinates by using $e^{i\phi(t)} e^{-i\varphi_0}$ instead of $e^{i\phi(t)}$ as predictable with respect to $\{\mathcal{F}_t\}$ (i.e., replacing $x(t)$ with $e^{i\phi(t)} e^{-i\varphi_0}$ in theorem 1). With this change, equation C.16 becomes

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t) - \varphi_0) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt.$$

$$\text{(C.17)}$$

Then we change the variable of the integral from $mt - \varphi_0$ to θ (and consequently dt to $\frac{1}{m} d\theta$) and use the following trigonometric identity,

$$\cos^2(\theta) = \frac{1}{2}\left(1 + \cos(2\theta)\right), \qquad\qquad\qquad\qquad \text{(C.18)}$$

to obtain

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} (1 + \cos(2\theta)) \exp\left(\kappa \cos(\theta)\right) d\theta$$

$$= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} \left(\exp\left(\kappa \cos(\theta)\right) + \cos(2\theta) \exp\left(\kappa \cos(\theta)\right)\right) d\theta.$$

Given that the integral is invariant to translations of the integration, we get

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[\int_0^{2\pi \gamma_T} \exp\left(\kappa \cos(\theta)\right) d\theta \right.$$

$$\left. + \int_0^{2\pi \gamma_T} \cos(2\theta) \exp\left(\kappa \cos(\theta)\right) d\theta \right]$$

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)\right] \qquad \text{(C.19)}$$

$$= \frac{2\pi \gamma_T}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa) \right] \tag{C.20}$$

$$= \frac{mT}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa) \right] \tag{C.21}$$

$$= \frac{1}{2\lambda_0 T I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa) \right]. \tag{C.22}$$

We can have a similar calculation for the imaginary part, $(\text{Cov}(Z))_{22}$, as well, but using the identity $\sin^2(\theta) = \frac{1}{2}(1 - \cos(2\theta))$ instead of equation A.20. The off-diagonal elements of the covariance matrix vanish due to symmetry of integrand.

Therefore, we showed that for a given $\kappa \geq 0$, scaled residual

$$Z' = e^{-i\varphi_0} \sqrt{K} \left(\widehat{\text{PLV}}_K - \text{PLV}^* \right)$$

converges to a zero mean complex gaussian with the following covariance:

$$\text{Cov} \begin{bmatrix} \text{Re}\{Z'\} \\ \text{Im}\{Z'\} \end{bmatrix} = \begin{bmatrix} \text{Re}\{Ze^{-i\varphi_0}\} \\ \text{Im}\{Ze^{-i\varphi_0}\} \end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}.$$

$\square$

Corollary 5. *Assume $\phi(t) = 2\pi kt/T$, with $k > 0$ integer, and a sinusoidal modulation of the intensity at frequency m/T, with $m > 0$ integer possibly different from k, phase shift φ_0, and modulation amplitude $\varkappa$ such that*

$$\lambda(t) = \lambda_0 \left(1 + \varkappa \cos\left(2\pi mt/T - \varphi_0\right)\right), \ \lambda_0 > 0, \ 0 \leq \varkappa \leq 1, \tag{C.23}$$

and the point process is homogeneous Poisson with rate λ_0. Then the expectation of the PLV estimate converges (for $K \mapsto +\infty$) to

$$\text{PLV}^* = \frac{1}{2} \varkappa e^{i\varphi_0} \delta_{km}, \tag{C.24}$$

where δ_{km} denotes the Kronecker symbol. Moreover, the asymptotic covariance of $Z = \sqrt{K} \left(\widehat{\text{PLV}}_K - \text{PLV}^* \right)$ is

$$\text{Cov} \begin{bmatrix} Re\{Z\} \\ Im\{Z\} \end{bmatrix} = \frac{1}{2\lambda_0 T} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{C.25}$$

Proof. Similar to corollary 2, we can derive the asymptotic PLV (see equation C.24) for this case, from the definition in equation 3.2. We use the

assumed phase $\phi(t) = 2\pi kt/T$ and apply the intensity function defined in equation C.23 in corollary 1:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{C.26}$$

$$= \frac{\int_0^T e^{i2\pi kt/T}\left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt}{\int_0^T \left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt}. \tag{C.27}$$

By using Euler's formula, we can write the second term in the numerator as weighted sum of exponentials $(\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix}))$,

$$\text{PLV}^* = \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa\int_0^T e^{i2\pi kt/T}\left(e^{i(2\pi mt/T - \varphi_0)} + e^{-i(2\pi mt/T - \varphi_0)}\right)dt}{\int_0^T dt + \int_0^T \varkappa\cos\left(2\pi mt/T - \varphi_0\right)dt} \tag{C.28}$$

$$= \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa\int_0^T e^{i2\pi(k+m)t/T}e^{i\varphi_0} + \varkappa\int_0^T e^{-i2\pi(k-m)t/T}e^{i\varphi_0}dt}{\int_0^T dt + \int_0^T \varkappa\cos\left(2\pi mt/T - \varphi_0\right)dt} \tag{C.29}$$

$$= \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa e^{i\varphi_0}\int_0^T e^{i2\pi(k+m)t/T} + \varkappa e^{i\varphi_0}\int_0^T e^{-i2\pi(k-m)t/T}dt}{\int_0^T dt + \varkappa\int_0^T \cos\left(2\pi mt/T - \varphi_0\right)dt}. \tag{C.30}$$

Given that $k, m > 0$ and we are integrating over full periods, all terms vanish except the last term in the numerator (if and only if $k = m$) and the first term in the denominator. Therefore we have,

$$\text{PLV}^* = \frac{1}{2}\frac{\varkappa e^{i\varphi_0}\int_0^T e^{-i2\pi(k-m)t/T}dt}{\int_0^T dt} \tag{C.31}$$

$$= \frac{1}{2}\varkappa e^{i\varphi_0}\delta_{km}. \tag{C.32}$$

We prove the first part of the corollary.

The covariance matrix of the asymptotic distribution can be derived by the procedure we used for the proof of corollary 2. We plug the rate $\lambda(t)$ assumed in the corollary (see equation C.23) and integrate on $[0, T]$,

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_0^T \cos^2(2\pi kt/T)\left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt, \tag{C.33}$$

and use the trigonometric identity, equation A.20, to get

$$(Cov(Z))_{11} = \frac{\lambda_0}{2\Lambda(T)^2} \int_0^T (1 + \cos(4\pi kt/T)) \, (1 + \varkappa \cos(2\pi mt/T - \varphi_0)) \, dt.$$

(C.34)

In the resulting equation,

$$(Cov(Z))_{11}$$
$$= \frac{\lambda_0}{2\Lambda(T)^2} \left[\int_0^T dt + \varkappa \int_0^T \cos(2\pi mt/T - \varphi_0) \, dt + \int_0^T \cos(4\pi kt/T) dt \right.$$
$$\left. + \varkappa \int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T - \varphi_0) \, dt \right],$$

(C.35)

all terms except the first one vanish. The second and third vanish as we integrate in the full period, and the last term vanishes given that

$$\int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T - \varphi_0) \, dt$$

$$= \cos(\varphi_0) \int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T) \, dt$$

$$+ \sin(\varphi_0) \int_0^T \cos(4\pi kt/T) \sin(2\pi mt/T) \, dt,$$

(C.36)

and k and m are integers.

Finally, given that $\Lambda(T) = \int_0^T \lambda(t) dt = \lambda_0 T$, we have

$$(Cov(Z))_{11} = \frac{1}{2\lambda_0 T}.$$

(C.37)

We have a similar calculation for the imaginary part, $(Cov(Z))_{22}$. The off-diagonal elements of the covariance matrix vanish due to the symmetry of he integrand.

Therefore, we showed that for the scaled residual,

$$Z = \sqrt{K} \left(\widehat{PLV}_K - PLV^* \right)$$

converges to a zero mean isotropic complex gaussian:

$$\sqrt{K} \left(\widehat{PLV}_K - PLV^* \right) \xrightarrow[K \to +\infty]{} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{2\lambda_0 T} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

$\square$

Appendix D: Circular Noise

We use random numbers drawn from the von Mises distribution to generate noise for the phase of an oscillation. Consider the oscillation $O^{orig}[t] = e^{2\pi i f t}$, where the bracket indicates the oscillation is sampled at equispaced discrete times $t = \{k\Delta\}_{k=1,...,q}$. Then $O[t]$ is a noisy version of this oscillation, which is perturbed in the phase

$$O[t] = e^{2\pi i f t} \exp(i\eta[t]), \tag{D.1}$$

where $\eta[t]$ is sampled i.i.d. from the zero-mean von Mises distribution $\mathcal{M}(0, \kappa)$ at each time t. Notably, κ is the dispersion parameter; therefore, larger κ correspond to smaller variance of the noise. In the simulation used in section 4.3, we use $\kappa = 10$.

In the simulation for the multivariate case, we use N_{osc}-dimensional vector of oscillations, $\boldsymbol{O}^{orig}[t] = \{O_j^{orig}[t]\}_{j=1,...,N_{osc}}$, and sample i.i.d. the noise for each oscillation, leading to the vector time series $\boldsymbol{\eta}[t]$. In this case, the noisy oscillations can be written as

$$\boldsymbol{O}[t] = \boldsymbol{O}^{orig}[t] \odot \exp(i\boldsymbol{\eta}[t]), \tag{D.2}$$

where $\odot$ is (entrywise) Hadamard product.

The advantage of such phase noise is to preserve the spectral content of the original oscillation better than conventional normal noise. Nevertheless, using conventional normal (white) noise (on both the real and imaginary parts of the oscillation) did not change the results significantly.

Appendix E: Tables of Parameters

The choice of parameters used in the figures in the main text. In all simulations, $\phi_0 = 0$.

Table 1: Parameters Used for Simulations in Figure 2.

Parameter	Description	A	B
f	Frequency	1 Hz	
K	Number of trials	5000	
T	Simulation length	5 s	
λ_0	Average firing rate	20 Hz	
N_S	Number of simulations	5000	
κ	Modulation strength	0	0.5

Table 2: Parameters Used for Simulations in Figure 3.

Parameter	Description	A	B	C	D
f	Frequency		1 Hz		
K	Number of trials		10		
T	Simulation length	0.75 s	0.5 s	1 s	x-axis
λ_0	Average firing rate		30 Hz		
N_S	Number of simulations		500		
κ	Modulation strength		0		

Table 3: Parameters Used for Simulations in Figure 5.

Parameter	Description	A1	A2	A3	B1	B2	B3
f	Frequency		5 oscillatory components, 11–15 Hz				
K	Number of trials			10			
T	Simulation length			11 s			
λ_0	Average firing rate			20 Hz			
N_S	Number of simulations			100			
κ	Modulation strength		0			0.15	
n_c	Number of LFP channels			100			
n_s	Number of spiking units	10	50	90	10	50	90
κ_{noise}	Dispersion parameter of phase noise			10			

Code Availability

The code to reproduce our simulation results is at https://github.com/shervinsafavi/safavi_neuralComp2021.

Acknowledgments

References

Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. New York: Springer.

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover.

Almog, A., Buijink, M. R., Roethler, O., Michel, S., Meijer, J. H., Rohling, J. H. T., & Garlaschelli, D. (2019). Uncovering functional signature in neural systems via random matrix theory. *PLOS Computational Biology*, *15*(5), e1006934.

Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). *An introduction to random matrices*. Cambridge: Cambridge University Press.

Ashida, G., Wagner, H., & Carr, C. E. (2010). Processing of phase-locked spikes and periodic signals. In S. Rotter& S. Grn (Eds.), *Analysis of parallel spike trains* (pp. 59–74). New York: Springer.

Aydore, S., Pantazis, D., & Leahy, R. M. (2013). A note on the phase locking value and its properties. *NeuroImage*, *74*, 231–244.

Bai, Z., & Yao, J.-f. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, *44* (3), 447–474.

Bai, Z., & Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica*, *18*, 425–442.

Banna, M., Merlevède, F., & Peligrad, M. (2015). On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Processes and Their Applications*, *125*(7), 2700–2726.

Benaych-Georges, F., & Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, *111*, 120–135.

Bhattacharjee, M., & Bose, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Annals of Statistics*, *44*(2), 598–628.

Billingsley, P. (1995). *Probability and measure*. New York: Wiley.

Brillinger, D. R. (1981). *Time series: Data analysis and theory*. Philadelphia: SIAM.

Bühlmann, P., Kalisch, M., & Meier, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annual Review of Statistics and Its Application*, *1*, 255–278.

Bun, J., Bouchaud, J.-P., & Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, *666*, 1–109.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, *7* (5), 446–451.

Buzsaki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Buzsaki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECOG, LFP and spikes. *Nat. Rev. Neurosci.*, *13* (6), 407–20.

Buzsaki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron*, *80*(3), 751–764.

Buzsaki, G., & Schomburg, E. W. (2015). What does gamma coherence tell us about inter-regional neural communication? *Nat. Neurosci.*, *18*, 484–489.

Capitaine, M., & Donati-Martin, C. (2016). *Spectrum of deformed random matrices and free probability.* arXiv:1607.05560.

Chafaï, D., & Tikhomirov, K. (2018). On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. *Probability Theory and Related Fields*, *170*(34), 847–889.

Chavez, M., Besserve, M., Adam, C., & Martinerie, J. (2006). Towards a proper estimation of phase synchronization from time series. *J. Neurosci. Methods*, *154*(1–2), 149–160.

Cole, S., & Voytek, B. (2019). Cycle-by-cycle analysis of neural oscillations. *Journal of Neurophysiology*, *122*(2), 849–861.

Cueva, C. J., Saez, A., Marcos, E., Genovesio, A., Jazayeri, M., Romo, R., . . . Fusi, S. (2020). Low-dimensional dynamics for working memory and time encoding. *PNAS, 117*, 23021–23032.

Dai, H., Wang, Y., Trivedi, R., & Song, L. (2016). Recurrent coevolutionary latent feature processes for continuous-time recommendation. In *Proceedings of the First Workshop on Deep Learning for Recommender Systems* (pp. 29–34). New York: ACM.

De, A., Valera, I., Ganguly, N., Bhattacharya, S., & Rodriguez, M. G. (2016). Learning and forecasting opinion dynamics in social networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, *29* (pp. 397–405). Red Hook, NY: Curran.

Deger, M., Helias, M., Boucsein, C., & Rotter, S. (2012). Statistical properties of superimposed stationary spike trains. *Journal of Computational Neuroscience*, *32*(3), 443–463.

Dickey, A. S., Suminski, A., Amit, Y., & Hatsopoulos, N. G. (2009). Single-unit stability using chronically implanted multielectrode arrays. *J. Neurophysiol.*, *102*(2), 1331–1339.

Doussal, P. L., Majumdar, S. N., & Schehr, G. (2016). Large deviations for the height in 1D Kardar-Parisi-Zhang growth at late times. *Europhysics Letters*, *113*(6), 60004.

Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.*, *14*(11), 770–785.

El Karoui, N. (2003). *On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity*. arXiv:0309355(math).

El Karoui, N. (2005). Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Physica Polonica. Series B, B35*(9), 2681–2697.

El Karoui, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Annals of Probability*, *35*(2), 663–714.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, *36*(6), 2757–2790.

Elsayed, G. F., & Cunningham, J. P. (2017). Structure in neural population recordings: An expected byproduct of simpler phenomena? *Nature Neuroscience*, *20*(9), 1310.

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, *48*(A), 367–378.

Ermentrout, B., & Pinto, D. (2007). Neurophysiology and waves. *SIAM News*, *40*(2).

Ermentrout, G. B., & Kleinfeld, D. (2001). Traveling electrical waves in cortex: Insights from phase dynamics and speculation on a computational role. *Neuron*, *29*(1), 33–44.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.*, *9*(10), 474–80.

Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, *88*, 220–35.

Fukushima, M., Chao, Z. C., & Fujii, N. (2015). Studying brain functions with mesoscopic measurements: Advances in electrocorticography for non-human primates. *Current Opinion in Neurobiology*, *32*, 124–131.

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, *94*, 978–984.

Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, *32*, 148–155.

Grosmark, A. D., & Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science*, *351*(6280), 1440–1443.

Grosmark, A. D., Mizuseki, K., Pastalkova, E., Diba, K., & Buzsáki, G. (2012). REM sleep reorganizes hippocampal excitability. *Neuron*, *75*(6), 1001–1007.

Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, *101*(3), 1126–1140.

Guhr, T., Müller-Groeling, A., & Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: Common concepts. *Physics Reports*, *299*(4–6), 189–425.

Hanson, F. B. (2007). *Applied stochastic processes and control for jump-diffusions: Modeling, analysis and computation*. Philadelphia: SIAM.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *33*(3), 438–443.

Herreras, O. (2016). Local field potentials: Myths and misunderstandings. *Front. Neural Circuits*, *10*, 101.

Hurtado, J. M., Rubchinsky, L. L., & Sigvardt, K. A. (2004). Statistical method for detection of phase-locking episodes in neural oscillations. *Journal of Neurophysiology*, *91*(4), 1883–1898.

Jiang, H., Bahramisharif, A., van Gerven, M. A. J., & Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *NeuroImage*, *118*, 359–367.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, *3*(4), 275–299.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, *29*, 295–327.

Johnstone, I. M., & Onatski, A. (2020). Testing in high-dimensional spiked models. *Annals of Statistics*, *48*(3), 1231–1254.

Juavinett, A. L., Bekheet, G., & Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, *8*, e47188.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., . . . Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236.

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 324–332). New York: ACM.

Kovach, C. K. (2017). A biased look at phase locking: Brief critical review and proposed remedy. *IEEE Transactions on Signal Processing*, *65*(17), 4468–4480.

Kritchman, S., & Nadler, B. (2009). Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, *57*, 3930–3941.

Krumin, M., Reutsky, I., & Shoham, S. (2010). Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front. Comput. Neurosci.*, *4*, 147.

Lepage, K. Q., Kramer, M. A., & Eden, U. T. (2011). The dependence of spike field coherence on expected intensity. *Neural Computation*, *23*(9), 2209–2241.

Li, Z., Cui, D., & Li, X. (2016). Unbiased and robust quantification of synchronization between spikes and local field potential. *J. Neurosci. Methods*, *269*, 33–8.

Liljenstroem, H. (2012). Mesoscopic brain dynamics. *Scholarpedia*, *7*(9), 4601.

Liptser, R. S., & Shiryaev, A. N. (2013a). *Statistics of random processes: I. General theory*. New York: Springer Science & Business Media.

Liptser, R. S., & Shiryaev, A. N. (2013b). *Statistics of random processes II: Applications*. New York: Springer Science & Business Media.

Liu, H., Aue, A., & Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *Annals of Statistics*, *43*(2), 675–712.

Louart, C., Liao, Z., & Couillet, R. (2018). A random matrix approach to neural networks. *Ann. Appl. Probab.*, *28*(2), 1190–1248.

Loubaton, P., & Vallet, P. (2011). Almost sure localization of the eigenvalues in a gaussian information plus noise model: Application to the spiked models. *Electronic Journal of Probability*, *16*, 1934–1959.

Maimon, G., & Assad, J. A. (2009). Beyond Poisson: Increased spike-time regularity across primate parietal cortex. *Neuron*, *62*(3), 426–440.

Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, *114*(4), 507–536.

Mastrogiuseppe, F., & Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, *99*(3), 609–623.

Mingo, J. A., & Speicher, R. (2017). *Free probability and random matrices*. New York: Springer.

Mitra, P. (2007). *Observed brain dynamics*. New York: Oxford University Press.

Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business*. New York: Springer.

Namaki, A., Ardalankia, J., Raei, R., Hedayatifar, L., Hosseiny, A., Haven, E., & Jafari, G. R. (2020). *Analysis of the global banking network by random matrix theory*. arXiv:200714447.

Nawrot, M. P., Boucsein, C., Molina, V. R., Riehle, A., Aertsen, A., & Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods*, *169*(2), 374–390.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27.

O'Leary, T., Sutton, A. C., & Marder, E. (2015). Computational models in the age of large datasets. *Current Opinion in Neurobiology*, *32*, 87–94.

Pennington, J., & Bahri, Y. (2017). Geometry of neural network loss surfaces via random matrix theory. In *Proceedings of the International Conference on Machine Learning* (pp. 2798–2806).

Pennington, J., & Worah, P. (2017). Nonlinear random matrix theory for deep learning. In I. Guyon, Y. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 2637–2646). Red Hook, NY: Curran.

Pennington, J., & Worah, P. (2019). Nonlinear random matrix theory for deep learning. *J. Stat. Mech.*, *2019*(12), 124005.

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, *77*(1–2), 1–37.

Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., . . . Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nature Neuroscience*, *21*, 903–919.

Peterson, E. J., & Voytek, B. (2018). *Healthy oscillatory coordination is bounded by single-unit computation*. bioRxiv:309427.

Protter, P. E. (2005). *Stochastic integration and differential equations*. New York: Springer.

Reimer, I. C., Staude, B., Ehm, W., & Rotter, S. (2012). Modeling and analyzing higher-order correlations in non-Poissonian spike trains. *Journal of Neuroscience Methods*, *208*(1), 18–33.

Rizoiu, M.-A., Lee, Y., Mishra, S., & Xie, L. (2017). *A tutorial on Hawkes processes for events in social media.* arXiv:1708.06401.

Rosenblum, M., Pikovsky, A., Kurths, J., Schäfer, C., & Tass, P. A. (2001). Phase synchronization: From theory to data analysis. In F. Moss & S. Gielen (Eds.), *Handbook of biological physics* (vol. 4, pp. 279–321). Amsterdam: Elsevier.

Safavi, S., Panagiotaropoulos, T., Kapoor, V., Ramirez-Villegas, J. F., Logothetis, N. K., & Besserve, M. (2020). *Uncovering the organization of neural circuits with generalized phase locking analysis*. bioRxiv. https://doi.org/10.1101/2020.12.09.413401

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., . . . Toyama, K. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLOS Computational Biology*, *5*(7).

Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation*, *15*(12), 2823–2842.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*(1), 334–350.

Sohn, H., Narain, D., Meirhaeghe, N., & Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, *103*, 934–907.

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142.

Tabibian, B., Valera, I., Farajtabar, M., Song, L., Schölkopf, B., & Gomez-Rodriguez, M. (2017). Distilling information reliability and source trustworthiness from digital traces. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 847–855). New York: ACM.

Timme, M., Geisel, T., & Wolf, F. (2006). Speed of synchronization in complex networks of neural oscillators: Analytic results based on random matrix theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *16*(1), 015108.

Tracy, C. A., & Widom, H. (2002). *Distribution functions for largest eigenvalues and their applications*. arXiv:0210034(math-ph).

Truccolo, W. (2016). From point process observations to collective neural dynamics: Nonlinear Hawkes process GLMs, low-dimensional dynamics and coarse graining. *Journal of Physiology–Paris*, *110*(4, Part A), 336–347.

Truccolo, W., Hochberg, L. R., & Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: Predicting single neuron spikes. *Nat. Neurosci.*, *13*, 105–111.

Veraart, J., Novikov, D. S., Christiaens, D., Ades-Aron, B., Sijbers, J., & Fieremans, E. (2016). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, *142*, 394–406.

Vinck, M., Battaglia, F. P., Womelsdorf, T., & Pennartz, C. (2012). Improved measures of phase-coupling between spikes and the local field potential. *J. Comput. Neurosci.*, *33*(1), 53–75.

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., & Pennartz, C. M. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage*, *51*(1), 112–22.

Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge: Cambridge University Press.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math*, *62*, 548.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, *67*, 325–327.

Williamson, R. C., Doiron, B., Smith, M. A., & Yu, B. M. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current Opinion in Neurobiology*, *55*, 40–47.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, *316*, 1609–1612.

Zarei, M., Jahed, M., & Daliri, M. R. (2018). Introducing a comprehensive framework to measure spike-LFP coupling. *Frontiers in Computational Neuroscience*, *12*.