

Multitaper Analysis of Semi-Stationary Spectra from Multivariate Neuronal Spiking Observations

Anuththara Rupasinghe and Behtash Babadi

Abstract—Extracting the spectral representations of neural processes that underlie spiking activity is key to understanding how brain rhythms mediate cognitive functions. While spectral estimation of continuous time-series is well studied, inferring the spectral representation of latent non-stationary processes based on spiking observations is challenging due to the underlying nonlinearities that limit the spectrotemporal resolution of existing methods. In this paper, we address this issue by developing a multitaper spectral estimation methodology that can be directly applied to multivariate spiking observations in order to extract the semi-stationary spectral density of the latent non-stationary processes that govern spiking activity. We establish theoretical bounds on the bias-variance trade-off of our proposed estimator. Finally, application of our proposed technique to simulated and real data reveals significant performance gains over existing methods.

Index Terms—Multivariate spiking observations, spectral density matrix, point processes, multitaper analysis.

I. INTRODUCTION

NEURAL oscillations are known to play a significant role in mediating the cognitive and motor functions of the brain [2]–[4]. The advent of high-density electrophysiology recordings [5]–[7] from multiple locations in the brain has opened a unique window of opportunity to probe these oscillations at the neuronal scale. In order to exploit such experimental data for inferring the mechanisms of brain function, spectral analysis techniques tailored for such neuronal spiking data are required [8].

While there exists a wide range of mathematical models for capturing neuronal spiking dynamics [9], statistical approaches based on the theory of point processes or generalized linear models have gained popularity in recent years [8], [10]–[13]. These models relate the statistics of the neural stimulus (input) and response (output) in a model-based fashion to facilitate further analyses such as estimation, prediction, and decoding. The input may consist of explicit (e.g., sensory stimuli) and latent (e.g., internal neural processes) portions. Existing methods for spectral analysis under the point process framework typically consider univariate spiking observations and take the input to be a latent second-order stationary process during the observation period [13]–[16]. Despite their relative success in capturing the spectral properties of spiking activity, these methods have several shortcomings to be addressed.

First of all, some of the existing methods first estimate the latent input processes in the time domain, followed by spectral analysis of these time-domain estimates [14], [15], [17]. The

time-domain estimation techniques vary from averaging across different trials [14] to estimating the latent process using Kalman filtering and smoothing [15], [17]. Due to the time-domain smoothing procedures used by these methods, the resulting power spectral density (PSD) estimates undergo distortion in the spectral domain.

Secondly, some of the existing methods that overcome the aforementioned challenge and aim at directly estimating the PSD without recourse to time-domain smoothing, can only operate on univariate time-series [13], [16]. As such, they are not able to capture the cross-spectral couplings of multiple spike trains, which is crucial in understanding the underlying circuit mechanisms of neural activity.

Finally, it is known that the brain oscillations that underlie neuronal spiking are non-stationary and may exhibit rapid changes corresponding to the brain state or behavioral dynamics [17], [18]. Existing methods apply sliding window adaptations of stationary spectral estimation methods to capture these changes in the spectral domain, and thus provide sub-optimal spectrotemporal resolutions due to the well-known limitations of sliding window analysis. Indeed, non-stationary time-series analysis has been well studied for multivariate continuous signals and various methods have been proposed to quantify the energy-frequency-time distributions [19]–[28]. One notable example is the evolutionary power spectral characterization [24], [25], which defines a time-varying spectral density matrix in order to quantify the local spectral energy distributions at each instant of time for a multivariate oscillatory process.

In light of the foregoing challenges, a unified framework for capturing the non-stationary spectral properties of multivariate neuronal spiking data with high spectrotemporal resolution is lacking, but highly desired due to the emerging demands of modern neuronal data analysis. In this work, we close this gap by developing a framework to estimate the semi-stationary spectral density (SSD) matrix of a multivariate non-stationary latent process, given spiking observations under the point process framework.

To address the first two challenges, we model the spiking observations as multiple realizations of Bernoulli processes with logistic links to the latent multivariate process. This model is motivated by its univariate counterpart, which has been previously utilized in the statistical analysis of spiking data [10], [29]–[31]. We then pose the problem of spectral estimation within a multitapering framework [32]–[36] to directly estimate the SSD from multivariate spiking observations, and thus obviating the need for intermediate time-domain smoothing.

To address the third challenge, we adopt the working hypothesis that the latent processes under study conform to the commonly-used regularity assumption in evolution-

The authors are with the Department of Electrical & Computer Engineering, University of Maryland, College Park, MD 20742. E-mails: {raar, behtash}@umd.edu. This work has been supported in part by the National Science Foundation Award No. 1807216 and the National Institutes of Health award No. 1U19NS107464-01. This work was presented in part at the IEEE Data Science Workshop, Minneapolis, MN, USA, June 2–5, 2019 [1].

ary spectral estimation known as semi-stationarity. We employ a state-space model to characterize the dynamics of the semi-stationary spectra [27] and derive an Expectation-Maximization (EM) algorithm for efficiently computing the maximum a posteriori (MAP) estimate of the Semi-stationary Spectral Density (SSD) matrix. We also establish theoretical bounds on the bias-variance performance of our proposed estimator, and recover the favorable asymptotic properties of the classical multitaper framework.

We compare the performance of our proposed method to existing techniques through simulation studies and application to experimentally-recorded neuronal data. We present simulated case studies using non-stationary multivariate autoregressive processes, whose dynamics are inspired by neural oscillations. These studies reveal that the proposed method outperforms two of the widely used methods for deriving spectral representations from spiking data. Finally, we apply our proposed estimator to multi-unit recordings from the rat cortical neurons during sleep [37], [38]. Our proposed method corroborates existing results on spectrotemporal dynamics of brain state transitions obtained by local field potential analysis, by directly inferring said dynamics from spike recordings at the neuronal scale and at high spectrotemporal resolutions.

In closing, our contributions can be summarized as follows: 1) Developing an algorithm for estimating the SSD matrix directly from spiking observations with no recourse to time-domain smoothing procedures, 2) Enhancing the spectrotemporal resolution by integrating the multi-tapering and state-space modeling frameworks, and 3) Establishing theoretical guarantees on the bias-variance performance of our proposed estimator.

II. PROBLEM FORMULATION

Let $N(t)$ and $H(t)$ denote the point process representing the number of spikes and spiking history of a neuron in $[0, t]$, respectively, where $t \in [0, T]$ and T denotes the observation duration. The Conditional Intensity Function (CIF) [39] of a point process $N(t)$ is defined as:

$$\lambda(t|H_t) := \lim_{\Delta \rightarrow 0} \frac{P[N(t + \Delta) - N(t) = 1|H_t]}{\Delta}. \quad (1)$$

To discretize the continuous process, we consider time bins of length Δ , small enough that the probability of having two or more spikes in an interval of length Δ is negligible. A choice of $\Delta \sim 1$ ms renders this assumption valid, due to the absolute refractory period property of biological neurons [10]. Considering the binary nature of spiking data, this discretized point process can be modeled by a Bernoulli process with success probability $\lambda_k := \lambda(k\Delta|H_k)\Delta$, for $1 \leq k \leq K$, where $K := T/\Delta$ is an integer (with no loss of generality). We refer to λ_k as CIF hereafter for brevity.

In a similar fashion, we consider spiking observations from an ensemble of J neurons, with CIFs $\{\lambda_{k,j}\}_{k=1}^K$, for $j = 1, 2, \dots, J$. Additionally, for each neuron, we assume that L independent realizations of spiking activity are observed. Thus, we represent the l^{th} realization corresponding to the CIF $\lambda_{k,j}$ by $n_{k,j}^{(l)}$. Further, we model $\lambda_{k,j}$ by a logistic link to the underlying latent random process, $x_{k,j}$. The logistic

function is often referred to as the *canonical link* function for a Bernoulli process in the generalized linear model and point process frameworks, and significantly facilitates procedures such as regression and maximum likelihood estimation [10]. Accordingly, we have

$$n_{k,j}^{(l)} \sim \text{Bernoulli}(\lambda_{k,j}), \quad (2)$$

where $\lambda_{k,j} = \ell(x_{k,j}) := 1/(1 + \exp(-x_{k,j}))$, with $\ell(\cdot)$ representing the logistic function. Our goal is to estimate the time-varying power spectral density of each process $x_{k,j}$, for $1 \leq j \leq J$, as well as the time-varying cross spectra between each pair of processes. To that end, we assume that the process $x_{k,j}$ is a *semi-stationary process* in the sense of Priestly [24], which implies a representation of the form,

$$x_{k,j} - \mu_{k,j} = \int_{-\pi}^{\pi} e^{ik\omega} A_{k,j}(\omega) dZ_j(\omega), \quad (3)$$

where $\mu_{k,j}$ is the mean of $x_{k,j}$, $A_{k,j}(\omega)$ is the time-varying amplitude function and $dZ_j(\omega)$ is an orthogonal increment process. The semi-stationarity assumption restricts the rate of change of $A_{k,j}(\omega)$ in the sense that for each fixed ω , the Fourier transform of $A_{k,j}(\omega)$ with respect to k , is highly concentrated around zero [21]. Note that this assumption is the basis for the widely-used short-time periodogram analysis as well as constructing estimators of the evolutionary spectra [24]. Given that the dynamics of the neuronal processes underlying spiking activity are typically much slower than the sampling rate, the assumption of semi-stationarity is commonly adopted in neural data analysis.

To obtain a discrete-parameter harmonic process, we approximate $Z_j(\omega)$ by a jump process over N frequency bins [13], and thereby replace $dZ_j(\omega_n)$ with $\frac{\pi}{N}(a_{j,n} + ib_{j,n})$, at $\omega_n = n\pi/N$, $1 \leq n \leq N-1$, where $a_{j,n}$ and $b_{j,n}$ are real-valued random variables. Since the random process is real, using the symmetry $Z_j(\omega) = Z_j(-\omega)$, we express the discretization of Eq. (3) as

$$x_{k,j} = \mu_{k,j} + \frac{2\pi}{N} \sum_{n=1}^{N-1} A_{k,j}(\omega_n) (a_{j,n} \cos(\omega_n k) - b_{j,n} \sin(\omega_n k)).$$

To explicitly utilize the semi-stationarity assumption, we consider a piece-wise constant approximation to $A_{k,j}(\omega)$, implying that the J -variate random process $\{x_{k,j}\}_{k,j=1}^{K,J}$ is jointly stationary in windows of small enough length W [27], and divide the total data duration K into M non-overlapping segments of length W , with $K = MW$. We thus have:

$$x_{k,j} = \mu_{m,j} + \frac{2\pi}{N} \sum_{n=1}^{N-1} (p_{m,j,n} \cos(\omega_n k) - q_{m,j,n} \sin(\omega_n k)),$$

for $(m-1)W + 1 \leq k \leq mW$, $1 \leq m \leq M$, where $p_{m,j,n}$ and $q_{m,j,n}$ are real-valued random variables. Defining $\mathbf{X}_{m,j} := [x_{(m-1)W+1,j}, x_{(m-1)W+2,j}, \dots, x_{mW,j}]^T$, $\mathbf{v}_{m,j} := [\frac{N}{2\pi}\mu_{m,j}, p_{m,j,1}, q_{m,j,1}, \dots, p_{m,j,N-1}, q_{m,j,N-1}]^T$, we have the compact representation $\mathbf{X}_{m,j} = \mathbf{A}_m \mathbf{v}_{m,j}$, where $\mathbf{A}_m$ is a $W \times (2N-1)$ matrix with the first column equal to $\frac{2\pi}{N}\mathbf{1}$, $(\mathbf{A}_m)_{w,2n} = \frac{2\pi}{N} \cos\left(\frac{n(m-1)W+w}{N}\pi\right)$ and $(\mathbf{A}_m)_{w,2n+1} = -\frac{2\pi}{N} \sin\left(\frac{n(m-1)W+w}{N}\pi\right)$, for $w = 1, 2, \dots, W$ and $n = 1, 2, \dots, N-1$.

The evolutionary spectrum of an oscillatory process

$x_{k,j}$, at frequency ω_n , is defined as $\psi_{k,j}(\omega_n) d\omega_n = |A_{k,j}(\omega_n)|^2 \mathbb{E}[|dZ_j(\omega_n)|^2]$ [24]. Moreover, for a J -variate stationary vector-valued orthogonal increment process $\mathbf{Z}(\omega_n) := [Z_1(\omega_n), \dots, Z_J(\omega_n)]^\top$, the spectral density matrix has been expressed as $\Psi(\omega_n) d\omega_n := \mathbb{E}[d\mathbf{Z}(\omega_n) d\mathbf{Z}(\omega_n)^H]$ [40]. Extending this notion to the semi-stationary spectra, we formulate the SSD matrix as:

$$\Psi_m(\omega_n) = \frac{\pi}{N} \mathbb{E}[(\mathbf{p}_{m,n} + i\mathbf{q}_{m,n})(\mathbf{p}_{m,n} - i\mathbf{q}_{m,n})^\top], \quad (4)$$

where $\mathbf{p}_{m,n} = [p_{m,1,n}, p_{m,2,n}, \dots, p_{m,J,n}]^\top$ and $\mathbf{q}_{m,n} = [q_{m,1,n}, q_{m,2,n}, \dots, q_{m,J,n}]^\top$.

For brevity, we concatenate the vectors $\mathbf{p}_{m,n}$ and $\mathbf{q}_{m,n}$ as $\mathbf{w}_{m,n} := [\mathbf{p}_{m,n}^\top, \mathbf{q}_{m,n}^\top]^\top$, $1 \leq n \leq N-1$, and define the vector $\mathbf{w}_m := [\mathbf{w}_{m,0}^\top, \mathbf{w}_{m,1}^\top, \dots, \mathbf{w}_{m,N-1}^\top]^\top$, with $\mathbf{w}_{m,0} := [\mu_{m,1}, \mu_{m,2}, \dots, \mu_{m,J}]^\top$. Thus, we see that the SSD matrix in Eq. (4), can be determined if $\mathbb{E}[\mathbf{w}_{m,n} \mathbf{w}_{m,n}^\top]$ is estimated for $1 \leq n \leq N-1$. As such, the task of estimating the semi-stationary spectra of the J -variate random process is reduced to computing $\mathbb{E}[\mathbf{w}_m \mathbf{w}_m^\top]$, for $m = 1, 2, \dots, M$, given the ensemble spiking data $\mathcal{D} = \{n_{k,j}^{(l)}\}_{l,k,j=1}^{L,K,J}$. To this end, we first develop a multitaper version of the foregoing SSD estimator to enhance spectral resolution, and then complement it with a suitable estimation algorithm within a Bayesian setting.

III. MULTITAPER ESTIMATION OF THE SSD MATRIX

A. The Proposed Multitaper Framework

The estimate $\Psi_m(\omega_n)$ in Eq. (4) is based on conventional Fourier analysis, which has its inherent bias and variance limitations. Multitapering is a technique widely used in power spectral estimation of stationary random processes, to overcome these limitations [32]. Given that the latent process $x_{k,j}$ is not directly observable, forming its multitaper estimate is not straightforward. We will indeed develop multitaper spectral estimates of $x_{k,j}$ based on its indirect observations $n_{k,j}$.

For a stationary continuous process, $x_1, x_2, \dots, x_K$ the classical multitaper estimator is given by:

$$S^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P |y^{(p)}(\omega)|^2 \quad (5)$$

with $y^{(p)}(\omega) = \sum_{k=1}^K \nu_k^{(p)} x_k e^{-i\omega k}$, where $\nu_k^{(p)}$ is the k^{th} time sample of the p^{th} discrete prolate spheroidal sequence (dpss) [33], for $1 \leq p \leq P$. The dpss tapers are a set of orthogonal tapers that maximally concentrate their spectral power within a resolution bandwidth of $[-\frac{\xi f_s}{K}, \frac{\xi f_s}{K}]$, for some positive constant ξ . Tapering with the dpss sequences results in approximately unbiased spectral estimates with minimal spectral leakage [34]–[36], and averaging the eigen-spectra $|y^{(p)}(\omega)|^2$, significantly reduces the variance of the estimated PSD [32], [34], [35].

While originally designed for univariate time series, multitapering has been likewise extended to multivariate time series [41]. Given a second order jointly stationary J -dimensional random process $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$, where, $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,J}]$, the multitaper cross-spectral estimate between the r^{th} process and the t^{th} process ($r, t \in \{1, 2, \dots, J\}$) has been defined as,

$$S_{r,t}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P y_r^{(p)}(\omega) (y_t^{(p)}(\omega))^* \quad (6)$$

where $y_r^{(p)}(\omega) := \sum_{k=1}^K \nu_k^{(p)} x_{k,r} e^{-i\omega k}$.

These classical estimators are not directly applicable in our framework, since the latent processes $x_{k,j}$ are not directly observable. Our objective is to estimate the semi-stationary spectra $\Psi_m(\omega_n)$ directly from the spiking observations, avoiding intermediate time-domain estimation of the latent processes. Thus, we need to develop an alternative approach that incorporates the effect of indirect observation of $x_{k,j}$ through spiking activity. The data log-likelihood in our model,

$$\log f(\mathcal{D}|\{x_{k,j}\}_{k,j=1}^{K,J}) = \sum_{k,j=1}^{K,J} L\{\bar{n}_{k,j} x_{k,j} - \log(1 + \exp(x_{k,j}))\}, \quad (7)$$

depends on the observations only through the ensemble average, $\bar{n}_{k,j} := \frac{1}{L} \sum_{l=1}^L n_{k,j}^{(l)}$. Hence, $\bar{n}_{k,j}$ is a sufficient statistic. Thus, we need to characterize the effect of tapering the latent time series $x_{k,j}$ on the observed ensemble average spiking $\bar{n}_{k,j}$. Note that tapering the observed process $\bar{n}_{k,j}$, i.e., $\nu_{k,j}^{(p)} \bar{n}_{k,j}$, is not admissible as the dpss tapers take negative values for $p > 1$, whereas spiking observations are non-negative. In addition, $\nu_{k,j}^{(p)} \bar{n}_{k,j}$ does not relate to the desired tapered latent process $\nu_{k,j}^{(p)} x_{k,j}$ in a straightforward fashion.

Consider the log-likelihood $\log f(\{n_{k,j}^{(l)}\}_{l=1}^L | x_{k,j}) = L\{\bar{n}_{k,j} x_{k,j} - \log(1 + \exp(x_{k,j}))\}$, in Eq. (7). Noting that

$$\left. \frac{\partial \log f(\{n_{k,j}^{(l)}\}_{l=1}^L | x_{k,j})}{\partial x_{k,j}} \right|_{x_{k,j} = \ell^{-1}(\bar{n}_{k,j})} = 0,$$

the logit of $\bar{n}_{k,j}$ defined by $\ell^{-1}(\bar{n}_{k,j}) := \log(\bar{n}_{k,j}/(1 - \bar{n}_{k,j}))$ is the maximum likelihood estimator of $x_{k,j}$. Accordingly, we use $\ell^{-1}(\bar{n}_{k,j})$ as the estimator of $x_{k,j}$. Furthermore, $\bar{n}_{k,j}$ converges almost surely to $\ell(x_{k,j})$, by the strong law of large numbers. To address the foregoing challenge, we propose

$$(\bar{n}_{k,j})^{(p)} := \ell(\nu_{k \bmod W+1}^{(p)} \ell^{-1}(\bar{n}_{k,j})), \quad (8)$$

to be the estimator of the ensemble average, had the latent process been tapered by the p^{th} dpss taper, i.e., had the spiking activity been governed by $\nu_{k \bmod W+1}^{(p)} x_{k,j}$. Note that when $\bar{n}_{k,j} = 0$ or $\bar{n}_{k,j} = 1$, the function $\ell^{-1}(\bar{n}_{k,j})$ is not defined. In such cases, we directly estimate $(\bar{n}_{k,j})^{(p)}$ by $\bar{n}_{k,j}$. The dpss tapers are sandwiched between $\ell(\cdot)$ and $\ell^{-1}(\cdot)$ in Eq. (8), which makes it an admissible non-negative ensemble average and is key to developing the multitaper estimator of the spectra of $x_{k,j}$. If one would switch the order of evaluating $\ell(\cdot)$ and tapering in Eq. (8), the process $(\bar{n}_{k,j})^{(p)}$ would reduce to $\nu_{k \bmod W+1}^{(p)} \bar{n}_{k,j}$, which is inadmissible for $p > 1$.

The problem is now reduced to computing the semi-stationary spectra of $\{x_{k,j}\}$ corresponding to each dpss taper, $\Psi_m^{(p)}(\omega_n)$, from $(\bar{n}_{k,j})^{(p)}$, for $p = 1, 2, \dots, P$. Then, the multitaper SSD estimate can be formed by averaging the P eigen-spectral estimates, $\Psi_m^{\text{mt}}(\omega_n) = \frac{1}{P} \sum_{p=1}^P \hat{\Psi}_m^{(p)}(\omega_n)$. In the next subsection, we first develop a method for estimating the SSD matrix of the untapered latent process $\{x_{k,j}\}$ from $\{\bar{n}_{k,j}\}$, and then extend it to the case of P tapers by replacing $\{\bar{n}_{k,j}\}$ with $\{(\bar{n}_{k,j})^{(p)}\}$, which gives the eigen-spectral SSD

estimate $\hat{\Psi}_m^{(p)}(\omega_n)$, for $p = 1, 2, \dots, P$.

B. MAP Estimation of the Parameters via the EM Algorithm

In order to capture the evolution of the SSD $\Psi_m(\omega_n)$, we impose a stochastic continuity constraint on the random variables $\mathbf{w}_m$ using the following state-space model:

$$\mathbf{w}_m = \Phi \mathbf{w}_{m-1} + \eta_m, \quad (9)$$

where the state transition matrix Φ is a constant matrix and $\eta_m \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_m)$. We consider the special case where Φ takes the form $\alpha \mathbf{I}$, for simplicity. Nevertheless, Φ can also be estimated from data within the same Expectation-Maximization [42] framework that follows next.

The parameters $\theta := \{\mathbf{Q}_m, 1 \leq m \leq M\}$ need to be estimated from the observations $\mathcal{D} = \{n_{k,j}^{(l)}\}_{k,j,l=1}^{K,J,L}$. Considering $\mathbf{w} = \{\mathbf{w}_m, 1 \leq m \leq M\}$ to be the hidden variable, we aim at recovering θ via MAP estimation. We consider a diagonal covariance matrix $\mathbf{Q}_m$, whose i^{th} diagonal entry is denoted by $Q_{m,i}$. This assumption is due to the orthogonality of the increment process $dZ_j(\omega)$ in Eq. (3) and eliminates undesired spectral coupling. In addition, it reduces the number of parameters to be estimated, thereby improving the convergence of the estimator. Given the typical sparsity of neuronal spiking in time and low number of observed realizations (L), an appropriate prior distribution on $\mathbf{Q}_m$ helps in reducing the estimation variance. Thus, we assume $\mathbf{Q}_m$ to be independent and identically distributed for $1 \leq m \leq M$, with a density:

$$f(\mathbf{Q}_m) \propto \exp \left(-\rho \sum_{j=1}^{2J} \sum_{n=1}^{N-2} \left(\log(Q_{m,J(2n-1)+j}) - \log(Q_{m,J(2n+1)+j}) \right)^2 \right).$$

This prior distribution encourages continuity in log scale of the spectral estimates corresponding to adjacent frequency bins, and can be controlled by appropriately selecting the hyper-parameter ρ . MAP estimation corresponds to maximizing the log-likelihood of the complete data $(\mathcal{D}, \mathbf{w})$, given by:

$$\begin{aligned} \log f(\mathcal{D}, \mathbf{w}, \theta) &= \log f_{\mathbf{w}|\theta}(\mathbf{w}|\theta) + \log f(\theta) + C_1 \\ &= -\frac{1}{2} \sum_{m=1}^M \left\{ \log |\mathbf{Q}_m| + (\mathbf{w}_m - \Phi \mathbf{w}_{m-1})^\top \mathbf{Q}_m^{-1} (\mathbf{w}_m - \Phi \mathbf{w}_{m-1}) \right\} \\ &\quad - \rho \sum_{m=1}^M \sum_{j=1}^{2J} \sum_{n=1}^{N-2} \left(\log(Q_{m,J(2n-1)+j}) - \log(Q_{m,J(2n+1)+j}) \right)^2 + C_2, \end{aligned}$$

where C_1 and C_2 represent terms that do not depend on θ . We next construct the EM algorithm for solving this MAP estimation problem:

1) *E Step*: Suppose that the current estimate of θ at the r^{th} iteration is given by $\hat{\theta}^{(r)}$. Then, the Q-function

$$\mathcal{Q}^{(r)} := \mathbb{E}[\log f(\mathcal{D}, \mathbf{w}, \theta) | \mathcal{D}, \hat{\theta}^{(r)}] \quad (10)$$

can be evaluated if the conditional expectations

$$\begin{aligned} \mathbf{w}_{m|M} &:= \mathbb{E}[\mathbf{w}_m | \mathcal{D}, \hat{\theta}^{(r)}], \\ \Sigma_{m|M} &:= \mathbb{E}[(\mathbf{w}_m - \mathbf{w}_{m|M})(\mathbf{w}_m - \mathbf{w}_{m|M})^\top | \mathcal{D}, \hat{\theta}^{(r)}], \\ \Sigma_{m,m-1|M} &:= \mathbb{E}[(\mathbf{w}_m - \mathbf{w}_{m|M})(\mathbf{w}_{m-1} - \mathbf{w}_{m-1|M})^\top | \mathcal{D}, \hat{\theta}^{(r)}], \end{aligned}$$

are known. To compute these conditional expectations we utilize the Fixed Interval Smoothing [43] and the Covariance Smoothing [44] algorithms. However, considering that the

forward model is not Gaussian in our formulation, we cannot directly use these algorithms to estimate $\mathbf{w}_{m|M}$ and $\Sigma_{m|M}$.

The density $f(\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\theta}^{(r)})$, with $\mathcal{D}_1^m = \{n_{k,j}^{(l)}\}_{k,j,l=1}^{mW,J,L}$, is proportional to the product of the two densities $f(\mathcal{D}_1^m | \{\mathbf{w}_s\}_{s=1}^m, \hat{\theta}^{(r)})$ and $f(\{\mathbf{w}_s\}_{s=1}^m | \hat{\theta}^{(r)})$, which are Binomial and Gaussian distributed, respectively. Utilizing the unimodality of the density of $\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\theta}^{(r)}$, we approximate it by a multivariate Gaussian, and derive the mean of the distribution, $\mathbf{w}_{m|m}^{(r)}$ by the mode of $\log f(\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\theta}^{(r)})$:

$$\begin{aligned} \argmax_{\mathbf{w}_m} \sum_{j,s,w=1}^{J,m,W} L \left\{ \bar{n}_{(s-1)W+w,j} (\mathbf{A}_s \mathbf{v}_{s,j})_w - \log(1 + \exp(\mathbf{A}_s \mathbf{v}_{s,j}_w)) \right\} \\ - \frac{1}{2} \sum_{s=1}^m \left\{ \log |\mathbf{Q}_s^{(r)}| + (\mathbf{w}_s - \Phi \mathbf{w}_{s-1})^\top (\mathbf{Q}_s^{(r)})^{-1} (\mathbf{w}_s - \Phi \mathbf{w}_{s-1}) \right\}, \quad (11) \end{aligned}$$

and its covariance, $\Sigma_{m|m}^{(r)}$ by the negative of the inverse Hessian of $\log f(\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\theta}^{(r)})$. Observing that the objective function is a combination of convex functions and is differentiable, we perform the above optimization for $\mathbf{w}_{m|m}^{(r)}$ using the Newton-Raphson method. Concurrently, we estimate $\Sigma_{m|m}^{(r)}$ using the Hessian matrix of the objective.

2) *M Step*: Due to the separability of the Q-function $\mathcal{Q}^{(r)}$ in terms of $\mathbf{Q}_m$'s, we can update $\mathbf{Q}_m$'s independently as:

$$\mathbf{Q}_m^{(r+1)} = \argmax_{\mathbf{Q}_m} \mathcal{Q}^{(r)}, \quad 1 \leq m \leq M. \quad (12)$$

Observing that $\mathcal{Q}^{(r)}$ is differentiable in $\mathbf{Q}_m$, we employ the multivariate Newton-Raphson algorithm to perform this maximization and derive the updates for $\mathbf{Q}_m$, $1 \leq m \leq M$.

An implementation of this EM procedure is outlined in Algorithm 1. Following convergence, we use the final estimates $\mathbf{w}_{m|M}$ and $\Sigma_{m|M}$ to evaluate the SSD matrix $\hat{\Psi}_m(\omega_n)$. As mentioned in Section III-A, the same EM procedure can be carried out for $\{\bar{n}_{k,j}^{(p)}\}$, for $p = 1, 2, \dots, P$ in order to estimate the multivariate eigen-spectra $\hat{\Psi}_m^{(p)}(\omega_n)$. Finally, the multitaper spectral density matrix is formed by averaging the eigen-spectral estimates as outlined in Algorithm 2. We refer to our proposed algorithm as the Point Process Multitaper Semi-stationary Spectral Density (PPMT-SSD) estimator.

IV. THEORETICAL ANALYSIS

In this section we derive bounds on the bias and variance of the proposed estimator. We first briefly review the corresponding bounds for the classical multitaper estimator (Eq. (5)). As proven in [45], the bias and variance of the multitaper estimate of a stationary process $x_1, x_2, \dots, x_K$ with a uniformly continuous PSD $S(\omega)$ are bounded as follows:

$$|\text{bias}(S^{\text{mt}}(\omega))| \leq (\sup_{\omega} S(\omega)) \left\{ 1 - \frac{1}{P} \sum_{p=1}^P c_p \right\} + o(1), \quad (13)$$

$$\text{Var}(S^{\text{mt}}(\omega)) = (1 + \beta(\omega)) \frac{1}{P^2} \sum_{p=1}^P c_p^2 S(\omega)^2 + \mathcal{O} \left(\frac{1}{P} \sum_{p=1}^P (1 - c_p) \right) + o(1), \quad (14)$$

as $K \rightarrow \infty$. Here c_p is the eigenvalue corresponding to the taper $\nu^{(p)}$ and $\beta(\omega) = 0$ if $\frac{\omega}{2\pi} \neq 0, 1/2 \bmod 1$ and is equal to 1 otherwise. It is evident that the multitaper estimator $\hat{S}^{\text{mt}}(\omega)$ is asymptotically unbiased.

Algorithm 1 Estimation of the Semi-stationary Spectral Density Matrix via the EM Algorithm

Inputs: Ensemble averages of the spiking observations $\{\bar{n}_{k,j}\}_{k,j=1}^{K,J}$, hyper-parameters ρ, ζ and α , max. number of EM iterations $R_{\max}$.

Outputs: Estimates of the SSD matrices $\hat{\Psi}_m(\omega_n)$ for $1 \leq m \leq M$, $1 \leq j \leq N - 1$

Initialization: Initial choice of $\mathbf{Q}_m^{(0)} = \zeta \mathbf{I}$, $\mathbf{w}_{0|0} = \mathbf{0}$, $\Sigma_{0|0} = \mathbf{0}$, $r = 1$.

```

1: for  $r \leq R_{\max}$  do
2:   Forward filter, for  $m = 1, 2, \dots, M$ 
      $\mathbf{w}_{m|m-1} = \Phi \mathbf{w}_{m-1|m-1}$ 
      $\Sigma_{m|m-1} = \Phi \Sigma_{m-1|m-1} \Phi^\top + \mathbf{Q}_m^{(r)}$ 
     Compute  $\mathbf{w}_{m|m}$  and  $\Sigma_{m|m}$  using the Newton's method
     as described in Eq. (11).
3:   Backward smoother, for  $m = M - 1, M - 2, \dots, 1$ 
      $\mathbf{B}_m = \Sigma_{m|m} \Phi^\top \Sigma_{m+1|m}^{-1}$ 
      $\mathbf{w}_{m|M} = \mathbf{w}_{m|m} + \mathbf{B}_m (\mathbf{w}_{m+1|M} - \mathbf{w}_{m+1|m})$ 
      $\Sigma_{m|M} = \Sigma_{m|m} + \mathbf{B}_m (\Sigma_{m+1|M} - \Sigma_{m+1|m}) \mathbf{B}_m^\top$ 
4:   Covariance smoother, for  $m = M - 1, M - 2, \dots, 1$ 
      $\Sigma_{m,m-1|M} = \Sigma_{m|M}^\top \mathbf{B}_{m-1}^\top$ 
5:   Update the  $\mathbf{Q}_m$ 's independently, for  $m = 1, 2, \dots, M$  using
     the multivariate Newton-Raphson method to solve
      $\mathbf{Q}_m^{(r+1)} = \underset{\mathbf{Q}_m}{\operatorname{argmax}} \mathcal{Q}^{(r)}$ .
6:   Set  $r \leftarrow r + 1$ 
7: end for
8: for  $1 \leq m \leq M$  do
9:    $\mathbf{R}_m = \Sigma_{m|M} + \mathbf{w}_{m|M} \mathbf{w}_{m|M}^\top$ 
10:  for  $1 \leq n \leq N - 1$  do
11:     $\mathbf{R}_m^n = (\mathbf{R}_m)^{(J(2n-1)+1:J(2n+1), J(2n-1)+1:J(2n+1))}$ .
12:     $\hat{\Psi}_m(\omega_n) = \frac{\pi}{N} \left\{ \mathbf{R}_m^n(1:J, 1:J) + \mathbf{R}_m^n(J+1:2J, J+1:2J) \right.$ 
       $\left. + i (\mathbf{R}_m^n(J+1:2J, 1:J) - \mathbf{R}_m^n(1:J, J+1:2J)) \right\}$ .
13:  end for
14: end for
15: Return  $\hat{\Psi}_m(\omega_n)$  for  $1 \leq m \leq M$ ,  $1 \leq n \leq N - 1$ 

```

Algorithm 2 Estimation of the Multitaper Semi-stationary Spectral Density Matrix (PPMT-SSD)

Inputs: Collection of ensemble averages of the observations $\{\bar{n}_{k,j}\}_{k,j=1}^{K,J}$, the set of P dpss tapers of length W $\{\nu_w^{(p)}\}_{w,p=1}^{W,P}$

Outputs: The multitaper estimates of the SSD matrices $\hat{\Psi}_m^{\text{mt}}(\omega_n)$ for $1 \leq m \leq M$, $1 \leq n \leq N - 1$

```

1: for  $p = 1, 2, \dots, P$  do
2:   for  $1 \leq w \leq W$ ,  $1 \leq m \leq M$ ,  $1 \leq j \leq J$  do
3:      $k = ((m-1)W + w)$ 
4:     if  $\bar{n}_{k,j} \neq 0$  and  $\bar{n}_{k,j} \neq 1$  then
5:        $(\bar{n}_{k,j})^{(p)} = \ell(\ell^{-1}(\bar{n}_{k,j}) \nu_w^{(p)})$ 
6:     else
7:        $(\bar{n}_{k,j})^{(p)} = \bar{n}_{k,j}$ 
8:     end if
9:   end for
10:  Compute the  $p^{\text{th}}$  tapered spectral density matrix estimate,
      $\hat{\Psi}_m^{(p)}(\omega_n)$  for  $1 \leq m \leq M$ ,  $1 \leq n \leq N - 1$ , using
     Algorithm 1, with  $\{\bar{n}_{k,j}^{(p)}\}_{k,j=1}^{K,J}$  as the input collection of
     ensemble averages.
11: end for
12: for  $1 \leq m \leq M$ ,  $1 \leq n \leq N - 1$  do
13:    $\hat{\Psi}_m^{\text{mt}}(\omega_n) = \frac{1}{P} \sum_{p=1}^P \hat{\Psi}_m^{(p)}(\omega_n)$ 
14: end for
15: return  $\hat{\Psi}_m^{\text{mt}}(\omega_n)$  for  $1 \leq m \leq M$ ,  $1 \leq n \leq N - 1$ 

```

We state our main theorem for a univariate second-order stationary process $x_1, x_2, \dots, x_K$, corresponding to the special case of $J = 1$, for the clarity of exposition. We later on provide

extensions to the multivariate and semi-stationary cases. In order to proceed with our theoretical analysis, we need to make two extra technical assumptions.

Assumption (1). From Eq. (7), the data likelihood for the univariate case can be expressed as

$$f(\mathcal{D}) := \int \exp \left(\sum_{k=1}^K L \left\{ \bar{n}_k x_k - \log(1 + \exp(x_k)) \right\} \right) \prod_{k=1}^K dx_k,$$

where $\bar{n}_k := \frac{1}{L} \sum_{l=1}^L n_k^{(l)}$. Given the nonlinear functional form of the integrand, we consider the saddle point approximation [46] of the integral, and take $\ell^{-1}(\bar{n}_k)$ as an estimator of x_k . Under this approximation, the proposed multitaper spectral estimator takes the simpler form:

$$\hat{S}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P |\hat{y}^{(p)}(\omega)|^2, \quad (15)$$

where $\hat{y}^{(p)}(\omega) := \sum_{k=1}^K \nu_k^{(p)} \ell^{-1}(\bar{n}_k) e^{-i\omega k}$.

Assumption (2). We assume that $|x_k| \leq B$ for all $k = 1, 2, \dots, K$, for some fixed upper bound B . In defining the bias and variance, we condition the expectations on the event $A := \{\bar{n}_k \mid \bar{n}_k \neq 0, \bar{n}_k \neq 1, 1 \leq k \leq K\}$, which is highly probable due to the absolute boundedness of x_k , for $L > 2(1 + \exp(B))$ (See Appendix A for details). We denote the conditional bias and variance by $\text{bias}_A(\cdot)$ and $\text{Var}_A(\cdot)$, respectively. Note that for the multivariate case, we naturally extend this assumption to $|x_{k,j}| \leq B$ for all k, j , and define the set A as $A := \{\bar{n}_{k,j} \mid \bar{n}_{k,j} \neq 0, \bar{n}_{k,j} \neq 1, 1 \leq k \leq K, 1 \leq j \leq J\}$.

Theorem 1 (Univariate Case). *Under the Assumptions (1) and (2) and for $L > \max\{2(1 + \exp(B)), 2B/\log(1 + \exp(B))\}$, the conditional bias and variance of $\hat{S}^{\text{mt}}(\omega)$ in Eq. (15) is bounded with respect to those of the direct multitaper estimate $S^{\text{mt}}(\omega)$ given in Eqs. (13) and (14) as:*

$$|\text{bias}_A(\hat{S}^{\text{mt}}(\omega))| \leq g_1 K \frac{\log L}{\sqrt{L}} + |\text{bias}(S^{\text{mt}}(\omega))|, \quad (16)$$

$$\text{Var}_A(\hat{S}^{\text{mt}}(\omega)) \leq \left\{ g_2 K \frac{\log L}{\sqrt{L}} + \sqrt{\text{Var}(S^{\text{mt}}(\omega))} \right\}^2 \quad (17)$$

where g_1 and g_2 are bounded constants depending on B , K and L , explicitly given in Appendix A.

Proof of Theorem 1. The proof of Theorem 1 is given in Appendix A. $\square$

Remark. Theorem 1 states that the cost of the indirect access to the process $\{x_k\}_{k=1}^K$ through spiking observations with L trials appears as excess terms in both the bias and variance, which would go to zero as $\frac{L}{K^2 \log^2 L} \rightarrow \infty$. Hence, for large enough number of realizations $\bar{L}$, one can expect a performance close to the direct multitaper estimate of $\{x_k\}_{k=1}^K$. While Assumption (2) on the boundedness of the time-series is natural in practical scenarios, as long as $|x_k| \leq \epsilon \log L$, for some $\epsilon < 1/2$, the excess bias and variance terms will be bounded by $\mathcal{O}(\log^2 L / L^{1/2-\epsilon})$, which implies that the bias and variance of the estimator $\hat{S}^{\text{mt}}(\omega)$ will converge to those of $S^{\text{mt}}(\omega)$ even under milder conditions.

The following corollary extends Theorem 1 to the multivariate case:

Corollary 1 (Stationary Multivariate Case). *Consider a second order jointly stationary J -variate process $\{\mathbf{x}_k\}_{k=1}^K$, where, $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,J}]^\top$. Suppose that the observations are binary spiking data, $\{n_{k,j}^{(l)}\}_{k,j,l=1}^{K,J,L}$ with $n_{k,j}^{(l)} \sim \text{Bernoulli}(\ell(x_{k,j}))$ and $\bar{n}_{k,j} := \frac{1}{L} \sum_{l=1}^L n_{k,j}^{(l)}$. Then, under Assumptions (1) and (2), the bias and variance of the multitaper cross-spectral estimate between the r^{th} and t^{th} processes, given by*

$$\hat{S}_{r,t}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P \hat{y}_r^{(p)}(\omega) (\hat{y}_t^{(p)}(\omega))^*$$

with $\hat{y}_r^{(p)}(\omega) = \sum_{k=1}^K \nu_k^{(p)} \ell^{-1}(\bar{n}_{k,r}) e^{-i\omega k}$, are bounded as follows:

$$\begin{aligned} |\text{bias}_A(\hat{S}_{r,t}^{\text{mt}}(\omega))| &\leq g'_1 K \frac{\log L}{\sqrt{L}} + |\text{bias}(S_{r,t}^{\text{mt}}(\omega))|, \\ \text{Var}_A(\hat{S}_{r,t}^{\text{mt}}(\omega)) &\leq \left\{ g'_2 K \frac{\log L}{\sqrt{L}} + \sqrt{\text{Var}(S_{r,t}^{\text{mt}}(\omega))} \right\}^2, \end{aligned}$$

where g'_1 and g'_2 are bounded constants that depend only on B , K , and L , explicitly given in Appendix B.

Proof. The proof is outlined in Appendix B. $\square$

It is not difficult to verify that $\text{bias}(S_{r,t}^{\text{mt}}(\omega))$ and $\text{Var}(S_{r,t}^{\text{mt}}(\omega))$ can be upper bounded in a similar fashion to Eqs. (13) and (14), with the true cross-spectra $S_{r,t}(\omega)$ replacing $S(\omega)$. Before extending the result of Corollary 1 to the semi-stationary case, we need an additional assumption:

Assumption (3). Given that Corollary 1 holds for large L , in this regime we relax the prior distribution on $\mathbf{Q}_m$ to be flat, i.e., $f(\mathbf{Q}_m) \propto 1$. Recall that the rationale for using a prior on $\mathbf{Q}_m$ in Section III-B was to reduce the variance of the estimates in the low spiking regime, i.e., small L .

Finally, combining Corollary 1 and the treatment of [27], we have the following corollary on the bias and variance of the PPMT-SSD estimator:

Corollary 2 (Semi-stationary Multivariate Case). *Suppose that the J -variate process in Corollary 1 is semi-stationary (jointly stationary within consecutive windows of length W). Let $\Psi_{m,r,t}(\omega_n)$ be the cross-spectra between the r^{th} and t^{th} processes over window m , $1 \leq m \leq K/W$, and $\hat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n)$ be the corresponding multitaper estimate obtained from spiking observations. Then, under Assumptions (1)–(3), the bias and variance of the proposed PPMT-SSD estimator at window m can be bounded as,*

$$\begin{aligned} |\text{bias}_A(\hat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n))| &\leq g''_1(\omega_n) W \frac{\log L}{\sqrt{L}} + |\Psi_{m,r,t}(\omega_n)| |1 - \kappa_m(\omega_n)| \\ &\quad + \kappa_m(\omega_n) \left\{ \sup_{\omega} \{|\Psi_{m,r,t}(\omega)|\} \left(1 - \frac{1}{P} \sum_{p=1}^P c_p \right) + o(1) \right\}, \\ \text{Var}_A(\hat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n)) &\leq \left\{ g''_2(\omega_n) W \frac{\log L}{\sqrt{L}} + \sqrt{\frac{2}{P}} \sup_{\omega} \{|\kappa_m(\omega)| |\Psi_{m,r,t}(\omega)|\} \right\}^2, \end{aligned}$$

where $g''_1(\omega)$, $g''_2(\omega)$ are bounded functions of B , L , W and $\kappa_m(\cdot)$ is a function of ω , explicitly given in Appendix B.

Proof. The proof is mainly based on Theorem 1 of [27]. The proof sketch is given Appendix B for brevity. $\square$

V. SIMULATION STUDIES

In this section, we present two simulation studies based on non-stationary AR processes, and compare the performance of the proposed PPMT-SSD estimator with respect to those of two existing methods and benchmarks.

A. Existing Methods and Benchmarks for Comparison

State-Space SSD (SS-SSD) Estimator: This estimator is based on [15], where the latent process x_k is modeled as a first-order autoregressive process, $x_k = x_{k-1} + \epsilon_k$, with $\epsilon_k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ for $1 \leq k \leq K$. Following an EM algorithm developed in [15], the MAP estimate of x_k given the observed data is obtained and its multitaper PSD is computed. For the multivariate non-stationary case, for each process $x_{k,j}$, $1 \leq j \leq J$, we assume joint stationarity within non-overlapping consecutive windows of length W . The MAP estimates of all processes are obtained and an estimate of the SSD matrix is derived using their PSDs within each window.

Peristimulus Time Histogram SSD (PSTH-SSD) Estimator: This estimator is derived by directly considering the ensemble mean of the spiking observations $\bar{n}_{k,j}$, referred to as the peristimulus time histogram (PSTH), as an estimate of the random signal $x_{k,j}$, for $1 \leq k \leq K$ and $1 \leq j \leq J$ [14]. With a similar joint stationarity assumption in windows of length W , the non-overlapping sliding window multitaper spectral estimate of the PSTH forms the PSTH-SSD estimator.

Benchmarks (True SSD & Oracle SSD): In order to benchmark our comparison, we consider the theoretical spectra of the AR processes derived using closed-form expressions (True SSD) as well as the non-overlapping sliding window direct multitaper estimates of the processes $x_{k,j}$ that have been used to generate the spikes. We refer to the latter benchmark as the Oracle SSD, as if an oracle could directly observe the latent processes and estimate their SSD.

B. Study 1: Estimating the SSD of a Latent Trivariate Process

In this study, we model the latent stimuli by AR processes, so that the ground truth is known for performance comparison. We generate three processes ($J = 3$) with spectral couplings, by considering different linear combinations of a set of AR(6) processes, $\{y_k^{(i)}\}_{k=1}^K$, $1 \leq i \leq 6$, where $y_k^{(i)}$ is tuned around the frequency f_i , with $f_0 = 0.0008$ Hz, $f_1 = 1.15$ Hz, $f_2 = 0.95$ Hz, $f_3 = 1.3$ Hz, $f_4 = 1.5$ Hz, $f_5 = 0.65$ Hz and $f_6 = 1.85$ Hz. These linear combinations are given as:

$$\begin{aligned} x_{k,1} &= y_k^{(1)} \cos\left(2\pi \frac{f_0}{f_s} k\right) + 1.2y_k^{(4)} + 1.2y_k^{(5)} u_{k-0.4K} + \sigma_{x1} \nu_{1,k} + x_{1,dc} \\ x_{k,2} &= 0.83 y_k^{(2)} + 0.83 y_{k-6}^{(4)} + 0.83 y_k^{(5)} + 0.83 y_k^{(6)} + \sigma_{x2} \nu_{2,k} + x_{2,dc} \\ x_{k,3} &= y_k^{(3)} + y_k^{(5)} + y_{k-10}^{(6)} u_{0.5K-1-k} + y_k^{(6)} u_{k-0.5K} + \sigma_{x3} \nu_{3,k} + x_{3,dc} \end{aligned}$$

where $x_{i,dc}$ are the DC components, u_k is the unit step function, $\nu_{i,k}$ is a zero mean white Gaussian noise with unit variance, and σ_i is a scaling standard deviation to set the SNR of all signals at 20 dB, for $i = 1, 2$ and 3. To induce

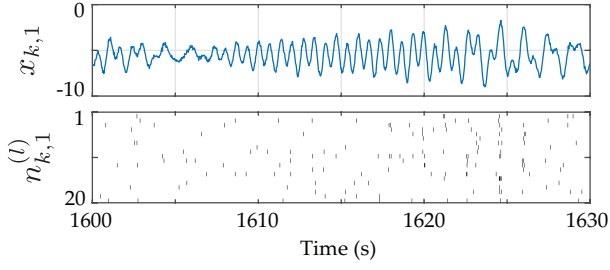


Fig. 1: Samples of the signal $x_{k,1}$ (top) and the raster plot of the corresponding spikes (bottom) from $t = 1600$ s to $t = 1630$ s.

non-stationarity, we have included an amplitude modulated component based on $y_k^{(1)}$ and an abruptly appearing version of $y_k^{(5)}$ in $x_{k,1}$. Likewise, we incorporated non-stationarity and coupling among the processes $\{x_{k,i}\}_{i=1}^3$, by adding time-varying weights to the different $y_k^{(i)}$ s using step functions.

All signals have been sampled at $f_s = 32$ Hz, for a total duration of 2000 seconds ($K = 64000$). We take the window length of stationary to be 100 seconds ($W = 3200$), resulting in a total number of $M = 20$ windows. We assume that $L = 20$ spike train realizations are observed from each latent process. All DC components have been set to -5.5 , so that the average spiking rate of the ensemble, denoted by f_r , is ≈ 0.28 spikes/s, consistent with the low spiking rate of experimentally recorded data. A 30 s sample window of the process $x_{k,1}$ and the corresponding spiking raster plot are shown in Fig. 1.

Fig. 2 shows the main results of this study, and is formatted as a grid with columns representing (from left to right) the True EDS, Oracle SSD, PPMT-SSD, SS-SSD, and PSTH SSD estimates, and rows representing $(\Psi_m)_{i,j}(\omega)$, i.e., the magnitude of the $(i,j)^{\text{th}}$ block of the SSD matrix for $i, j = 1, 3$. Furthermore, for a closer inspection, the magnitude of the spectra corresponding to a window of $t = 700$ s to $t = 800$ s for $(\Psi_m)_{i,j}(\omega)$, for $i, j = 1, 2$ are shown in Fig. 3. It can be observed that the proposed PPMT-SSD estimator (Fig. 2C) results in much less background noise compared to all the others, while precisely capturing the evolution of the spectra and properly resolving the various spectral components. The latter is more evident from Fig. 3, where the PPMT-SSD (black trace) closely matches the true SSD (blue trace) on par with the Oracle SSD (red trace), while the SS-SSD (green trace) and PSTH-SSD (orange trace) show significant bias and variability.

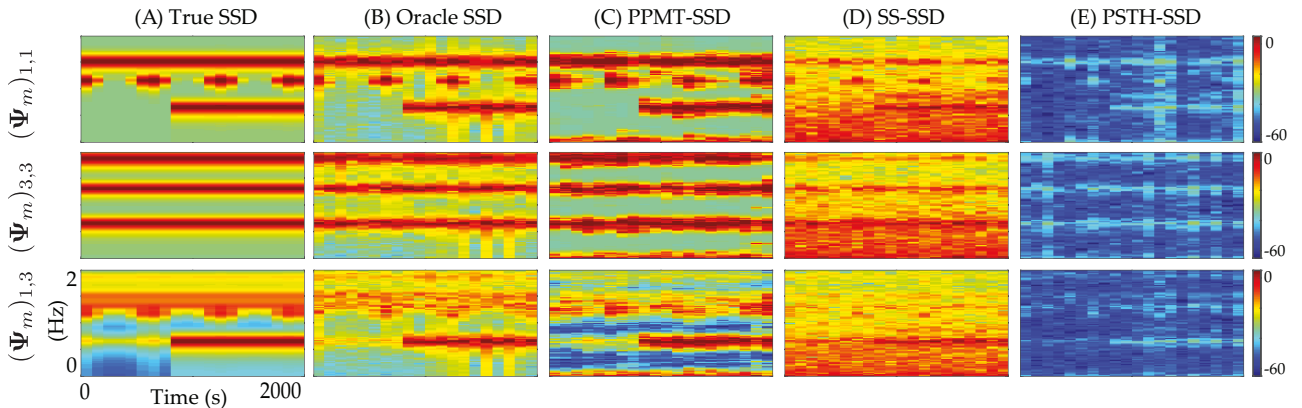


Fig. 2: Estimation results for Study 1. Each panel shows the magnitude spectra in dB scale. Columns from left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD. Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{3,3}(\omega)$ and $(\Psi_m)_{1,3}(\omega)$.

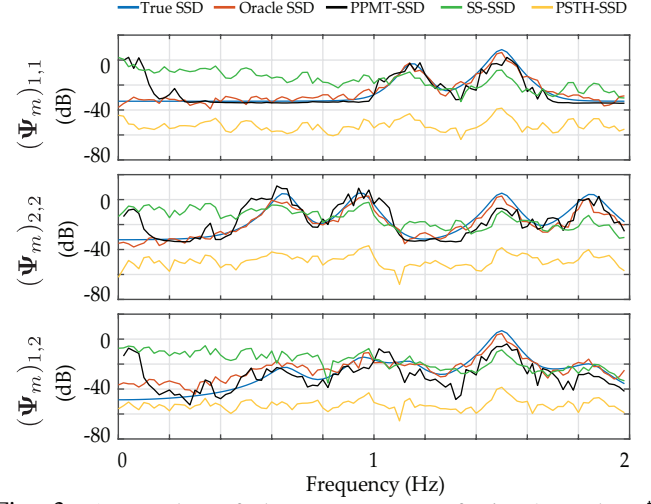


Fig. 3: A snapshot of the spectrograms of Fig. 2 at the 8th window ($t = 700$ s– 800 s). Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

It is worth mentioning that the erroneous spectral peak near the DC component in Fig. 3 (black trace) is due to the estimation error of the DC component in the low spiking regime of our setting, and is mitigated as the spiking rate increases.

Due to time-domain smoothing carried out by the SS-SSD estimator, the SSD rapidly decays with frequency (Fig. 3, green trace). As such, the SS-SSD estimate (Fig. 2D) heavily amplifies non-existing low frequency components that arise from the intrinsic noise in spiking observations, while suppressing the higher frequency components that exist in the true SSD (Fig. 2A). Similarly, the PSTH-SSD estimate shown in Fig. 2E fails to capture most of the spectrotemporal features of the SSD, since it does not account for the binary nature of the observations.

To quantify the performance of these estimators, we repeated this numerical experiment for a total of 50 trials, generating independent realizations of the AR processes and spiking observations in each trial. The average Mean Squared Error (MSE) with respect to the True SSD (in dB scale) and the average spectral leakage across trials are presented in Table I. To compute the spectral leakage, we first define the *in-band* spectra by the components of the True SSD above -10 dB, and then take the relative power of each estimate outside the in-band spectra as the spectral leakage. The MSE

values are normalized with respect to the total power of the True SSD (in dB scale). The proposed PPMT-SSD estimator achieves the lowest MSE, followed by the SS-SSD estimator with a significant gap. The PSTH-SSD estimator exhibits the poorest performance, in terms of both average MSE and variance, which is also visually evident in Fig. 3. The spectral leakage of the SS-SSD and PSTH-SSD estimates are an order of magnitude higher than that of the PPMT-SSD, which is comparable to that of the Oracle SSD.

Table I: Comparison of MSE and Spectral Leakage in Study 1

Estimation method	MSE	Spectral Leakage
Oracle SSD	0.0490 ± 0.0028	$6.79\% \pm 0.36\%$
PPMT-SSD	0.1868 ± 0.0075	$7.69\% \pm 0.89\%$
SS-SSD	0.3906 ± 0.0036	$67.06\% \pm 1.78\%$
PSTH-SSD	1.4777 ± 0.0315	$46.59\% \pm 0.89\%$

It is noteworthy that the superior performance of PPMT-SSD comes at the cost of higher computational complexity: Algorithm 2 has a computational complexity of $\mathcal{O}(JPMWN^2)$, whereas the complexity of SS-SSD and PSTH-SSD are considerably lower at $\mathcal{O}(JPMW \log W)$. As such, the performance gain achieved by the PPMT-SSD estimator comes at the cost of a factor of $\mathcal{O}(\frac{N^2}{\log W})$ increase in computational complexity. In the spirit of easing reproducibility, a MATLAB implementation that regenerates the data, results and figures outlined in this section has been made publicly available on the open source repository GitHub [47].

C. Robustness to the Number of Realizations and Spiking Rate

Here, we evaluate the robustness of the various algorithms with respect to two main data attributes: average spiking rate (f_r) and number of realizations (L). For the same latent processes considered in Study 1, we first vary L , while fixing the average spiking rate at $f_r = 0.28$ spikes/s. Next, we fix $L = 10$, and vary f_r by changing the DC components of the latent processes. Finally, we consider $L = 1$ and vary f_r . The corresponding MSE and spectral leakage performance curves are shown in Fig. 4A, B, and C, respectively.

The PPMT-SSD estimates uniformly outperform the PSTH-SSD estimates in terms of both MSE and spectral leakage. They similarly outperform the SS-SSD estimates, except in terms of the MSE for the case of $L = 1$ and $f_r \leq 1$ spikes/s (Fig. 4A and C). The spectrograms shown in Fig. 2 correspond to the case (a) marked in Fig. 4A ($L = 20$, $f_r = 0.28$ spikes/s). For visual comparison, in Fig. 5 we

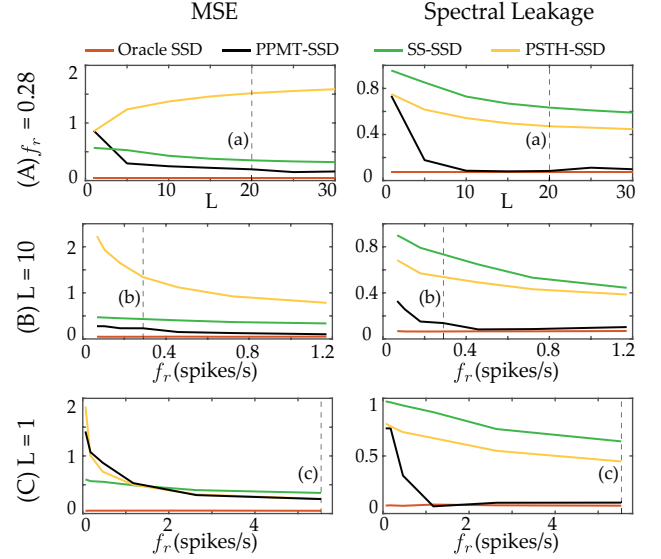


Fig. 4: Analysis of the robustness of the estimates in Study 1 to L (number of realizations) and f_r (average spiking rate). MSE (left) and Spectral Leakage (right) performance curves are computed by varying: (A) L , for $f_r = 0.28$ spikes/s, (B) f_r , for $L = 10$, and (C) f_r , for $L = 1$.

also depict the spectrograms of two other cases from Fig. 4: ($L = 10$, $f_r = 0.28$ spikes/s), marked as case (b) in Fig. 4B, and ($L = 1$, $f_r = 5.5$ spikes/s) marked as case (c) in Fig. 4C. Overall, all estimates improve by increasing L and/or the spiking rate, as expected. It is noteworthy that all estimators, including PPMT-SSD, perform poorly when both f_r and L are small (Fig. 4C). However, even in this regime, the PPMT-SSD estimator exhibits significantly lower spectral leakage, even for $L = 1$. In conclusion, maintaining a trade-off between L and the spiking rate results in substantially improved performance.

D. Choice of Design Parameters and Hyper-parameters

We have chosen $N = 800$ bins in order to have a densely sampled spectral representation. Note that the choice of N directly affects the computational complexity of the estimator. In cases where the spectral range is known *a priori* (e.g., $[0, 2]$ Hz in our simulation setting), the computational complexity can be significantly reduced by using only the frequency bins in the relevant range when constructing the matrix A [13].

The time-bandwidth product of the multitaper framework has been chosen as 2 ($\xi = 2$), and we have used the first three tapers. Note that the design bandwidth of the multitaper framework $\frac{\xi f_s}{W}$ must be chosen to match the spectral spacing

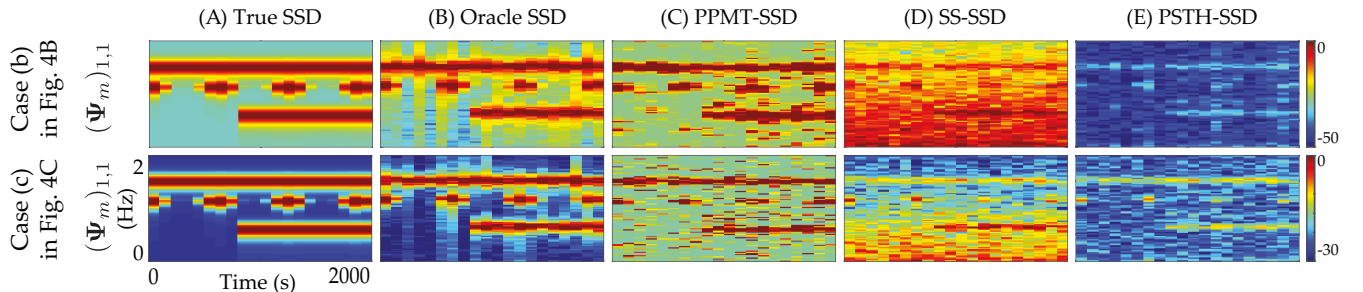


Fig. 5: Estimated $(\Psi_m)_{1,1}(\omega)$ for two selected cases from Fig. 4. Each panel shows the magnitude spectra in dB scale. Top: Case (b) in Fig. 4B, $L = 10$ and $f_r = 0.28$ spikes/s. Bottom: Case (c) in Fig. 4C, $L = 1$ and $f_r = 5.5$ spikes/s. Columns, left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD.

$\frac{f_s}{2N}$. Thus, ξ could be determined based on W and N . For any given ξ , the number of tapers (P) with high spectral concentration within the main lobe is bounded by $2\xi - 1$ [34].

The hyper-parameters in Algorithm 1 are ρ , ζ and α . By appropriately adjusting the magnitude of the state transition parameter $\alpha \in [0, 1]$, using cross-validation or data-driven tuning within the EM framework, the degree of temporal dependency between adjacent time windows can be controlled. Thus, α was chosen as 0.4. The optimal choices of ρ and ζ depend on the average spiking rate and the number of realizations (L). Accordingly, we have set $\rho = 0.2$ and $\zeta = 0.02$ in Study 1. In general, the parameter ζ affects the average noise floor of the resulting spectral estimates and ρ controls the degree of fluctuations. When the spiking data are highly sparse (i.e., average spiking rate and L are small), choosing higher values of ζ and lower values of ρ results in more robust estimates. These parameters can also be systematically tuned through cross-validation via Monte Carlo methods as shown in [13].

E. Study 2: Estimating the SSD of a Bivariate Process, with Only One Directly Observable Component

While the previous study was a natural choice for performance comparison, this study is of particular interest in the joint analysis of neural spiking and continuous signals, such as the local field potential (LFP). The LFP signal corresponds to the electrical field potential measured at the cortical surface, and mesoscale dynamics of cortical activity. We consider a bivariate random process, whose first component $x_{k,1}$ is observed through spiking activity $\{n_{k,1}^{(l)}\}_{k,l=1}^{K,L}$, while its second component $x_{k,2}$ is directly observable in i.i.d. zero-mean Gaussian noise, i.e., $\tilde{x}_{k,2} := x_{k,2} + \nu_k$, with $\nu_k \sim \mathcal{N}(0, \sigma_\nu^2)$. Explicitly, the two processes are given by,

$$x_{k,1} = y_k^{(1)} \cos\left(2\pi \frac{f_0}{f_s} k\right) + y_k^{(4)} + y_k^{(7)} + \sigma_{x1}\nu_{1,k} + x_{1,dc}$$

$$x_{k,2} = 0.83 y_k^{(2)} + 0.83 y_{k-6}^{(4)} + 0.83 y_k^{(5)} + 0.83 y_k^{(6)} + \sigma_{x2}\nu_{2,k} + x_{2,dc}.$$

The process $x_{k,2}$ here is exactly the same as that in Section V-B. To induce additional non-stationarity, we slightly modified the process $x_{k,1}$ by including an additional frequency modulated component. The AR component $y_k^{(7)}$, has been tuned around the frequency f_7 , which changes by decrements of 0.06 Hz every 200 seconds, starting at 0.9 Hz at $t = 0$ s.

The SSD matrix can be estimated by Algorithm 2 with a minor modification in the forward filtering step (step 2) of Algorithm 1: Given that the second process is directly observable, the distribution $f(\mathcal{D}_1^m | \{\mathbf{w}\}_1^m, \boldsymbol{\theta}^{(r)})$ needs to be modified, and accordingly, the log-posterior in Eq. (11) changes to:

$$\sum_{s,w=1}^{m,W} L \left\{ \bar{n}_{(s-1)W+w,1}(\mathbf{A}_s \mathbf{v}_{s,1})_w - \log(1 + \exp(\mathbf{A}_s \mathbf{v}_{s,1})_w) \right\}$$

$$- \sum_{s,w=1}^{m,W} \frac{1}{2\sigma_\nu^2} \left(\tilde{x}_{(s-1)W+w,2} - (\mathbf{A}_s \mathbf{v}_{s,2})_w \right)^2 - \frac{1}{2} \sum_{s=1}^m (\mathbf{w}_s - \Phi \mathbf{w}_{s-1})^T (\mathbf{Q}_s^{(r)})^{-1} (\mathbf{w}_s - \Phi \mathbf{w}_{s-1}).$$

Fig. 6 shows the corresponding estimation results, similarly formatted as in Fig. 3. Note that in this case, we take the SS-SSD and PSTH-SSD estimates of $(\Psi_m)_{2,2}(\omega)$ to be the same as its Oracle SSD estimate, given that these methods are based on estimating the process $x_{k,2}$ in time domain (which is directly observable here). Similar to previous study, the proposed PPMT-SSD estimator (Fig. 6C) captures the dynamics of the spectra $(\Psi_m)_{1,1}(\omega)$ and $(\Psi_m)_{1,2}(\omega)$ accurately, closely matching the True SSD (Fig. 6A). As before, the SS-SSD estimator (Fig. 6D) is not able to capture the SSD dynamics, especially at high frequencies. Though some frequency components at certain time windows are detected by the PSTH-SSD estimates (Fig. 6E), most of the frequency content is concealed by background noise.

VI. APPLICATION TO EXPERIMENTALLY-RECORDED DATA FROM THE RAT CORTICAL NEURONS DURING SLEEP

Finally, we apply our proposed PPMT-SSD estimator to multi-unit recordings from the rat cortical neurons during sleep (data from [37], publicly available in the Collaborative Research in Computational Neuroscience data sharing website [38]). The data set includes spiking activities of putative pyramidal cells (pE) and putative interneurons (pI), as well as LFP signals recorded from deep layers of frontal cortical areas [38]. In order to examine the homeostatic effects of sleep on cortical activity, three main brain states of waking (WAKE), rapid eye movement (REM) sleep, and non-rapid eye movement (nonREM) sleep have been identified and labeled throughout the recordings (See [37] for details). Analyzing the dynamics of neuronal activity during REM and nonREM sleep provides insights into how these brain states differentially contribute to homeostasis. For instance, it has been shown

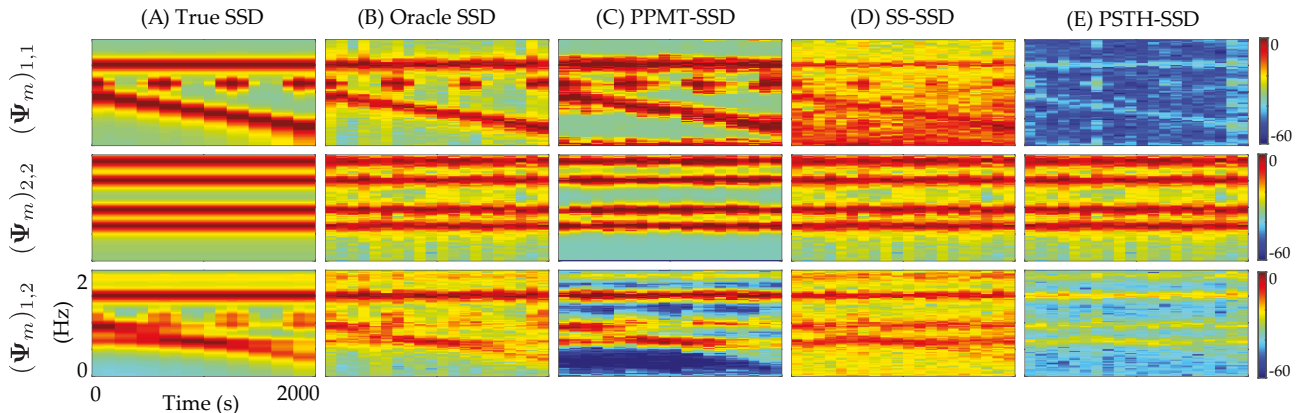


Fig. 6: Estimation results for Study 2. Each panel shows the magnitude spectra in dB scale. Columns, left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD. Rows, top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

in [37] that the LFP signals exhibit increased delta activity (0.5–4 Hz) during nonREM sleep, as compared to WAKE and REM sleep episodes.

While the results of [37] based on spectrotemporal analyses of the LFPs give significant insights into the cortical dynamics during sleep at the mesoscale, our proposed methodology could be utilized to capture the spectrotemporal dynamics of neuronal activity at the neuronal scale and across the pE and pI cell populations. To this end, we consider a bivariate setting ($J = 2$), where spiking observations of pE and pI cells are represented by $n_{k,1}^{(l)}$ and $n_{k,2}^{(l)}$, respectively. We choose 10 spike trains ($L = 10$) from each cell type for the analysis, for a total observation duration of 35 minutes (data from animal BWRat19, premotor cortex/M2 area [38]). Note that we have assumed the activity of the L neurons to constitute independent realizations of a process governed by the same underlying intensity. This assumption is motivated by the fact that cortical neurons are known to phase-lock to global oscillatory signals (governed by subcortical processes) under brain states such as anesthesia and sleep [17], [37], [48], and exhibit the so-called UP-DOWN state dynamics [49], [50].

The average spiking rates of the pE and pI populations amount to $f_r = 2.92$ spikes/s and $f_r = 7.86$ spikes/s, respectively. The spike trains are sampled at $f_s = 64$ Hz. The first 7.5 minutes of the observation duration pertain to the WAKE state, followed by a sleep episode consisting of REM and nonREM epochs. Fig. 7 shows a zoomed-in view of the raster plots of the two neuronal ensembles corresponding to a 30 s window starting at $t = 11$ min.

We assume stationarity within windows of length 30 s, resulting in total $M = 70$ non-overlapping time windows. Focusing our analysis to the delta band (0.5–4 Hz) [37], we estimate the SSDs up to 4 Hz while $N = 512$. Further, we choose $\xi = 2$, and use the first three tapers in the analysis. We set $\alpha = 0.1$, $\zeta = 0.02$ and $\rho = 0.2$. Fig. 8 shows the results of our analysis. The first three columns (from left to right) represent the PSTH-SSD, SS-SSD and PPMT-SSD

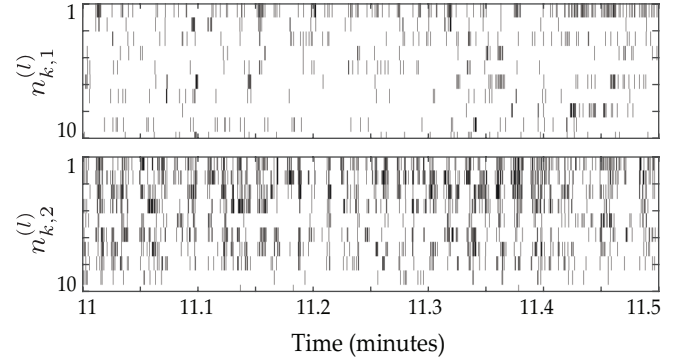


Fig. 7: Spike raster plots of pE cells (top: $n_{k,1}^{(l)}$) and pI cells (bottom: $n_{k,2}^{(l)}$) within the time window of 11–11.5 min, from the data used in the spectrotemporal analysis of Section VI.

estimates. The fourth column illustrates the magnitude spectra corresponding to the time window of 11–11.5 min shown in Fig. 7. Rows from top to bottom correspond to the estimates of $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$, respectively.

The PSTH-SSD estimate (Fig. 8A) is heavily attenuated and concealed by background noise (first column, top panel, and orange traces in the fourth column) and does not capture the spectral variations across REM-nonREM transitions. On the other hand, the SS-SSD estimate (Fig. 8B) captures these variations at the expense of significantly amplifying a wide band of low frequencies and introducing spurious background noise (green traces in the fourth column). The PPMT-SSD estimates (Fig. 8C), however, provide a denoised and well-delineated spectrotemporal representation within and across the brain states. Both pE and pI cells (Fig. 8C, first and second row, respectively), exhibit increased power in the delta band over the nonREM epochs. This is consistently observed in the three main nonREM epochs of 7.5–14 min, 14.5–18 min, and 19–29 min, manifestly within the 0.5–2 Hz frequency band. The cross-spectral coupling between the two cell groups (Fig. 8C, third row) similarly increases within the 0.5–2 Hz band during nonREM epochs, as compared to the REM and WAKE

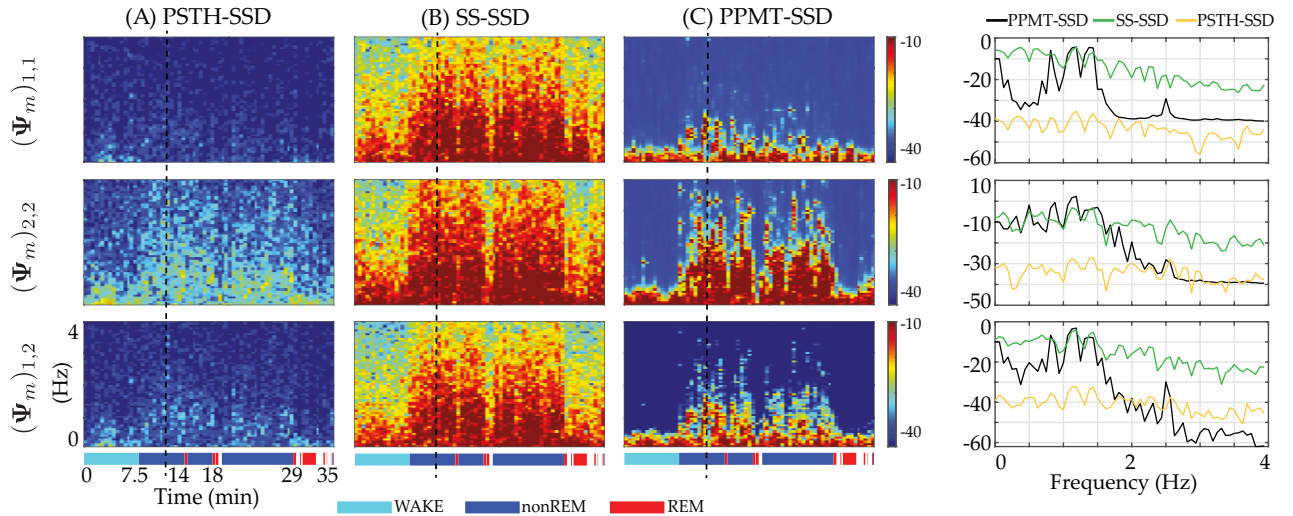


Fig. 8: SSD analysis of rat cortical activity during sleep. Each panel shows the magnitude spectra in dB scale. The first three columns from left to right: (A) PSTH-SSD, (B) SS-SSD, and (C) PPMT-SSD. State labels of WAKE (cyan), nonREM (blue) and REM (red) are indicated at the bottom of each column. The fourth column shows a snapshot of the spectrograms at the 23rd window (11–11.5 min), whose endpoint is marked by dashed vertical lines in the first three columns. Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

states. These observations are consistent with the changes in the LFP spectra across REM/nonREM transitions reported in [37], which indicate steady increase of delta power during nonREM episodes. Note that the spectral power in the 2–4 Hz band exhibits fluctuation during the nonREM epochs, with increased power in only some of the time windows. This variability is likely due to the non-rhythmic and burst-like spiking activity of the ensemble, perhaps modulated by finer state dynamics endogenous to the nonREM episodes.

The spectral snapshots shown in the fourth column of Fig. 8 show a consistent performance pattern as in Fig. 3 from our simulation studies: the spectral peaks are clearly distinguishable in the PPMT-SSD estimates, whereas the SS-SSD and PSTH-SSD estimates flatten the peaks across a wide band of frequencies, with the former undergoing significant attenuation. It is noteworthy that the attenuation of the PPMT-SSD estimates $(\Psi_m)_{1,1}(\omega)$ and $(\Psi_m)_{1,2}(\omega)$ (Fig. 8, fourth column, first and third rows) around 0.5 Hz is likely due to the increase of burst-like activity of the pE neurons in the particular window chosen for generating the snapshots. This attenuation is not consistent across windows, as evident in the spectrogram plots (Fig. 8C, first and third rows), whereas the increased activity of the pI neurons in the 0.5–2 Hz band is persistent during the nonREM episodes (Fig. 8C, second row). Our analysis validates the utility of PPMT-SSD estimation as an alternative to existing methods, which excels in providing spectrotemporal characterizations of multivariate spiking data with high resolution and favorable denoising performance.

VII. CONCLUDING REMARKS

In this work, we proposed a methodology for estimating the SSD matrix of a multivariate non-stationary latent process *directly* from binary spiking observations, thus obviating the need for intermediate time-domain smoothing procedures used by existing techniques, which in turn result in biased estimates of the spectra. To this end, we integrated techniques from state-space modeling, multitaper analysis and point processes. We established theoretical bounds on the bias and variance performance of the proposed estimator, and compared its performance with the existing techniques through application to simulated and experimentally-recorded neural data. Our simulation studies confirmed our theoretical analysis and revealed the favorable performance of our proposed method over existing approaches. Our application to real data provided a highly-resolved characterization of the spectrotemporal dynamics of cortical activity during sleep at the neuronal scale. It is worth noting that in the emerging neural recording technologies, such as two-photon calcium and voltage imaging, the LFP signals are not recorded. Spiking activity, however, can be recovered from these recordings using signal deconvolution techniques, which renders our proposed methodology highly desirable in analyzing these data. A key limitation of our proposed method is the assumption of semi-stationarity, which limits its applicability to more general classes of non-stationary processes. Nevertheless, this assumption seems to be plausible for neuronal spiking data under anesthesia or sleep, where the underlying cortical activity pertains to slowly-varying states,

as well as other types of slowly-varying binary oscillatory processes such as heart beat data [51]. Our methodology can also be extended to infer non-stationary network-level properties such as the frequency domain Granger-Geweke causality [52], [53].

APPENDIX A PROOF OF THEOREM 1

Proof. Let $S(\omega)$ be the PSD of the process $\{x_k\}_{k=1}^K$. Then,

$$\begin{aligned} |\text{bias}(\hat{S}^{\text{mt}}(\omega))| &:= |\mathbb{E}[\hat{S}^{\text{mt}}(\omega)] - S(\omega)| \\ &\stackrel{(a)}{\leq} |\mathbb{E}[\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)]| + |\text{bias}(S^{\text{mt}}(\omega))|, \end{aligned} \quad (18)$$

where (a) follows from the triangle inequality. Further,

$$\begin{aligned} \text{Var}(\hat{S}^{\text{mt}}(\omega)) &:= \mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - \mathbb{E}[\hat{S}^{\text{mt}}(\omega)]|^2] \\ &\leq \mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2] + \text{Var}(S^{\text{mt}}(\omega)) \\ &\quad + 2\mathbb{E}[(\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega))(S^{\text{mt}}(\omega) - \mathbb{E}[S^{\text{mt}}(\omega)])] \\ &\stackrel{(b)}{\leq} \left\{ \sqrt{\mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2]} + \sqrt{\text{Var}(S^{\text{mt}}(\omega))} \right\}^2, \end{aligned} \quad (19)$$

where (b) follows from the Cauchy-Schwarz inequality. Thus, the desired bounds on the bias and variance can be established through bounding the first and second moments of $(\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega))$. The first moment can be bounded by

$$\begin{aligned} |\mathbb{E}[\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)]| &\stackrel{(c)}{\leq} \frac{1}{P} \sum_{p=1}^P \left| \mathbb{E} \left[|\hat{y}^{(p)}(\omega)|^2 - |y^{(p)}(\omega)|^2 \right] \right| \\ &\stackrel{(d)}{\leq} \frac{1}{P} \sum_{p=1}^P \sum_{k=1}^K \sum_{m=1}^K \left| \mathbb{E}[\ell^{-1}(\bar{n}_k)\ell^{-1}(\bar{n}_m) - x_k x_m] \nu_k^{(p)} \nu_m^{(p)} e^{-i\omega(m-k)} \right| \\ &\stackrel{(e)}{\leq} \frac{1}{P} \sum_{p=1}^P \sum_{k=1}^K \left| \mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2] \right| (\nu_k^{(p)})^2 \\ &\quad + \frac{1}{P} \sum_{p=1}^P \sum_{k \neq m} \left| \mathbb{E}[\ell^{-1}(\bar{n}_k)\ell^{-1}(\bar{n}_m) - x_k x_m] \right| \nu_k^{(p)} \nu_m^{(p)}, \end{aligned} \quad (20)$$

where (c) and (d) follow from the triangle inequality and (e) follows by bounding the complex sinusoid. The main technical difficulty in further development of the bounds is due to the fact that $\ell^{-1}(z)$ does not have a Taylor series expansion for $z \in (0, 1)$. We thus have to find other algebraically useful bounds. To this end, we need to establish the following technical lemma.

Lemma 1. *Consider the event $A = \{\bar{n}_k \mid \bar{n}_k \neq 0, \bar{n}_k \neq 1, 1 \leq k \leq K\}$. The following inequality holds true for all $\bar{n}_k \in A$:*

$$\varepsilon(\bar{n}_k) := |\ell^{-1}(\bar{n}_k) - x_k| \leq g(x_k, L) |\bar{n}_k - \lambda_k|,$$

where

$$g(x_k, L) = \max \left\{ \frac{1}{\lambda_k(1-\lambda_k)}, \frac{|\log(L-1) + x_k|}{|\lambda_k - 1/L|}, \frac{|\log(L-1) - x_k|}{|1 - 1/L - \lambda_k|} \right\}.$$

Proof of Lemma 1. First, consider the case $\lambda_k \leq 0.5$. We bound the function $\varepsilon(\bar{n}_k)$ in a piece-wise fashion as follows. Note that $\ell^{-1}(\bar{n}_k)$ is convex for $\bar{n}_k \geq 0.5$ and concave for $\bar{n}_k \leq 0.5$. Thus, it immediately follows that for $\bar{n}_k \leq \lambda_k$, $\varepsilon(\bar{n}_k)$ is convex and hence,

$$\varepsilon(\bar{n}_k) \leq \frac{|\log(L-1) + x_k|}{|\lambda_k - 1/L|} (\lambda_k - \bar{n}_k). \quad (21)$$

Furthermore, for $\lambda_k \leq \bar{n}_k \leq 0.5$, $\varepsilon(\bar{n}_k)$ is concave, and hence is bounded by the tangent at λ_k as follows.

$$\varepsilon(\bar{n}_k) \leq \frac{1}{\lambda_k(1-\lambda_k)} (\bar{n}_k - \lambda_k) \quad (22)$$

Finally, for the case of $\bar{n}_k \geq 0.5$, consider the line

$$h(\bar{n}_k) := \frac{|\log(L-1) - x_k|}{|1 - 1/L - \lambda_k|} (\bar{n}_k - \lambda_k). \quad (23)$$

From the convexity of $\varepsilon(\bar{n}_k)$, $h(\bar{n}_k)$ upper bounds $\varepsilon(\bar{n}_k)$ for $\bar{n}_k \geq 0.5$, since $h(0.5) \geq \varepsilon(0.5)$ for $\lambda_k \leq 0.5$. Combining the piece-wise bounds in Eqs. (21), (22) and (23), we conclude the claim in Lemma 1 for $\lambda_k \leq 0.5$. Due to the symmetry of $\varepsilon(\bar{n}_k)$, through a similar argument, the bound can be established for $\lambda_k > 0.5$, which concludes the proof. $\square$

Given that $|x_k| \leq B$ and assuming that L is large enough so that $L \geq 2(1 + \exp(B))$, we can further simplify the bound of Lemma 1. We have:

$$\begin{aligned} g(x_k, L) &\leq \max \left\{ \exp(B) (1 + \exp(-B))^2, \frac{|\log(L-1) + B|}{(1/(1 + \exp(B)) - 1/L)} \right\} \\ &\leq \max \left\{ \exp(B) (1 + \exp(-B))^2, 4(1 + \exp(B)) \log L \right\}. \end{aligned}$$

Thus, for sufficiently large L , we conclude that,

$$\varepsilon(\bar{n}_k) \leq 4(1 + \exp(B)) \log L |\bar{n}_k - \lambda_k|. \quad (24)$$

Now, consider the expectations in the bounds of Eq. (20). Using iterated conditioning,

$$\begin{aligned} |\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2]| &= |\mathbb{E}[\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 | x_k] - x_k^2]| \\ &= |\mathbb{E}[2x_k \mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k) | x_k] + \mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 | x_k]]| \\ &\stackrel{(f)}{\leq} \mathbb{E}[2|x_k| \mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| | x_k]] + \mathbb{E}[\mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 | x_k]], \quad (25) \end{aligned}$$

where (f) follows from triangle and Jensen's inequalities. In order further simplify these bounds, we invoke the result of Lemma 1. First, note that $\ell^{-1}(\bar{n}_k)$ is unbounded in the complement of event A . Provided $|x_k| \leq B$, $\mathbb{P}(\bar{n}_k \neq 0)$ and $\mathbb{P}(\bar{n}_k \neq 1)$ can be lower bounded by $1 - \exp(-L \log(1 + \exp(-B)))$, which implies that

$$\mathbb{P}(A) \geq 1 - 2\exp(-L \log(1 + \exp(-B))).$$

Therefore, for sufficiently large L , we see that $\mathbb{P}(A)$ is exponentially close to 1. Thus, hereafter we condition the expectations on the highly probable event A . From Eq. (24), we get

$$\begin{aligned} \mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| | x_k, A] &\leq 4(1 + \exp(B)) \log L \\ &\quad \mathbb{E}[|\bar{n}_k - \lambda_k| | x_k, A]. \end{aligned}$$

Note that the random variable $n_k = L\bar{n}_k$ is the sum of L independent Bernoulli random variables given x_k . Thus, given x_k , $n_k \sim \text{Binomial}(L, \lambda_k)$. Accordingly,

$$\begin{aligned} \mathbb{E}[|\bar{n}_k - \lambda_k| | x_k, A] &= \mathbb{E}[|\bar{n}_k - \lambda_k| \mathbb{1}_A | x_k] / \mathbb{P}(A) \\ &\stackrel{(g)}{\leq} \sqrt{\mathbb{E}[(\bar{n}_k - \lambda_k)^2 \mathbb{1}_A | x_k]} / \mathbb{P}(A) \\ &\stackrel{(h)}{\leq} \sqrt{\lambda_k(1 - \lambda_k)} / (\sqrt{L} \mathbb{P}(A)), \quad (26) \end{aligned}$$

where (g) follows from the Jensen's inequality and (h) follows from substituting expression for the variance of a binomial

random variable. Further, note that $\lambda_k(1 - \lambda_k) \leq 1/4$, for $\lambda_k \in [0, 1]$ and $\mathbb{P}(A) \geq 1/2$, if $L \geq 2B/\log(1 + \exp(B))$, which is satisfied for large enough L . Thus, combining the bounds in Eqs. (26) and (24), we get,

$$\mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| | x_k, A] \leq 4(1 + \exp(B)) \frac{\log L}{\sqrt{L}}. \quad (27)$$

By a similar argument we can show that,

$$\mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 | x_k, A] \leq 8(1 + \exp(B))^2 \left(\frac{\log L}{\sqrt{L}} \right)^2.$$

Thus, the expectation in Eq. (25) is bounded as:

$$\begin{aligned} |\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2 | A]| &\leq 8(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \\ &\quad \times \left(B + (1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right). \quad (28) \end{aligned}$$

Following a similar argument, one can show for $n \neq m$,

$$\begin{aligned} |\mathbb{E}[\ell^{-1}(\bar{n}_k) \ell^{-1}(\bar{n}_m) - x_k x_m | A]| &\leq 8(1 + \exp(B)) \\ &\quad \times \frac{\log L}{\sqrt{L}} \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right). \quad (29) \end{aligned}$$

Finally, using the bounds of Eqs. (28) and (29) and noting that $\sum_{k=0}^{K-1} (\nu_k^{(p)})^2 = 1$, for all $0 \leq p \leq P$, we can upper bound the expectation in Eq. (20) as,

$$|\mathbb{E}[\widehat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega) | A]| \leq g_1 K \frac{\log L}{\sqrt{L}}, \quad (30)$$

where

$$g_1 := 8(1 + \exp(B)) \left\{ \left(\frac{1}{K} + 1 \right) B + \left(\frac{1}{K} + 2 \right) (1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right\}.$$

This concludes the proof of the bound on bias. Following similar bounding techniques, the second moment in Eq. (19) can be bounded by:

$$\sqrt{\mathbb{E}[|\widehat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2 | A]} \leq g_2 K \frac{\log L}{\sqrt{L}}, \quad (31)$$

where

$$\begin{aligned} g_2 &:= 4(1 + \exp(B)) \left\{ \frac{\sqrt{2}}{K} \left[\sqrt{\frac{13 \log L}{3 \sqrt{L}}} (1 + \exp(B)) + B \right] \right. \\ &\quad \left. + \left[4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2 \right]^{1/2} \right\}. \end{aligned}$$

This concludes the proof of Theorem 1. $\square$

APPENDIX B PROOF OF COROLLARIES 1 AND 2

Proof of Corollary 1. Proof of Corollary 1 follows the proof of Theorem 1 closely, with the natural extension to the multivariate case. Following the proof of Theorem 1, the constants g'_1 and g'_2 in this case are given by:

$$\begin{aligned} g'_1 &:= 8(1 + \exp(B)) \times \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right) \\ g'_2 &:= 4(1 + \exp(B)) \sqrt{4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2} \end{aligned}$$

$\square$

Proof of Corollary 2. As for Corollary 2, we work under the technical assumptions of Theorems 1 and 2 in [27]. Following [27], we assume that in Eq. (9), $\mathbf{Q}_m = \mathbf{Q}$ for all m in this proof, and that the EM algorithm finds estimates of $\mathbf{Q}$

and α close to their true value (for large enough K). Then, under Assumptions (1) and (2), we identify the effective observation $\tilde{\mathbf{y}}_m^{(p)}$ corresponding to the p^{th} taper at window m in [27] by the concatenation of $\nu_w^{(p)} \text{logit}(\tilde{n}_{(m-1)W+w,j})$ for $w = 1, 2, \dots, W$ and $j = 1, 2, \dots, J$ in a vector of length WJ . We also assume, without loss of generality that $W = uN$, for some integer u . Then, we denote by Σ_∞ the steady state covariance of the backward smoother, and $\Lambda := \alpha \Sigma_\infty (\alpha^2 \Sigma_\infty + \mathbf{Q})^{-1}$ and $\Gamma = (\alpha^2 \Sigma_\infty + \mathbf{Q})[\mathbf{I} - uW((\alpha^2 \Sigma_\infty + \mathbf{Q})^{-1} + uW\mathbf{I})^{-1}]$, as in [27]. Note that these matrices are $NJ \times NJ$ in our case. Under the same assumptions [27], we consider them to be diagonal with the i^{th} diagonal element being γ_i and η_i respectively. Then, following the proof of Theorem 1, Corollary 1 and those of Theorems 1 and 2 in [27], it can be shown that the statement of the corollary holds with the constants:

$$g_1''(\omega_n) := 8(1 + \exp(B)) \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right) \\ \times \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|}, \\ g_2''(\omega_n) := 4(1 + \exp(B)) \sqrt{4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2} \\ \times \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|},$$

and

$$\kappa_m(\omega_n) := \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|} \alpha^{|s-s'|}.$$

We refer the interested readers to [27] for detailed derivations. $\square$

REFERENCES

- [1] A. Rupasinghe and B. Babadi, "Multitaper analysis of evolutionary spectral density matrix from multivariate spiking observations," in *2019 IEEE Data Science Workshop (DSW)*, Minneapolis, MN, June 2–5, 2019.
- [2] A. Schnitzler and J. Gross, "Normal and pathological oscillatory communication in the brain," *Nat. Rev. Neurosci.*, vol. 6, pp. 285–296, 2005.
- [3] P. J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology," *Neuron*, vol. 52, no. 1, pp. 155 – 168, 2006.
- [4] L. M. Ward, "Synchronous neural oscillations and cognitive processes," *Trends in Cognitive Sciences*, vol. 7, no. 12, pp. 553 – 559, 2003.
- [5] R. P. Vertes and R. W. Stackman, *Electrophysiological Recording Techniques*. Humana Press, 2011, vol. 54.
- [6] J. J. Jun, N. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. Anastassiou, A. Andrei, C. Aydin, M. Barbic, T. Blanche, V. Bonin, J. Couto, B. Dutta, S. Gratiy, D. Gutnisky, M. Husser, B. Karsh, and T. D. Harris, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, pp. 232–236, 11 2017.
- [7] J. Du, T. J. Blanche, R. R. Harrison, H. A. Lester, and S. C. Masmanidis, "Multiplexed, high density electrophysiology with nanofabricated neural probes," *PLOS ONE*, vol. 6, no. 10, pp. 1–11, 10 2011.
- [8] E. N. Brown, R. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature Neuroscience*, vol. 7, pp. 456–61, 06 2004.
- [9] A. V. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, "Modeling single-neuron dynamics and computations: a balance of detail and abstraction," *science*, vol. 314, no. 5796, pp. 80–85, 2006.
- [10] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *Journal of Neurophysiology*, vol. 93, no. 2, pp. 1074–1089, 2005.
- [11] L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," *Network: Computation in Neural Systems*, vol. 15, no. 4, pp. 243–262, 2004.
- [12] L. Paninski, J. Pillow, and J. Lewi, "Statistical models for neural encoding, decoding, and optimal stimulus design," *Progress in brain research*, vol. 165, pp. 493–507, 2007.
- [13] S. Miran, P. L. Purdon, E. N. Brown, and B. Babadi, "Robust estimation of sparse narrowband spectra from neuronal spiking data," *IEEE Trans. on Biomedical Engineering*, vol. 64, no. 10, pp. 2462–2474, Oct 2017.
- [14] D. M. Halliday and J. R. Rosenberg, *Time and Frequency Domain Analysis of Spike Train and Time Series Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 503–543.
- [15] A. C. Smith and E. N. Brown, "Estimating a state-space model from point process observations," *Neural Computation*, vol. 15, no. 5, pp. 965–991, 2003.
- [16] P. Das and B. Babadi, "Multitaper spectral analysis of neuronal spiking activity driven by latent stationary processes," *Signal Processing*, vol. 170, no. 107429, 2020.
- [17] L. D. Lewis, V. S. Weiner, E. A. Mukamel, J. A. Donoghue, E. N. Eskandar, J. R. Madsen, W. S. Anderson, L. R. Hochberg, S. S. Cash, E. N. Brown, and P. L. Purdon, "Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness," *Proc. Natl. Acad. Sci.*, vol. 109, no. 49, pp. 19 891–19 892, 2012.
- [18] S. Grün, M. Diesmann, and A. Aertsen, "Unitary events in multiple single-neuron spiking activity: II. Nonstationary data," *Neural Computation*, vol. 14, no. 1, pp. 81–119, 2002.
- [19] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, 1998.
- [20] L. Cohen, *Time-frequency Analysis: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.
- [21] M. B. Priestley, *Spectral analysis and time series*. Academic Press London ; New York, 1981.
- [22] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Trans. on Information Theory*, vol. 52, no. 3, pp. 1067–1086, March 2006.
- [23] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of non-stationary processes," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, December 1985.
- [24] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 27, no. 2, pp. 204–237, 1965.
- [25] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes," *IEEE Trans. on Signal Processing*, vol. 45, no. 6, pp. 1520–1534, June 1997.
- [26] J. Hammond and P. White, "The analysis of non-stationary signals using time-frequency methods," *Journal of Sound and Vibration*, vol. 190, no. 3, pp. 419 – 447, 1996.
- [27] P. Das and B. Babadi, "Dynamic Bayesian multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 66, no. 6, pp. 1394–1409, March 2018.
- [28] S.-E. Kim, M. K. Behr, D. Ba, and E. N. Brown, "State-space multitaper time-frequency analysis," *Proc. Natl. Acad. Sci.*, vol. 115, no. 1, pp. E5–E14, 2018.
- [29] E. N. Brown, R. Barbieri, U. T. Eden, and L. M. Frank, "Likelihood methods for neural spike train data analysis," *Computational neuroscience: A comprehensive approach*, pp. 253–286, 2003.
- [30] U. T. Eden, L. M. Frank, R. Barbieri, V. Solo, and E. N. Brown, "Dynamic analysis of neural encoding by point process adaptive filtering," *Neural computation*, vol. 16, no. 5, pp. 971–998, 2004.
- [31] A. Sheikhattar, J. B. Fritz, S. A. Shamma, and B. Babadi, "Recursive sparse point process regression with application to spectrotemporal receptive field plasticity analysis," *IEEE Trans. on Signal Processing*, vol. 64, no. 8, pp. 2026–2039, 2015.
- [32] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.
- [33] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty V: the discrete case," *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May 1978.
- [34] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.
- [35] B. Babadi and E. N. Brown, "A review of multitaper spectral analysis," *IEEE Trans. on Biomedical Engineering*, vol. 61, pp. 1555–1564, 2014.
- [36] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 40, no. 12, pp. 2941–2946, Dec 1992.

- [37] B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, and G. Buzsáki, "Network homeostasis and state dynamics of neocortical sleep," *Neuron*, vol. 90, no. 4, pp. 839 – 852, 2016.
- [38] —, "Multi-unit spiking activity recorded from rat frontal cortex (brain regions mPFC, OFC, ACC, and M2) during wake-sleep episode wherein at least 7 minutes of wake are followed by 20 minutes of sleep," *CRCNS.org*, 2016. [Online]. Available: <http://dx.doi.org/10.6080/K02N506Q>
- [39] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007, vol. 2.
- [40] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Berlin, Heidelberg: Springer-Verlag, 1986.
- [41] A. T. Walden, "A unified view of multitaper multivariate spectral estimation," *Biometrika*, vol. 87, no. 4, pp. 767–788, 2000.
- [42] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of time series analysis*, vol. 3, no. 4, pp. 253–264, 1982.
- [43] H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, aug 1965.
- [44] P. D. Jong and M. J. Mackinnon, "Covariances for smoothed estimates in state space models," *Biometrika*, vol. 75, no. 3, pp. 601–602, 1988.
- [45] K. Lij and M. Rosenblatt, "Prolate spheroidal spectral estimates," *Statistics & Probability Letters*, vol. 78, no. 11, pp. 1339 – 1348, 2008.
- [46] C. Goutis and G. Casella, "Explaining the saddlepoint approximation," *The American Statistician*, vol. 53, no. 3, pp. 216–224, 1999.
- [47] "Multitaper Analysis of Evolutionary Spectra from Multivariate Spiking Observations MATLAB Codes," 2019. [Online]. Available: <https://github.com/Anuththara-Rupasinghe/PPMT-ESD-Estimation>
- [48] S. Chauvette, S. Crochet, M. Volgushev, and I. Timofeev, "Properties of slow oscillation during slow-wave sleep and anesthesia in cats," *Journal of Neuroscience*, vol. 31, no. 42, pp. 14998–15008, 2011.
- [49] V. V. Vyazovskiy, B. A. Riedner, C. Cirelli, and G. Tononi, "Sleep homeostasis and cortical synchronization: II. A local field potential study of sleep slow waves in the rat," *Sleep*, vol. 30, no. 12, pp. 1631–1642, 2007.
- [50] V. Crunelli and S. W. Hughes, "The slow (< 1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators," *Nature Neuroscience*, vol. 13, no. 1, p. 9, 2010.
- [51] R. Barbieri, E. C. Matten, A. A. Alabi, and E. N. Brown, "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 288, no. 1, pp. H424–H435, 2005.
- [52] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American statistical association*, vol. 77, no. 378, pp. 304–313, 1982.
- [53] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.



Behtash Babadi (S'08–M'14) received the B.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran in 2006, and the M.Sc. and Ph.D. degrees in Engineering Sciences from Harvard University, Cambridge, MA in 2008 and 2011, respectively. From 2011 to 2013, he was a post-doctoral fellow at the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology, Cambridge, MA as well as at the Department of Anesthesia, Critical Care and Pain Medicine at Massachusetts General Hospital, Boston, MA. He joined the Department of Electrical and Computer Engineering at the University of Maryland, College Park, MD in 2014, where he is currently an Associate Professor, with a joint appointment in the Institute for Systems Research. He received an NSF CAREER Award in 2016. His research interests include statistical and adaptive signal processing, neural signal processing, and systems neuroscience.



Anuththara Rupasinghe (S'16) received the B.Sc. degree (first class honors) in Electrical and Electronic Engineering from the University of Peradeniya, Sri Lanka in 2016. She is currently working towards the Ph.D. degree in Electrical and Computer Engineering at the University of Maryland, College Park, MD, USA, where she received the A. James Clark School of Engineering Distinguished Graduate Fellowship in 2017 and the Outstanding Teaching Assistant Award in 2018. Her research interests include statistical signal processing, machine learning and computational neuroscience.