A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects

Wilson Truccolo,[1] **Uri T. Eden,**[2,3] **Matthew R. Fellows,**[1] **John P. Donoghue,**[1] **and Emery N. Brown**[2,3]

[1]*Neuroscience Department, Brown University, Providence, Rhode Island;* [2]*Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston; and* [3]*Division of Health Sciences and Technology, Harvard Medical School/Massachusetts Institute of Technology, Cambridge, Massachusetts*

Truccolo, Wilson, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93: 1074–1089, 2005. First published September 8, 2004; doi:10.1152/jn.00697.2004. Multiple factors simultaneously affect the spiking activity of individual neurons. Determining the effects and relative importance of these factors is a challenging problem in neurophysiology. We propose a statistical framework based on the point process likelihood function to relate a neuron's spiking probability to three typical covariates: the neuron's own spiking history, concurrent ensemble activity, and extrinsic covariates such as stimuli or behavior. The framework uses parametric models of the conditional intensity function to define a neuron's spiking probability in terms of the covariates. The discrete time likelihood function for point processes is used to carry out model fitting and model analysis. We show that, by modeling the logarithm of the conditional intensity function as a linear combination of functions of the covariates, the discrete time point process likelihood function is readily analyzed in the generalized linear model (GLM) framework. We illustrate our approach for both GLM and non-GLM likelihood functions using simulated data and multivariate single-unit activity data simultaneously recorded from the motor cortex of a monkey performing a visuomotor pursuit-tracking task. The point process framework provides a flexible, computationally efficient approach for maximum likelihood estimation, goodness-of-fit assessment, residual analysis, model selection, and neural decoding. The framework thus allows for the formulation and analysis of point process models of neural spiking activity that readily capture the simultaneous effects of multiple covariates and enables the assessment of their relative importance.

INTRODUCTION

Understanding what makes a neuron spike is a challenging problem, whose solution is critical for deciphering the nature of computation in single cells and neural ensembles. Multiple factors simultaneously affect spiking activity of single neurons and thus assessing the effects and relative importance of each factor creates the challenge. Neural activity is often studied in relation to 3 types of covariates. First, spiking activity is associated with extrinsic covariates such as sensory stimuli and behavior. For example, the spiking activity of neurons in the rat hippocampus is associated with the animal's position in its environment, the theta rhythm, theta phase precession, and the animal's running velocity (Frank et al. 2002; Mehta et al. 1997, 2000; O'Keefe and Dostrovsky 1971; O'Keefe and Recce

1993). Retinal neurons respond to light intensity and light contrast, and V1 neurons are influenced by the spatiotemporal structure outside their classic receptive fields (Knierim and Vanessen 1992; Sillito et al. 1995; Vinje and Gallant 2000). The spiking activity of neurons in the arm region of the primary motor cortex (MI) is strongly associated with several covariates of motor behavior such as hand position, velocity, acceleration, and generated forces (Ashe and Georgopoulos 1994; Fu et al. 1995; Scott 2003). Second, the current spiking activity of a neuron is also related to its past activity, reflecting biophysical properties such as refractoriness and rebound excitation or inhibition (Hille 2001; Keat et al. 2001; Wilson 1999).

Third, current capabilities to record the simultaneous activity of multiple single neurons (Csicsvari et al. 2003; Donoghue 2002; Nicolelis et al. 2003; Wilson and McNaughton 1993) make it possible to study the extent to which spiking activity in a given neuron is related to concurrent ensemble spiking activity (Grammont and Riehle 1999, 2003; Hatsopoulos et al. 1998, 2003; Jackson et al. 2003; Maynard et al. 1999; Sanes and Truccolo 2003). Therefore, a statistical modeling framework that allows the analysis of the simultaneous effects of extrinsic covariates, spiking history, and concurrent neural ensemble activity would be highly desirable.

Current studies investigating the relation between spiking activity and these 3 covariate types have used primarily linear (reverse correlation) or nonlinear regression methods (e.g., Ashe and Georgopoulos 1994; Fu et al. 1995; Luczak et al. 2004). Although these methods have played an important role in characterizing the spiking properties in many neural systems, 3 important shortcomings have not been fully addressed. First, neural spike trains form a sequence of discrete events or point process time series (Brillinger 1988). Standard linear or nonlinear regression methods are designed for the analysis of continuous-valued data and not point process observations. To model spike trains with conventional regression methods the data are frequently smoothed or binned, a preprocessing step that can alter their stochastic structure and, as a consequence, the inferences made from their analysis. Second, although it is accepted that extrinsic covariates, spiking history, and neural ensemble activity affect neural spiking, current approaches make separate assessments of these effects, thereby making it difficult to establish their relative importance. Third, model goodness-of-fit assessments as well as the analysis of neural

ensemble representation based on decoding should be carried out using methods appropriate for the point process nature of neural spike trains.

To address these issues, we present a point process likelihood framework to analyze the simultaneous effects and relative importance of spiking history, neural ensemble, and extrinsic covariates. We show that this likelihood analysis can be efficiently conducted by representing the logarithm of the point process conditional intensity function in terms of linear combinations of general functions of the covariates and then using the discrete time point process likelihood function to fit the model to spike train data in the generalized linear model (GLM) framework. Because the discrete time point process likelihood function is general, we also show how it may be used to relate covariates to neural spike trains in a non-GLM model. We illustrate the methods in the analysis of a simulated data example and an example in which multiple single neurons are recorded from MI in a monkey performing a visuomotor pursuit-tracking task.

METHODS

In this section we present the statistical theory underlying our approach. First, we define the conditional intensity function for a point process. Second, we present a discrete time approximation to the continuous time point process likelihood function, expressed in terms of the conditional intensity function. Third, we show that when the logarithm of the conditional intensity is a linear combination of functions of the covariates, the discrete time point process likelihood function is equivalent to the likelihood of a GLM under a Poisson distribution and log link function. Alternatively, if the point process is represented as a conditionally independent Bernoulli process and the probability of the events is modeled by a logistic function, then the likelihood function is equivalent to the likelihood of a GLM under a Bernoulli distribution and a logistic link function. Fourth, we present several forms of conditional intensity models for representing spiking history, neural ensemble, and extrinsic covariate effects. Finally, we define our approach to maximum likelihood estimation, goodness-of-fit assessment, model comparison, residuals analysis, and decoding from point process observations by combining the GLM framework with analysis methods for point processes.

A *point process* is a set of discrete events that occur in continuous time. For a neural spike train this would be the set of individual spike times. Given an observation interval (0, T), a sequence of J spike times $0 < u_1 < \cdots < u_j < \cdots < u_J \leq T$ constitutes a point process. Let $N(t)$ denote the number of spikes counted in the time interval (0, t] for $t \in (0, T)$. We define a single realization of the point process during the time interval (0, t] as $N_{0:t} = \{0 < u_1 < u_2 < \cdots < u_j \leq t \cap N(t) = j\}$ for $j \leq J$.

Conditional intensity function

A stochastic neural point process can be completely characterized by its conditional intensity function $\lambda(t \mid H(t))$ (Daley and Vere-Jones 2003), defined as

$$\lambda(t \mid H(t)) = \lim_{\Delta \to 0} \frac{P[N(t + \Delta) - N(t) = 1 \mid H(t)]}{\Delta} \quad (1)$$

where $P[\cdot \mid \cdot]$ is a conditional probability and $H(t)$ includes the neuron's spiking history up to time t and other relevant covariates. The conditional intensity is a strictly positive function that gives a history-dependent generalization of the rate function of a Poisson process. From *Eq. 1* we have that, for small Δ, $\lambda(t \mid H(t))\Delta$ gives approximately the neuron's spiking probability in the time interval (t,

$t + \Delta$]. Because defining the conditional intensity function completely defines the point process, to model the neural spike train in terms of a point process it suffices to define its conditional intensity function. We use parametric models to express the conditional intensity as a function of covariates of interest, therefore relating the neuron's spiking probability to the covariates. We use $\lambda(t \mid \theta, H(t))$ to denote the parametric representation of the conditional intensity function in *Eq. 1*, where θ denotes an unknown parameter to be estimated. The dimension of θ depends on the form of the model used to define the conditional intensity function.

A discrete time representation of the point process will facilitate the definition of the point process likelihood function and the construction of our estimation algorithms. To obtain this representation, we choose a large integer K and partition the observation interval (0, T) into K subintervals $(t_{k-1}, t_k]_{k=1}^K$ each of length $\Delta = TK^{-1}$. We choose large K so that there is at most one spike per subinterval. The discrete time versions of the continuous time variables are now denoted as $N_k = N(t_k)$, $N_{1:k} = N_{0:t_k}$, and $H_k = H(t_k)$. Because we chose large K, the differences $\Delta N_k = N_k - N_{k-1}$ define the spike train as a binary time series of zeros and ones. In discrete time, the parametric form of the conditional intensity function becomes $\lambda(t_k \mid \theta, H_k)$.

Point process likelihood and GLM framework

Because of its several optimality properties, we choose a likelihood approach (Pawitan 2001) for fitting and analyzing the parametric models of the conditional intensity function. As in all likelihood analyses, the likelihood function for a continuous time point process is formulated by deriving the joint probability density of the spike train, which is the joint probability density of the J spike times $0 < u_1 < u_2 < \cdots < u_J \leq T$ in (0, T). For any point process model satisfying *Eq. 1*, this probability density can be expressed in terms of the conditional intensity function (Daley and Vere-Jones 2003). Similarly, in the discrete time representation, this joint probability density can be expressed in terms of the joint probability mass function of the discretized spike train (see APPENDIX *Eqs. A1* and *A2*) and is expressed here as a product of conditionally independent Bernoulli events (Andersen et al. 1992; Berman and Turner 1992; Brillinger 1988; Brown et al. 2003)

$$P(N_{1:K} \mid \theta) = \prod_{k=1}^K [\lambda(t_k \mid \theta, H_k)\Delta]^{\Delta N_k}[1 - \lambda(t_k \mid \theta, H_k)\Delta]^{1 - \Delta N_k} + o(\Delta^J) \quad (2)$$

where the term $o(\Delta^J)$ relates to the probability of observing a spike train with 2 or more spikes in any subinterval $(t_{k-1}, t_k]$. From *Eqs. A3–A5* in the APPENDIX, it follows that *Eq. 2* can be reexpressed as

$$P(N_{1:K} \mid \theta) = \exp\left\{ \sum_{k=1}^K \log [\lambda(t_k \mid \theta, H_k)\Delta]\Delta N_k - \sum_{k=1}^K \lambda(t_k \mid \theta, H_k)\Delta \right\} + o(\Delta^J) \quad (3)$$

If we view *Eq. 3* as a function of θ, given the spike train observations $N_{1:K}$, then this probability mass function defines our discrete time point process likelihood function and we denote it as $L(\theta \mid H_K) = P(N_{1:K} \mid \theta)$. From *Eq. A6* in the APPENDIX, it can be seen that *Eq. 3* is a discrete time approximation to the joint probability density of a continuous time point process.

To develop a computationally tractable and efficient approach to estimating θ we note that for any subinterval $(t_{k-1}, t_k]$, the conditional intensity function is approximately constant so that, by *Eq. 3*, $P(\Delta N_k) = \exp\{\log [\lambda(t_k \mid \theta, H_k)]\Delta N_k - \lambda(t_k \mid \theta, H_k)\Delta\}$ is given by the Poisson probability mass function. Because Δ is small, this is equivalent to the Bernoulli probability $P(\Delta N_k) = [\lambda(t_k \mid \theta, H_k)\Delta]^{\Delta N_k} [1 - \lambda(t_k \mid \theta, H_k)\Delta]^{1 - \Delta N_k}$ in *Eq. 2*. If we now express the logarithm of the conditional intensity function as a linear combination of general functions of the covariates

$$\log \lambda(t_k \mid \theta, H_k) = \sum_{i=1}^{q} \theta_i g_i[v_i(t_k + \tau)] \tag{4}$$

where g_i is a general function of a covariate $v_i(t_k)$ at different time lags τ, and q is the dimension of the estimated parameter θ, then *Eq. 3* has the same form as the likelihood function for a GLM under a Poisson probability model and a log link function (see APPENDIX, *Eqs. A7–A8*). Thus, maximum likelihood estimation of model parameters and likelihood analyses can be carried out using the Poisson–GLM framework. Alternatively, if we extend the results in Brillinger (1988), we obtain

$$\log \{[1 - \lambda(t_k \mid \theta, H_k)\Delta]^{-1}[\lambda(t_k \mid \theta, H_k)\Delta]\} = \sum_{i=1}^{q} \theta_i g_i[v_i(t_k + \tau)] \tag{5}$$

then *Eq. 2* has the same form as the likelihood function for a GLM under a Bernoulli probability distribution and a logistic link function (*Eqs. A9* and *A10*). Thus, maximum likelihood estimation of model parameters and likelihood analyses can also be carried out using the Bernoulli–GLM framework (see also Kass and Ventura 2001). In other words, for Δ sufficiently small (i.e., at most one spike per time subinterval), likelihood analyses performed with either the Bernoulli or the Poisson model are equivalent. However, because we are interested in modeling the conditional intensity function directly, instead of the probability of events in our discrete time likelihoods, we used the Poisson–GLM framework in our analyses.

Therefore, we can take advantage of the computational efficiency and robustness of the GLM framework together with all of the analysis tools from the point process theory: goodness-of-fit based on the time rescaling theorem, residual analysis, model selection, and stochastic decoding based on point process observations. We refer to this combined framework as the *point process–GLM framework*. This framework covers a very large class of models because *Eq. 4* allows for general functions of the covariates and of interaction terms consisting of combinations of the covariates. An application of GLM analysis to spike train data, without the support of the derived relations between the point process and GLM likelihood functions, would remain purely heuristic in nature.

Finally, *Eqs. 2* and *3* are generally applicable discrete time approximations for the point process likelihood function. Thus, when a parametric model of the conditional intensity function cannot be expressed in terms of either *Eq. 4* or *Eq. 5*, the GLM framework may be replaced with standard algorithms for computing maximum likelihood estimates (Pawitan 2001).

Models for the conditional intensity function

We formulate specific models for the conditional intensity function that incorporate the effects of spiking history, ensemble, and extrinsic covariates. For the exposition in the remainder of this section, we extend our notation to include the neural ensemble activity. Consider an observation time interval $t \in (0, T]$ with corresponding sequences of J^c spike times $0 < u_1^c < \cdots < u_j^c < \cdots < u_{J^c}^c \le T$, for $c = 1, \ldots,$ C recorded neurons. Let $N_{1:K}^{1:C} = \cup_{c=1}^{C} N_{1:K}^c$ denote the sample path for the entire ensemble.

CONDITIONAL INTENSITY MODELS IN THE POINT PROCESS–GLM FRAMEWORK. The general form for the conditional intensity function we use to model a single cell's spiking activity is

$$\lambda(t_k \mid N_{1:k}^{1:C}, \mathbf{x}_{k+\tau}, \theta) = \lambda_I(t_k \mid N_{1:k}, \theta_I)\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E)\lambda_X(t_k \mid \mathbf{x}_{k+\tau}, \theta_X) \tag{6}$$

where $\theta = \{\theta_X, \theta_E, \theta_I\}$, $\lambda_I(t_k \mid N_{1:k}, \theta_I)$ is the component of the intensity function conditioned on the spiking history $N_{1:k}$ of the neuron whose intensity is being modeled, $\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E)$ is the component related to the ensemble history contribution, and $\lambda_X(t_k \mid$

$\mathbf{x}_{k+\tau}, \theta_X)$ is the component related to an extrinsic covariate $\mathbf{x}_{k+\tau}$, where τ is an integer time shift. Note that the term H_k, used in the previous section, is now replaced by more specific information according to the model.

We consider the following specific models for each of these 3 covariate types. We begin with a model incorporating the spiking history component.

The spiking history component is modeled as

$$\lambda_I(t_k \mid N_{1:k}, \theta_I) = \exp\left\{\gamma_0 + \sum_{n=1}^{Q} \gamma_n \Delta N_{k-n}\right\} \tag{7}$$

where Q is the order of the autoregressive process, γ_n represents the autoregressive coefficients, and γ_0 relates to a background level of activity. This model is henceforth referred to as the autoregressive spiking history model. We apply Akaike's standard information criterion (AIC, see *Eq. 16* below) to estimate the parameter Q. We expect this autoregressive spiking history model to capture mostly refractory effects, recovery periods, and oscillatory properties of the neuron.

The contributions from the ensemble are expressed in terms of a regression model of order R

$$\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E) = \exp\left\{\beta_0 + \sum_{c} \sum_{r=1}^{R} \beta_r^c \Delta N_{k-r}^c\right\} \tag{8}$$

where the first summation is over the ensemble of cells with the exception of the cell whose conditional intensity function is being modeled. Thus the above model contains $R \times (C - 1)$ parameters plus one additional parameter for the background level. Note that the coefficients in the ensemble model capture spike effects at 1-ms time resolution and in this way they may reflect lagged-synchrony between spikes of the modeled cell and other cells in the ensemble. Alternatively, to investigate ensemble effects at lower time precision, we consider the ensemble rates model

$$\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E) = \exp\left\{\beta_0 + \sum_{c} \sum_{r=1}^{R} \beta_r^c (N_{k-(r-1)W}^c - N_{k-rW}^c)\right\} \tag{9}$$

where the term $N_{k-(r-1)W}^c - N_{k-rW}^c$ is the spike count in a time window of length W covering the time interval $(t_{k-rW}, t_{k-(r-1)W}]$. The coefficients in this model may reflect spike covariances on slow time scales.

In our application to MI data, the extrinsic covariate $\mathbf{x}_{k+\tau}$ will specify the hand velocity. To model this component we employ a variation of the Moran and Schwartz (1999) model, henceforth referred to as the velocity model

$$\lambda_X(t_k \mid \mathbf{x}_{k+\tau}, \theta_X) = \exp\{\alpha_0 + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})]\} \tag{10}$$

where $|V_{k+\tau}|$ and $\phi_{k+\tau}$ are, respectively, the magnitude and angle of the 2-D hand velocity vector in polar coordinates at time $t_{k+\tau}$. In this model $\mathbf{x}_{k+\tau} = [|V_{k+\tau}|, \phi_{k+\tau}]'$. For illustration purposes, we have considered only a single, fixed-time shift τ in the above model. Based on previous results (Paninski et al. 2004) we set $\tau = 150$ ms. A much more generic model form including linear or nonlinear functions of covariates at many different time lags could be easily formulated.

The most complex conditional intensity function models we investigate are the autoregressive spiking history plus velocity and ensemble activity, and the autoregressive spiking history plus velocity and ensemble rates models. For the former, the full conditional intensity function model is given by

$$\lambda(t_k \mid N_{1:k}^{1:C}, \mathbf{x}_{k+\tau}, \theta) = \exp\{\mu + \sum_{n=1}^{Q} \gamma_n \Delta N_{k-n}$$

$$+ \sum_{c} \sum_{r=1}^{R} \beta_r^c \Delta N_{k-r}^c + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})]\} \quad (11)$$

where μ relates to the background activity.

It should be noticed that although these models are in the "generalized linear" model class, the relation between the conditional intensity function and spiking history, ensemble, and extrinsic covariates can be highly nonlinear. These models are linear only after the transformation of the natural parameter (here the conditional intensity function) by the log link function and only with respect to the model parameters being estimated. As seen in *Eq. 4*, general functions (e.g., quadratic, cubic, etc.) of the actual measured covariates can be used.

NON-GLM CONDITIONAL INTENSITY FUNCTION MODEL. To illustrate the generality of the proposed point process framework, we construct and analyze a non-GLM conditional intensity function model that also incorporates effects of spiking history, neural ensemble, and extrinsic covariates. Additionally, this example demonstrates a procedure for obtaining a conditional intensity function by first modeling the interspike interval (ISI) conditional probability density function. The conditional intensity is obtained from the ISI probability density model using the relation (Brown et al. 2003)

$$\lambda(t_k \mid \theta, H_k) = \frac{p(t_e \mid \theta, H_k)}{1 - \int_{u_{N_{k-1}}}^{t_k} p[t \mid \theta, H(t)]\mathrm{d}t} \quad (12)$$

where $t_e = t_k - u_{N_{k-1}}$ is the elapsed time since the most recent spike of the modeled cell before time t_k and $p(t_e \mid \theta, H_k)$ is the ISI probability density, specified here by the inhomogeneous inverse Gaussian (Barbieri et al. 2001). This probability density is given in *Eq. A11* in the APPENDIX. This density is specified by a time-varying scaling parameter $s(t_k \mid \cdot)$ that, in our application to MI spiking data, captures the velocity and ensemble rates covariate effects

$$s(t_k \mid \mathbf{x}_{t+\tau}, N_{1:k}^{1:C}, \theta_X, \theta_E) = \exp\left\{ \mu + \sum_{c} \sum_{r=1}^{R} \beta_r^c (N_{k-(r-1)W}^c - N_{k-rW}^c) \right.$$

$$\left. + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})] \right\} \quad (13)$$

and a location parameter ψ. The set of parameters defining the inhomogeneous inverse Gaussian density and therefore the conditional intensity function is denoted $\theta = \{\theta_X, \theta_E, \psi\}$. This model (*Eqs. 12, 13,* and *A11*) is henceforth referred to as the inhomogeneous inverse Gaussian plus velocity and ensemble rates model. The history dependence in this model extends back to the time of the previous, most recent spike.

Maximum likelihood parameter estimation

Maximum likelihood parameter estimates for the models in the point process–GLM framework were efficiently computed using the iterative reweighted least squares (IRLS) algorithm. This method is the standard choice for the maximum likelihood estimation of GLMs because of its computational simplicity, efficiency, and robustness. IRLS applies the Newton–Raphson method to the reweighted least squares problem (McCullagh and Nelder 1989). Given the conditional intensity model in *Eq. 4*, the log-likelihood function is strictly concave. Therefore, if the maximum log-likelihood exists, it is unique (Santner and Duffy 1989). Confidence intervals and *p*-values were obtained following standard

computations based on the observed Fisher information matrix (Pawitan 2001). Statistically nonsignificant parameters (e.g. $P \geq 0.001$) were set to zero for all of the models. In the non-GLM case, the inhomogeneous inverse Gaussian model was fit by direct maximization of the likelihood function using a quasi-Newton method (IMSL, C function *min_uncon_multivar*, from Visual Numerics, 2001). For the data sets used here, the most intensive computations involved operations on large matrices of size about $10^6 \times 200$. Algorithms were coded in C and run on dual-processor 3.9-GHz IBM machines, 2 GB of RAM memory. Standard GLM estimation using IRLS is also available in several statistical packages (S-Plus, SPSS, and Matlab Statistics toolbox).

Goodness-of-fit, point process residual analyses and model comparison

KOLMOGOROV–SMIRNOV (K-S) TEST ON TIME RESCALED ISIS. Before making an inference from a statistical model, it is crucial to assess the extent to which the model describes the data. Measuring quantitatively the agreement between a proposed model and a spike train data series is a more challenging problem than for models of continuous-valued processes. Standard distance measures applied in continuous data analyses, such as average sum of squared errors, are not designed for point process data. One alternative solution to this problem is to apply the time-rescaling theorem (Brown et al. 2002; Ogata 1988; Papangelou 1972) to transform point processes like spike trains into continuous measures appropriate for goodness-of-fit assessment. Once a conditional intensity function model has been fit to a spike train data series, we can compute rescaled times z_j from the estimated conditional intensity and from the spike times as

$$z_j = 1 - \exp\left\{ -\int_{u_j}^{u_{j+1}} \lambda(t \mid H(t), \hat{\theta})\mathrm{d}t \right\} \quad (14)$$

for $j = 1, \ldots, J - 1$, where $\hat{\theta}$ is the maximum likelihood estimate of the parameters. The z_j values will be independent uniformly distributed random variables on the interval $[0, 1]$ if and only if the conditional intensity function model corresponds to the true conditional intensity of the process. Because the transformation in *Eq. 14* is one to one, any statistical assessment that measures the agreement between the z_j values and a uniform distribution directly evaluates how well the original model agrees with the spike train data. To construct the K-S test, we order the z_j values from smallest to largest, denoting the ordered values as $z_{(j)}$, and then plot the values of the cumulative distribution function of the uniform density function defined as $b_j = (j - 1/2)/J$ for $j = 1, \ldots, J$ against the $z_{(j)}$. We term these plots *K-S plots*. If the model is correct, then the points should lie on a 45° line. Confidence bounds for the degree of agreement between a model and the data may be constructed using the distribution of the Kolmogorov–Smirnov statistic (Johnson and Kotz 1970). For moderate to large sample sizes the 95% confidence bounds are well approximated by $b_j \pm 1.36 \cdot J^{-1/2}$ (Johnson and Kotz 1970).

To assess how well a model performs in terms of the original ISIs ($ISI_j = u_j - u_{j-1}$), we relate the ISIs to the computed z_j values in the following manner. First, the empirical probability density of the z_j values is computed, and the ratio of the empirical to the expected (uniform) density is calculated for each bin in the density. Second, the ISI values in the data are rounded to integer milliseconds and collected into bins. For these ISIs, all the corresponding z_j values as well as the ratios of empirical to expected densities in the related bins are obtained. This correspondence between ISIs and z_j values is easily available from *Eq. 14*. Third, we compute the mean ratio R (i.e., the mean of all the ratios for this particular ISI value). A mean ratio $R > 1$ ($R < 1$) implies that there are more (less) rescaled ISIs of length z_j than expected and that the intensity is being underestimated (overestimated), on average, at this particular ISI value.

If the model is correct, the z_j values should be not only uniformly distributed, but also independent. Thus, even when the K-S statistic is small, we still need to show that the rescaled times are independent. Here we assess independence up to 2nd-order temporal correlations by computing the autocorrelation function of the transformed rescaled times. As a visualization aid, we plot z_{j+1} against z_j.

POINT PROCESS RESIDUAL ANALYSIS. A standard approach in goodness-of-fit analysis is to examine structure in the data that is not described by the model. For continuous valued data, this is done by analyzing the residual error (i.e., the difference between the true and predicted values). For point process data, a different definition of residuals is needed to relate the conditional intensity function to the observed spike train data. The point process residual (Andersen et al. 1992) over nonoverlapping moving time windows is defined as

$$M(t_k) = \sum_{i=k-B}^{k} \Delta N_i - \int_{t_{k-B}}^{t_k} \lambda(t \mid H(t), \hat{\theta}) \mathrm{d}t \qquad (15)$$

for $k - B \geq 1$. In the application to MI data, we will look for relations between the point process residual and motor covariates (e.g., speed or direction) by computing their cross-correlation functions. Existence of correlations would imply that there is some structure left in the residuals that is not captured by the conditional intensity function model.

MODEL SELECTION. An additional tool for comparing models comes from the statistical theory of model selection (Burnham and Anderson 2002). The idea consists of choosing the best models to approximate an underlying process generating the observed data, a process whose complexity can be potentially infinite dimensional. To achieve this goal, we adopt Akaike's standard information criterion (AIC) (Akaike 1973). This criterion also provides a way to rank different candidate models. The AIC was originally derived as an estimate of the expected relative Kullback–Leibler distance (Cover and Thomas 1991) between a distribution given by an approximating model and the distribution of the true underlying process generating the data. This criterion is formulated as

$$AIC(q) = -2 \log L(\hat{\theta} \mid H_k) + 2q \qquad (16)$$

where $L(\hat{\theta} \mid H_K)$ is the likelihood function, $L(\hat{\theta} \mid H_K) = P(N_{1:K} \mid \hat{\theta}, H_K)$; $\hat{\theta}$ is the maximum likelihood estimate of the model parameters $\hat{\theta}$; and q is the total number of parameters in the model. By this criterion, the best model is the one with the smallest AIC, implying that the approximate distance between this model and the "true process" generating the data is smallest. The AIC is frequently interpreted as a measure of the trade-off between how well the model fits the data and the number of parameters required to achieve that fit, or of the desired trade-off between bias and variance (Burnham and Anderson 2002). An equivalence between AIC and cross-validation for the purpose of model selection has been established (Stone 1977). AIC can be applied to both nested and nonnested models, and to models with different distributions in their stochastic component. We compute AIC values for different models to guide our model comparison. Specifically, we provide the difference between the AIC of all of the models with respect to the AIC of the best model. We also use the AIC to estimate the order of the autoregressive spiking history component in *Eq. 7*.

Neural decoding analysis by state estimation with point process observations

Beyond assessing the goodness-of-fit of a single cell model with respect to its individual spike train data, we also analyze the ability of the model, over the entire cell population, to decode an m-dimensional extrinsic covariate $\mathbf{x}_{k+\tau}$. Such decoding will use the spike times of the entire ensemble of cells and the corresponding conditional intensity

function for each of these cells. We thus perform a state estimation of $\mathbf{x}_k$ based on point process observations and thereby assess the ensemble coding properties of the cell population. The estimated extrinsic covariate will be given by the posterior mode after a Gaussian approximation to the Bayes–Chapman–Kolmogorov system (Eden et al. 2004).

For the particular type of hand kinematics data described above, we model $\mathbf{x}_k$ as a Gaussian autoregressive process of order 1, henceforth $AR(1)$, given by

$$\mathbf{x}_{k+\tau} = \mu_{\mathbf{x}} + \mathbf{F}\mathbf{x}_{k+\tau-1} + \varepsilon_{k+\tau} \qquad (17)$$

where $\mu_{\mathbf{x}}$ is an m-dimensional vector of mean parameters, $\mathbf{F}$ is an $m \times m$ state matrix, and ε_k is the noise term given by a zero mean m-dimensional white Gaussian vector with $m \times m$ covariance matrix $\mathbf{W}_\varepsilon$. The matrices $\mathbf{F}$ and $\mathbf{W}_\varepsilon$ are fitted by maximum likelihood.

The point process observation equation is expressed in terms of the modeled conditional intensity functions $\lambda^c(t_k \mid \cdot)$ for each of the C cells entering the decoding. As an example, for intensity functions conditioned on a motor covariate $\mathbf{x}_{k+\tau}$ and intrinsic spiking history $N_{1:k}^C$, we have the following recursive point process filter.

One step prediction

$$\mathbf{x}_{k+\tau|k+\tau-1} = \mu_{\mathbf{x}} + \mathbf{F}\mathbf{x}_{k+\tau-1|k+\tau-1} \qquad (18)$$

One-step prediction covariance

$$\mathbf{W}_{k+\tau|k+\tau-1} = \mathbf{F}\mathbf{W}_{k+\tau-1|k+\tau-1}\mathbf{F} + \mathbf{W}_\varepsilon \qquad (19)$$

Posterior covariance

$$\mathbf{W}_{k+\tau|k+\tau} = \left[\mathbf{W}_{k+\tau|k+\tau-1}^{-1} + \sum_{c=1}^{C} [\nabla \log \lambda^c (t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)] \right.$$
$$\lambda^c(t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)\Delta[\nabla \log \lambda^c (t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)]'$$
$$- \sum_{c=1}^{C} \nabla^2 \log \lambda^c (t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)$$
$$\left. [\Delta N_{1:k}^c - \lambda^c(t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)\Delta] \right]^{-1} \qquad (20)$$

Posterior mode

$$\mathbf{x}_{k+\tau|k+\tau} = \mathbf{x}_{k+\tau|k+\tau-1} + \mathbf{W}_{k+\tau|k+\tau}$$
$$\times \sum_{c=1}^{C} \nabla \log \lambda^c(t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c) [\Delta N_{1:k}^c - \lambda^c(t_k \mid N_{1:k}^c, \mathbf{x}_{k+\tau|k+\tau-1}, \hat{\theta}^c)\Delta]$$

$$(21)$$

The term $\nabla(\nabla^2)$ denotes the m-dimensional vector ($m \times m$ matrix) of first (second) partial derivatives with respect to $\mathbf{x}_{k+\tau}$, and $\mathbf{W}_{k+\tau|k+\tau}$ is the posterior covariance matrix of $\mathbf{x}_{k+\tau}$. Similarly, decoding equations based on other models of the conditional intensity function can be obtained. The derivation of the recursive point process filter is based on the well-established (Mendel 1995; Kitagawa and Gersh 1996) relation between the posterior probability density and the Chapman–Kolmogorov (one-step prediction) probability density, and on a Gaussian approximation of the posterior density (for details see Eden et al. 2004). The Gaussian approximation results from a 2nd-order Taylor expansion of the density and it is a standard first approach for approximating probability densities (Tanner 1996; Pawitan 2001). Nonetheless, the spiking activity enters into the computations in a very non-Gaussian way through the point process model instantiated by the conditional intensity function.

The amount of uncertainty in the algorithm about the true state of the decoded parameter is related to the matrix $\mathbf{W}_{k+\tau|k+\tau}$. Confidence

regions and coverage probability for the decoding can thus be obtained as follows. At time $k\Delta t$ an approximate 0.95 confidence region for the true covariate $\mathbf{x}_{k+\tau}$ may be constructed as

$$(\mathbf{x}_{k+\tau} - \mathbf{x}_{k+\tau|k+\tau})'\mathbf{W}_{k+\tau|k+\tau}^{-1}(\mathbf{x}_{k+\tau} - \mathbf{x}_{k+\tau|k+\tau}) \le \chi_{0.95}^2(m) \qquad (22)$$

for $k = 1, 2, \ldots, K$, where $\chi_{0.95}^2(m)$ gives the 0.95 quantile of the χ^2 distribution with degrees of freedom equal to the dimension m of the covariate. The coverage probability up to time t_k is given by s_k/k where s_k is the number of times the true covariate is within the confidence regions during the time interval $(0, k\Delta]$. In the decoding analysis we compute the mean of the coverage probability over the entire decoding period. A Monte Carlo simulation is employed to obtain the confidence intervals and coverage probability for covariate in polar coordinates. We first use the estimated posterior covariance matrix to generate 10^4 Gaussian-distributed samples centered at the current covariate estimates in Cartesian coordinates. Second, these random samples are converted to polar coordinates. Finally, the confidence intervals are then computed from the distribution of the random samples in polar coordinates.

RESULTS

The proposed point process framework is illustrated with 2 examples. The first one is applied to simulated neural spike data and the second to multiple single units simultaneously recorded from monkey primary motor cortex. For the discrete time representation of the neural point process we set $\Delta = 1$ ms.

Simulation study

The goal of the simulation study is 2-fold. First, we illustrate the main properties of the model in *Eq. 11* containing the autoregressive history, neural ensemble history, and motor covariate effects. Second, we demonstrate that the parameters of the simulated model are accurately recovered from relatively small spike data sets by maximum likelihood estimation implemented with the IRLS algorithm.

The conditional intensity functions of 6 neurons (A, B, C, D, E, F) were simulated using methods as described in Ogata (1981). The intensity of 5 of them (B–F) was given by the velocity model (*Eq. 10*); that is, the neurons were modeled as inhomogeneous Poisson processes with mean background spiking rates of 17, 16, 9, 8, and 7 Hz, respectively, and inhomogeneity introduced by the modulating hand velocity signal. Different velocity tuning functions were used for the set of cells. The hand velocity signal was sampled from actual hand trajectories performed by a monkey (see *Application to MI spiking data*, below). The conditional intensity function for neuron A was given by the autoregressive spiking history plus ensemble and velocity model (*Eq. 11*). The background mean firing rate of this neuron was set to 10 Hz. The autoregressive spiking history component contained 120 coefficients covering 120 ms of spiking history (see Fig. 2*B*). The autoregressive coefficients mimicked the effects of refractory–recovery periods and rebound excitation. From the ensemble of 5 neurons, only 2 contributed excitatory (neuron B) and inhibitory (neuron C) effects at 3 time lags (-1, -2, and -3 ms).

The simulation scheme worked as follows. Starting with the initial simulation time step, first the conditional intensity functions were updated and then, at the same time step, the spiking activities for all of the cells were simulated. The simulation then moved to the next time step. The conditional intensity

functions were updated based on the past intrinsic and ensemble spiking history (neuron A only) and on the current hand velocity state (all neurons).

The main features of the conditional intensity function model in *Eq. 11* can be observed in Fig. 1, where the simulated conditional intensity function of neuron A and its own spiking activity are plotted together with the activity of the other 5 neurons and the contribution of velocity signal. The simulated conditional intensity function clearly shows the dependence on spike history: after a spike, the intensity drops to almost zero and slowly recovers, reaching a period of higher than background spiking probability at about 20 ms after the spike. Fast excitatory and inhibitory effects follow the spikes of neurons B and C. Spiking history, neural ensemble, and velocity modulate each other's contributions in a multiplicative fashion.

From the simulated ensemble spike trains and from the velocity signal, we then estimated the conditional intensity function generating the spiking activity of neuron A. The data set entering the estimation algorithm thus consisted of 6 simulated spike trains, each 200 s long, and of the hand velocity time series in polar coordinates. The spike train for the modeled neuron A contained 2,867 spikes. The parametric model for the estimated conditional intensity function consisted of a background mean, 120 autoregressive coefficients, and 5 regressive coefficients for each of the other 5 neurons (i.e., 25 coefficients in total). Each set of 5 coefficients related to spiking activity at lags $-1, -2, \ldots, -5$ ms. The IRLS
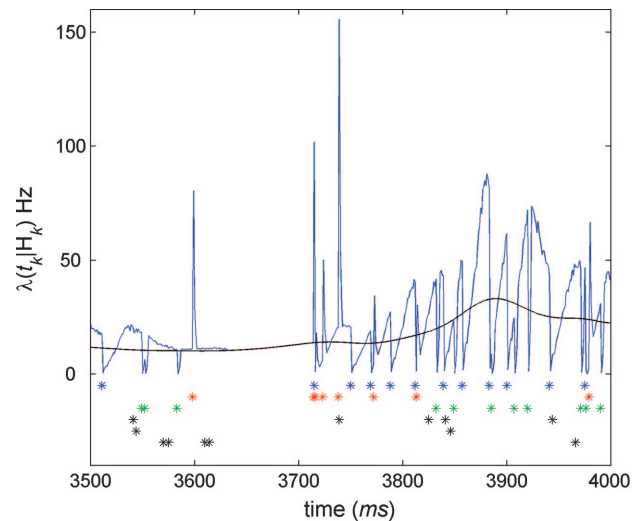

FIG. 1. Simulated conditional intensity function model. Conditional intensity function, modeled as in *Eq. 1*, was simulated and used to generate a spike train (neuron A, blue asterisks mark the times of the spike events). In this model, the intensity (blue curve) was conditioned on the past spiking history, the spikes of 2 other neurons (neuron B, excitatory, red asterisks; neuron C, inhibitory, green asterisks), and on hand velocity. Past spiking history effect was modeled by a 120-order autoregressive process carrying a refractory period, recovery, and rebound excitation. Coefficient values were based on parameters estimated from a primary motor cortex (MI) cell (see Fig. 5*B*). The conditional intensity function resulting from the contribution of only hand velocity is shown by the black line. Three other cells were also simulated (neurons D, E, and F; black asterisks). Neurons B–F were modeled as inhomogeneous Poisson processes modulated according to the velocity model (*Eq. 10*). All cells had different preferred movement directions. Spiking history, ensemble, and velocity modulated each other in a multiplicative fashion. Simulated ensemble spike trains together with hand velocity were used to estimate the parameters for the conditional intensity function model of neuron A (see Fig. 2).

algorithm converged in 12 iterations (tolerance was set to 10^{-6}). Statistically nonsignificant parameters ($P > 0.001$) were set to zero (see METHODS section). The true model parameters used in the simulation of neuron A were accurately recovered (Fig. 2, *B* and *C*), with the estimated model passing the K-S goodness-of-fit test (Fig. 2*D*). Parameter estimation on smaller data sets (about 50 s of data) led to similar successful fittings.

Application to MI spiking data

Experimental data were obtained from the MI area of a behaving monkey. Details of the basic recording hardware and protocols are available elsewhere (Donoghue et al. 1998; Maynard et al. 1999). After task training, a Bionic Technologies LLC (BTL, Salt Lake City, UT) 100-electrode silicon array was implanted in the area of MI corresponding to the arm representation. One monkey (*M. mulatta*) was operantly conditioned to track a smoothly and randomly moving visual target. The monkey viewed a computer monitor and gripped a 2-link, low-friction manipulandum that constrained hand movement to a horizontal plane. The hand (*x, y*) position signal was digitized and resampled to 1 kHz. Low-pass–filtered finite
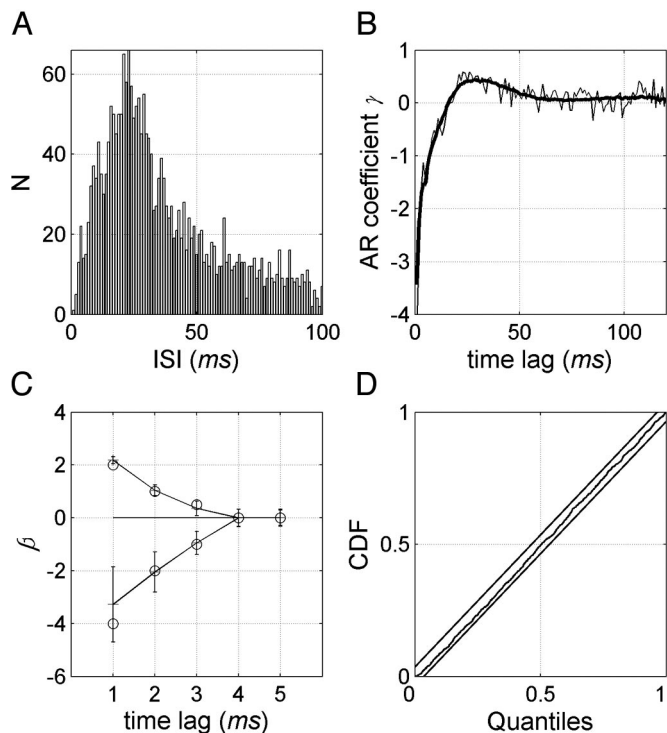

FIG. 2. Model parameter estimation by iteratively reweighted least squares (IRLS) in the generalized linear model (GLM) framework. Spike trains of the 6 simulated cells (each lasting 200 s; see Fig. 1) together with hand velocity produced the data set for the estimation of the parameters of the conditional intensity function model for neuron A. Spike train of neuron A consisted of 2,867 spikes. *A*: ISI distribution for neuron A. *B*: true autoregressive coefficients (thick curve) and the estimated ones. *C*: ensemble covariate in the estimated model contained 5 coefficients per cell (covering 5 ms of the past). Only the 3 coefficients for the excitatory and inhibitory cells were significantly different from zero ($P < 0.001$). Bars indicate the 95% confidence intervals. Small circles show the location of the true coefficients. *D*: Kolmogorov–Smirnov (K-S) goodness-of-fit test shows that the estimated conditional intensity model passed the test (the 95% confidence region is given by the parallel lines). Coefficients for the velocity covariate $\alpha_1 = 0.1$ and $\alpha_2 = -0.05$ were estimated as $\hat{\alpha}_1 = 0.105$ and $\hat{\alpha}_2 = -0.045$.

differences of position data were used to obtain hand velocities. Some 130 trials (8–9 s each) were recorded. More details about the statistical properties of the distributions for hand position and velocity, spiking sorting methods, and other task details can be found in Paninski et al. (2004).

Models including 1, 2, or all of the 3 types of covariates were analyzed. To start, we focus on the analysis of the velocity and the autoregressive spiking history plus velocity models. Later, we also compare these 2 models using neural decoding based on the observation of the entire ensemble of cells. For this reason, we analyzed these 2 models for each of the 20 cells in the ensemble. More detailed analysis involving K-S plots, point process residuals, and AIC model comparison will be illustrated for one typical cell.

K-S GOODNESS-OF-FIT ANALYSIS FOR THE VELOCITY AND THE AUTOREGRESSIVE SPIKING HISTORY PLUS VELOCITY MODELS. The tuning functions obtained from the velocity model (*Eq. 10*) are shown in Fig. 3. This model was statistically significant for all of the cells. Preferred direction was diverse across cells, covering the range of possible directions. The corresponding K-S plots are shown in Fig. 4. The quantiles refer to the $z_{(j)}$s (*Eq. 14*) and the cumulative distribution function (CDF) refers to the expected uniform distribution for the case when the estimated conditional intensity model was equivalent to the true one. The velocity model tends to overestimate (underestimate) the conditional intensity at lower (middle) quantiles. Introduction of the autoregressive spiking history component (*Eq. 7*) in the velocity model greatly improved the explanation of the spiking process, almost completely eliminating both the over- and underestimation of the intensity. The maximum order of the autoregressive component was about 120 (i.e., the component incorporated history effects spanning over 120 ms in the past). The most significant history effects extended to 60 ms in the past. For the majority of the cells, this component seemed to capture mostly 3 main history effects: refractory and recovery periods followed by an increase in the firing probability around 20 ms after a spike (see Fig. 5*B*). It should be noticed that the autoregressive coefficients could have also reflected dynamical network properties of nonmeasured neural ensembles such as networks of excitatory and inhibitory neurons where the modeled cell is embedded, or nonmeasured fast extrinsic covariates. No significant differences in the K-S plots were observed between a pure autoregressive history model and the autoregressive history plus velocity models (not shown).

Figure 5, *C* and *D* summarize the above observations for a typical cell (cell 75a, 29,971 spikes over 130 trials, i.e., ~1,040 s) in this example set and relate the fitting problems of the velocity model to the original nontime rescaled ISIs. In the velocity model, the intensity is overestimated (mean ratio $R < 1$) for ISIs below 10 ms and underestimated (mean ratio $R > 1$) for ISIs in the interval 10 to about 40 ms (Fig. 5*D*). The overestimation is likely a reflection of a refractory–recovery period (up to ~10 ms) after the cell has spiked, which is not captured by the velocity model. The underestimation reflects a period of increased firing probability that follows the recovery period. These 2 different regimes are reasonably well captured by the coefficients of the autoregressive component (see Fig. 5*B*), thus resulting in the improved fit observed for the autoregressive spiking history plus velocity model. Introduction of
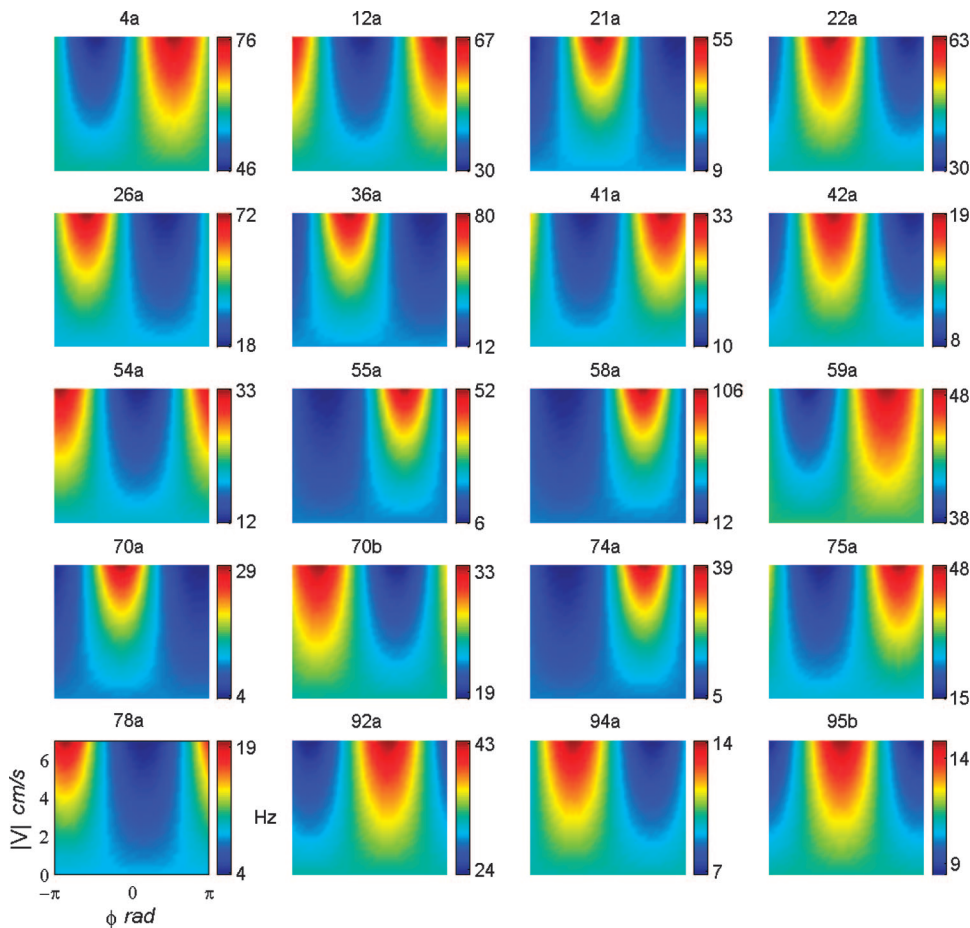
FIG. 3. Velocity tuning functions. Conditional intensity function values, based on the velocity model, are expressed by pseudo-color maps. Velocity is given in polar coordinates, with ϕ representing the movement direction. Each subplot relates to a particular cell, with cells' labels given at the *top*.

the autoregressive component makes the observed density for the z_j values much closer to the expected uniform density (Fig. 5*C*).

The K-S statistic measures how close rescaled times are to being uniformly distributed on [0, 1]. In addition, a good model should also generate independent and identically distributed rescaled times. To illustrate this point, we checked for temporal correlations at lag 1 in the time-rescaled ISIs (Fig. 6). As expected, some temporal structure remains in the case of the velocity model ($r^2 = 0.25$, $P = 10^{-6}$), whereas this structure is effectively insignificant for the velocity plus autoregressive spiking history ($r^2 = 0.002$, $P = 10^{-6}$). The cross-correlation function computed over a broad range of lags was consistent with this result.

POINT PROCESS RESIDUAL ANALYSIS.   Even though the parameters for the velocity model were statistically significant, the K-S plot analysis showed that the velocity model fell short of explaining the entire statistical structure in the observed single-cell spiking activity. It thus remains to be seen how well this model captured the relationship between hand velocity and spiking activity. Besides neural decoding, another approach to address this problem is to measure the correlations among the point process residuals as defined in *Eq. 15* and the movement velocity. Existence of correlations would imply that there is some structure left in the residuals that is not captured by the velocity model. On the other hand, a decrease in the correlation level with respect to some other model would imply that the

velocity model does capture some of the structure in the spiking activity related to hand velocity.

We computed the correlations for the residuals from the autoregressive spiking history model (*Eq. 7*) and compared them to the correlations for the residuals from the velocity and from the autoregressive spiking history plus velocity model. Residuals were computed for nonoverlapping 200-ms moving windows (Fig. 7). Cross-correlation functions were computed between the residuals and the mean of the kinematic variables. Mean (*x, y*) velocities were computed for each time window and were used to obtain, in polar coordinates, the respective mean movement speed and direction. In the autoregressive model case, peak cross-correlation values between the residuals and direction, speed, and velocities in *x* and *y* coordinates were 0.29, 0.10, −0.17, and 0.50, respectively. For the autoregressive spiking history and velocity model, the peak cross-correlation values for the same variables were 0.08, 0.06, −0.12, and 0.28. This suggests that, for this particular neuron, the velocity model captures a significant amount of information about hand velocity available in the spiking activity. Nonetheless, it is also clear that there is a residual structure in the spiking activity that is statistically related to the hand velocity in Cartesian coordinates and that is not captured by the autoregressive spiking history plus velocity model. Furthermore, the cross-correlation functions for both the velocity and the autoregressive spiking history plus velocity model show no significant differences, which suggests that the autoregressive
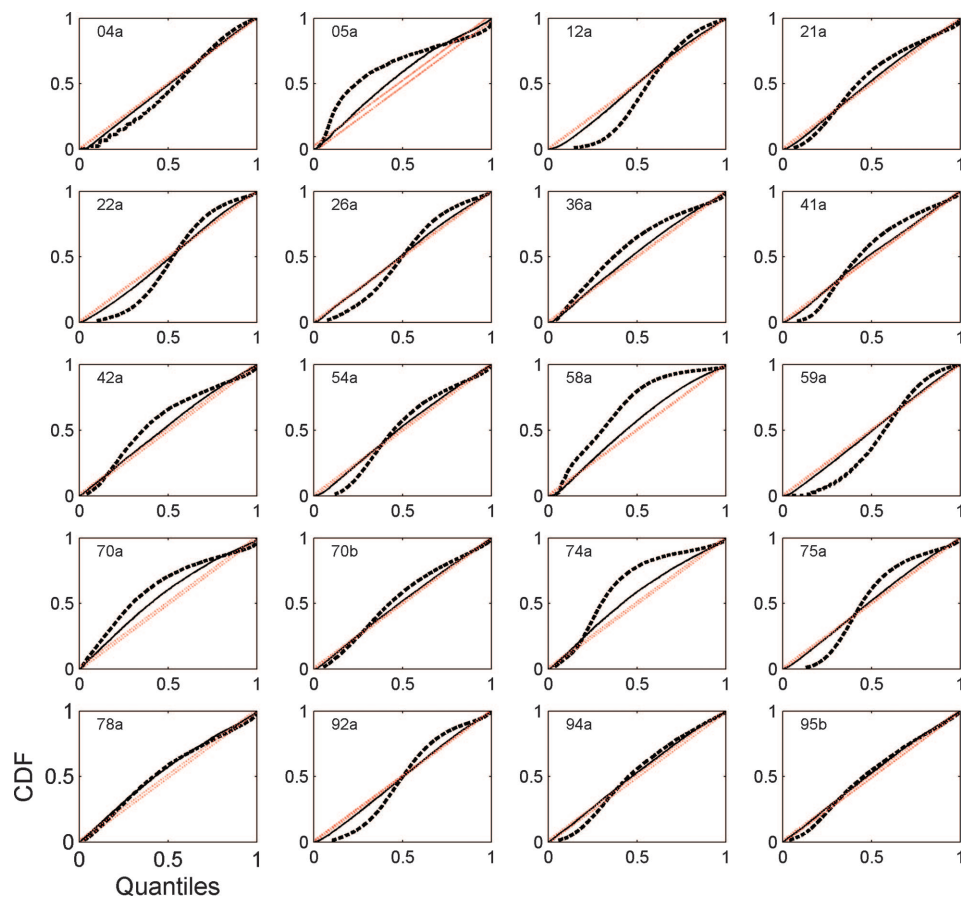
FIG. 4. K-S plots for the velocity model and the autoregressive spiking history plus velocity model. Because the K-S plots are constructed from a uniform distribution on the interval [0, 1], the 50th and 100th percentiles correspond, respectively, to quantiles 0.5 and 1.0 on both the horizontal and vertical axes. Two-sided 95% confidence error bounds of the K-S statistics are displayed for each cell (45° red lines). Visual inspection alone already reveals that, for most of the cells, the autoregressive spiking history plus velocity model (solid curve) improves the fit considerably.

component does not carry additional statistical information about hand velocity.

MODEL COMPARISON. We compared partial and complete (i.e., 1, 2, or 3 covariate types) conditional intensity function models for cell 75a. The ensemble model included spiking activity of each cell at 4 time lags (-1, -2, -3, and -4 ms). The ensemble rates included the spike counts of each cell at 3 nonoverlapping and lagged time windows. The length of each of the time windows, specified by the parameter W in *Eq. 9*, was 50 ms. The K-S plots in Fig. 8 reveal that the autoregressive spiking history plus velocity and the autoregressive spiking history plus velocity and ensemble rates models provided the best fits among all of the models for this specific cell. The inhomogeneous inverse Gaussian plus velocity and ensemble rates model performed better than the velocity, ensemble, and ensemble rates models. Inspection of the coefficients for the ensemble and ensemble rates models showed that the dependencies were statistically significant for many of the cells in the ensemble. Individual cells contributed either positive or negative effects to the conditional intensity function and the effective ensemble contribution to the modulation of the conditional intensity function could reach tens of hertz.

In the above K-S plot comparisons, some of the models had K-S statistics far from the 95% confidence intervals or nearly identical to those from other models, making a clear comparison difficult. The AIC analysis was then used to provide a more detailed comparison, as well as to take the complexity of the model (i.e., number of parameters) into consideration in the model comparison. Figure 9 shows the ranked models in terms

of their difference with respect to the AIC of the best model. In this context, models with lower AIC difference values are considered better models.

Overall, this criterion provided a fine model ranking and suggested that models containing the autoregressive spiking history component performed better in each instance. Among the alternative models for spiking history, the autoregressive spiking history model performed better than the conditional intensity model based on the inhomogeneous inverse Gaussian ISI distribution model (*Eqs. 12*, *13*, and *A11*), both in the AIC and K-S goodness-of-fit analyses. Also, the ensemble rates model did better than models containing only the velocity covariate or the ensemble covariate at fine temporal precision.

VELOCITY AND MOVEMENT DIRECTION DECODING ANALYSIS. The velocity (*Eq. 10*) and the autoregressive spiking history plus velocity models were used in the neural decoding of hand velocity. Models were fit to a training data set (120 trials, about 8–9 s each) and applied to decoding on a different test data set (10 trials, again about 8–9 s each). The state matrix $\mathbf{F}$ for the $AR(1)$ state process (*Eq. 17*) was estimated to be diagonal with nonzero terms approximately equal to 0.99, and the noise covariance matrix $\mathbf{W}_\varepsilon$ to be diagonal with nonzero entries equal to 0.01. Figure 10 shows the resulting decoding of movement direction and, in Cartesian coordinates, the estimated (x, y) velocities for a single test trial based on the velocity model. Overall, decoding of movement direction was remarkably good. Decoded (x, y) velocities captured mostly slower fluctuations. To compare the decoding performance of the 2 models, we computed the coverage probability and the
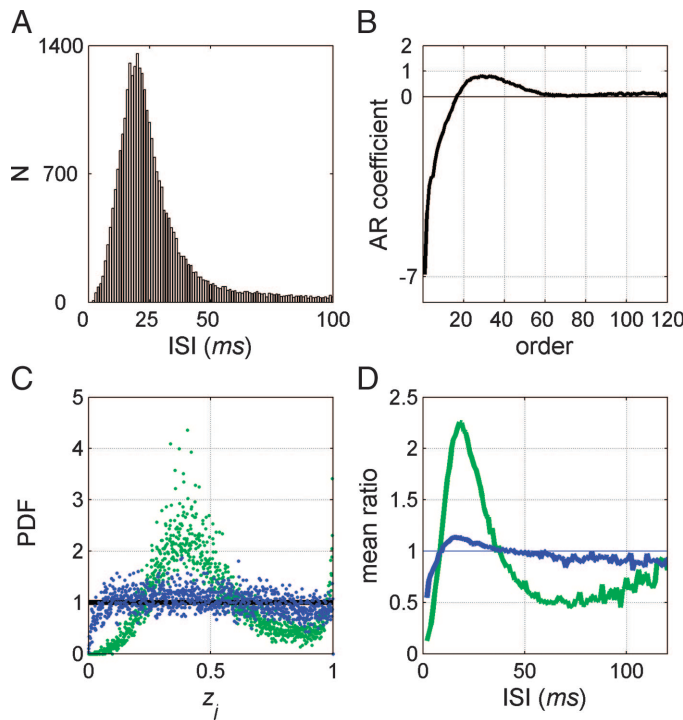
FIG. 5. Contribution of the autoregressive spiking history component. Cell 75a is chosen to illustrate how the addition of the autoregressive component improves the model's fit. *A*: ISI histogram. *B*: estimated coefficients of the autoregressive component. Autoregressive component incorporates a recovery period after the cell spikes, which lasts for about 18 ms (negative coefficients). Cell's firing probability then starts to increase, peaking at about 25 ms after a spike. *Order* refers to the order of the AR coefficient representing increasing times since the last spike. *C*: histogram for the transformed times z_j for both models (green: velocity model; blue: autoregressive spiking history plus velocity model). Black line shows the expected uniform distribution for the case where the estimated intensity function is close enough to the true intensity function underlying the neural point process. *D*: mean ratio of observed to expected z_j values indicates that the velocity model overestimates, on average, the intensity function for periods up to about 10 ms after a spike, while it tends to underestimate the intensity for periods between 10 and 40 ms. Introduction of the negative (positive) autoregressive coefficients almost completely eliminates the over (under) estimation of the conditional intensity function based on the velocity model alone.

decoding error. Table 1 gives the mean values (across time and trials) for the coverage probabilities of the bivariate estimate (velocity magnitude and movement direction) and the coverage probabilities of the univariate estimate (velocity magnitude or movement direction). Mean coverage probability for the movement direction estimate was 0.94 for the velocity model. For the same model, coverage probabilities for the bivariate estimate and velocity magnitude were much smaller, consistent with the observation that the estimated velocities in Cartesian coordinates captured mostly slow fluctuations. Mean coverage probability, mean and median decoding errors, and confidence intervals for the decoding errors were not significantly different between the 2 models.

DISCUSSION

An important problem in neurophysiology is determining the factors that affect a neuron's spiking behavior. To address this question we have presented a point process statistical framework that allowed us to characterize simultaneously the effects of several covariates on the spiking activity of an individual
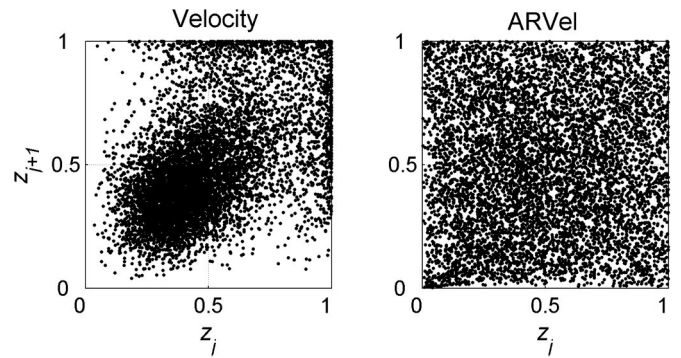

FIG. 6. Temporal correlations in the time-rescaled ISIs. Scatter plots are shown for consecutive z_j values from the velocity and autoregressive spiking history plus velocity (ARVel) models applied to cell 75a. Clearly, the autoregressive spiking history plus velocity model presents a more independent rescaled distribution. Corresponding correlation coefficients are 0.25 ($P = 10^{-6}$) for the velocity model and 0.002 ($P = 10^{-6}$) for the autoregressive spiking history plus velocity model. Cross-correlation functions computed over a broad range of lags led to similar results. Thus, in addition to improving the fit in the K-S plots, the introduction of the autoregressive component also eliminates temporal correlations among the rescaled times observed for the velocity model.

neuron. The 3 types of covariates we considered were the neuron's spiking history, past neural ensemble activity, and extrinsic covariates such as stimuli or behavior. Because de-
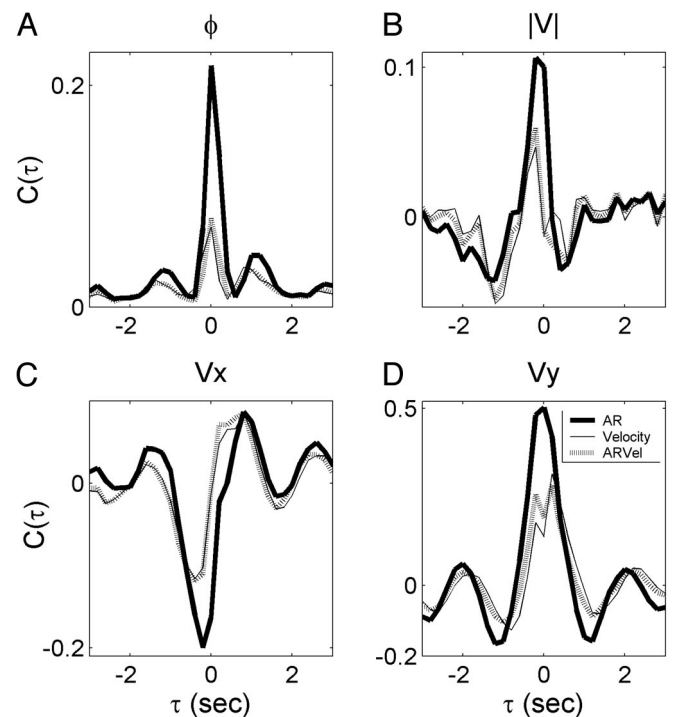

FIG. 7. Point process residual analysis. *A*: cross-correlation function $C(\tau)$ between the hand-movement direction and the residuals from the autoregressive spiking history (AR, thick curve), the velocity (thin curve), and the autoregressive spiking history plus velocity (ARVel, dashed curve) models applied to cell 75a. *B–D*: cross-correlations functions between the residuals and speed, and velocity in Cartesian coordinates (V_x and V_y). Correlations are significantly reduced for the velocity model in comparison to the autoregressive spiking history model. Nonetheless, there remains some structure in the point process residual that is related to the hand velocity but was not captured by the velocity model. Correlations for the velocity model were practically identical to the autoregressive spiking history plus velocity model, suggesting that the autoregressive component does not provide additional information about velocity (see text for details).
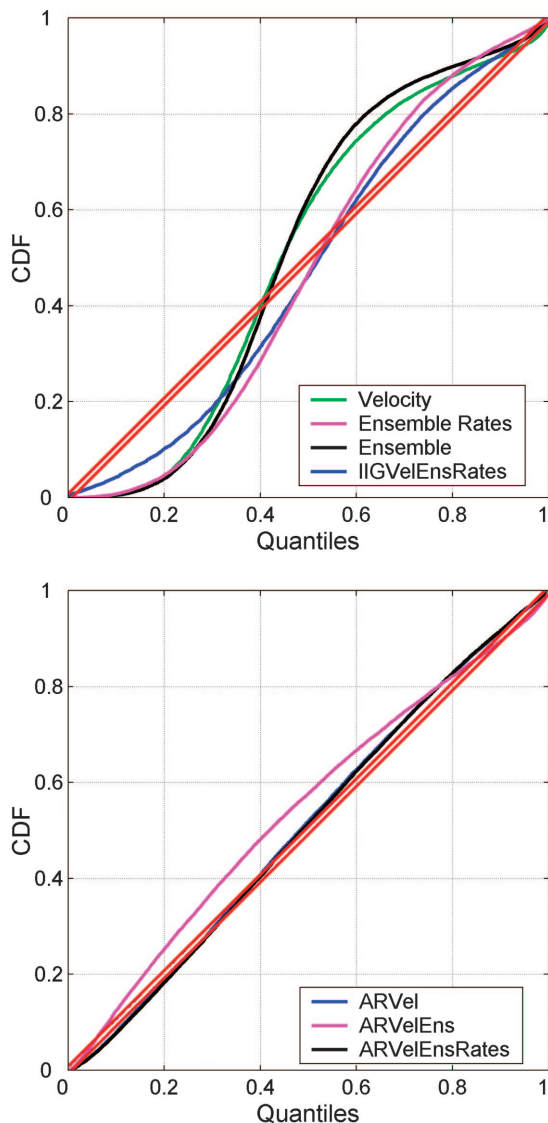
FIG. 8. Goodness-of-fit assessment (K-S plots) of alternative models. For comparison purposes, the models are shown in 2 groups. *Top*: the velocity, ensemble, ensemble rates, and the inhomogeneous inverse Gaussian plus velocity and ensemble rates models (IIGVelEnsRates) are compared. *Bottom*: the autoregressive spiking history plus velocity model (ARVel) is compared to 2 other models that add the ensemble (ARVelEns) or the ensemble rates component (ARVelEnsRates). K-S plots for the ARVel and the ARVelEnsRates partially overlap. (Cell 75a).

fining the conditional intensity function defines a point process model, the explanatory contributions of the covariates were assessed by constructing conditional intensity function models in terms of these covariates and by using a likelihood-based estimation approach to fit these models to data and to assess goodness-of-fit.

Analyses that measure the simultaneous effects of different covariates are crucial because the covariates modulate the neural spiking activity at the same time. Currently used analysis methods do not allow for the modeling of the simultaneous effects of spiking history, neural ensemble, and extrinsic covariates on the spike train treated as a point process. To evaluate the relation between spiking activity and covariates, the spike train data are frequently transformed to a rate function and the relation between the rate function and the covariate

is then assessed using regression methods (e.g., Ashe and Georgopoulos 1994; Luczak et al. 2004). The relation between a neuron's spiking activity, spiking history, and concurrent neural activity are usually assessed using autocorrelation and pairwise cross-correlation analyses performed directly on the spike train (e.g., Hatsopoulos et al. 1998). The use of different methods to assess individually the importance of these covariates precludes an analysis of the neural point process in which the relative importance of all covariates is assessed and may also lead to a misleading estimate of the covariate effects. For example, spiking history effects can interfere with the accurate estimation of extrinsic covariate effects in spike triggered averages and reverse correlation methods (Aguera y Arcas and Fairhall 2003; Aguera y Arcas et al. 2003). Additionally, current analysis techniques also assess the contribution of ensemble covariates separately. For instance, pairwise cross-correlation analyses measure the statistical association of a single neuron to each member of the ensemble separately but not the association between the single neuron's activity and the entire observed ensemble.

The key to our likelihood approach is representing the conditional intensity function of a single neuron in terms of the covariates. In this way, the covariates are directly related to the probability that the neuron spikes. Although this formulation can be used generally to analyze the relation between covariates and neural spiking activity, it usually requires writing a new algorithm or function to carry out the maximum likelihood
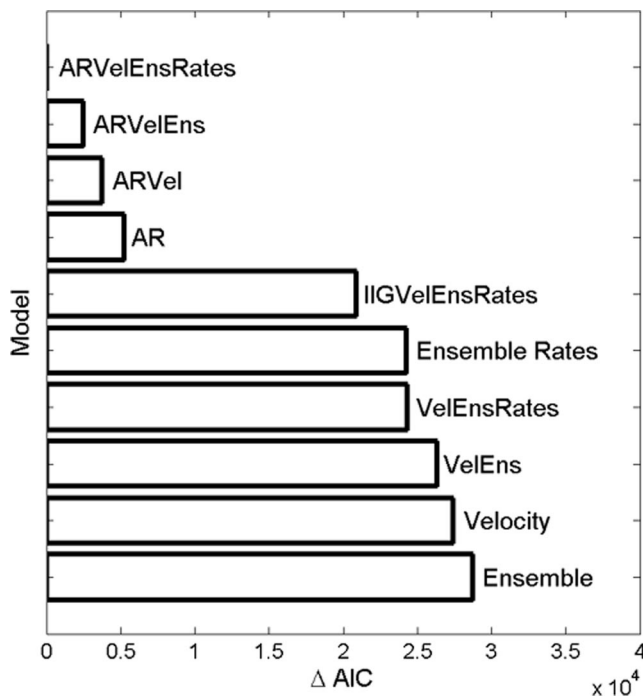

FIG. 9. Akaike's standard information criterion (AIC) model comparison. For convenience, we plot the differences of the AICs, denoted by ΔAIC of all of the models with respect to the AIC of the best model. Following this criterion and convention, better models have smaller AIC differences. Model labels: autoregressive spiking history (AR), autoregressive spiking history plus velocity (ARVel), autoregressive spiking history plus velocity and ensemble (ARVelEns), autoregressive spiking history plus velocity and ensemble rates (ARVelEnsRates), inhomogeneous inverse Gaussian plus velocity and ensemble rates models (IIGVelEnsRates), velocity plus ensemble (VelEns), and velocity plus ensemble rates (VelEnsRates). See METHODS section for model details. (Cell 75a).
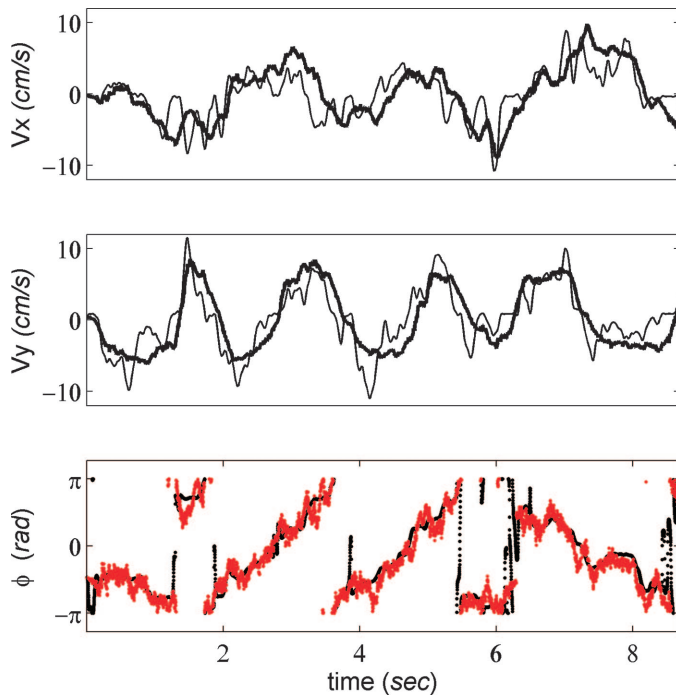
FIG. 10. Neural decoding of (x, y) velocities and movement direction by the point process filter. Estimated velocities (thick curve) and direction (red dots), together with true velocities and direction, are shown for a single decoded test trial. Time was discretized at a 1-ms resolution. At every millisecond, an estimate of the velocity parameters was obtained based on the state of the cells in the ensemble. All 20 recorded cells were used in the decoding. Conditional intensity function for each cell was given by the velocity model. Original decoding was done in polar coordinates. From a total of 130 trials, 120 trials were used for model fitting and 10 test trials for neural decoding. See Table 1 for summary statistics over the entire ensemble of test trials.

estimation with each new model formulation. To make our point process likelihood approach more broadly applicable, we showed that by representing the logarithm of the conditional intensity function as a linear combination of general functions of the covariates, the conditional intensity function model could be efficiently fit to neural spike train data using the GLM framework under a Poisson distribution and a log link function (*Eqs. 3*, *A6*, and *A8*). We also showed that, equivalently, if the point process is represented as a conditionally dependent Bernoulli process (*Eqs. 2* and *A2*) and the probability of the events is modeled by a logistic function, then the point process model could also be fit using the GLM framework under a Bernoulli distribution and a logistic link function (*Eqs. 5* and *A10*). Regarding the time discretization, we note that the 1-ms discretization interval we chose should not be interpreted as the absolute discretization interval for all applications of the point process–GLM framework. How fine the discretization interval should be will depend on the particular problem under study.

Our use of the Poisson distribution to fit point process models does not mean that we are assuming that our original data are Poisson. We are rather exploiting for computational purposes the fact that all point process likelihoods that admit a conditional intensity function, including those that are history dependent, have the same mathematical form given by *Eq. A6*. Our analysis makes explicit the relation between likelihood methods for point processes, conditional Bernoulli processes, and GLM model fitting with Poisson or Bernoulli distributions.

This relation provides a justification for using this GLM framework to analyze spike trains as point process data.

The point process–GLM framework for analyzing neural spike train data has several important advantages. First, this framework allows us to formulate complex models to relate the spiking activity to covariates. The fact that *Eq. 4* is written in terms of general functions of the covariates provides the framework with a very large class of possible models. Second, the GLM framework is part of several standard mathematical and statistical packages (e.g., Splus, Matlab, SPSS), so that the approach is readily accessible to experimentalists analyzing their data. Even though likelihood-based methods are highly desirable because of several optimality properties, the biggest impediment to their widespread use is the lack of readily available software in which a flexible class of neural spike train models can be easily applied to data. The point process–GLM framework offers a practical, broadly applicable solution to the computational problem of fitting potentially complex point process models for neural spike trains by maximum likelihood.

Third, the point process–GLM framework makes it possible to apply a set of goodness-of-fit tools for point processes not available in the GLM. These are the point process residuals analysis, goodness-of-fit tests based on the time-rescaling theorem, and decoding from point process observations. Fourth, the model selection and goodness-of-fit methods available in GLM are extended to spike train data. Thus we have a set of complementary methods to assess the extent to which proposed models explain the structure in neural spike trains. Although we used the point process–GLM framework to carry out most of the analysis, we also illustrated with the conditional intensity function based on the inhomogeneous inverse Gaussian model how non-GLM point process likelihood models may be used to analyze neural spike train data. Finally, by analogy with the way in which linear regression methods are used to analyze the relation between a continuous dependent variable and a set of candidate explanatory variables, the ready availability of software to implement GLMs also makes it possible for neurophysiologists to quickly assess the relevance of a wide range of covariates before proceeding to construct more specific models that may require non-GLM algorithms to carry out the model fitting.

A key objective of the proposed framework is to provide tools for the assessment of the relative importance of the covariates on neural spiking activity. We showed how this objective is accomplished by analyzing the goodness-of-fit of

TABLE 1. *Mean coverage probabilities, mean, and median errors for the velocity, and the autoregressive spiking history plus velocity models*

	Velocity	(AR) History + Velocity
$(\lvert V \rvert, \phi)$: Mean cov. prob.	0.30 ± 0.10	0.24 ± 0.08
$\lvert V \rvert$: Mean cov. prob.	0.32 ± 0.10	0.27 ± 0.08
ϕ: Mean cov. prob.	0.94 ± 0.02	0.90 ± 0.04
(V_x, V_y): Mean error	3.60 ± 0.3	3.60 ± 0.4
(V_x, V_y): Median error	3.43	3.38
(V_x, V_y): Error (95% CI)	$(0, 7.8]$	$(0, 7.8]$
ϕ: Mean error	$0.15\pi \pm 0.1$	$0.14\pi \pm 0.1$
ϕ: Median error	0.11π	0.11π
ϕ: Error (95% CI)	$(0, 0.65\pi]$	$(0, 0.63\pi]$

CI denotes the confidence interval. Velocity is given in cm/s.

each model we proposed. Because no method provides a complete assessment of goodness-of-fit, we used the standard statistical approach of using multiple complementary measures that evaluate different aspects of the model's agreement with the data. We applied 4 complementary types of goodness-of-fit techniques in our example data set: Kolmogorov–Smirnov tests based on the time-rescaling theorem, model selection by AIC, point process residual analysis, and neural spike train decoding based on point process observations.

First, the K-S analyses performed on the example data problems were useful in providing a sense of how close our models were to capturing the stochastic structure in the example data sets. The fact that some of our models captured most of the statistical structure in the spiking activity (Figs. 4 and 8) suggests that developing a parsimonious statistical model of MI activity is a realistic goal. The K-S plots also highlighted the importance of the spiking history. The fact that the coefficients had a structure that seemed to reflect mostly refractory and recovery periods together with rebound excitation in short time scales (effectively shorter than 60 ms) suggests that the autoregressive component successfully captured important history effects.

Second, the AIC analysis provided additional information for model comparison when the K-S plots did not distinguish between different models. By finely ranking the different models, the AIC analysis allowed for the assessment of the distinct effects and relative importance of the 3 types of covariates. Our choice of the AIC for model comparison was motivated by the fact that AIC was derived to be an estimate of the expected relative Kullback–Leibler distance between the distributions generated by the model and the distribution of the underlying stochastic process generating the data. The "true" underlying model does not need to be in the set of tested models for the AIC analysis to suggest the most appropriate model, a situation that we believe is more often the rule than the exception in biology. Because the AIC is penalized with increasing numbers of model parameters, its use is more appropriate than the use of the data likelihood itself in preventing overfitting of the data by the model. Additionally, an equivalence between AIC and cross-validation in model comparison problems has been established (Stone 1977), and model-based penalty methods (e.g., AIC) outperform cross-validation in important aspects when assessing prediction performance (Efron 2004). We also computed the Bayesian information criterion (BIC) (Schwarz 1978), an alternate, more conservative criterion function for each of our models and found that it yielded the same model rankings as did the AIC. When comparing complex models, another protection against overfitting comes from having large quantities of data as compared to the number of model parameters. In our example the $>10^6$ data observations far outnumbered the maximum 200 or so parameters in our most complex models. Apparent inconsistencies between AIC and K-S analyses could occur in cases where, for example, the conditional intensity model is more likely to produce the observed spike train as a whole, but is less accurate in describing a specific aspect of the data structure, such as the regime of small ISIs. This might have been the case when comparing the AIC and K-S plots results for the autoregressive spiking history plus velocity and ensemble model and a simpler autoregressive spiking history plus velocity model (Figs. 8 and 9).

Third, we illustrated how the point process residual analysis can be used to assess the contribution of an extrinsic covariate to a single neuron's spiking activity. In the illustration example, the residual analysis showed that the introduction of the velocity covariate captured a significant amount of the statistical structure related to hand velocity available in the spiking activity of a single neuron. Yet, for the particular cell chosen in this study, the analysis was also able to reveal that there still was a significant amount of structure in the residuals that was correlated to hand velocity but that was not captured by this specific form of the velocity model. Cross-correlation analysis of the point process residuals and extrinsic covariates is thus an important tool for assessing whether a particular model has captured well the effects of the covariate on the spiking activity. The ideal model should produce a residual with no significant correlations to the modeled covariate. It should also be noted that, unlike the decoding analysis, the point process residual analysis is not dependent on the properties of a decoding algorithm.

Fourth, complementing the above 3 goodness-of-fit analysis tools, the spike train decoding allowed for the goodness-of-fit assessment at the neural ensemble level. In conjunction with understanding what makes a neuron spike, we are also interested in assessing how well a model captures the representation of an extrinsic covariate at the ensemble level. At present, decoding is the only technique we have for assessing goodness-of-fit at this level. The key elements for assessing goodness-of-fit in the decoding analysis were the predicted signal and its confidence intervals and coverage probability, and especially the estimation error and its confidence intervals. The confidence intervals and the coverage probability based on the estimated posterior covariance matrix provided an estimate of the amount of uncertainty in the decoding algorithm, whereas the decoding error and its distribution provided a measure of the algorithm's actual performance. For this reason, the mean coverage probability should be interpreted in conjunction with the mean decoding error. As suggested by the narrow distribution of the decoding error and approximately 0.95 mean coverage probability (Table 1), hand movement direction was remarkably well decoded. Velocity estimates in Cartesian coordinates captured reasonably well slow fluctuations in the measured hand velocity. We also illustrated how to assess the contribution of the autoregressive spiking history component to neural decoding. The autoregressive spiking history plus velocity and the velocity models performed similarly well. This preliminary result suggests that short time dynamics captured by the autoregressive component did not play a crucial role in decoding hand velocity or movement direction in this data set. Given that the models used in the decoding analysis did not include the ensemble or ensemble rates covariate, the ensemble decoding assumed independent encoder cells. Nonetheless, the framework also allows the assessment of the contribution of interaction patterns in the neural ensemble to decoding. That could be easily achieved by extending the conditional intensity models to include the ensemble covariates. This analysis is beyond the scope of this paper and will be addressed elsewhere.

In summary, the above 4 complementary goodness-of-fit and model selection analyses are an essential step for achieving our primary objective of assessing the effects and relative importance of the modeled covariates. The proposed point process

framework provides a starting point for building and analyzing complex models of spiking activity including the 3 types of covariates discussed in this paper. In particular, the point process–GLM framework provides a systematic approach to neural spike train data analysis analogous to that for continuous-valued variables under the linear Gaussian regression framework.

We foresee several potential improvements and extensions of this framework. Although we were able to fit models with hundreds of parameters, much larger models will require the development of efficient algorithms for both GLM and non-GLM computations. Also, to analyze the relation between the simultaneous firing of an entire ensemble relative to its spiking history and a set of extrinsic covariates, we can extend the framework by using multivariate point process likelihoods (Chornoboy et al. 1988). A multivariate likelihood model will facilitate the study of the independence, redundancy, and synergy in the ensemble representation. Finally, multielectrode devices (Csicsvari et al. 2003; Donoghue 2002; Nicolelis et al. 2003) now make possible the simultaneous recordings of multiple single cells from many different brain areas. We foresee the proposed framework as a valuable tool for investigating how interacting brain regions represent, compute, and process information. We are currently applying this framework to the analysis of parietal and MI spiking activity in monkeys performing visuomotor tracking tasks and hippocampus activity in rats performing a range of learning tasks.

APPENDIX

Continuous and discrete time point process likelihood function

The likelihood of a neural spike train, like that of any statistical model, is defined by finding the joint probability density of the data. We show below that the joint probability of any point process is easy to derive from the conditional intensity function. We show that the point process likelihood function in *Eqs. 2* and *3* gives a discrete time approximation of the likelihood function for a continuous time point process (*Eq. A6* below).

Let $0 < u_1 < u_2, \ldots, u_J < T$ be a set of neural spike train measurements. Using the discrete time representation given in the METHODS section, define the events

$$A_k = \{\text{spike in } (t_{k-1}, t_k] \mid H_k\}$$
$$E_k = \{A_k\}^{\Delta N_k} \{A_k^c\}^{1-\Delta N_k}$$
$$H_k = \left\{ \bigcap_{j=1}^{k-1} E_j \right\} \tag{A1}$$

for $k = 1, \ldots, K$ and where A_k^c is the complement of A_k. For simplicity, H_k includes only the intrinsic history of the process. It can be easily extended to incorporate neural ensemble activity and other extrinsic covariates. By construction of the partition of the interval (0, T], introduced in the METHODS section, we must have $u_j \in (t_{k_j-1}, t_{k_j}]$, $j = 1, \ldots, J$, for a subset of the intervals satisfying $k_1 < k_2 < \cdots < k_J$. The remaining $K - J$ intervals have no spikes.

The probability of exactly J events occurring within the intervals $(t_{k_j-1}, t_{k_j}]_{j=1}^J$ in (0, T], may then be computed as

$$P(N_{1:K}) = P(u_j \in (t_{k_j-1}, t_{k_j}], j = 1, \ldots, J, \cap N(T) = J)$$

$$= \prod_{k=1}^{K} P(A_k)^{\Delta N_k} P(A_k^c)^{1-\Delta N_k} \tag{A2}$$

by the definition of A_k and E_k in *Eq. A1*.

The spike train thus forms a sequence of conditionally independent Bernoulli trials, with the probability of a spike in the kth time interval given by $P(A_k)$. In any interval $(t_{k-1}, t_k]$ we have

$$P(A_k) = \lambda(t_k \mid H_k)\Delta + o(\Delta)$$

$$P(A_k^c) = 1 - \lambda(t_k \mid H_k)\Delta + o(\Delta) \tag{A3}$$

Substituting *Eq. A3* into *Eq. A2* yields

$$P(N_{1:K}) = \prod_{k=1}^{K} [\lambda(t_k \mid H_k)\Delta]^{\Delta N_k}[1 - \lambda(t_k \mid H_k)\Delta]^{1-\Delta N_k} + o(\Delta^J) \tag{A4}$$

which is *Eq. 2*. For small Δ, $[1 - \lambda(t_k)\Delta] \approx \exp\{-\lambda(t_k)\Delta\}$ and $\log [\lambda(t_k)\Delta[1 - \lambda(t_k)\Delta]^{-1}] \approx \log (\lambda(t_k)\Delta)$, therefore we obtain

$$P(N_{1:K}) = \prod_{k=1}^{K} [\lambda(t_k \mid H_k)\Delta]^{\Delta N_k}[1 - \lambda(t_k \mid H_k)\Delta]^{-\Delta N_k} \prod_{k=1}^{K} [1 - \lambda(t_k \mid H_k)\Delta] + o(\Delta^J)$$

$$= \prod_{k=1}^{K} \left[\frac{\lambda(t_k \mid H_k)\Delta}{1 - \lambda(t_k \mid H_k)\Delta} \right]^{\Delta N_k} \prod_{k=1}^{K} \exp\{-\lambda(t_k \mid H_k)\Delta\} + o(\Delta^J)$$

$$= \exp\left\{ \sum_{k=1}^{K} \log [\lambda(t_k \mid H_k)\Delta]\Delta N_k - \sum_{k=1}^{K} \lambda(t_k \mid H_k)\Delta \right\} + o(\Delta^J) \tag{A5}$$

The probability density of these J exact spikes in (0, T], given by $p(N_{0:T}) = \lim_{\Delta \to 0} P(N_{1:K})/\Delta^J$, is then obtained as

$$P(N_{0:T}) = \lim_{\Delta \to 0} \frac{\exp\left\{ \sum_{k=1}^{K} \log [\lambda(t_k \mid H_k)\Delta]\Delta N_k - \sum_{k=1}^{K} \lambda(t_k \mid H_k)\Delta \right\} + o(\Delta^J)}{\Delta^J}$$

$$= \lim_{\Delta \to 0} \frac{\exp\left\{ \sum_{k=1}^{K} \log \lambda(t_k \mid H_k)\Delta N_k - \sum_{k=1}^{K} \lambda(t_k \mid H_k)\Delta \right\}\Delta^J + o(\Delta^J)}{\Delta^J}$$

$$= \exp\left\{ \int_0^T \log \lambda(t \mid H(t))dN(t) - \int_0^T \lambda(t \mid H(t))dt \right\} \tag{A6}$$

which is the joint probability density of the point process spike train in continuous time (Brown et al. 2003; Daley and Vere-Jones 2003). Note that we could have derived the likelihood for the continuous time point process (and therefore also *Eq. 3*) by a generalization of the continuous time Poisson process (Daley and Vere-Jones 2003), without resorting to representing the neural point process as a conditional Bernoulli process. We formulated the spike train joint probability in terms of *Eq. A2* only to show (see below) the equivalence between Poisson and Bernoulli–GLMs when Δ is sufficiently small.

The Poisson and Bernoulli–GLMs

We briefly define a generalized linear model and show that for small enough Δ, the Bernoulli and Poisson–GLMs are equivalent in the modeling of spiking train data.

Two main aspects characterize a generalized linear model of a random variable y (McCullagh and Nelder 1989). First, the modeled random variable y has a distribution in the exponential family. Among several members of this family are the Gaussian, the Poisson, and the Bernoulli distribution. The exponential family has the general form

$$f(y \mid \Theta, \phi) = \exp\{[y\Theta - b(\Theta)]/a(\phi) + c(y, \phi)\} \quad (A7)$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are some specific functions. If ϕ is known, this is an exponential-family model with canonical parameter Θ.

For the particular case of the Poisson distribution, $\Theta = \log\lambda$, $b(\Theta) = e^{\Theta}$, $a(\phi) = \phi = 1$, $c(y, \phi) = -\log(y!)$. The location and scale parameters are λ and ϕ, respectively. Thus the distribution in *Eq. A7* can be expressed as $f(y \mid \lambda) = \exp\{y \log \lambda - \lambda - \log(y!)\} = \lambda^y e^{-\lambda}/y!$. Note that the canonical parameter Θ has, in the Poisson case, a natural representation in terms of the logarithm of the parameter λ. The joint probability distribution for an independently distributed data set $\mathbf{y} = \{y_k\}_{k=1}^{K}$ becomes

$$f(\mathbf{y} \mid \Theta, \phi) \propto \exp\left\{\sum_{k=1}^{K} y_k \log \lambda_k - \sum_{k=1}^{K} \lambda_k\right\} \quad (A8)$$

If the rate function λ_k of this Poisson process is generalized by the conditional intensity function (*Eq. 1*); and $y_k = \Delta N_k$, $\Delta = 1$, then *Eq. A8* has the same form as the typical general likelihood function for any discrete time point process (*Eqs. 3* and *A5*).

For the Bernoulli case, we let $y = \{0, 1\}$ with the probability of success denoted by P, and set $\Theta = \log([1 - P]^{-1}P)$, $b(\Theta) = \log(1 + e^{\Theta})$, $a(\phi) = \phi$, $c(\cdot) = 1$, and $\phi = 1$. Thus, for single realizations we have $p(y \mid P) = P^y(1 - P)^{1-y}$. Given an independently distributed data set $\mathbf{y} = \{y_k\}_{k=1}^{K}$, the likelihood function under the Bernoulli distribution becomes

$$f(\mathbf{y} \mid \Theta, \phi) = \prod_{k=1}^{K} [P_k]^{y_k}[1 - P_k]^{1-y_k} \quad (A9)$$

By letting $[P_k]^{y_k} = P(A_k)^{\Delta N_k}$ we obtain *Eq. A2*.

Second, the defining feature of a generalized linear model follows. The canonical parameter of the exponential family is expressed in a linear form by a transformation given by a monotonic differentiable function. In the Poisson case, if the canonical parameter is modeled as a linear combination of general functions of covariates $\mathbf{v}$ of interest (that is, $\Theta = \log \lambda(\theta, \mathbf{v}) = \Sigma_{i=1}^{q} \theta_i g_i(v_i)$ as in *Eq. 4* or equivalently as $\lambda(\theta, \mathbf{v}) = \exp\{\Sigma_{i=1}^{q} \theta_i g_i(v_i)\}$ as in *Eqs. 5–10*), then $f(y \mid \Theta, \phi) \propto \exp\{y \log \lambda(\theta, \mathbf{v}) - \lambda(\theta, \mathbf{v})\}$ gives the distribution for a GLM under a Poisson distribution and a log link function. In the Bernoulli case, if $\Theta = \log([1 - P(A_k \mid \theta, H_k)]^{-1}P(A_k \mid \theta, H_k))$ is modeled as linear combination of general functions of the covariates, then *Eqs. 2* and *A2* give the likelihood function for a GLM under a Bernoulli distribution and a *logistic* link function.

Finally, we establish the relation between the Poisson and Bernoulli–GLMs in the context of neural point process models. After making explicit the parametric model of the conditional intensity function, we have the probability of a spike event in the time interval $(t_{k-1}, t_k]$ given by $P(A_k \mid \theta, H_k) = \lambda(t_k \mid \theta, H_k)\Delta + o(\Delta)$. For small Δ

$$\log\left[\frac{P(A_k \mid \theta, H_k)}{1 - P(A_k \mid \theta, H_k)}\right] \approx \log[\lambda(t_k \mid \theta, H_k)\Delta] \quad (A10)$$

Therefore, for small enough Δ, the Bernoulli and Poisson–GLMs are equivalent when applied to the modeling of spiking train data.

The IIG model for the ISI probability density

For a particular cell, let $t_e = t_k - u_{N_{k-1}}$ denote the time elapsed since the last spike $u_{N_{k-1}}$. The inhomogeneous inverse Gaussian ISI probability density function conditioned on the motor covariate and neural ensemble activity is defined as

$$p(t_e \mid G) = \frac{s(t_k \mid \cdot)}{\left[2\pi\left[\int_{u_{N_{k-1}}}^{t_k} s(t \mid \cdot)dt\right]^3\right]^{1/2}} \exp\left\{-\frac{1}{2}\frac{\left(\int_{u_{N_{k-1}}}^{t_k} s(t \mid \cdot)dt - \psi\right)^2}{\psi^2 \int_{u_{N_{k-1}}}^{t_k} s(t \mid \cdot)dt}\right\} \quad (A11)$$

where $G = \{u_{N_{k-1}}, \mathbf{x}_{k+\tau}, N_{1:k}^{1:C}, \theta\}$, $\psi > 0$ is the location parameter and $s(t_k \mid \cdot) > 0$ is the scaling parameter at time t_k, conditioned on the extrinsic and ensemble rates covariates as given in *Eq. 13*.

REFERENCES

Aguera y Arcas B, and Fairhall AL. What causes a neuron to spike? *Neural Comput* 15: 1789–1807, 2003.

Aguera y Arcas B, Fairhall AL, and Bialek W. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comput* 15: 1715–1749, 2003.

Akaike H. Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, edited by Petrov BN and Csaki F. Budapest: Akademiai Kiado, 1973, p. 267–281.

Andersen PK, Borgan O, Gill RD, and Keiding N. *Statistical Models Based on Counting Processes.* New York: Springer-Verlag, 1992.

Ashe J and Georgopoulos AP. Movement parameters and neural activity in motor cortex and area-5. *Cereb Cortex* 4: 590–600, 1994.

Barbieri R, Quirk MC, Frank LM, Wilson MA, and Brown EN. Construction and analysis on non-Poisson stimulus-response models of neural spiking activity. *J Neurosci Methods* 105: 25–37, 2001.

Berman M and Turner TR. Approximating point process likelihoods with GLIM. *Appl Stat* 41: 31–38, 1992.

Brillinger DR. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol Cybern* 59: 189–200, 1988.

Brown EN, Barbieri R, Eden UT, and Frank LM. Likelihood methods for neural data analysis. In: *Computational Neuroscience: A Comprehensive Approach*, edited by Feng J. London: CRC, 2003.

Brown EN, Barbieri R, Ventura V, Kass RE, and Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput* 14: 325–2346, 2002.

Burnham KP and Anderson DR. *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach* (2nd ed.). New York: Springer-Verlag, 2002.

Chornoboy ES, Schramm LP, and Karr AF. Maximum likelihood identification of neural point process systems. *Biol Cybern* 59: 265–275, 1988.

Cover TM and Thomas JA. *Elements of Information Theory.* New York: Wiley, 1991.

Csicsvari J, Henze DA, Jamieson B, Harris KD, Sirota A, Bartho P, Wise KD, and Buzsaki G. Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol* 90: 1314–1323, 2003.

Daley D and Vere-Jones D. *An Introduction to the Theory of Point Processes.* New York: Springer-Verlag, 2003.

Donoghue JP. Connecting cortex to machines: recent advances in brain interfaces. *Nat Neurosci* 5: 1085–1088, 2002.

Donoghue JP, Sanes JN, Hatsopoulos NG, and Gaal G. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *J Neurophysiol* 79: 159–173, 1998.

Eden UT, Frank LM, Barbieri R, Solo V, and Brown EN. Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput* 16: 971–998, 2004.

Efron B. The estimation of prediction error: Covariance penalties and cross-validation. *J Am Stat Assoc* 99: 619–632, 2004.

Frank LM, Eden UT, Solo V, Wilson MA, and Brown EN. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *J Neurosci* 22: 3817–3830, 2002.

Fu QG, Flament D, Coltz JD, and Ebner TJ. Temporal encoding of movement kinematics in the discharge of primate primary motor and premotor neurons. *J Neurophysiol* 73: 836–854, 1995.

Grammont F and Riehle A. Precise spike synchronization in monkey motor cortex involved in preparation for movement. *Exp Brain Res* 128: 118–122, 1999.

Grammont F and Riehle A. Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. *Biol Cybern* 88: 360–373, 2003.

Hatsopoulos NG, Ojakangas CL, Paninski L, and Donoghue JP. Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc Natl Acad Sci USA* 95: 15706–15711, 1998.

Hatsopoulos NG, Paninski L, and Donoghue JP. Sequential movement representations based on correlated neuronal activity. *Exp Brain Res* 149: 478–486, 2003.

Hille B. *Ion Channels of Excitable Membrane* (3rd ed.). Sunderland, MA: Sinauer Associates, 2001.

Jackson A, Gee VJ, Baker SN, and Lemon RN. Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron* 38: 115–125, 2003.

Johnson A and Kotz S. *Distributions in Statistics: Continuous Univariate Distributions.* New York: Wiley, 1970.

Kass RE and Ventura V. A spike-train probability model. *Neural Comput* 13: 1713–1720, 2001.

Keat J, Reinagel P, Reid RC, and Meister M. Predicting every spike: a model for the responses of visual neurons. *Neuron* 30: 803–817, 2001.

Kitagawa G and Gersh W. *Smoothness Priors Analysis of Time Series.* New York: Springer-Verlag, 1996.

Knierim JJ and Vanessen DC. Neuronal responses to static texture patterns in area-V1 of the alert macaque monkey. *J Neurophysiol* 67: 961–980, 1992.

Luczak A, Hackett TA, Kajikawa Y, and Laubach M. Multivariate receptive field mapping in marmoset auditory cortex. *J Neurosci Methods* 136: 77–85, 2004.

Maynard EM, Hatsopoulos NG, Ojakangas CL, Acuna BD, Sanes JN, Normann RA, and Donoghue JP. Neuronal interactions improve cortical population coding of movement direction. *J Neurosci* 19: 8083–8093, 1999.

McCullagh P and Nelder JA. *Generalized Linear Models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC, 1989.

Mehta MR, Barnes CA, and McNaughton BL. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc Natl Acad Sci USA* 94: 8918–8921, 1997.

Mehta MR, Quirk MC, and Wilson MA. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* 25: 707–715, 2000.

Mendel JM. *Lessons in Estimation Theory for Signal Processing, Communications, and Control.* Englewood Cliffs, NJ: Prentice Hall, 1995.

Moran DW and Schwartz AB. Motor cortical representation of speed and direction during reaching. *J Neurophysiol* 82: 2676–2692, 1999.

Nicolelis MA, Dimitrov D, Carmena JM, Crist R, Lehew G, Kralik JD, and Wise SP. Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc Natl Acad Sci USA* 100: 11041–11046, 2003.

Ogata Y. On Lewis' simulation method for point processes. *IEEE Trans Inform Theory* IT-27: 23–31, 1981.

Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83: 9–27, 1988.

O'Keefe J and Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34: 171–175, 1971.

O'Keefe J and Recce ML. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3: 317–330, 1993.

Paninski L, Fellows MR, Hatsopoulos NG, and Donoghue JP. Spatiotemporal tuning of motor neurons for hand position and velocity. *J Neurophysiol* 91: 515–532, 2004.

Papangelou F. Integrability of expected increments of point processes and a related change of scale. *Trans Am Math Soc* 165: 483–506, 1972.

Pawitan Y. *In All Likelihood: Statistical Modeling and Inference Using Likelihood.* London: Oxford Univ. Press, 2001.

Sanes JN and Truccolo W. Motor "binding": do functional assemblies in primary motor cortex have a role? *Neuron* 38: 3–5, 2003.

Santner TJ and Duffy DE. The statistical analysis of discrete data. New York: Springer-Verlag, 1989.

Schwarz G. Estimating the dimension of a model. *Ann Stat* 6: 461–464, 1978.

Scott SH. The role of primary motor cortex in goal-direction movements: insights from neurophysiological studies on non-human primates. *Curr Opin Neurobiol* 13: 671–677, 2003.

Sillito AM, Grieve KL, Jones HE, Cudeiro J, and Davis J. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378: 492–496, 1995.

Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J R Stat Soc B* 39: 44–47, 1977.

Tanner MA. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* New York: Springer-Verlag, 1996.

Vinje WE and Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.

Wilson HR. Simplified dynamics of human and mammalian neocortical neurons. *J Theor Biol* 200: 375–388, 1999.

Wilson MA and McNaughton BL. Dynamics of the hippocampal ensemble code for space. *Science* 261: 1055–1058, 1993.