MIT Open Access Articles

A Novel Nonparametric Approach for Neural Encoding and Decoding Models of Multimodal Receptive Fields

A Novel Nonparametric Approach for Neural Encoding and Decoding Models of Multimodal Receptive Fields

Rahul Agarwal
rahul.jhu@gmail.com
Department of Biomedical Engineering, Johns Hopkins University, Baltimore,
MD 21218, U.S.A.

Zhe Chen
zhe.chen3@nyumc.org
Department of Psychiatry and Department of Neuroscience and Physiology,
New York University School of Medicine, New York, NY 10016, U.S.A.

Fabian Kloosterman
fabian.kloosterman@nerf.be
Neuro-Electronics Research Flanders, Leuven, Belgium; Imec, Leuven, Belgium;
and Brain and Cognition Research Unit, KU Leuven 3000, Belgium

Matthew A. Wilson
mwilson@mit.edu
Picower Institute for Learning and Memory, MIT, Cambridge, MA 02139, U.S.A.

Sridevi V. Sarma
ssarma2@jhu.edu
Department of Biomedical Engineering, Johns Hopkins University, Baltimore,
MD 21218, U.S.A.

Pyramidal neurons recorded from the rat hippocampus and entorhinal cortex, such as place and grid cells, have diverse receptive fields, which are either unimodal or multimodal. Spiking activity from these cells encodes information about the spatial position of a freely foraging rat. At fine timescales, a neuron's spike activity also depends significantly on its own spike history. However, due to limitations of current parametric modeling approaches, it remains a challenge to estimate complex, multimodal neuronal receptive fields while incorporating spike history dependence. Furthermore, efforts to decode the rat's trajectory in one- or two-dimensional space from hippocampal ensemble spiking activity have mainly focused on spike history–independent neuronal encoding models. In this letter, we address these two important issues by extending a recently introduced nonparametric neural encoding framework that allows modeling both complex spatial receptive fields and spike history dependencies.

Using this extended nonparametric approach, we develop novel algorithms for decoding a rat's trajectory based on recordings of hippocampal place cells and entorhinal grid cells. Results show that both encoding and decoding models derived from our new method performed significantly better than state-of-the-art encoding and decoding models on 6 minutes of test data. In addition, our model's performance remains invariant to the apparent modality of the neuron's receptive field.

1 Introduction

A fundamental goal in neuroscience is to establish a functional relationship between a sensory stimulus or induced behavior and the associated neuronal response. For example, in a rodent experiment (O'Keefe, 1979; Hasselmo, 2008), microelectrodes are implanted into a rat's hippocampus, and the entorhinal cortex and ensembles of neurons' activities are recorded while the animal freely forages in an open arena (see Figure 1A). The goals of such an experiment are to understand how a single neuron or population of neurons encodes spatial information of the rat's position and to determine if we can decode the animal's position based on their spiking activity (Brown, Frank, Tang, Quirk, & Wilson, 1998; Kloosterman, Layton, Chen, & Wilson, 2014).

In the experiment described above, pyramidal neurons recorded from the rat entorhinal cortex and hippocampus have shown complex grid-like or place-like spiking patterns as a function of the rat's position in two-dimensional (2D) environments (O'Keefe, 1979; Moser, Kropff, & Moser, 2008; Hasselmo, 2008) (see Figure 1). However, it is well known that these neurons fire in a non-Poisson fashion and their spiking also depends on their own spiking history. For instance, bursting patterns have been found to be a significant component of hippocampal neuronal spiking activity (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Fenton & Muller, 1998). Thus, encoding and decoding models that capture spike history dependencies will likely yield better performance (Truccolo, Eden, Fellows, Donoghue, & Brown, 2005; Pillow et al., 2008). However, this well-known intrinsic phenomenon has not been fully explored in the rodent hippocampal decoding literature (Brown et al., 1998; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998; Barbieri et al., 2004; Santaniello, Montgomery, Gale, and Sarma, 2012; Kloosterman et al., 2014). The fact that the receptive field has a complex and multimodal shape (especially in 2D space) complicates such modeling. Parametric models may be insufficient to fully characterize a complex, multimodal spatial receptive field structure and history dependence simultaneously.

Nonparametric modeling assumes no explicit probability distribution (i.e., distribution free) and lets data speak for themselves; therefore, it is more flexible in statistical data analysis. In the literature, several nonparametric methods have been developed for modeling neuronal spike activity
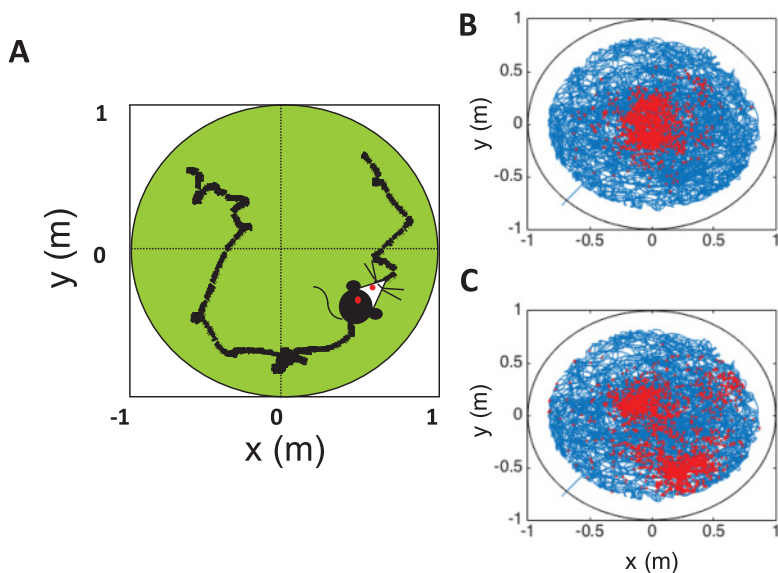
Figure 1: Spiking activity of unimodal and multimodal place cell. (A) A schematic of the circular arena in which the rat is freely foraging. Spiking activity of a unimodal place cell (B), and a multimodal place cell (C). The rat's run trajectory (for 36.36 min) is marked by blue lines. Red dots mark the rat's spatial position when the cell spiked.

(Chen, Kloosterman, Layton, & Wilson, 2012; Kloosterman et al., 2014; Rad & Paninski, 2010; Coleman & Sarma, 2010). Specifically, Chen et al. (2012) and Kloosterman et al. (2014) proposed an encoding framework based on kernel density estimation methods for capturing complex dependencies between spiking features and the rat's position in a 1D environment. However, their model did not incorporate the spike history dependence. In another work, Theis, Chagas, Arnstein, Schwarz, and Bethge (2013) used a semiparametric mixtures-of-gaussian model to estimate ganglion cell responses that incorporate spike history effects and showed improvement over standard parametric models. Nonparametric methods, such as histograms and radial basis functions, have also been used for dimensionality reduction in fitting neuronal response functions (Sharpee, Rust, & Bialek, 2004; Williamson, Sahani, & Pillow, 2015) .

   Inspired by these works, we applied nonparametric algorithms to develop encoding models that use both spatial and spike history information for neurons recorded from a rat's hippocampus and entorhinal cortex. Further, we proposed a novel nonparametric decoding algorithm and applied it to reconstruct the rat's trajectory using spiking activity from these cells. Our nonparametric framework is based on a novel and efficient nonparametric

maximum likelihood (ML) density estimator over a set of densities whose square root is band-limited (BL) (Agarwal, Chen, & Sarma, 2015). The BLML procedure uses convex optimization with quadratic time complexity for encoding (for details, see appendix A) and linear time complexity (in terms of number of neurons per time step) for decoding.

Finally, we use 6 minutes of unseen data from these 53 cells to evaluate the performance of our nonparametric models and compared them with state-of-the-art parametric models with and without spike history covariates. Results show that the performance of our nonparametric encoding and decoding models with spike history is significantly better than the performance of state-of-the-art parametric models. The result remained robust across the population of neurons with both unimodal and multimodal receptive fields.

2 Methods

2.1 Notation. Let t denote a continuous-time index, and let $c = 1, \ldots, m$ denote the neuronal index among the population. We use the lowercase p to denote the probability density function (pdf) and the uppercase P to denote the probability. Further, the terms *unimodal* and *multimodal cells* are used throughout, but the cells are not definitively defined as such.

2.2 Data Collection. In the experimental protocol, a Long-Evans rat was freely foraging in an open field arena ($\sim$1 m radius) for a period of 36.36 minutes. Custom microelectrode drives with variable numbers of tetrodes were implanted in the rat's medial entorhinal cortex and dorsal hippocampal CA1 area. Spikes were acquired with a sampling rate of 31.25 kHz and filter settings of 300 Hz to 6 kHz. Two infrared diodes alternating at 60 Hz were attached to the microelectrode array drive of the animal for position tracking. Spike sorting was accomplished using a custom manual clustering program (Xclust, M. A. Wilson). All procedures were approved by the MIT Institutional Animal Care and Use Committee.

In this selected recording session, $m = 74$ neurons were recorded from both entorhinal cortex ($m = 8$) and dorsal hippocampus ($m = 66$). Out of these 74 neurons, 21 (6 from entorhinal cortex) neurons were discarded, as they fired either independent of position (x, y) (i.e., putative interneurons) or had very low firing rates (<0.175 Hz). Out of the remaining $m = 53$ neurons, $m = 27$ (1 from entorhinal cortex, 26 from hippocampus) were unimodal and $m = 26$ multimodal (1 from entorhinal cortex, 25 from hippocampus). The x and y coordinates of the rat's position when these cells spiked are shown as scatter plots in Figure 2.

2.3 Neuronal Encoding and Decoding Models. The encoding model describes the propensity of a single cell to spike over time as a function of external and internal covariates. As a neuronal spike train is a stochastic
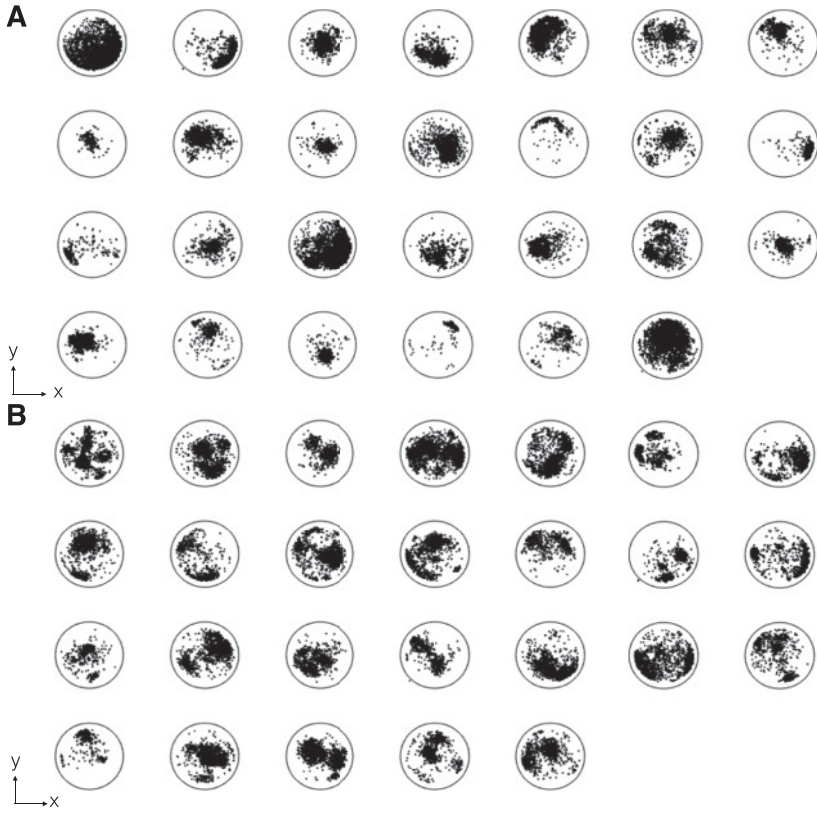
A

y
└──→ x

B

y
└──→ x

Figure 2: Spike scatter plot for spiking activity. (A) Unimodal hippocampal place cells ($m = 27$) and (B) multimodal hippocampal place cells and entorhinal grid cells ($m = 26$). Each plot corresponds to the spiking activity of a single neuron displayed in (x, y) coordinates in the circular arena. Each dot corresponds to (x, y) coordinates of rat's position when the cell spiked.

point process (i.e., random binary events on a continuum), the goal of encoding is to characterize the conditional intensity function (CIF) (Cox & Isham, 2000; Agarwal & Sarma, 2011, 2012a, 2012b). In this letter, we assume that the spiking activity of each place and grid cell depends only on the rat's position and the cell's own spike history. Therefore, we write the CIF for neuron c as (Brown, Barbieri, Ventura, Kass, & Frank, 2002; Truccolo et al., 2005)

$$\lambda_c(t|x_t, y_t, \mathcal{H}_c(t))$$

$$\triangleq \lim_{\Delta t \to 0} \frac{P(\text{spike within time interval } (t, t + \Delta t]|x_t, y_t, \mathcal{H}_c(t))}{\Delta t}, \quad (2.1)$$

where (x_t, y_t) denotes the rat's position at time t in the two-dimensional (2D) environment, and the vector $\mathcal{H}_c(t)$ consists of history covariates for cell c up to time t as in Kass and Ventura (2001), Sarma et al. (2012), Sarma et al. (2010), Santaniello et al. (2012), Kahn, Sheiber, Thakor, and Sarma (2011), and Agarwal, Sarma, Thakor, Schieber, and Massaquoi (2015).

The decoding model is the inverse of the encoding model and is used to predict the external covariates, at any given time t, from the spiking activity of cells up to the time t. In this letter, we realize the decoding model by using the maximum a posteriori (MAP) estimator that maximizes the posterior probability $P(x, y | \tau_1(t), \ldots, \tau_m(t)) \propto P(x, y) \mathcal{L}(\tau_1(t), \ldots, \tau_m(t) | x, y)$, where $\tau_c(t)$ denotes the time since the last spike for cell c at time t and $\mathcal{L}(\tau_1(t), \ldots, \tau_m(t) | x, y)$ denotes the likelihood.

As the likelihood is a high-dimensional object, we simplify as follows. At any given time t, we assume that the likelihood of observing the time since the last spike, $\tau_c(t)$, from the cth neuron ($c = 1, \ldots, m$) given the rat's spatial position (x, y) at time t, has a factorial form

$$\mathcal{L}(\tau_1(t), \ldots, \tau_m(t) | x, y) = \prod_{c=1}^{m} P(\tau_c(t) | x, y), \tag{2.2}$$

where we assume that the neurons fire conditionally independently given the position such that $P(\tau_c(t), \tau_l(t) | x, y) = P(\tau_c(t) | x, y) P(\tau_l(t) | x, y)$ for any $c \neq l$. This allows us to write the MAP estimator as

$$(\hat{x}_t, \hat{y}_t) = \arg\max_{x, y} \ P(x, y) \prod_{c=1}^{m} P(\tau_c(t) | x, y). \tag{2.3}$$

For simplicity, we do not consider the temporal prior $P(x_t, y_t | x_{t-1}, y_{t-1})$ in the course of computing the posterior. In sections 2.3.1 and 2.3.2, we use both parametric and nonparametric methods to estimate the encoding and decoding models for neuronal spike trains.

2.3.1 Parametric Models. Encoding. In the context of a generalized linear model (GLM) framework, we assume that the CIF has a log-linear form. Namely, the logarithm of the CIF is linear with respect to its parameters (Truccolo et al., 2005). Inspired by the past successes of using gaussian and Zernike polynomials in modeling 2D hippocampal place cells (Barbieri et al., 2004), these two representations are chosen to estimate the CIFs. In addition, spike history dependencies are empirically modeled with spike count history covariates h_j; $j = 1, \ldots, J$ that correspond to the number of spikes within the past interval $(t - 2j, t - 2j + 2]$ ms. Specifically, the CIF for each neuron c is modeled in the following two ways:

- Quadratic GLM (GLMquad)

$$\log \lambda_c(t|x_t, y_t, \mathbf{h}_c(t)) = \alpha_{1c} + \alpha_{2c}x_t + \alpha_{3c}y_t + \alpha_{4c}x_t^2 + \alpha_{5c}y_t^2$$

$$+ \alpha_{6c}x_t y_t + \sum_{j=1}^{J} \beta_{jc}h_{jc}(t) \tag{2.4}$$

- Zernike GLM (GLMzern)

$$\log \lambda_c(t|x_t, y_t, \mathbf{h}_c(t)) = \sum_{u,v} \gamma_{uvc}\, \chi(u,v;x_t,y_t) + \sum_{j=1}^{J} \beta_{jc}h_{jc}(t), \tag{2.5}$$

where $\{\alpha_{1c}, \ldots, \alpha_{6c}\}, \{\beta_{jc}, \gamma_{uvc}\}$ are the unknown GLM parameters for GLMquad and GLMzern, respectively; $\chi(u,v;x_t,y_t)$ denotes the fourth-order Zernike polynomials as used in Barbieri et al. (2004), and $\mathbf{h}_c(t) \triangleq [h_{1c}(t), \ldots, h_{Jc}(t)]^\top$ denotes the spike history for cell c at time t. Note that since the log of the CIF is modeled as a quadratic function in GLMquad, the CIF has a gaussian shape.

Decoding. In light of equations 2.2 and 2.3, $P(\tau_c(t)|x,y)$ is evaluated as (Barbieri et al., 2001)

$$P(\tau_c(t)|x,y) = \lambda_c(t|x,y,\mathbf{h}_c(t))e^{-\int_{t-\tau_c(t)}^{t} \lambda_c(s|x,y,\mathbf{h}_c(s))ds}, \tag{2.6}$$

where the integration in equation 2.6 is computed as

$$\int_{t-\tau_c(t)}^{t} \lambda_c(s|x,y,\mathbf{h}_c(s))ds = \int_{t-\tau_c(t)}^{t} e^{\sum_j \alpha_{jc}\zeta_j(x,y)} e^{\sum_j \beta_{jc}h_{jc}(s)} ds \tag{2.7}$$

$$\simeq e^{\sum_j \alpha_{jc}\zeta_j(x,y)} \int_{t-\tau_c(t)}^{t} e^{\sum_j \beta_{jc}h_{jc}(s)} ds \tag{2.8}$$

$$\triangleq e^{\sum_j \alpha_{jc}\zeta_j(x,y)} \lambda_{\tau,c}(\tau_c(t)), \tag{2.9}$$

where $\{\zeta_j\}$ denote the spatial position covariates in either GLMquad or GLMzern. In writing the above probabilities and equations, it is assumed that the rat's spatial position (x,y) remains unchanged at the fine timescales of neuronal spiking. Note that $\lambda_{\tau,c} \triangleq \int_{t-\tau_c(t)}^{t} e^{\sum_j \beta_{jc}h_{jc}(s)} ds$ can be directly estimated from the data as the spike history is known at any time during the decoding process.

A uniform prior $P(x,y)$ is used for MAP estimation in equation 2.3. A similar encoding and decoding formulation can be used for spike history–independent models, except that the hyperparameter J is set to zero. The MAP estimates $(\hat{x}_t, \hat{y}_t)$ derived from both spike history–dependent and

independent models are then smoothed out by filtering bidirectionally by a two-pole Butterworth filter with cutoff frequency 0.35 Hz (for containing 95% power of (x, y) coordinates) to obtain the final smooth estimate $(\tilde{x}_t, \tilde{y}_t)$.

Model estimation and validation. The parameters of history-independent models are estimated by maximum likelihood regression, whereas the parameters for history-dependent models are estimated using ℓ_2-regularized GLM ridge regression (with regularization parameter ρ). The log-linear assumption on the parameters results in a convex likelihood function that allows for efficient computation of the ML estimate (McCullagh, 1984; Truccolo et al., 2005). We use a conjugate gradient algorithm to solve the convex optimization problem.

For simplicity, despite the shape differences in receptive fields, we assume that all neuronal place receptive fields have the same level of history dependence and number of parameters. Therefore, we use the one-vector of hyperparameters $[J, \rho]$ for all cells. These hyperparameters are then chosen by maximizing the R^2 statistic computed as a measure of decoding performance on the validation data set using the hill-climbing algorithm (Yuret & De La Maza, 1993), which guarantees convergence to local maxima. In particular, the first parameter is varied while fixing the second parameter to some initial value. A maximum decoding performance is obtained in the first dimension, and the first parameter is then fixed to the corresponding value. The second is then varied, and its respective value that provides maximum decoding performance is obtained. The process is repeated until a local maximum in all dimensions is found—namely, no improvement in decoding performance is observed by varying any of the two parameters in any direction.

Out of 30.36 minutes of training data, the first 85% is used for model building, and the remaining 15% is used to validate hyperparameters. Once the hyperparameters are validated, the model is retrained in the entire 30.36 min of training data and assessed on the new 6 minute test data.

2.3.2 Nonparametric Models. Encoding. Let $x_t, y_t, h_c(t) \triangleq \log \tau_c(t)$, and $s_c(t)$ be the outcomes of random variables X, Y, H_c, S_c, respectively, where $s_c(t) \triangleq 1$ if and only if cell c fires an action potential in $\lim_{\Delta t \to 0}(t, t + \Delta t]$ and 0 otherwise.

Then the CIF for neuron c can be written in the following equivalent form as the ratio of two densities (Kloosterman et al., 2014):

$$\lambda_c(t|x_t, y_t, h_c(t)) \simeq \frac{N}{T} \frac{p_{X,Y,H_c|S_c}(x_t, y_t, h_c(t))}{p_{X,Y,H_c}(x_t, y_t, h_c(t))}, \tag{2.10}$$

where N denotes the total number of spikes within total duration T. The numerator $p_{X,Y,H|S_c}(x, y, h) \triangleq P(X = x, Y = y, H_c = h|S_c = 1) \geq 0$ denotes a 3D conditional pdf of spatial position and history condition on neuronal

spiking, and the denominator $p_{X,Y,H_c}(x, y, h) \triangleq P(X = x, Y = y, H_c = h) > 0$ denotes a 3D joint pdf of the spatial position and spike history. Note that the time and cell indices are dropped from x, y, and h in these definitions as the densities are defined on space of X, Y, and H_c, which take any particular value x_t, y_t, and $h_c(t)$ for any given cell c at any given time t. The use of the logarithm in the history function allows for a smoother dependence of $\lambda_c(t)$ on τ_c, which in turn captures high-frequency components in the CIF due to refractoriness (i.e., a sharp decrease in $\lambda_c(t)$ right after emitting a spike) and bursting.

To estimate $p_{X,Y,H_c}(x, y, h)$ and $p_{X,Y,H_c|S_c}(x, y, h)$, we apply the probability product rule:

$$\frac{p_{X,Y,H_c|S_c}(x, y, h)}{p_{X,Y,H_c}(x, y, h)} = \frac{p_{X,Y|S_c}(x, y)}{p_{X,Y}(x, y)} \frac{p_{H_c|X,Y,S_c}(h)}{p_{H_c|X,Y}(h)}. \tag{2.11}$$

Let $\lambda_{xy,c}(x, y) \triangleq \frac{p_{X,Y|S_c}(x,y)}{p_{X,Y}(x,y)}$. We rewrite equation 2.10 as

$$\lambda_c(t|x, y, h) = \lambda_{xy,c}(x, y) \frac{N}{T} \frac{p_{H_c|X,Y,S_c}(h)}{p_{H_c|X,Y}(h)}$$

$$\approx \lambda_{xy,c}(x, y) \frac{N}{T} \frac{p_{H_c|\lambda_{xy,c},S_c}(h)}{p_{H_c|\lambda_{xy,c}}(h)}$$

$$= \lambda_{xy,c}(x, y) \frac{N}{T} \frac{p_{H_c\lambda_{xy,c}|S_c}(h, \lambda_{xy,c}(x, y))}{p_{H_c\lambda_{xy,c}}(h, \lambda_{xy,c}(x, y))} \times \frac{p_{\lambda_{xy,c}}(\lambda_{xy,c}(x, y))}{p_{\lambda_{xy,c}|S_c}(\lambda_{xy,c}(x, y))}$$

$$\triangleq \lambda_{xy,c}(x, y)\lambda_{h,c}(h, \lambda_{xy,c}(x, y)). \tag{2.12}$$

In the second step, it is assumed that $\lambda_{xy,c}(x, y)$ conveys the same information about h as knowing (x, y). This assumption is reasonable as spiking history h depends on the spatial coordinates (x, y) through the spiking patterns of the neuron, which are characterized by $\lambda_{xy,c}(x, y)$. This assumption reduces a 3D into a 2D density estimation problem. The validity of this assumption is tested by performing goodness-of-fit tests and decoding analysis on the held-out test data. A direct way to verify the validity of this assumption is to calculate the dependencies between H_c and $\lambda_{xy,c}$ and H_c and (x, y). (For details, see Agarwal, Sacre, & Sarma, 2015.)

Further, to estimate $\lambda_{h,c}(h, \lambda_{xy,c}(x,y)) \triangleq \frac{N}{T} \frac{p_{H_c\lambda_{xy}|S_c}(h\lambda_{xy}(x,y))}{p_{H\lambda_{xy}}(h\lambda_{xy,c}(x,y))} \times \frac{p_{\lambda_{xy,c}}(\lambda_{xy,c}(x,y))}{p_{\lambda_{xy,c}|S_c}(\lambda_{xy,c}(x,y))}$
in equation 2.12, we use the following nonlinear transformations

$$p_{H\lambda_{xy,c}|S_c}(h, \lambda_{xy,c}(x,y)) \longrightarrow p_{Hg(\lambda_{xy,c})|S_c}(h, g(\lambda_{xy,c}(x,y))),$$

$$p_{H\lambda_{xy,c}}(h, \lambda_{xy,c}(x,y)) \longrightarrow p_{Hg(\lambda_{xy,c})}(h, g(\lambda_{xy,c}(x,y))),$$

$$p_{\lambda_{xy,c}|S_c}(\lambda_{xy,c}) \longrightarrow p_{g(\lambda_{xy,c})|S_c}(g(\lambda_{xy,c})),$$

$$p_{\lambda_{xy,c}}(\lambda_{xy,c}(x,y)) \longrightarrow p_{g(\lambda_{xy,c})}(g(\lambda_{xy,c}(x,y))),$$

where $g(\lambda_{xy,c}) \triangleq \log(\delta + \lambda_{xy,c})$, where $\delta > 0$ is a very small constant that aims to prevent numerical underflow. This nonlinear transformation is motivated by the fact that $\lambda_{xy,c} > 0$ and hence has a nonsmooth pdf at $\lambda_{xy,c} = 0$, which is difficult to estimate using nonparametric methods. The transformation $g(\cdot)$ is aimed to smooth out the density of $\lambda_{xy,c}$ for an easier nonparametric estimation. Note that for any one-to-one transformation $g(\cdot)$, the change of variables does not affect the estimation $\lambda_{h,c}(h, \lambda_{xy,c}(x,y))$ due to the axioms of probability (see appendix B).

To estimate history-independent encoding models, we set $\lambda_{h,c}(h, \lambda_{xy,c}(x, y)) = \frac{N}{T}$.

Finally, we define the place receptive field (RF) as the expected value of $\lambda_c(x, y, h)$ with respect to the probability distribution $p_{H_c}(h)$:

$$\text{RF}_c(x, y) = \mathbb{E}_h[\lambda_c(x, y, h)] = \lambda_{xy,c}(x, y) \int \lambda_{h,c}(h, \lambda_{xy,c}) p_{H_c}(h) dh. \quad (2.13)$$

Therefore, the estimation of the RF consists of three steps: (1) estimate $\lambda_{xy,c}(x, y)$; (2) estimate $\lambda_{h,c}(h, \lambda_{xy,c})$ (note that step 2 is dependent on step 1, but step 1 is independent of step 2, and no iterative solution is required); and (3) estimate the expected value and product according to equation 2.13.

Decoding. Assuming a change of variable: $h_c \triangleq \log \tau_c$, instead of using equation 2.3, we use an equivalent yet simpler form of the MAP estimator:

$$(\hat{x}_t, \hat{y}_t) = \text{argmax}_{x,y} \ p_{X,Y}(x, y) \prod_{c=1}^{m} p_{H_c|X,Y}(h_c(t)|x, y), \quad (2.14)$$

which yields the identical estimator due to rules of probability: $p_{H_c|X,Y}(h_c(t)|x, y) = p_{\tau_c(t)|X,Y}(\tau_c|x, y) \tau_c(t)$, where $\tau_c = \exp(h_c) = |\frac{d}{dh_c} \exp(h_c)|$ (see equation B.3 in appendix B) has the same dependence on (x, y). To simplify equation 2.14, we assume that $\lambda_{xy,c}$ is a sufficient statistic (Lehmann

& Scheffé, 1950) for h_c for the cth neuron. That is, knowing (x, y) does not convey any additional information about h_c if $\lambda_{xy,c}(x, y)$ is known. This assumption leads to

$$(\hat{x}_t, \hat{y}_t) = \text{argmax}_{x,y} \; p_{X,Y}(x, y) \prod_{c=1}^{m} p_{H_c|\lambda_{xy,c}}(h_c(t)|\lambda_{xy,c}(x, y)). \tag{2.15}$$

Now, the density $p_{H_c|\lambda_{xy,c}}(h_c(t)|\lambda_{xy,c})$ ((x, y) has been dropped from $\lambda_{xy,c}(x, y)$ for brevity of notation) is estimated as

$$p_{H_c|\lambda_{xy,c}}(h_c(t)|\lambda_{xy,c}) = \frac{p_{H_c\lambda_{xy,c}}(h_c(t), \lambda_{xy,c})}{p_{\lambda_{xy,c}}(\lambda_{xy,c})} = \frac{p_{H_c g(\lambda_{xy,c})}(h_c(t), g(\lambda_{xy,c}))}{p_{g(\lambda_{xy,c})}(g(\lambda_{xy,c}))},$$

$$\tag{2.16}$$

where $g(\lambda_{xy,c}) \triangleq \log(\delta + \lambda_{xy,c})$ is a one-to-one nonlinear transformation. The second equality is derived from rules of probability (see equations B.4 and B.5 in appendix B). The densities $p_{X,Y}(x, y)$, $p_{H_c g(\lambda_{xy,c})}(h_c(t)g(\lambda_{xy,c}))$, and $p_{g(\lambda_{xy,c})}(g(\lambda_{xy,c}))$ are estimated nonparametrically, as described in the next section.

For spike history–independent decoding models, the $P(\tau_c(t)|x, y)$ in equation 2.3 is first replaced by $P(\tau_c(t)|\lambda_{xy,c}(x, y))$ using the sufficient statistic assumption. Then $P(\tau_c(t)|\lambda_{xy,c}(x, y))$ is calculated using pseudo-equilibrium assumption (as done in GLM decoding): $P(\tau_c(t)|\lambda_{xy,c}(x, y)) = \frac{N}{T}\lambda_{xy,c}(x, y)e^{-\frac{N}{T}\lambda_{xy,c}(x,y)\tau_c(t)}$.

The MAP estimators $(\hat{x}, \hat{y})$ are computed by maximizing equation 2.3 or 2.14 over a 2D grid of coordinate space. These MAP estimators are then smoothed by filtering bidirectionally by a two-pole Butterworth filter with cutoff frequency 0.35 Hz (for containing 95% power of true (x, y) coordinates) to obtain the final smooth estimate $(\tilde{x}_t, \tilde{y}_t)$. Figure 3 provides a summary of the encoding and decoding processes.

Model estimation and validation. Note that the rate functions $\lambda_{xy,c}(x, y)$ and $\lambda_{h,c}(h, \lambda_{xy,c})$ appear as products and ratios of densities. These densities are estimated using both second-order kernel density estimators (KDE) (Silverman, 1986) and our recently developed nonparametric band-limited ML estimator BLML (Agarwal, Chen et al., 2015) (see also appendix A). Specifically, the two nonparametric density estimators are defined as follows:

- Kernel density estimators (KDE): KDE assumes that the pdf is a linear combination of kernel functions centered at the data samples as follows:
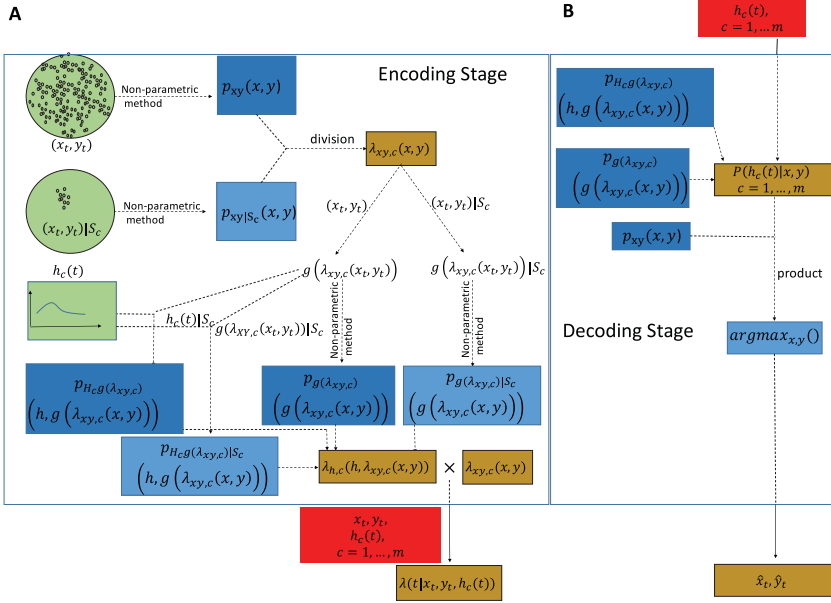
Figure 3: Encoding and decoding processes under the nonparametric framework. (A) Encoding process and (B) decoding process. Green boxes denote training data, blue boxes denote densities estimated nonparametrically using training data, yellow boxes denote important functions computed using these densities, and red boxes denote test data. The CIF and decoding estimates are obtained by applying functions in tan boxes on the test data. Dark blue boxes denote the densities used in both encoding and decoding processes. $g(\cdot)$ is a nonlinear function to smooth out dependence on $\lambda_{xy,c}$ for an easier nonparametric estimation.

$$p(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{b}}(\mathbf{x} - \mathbf{x}_i). \tag{2.17}$$

Here, the vectors $\{\mathbf{x}_i\}_{i=1}^{n} \in \mathbb{R}^d$ are the data samples, and $K_{\mathbf{b}}$ is a d-dimensional gaussian kernel with bandwidth vector $\mathbf{b} \in \mathbb{R}^d$. We have used the isotropic kernel, $\mathbf{b} = [b_1, \dots, b_d]$, with an optimal bandwidth $b_i = constant_i \times n^{-0.2}$ (Silverman, 1986), where $constant_i$s are determined by a cross-validation procedure.

- Band-limited maximum likelihood (BLML) estimation, which assumes that the pdf square root is band-limited and smooth (Agarwal, Chen et al., 2015). The BLML estimator has the following form:

$$p(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}_i \operatorname{sinc}_{\mathbf{f}_c}(\mathbf{x} - \mathbf{x}_i) \right)^2. \tag{2.18}$$

Here, the vectors $\{\mathbf{x}\}_{i=1}^n \in \mathbb{R}^d$ are the data samples, $\mathbf{f_c}$ denotes the vector of cutoff frequencies, $\mathrm{sinc}_{\mathbf{f_c}}(\mathbf{x}) \triangleq \prod_{j=1}^d \frac{\sin(\pi f_{cj} x_j)}{\pi x_j}$, and the vector $\hat{\boldsymbol{\alpha}} \triangleq [\hat{\alpha}_1, \cdots, \hat{\alpha}_n]^\top$ is computed by solving the following optimization problem (see appendix A for details):

$$\hat{\boldsymbol{\alpha}} = \underset{\rho_n(\mathbf{c})=0}{\arg\max}\left(\prod \frac{1}{\alpha_i^2}\right), \qquad (2.19)$$

where, $\rho_{ni}(\boldsymbol{\alpha}) \triangleq \sum_{j=1}^n \alpha_j \mathrm{sinc}_{\mathbf{f_c}}(\mathbf{x}_i - \mathbf{x}_j) - \frac{n}{\alpha_i}$. Note that BLML appears similar to KDE, but it is different in the following ways: (1) BLML is a non-parametric ML estimator, while KDE is not optimal in any simple statistical sense; (2) the summation in the BLML estimator is squared; and (3) the coefficients in front of the sinc functions in the BLML estimator are computed by solving an optimization problem that falls out of the ML procedure. The BLML estimator is computed using the BLMLQuick algorithm (Agarwal, Chen et al., 2015) defined in appendix A.

For simplicity, despite the shape differences in place receptive fields, we assume that all neuronal place receptive fields have the same level of smoothness in the x and y directions across the entire population. Namely, we use the one-vector of bandwidths $(b_x, b_{\lambda_{xy}}, b_h)$ (for the KDE estimator) and cutoff frequencies $(f_x, f_{\lambda_{xy}}, f_h)$ (for the BLML estimator) for the entire neuronal population. To select these hyperparameters of bandwidths and cutoff frequencies, performance as measured by goodness-of-fit on validation data was evaluated as in Kloosterman et al. (2014). Specifically, the combination of bandwidths and cutoff frequencies that produces the best decoding performance on test data (as measured by R^2 statistic) is selected. As the optimization on cutoff frequency is a 3D problem, a hill-climbing strategy (Yuret & De La Maza, 1993) (as done for parametric models) is used to find cutoff frequencies and bandwidths that jointly maximize decoding performance locally. Particularly, the first parameter is varied while fixing the second and third parameter to some initial values. A maximum decoding performance is obtained in the first dimension, and the first parameter is then fixed to the corresponding value. The second and third parameters are then varied, and their respective values that provide maximum decoding performance are obtained. The process is repeated until a local maximum in all dimensions is found: namely, no improvement in decoding performance is observed by varying any of the three parameters in any direction.

Similar to GLM estimation, out of 30.36 minutes of training data, the first 85% is used for model building and the remaining 15% to validate hyperparameters. Once the hyperparameters are validated, the model is retrained on the entire 30.36 minutes of training data and assessed on the new 6 minute test data.

2.4 Performance Evaluation

2.4.1 Encoding Models. The goodness-of-fit of the eight estimators (`GLMquad`, `GLMzern`, `KDE`, `BLML` with and without spike history) of $\lambda(t)$ is assessed by using the time-rescaling theorem and the Kolmogorov-Smirnov (KS) statistic (Brown et al., 2002) on 6 minutes of test data. Particularly, we tested if rescaled times are distributed as a Poisson process with rate 1. The similarity between the two cumulative distribution functions is quantified using the normalized (by $\frac{1.36}{\sqrt{N}}$, $N =$ number of spikes) KS statistic. Due to normalization, a value of $KS > 1$ indicates that the estimated λ is too extreme ($p < 0.05$, one-sample KS test) to generate the test data. The closer the normalized KS statistic is to 0, the better is the estimate.

Since the KS statistic may be insufficient to assess non-Poisson neural encoding models (Pillow, 2009; Gerhard & Gerstner, 2010), we also compute the negative log likelihood of 6 minute test data for all tested models:

$$\log l = \sum_{t=1}^{T} \lambda_c(t)\Delta t - \sum_{t=1}^{T} s_c(t) \log(\lambda_c(t)\Delta t), \qquad (2.20)$$

where $\Delta t = 1$ ms is used. A smaller negative log-likelihood value on test data implies better data predictability. For the purpose of comparison and better display, the negative loglikelihoods for all models are plotted against the result of a baseline model (`GLMquad` with no spike history). In the reported negative log-likelihood statistics, all values have been normalized by 200 and then shifted up by 3.

2.4.2 Decoding Models. The decoding model performance is assessed by using the R^2 statistic of estimates obtained from 6 minutes of test data. In addition, box plots with median, 5%, 25%, 75%, and 95% quantiles of the decoding errors $\sqrt{(\tilde{x}_t - x_t)^2 + (\tilde{y}_t - y_t)^2}$ for all the eight models are presented.

All parametric and nonparametric methods are implemented in customized software written in Matlab2014a (MathWorks, Natick, MA).

3 Results

3.1 Hyperparameter Validation.
The validated hyperparameters of parametric and nonparametric models are summarized in Table 1. Note that spike history–independent parametric models contain no hyperparameter.

3.2 Encoding: CIF Estimation.
Figure 4 shows the estimated CIFs using GLM methods for a unimodal and multimodal place cell, respectively. In Figure 4A, the `GLMquad` method with and without history are compared.

Table 1: Model Hyperparameters Selected from Validation of Experimental Data.

Model	Without Spike History	With Spike History
GLMquad	NA	$[J, \rho] = [35, 0]$
GLMzern	NA	$[J, \rho] = [30, 3.24]$
KDE	$b_x = 0.667n^{-0.2}$	$[b_x, b_{\lambda_{xy}}, b_h] = [0.13, 0.33, 0.24]n^{-0.2}$
BLML	$f_x = 3.6$	$[f_x, f_{\lambda_{xy}}, f_h] = [4.4, 3.0, 3.6]$

As can be seen in Figure 4A(ii,iii), both methods capture the position dependence. However, as shown in Figure 4A(v), the $\lambda(t)$ estimated using GLMquad with no spike history dependence (GLMquadNH) fails to capture bursting and refractoriness phenomena, as evidenced in the KS statistic (2.01 versus 1.34 with spike history dependence) and negative log likelihood (3 versus 2.28 with spike history dependence) (see Figure 4A(vi)).

In Figure 4B, the performance of GLMquad and GLMzern (both with history dependence) on a multimodal place cell data is shown. It can be seen that GLMquad method fails to capture the spatial complexity in firing patterns of the multimodal place cell. Although, GLMzern did better in capturing the spatial complexity, KS = 2.2 implies scope for improvement in model fitting (see Figure 4B(vi)).

The encoding results are summarized in Table 2. In contrast to the GLM, the nonparametric encoding models yield better performance for the multimodal place cell data for the same multimodal cell. Figure 5B shows that the KS statistics for both BLML (KS = 0.58) and KDE (KS = 0.84) are lower than 1. Similarly, the negative log-likelihood values (0.85 for BLML and 0.98 for KDE) are better than those of GLMquad (1.83) and GLMzern (1.43). The estimates for $\lambda_{xy,c}(x, y)$ using BLML and KDE are shown in Figures 5C and 5E, respectively. First, it can be seen that for both place and grid cells, $\lambda_{xy,c}(x, y)$ is larger wherever the neuron spikes more, verifying the "place-like" and "grid-like" structures. Second, the BLML method produces a smoother $\lambda_{xy,c}(x, y)$ than the KDE method. Figures 5D and 5F show $\lambda_h(h, \lambda_{xy,c})$ for the BLML and KDE methods, respectively.

Figure 5G shows the estimated CIF over time from the nonparametric KDE and BLML methods. Both methods are successful in capturing refractoriness and bursting in the neuronal activity. Specifically, refractoriness is captured by smaller values near $\tau = 0$ across all λ_{xy}, and bursting is captured by sudden large values near $\tau = 4$ ms across all λ_{xy} (see Figures 5D and 5F). The values achieved at $\tau = 4$ and smaller λ_{xy} are higher than the values achieved at $\tau = 4$ ms and larger $\lambda_{xy,c}$. This suggests that even if $\lambda_{xy,c}$ is small, $\lambda_c(t) = \lambda_{xy,c}(x_t, y_t)\lambda_h(h_t, \lambda_{xy,c}(x_t, y_t))$ can still reach a significant value if a spike occurred in the past 4 ms. This in turn makes the values of $\lambda(t)$ after
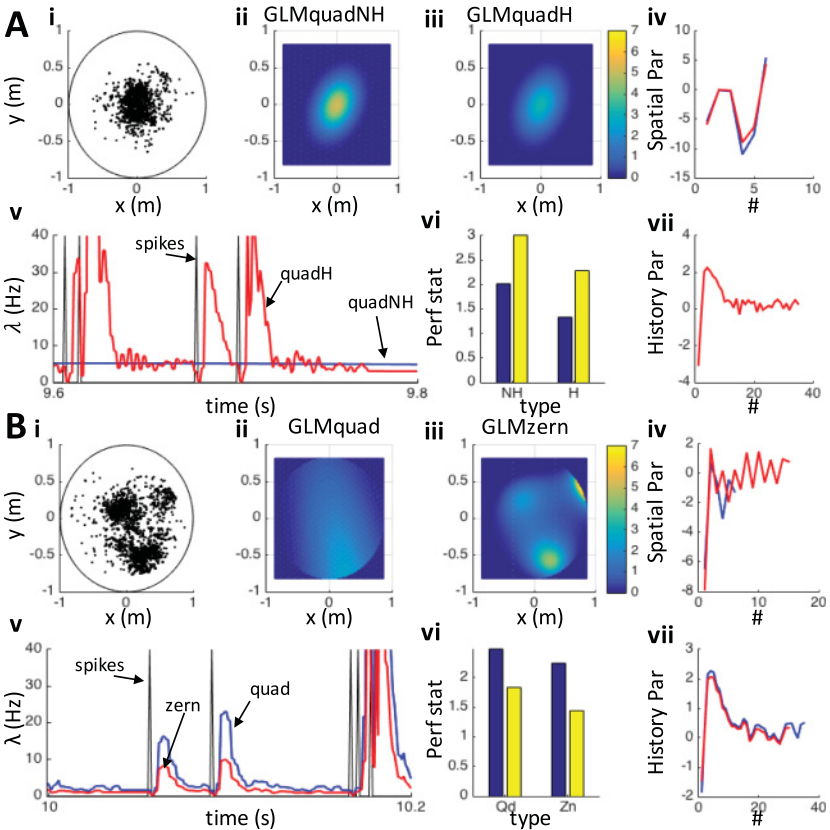
Figure 4: Performance of GLM methods on a (A) unimodal place cell and (B) multimodal place cell. (i) The rat's spatial position when the cells spiked are marked by dots. (ii, iii) Estimated dependence of CIF on (x, y) coordinates, using different GLMs. (iv) Estimated spatial parameters for different models. (v) Estimated CIF as a function of time. (vi) Performance metrics of different models evaluated on test data, KS statistic (blue bar), negative log likelihood (yellow bar). `GLMquadNH`, `GLMquadH`, and `GLMzernH` denote the respective GLM without and with spike history.

4 ms of any spike almost constant, which is therefore independent of (x, y). These findings confirm that bursting in place and grid cells is an internal phenomenon that is not affected much by the external stimulus.

Figure 6 shows the recorded and simulated activity (using the `BLML` model) for a multimodal place cell shown in Figure 5. It can be seen that at both macro- and microlevels, the nonparametric `BLML` model is able to reproduce the recorded activity; the spatial dependence is nicely captured as

Table 2: Comparison of Performance Metrics (on Test Data) of Parametric and Nonparametric Methods in Encoding and Decoding.

Model	Median KS Statistic	Median Negative Log Likelihood	Median Decoding Error (cm)	R^2
Unimodal cells ($m = 27$)				
GLMquad (w/o, w/ history)	4.33, 2.04	3.00, 2.36	23.59, 24.01	0.40, 0.44
GLMzern (w/o, w/ history)	3.80, 1.88	2.90, 2.28	24.21, 28.25	0.27, 0.32
KDE (w/o, w/ history)	4.41, **0.99**	2.91, 2.29	25.67, **15.39**	0.40, 0.80
BLML (w/o, w/ history)	4.04, 1.01	2.94, **1.82**	22.29, 15.51	0.55, **0.83**
Multimodal ($m = 26$)				
GLMquad (w/o, w/ history)	5.74, 2.50	3.00, 2.00	29.70, 30.81	0.36, 0.34
GLMzern (w/o, w/ history)	4.65, 2.25	2.72, 1.83	21.58, 23.62	0.44, 0.55
KDE (w/o, w/ history)	4.86, 1.63	2.70, 2.07	19.07, **14.12**	0.64, **0.83**
BLML (w/o, w/ history)	4.65, **1.30**	2.66, **1.22**	19.99, 14.42	0.62, **0.83**
All combined ($m = 53$)				
GLMquad (w/o, w/ history)	4.82, 2.36	3.00, 2.15	24.24, 23.63	0.47, 0.50
GLMzern (w/o, w/ history)	4.27, 2.20	2.82, 1.94	20.04, 23.40	0.47, 0.59
KDE (w/o, w/ history)	4.58, 1.35	2.80, 2.25	21.75, 12.85	0.64, 0.88
BLML (w/o, w/ history)	4.25, **1.09**	2.79, **1.52**	18.21, **12.39**	0.72, **0.89**

Note: The best performance in each column is shown in bold.

shown in Figure 6B, along with microlevel bursting, as shown in Figure 6D. Further, it can be seen in the zoom-in view in Figure 6C that there is high variability (Fenton & Muller, 1998; Moser et al., 2008) between the spike counts on different trajectories that pass through the region, even though the $\lambda_{xy,c}$ is almost constant over this small region. This effect is replicated in simulated activity due to incorporation of spike history. Statistical dependencies on the spike history, particularly bursting, makes the neuron either spike at least three of four times if it fires once or the neuron does not fire at all, resulting in increased variance.

For the entire population of unimodal place, cells, multimodal place, and grid cells, models with spike history are significantly better as measured by KS statistics and predictive log likelihoods than the models without spike history for BLML (KS statistic, $P = 2.5e$-10, log likelihood $P = 2.39e$-10, paired Wilcoxon signed-rank test), KDE (KS statistic $P = 2.4e$-10, log likelihood $P = 1.94e$-05), GLMquad (KS statistic $P = 2.4e$-10, log likelihood $P = 3.57e$-08), and GLMzern (KS statistic $P = 2.5e$-10, log likelihood $P = 2.39e$-10). Further, BLML methods with spike history (median KS statistics $= 1.09$, median negative log likelihood $= 1.52$) perform significantly better than the KDE models with spike history (median KS statistics 1.35, negative log likelihood 2.25), GLMquad with spike history (median KS statistics 2.36, negative log likelihood 2.15), and GLMzern with spike history (median KS statistics 2.2, negative log likelihood 1.94) methods (respective P-values for
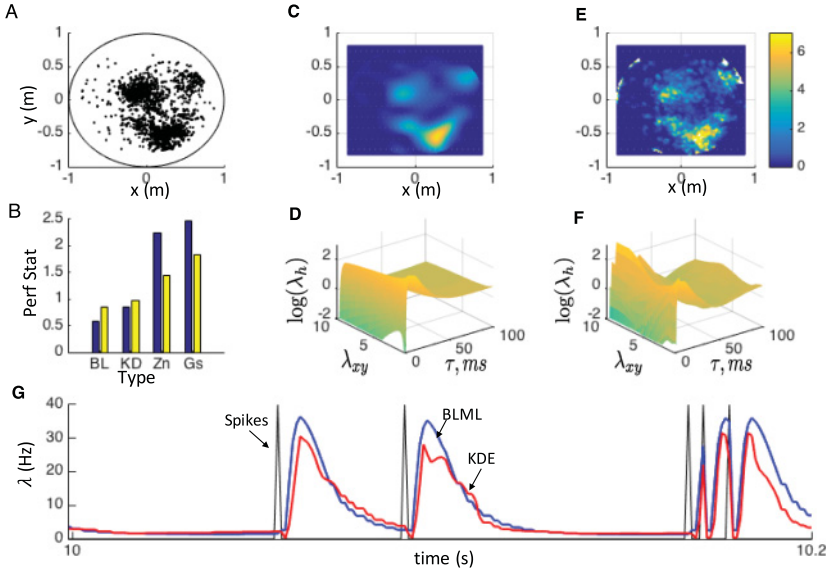
Figure 5: Comparison of BLML, KDE, and GLM methods with spike history dependence for a multimodal place cell. (A) Each black dot marks the rat's spatial location when the cell spiked. (B) The performance stats for BLML, KDE, GLMquad, and GLMzern using test data (blue bars KS statistic, yellow bars negative log likelihood). (C,E) Estimated $\lambda_{xy,c}(x, y)$ for BLML and KDE, respectively. (D,F) Estimated $\lambda_{h,c}(h, \lambda_{xy,c})$ for BLML and KDE, respectively. (G) Estimated CIF as a function of time for the two nonparametric encoding methods.

KS-statistic and negative log likelihood: $P = 2.19\text{e-}05$ and $2.7\text{e-}10$, $P = 1.03\text{e-}09$ and $2.5\text{e-}10$, $P = 1.03\text{e-}09$ and $4.7\text{e-}10$, paired Wilcoxon signed-rank test). Box plots showing the KS statistic and negative log likelihood values for all eight statistical models are shown in Figures 7 and 8. As expected, the improvement in CIF estimation using nonparametric methods is greater for the multimodal place cell and grid cell populations (the gap between the median negative log likelihood is 0.61 between BLML and GLMzern with spike history) as compared to unimodal place cell population (the gap between the median negative log likelihood is 0.45 between BLML and GLMzern with spike history).

3.3 Decoding. The decoding results are summarized in Table 2. In addition, Figure 9 shows a summary of decoding errors $\sqrt{(\tilde{x}_t - x_t)^2 + (\tilde{y}_t - y_t)^2}$ obtained using 6 minutes of test data, over the population of unimodal place cells, multimodal place, and grid cells and both using BLML, KDE, GLMquad, and GLMzern with and without spike history covariates. Overall,
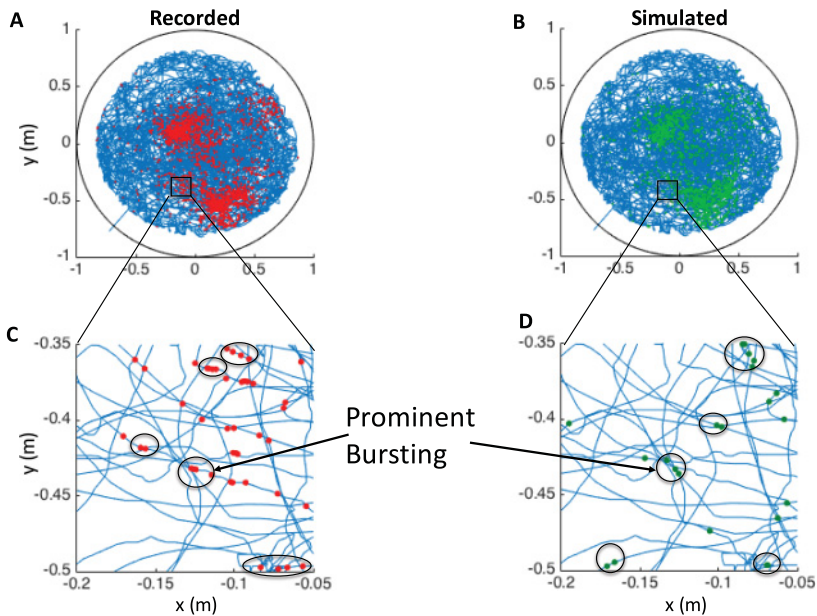
Figure 6: Simulated activity for a multimodal place cell using BLML. (A,C) The recorded spiking activity of the multimodal place cell shown in Figure 5 and its zoom-in view. (B,D) The simulated spiking activity of the multimodal place cell and its zoom-in view. Each red dot marks the spike with respect to spatial coordinates.

models with spike history (respective R^2 statistics: 0.89, 0.88, 0.50, 0.59; respective median decoding error: 12.39, 12.85, 23.62, 23.40 cm) perform significantly better than their counterparts without spike history (respective R^2: 0.72, 0.64, 0.47, 0.47; respective median decoding error: 18.21, 21.75, 24.24, 20.04 cm; respective P-values of signed-rank test: 9.1e-97, 1.7e-135, 7.4e-07, 1.6e-05). Note that although the R^2 value is better for GLMzern with spike history (0.59 versus 0.47), the median decoding error is slightly worse (23.4 versus 20.04 cm). More important, BLML with spike history performs the best among all four methods (respective P-values of signed-rank test: 0.02, 5.7e-165, and 4.0e-203 for KDE, GLMquad, and GLMzern in decoding error). Similarly, BLML without spike history also performs significantly better than KDE, GLMquad, and GLMzern in decoding errors (respective P-values: 8.6e-27, 9.9e-54, and 1.01e-16). From Table 2, it can be seen that BLML with or without spike history also performs the best in decoding the rat's position only using unimodal cells or multimodal cells. Interestingly, GLMquad has better performance than GLMzern in R^2 for unimodal cells, yet worse performance for multimodal cells.
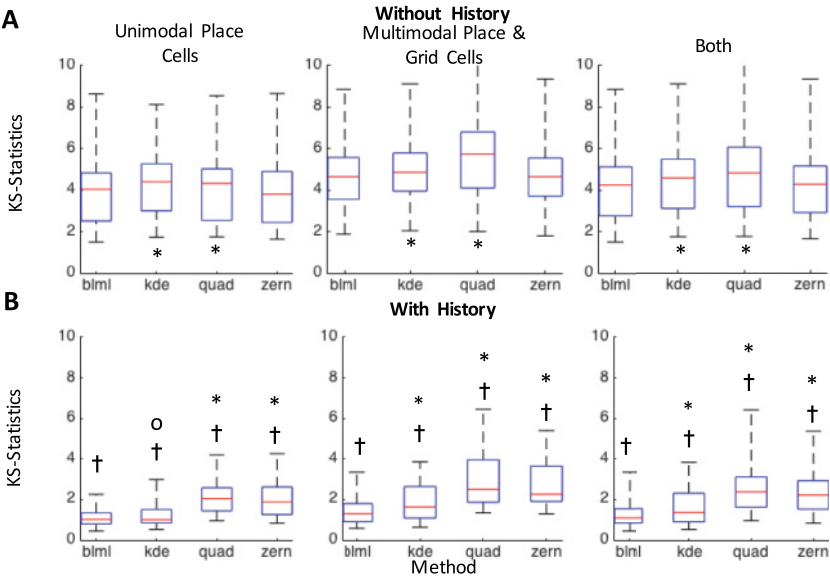
Figure 7: Comparison of KS-statistics values. Box plots for KS statistics for the BLML, KDE, GLMzern, and GLMquad methods (A) without spike history dependence and (B) with spike history dependence for unimodal place cells, multimodal place cells and grid cells and both combined. * and ∘ denote $P < 0.01$ and $P < 0.05$, respectively, for comparisons between the mean KS statistics achieved by BLML and the indicated method. † means $P < 0.01$ for comparisons between the mean KS statistics achieved by a method without and with spike history dependence. All P-values are calculated using the Wilcoxon signed-rank test.

As an illustration, Figure 10 shows the reconstructed rat's trajectory calculated using BLML and GLMzern with spike history, respectively, for a 1 minute period in the test data. The animal's actual position (black) is overlaid on the estimates (green; red), and each subpanel shows the reconstruction for a 20 second period. A video clip for demonstration of our decoding performance in experimental data (using BLML) is provided in the supplementary material.

4 Discussion and Conclusion

4.1 Nonparametric Modeling: Flexibility and Robustness. In this letter, we develop a novel nonparametric approach to study encoding properties and decoding performance in hippocampal place cells and entorhinal grid cells. Our approach is inspired by the nonparametric encoding framework previously proposed in Chen et al. (2012) and Kloosterman et al.
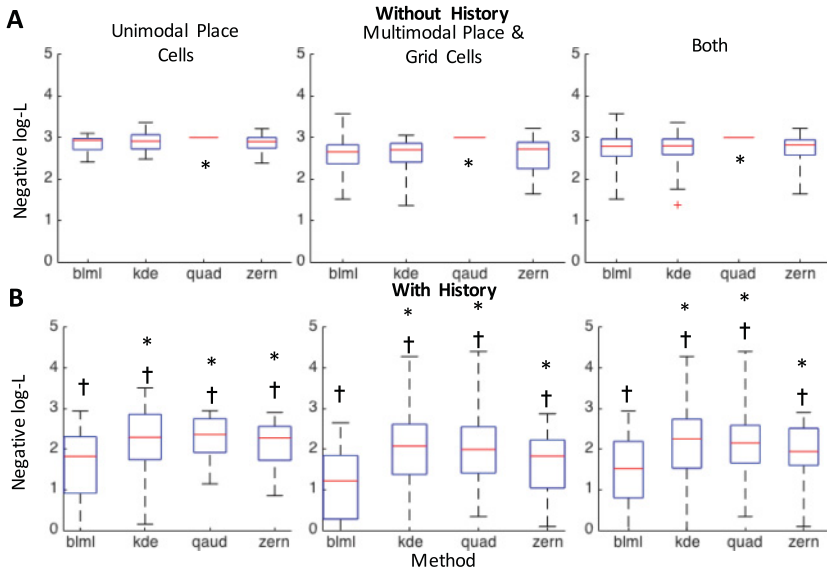
Figure 8: Comparison of predictive negative log-likelihood values. Box plots for negative log-likelihood values for BLML, KDE, GLMquad, and GLMzern methods (A) without spike history dependence and (B) with spike history dependence for unimodal place cells, multimodal place cells and grid cells and both combined. * denote $P < 0.01$ for comparisons between the mean KS statistics derived from BLML and the other indicated method. † means $P < 0.01$ for comparisons between the mean KS statistics achieved by a method without and with spike history dependence. All P-values are calculated using the Wilcoxon signed-rank test.

(2014). The basic idea of the nonparametric encoding framework is to convert the encoding or decoding problem to a density estimation problem. In this letter, two nonparametric density estimators are constructed: the BLML estimator Agarwal, Chen et al. (2015) and the standard KDE. The performance of these two nonparametric models is then compared to parametric GLM models (using either quadratic or Zernike polynomials). The BLML estimator outperforms both parametric methods. The BLML model has the best performance for encoding models, whereas KDE methods are comparable to BLML in decoding performance. Both BLML and KDE methods outperformed parametric models in both encoding and decoding performances.

The proposed neural encoding models also capture the statistical dependence of spike history on current spiking activity. This allows modeling of neuronal spiking at a millisecond timescale, and, hence, both bursting and refractoriness phenomenon are captured successfully. Further, our result can also explain the observed extra variance in spiking activity, which cannot be accounted sufficiently by an inhomogeneous Poisson process
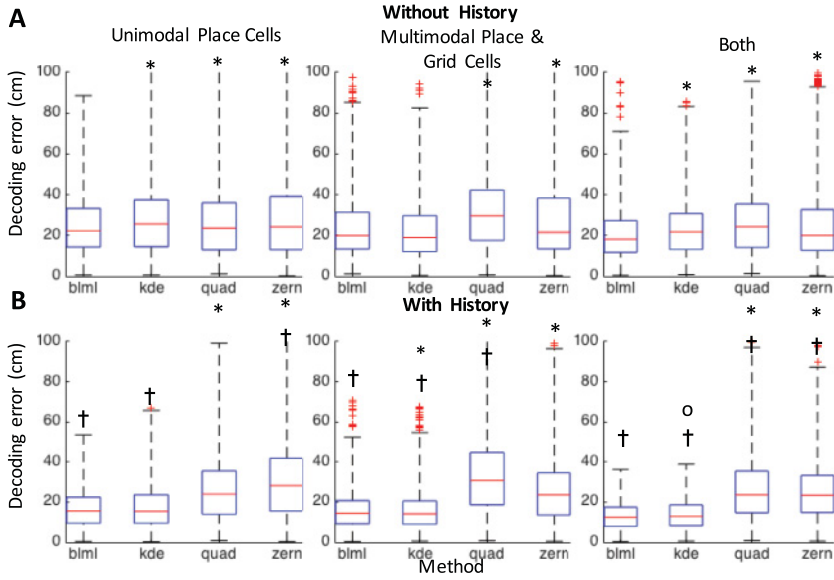
Figure 9: Comparison in decoding error performance. Box plots of neural decoding errors for the BLML, KDE, GLMzern, and GLMquad methods without (A) and with (B) spike history dependence for unimodal place cells, multimodal place cells, and grid cells and both combined. * denotes $P < 0.01$ for comparisons between the mean KS statistics derived from BLML and the other indicated method. † means $P < 0.01$ for comparisons between the mean KS statistics achieved by a method without and with spike history dependence. All P-values are calculated using Wilcoxon signed rank-test.

(Fenton & Muller, 1998; Moser et al., 2008). This improved the decoding performance of our methods over state-of-the-art decoding methods that ignore spike history dependence and hence assume incorrectly that neuronal spiking is an inhomogeneous Poisson process phenomenon.

Due to the nonparametric nature of the BLML estimator, the same model structure (with the same cutoff frequency) is able to characterize a variety of receptive field types observed in both place and grid cells. This modeling flexibility is an important step forward towards a practical solution of neural decoding, required in brain-machine interfaces. Currently, neural decoding via implantable devices suffers from random drifts in recordings and neuronal firing over time. Such drifts may include changes in firing patterns of the cell itself due to change in neural structure or environment (Moser et al., 2008; Frank, Stanley, & Brown, 2004; Muller & Kubie, 1987). Therefore, to achieve robust decoding, the encoding models should be flexible so that they can model a variety of receptive fields. As shown in this letter, the nonparametric method constructed by the BLML estimator is
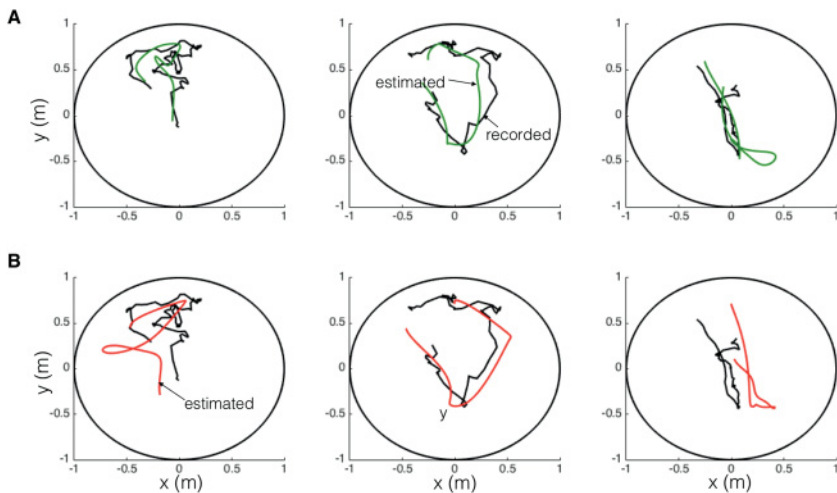
Figure 10: Comparison of trajectory decoding using nonparametric (BLML) and parametric (GLMzern) decoding approaches. Reconstructed rat's run trajectory calculated using (A) BLML and (B) GLMzern (both with spike history covariates), respectively, for a 1 minute period in test data. Each subpanel shows the reconstruction for a 20 second period.

statistically and computationally efficient, flexible enough to characterize a variety of receptive field types in the encoding phase, and may remain robust if the receptive field structure changes over time.

More important, our neural decoding results are the first attempt to decode the rat's trajectory in a 2D space using multimodal place receptive fields (from either place cells or grid cells) together with spike history. The popular parametric GLM framework is insufficient for modeling such multimodal place and grid cells with or without spike history (Barbieri et al., 2004; Brown et al., 2002). Previous attempts using nonparametric models mainly concentrated on decoding from unsorted multiunit activity, and the decoding was achieved in a 1D environment (Kloosterman et al., 2014) without using information stored in spiking history.

4.2 Fourier Representation in Place Receptive Fields. Recently, researchers (Kubie & Fox, 2015; Ormond & McNaughton, 2015) have hypothesized that the complexity in hippocampal place receptive fields is formed as a result of a Fourier-like summation of multiple receptive fields of grid cells (see Figure 11), and the grid cells' receptive field can be closely approximated by a 2D cosine function with a specific spatial frequency. These grid cells are generally arranged in anatomical regions called modules (see Figure 11A) that contain the grid cells with similar spatial frequencies
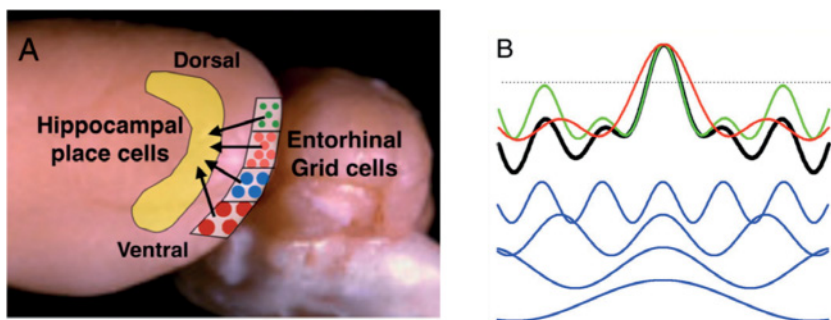
Figure 11: The Fourier hypothesis. (A) A lateral view of rat brain illustrating medial entorhinal cortex projection to the hippocampus. The grid cells are arranged in modules within the medial entorhinal cortex. Each module contains grid cells that have receptive fields with similar spatial frequency. The hippocampal place cells receive synaptic projections from grid cells in different modules. (B) The anatomical connections suggest that the hippocampal place cells linearly sum up the inputs from grid fields to generate a place field. (Figure courtesy of Kubie & Fox, 2015, reprinted by permission of PNAS.)

(Stensola et al., 2012). The hippocampal place cells receive projections from grid cells across multiple modules (each with distinct spatial frequencies). A toy example of such a sum is shown in Figure 11B. This hypothesis is called the Fourier hypothesis. If the Fourier hypothesis is biologically plausible, then it will also explain the superior performance of the nonparametric BLML estimator, which assumes that place and grid fields are supported by sinusoidal functions as in the Fourier representation. However, the superior performance of the BLML estimator in this investigation may also be due to superiority of BLML density estimators. It remains to be tested which of the aforementioned is true.

4.3 Extension and Future Work. Several extensions of our work can be considered in future investigations. First, in neural encoding, we can in principle incorporate other informative neural covariates, such as the power or phase of the local field potential (Agarwal & Sarma, 2010, 2012a; Agarwal et al., 2014). The expectation is that, this will further increase the performance of the encoding model. In addition, finding an efficient procedure for bandwidth parameter selection is important to obtain good performance. Details on this investigation can be found in our previous work (Agarwal, Chen et al., 2015). Second, we can formulate the neural decoding problem within a state-space framework (Brown et al., 1998; Chen, 2013, 2015), by considering a smooth temporal prior. This will further improve the decoding accuracy. Finally, our methodology is not limited to rodent hippocampal

place cells or grid cells; it can be extended to other modalities, such as the retina, lateral geniculate nucleus, primary visual cortex, and primary motor cortex.

Appendix A: `BLMLQuick` Algorithm for `BLML` Estimator ⎯⎯⎯⎯⎯⎯⎯

To make this letter self-contained, we present a short introduction of the `BLML` estimator and `BLMLQuick` algorithm. The detailed results are provided in (Agarwal, Chen et al., 2015).

A.1 Formulation of the `BLML` Estimator. Consider a pdf, $p(x) \triangleq v^2(x)$, of a random variable $x \in \mathbb{R}$ (unit: dimensionless) with Fourier transform $\mathcal{V}(\omega) \triangleq \int v(x)e^{-j\omega x}\,dx$ which has support in $(\frac{-\omega_c}{2}, \frac{\omega_c}{2})$ (unit: radian), namely, $\omega_c = 2\pi f_c \in \mathbb{R}$, is the cutoff frequency of $\mathcal{P}(\omega)$ (where f_c denotes the corresponding cutoff frequency in unit Hz). Then,

$$\mathbb{U}(\omega_c) = \left\{ v^2 : \mathbb{R} \to \mathbb{R}^+ \int v^2(x)dx = 1, \mathcal{V}(\omega) = 0 \ \forall |\omega| > \frac{\omega_c}{2} \right\} \qquad \text{(A.1)}$$

Define,

$$\mathbb{V}(\omega_c) = \left\{ v : \mathbb{R} \to \mathbb{R} \mid v(x) = \sqrt{p(x)}, \ p \in \mathbb{U}(\omega_c) \right\}. \qquad \text{(A.2)}$$

and

$$\mathbb{W}(\omega_c) = \left\{ \mathcal{V} : \mathbb{R} \to \mathbb{C} \mid \mathcal{V}(\omega) = \int v(x)e^{-j\omega x}dx, \ v \in \mathbb{V}(\omega_c) \right\}, \qquad \text{(A.3)}$$

where $j = \sqrt{-1}$. Note that $\mathbb{V}(\omega_c)$ and $\mathbb{W}(\omega_c)$ are Hilbert spaces with the inner product defined as $\langle a, b \rangle = \int a(x)b^*(x)dx$, $\langle a, b \rangle = \frac{1}{2\pi} \int a(\omega)b^*(\omega)d\omega$, respectively. The norm $||a||_2^2 = \langle a, a \rangle$ is defined for both spaces. Further, note that for all elements in $\mathbb{V}(\omega_c)$ and $\mathbb{W}(\omega_c)$, $||a||_2^2 = \langle a, a \rangle = 1$.

A.2 The Likelihood Function for Band-Limited PDFs. Now consider a random variable, $x \in \mathbb{R}$, with unknown pdf $p(x) \in \mathbb{U}(\omega_c)$ and its n independent realizations $x_1, x_2, \ldots, x_n$. The likelihood $\mathcal{L}(x_1, \ldots, x_n)$ of observing $x_1, \ldots, x_n$ is

$$\mathcal{L}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} v^2(x_i), \ v \in \mathbb{V}(\omega_c) \qquad \text{(A.4a)}$$

$$= \prod_{i=1}^{n} \left(\frac{1}{2\pi} \int \mathcal{V}(\omega)e^{j\omega x_i}d\omega \right)^2, \ \mathcal{V} \in \mathbb{W}(\omega_c). \qquad \text{(A.4b)}$$

Defining

$$b_i(\omega) \triangleq \begin{cases} e^{-j\omega x_i} & \forall \, \omega \in \left(-\frac{\omega_c}{2}, \frac{\omega_c}{2}\right) \\ 0 & \text{otherwise} \end{cases} \tag{A.5}$$

yields

$$\mathcal{L}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(\langle \mathcal{V}(\omega), b_i(\omega)\rangle\right)^2 \triangleq \mathcal{L}[\mathcal{V}]. \tag{A.6}$$

Further, consider $\hat{\mathcal{V}}(\omega)$ that maximizes the likelihood function (see equation A.6):

$$\hat{\mathcal{V}} = \underset{\mathcal{V} \in \mathbb{W}(\omega_c)}{\arg \max}(\mathcal{L}[\mathcal{V}]). \tag{A.7}$$

Then the BLML estimator is

$$\hat{p}(x) = (\hat{v}(x))^2 = \left(\frac{1}{2\pi} \int \hat{\mathcal{V}}(\omega) e^{j\omega x} d\omega\right)^2. \tag{A.8}$$

Theorem 1 (Agarwal, Chen et al., 2015). *Consider n independent samples of an unknown band-limited pdf, $p(x)$, with assumed cutoff frequency f_c (unit: Hz). Then the BLML estimator of $p(x)$ is given as*

$$\hat{p}(x) = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}_i \, \frac{sin(\pi f_c(x - x_i))}{\pi(x - x_i)}\right)^2, \tag{A.9}$$

where $\hat{\boldsymbol{\alpha}} \triangleq [\hat{\alpha}_1, \cdots, \hat{\alpha}_n]^\top$ and

$$\hat{\boldsymbol{\alpha}} = \underset{\rho_n(\boldsymbol{\alpha})=0}{arg \, max}\left(\prod_{i=1}^{n} \frac{1}{\alpha_i^2}\right). \tag{A.10}$$

Here $\rho_{ni}(\boldsymbol{\alpha}) \triangleq \frac{1}{n} \sum_{k=1}^{n} \alpha_k \frac{sin(\pi f_c(x_i - x_k))}{\pi(x_i - x_k)} - \frac{1}{\alpha_i}$ ($\forall \, i, k = 1, \ldots, n$).

A.3 Choice of Optimal Orthant. Agarwal, Chen et al. (2015), showed that the asymptotic solution of equation A.10 lies in the orthant with indicator vector $\alpha_{0i} = 1 \; \forall i = 1, \ldots, n$ if $p(x)$ is band-limited and $p(x) > 0$ (i.e., strictly positive $\forall \, x \in \mathbb{R}$). Therefore, the optimal orthant vector can be chosen trivially (i.e., $\boldsymbol{\alpha}_0 = \pm[1, 1, \ldots, 1]^\top$), and then $\boldsymbol{\rho}_n(\boldsymbol{\alpha}) = 0$ is solved in that

orthant to compute $\hat{\alpha}$. It is important to note that the asymptotic proper-
ties start to manifest for sample sizes as far as $n = 100$, and the condition
$p(x) > 0$ is obeyed by many pdfs encountered in nature. Therefore, the two
conditions are very less restrictive.

A.4 BLMLQuick Algorithm. Consider a function $\bar{p}(x)$ such that

$$\bar{p}(x) = f_s \int_{x - \frac{0.5}{f_s}}^{x + \frac{0.5}{f_s}} p(\tau)d\tau, \tag{A.11}$$

where $p \in \mathbb{U}(\omega_c)$ and $f_s > 2f_c$ is the sampling frequency. It is easy to verify
that $\bar{p}(x)$ is also a pdf and $\bar{p} \in \mathbb{U}(\omega_c)$. Now consider samples $\bar{p}[\kappa] \triangleq \bar{p}(\kappa/f_s)$
(where κ denotes an integer scale). Clearly these samples are related to $p(x)$
as

$$\bar{p}[\kappa] = \int_{\frac{\kappa - 0.5}{f_s}}^{\frac{\kappa + 0.5}{f_s}} p(x)dx. \tag{A.12}$$

Further consider $\bar{x}_i$'s computed by binning from independent and identi-
cally distributed (i.i.d.) observations $\{x_i\}_{i=1}^n$ of random variable from $p(x)$,
as

$$\bar{x}_i = f_s \left\lfloor \frac{x_i}{f_s} + 0.5 \right\rfloor, \tag{A.13}$$

where $\lfloor \rfloor$ denotes the function that returns the greatest integer of the
input variable. Note that $\bar{x}_i$ are the i.i.d. observations sampled from
$\tilde{p}(x) = \sum_\kappa \bar{p}[\kappa]\delta(x - \frac{\kappa}{f_s})$. Now since $f_s > 2f_c$, the BLML estimate for $\tilde{p}(x)$
should converge to $\bar{p}(x)$ due to Nyquist's sampling theorem (Marks, 1991).
We called this BLML estimator BLMLQuick. Assuming that the rate of conver-
gence for BLML is $\mathcal{O}(n^{-1})$, then if f_s is chosen such that $||p - \bar{p}||_2^2 = \mathcal{O}(n^{-1})$,
the BLMLQuick will also converge with a rate of $\mathcal{O}(n^{-1})$. This happens at
$f_s = f_c n^{0.25} > f_c$ also $f_s > 2f_c$ if $n > 16$.

A.5 Implementation and Computational Complexity. First consider
the following theorem (Agarwal, Chen et al., 2015).

Theorem 2. *Consider n i.i.d observations $\{x_i\}_{i=1}^n$ of random variable x with pdf*
$p(x)$ that has a $\frac{1}{|x|^r}$ tail. Then $Pr(min(\{x_i\}_{i=1}^n) < -(\frac{n}{(r-1)\epsilon})^{\frac{1}{r-1}}) \simeq 1 - e^{-\epsilon} \simeq \epsilon$ for
large n.

Let $i = 1, \ldots, n$ denote the index for samples and $b = 1, \ldots, B$ denote the
index for bins. The implementation of the BLMLQuick algorithm consists of
the following steps:

1. Compute $\{\bar{x}_b, n_b\}_{b=1}^{B}$ from $\{x_i\}_{i=1}^{n}$. This step has a computational complexity of $\mathcal{O}(n)$.
2. Sort $\{\bar{x}_b, n_b\}_{b=1}^{B}$ and construct $\mathbf{S}$: $s_{ab} = s(\bar{x}_a - \bar{x}_b)$ ($\forall a, b = 1, \ldots, B$) and $\bar{\mathbf{S}} = \mathbf{S} \times \text{diag}(\{n_b\}_{b=1}^{B})$. Note that $\mathbf{S}$ is a block-Toeplitz matrix (Toeplitz arrangements of blocks and each block is Toeplitz). This step has a computational complexity of $\mathcal{O}(B^2)$ (Akaike, 1973).
3. Use convex optimization algorithms to solve $\boldsymbol{\rho}_n(\boldsymbol{\alpha}) = \mathbf{0}$. Newton's method should take a finite number of iterations to reach a given tolerance ϵ since the cost function is self-concordant (Boyd & Vandenverghe, 2004). Therefore, the computational complexity of optimization is the same as the computational complexity of one iteration. The complexity of one iteration is the same as the complexity of calculating

$$\left(\text{diag}\left(\{1/\alpha_b^2\}_{b=1}^{B}\right) + \mathbf{S} \times \text{diag}\left(\{n_b\}_{b=1}^{B}\right)\right)^{-1}$$
$$= \left(\text{diag}\left(\{1/(\alpha_b^2 n_b)\}_{b=1}^{B}\right) + \mathbf{S}\right)^{-1} \text{diag}\left(\{n_b\}_{b=1}^{B}\right)^{-1}. \qquad \text{(A.14)}$$

   As the matrix $\left(\text{diag}\left(\{1/(\alpha_b^2 n_b)\}_{b=1}^{B}\right) + \mathbf{S}\right)$ has a block-Toeplitz structure, the Akaike algorithm (Akaike, 1973) can be used to evaluate each iteration of Newton's method in $\mathcal{O}(B^2)$. Note that Simulations show that $\mathbf{S}$ can be approximated accurately (to machine accuracy) by a low-rank matrix (e.g., with a rank order $R = 20$ for $B = 1000$); therefore, the inversion can be performed in $\mathcal{O}(R^2 + RB)$. Further, in some cases, one may end up with a large B (e.g., if the true pdf has heavy tails) so that storing the Hessian matrix becomes expensive. In such cases, a quasi-Newton or gradient-descent optimization algorithm can be used that computes the `BLML` estimator fairly quickly.
4. Evaluate the `BLMLQuick` estimate $p(x) = (\frac{1}{n} \sum_{b=1}^{B} n_b \alpha_b s(x - x_b))^2$ at l given points. This step has a computational complexity of $\mathcal{O}(Bl)$.

In summary, the total computational complexity of `BLMLQuick` is $\mathcal{O}(n + B^2 + lB)$. Substituting $B \leq \mathcal{O}(n^{\frac{1}{r-1}}) f_s \leq \mathcal{O}(f_c n^{\frac{1}{r-1}+0.25})$ (from theorem 2), yields the total computational complexity $\mathcal{O}(n + f_c^2 n^{\frac{2}{r-1}+0.5} + f_c l n^{\frac{1}{r-1}+0.25})$, where $\frac{1}{|x|^r}$ describes the tail of the pdf (the gaussian pdf has an infinite value of r) and l is the number of points being evaluated.

Appendix B: Change of Variables

Let $X \in \mathbb{R}, Y \in \mathbb{R}, U \in \mathbb{R}$, and $Y = g(X)$, where $g(\cdot)$ is a one-to-one nonlinear function with inverse image $g^{-1}(\cdot)$.

When $g(\cdot)$ is an increasing function, the cumulative distribution function (cdf) of the transformed variable is

$$F_Y(y) = \Pr\{Y \leq y\} = \Pr\{g(X) \leq y\} = \Pr\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$
$$\text{(B.1)}$$

When $g(\cdot)$ is an increasing function, the cdf of the transformed variable is

$$F_Y(y) = \Pr\{Y \le y\} = \Pr\{g(X) \le y\} = \Pr\{X \ge g^{-1}(y)\} = 1 - F_X(g^{-1}(y)).$$
(B.2)

When equations B.1 and B.2 are combined, the pdf of the transformed variable is

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$
(B.3)

Considering the ratio between a conditional pdf and a marginal pdf, we have

$$\frac{p_Y(y|z)}{p_Y(y)} = \frac{p_X(g^{-1}(y)|z) \left| \frac{d}{dy} g^{-1}(y) \right|}{p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|}$$

$$= \frac{p_X(g^{-1}(y)|z)}{p_X(g^{-1}(y))} = \frac{p_X(x|z)}{p_X(x)},$$
(B.4)

where it is assumed that $p_Y(y) > 0$ and $p_X(x) > 0$. Similarly,

$$\frac{p_{UY}(u,y|z)}{p_{UY}(u,y)} = \frac{p_{UX}(u,g^{-1}(y)|z) \left| \frac{d}{dy} g^{-1}(y) \right|}{p_{UX}(u,g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|}$$

$$= \frac{p_{UX}(u,g^{-1}(y)|z)}{p_{UX}(u,g^{-1}(y))} = \frac{p_{UX}(u,x|z)}{p_{UX}(u,x)},$$
(B.5)

where $p_{UY}(u,y) > 0$, $p_{UX}(u,x) > 0$. Therefore, the density ratio is invariant to the one-to-one nonlinear transform.

Acknowledgments

References

Agarwal, G., Stevenson, I. H., Berényi, A., Mizuseki, K., Buzsáki, G., & Sommer, F. T. (2014). Spatially distributed local fields in the hippocampus encode rat position. *Science*, *344*(6184), 626–630.

Agarwal, R., Chen, Z., & Sarma, S. V. (2015). *Nonparametric estimation of band-limited probability density functions*. arXiv:1503.06236v1. http://arxiv.org/pdf/1503 .06236v1.pdf

Agarwal, R., Sacre, P., & Sarma, S. V. (2015). *Mutual dependence: A novel method for computing dependencies between random variables*. arXiv preprint arXiv:1506.00673.

Agarwal, R., & Sarma, S. V. (2010). Restoring the basal ganglia in Parkinson's disease to normal via multi-input phase-shifted deep brain stimulation. In *Proc. of the Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1539–1542). Piscataway, NJ: IEEE.

Agarwal, R., & Sarma, S. V. (2011). An analytical study of relay neuron's reliability: Dependence on input and model parameters. In *Proc. of the Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2426–2429). Piscataway, NJ: IEEE.

Agarwal, R., & Sarma, S. V. (2012a). The effects of DBS patterns on basal ganglia activity and thalamic relay. *Journal of Computational Neuroscience*, *33*(1), 151–167.

Agarwal, R., & Sarma, S. V. (2012b). Performance limitations of relay neurons. *PLoS Computational Biology*, *8*(8), e1002626.

Agarwal, R., Sarma, S. V., Thakor, N. V., Schieber, M. H., & Massaquoi, S. (2015). Sensorimotor gaussian fields integrate visual and motor information in premotor neurons. *Journal of Neuroscience*, *35*(25), 9508–9525.

Akaike, H. (1973). Block Toeplitz matrix inversion. *SIAM J. Appl. Math.*, *24*(2), 234–241.

Barbieri, R., Frank, L., Nguyen, D., Quirk, M., Solo, V., Wilson, M., & Brown, E. (2004). Dynamic analyses of information encoding in neural ensembles. *Neural Computation*, *16*(2004), 277–307.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., & Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, *105*(1), 25–37.

Boyd, S., & Vandenverghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.

Brown, E., Barbieri, R., Ventura, V., Kass, R., & Frank, L. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.*, *14*(2), 325–346.

Brown, E., Frank, L., Tang, D., Quirk, M., & Wilson, M. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*(18), 7411–7425.

Chen, Z. (2013). An overview of Bayesian methods for neural spike train analysis. *Computational Intelligence and Neuroscience*, *2013*, 251905.

Chen, Z. (Ed.) (2015). *Advanced state space methods in neural and clinical data*. Cambridge: Cambridge University Press.

Chen, Z., Kloosterman, F., Layton, S., & Wilson, M. A. (2012). Transductive neural decoding for unsorted neuronal spikes of rat hippocampus. In *Proc. of the Conference*

of the IEEE Engineering in Medicine and Biology Society (pp. 1310–1313). Piscataway, NJ: IEEE.

Coleman, Todd P., & Sarma, Sridevi V. (2010). A computationally efficient method for nonparametric modeling of neural spiking activity with point processes. *Neural Comput., 22*(8), 2002–2030.

Cox, D., & Isham, V. (2000). *Point processes*. London: Chapman and Hall/CRC.

Fenton, A. A., & Muller, R. U. (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proceedings of the National Academy of Sciences USA*, *95*(6), 3182–3187.

Frank, L. M., Stanley, G. B., & Brown, E. N. (2004). Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*, *24*(35), 7681–7689.

Gerhard, F., & Gerstner, W. (2010). Rescaling, thinning or complementing? On goodness-of-fit procedures for point process models and generalized linear models. In J. D. Lafferty, C. K. J. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Cullotta (Eds.), *Advances in neural information processing systems, 23*. Cambridge, MA: MIT Press.

Hasselmo, M. E. (2008). Grid cell mechanisms and function: Contributions of entorhinal persistent spiking and phase resetting. *Hippocampus*, *18*(12), 1213–1229.

Kahn, K., Sheiber, M., Thakor, N., & Sarma, S. V. (2011). Neuron selection for decoding dexterous finger movements. In *Proc. of the Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE.

Kass, R., & Ventura, V. (2001). A spike train probability model. *Neural Comput.*, *13*, 1713–1720.

Kloosterman, F., Layton, S., Chen, Z., & Wilson, M. (2014). Bayesian decoding using unsorted spikes in the rat hippocampus. *J. Neurophysiol.*, *111*(1), 217–227.

Kubie, J. L., & Fox, S. E. (2015). Do the spatial frequencies of grid cells mold the firing fields of place cells? *Proceedings of the National Academy of Sciences USA*, *112*(13), 3860–3861.

Lehmann, E. L., & Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation: Part I. *Sankhyā: The Indian Journal of Statistics*, *10*, 305–340.

Marks, I. R. (1991). *Introduction to shannon sampling and interpolation theory*. New York: Springer-Verlag.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, *16*(3), 285–292.

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, *31*, 69–89.

Muller, R. U., & Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, *7*(7), 1951–1968.

O'Keefe, J. (1979). A review of the hippocampal place cells. *Progress in Neurobiology*, *13*(4), 419–439.

Ormond, J., & McNaughton, B. L. (2015). Place field expansion after focal MEC inactivations is consistent with loss of Fourier components and path integrator gain reduction. *Proceedings of the National Academy of Sciences USA*, *112*(13), 4116–4121.

Pillow, J. W. (2009). Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems, 22* (pp. 1473–1481), Cambridge, MA: MIT Press.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature, 454*(7207), 995–999.

Rad, K. R., & Paninski, L. (2010). Efficient, adaptive estimation of two-dimensional firing rate surface via Gassign process methods. *Network, 21*(3–4), 142–168.

Santaniello, S., Montgomery, Jr. E. B., Gale, J. T., & Sarma, S. V. (2012). Non-stationary discharge patterns in motor cortex under subthalamic nucleus deep brain stimulation. *Front. Integr. Neurosci., 6*(35), 1–13.

Sarma, S., Cheng, M., Eden, U., Williams, Z., Brown, E., & Eskandar, E. (2012). The effects of cues on neurons in the basal ganglia in Parkinson's disease. *Front. Integr. Neurosci., 6*(40).

Sarma, S. V., Eden, U. T., Cheng, M. L., Williams, Z., Hu, R., Eskandar, E. N., & Brown, E. N. (2010). Using point process models to compare neural spiking activity in the sub-thalamic nucleus of Parkinson's patients and a healthy primate. *IEEE Trans. Biomed. Engr., 57*(6), 1297–1305.

Sharpee, T., Rust, N. C., & Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Computation, 16*, 223–250.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: CRC Press.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature, 492*(7427), 72–78.

Theis, L., Chagas, A. M., Arnstein, D., Schwarz, C., & Bethge, M. (2013). Beyond GLMs: A generative mixture modeling approach to neural system identification. *PLoS Computational Biology, 9*(11), e1003356.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology, 93*(2), 1074–1089.

Williamson, R. S., Sahani, M., & Pillow, J. W. (2015). The equivalence of information-theoretic and likelihood-based methods for neural dimensionality reduction. *PLoS Computational Biology, 11*(4), e1004141.

Yuret, D., & De La Maza, M. (1993). Dynamic hill climbing: Overcoming the limitations of optimization techniques. In *The Second Turkish Symposium on Artificial Intelligence and Neural Networks* (pp. 208–212). N.p.

Zhang, K., Ginzburg, I., McNaughton, B., & Sejnowski, T. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology, 79*, 1017–1044.