Sparse model-based estimation of functional
dependence in high-dimensional field and spike
multiscale networks

To cite this article: Ramin Bighamian et al 2019 J. Neural Eng. 16 056022

View the article online for updates and enhancements.
Sparse model-based estimation of functional dependence in high-dimensional field and spike multiscale networks

Ramin Bighamian\textsuperscript{1}, Yan T Wong\textsuperscript{2, \*}, Bijan Pesaran\textsuperscript{3} and Maryam M Shanechi\textsuperscript{1, \*}\textsuperscript{4}

\textsuperscript{1} Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States of America
\textsuperscript{2} Department of Physiology, and Electrical and Computer Systems Engineering, Monash University, VIC, Australia
\textsuperscript{3} Center for Neural Sciences, New York University, New York, NY, United States of America
\textsuperscript{4} Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, United States of America

E-mail: shanechi@usc.edu

Received 5 December 2018, revised 22 April 2019
Accepted for publication 17 May 2019
Published 10 September 2019

Abstract

\textbf{Objective.} Behavior is encoded across multiple scales of brain activity, from binary neuronal spikes to continuous fields including local field potentials (LFP). Multiscale models need to describe both the encoding of behavior and the conditional dependencies in simultaneously recorded spike and field signals, which form a high-dimensional multiscale network. However, learning spike-field dependencies in high-dimensional recordings is challenging due to the prohibitively large number of spike-field signal pairs, which makes standard learning techniques subject to overfitting. \textbf{Approach.} We present a sparse model-based estimation algorithm to learn these multiscale network dependencies. We develop a multiscale encoding model consisting of a point process model of binary spikes for each neuron whose firing rate is a function of the LFP network features and behavioral states. Doing so, spike-field dependencies constitute the model parameters to be learned. We resolve the parameter learning challenge by forming a constrained optimization problem to maximize the likelihood with an L1 penalty term that eases the detection of significant spike-LFP dependencies. We then apply the Akaike information criterion (AIC) to force a sparse number of nonzero dependency parameters in the model. \textbf{Main results.} We validate the algorithm using simulations and spike-field data from two non-human primates (NHP) in a 3D motor task with motor cortical recordings and a pro-saccade visual task with prefrontal recordings. We find that by identifying a model with a sparse set of dependency parameters, the algorithm improves spike prediction compared with models without dependencies. Further, the algorithm identifies significantly fewer dependency parameters compared with standard methods while improving their spike prediction likely due to detecting fewer spurious dependencies. Also, spike prediction on any electrode improves by including LFP features from all electrodes compared with using only those on the same electrode. Finally, unlike standard methods, the algorithm uncovers patterns of spike-field network dependencies as a function of distance, brain region, and frequency band. \textbf{Significance.} This algorithm can help study functional dependencies in high-dimensional spike-field networks and leads to more accurate multiscale encoding models.
Keywords: functional dependence, multiscale data, neural encoding models, spikes, local field potentials (LFP)

(Some figures may appear in colour only in the online journal)

1. Introduction

Neural representations of behavior involve multiple spatial and temporal scales of neural activity from binary spiking activity of individual neurons to continuous field activity including local field potentials (LFP) or electrocorticogram (ECoG), which pool activity across large numbers of neurons [1–8]. Simultaneously recorded spike and field signals on an array of electrodes constitute a high-dimensional multiscale network and allow for modeling the encoding of behavior across multiple spatiotemporal scales. A critical aspect of building a multiscale encoding model is to learn the functional dependencies across scales, i.e. between spikes and fields in the network [9, 10]. Investigation of these dependencies can advance our understanding of neural network representations, lead to more accurate encoding models for behavior, and inform future neurotechnologies such as motor brain-machine interfaces (BMIs) [11–23] or other closed-loop systems such as deep brain stimulation (DBS) systems for neurological and neuropsychiatric disorders [24–32]. Such investigations require developing a rigorous mathematical framework to learn a model that describes not only the encoding of behavior, but also the dependencies that exist between spikes and fields in high-dimensional multiscale network recordings.

Prior studies have largely focused on studying dependencies between a pair of spike and LFP signals using correlation-based measures such as spike-LFP coherence in different behavioral tasks (e.g. [5, 33–36]). Spike-LFP coherence measures linear predictability between the LFP signal from one electrode and spike times on another electrode as a function of frequency, and thus provides a valuable tool to assess the degree of pairwise dependence. However, computing the coherence does not aim to model the simultaneous modulation of LFP and spikes due to a common behavior, whose variation will cause covariation in spikes and LFPs. Indeed, part of the computed coherence could be due to common behavioral modulation rather than conditional dependencies between spike and LFP signals—i.e. dependencies that persist even when the effect of a common behavior is removed or conditioned on. While it could be possible to mitigate this issue by studying stationary or epoched behaviors—i.e. keeping the behavior fixed in the interval used for coherence analyses—it is not easy to do so for rich behaviors that are non-stationary or un-epoched, for example when studying the motor cortex during a continuously-evolving movement (see discussions section 4.2).

Additionally, studying dependencies across a network of recorded spike and LFP signals on an array of electrodes requires a simultaneous assessment of coherences between the spikes on a given electrode and hundreds to thousands of LFP features at different electrodes and frequency bands. Such a simultaneous assessment involves adequately addressing multiple comparison issues because of the large number of coherence pairs that are computed separately. Also, to get a more accurate measure of direct dependency between the spikes on an electrode and any recorded LFP signal, the effect of other LFP signals on the spikes should be considered simultaneously, which may not be easy with a pairwise measure. Finally, computing correlation-based measures does not provide an encoding model to describe the representation of behavior as well as conditional dependencies in the spike-field multiscale network. Recent studies have taken important steps by developing models to investigate the spike-LFP conditional dependencies during behavior but they have focused on learning these dependencies on a single electrode [37, 38]. As neural representations involve large-scale neural networks, new algorithms are required to learn multiscale encoding models that include both the effect of behavior and the conditional dependencies between the spikes on a given electrode and network field features across all electrodes on an array (see also discussions section 4.2).

Building multiscale encoding models that incorporate dependencies between spikes and fields in a large network poses challenging computational problems. First, multiscale modeling of spikes and fields requires taking into account the different statistical profiles of these signals. In particular, spikes are binary-valued time-series with a 1 representing a spike and a 0 lack thereof. In contrast, fields are continuous-valued time-series. Recent work has developed multiscale decoders for spike-field activity that take into account these differences but has done so under the assumption that spikes and fields are independent conditioned on the behavior [6, 7]. While this simplifying assumption allows for developing tractable real-time decoders, it does not model the conditional dependencies that exist between spikes and fields and are essential to our understanding of neural representations [10, 39]. Second, there is a prohibitively large number of dependencies corresponding to all possible pairs of spike and field feature signals in the recordings that need to be learned. This is because current technology can record spikes and fields simultaneously from tens to hundreds of electrodes. Finally, the duration of data recorded for model training is often limited relative to the number of dependencies that need to be learned. Together, these aspects create a challenging learning problem that is prone to overfitting.

In this paper, we resolve the above challenges by developing a sparse model-based estimation algorithm to identify conditional dependencies in high-dimensional spike-field networks during behavior. We identify these dependencies through a multiscale encoding model of spikes and field signals by building a combined likelihood function. To incorporate dependencies, we model the spikes as binary point processes whose instantaneous firing rate depends not only on behavior but also on all field features in the network (i.e. on all electrodes). Thus spike-field dependencies are characterized
as parameters in this rate function, which need to be learned. To address the overfitting problem, we introduce a sparsity constraint in an optimization problem that solves for the large number of model dependency parameters by maximizing the likelihood of spike observations. We first employ a sparsity-promoting L1 penalty to ease the detection of field features with a significant contribution to neuronal spikes. We then use the Akaike information criterion (AIC) to further enforce sparsity by incorporating only a small number of dependencies into the model based on a trade-off between model complexity and accuracy in predicting spike events. We validate the sparse model-based estimation algorithm using both simulated data and experimental data collected from two non-human primates (NHP) under two different recording and behavioral setups: motor cortical recordings during a 3-dimensional (3D) motor task and prefrontal recordings during a pro-saccade visual task.

Despite the large number of possible dependencies, our algorithm identifies an encoding model that (1) improves spike prediction compared with models without dependencies, (2) robustly identifies only a sparse set of dependencies over the large network unlike standard methods, and (3) improves spike prediction of these standard methods likely due to detecting fewer spurious dependencies. Using the algorithm, we also find that spike prediction on a given electrode is improved when using LFP features from other electrodes in addition to those on the same electrode. We further use our algorithm to investigate spike-field network dependencies as a function of distance, brain region, and LFP frequency band. We find that different frequency bands differentially contribute to spike prediction. Also, the spike-field dependencies attenuate with distance in the network for high-frequency LFP features but this attenuation is less pronounced for low-frequency LFP features. Finally, without imposing the sparsity constraint, none of these trends are uncoverable likely due to the large number of falsely-detected dependencies. Taken together, the developed algorithm can learn sparse multiscale neural encoding models for high-dimensional spike-field networks. It can thus serve as a new tool to decipher conditional spike-field dependencies while dissociating the effect of behavioral modulation, and while simultaneously modeling the effect of the large number of LFP signals from all electrodes on the spikes from any electrode. The algorithm may also guide neurotechnologies by providing more accurate encoding models.

2. Methods

In this section, we first present the multiscale encoding model. We then develop the sparse model-based estimation algorithm to identify significant spike-field conditional dependencies in a recorded network of spike and field signals during behavior (figure 1). The sparse model-based estimation algorithm consists of a sparsity promoting optimization step combined with AIC model selection as described in sections 2.2 and 2.3, respectively. We also describe the simulation setup and the experimental data used for model validation.

2.1. Multiscale encoding model

We denote the binary spike event of neuron $c$ at time $t$ by $N_c^t$, which is 1 if a spike occurs and 0 otherwise. We denote all the field features in the network by $\mathbf{y}$ (figure 1(B)). We can then write the multiscale encoding model using a combined likelihood function as

$$p(N_c^t, \mathbf{y}, \mathbf{x}) = p(N_c^t | \mathbf{y}, \mathbf{x}_c) p(\mathbf{y}, \mathbf{x}_c)$$

where $\mathbf{x}_c$ denotes the brain state or behavior (e.g. intended kinematics) at time $t$. The field features can for example be taken as power features at different frequency bands. Writing the combined likelihood function in this way, we can summarize the multiscale dependencies between a given neuron $c$ and the field features (figure 1(B)) in the conditional likelihood function for spikes, i.e. in $p(N_c^t | \mathbf{y}, \mathbf{x}_c)$. This conditional likelihood function links neuronal spikes to the behavior and field features simultaneously. In turn, $p(\mathbf{y}, \mathbf{x}_c)$ can be modeled as desired for example as a linear function of behavior with Gaussian noise [7, 40–46]. Thus the challenge in learning the multiscale encoding model is to estimate the large number of conditional dependency parameters in $p(N_c^t | \mathbf{y}, \mathbf{x}_c)$, which is the focus of our work.

We model the spiking activity of each individual neuron $c$ as a point process characterized by the instantaneous firing rate function $\lambda_c(t)$ [47–51]. For a given task with behavioral states $\mathbf{x}$, and field features $\mathbf{y}$, the firing rate for neuron $c$ at time $t$ is given by

$$\lambda_c(t | \mathbf{x}, \mathbf{y}) = \exp(\beta_c + \alpha_c^T \mathbf{x}_c + \omega_c^T \mathbf{y}_c),$$

where $\beta_c$ is the baseline firing rate, $\alpha_c$ models the modulation of firing rates with behavioral states, and $\omega_c$ are the parameters modeling the dependencies between all the field features in the network and the spiking activity and are high-dimensional. Note that since our goal is to study dependencies between spikes and fields and also motivated by prior work [46–49, 51–56], here we assume that spiking activity of different neurons are conditionally-independent conditioned on the behavior and field features. We bin the spikes in sufficiently small time-bins $\Delta$, containing at most one spike. The point process likelihood function for neuron $c$ is then given by [48, 49]

$$p(N_c^t | \mathbf{x}_c, \mathbf{y}_t) = (\lambda_c(t | \mathbf{x}_c, \mathbf{y}_t)\Delta)^{N_c^t} \exp(-\lambda_c(t | \mathbf{x}_c, \mathbf{y}_t)\Delta).$$

Our goal in this work is to estimate the dependencies by developing a method to learn the set of parameters $\theta_c = \{\beta_c, \alpha_c, \omega_c\}$ for each neuron using the measurements of spikes from neuron $c$, field features all over the network (i.e. on all electrodes), and any observable behavioral states (e.g. kinematics). This is a difficult learning problem since the number of spike-field dependencies across the network (i.e. dimension of $\omega_c$) is prohibitively large (e.g. about 700 per neuron in our motor cortical recordings).

2.2. Sparsity-promoting optimization to recover the dependencies

One approach to learn the model parameters is to use maximum-likelihood estimation (MLE), which solves for the
parameter values that maximize the likelihood of observed spikes under the model in (3) as

\[
\mu_{c}^{ML} = \arg \max_{\mu_c} \sum_{i=1}^{T} \log(p(N_{c, i} | \mathbf{x}_i, y_i)).
\]  

(4)

However, given the hundreds of dependency parameters for the network, MLE will lead to overfitting and thus obtaining spurious dependencies between spikes and fields. We thus re-formulate the MLE by introducing a sparsity-promoting penalty within the maximization. This penalty allows us to uncover the most significant dependency terms using the available training data. Specifically, we use an L1 penalty term—which penalizes the L1 norm of the dependency parameters—thus biasing the estimated model parameters to be small. This penalty alters the MLE to identify the subset of field features with the most contribution to

Figure 1. Diagram for estimation of sparse functional dependencies between neuronal spikes and high-dimensional network field features. (A) We validate the algorithm in three scenarios: (1) simulated spike-field data in a center-out reaching kinematic task, (2) NHP motor cortical spike-LFP recordings during an arm movement task, and (3) NHP prefrontal spike-LFP recordings during a pro-saccade visual task. (B) Spike and LFP data are recorded from an array of electrodes in the motor or prefrontal brain regions. For illustration purposes, four LFP signals \( \{y_i\} \) and one spike signal \( N_c \) are shown. (C) L1-penalization is employed to ease the detection of LFP features with significant large contribution to neuronal spikes. A hypothetical problem with dependencies \( \omega_c = [\omega_1, \omega_2]^T \) is shown where \( \omega_1 \) and \( \omega_2 \) are representative insignificant (insensitive) and significant (sensitive) dependency terms, respectively. For illustrative purposes, the contours of the log-likelihood function \( L \) (in (3)) around the dependencies identified by maximum-likelihood estimation (i.e. by (4)) are shown in blue (the log-likelihood decreases as we go away from the maximum-likelihood estimation). The triangle-shaped L1 penalty is shown in red. The sparse estimation is close to the point at which the blue contours first touch the triangle-shaped L1 penalty, which gives the solution to (5). The output of this step is the estimation of functional dependence parameters \( \bar{\omega}_c \). (D) AIC is applied to further enforce sparsity in the identified dependencies from step (C). Dependencies are identified by considering the tradeoff between the model spike prediction performance and model complexity. The output of this step is the estimated sparse set of functional dependence parameters \( \bar{\omega}^* \). (E) Estimated model with identified sparse dependencies is evaluated in cross-validation and on a test dataset. In our hypothetical example, red crosses indicate the time points at which spikes happen. Black dashed line stands at the upper level at the time points of model predicted spikes, for a single spike detection threshold on the model-predicted firing rate. The area under the receiver operating characteristic curve (obtained by trying multiple spike detection thresholds) is used to compute the spike prediction performance (section 2.6).
spiking activity (e.g. $\omega_2$ in figure 1(C) as a representative significant dependency term) and to focus on fitting their dependency parameters instead of all dependency parameters. The sparsity-promoting optimization is written as (see figure 1(C)):

$$\mu_{c}^{L1} = \{\beta_i, \alpha_i, \omega_i\}^{L1}$$

$$= \arg \max_{\mu_i} \left( \sum_{i=1}^{T} \log(p(N_t^c[x_i, y_i])) - \gamma \|\omega_c\|_1 \right)$$

(5)

where $\gamma > 0$ is the L1 penalty term, $\|\cdot\|_1$ represents the L1 norm operator, and field features $y_i$ are z-scored to have a mean of zero and standard deviation of one. The magnitudes of estimated $\omega_c$ parameters indicate the scaled contribution of field features to the neuronal spikes, i.e. the dependencies. We solve this maximization problem using a gradient-based optimization search function [57].

### 2.3. Model selection via AIC

While the insignificant functional dependency parameters in $\omega_c$ identified under the sparsity-promoting penalty term in (5) are biased to be small, they will still all be nonzero terms. Thus, a model selection criterion is needed to only include the dependency parameters that lead to a meaningful increase in spike prediction performance. To identify only the significant dependencies and consequently build a parsimonious model with a limited number of dependencies, i.e. nonzero $\omega_c$ elements, we further enforce model sparsity by applying AIC [58]. AIC is employed as the selection criterion to determine the appropriate level of sparsity. AIC adds dependency parameters to the model one-by-one in order of their magnitude in the estimated $\omega_c$ in (5). It terminates adding new dependency parameters as soon as they show trivial improvement in spike prediction or equivalently when the minimum AIC is reached. In other words, the termination occurs based on a trade-off between the goodness of spike prediction performance and model complexity. The full procedure is designed as follows.

Given $I$ total field features in the network, we build $I$ models corresponding to 1 to $I$ dependency parameters. We thus learn $\omega^{(i)}_c$, for $i = 1, \ldots, I$, where $i$ indicates the number of nonzero dependencies included in the model. For the $i$th model, we select which $i$ dependency parameters to include by picking the largest $i$ dependencies fitted in (5) with the penalty term. For example, $i = 1$ for each neuron $c$ corresponds to including only the field feature whose corresponding dependency parameters in the $\omega_c$ fitted from (5) has the maximum magnitude. For each $i$, we then re-fit the model parameters $\mu^{AIC(i)}_c = \{\beta^{(i)}_c, \alpha^{(i)}_c, \omega^{(i)}_c\}$ using (5). However, this time, $\omega^{(0)}_c$—with the same dimension as $\omega_c$—only has $i$ nonzero spike-field dependencies, which are associated with the largest $i$ contributing field features in $\omega_c$ (note in refitting the $i$th model, we only include the $i$ field features with the largest dependencies, which is equivalent to setting all the other dependency parameters to zero). For the $i$th refitted model, we compute the AIC as

$$AIC^{(i)}_c = 2i - 2 \sum_{t=1}^{T} \log(p(N_t^c[x_i, y_i]; \mu^{AIC(i)}_c))$$

$$= 2i - 2 \sum_{t=1}^{T} [N_t^c(\log(\Delta) + \beta + \alpha^T_c x_t + \omega^T_c y_t)$$

$$- \Delta \exp(\beta + \alpha^T_c x_t + \omega^T_c y_t)]$$

(6)

where the evaluation in the equation above is done at $\mu^{AIC(i)}_c = \{\beta^{(i)}_c, \alpha^{(i)}_c, \omega^{(i)}_c\}$.

Once AIC is computed, the iteration number $i^*$ that gives the minimum AIC indicates the optimum number of contributing field features. We define $\omega^{*}_c = \omega^{(i^*)}_c$, and the $i^*$ nonzero $\omega^{*}_c$ elements denote the parameters associated with the spike-field dependencies for neuron $c$ (figure 1(D)). For computational efficiency in our data analyses here, we only re-fit the model parameters for $i \leq 80$.

### 2.4. Behavioral tasks and neural data recording

We test the sparse model-based estimation algorithm on a set of simulated data as well as experimental data collected from two NHPs as described below.

#### 2.4.1. Simulated data

We simulated spike trains for 30 neurons and 150 field features during a center-out-and-back reaching task, where 2D position in the task traces out the path shown in figure 1(A(1)) over the course of 176 s for 11 cycles. Each cycle includes eight center-out-and-back cursor movements, going from the center to one of eight peripheral targets and then back to the center [7, 46]. In the simulations, we pick the model parameters in (2) to get a maximum firing rate of 60–70 Hz for each neuron. We also pick the parameters such that the contribution to spike modulation from behavioral states versus field features is comparable. Thus, in our simulations, spikes are assumed to be equally modulated by behavior and LFPs. We simulate the spikes according to the point process model in (3) [59]. We take the behavioral state $x$, to be the 4-dimensional vector of positions and velocities in the horizontal and vertical directions. We relate each neuron to ten random field features. We select the time-bin $\Delta$ to be 1 ms (such that it at most contains one spike).

#### 2.4.2. Experimental data

To provide a comprehensive test of the modeling approach and to study the trends in conditional dependencies between spikes and LFP features uncovered by this approach, we use NHP data in two behavioral tasks and two neural systems: motor cortical recordings during a 3D motor task (figure 1(A(2))) and prefrontal recordings during a pro-saccade task (figure 1(A(3))). We now describe the two experimental setups.

**Motor task.** This dataset consists of seven sessions recorded from a Rhesus macaque [60]. The subject was trained and instructed to perform a reach to objects located at different locations in 3D space for a liquid reward (figure 1(A(2))). An array consisting of 137 electrodes (Gray Matter Research,
USA) covering the dorsal premotor cortex (PMd), ventral premotor cortex (PMv), primary motor cortex (M1), and prefrontal cortex (PFC) was used for recordings. Neighboring electrodes were spaced 1.5 mm apart in each of the X and Y directions while their depth in the Z direction was adjusted by the experimenter. While performing the task with the contra-lateral arm, reflective body markers attached to the subject’s skin tracked the movement of the arm, and were captured by the near infrared cameras with sampling frequency of 100 frames s$^{-1}$ (Motion Analysis Corp., USA). A set of seven joint angles corresponding to the shoulder, elbow and wrist movements were used as the behavioral state covariates $s_{x}$ in (2) (shoulder elevation, elevation angle, shoulder rotation, elbow flexion, pro supination, wrist flexion, and wrist deviation).

Neural data was recorded with a sampling rate of 30kHz. As described in [60], LFP signals were extracted by applying a low-pass filter with 400 Hz cut-off frequency on the raw signals; spikes were obtained by band-pass filtering the raw signal from 0.3–6.6kHz and then identifying the threshold crossings below the mean filtered signal. The threshold was set to 3.5 standard deviations [60]. We eliminated low-quality noisy spike channels by considering 30 neurons with the best spike prediction performance in response to movement (see section 2.6 for spike prediction performance). The pairwise distance from the selected neurons to the network LFP electrodes spanned the range of 0–27.9 mm. Neural signals were down-sampled to 100 Hz. More details about this task and neural pre-processing can be found in [60].

**Pro-saccade visual task.** This dataset consists of five sessions recorded from a second Rhesus macaque [61]. This NHP performed a visually-guided oculomotor delayed response task with a fixed head position. An electrode array consisting of 32 electrodes covering PFC was used for recordings (Gray Matter Research, USA). Neighboring electrodes were spaced 1.5 mm in the X and Y directions while their depth in the Z direction was adjusted by the experimenter. Each trial started with the monkey fixating at the center of the screen for a baseline period (500–800 ms), right after illumination of a central fixation target. Next, the animal was shown a spatial cue at a peripheral target location. After a delay period (1000–1500 ms), the central fixation square was extinguished and followed by a Go signal for the animal to start saccading to the target location. Within the delay period and saccade, the spatial cue remained illuminated. The animal was given liquid reward upon a successful trial. The spatial cue was presented at one location on a grid of eight locations that were spaced on the corners of the screen as well as centers of the screen sides. Spatial cue locations were interleaved trial-by-trial in equal proportions. Eye position was monitored with an infrared optical eye tracking system with a sampling rate of 120 Hz. A trial was not successful if the animal’s gaze was not aligned within 2° of the center of the fixation target before the Go command or within 2° of the center of a spatial cue target following an initially correct saccade after Go. Within the successful trials, we analyzed the continuous neural data associated with three consecutive epochs: 0.5 s before the Go cue until the Go cue appears (−0.5), during the saccade (sac), and within 0.5 s after the end of saccade (+0.5). Due to evident task-relevant changes in the spiking activity between epochs, three different levels of baseline firing rate each associated with one epoch were considered for the point process model of spikes in (2), i.e. $\beta_{c} = \beta_{c, t} \in \{ \beta_{c,0.5}, \beta_{c,sac}, \beta_{c,0.5} \}$.

Spike events were classified from the raw signal during each recording session and resorted offline as described in [61]. Spikes were obtained by high-pass filtering the raw signal at 300 Hz while LFP signals were computed by low-pass filtering at the same frequency. The sampling rate of the raw neural signal was 30kHz. Single-unit activity was isolated by thresholding at 3.5 standard deviations below the mean, performing a principal component analysis (PCA) of putative spike waveforms, over-clustering these waveforms in PCA space using kmeans and then merging clusters based on visual inspection as described in [61]. Among the 32 neurons in each session, on average only ten sorted neurons showed isolated clusters and were subject to further analysis. The pairwise distance from the selected neurons to the network LFP electrodes spanned the range of 0–10.5 mm. For each sorted neuron, the preferred direction was identified as the direction in which the neuron shows the maximum number of spikes. For each preferred direction, neural signals were down-sampled to 500 Hz before they were used in this work. More details about this task and neural pre-processing can be found in [61].

### 2.5. LFP feature extraction

LFP features in the experimental tasks were taken to be the log power of z-scored LFP signals within different frequency bands. The LFP frequency spectrum was divided into the following bands, similar to prior studies [62]: $\theta$ (4–8 Hz), $\alpha$ (8–12 Hz), $\beta$ (16–24 Hz), low $\gamma$ (25–55 Hz), and high $\gamma$ (65–140 Hz). The LFP power across time for each band was computed by using a sliding time window of 300 ms.

### 2.6. Model evaluation

We test the sparse model-based estimation algorithm using cross-validation. For the simulated data, we use nine cycles of the center-out-and-back reaching task to train the model and the remaining two cycles to test it. For the experimental data, a two-fold cross-validation was performed in each session. In both simulations and experimental data analysis, we compare three classes of encoding models: (1) an encoding model without incorporating the conditional dependencies (i.e. using only the behavior), (2) an encoding model with both behavior and conditional dependency terms estimated using MLE (i.e. without the penalty but using AIC for model selection), and (3) an encoding model with the behavior and conditional dependency terms estimated using the sparse model-based estimation algorithm. We used various performance criteria for the simulated and experimental data as detailed below.

#### 2.6.1. Simulated data performance measures

Two sparsity penalty levels, $\gamma = 0$ and $\gamma = 150$, were considered to evaluate the effectiveness of the sparsity penalty. We refer to estimations with $\gamma = 0$ as unpenalized estimation and to...
estimations with a non-zero $\gamma$ as penalized estimation. We report the results for both sparsity penalty levels and also for the encoding model solely modulated by kinematics, i.e. for $\omega_e = 0$. Since in simulations we know the ground-truth dependency terms, as our measure of performance, we report the true positives (TP) and the false positives (FP) in estimating dependencies as well as the number of retained field features, i.e. number of dependencies, in the estimated models. As an aggregate measure combining TP and FP, we report the F-score [63]. We also report how well the estimated models predict the spike events (0 or 1) for different neurons (figure 1(E))—i.e. the spike prediction performance—using the area under the receiver operating characteristic curve (AUC) to compute the predictive power (PP), defined as $2\text{AUC}-1$ [1, 38]. A perfect predictor gives a PP score of 1, while one with random guesses has a PP score of 0.

2.6.2. Experimental data performance measures. Conditional dependencies were identified with unpenalized MLE and sparse model-based estimation algorithm. The level of the penalty imposed in the sparse model-based estimation was tuned for each session using the method explained in section 2.7. We first find the PP scores for the encoding model with no dependencies, i.e. $\omega_e = 0$, when spikes are solely modulated by their task-epoch-dependent baseline activity in the pro-saccade task $(\beta_{e,j})$ or by the baseline activity together with the kinematics in the motor task $(\beta_e, \mathbf{x}_i)$. As our baseline for comparisons, we use the PP score for this model with no dependencies. We then report the PP score improvement over this baseline, denoted by $\%PP$, for the two models with dependencies estimated with the MLE and with the sparse model-based algorithm, respectively. Note that in the pro-saccade task, PP scores for the model with no dependencies may be close to zero. Thus we computed the $\%PP$ improvement across neurons with baseline PP score of at least 0.05. We also find the number of detected dependencies or equivalently the number of retained LFP features in these two models. We expect that, compared to the model with no dependencies, an optimum sparse estimation of dependencies would lead to a model with significantly larger PP score, which is achieved with a small number of retained LFP features.

In addition to PP score, $\%PP$, and number of LFPs retained, we also report the pattern of dependencies in the spike-LFP network uncovered by our sparse model-based estimation algorithm. In particular, we quantify (1) the spatial pattern of conditional dependencies between LFP and neuronal spiking activity (i.e. the LFP influence on spikes as a function of distance between the spike and LFP electrodes), (2) the spatial pattern of spike-LFP dependencies for different LFP frequency bands, and (3) the spatial pattern for positive versus negative spike-LFP dependencies. For the motor task in which we have recordings from multiple motor cortical regions, we also investigate the pattern of spike-LFP dependencies among different regions. Finally, to further examine the importance of LFP contribution from the entire network (i.e. all electrodes) to the neuronal spikes on a given electrode, we repeat the sparse model-based estimation algorithm but this time we only study dependencies between neurons and LFP features recorded on the same electrode. We compare the resulting spike PP score improvement using the LFP features on the same electrode that records the spikes to that using all the LFP features on all electrodes.

2.7. Tuning of L1-sparsity penalty level

We tuned the sparsity penalty term $\gamma$ for each session of experimental data using an inner-level cross-validation within the training set. In particular, we performed a 5-fold cross validation analysis with three different penalties, $\gamma \in \{50, 100, 150\}$ on half of the data (training set) in each session. For each penalty, we computed the average (over folds) number of identified dependencies denoted by $N$ (equivalent to the average number of retained LFP features) as well as the estimated PP denoted by $PP$. Then, we formed their weighted sum as

$$M = \frac{\mu_1(1 - N/I) + \mu_2PP}{\mu_1 + \mu_2} \quad (7)$$

where $I$ is the maximum possible number of spike-LFP dependencies, i.e. 80. In this work, we assumed $\mu_1 = \mu_2 = 0.5$. Thus to balance the PP and number of retained LFP features, in each session, we selected the penalty $\gamma$ that achieved the maximum $M$.

2.8. Examining the spatial patterns of dependencies

We examine the spatial patterns of dependencies, i.e. the dependency between LFP features and spikes as a function of the distance between their electrodes in the 3D volume of brain tissue. However, the distances between spike-LFP electrode constitutes a large set of unique and different distances. Thus we discretize the 3D volume into rings around a given neuron such that we can obtain a statistically reliable dependency pattern by averaging the dependency parameters for the LFP features within each ring.

For a neuron recorded on a given electrode, there are also LFP features recorded on the same electrode, which provide the zero-distance dependencies. For zero-distance patterns, we just average the results across all neurons. While we apply commonly-used filtering procedures to obtain LFP and spike signals [64, 65], part of the LFP contribution to spikes on the same electrode may be due to spike leakage [65]. Dissociating same-electrode LFP contributions to spikes that are due to leakage from those that are due to other sources such as synaptic events is an important topic [64, 66], which is not the focus of our study. Here, our goal is to show that we can identify the LFP contribution to spikes regardless of their source. Further, we use the same-electrode (zero-distance) LFP modeling mainly to compare with when the LFP features from all electrodes are modeled to show that there are significant LFP contributions to spikes beyond that on the same electrode (see also discussions section 4.5).

We group the rest of the dependency terms at nonzero distances in rings around the neurons with 1 mm thickness and
inner radii that increase in 1 mm increments (starting from 0 mm). For each neuron, the dependencies for each ring are then averaged across LFP features in that ring and summarized at a distance equal to the ring’s outer radius. The results for each ring are finally averaged across all neurons. To ensure each ring contains enough number of features, we report the results for the features placed within \(d \leq 17\) mm in the motor task and \(d \leq 8\) mm in the pro-saccade task since the number of LFP channels thereafter drops significantly (less than the average number of LFP channels contained in a given ring).

2.9. Measures for spike-LFP dependency quantification over distance and frequency band

To assess how the identified dependencies contribute to spike prediction over distance, for each neuron \(c\) we introduce a measure called normalized power contribution of LFP features, defined as

\[
\chi(d) = \frac{\|\omega_c(d)^T y_c(d)\|_2}{\|\omega_c(0)^T y_c(0)\|_2},
\]

where superscript \(d\) denotes distance from neuron \(c\), \(y_c(d)\) denotes all LFP features at that distance, \(\omega_c(d)\) denotes the dependency terms corresponding to these LFP features, and \(\| \cdot \|_2\) denotes the 2-norm. This quantity measures how much the LFP features at a given nonzero distance \(d\) contribute to spike prediction of a given neuron compared with the LFP features on the same electrode on which the neuron is recorded. \(\chi(d)\) is computed for all distances surrounded by each ring, then averaged and summarized at 1 mm increments, as elaborated above in section 2.8.

We also define a similar measure of normalized power contribution of features over different frequency bands \(f\), as

\[
\chi(df) = \frac{\|\omega_c(df)^T y_c(df)\|_2}{\|\omega_c(0)^T y_c(0)\|_2},
\]

where the additional superscript \(f\) indicates the particular frequency band of LFP features and their corresponding dependency parameters. At a given distance \(d\) from a given neuron, this quantity measures the normalized contribution of different LFP frequency bands to spike prediction of that neuron. We compute this measure separately for positive and negative dependencies \(\omega_c(df)^+\), \(\chi(df)^+\) is also computed for all distances surrounded by each ring, then averaged and summarized at 1 mm increments.

Finally, to assess how the identified dependencies distribute over distance, we report the proportion of contributing LFP features, defined as the ratio (%) of the number of contributing features at a particular distance to the total number of features at the same distance (again computed in rings at 1 mm increments).

3. Results

We demonstrate the sparse model-based estimation algorithm on the simulated data and on two experimental datasets. We also investigate the distribution of estimated dependencies over distance and over spectral frequency bands. Finally, we examine the pattern of dependencies between different brain regions for the motor task.

### Table 1. Average spike predictive power (PP) and its improvement over the model with no dependency terms (%PP), number of retained field features per neuron, TP/TN, and F-score for the simulation data under \(\gamma = 150\) (i.e. estimated spike-field dependency with penalty), \(\gamma = 0\) (i.e. estimated spike-field dependency without penalty) and \(\omega_c = 0\) (i.e. the model with no dependency terms).

<table>
<thead>
<tr>
<th>(\gamma)</th>
<th>PP</th>
<th>%PP</th>
<th>Retained LFP</th>
<th>TP/FP</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ± SEM</td>
<td>Mean ± SEM</td>
<td>Mean ± SEM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma = 150)</td>
<td>0.29 ± 0.02\textsuperscript{a,b}</td>
<td>109 ± 21\textsuperscript{b}</td>
<td>16 ± 1\textsuperscript{a}</td>
<td>240/249</td>
<td>0.61</td>
</tr>
<tr>
<td>(\gamma = 0)</td>
<td>0.27 ± 0.02\textsuperscript{b}</td>
<td>94 ± 24\textsuperscript{b}</td>
<td>29 ± 17\textsuperscript{b}</td>
<td>271/605</td>
<td>0.46</td>
</tr>
<tr>
<td>(\omega_c = 0)</td>
<td>0.15 ± 0.01</td>
<td>0 ± 0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\textsuperscript{a}p < 0.05 with respect to \(\gamma = 0\) (paired t-test, \(N = 30\)).

\textsuperscript{b}p < 0.05 with respect to \(\omega_c = 0\) (paired t-test, \(N = 30\)).

3.1. Algorithm validation on simulated data

We first used the simulated data to validate the sparse model-based estimation algorithm. First, we found that the model with dependencies outperformed the model with no dependencies in spike prediction performance (i.e. PP score). Comparing the average results among the three models quantified in table 1, the model with dependencies improves the PP score of spiking activity compared to the case with no dependencies regardless of the sparsity penalty level used (see also figure 2). From table 1 and figure 2(a), PP scores for the model modulated by both field features and kinematics under both \(\gamma = 0\) and \(\gamma = 150\) are significantly larger than PP for the model of spikes solely affected by kinematics (\(P < 10^{-5}\), paired t-test, \(N = 30\)), resulting in average improvements of 94% and 109%, respectively.

Second, we found that using the sparsity penalty results in more accurate detection of the spike-field dependencies in the network (table 1 and figure 3). Figure 3(b) shows that without the penalty term that is incorporated in the sparse model-based estimation algorithm, a large portion of the non-existing dependency terms (i.e. those that are equal to zero) are estimated with large values that are indistinguishable from the estimated values for existing dependency terms. Figure 3(a) shows that including the sparsity penalty alleviates this issue by lowering the estimated value of many of these non-existing dependency terms (i.e. estimated spike-field dependency without penalty), \(\omega_c = 0\) (i.e. the model with no dependency terms).
the number of true dependencies detected by the unpenalized model showed only 13% improvement compared with that with the penalized model. Consequently, the F-score for the sparse model-based estimation algorithm was 0.61 while it was only 0.46 without the penalty term (table 1).

Third, despite detecting much fewer dependency terms, the penalized estimation of dependencies does not degrade spike prediction performance, and indeed even improves it. Figure 2(b) shows that the number of dependencies identified under $\gamma = 150$ is significantly smaller than that under $\gamma = 0$ ($P = 10^{-9}$, paired t-test, $N = 30$; see table 1). Importantly, despite the smaller number of retained dependencies, the PP score using $\gamma = 150$ compared to $\gamma = 0$ does not degrade and actually even improves ($P = 0.047$, paired t-test, $N = 30$; see table 1 and figure 2(a)). These results suggest that without the penalty term many of the detected dependencies are spurious and due to overfitting.

It is noted that simulated spike channels are sparsely connected to LFP channels (each spike channel is only connected to ten random LFP channels), and thus identifying these sparse dependencies can significantly improve the PP without loss of performance due to potential small contributions from the remaining insignificant LFPs (see large %PP in table 1 under the sparse (i.e. penalized) model-based estimation of dependencies).

3.2. Sparse model-based estimation algorithm improves spike prediction performance on NHP data

We examined the model performance using the experimental data in both motor and pro-saccade tasks.

First, we found that in both tasks and regardless of the sparsity penalty level included for dependency estimation, the model with spike-LFP dependencies had better average spike prediction performance (i.e. PP score) compared to the model with no dependencies (i.e. with $\omega_s = 0$ in equation (2)) (table 2 and figure 4). In the motor task (figure 4(a) and table 2), the PP improvements (i.e. %PP) averaged across sessions with and without the penalty term were 31% and 13%, respectively ($P = 10^{-15}$ and $P = 0.03$, respectively, paired t-test, $N = 14$). From figure 4(a), with the penalty term, %PP is highly significantly positive for every session individually as well ($P < 10^{-5}$, paired t-test, $N = 60$). Similarly, in the pro-saccade task (see figure 4(c) and table 2), PP improvements averaged across sessions with and without the penalty term were 24% and 14%, respectively ($P = 10^{-5}$ and $P = 2 \times 10^{-4}$, respectively, paired t-test, $N = 10$). Figure 4(c) shows that with the penalty term, %PP is significantly positive in every session as well ($P < 0.02$, $14 \leq N \leq 28$). Note that while LFP features are themselves correlated, for estimating the dependency parameters with the spikes, they are observable variables in the model. We performed a control analysis in which we first decorrelated the LFP features before enforcing sparsity to see if this step changed the spike prediction performance compared to when LFP features were intact in the motor task sessions. We found that adding this decorrelation step before enforcing sparsity does not significantly change the spike prediction performance ($P = 0.3$, paired t-test $N = 7$).

Second, estimating the spike-LFP dependencies with the sparse model-based estimation algorithm resulted in improved spike prediction performance compared to estimating them with MLE (i.e. without the penalty term). Compared with the latter, the PP scores averaged across sessions and obtained by the sparse model-based estimation algorithm were significantly improved by 22% in the motor task and 10.9% in the pro-saccade task ($P < 7 \times 10^{-4}$, paired t-test, $N = 14$ and $N = 10$, respectively; figures 4(a) and (c) and table 2). Note in the pro-saccade task, PP scores for MLE may be close to zero; thus the %PP improvement is computed across the neurons with unpenalized PP score of at least 0.05. Moreover, while estimating the dependencies using the sparse model-based estimation algorithm resulted in a highly significant %PP improvement compared with the model with no dependency in every session, this was not the case for MLE. In particular, including dependencies estimated by MLE did not significantly improve the PP in one session in the motor task (second session, $P = 0.11$, paired t-test, $N = 60$) and in two sessions in the pro-saccade task (first and fifth sessions, $P > 0.05$, paired
t-test, \( N = 20 \) and 18, respectively). These results suggest that the sparse model-based estimation algorithm computes the dependencies more accurately compared with MLE, which does not incorporate the penalty term.

Third, the sparse model-based estimation algorithm improved spike prediction performance compared with MLE while using a significantly smaller number of spike-LFP dependency terms. Averaged across sessions (figures 4(b) and (d) and table 2), the number of dependency terms retained in the sparse model-based algorithm was as small as 9% and 26% of those used in the MLE in the motor and pro-saccade tasks, respectively \((P < 6 \times 10^{-11}\), paired t-test, \( N \) for motor and pro-saccade tasks are 14 and 10, respectively). These results again suggest that the sparse model-based estimation algorithm detects a sparse number of significant dependencies and that MLE likely detects many spurious dependencies leading to degraded spike prediction.

### 3.3. Sparse model-based estimation algorithm uncovers a decreasing pattern for the contribution of LFPs to spikes over distance in the network

We used the sparse model-based estimation algorithm as a tool to investigate whether the identified dependencies show any specific spatial pattern over distance from a recorded neuron. Figure 5 presents how the identified spike-LFP dependencies by the algorithm distribute (figures 5(a) and (c)) and contribute (figures 5(b) and (d)) to spikes over distance. Results are averaged across all the recording sessions in the motor (figures 5(a) and (b)) and pro-saccade (figures 5(c) and (d)) tasks and presented in 1 mm increments around neurons—each increment is summarized and shown as a red square (see section 2.8). To show the distribution, figures 5(a) and (c) shows the proportion of contributing LFP features and to further show the contribution, figures 5(b) and (d) present
the normalized power contribution of LFP features $\chi^{(d)}$, both averaged across all the cross-validation folds (see section 2.9 for definition of measures). In addition to comparing these measures at different distances, to further quantify how LFP features impact the neuronal spikes over distance, we performed a hypothesis test on coefficients of a regression model defined as an exponential function $\rho_1 + e^{-(d-\rho_2)}$, which was fitted to the proportion of contributing LFP feature curves in figures 5(a) and (c) and to the normalized power contribution of LFP feature curves in figures 5(b) and (d).

First, to study the distribution over distance, we compared the proportion of contributing LFP features to spikes at different distances. We found that the proportion of LFP features contributing to spikes sharply drops versus distance in the motor task, while spikes are influenced by an approximately exponential function $\rho_1 + e^{-(d-\rho_2)}$ in the pro-saccade task sharply decays with the spike-LFP distance in the network, while the decay in the pro-saccade task was not as sharp (figures 5(b) and (d)). In particular, in the motor task, the zero-distance LFP contribution was larger than all other distances ($P < 10^{-13}$, two-sample t-test with Bonferroni correction for multiple comparisons), while it was more comparable to the nonzero distances in the pro-saccade task with a statistically significant drop observed only at the two largest distances ($i.e.$ $d = 7$ and 8 mm) ($0.03 < P < 0.40$, two-sample t-test with Bonferroni correction for multiple comparison). We further confirmed this result using our regression analysis. We observed that the slope of the fitted exponential curve at $d = 0$ for the motor task was almost 1.5 times that in the pro-saccade task—$−0.62$ versus $−0.41$, respectively—again suggesting that the contribution of LFP features on spiking activity has a sharper drop over distance in the motor task (see figures 5(b) and (d)). This result is consistent with the trend for the proportion of contributing LFP features, which has a sharp drop in the motor task but has a non-decreasing trend in the pro-saccade task (figures 5(a) and (c)).

It is important to note that the patterns uncovered above were not uncoverable using MLE (i.e. without the sparsity penalty), likely due to the large number of false positive dependencies detected (see section 3.1). Thus these results show the utility of the sparse model-based estimation algorithm as a scientific tool to uncover patterns of conditional dependencies simultaneously across a large multiscale network (i.e. many electrodes on the array). Finally, we made the above comparison between the motor and pro-saccade tasks to later link the sharper spatial attenuation of LFP influence on spikes in the motor task to the more dominant contribution of high-frequency LFP features in that task versus the more balanced contribution of low- and high-frequency features in the pro-saccade task (see sections 3.5.3 and 4.3).

### Table 2. Average spike predictive power (PP) and its improvement over the model with no dependency terms (%PP), and number of retained LFP features per neuron for motor and pro-saccade tasks under $\gamma \neq 0$ (i.e. estimated spike-LFP dependency with penalty), $\gamma = 0$ (i.e. estimated spike-LFP dependency without penalty) and $\omega_c = 0$ (i.e. the model with no dependency terms).

<table>
<thead>
<tr>
<th>Measure</th>
<th>PP Mean ± SEM</th>
<th>%PP Mean ± SEM</th>
<th>Retained LFP Mean ± SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor task</td>
<td>$\gamma \neq 0$</td>
<td>$0.26 ± 0.006_{ab}$</td>
<td>$31 ± 1.4_{ab}$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0$</td>
<td>$0.23 ± 0.005_b$</td>
<td>$13 ± 1.7_b$</td>
</tr>
<tr>
<td></td>
<td>$\omega_c = 0$</td>
<td>$0.20 ± 0.005$</td>
<td>$0 ± 0$</td>
</tr>
<tr>
<td>Pro-saccade task</td>
<td>$\gamma \neq 0$</td>
<td>$0.29 ± 0.021_{ab}$</td>
<td>$24 ± 5_{ab}$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 0$</td>
<td>$0.27 ± 0.02_b$</td>
<td>$14 ± 6_b$</td>
</tr>
<tr>
<td></td>
<td>$\omega_c = 0$</td>
<td>$0.24 ± 0.02$</td>
<td>$0 ± 0$</td>
</tr>
</tbody>
</table>

$^a p < 0.05$ with respect to $\gamma = 0$ (paired t-test, $N \in \{14,10\}$).

$^b p < 0.05$ with respect to $\omega_c = 0$ (paired t-test, $N \in \{14,10\}$).

3.4. Sparse model-based estimation algorithm uncovers the spectral distribution of spike-LFP dependencies

We used the sparse model-based estimation algorithm to study whether the spike-LFP dependencies across the network show any specific pattern as a function of frequency bands of LFP features.

In the motor task, we found that the majority of LFP features that are identified to contribute to neuronal spikes are from the high $\gamma$ band as shown in figure 6(a). Also, for the motor task, 94% of the studied neurons were influenced by at least one high $\gamma$ LFP feature, while this proportion becomes as small as 24%, 15%, 29% and 28% for the $\theta$, $\alpha$, $\beta$, and low $\gamma$ LFP features, in order. Indeed these proportions are different across different bands ($P = 2 \times 10^{-154}$, Kruskal–Wallis test
between five groups, \(N = 420\) and the proportion for the high \(\gamma\) band is significantly higher than every other band in the motor task (\(P < 8 \times 10^{-8}\), Wilcoxon rank sum tests with Bonferroni correction for multiple comparison, \(N = 420\)), demonstrating its strong conditional dependency with spikes. Note that these proportions are computed by finding the contributing LFP features across all electrodes (i.e. at all distances from the neuron). Thus these proportions dissociate the overall influence of the LFP network on spikes in different frequency bands (see also discussion section 4.4).

In comparison, in the pro-saccade task, spiking activity is more comparably affected by the LFP features in high and low frequency regimes, i.e. high \(\gamma\) and \(\theta\). The number of contributing LFP features per neuron was highest for high \(\gamma\) and \(\theta\) as shown in figure 6(b). Also, the proportion of the studied neurons that were influenced by at least one feature in a given band were 65%, 49%, 36%, 39%, 66% for \(\theta\), \(\alpha\), \(\beta\), low \(\gamma\), and high \(\gamma\), in order. Thus there were differences in what proportion of neurons were affected by different frequency bands (see also discussion section 4.4).

There are two additional points to emphasize. First, while we built an encoding model that modeled the effect of the entire LFP network from all electrodes on a given neuron simultaneously, our results are qualitatively consistent with prior pairwise coherence analyses in the pro-saccade task showing that among the low frequency LFP features in PFC, \(\theta\) band is the major contributor to neuronal spikes [68, 69]. This consistency further validates our network modeling framework. Second, importantly, again none of these patterns were uncoverable with MLE, suggesting the importance of the sparse model-based approach in obtaining accurate encoding models and detecting non-spurious dependencies.

### 3.5. Spectro-spatial distribution of spike-LFP dependencies over the entire network

After examining the spatial and spectral distribution of dependencies separately, we studied their spectro-spatial distribution simultaneously. In these analyses, we also distinguished positive from negative dependencies, i.e. separately computed our proportion and power contribution measures for positive bands differentially influence spiking activity, with high \(\gamma\) and \(\theta\) bands having the most influence in the pro-saccade task (figure 6).
and negative $\omega(d_f)$ (see equation (9)). In particular, we studied both the proportion of contributing LFP features per neuron (figures 7(a) and 8(a)) and their power contribution to each neuron $\chi(d_f)$ in equation (9) (figures 7(b) and 8(b)) at different frequency bands over a wide distance around the neurons. In all the analyses that follow, we used the sparse model-based estimation algorithm. None of the following patterns were uncoverable using MLE (i.e. without the sparsity penalty), again likely due to the large number of spurious dependencies with MLE.
3.5.1. Distribution of positive and negative spike-LFP dependencies across LFP frequency bands. We found that high frequency network LFP features in the motor and pro-saccade tasks are mostly positively correlated with spikes (i.e. have a positive $\omega^*(df)$ associated with them), while the lower frequency bands in the network showed both positive and negative correlation to spikes (figures 7 and 8). Comparing the proportion of positive and negative contributing LFP features at each frequency band in the motor task, we found that estimated dependencies at $\theta$ and high $\gamma$ bands are mostly positively affecting the neuronal spiking activity ($P < 0.01$, paired t-test, $N = 17$), dependencies at $\beta$ and low $\gamma$ bands are dominantly negatively affecting the spikes ($P < 0.007$, paired t-test, $N = 17$), and no significant difference was seen between the positive and negative dependencies at $\alpha$ band ($P = 0.09$, paired t-test, $N = 17$) (see figure 7(a)). For the prosaccade task (see figure 8(a)), LFP features at $\theta$, $\alpha$, and high $\gamma$ are largely positively related to neuronal spikes ($P < 0.008$, paired t-test, $N = 8$), while neurons are predominantly influenced negatively by $\beta$ and low $\gamma$ features ($P < 0.03$, paired t-test, $N = 8$). Same observations were made using the power contribution of LFPs (see figures 7(b) and 8(b)). Note that these results are again quantifying the positivity and negativity of the contributions across the entire network (i.e. at all spatial distance for the motor task, estimated by the sparse model-based estimation algorithm. The results are separately shown for LFP features that positively or negatively contribute to spikes. SEM is presented as vertical dashed lines. Proportion of contributing LFPs is the proportion (%) of contributing features at a particular distance among the entire set of features at the same distance. Power contribution of LFP features is $\chi^{(df)}$ as defined in section 2.9, where superscript $d$ and $f$ denote distance from the neuron and frequency band, respectively. Proportion of contributing LFP features and power contribution of LFP features are averaged between all cross-validation folds and summarized at 1 mm incremental distances.
distances from a given neuron). Finally, our results at zero-distance are consistent with prior work that studies relationships between spikes and LFPs on the same electrode [37] (see also discussion section 4.4).

3.5.2. Spatial distribution of spike-LFP dependencies as a function of LFP frequency bands. We also found that the proportion of contributing high frequency LFP features to spikes significantly drops over distance, while change over distance is not pronounced for the low frequency LFP features, indicating the long-range impact of low frequency features to neuronal spikes (figures 7 and 8). In each task, for each frequency band, we first found which one of the negative versus positive dependencies had the dominant contributing proportion, i.e. compared the blue versus red proportions in figures 7(a) and 8(a)) (see section 3.5.1). We then studied the pattern of the dominant proportions for each frequency band (i.e. either red or blue) over distance. In particular, for each frequency band, we computed how much the proportion of contributing LFPs drops from zero distance to maximum distance, i.e. 18 mm in the motor and 8 mm in the pro-saccade task. We observed that the proportion of contributing LFPs showed a significant drop over distance in high $\gamma$ band ($P = 1 \times 10^{-95}$ in the motor and $P = 4 \times 10^{-4}$ in the pro-saccade task, two-sample Wilcoxon rank sum tests), while it remained mostly constant or showed a less pronounced change in the lower frequency bands (motor: $P = 0.01$ for $\theta$ and $P > 0.11$ for $\{\alpha, \beta, \text{low } \gamma\}$; pro-saccade: $P > 0.052$ for $\{\theta, \alpha, \beta, \text{low } \gamma\}$, two-sample Wilcoxon rank sum tests).

We further confirmed these results using a statistical regression analysis by fitting linear models to proportion of positive or negative contributing LFP features to spikes—again, whichever has the dominant proportion, see section 3.5.1—at different frequency bands (figures 7(a) and 8(a)). First we

![Figure 8. Proportion of contributing LFP features (a) and power contribution of LFP features (b) at different frequency bands over spatial distance for the pro-saccade task, estimated by the sparse model-based estimation algorithm. Figure convention is the same as in figure 7.](image-url)
examined the high-frequency bands. In both the motor and the pro-saccade tasks, the proportion of positively-contributing features at high \( \gamma \) band showed a significant drop over distance with the slopes of \(-0.33\) and \(-0.39\), respectively (\(P < 0.035\), F-test, \(N = 17\) in the motor and \(N = 8\) in the pro-saccade task). These drops were from their statistically nonzero intercepts of 4.7 and 4.4, respectively (\(P < 1 \times 10^{-3}\), F-test, \(N = 17\) in the motor and \(N = 8\) in the pro-saccade task). For the lower frequency bands, in the pro-saccade task, the proportion of positively-contributing features at \(\theta\) and \(\alpha\) bands and negatively-contributing features at low \(\gamma\) band did not show significant changes over distance (\(P > 0.11\), F-test, \(N = 8\)) from their statistically significant nonzero intercepts of 1.2, 1.2, and 0.6, respectively (\(P < 0.05\), F-test, \(N = 8\)). On the other hand, the proportion of negatively-contributing features at \(\beta\) band showed a significant increase over distance (slope 0.1; \(P = 0.02\), F-test, \(N = 8\)) from a statistically insignificant intercept of 0.4 (\(P = 0.06\), F-test, \(N = 8\)). Note that due to the small proportion of contributing features at the first four frequency bands in the motor task (< 1%, see figure 7(a)), statistical regression for them could not be performed. Similar patterns were observed in the power contribution of LFPs (figure 7(b) and 8(b)) that are not discussed for conciseness. Taken together, the regression analysis confirmed that the proportion of contributing features in high frequency bands sharply decays over distance while those of lower frequency bands remain rather constant, indicating the long-range impact of low frequency features on neurons.

Finally, while the proportion of high \(\gamma\) contributing LFPs showed a significant decay over distance in the motor task, these high \(\gamma\) features still constituted the most dominant contribution to spikes when considering the whole network and all frequency bands (pairwise comparison of the proportion of contributing features at different bands: \(P < 0.039\), paired t-test with Bonferroni correction for multiple comparison, \(N = 17\); figure 7(a)). In comparison, in the pro-saccade task, neurons are comparably influenced by high \(\gamma\) and low frequency bands (\(P = 0.3\) for \(\theta\), \(P = 0.1\) for \(\alpha\), paired t-test with Bonferroni correction for multiple comparison, \(N = 8\)), and influenced significantly less by the middle frequency bands (\(P < 0.043\) for \(\beta\) and low \(\gamma\) bands, paired t-test with Bonferroni correction for multiple comparison, \(N = 8\)). These results again confirm the importance of spike modulation by high frequency features in the motor task and the importance of both low and high frequency features in the pro-saccade task (see also figure 6).

3.5.3. Spectral distribution of dependencies in the motor and pro-saccade tasks explains their different spatial dependency patterns. Why does the LFP contribution to spikes drop much faster in the motor task compared with the pro-saccade task (figure 5(b) versus (d))? We found that differences in the spectral distribution of the contributing LFP features in the two tasks may explain this observation. In particular, one reason for this could be the more dominant role of high-frequency features on spikes in the motor task compared to the pro-saccade task.

Figure 9. Pattern of conditional dependencies between spikes and LFP features within the four brain regions in the motor task, estimated by the sparse model-based estimation algorithm. Results are averaged across neurons in each region and across all the cross-validation folds. Each subfigure counts the average number of LFP dependencies coming from the four different regions per neuron in a single region: (a) M1 neurons, (b) PMd neurons, (c) PMv neurons, (d) PFC neurons. Vertical axis indicates where the neuron is located at. Horizontal axis indicates where the retained LFP features are located at. For example (a) shows the average number of dependencies of LFPs in M1, PMd, PMv, and PFC to a neuron in M1 and similarly for other subfigures.
task and the observation that high-frequency features in general have a short-range impact on spikes. First, the detected dependencies in the motor task were mostly from the high frequency regime, i.e. high $\gamma$, while those in the pro-saccade task were comparably from both low and high frequency regimes, i.e. $\theta$ and high $\gamma$ (figure 6). Second, in both the motor and the pro-saccade tasks, we showed that the proportion and contribution of high $\gamma$ dependencies sharply attenuate over distance and mostly influence the neurons placed within a short distance around them (figures 7 and 8). This observation is also consistent with prior work, such as [70, 71]. In contrast, the proportion and contribution of dependencies at lower frequency bands, in particular $\theta$ band in the pro-saccade task, stays relatively constant over distance, i.e. low frequency LFP features almost similarly influence the neurons over the network regardless of how far they are (see section 3.5.2 and figures 5 and 7). In other words, high $\gamma$ features have a short-range impact on neuronal spikes while low frequency features have a long-range impact on the network of neurons. Thus together, the first and second set of observations suggest that the contribution of dependencies in the motor task sharply attenuates with the spike-LFP distance in the network because most of these dependencies are at the short-range high $\gamma$ band. Also, the contribution of dependencies for the pro-saccade task does not decay sharply over distance because of the dominant role of long-range low frequency dependencies on neuronal spikes in the PFC (see also discussion section 4.3).

### 3.6. Sparse model-based estimation algorithm uncovers the pattern of brain regional functional dependencies

Since in the motor task we had recordings from different brain regions (PMd, PMv, M1, PFC), we further dissociated and assessed the inter- and intra-regional dependency patterns between neurons and LFP features in the motor task. We found that a given neuron in each brain region had the largest number of dependencies with LFP features in that same region (figure 9), as expected from the spatial distribution of dependencies (figures 5 and 7). After intra-regional dependencies that were most dominant, in terms of the secondary effects, neurons in M1 are influenced by PMd, while LFP features in PFC largely contribute to the neurons in PMd and PMv (figure 9). These pattern of dependencies found by the sparse model-based estimation algorithm between different brain regions is consistent with prior evidence about information flow in the motor cortical regions [72, 73] (see discussions section 4.7), which further validates our modeling approach.

### 3.7. The importance of network modeling I: spike prediction on a given electrode is improved using the network of LFP features on all electrodes compared with using only the LFP features on the same electrode

We further found the importance of dependencies across the whole network in modulating spikes by comparing to when the sparse model-based estimation algorithm used only zero-distance LFP features (i.e. LFP features on the same electrode that records the spikes) (figure 10). We found that compared with the LFP features on the same electrode, features from the network significantly improve spike prediction, i.e. PP scores, in both tasks. In the motor task, compared with the model with no dependencies, including the zero-distance LFP features improved the PP score by 20.6%. In comparison, when including all network LFP features from all electrodes, this improvement was significantly higher at 31% (table 2; $P = 7 \times 10^{-12}$, paired t-test, $N = 14$). In the pro-saccade task, compared with the model with no dependencies, including the zero-distance LFP features improved the PP score by 15.9%. Again, in comparison, when including all network LFP features from all electrodes, this improvement was significantly higher at 24% (table 2; $P = 9 \times 10^{-4}$, paired t-test, $N = 10$).

![Figure 10](image-url)
3.8. The importance of network modeling II: spike prediction on a given electrode is improved using the network of LFP features on all electrodes compared with using only the nearest LFP features

It was shown that the sparse model-based estimation algorithm can improve the spike PP using on average seven and ten retained dependency parameters per spike channel in the motor and pro-saccade visual tasks, respectively (see table 2). We performed a control analysis to study to what extent spike prediction performance would be different if we instead used a comparable number of LFP features that were physically closest to a given spike channel. We performed the same analysis with 10, 15, and 20 nearest LFP features in a spike channel’s proximity. The spike prediction performance is obtained from (5) using all the closest 10, 15, and 20 LFP features in this analysis.

We found that modeling the dependencies across the entire network using the sparse model-based algorithm improved %PP compared to just using a similar number of closest LFP features. Averaged among all the sessions, the full network LFP in both motor and pro-saccade visual tasks leads to significantly larger spike PP improvement ($P < 0.05$, paired t-test, $N$ for motor and pro-saccade tasks are seven and five, respectively). Moreover, as expected, this level of improvement was larger in the pro-saccade than the motor task (see figure 11; motor task $0.02 < P < 0.04$, paired t-test, $N = 7$, pro-saccade task $P < 0.004$, paired t-test, $N = 5$). This is because our algorithm had already identified that most of the contributing LFP features are the nearby features in the motor task but in contrast the contributing LFP features come from distributed distances in the pro-saccade task (figures 5, 7 and 8). Consistently, we see from figure 11(a) that in five out of seven motor kinematic sessions the spike %PP improvement using the full network LFP is similar to using the nearest LFP features. However, in the pro-saccade task, the full network LFP shows a significant improvement in %PP compared with using the nearest LFP features in four out of five sessions (figure 11(b)). The above results again suggest the importance of using the sparse model-based algorithm to capture the effect of the full network on the spiking activity of a given channel to obtain more accurate encoding models. These results also confirm the findings of the sparse model-based estimation algorithm that the LFPs in a close proximity of spike channels are the large contributors to spike prediction in the motor kinematic task but in contrast in the pro-saccade task, LFPs distributed over the entire network contribute to spike prediction (figure 11).

4. Discussion

Toward the goal of estimating accurate encoding models from data that take into account both behavior and the conditional dependencies between spikes and fields, we developed a sparse model-based estimation algorithm. We modeled the neuronal spiking activity as binary point processes whose instantaneous rate function was modulated not only by behavior but also by the field features. The conditional dependencies were thus parameters in this model to be estimated. To resolve the challenge of estimating a prohibitively large number of spike-field dependencies from data, we formulated a maximum-likelihood estimation problem with a sparsity-promoting L1 penalty on the dependency parameters. We then used AIC to further enforce model sparsity by detecting the field features with important dependency terms.

We validated the sparse model-based estimation algorithm using simulated data as well as NHP data in two distinct tasks: motor task with motor cortical recordings and pro-saccade task with prefrontal recordings. To show the biological consistency of our results, we further used the algorithm as a
tool to comprehensively study the spectro-spatial properties of functional dependencies between a given neuron and the entire set of field features on all electrodes on the array simultaneously and during behavior. We found that including the conditional dependencies improved the spike prediction performance. Further, the encoding model estimated by the algorithm used significantly fewer dependency parameters compared with the model estimated with MLE (i.e. without the penalty term), while at the same time improving spike prediction performance compared with MLE. Also, the algorithm uncovered spectro-spatial patterns for dependencies over a large spatial network (i.e. over the entire array of electrodes), over different field frequency regimes, and between different brain regions. In contrast, these patterns were masked using the MLE. Finally, spike prediction on a given electrode was improved by modeling the contribution of LFP features on all electrodes compared to when modeling only features on the same electrode. Taken together, the sparse model-based estimation algorithm can serve as a new tool to identify and study conditional dependencies between a given neuron and to build more accurate multiscale encoding models.

4.1. The algorithm improved spike prediction while using significantly fewer dependencies

The sparse model-based estimation algorithm provides a sparse solution to the problem of estimating functional dependence in a high-dimensional spike-LFP network. It thus avoids the overfitting problem by detecting and then only including the LFP features with significant contribution to spikes in the encoding model. Our simulations demonstrated that the sparse model-based estimation of dependencies leads to significantly fewer FP in detecting which field features have conditional dependencies with the spikes (table 1).

Our experimental results across both the motor and the pro-saccade tasks reveal that sparse estimation of conditional dependencies significantly improves the spike prediction performance. These results suggest that even conditioned on behavior, there are remaining dependencies between spikes and LFPs. Also, estimating the conditional dependencies with the sparse model-based algorithm resulted in much fewer dependencies to be incorporated in the encoding model compared with estimating them with MLE (i.e. without the penalty term). Importantly, despite the significantly fewer dependencies included, the spike prediction performance not only was not degraded, but also was significantly improved (tables 1 and 2). This result suggests that without the penalty term, MLE identifies many spurious dependencies (i.e. FP) that introduce noise and inaccuracy in spike prediction.

4.2. Comparison with prior measures of spike-LFP relationship

In addition to providing a more accurate multiscale encoding model, the proposed algorithm can also serve as a tool to study conditional dependencies between spikes and fields while taking into account common behavioral modulations and considering the spike modulation due to all field features on an array of electrodes simultaneously. To study the relationship between spike and LFP features, the majority of prior work have used tools that are formed on the basis of spike-triggered average of the LFP (stLFP) [10, 39, 67] and spike-LFP coherence concepts [5, 33, 35, 36]. Specifically, stLFP is a tool in the time-domain to assess the strength of activity reflected in the LFP features at a channel correlated with spiking activity at another channel [10], while spike-LFP coherence is a frequency dependent measure of linear association between spike timing and LFP features. While these techniques are quite important in studying pairwise correlations, there are various additional functionalities that the proposed sparse model-based estimation algorithm offers.

First, prior correlation-based methods do not aim to directly dissociate the correlations due to spikes and LFP being modulated by a common behavior (unconditional dependencies) from correlations due to dependencies between these signals, i.e. those relationships that persist when common behavioral modulation is removed (conditional dependencies). In particular, at least part of the correlations detected using these techniques could be due to both spikes and LFPs encoding a common behavior, and thus co-varying according to variations in the common behavioral states. While it could be possible to address this issue for epoched and fixed behaviors—when behavior is fixed throughout the window for coherence analysis—, it is not easy to address it for rich non-stationary behaviors, for example when studying the motor cortex during continuously evolving movements. Second, these correlation methods do not aim to provide an encoding model for simultaneous representation of behavior and dependency terms. Finally, these correlation methods are pairwise rather than network-based. While it is possible to compute the coherence between every single spike-LFP pair to compute a coherence matrix for the network, doing so is prone to multiple comparison challenges given the large number of separately computed pairs. Also, to get a more accurate measure of direct dependency between spikes and a given LFP signal, it is important to simultaneously take into account the effect of other LFP signals on the spikes, so that the effect of each LFP signal can be dissociated from the rest. More recently, model-based techniques have been developed that can study the dependency terms under a behavioral task [37, 38]. These tools, however, focused on estimating the dependency terms on a single electrode rather than the entire network of electrodes. The sparse model-based estimation algorithm allowed us to include the effect of all the LFP features across the network of electrodes. We showed that including the network field features (i.e. all electrodes) improves spike prediction compared with only including the LFP features on the same electrode (figure 10, section 3.7). Thus, in addition to enabling the study of spectro-spatial patterns of dependencies across the network, our algorithm estimates sparse dependency terms across the network that account for spiking variability not explained by behavior and same-electrode LFP features to significantly improve spike prediction.
4.3. Spatial distribution of contributing LFP features to spikes

The sparse model-based estimation algorithm allows for identification of the functional dependencies within a large spatial network and during behavior. By simultaneously modeling the contribution of behavior, if any, as well as all the LFP features in the network (i.e. on all electrodes) on the spiking activity from a given electrode, we found that the LFP contribution to spikes in PFC and motor cortex drops as the spike-LFP distance increases (see figures 5(b) and (d)). Studies on pairwise coherence analysis in the visual cortex when presenting visual stimuli have shown that spike-LFP coherence drops as a function of distance between the spike and LFP channels [39, 67]. Although using coherence and not building an encoding model of behavior, this prior observation may corroborate our findings and further validate our algorithm in estimating encoding models.

Our results also show that the contribution of LFP features on the spiking activity shows a much larger drop over distance in the motor task versus the pro-saccade task (see figures 5(b) and (d)). One reason for this observation could be the more dominant contribution of high-frequency features to spikes in the motor task compared to the pro-saccade task. Indeed, prior studies have indicated that the effects of high frequency field features are subject to steeper attenuation with distance [70, 71] compared with low-frequency field features, as was also observed in our analyses (figures 7 and 8). Further, our results demonstrated that the motor cortical neurons during the motor task were largely affected by high-frequency LFP features (figures 6(a) and 7; see very small contribution from low frequency features) in contrast to PFC neurons in the pro-saccade task, which had significant contributions from both low and high frequencies (figures 6 and 8). The more dominant role of high-frequency features in the motor task and the observation that high-frequency field influence on spikes decays faster over distance could be the reason for our observation that the LFP feature contribution to spikes in the motor task has a much steeper drop over distance compared with the pro-saccade task (figures 5(b) and (d); see section 4.4 below for more details)

Finally, in the pro-saccade task, the proportion of contributing LFP features (and the normalized power contribution of LFPs to some extent) did not attenuate over distance (figures 5(c) and (d)). This may be due to the large contribution of low frequency features, which have a long-range impact on PFC spikes in the pro-saccade task (figures 6(b) and 8).

These observations were also confirmed by analyzing the nearest LFP features (see section 3.8). We found that the nearest LFP features to a spike channel can lead to a spike PP that was similar to the one obtained by using the full network LFP in the motor task. But in contrast, in the pro-saccade task, the spiking PP obtained was significantly smaller when using the nearest LFP features compared to when modeling the entire network LFP using the sparse model-based method.

4.4. Spectral contribution of LFP features to spikes and their spatial patterns

The sparse model-based estimation algorithm allowed us to uncover the spectral patterns of LFP feature contribution to spikes. The high frequency LFP features showed significant contribution to the neuronal spikes. In particular, high γ features showed the most influence on neurons in the motor task, and together with the θ band had the largest influence on neurons in the pro-saccade task (figures 6–8). The large high-frequency contribution was mostly concentrated within a short distance from the neuron and dropped as the spike-LFP distance became nonzero (figures 7 and 8). This result is consistent with results from prior work that studied pairwise spike-LFP coherence and showed the high correlation between spikes and high frequency LFP features immediately adjacent to the recorded neuron [39, 67]. Based on the prior studies [65, 74, 75], LFP signals represent the action potentials and synaptic activity of a few thousand neurons depending on the diameter of the electrode. Lower frequency LFP features are compound signals of slower events, such as synaptic events, from a very large population of cells, while higher frequency LFP features reflect the action potentials in smaller populations [74, 75]. These results may suggest that high frequency LFP features are dominated by current from action potentials, and consequently that a high correlation between the spiking activity and high frequency LFP features in their close proximity would be expected [10, 65, 75].

In comparison to the high frequency features, the contribution of the low frequency features showed a broader long-range impact on the neuronal spikes (see figures 7 and 8). Indeed, compared to the action potentials, synaptic events attenuate less with distance and propagate over large distances in extracellular space [70, 71, 76], which is consistent with the pattern uncovered by the sparse model-based estimation algorithm. Taken together, consistent with prior studies mentioned above, our analysis revealed that for close neighboring electrodes to a neuron, the correlation between spikes and LFP signal is much more dominant for high frequency LFP features compared with low frequency LFP features. Furthermore, our statistical regression test showed that the contribution of low frequency field features to spikes attenuates less over distance (see section 3.5.2).

In the pro-saccade task, we found that after high γ, θ features show the most number of dependencies with neurons (see figure 6), which is again inline with earlier studies using pairwise coherence analysis [68, 69] and thus further validates our modeling framework. Womelsdorf et al studied the spike-LFP phase consistency for different LFP frequency bands and showed the important role of the θ band activity in modulating the timing of spikes in the pro-saccade visual task [68, 69]. They showed that spike-LFP phase consistency in PFC tends to be stronger when LFP θ activity is stronger, leading to a larger number of neuronal spikes during a pro-saccade. Interestingly, this is in agreement with our results (see figures 6 and 8), where the dependencies identified at θ band are positive, suggesting the positive correlation of θ band LFP power with spike firing rate. In addition, our results showed that, together with the low γ features, the intermediate frequency band β contributes the least to the spiking activity (figure 6). Prior work have suggested that β LFP features only occur as transient events during decision making and working memory tasks [77], which may be the reason for weak
relationship of $\beta$ LFP activity with spikes in our study. In particular, Spitzer and Haegens [77] in a review of prior work discussed the short-lived nature of beta modulations recorded from cortical areas, such as PFC, during decision making and working memory tasks.

We showed that high $\gamma$ features in the motor cortex and PFC (during the motor and pro-saccade tasks, respectively) had positive dependencies with spikes, suggesting their positive correlation over distance (figures 7 and 8). In contrast, the lower frequency bands across the network showed both positive and negative conditional dependencies to spikes. Manning et al [37] studied the relation between spikes firing rate and LFP features recorded from the same electrode at widespread brain regions, including the frontal cortex, occipital and parietal cortices, amygdala, hippocampus, and parahippocampal region, in subjects under a virtual navigation game. Even though they only explored the relationships on the same electrode and did not focus on the spatial spike-LFP relationship or the behavioral modulation, they observed large positive spike-LFP relationships in the $\gamma$ band, while finding less pronounced correlation with spikes in both positive and negative directions in the remaining frequency bands. Thus, despite modeling the dependencies between a neuron and all LFP features on all electrodes simultaneously, our same-electrode results (corresponding to the zero-distance points in figures 7 and 8) are consistent with this prior work, i.e., our algorithm finds a similar positive pattern of dependencies on the same electrode as this prior work, thus further validating the algorithm.

4.5. LFP contributions to spikes on the same electrode

Similar to the traditional approach for separating spikes from field signals [64, 65], in this work we used the neural signal components in frequencies above a predetermined threshold to extract the spikes and obtained the LFP features from the lower frequency components. Although this approach is widely used [64, 65], the extracted LFP features may include contributions from spikes on the same electrode, sometimes referred to as spike leakage [64–66, 76]. Studying the source of LFP contribution to spikes on the same electrode and dissociating the contribution due to leakage from that due to other synaptic sources is an important topic, which was not the focus of our approach. Instead, our goal is to develop an algorithm that can identify the conditional dependencies regardless of their source. Using the algorithm, we showed that spike prediction on a given electrode improves when modeling the contribution of LFP features on all electrodes compared to when modeling only LFP features on the same electrode as the spikes (figure 10, section 3.7). This result suggests that, regardless of whether leakage is a dominant source of LFP contribution to spikes on the same electrode or not, modeling the contribution of LFP features on other electrodes (i.e., network LFP features) improves spike prediction. This result also suggests that there are significant spike-field dependencies across the network beyond those on the same electrode.

While not our focus, determining the source of LFP contribution to spikes on the same electrode (and dissociating the contribution due to leakage) is an important but distinct problem to study [64, 66]. Our method can provide a new tool to study this problem in future studies. For example, this can be done by using our method on spike and LFP features computed while trying to minimize leakage on the same electrode [76] and by then comparing the results to when the method is applied on traditionally-computed spike and LFP features. Here the distance between our electrodes was at least 1.5 mm. Future electrode array technologies with dense electrodes (tens of microns apart) may allow for studying the LFP contribution to spikes at much shorter distances while avoiding the leakage issue. This can be done by modeling the spike-field dependency on adjacent electrodes with much shorter distances (much shorter than 1.5 mm), which are still far enough for the spike leakage between them to be negligible.

As discussed above, the spike leakage may impact the zero-distance spike-LFP dependencies (see also section 2.8). It is worth noting though that several observations in this study are not fully consistent with spike leakage being a dominant source of spike-field dependencies on the same electrode, as also has been suggested in prior work [66]. First, we observed differences between the motor and pro-saccade tasks in terms of the spatial distribution of contributing LFP features to spikes and in terms of the contribution of same-electrode LFP features to spikes. While there is an attenuation in the proportion of contributing LFP features over distance in the motor task, in the pro-saccade task this proportion is comparable at different distances (figure 5(a) versus (c)). Second, it has been suggested that a wide range of LFP frequency regimes on the same electrode can be affected by leakage (e.g., leakage may affect LFPs down to around 10 Hz [64]). Despite this, we found that in the pro-saccade task the contribution of high-frequency high $\gamma$ and low-frequency $\theta$ bands are comparable and larger than other middle bands (figure 6(b)). Also most of the contributing LFP features at zero-distance were from the high $\gamma$ band (figures 7(a) and 8(a)).

Interpreting what these observations imply regarding the source of LFP contribution to spikes on the same electrode is an important area of future investigation.

Finally, differences between the motor and pro-saccade data pre-processing, e.g., spike sorting methods (see section 2.4.2), may to some extent account for differences observed between motor and pro-saccade tasks especially in term of functional dependencies between spikes and LFPs on the same electrode (i.e., zero-distance). While a direct comparison between motor and prosaccade tasks is not the main goal of our study, future studies can apply the presented sparse model-based algorithm on datasets with consistent sorting to further investigate this question.

4.6. Sparse model-based estimation algorithm was essential in uncovering spectro-spatial network dependency patterns

The sparse model-based estimation algorithm was necessary in uncovering the pattern of spike-field conditional dependencies over distance and frequency bands. In particular, we also identified the spike-field conditional dependencies for both experimental tasks using MLE (i.e., without the sparsity penalty). We found that the MLE could not find any specific pattern for
the conditional dependencies. For example, encoding models estimated by the MLE indicated that the field features have the same contribution to the spiking activities regardless of their distance from neurons, which is inconsistent with prior evidence in the literature [39, 67]. Consistently, compared to the proposed algorithm, the number of FP (spike-field dependencies incorrectly identified) was increased by a factor of 2.4 using MLE in the simulations while the number of TP only changed by 13%. Prior work exploring connectivity between spike trains of ensembles of neurons has similarly found MLE to be unreliable in the presence of limited data, thus indicating the importance of regularization [78]. Taken together, the large number of falsely detected spike-field dependencies in the multiscale network using MLE will likely mask any spectro-spatial patterns in analyzing spike-field network dependencies.

4.7 The contribution of LFP features in different brain regions to spikes

The sparse model-based estimation algorithm allowed us to identify the functional dependencies in a large spatial network and thus to examine them between different brain regions in the motor task (in which we had recorded from different regions). First, we found that the spiking activity in each brain region, i.e. M1, PMd and PMv, was most influenced by the LFP features in the same region. This is consistent with our results on spatial contribution of fields to neurons, which indicated that neurons are most influenced by the LFP features within a short distance (figure 5(a)). Second, after intra-regional dependencies, the most significant dependencies for neurons in M1 came from LFP features in PMd and PMv. In comparison, LFP features in PFC largely contributed to neurons in PMd and PMv. These results are indeed in agreement with prior evidence on the information flow in the motor cortical areas [72, 73, 79]. For example, it has been suggested that PFC plans movements, premotor cortex (PMd and PMv) organizes sequences of movements, and eventually M1 executes the intended movements [72, 73, 79, 80]. This prior evidence may suggest that information flow is from prefrontal to premotor cortex and then to motor cortex, which is consistent with what our algorithm finds while modeling the conditional dependencies between the entire network of LFP features and the spiking activity of a given neuron. These results further show the biological consistency of the algorithm in detecting conditional dependencies and incorporating them in multiscale encoding models.

Acknowledgment

The authors acknowledge support of the Army Research Office (ARO) under contract W911NF-16-1-0368. This is part of the collaboration between US DOD, UK MOD and UK Engineering and Physical Research Council (EPSRC) under the Multidisciplinary University Research Initiative (MURI). The authors also acknowledge support of the Office of Naval Research (ONR) Young Investigator Program (YIP) under N00014-19-1-2128.

References

[17] Donoghue J P 2008 Bridging the brain to the world: a perspective on neural interface systems Neuron 60 511–21

ORCID iDs

Yan T Wong https://orcid.org/0000-0002-5281-8221
Maryam M Shanechi https://orcid.org/0000-0002-0544-7720


[27] Bolus M, Willatts A, Whitmire C, Rozell C and Stanley G 2017 Broadband shifts in LFP power spectra are correlated with single-neuron spiking variability in motor cortex during movement execution Frontiers Syst. Neurosci. 9 89


[34] Wang W et al 2013 An electrocorticographic brain interface in an individual with tetraplegia PLoS One 8 e55344


[43] Shanechi M M, Wornell G W, Williams Z and Brown E N 2010 A parallel point-process filter for estimation of goal-directed movements from neural signals 2010 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (Dallas, TX, USA) pp 521–4


