Algorithms for the analysis of ensemble neural spiking activity using simultaneous-event multivariate point-process models

Demba Ba[1,2]*, Simona Temereanca[3] and Emery N. Brown[1,2,4]

[1] Department of Anesthesia, Critical Care and Pain Medicine, Harvard Medical School, Massachusetts General Hospital, Charlestown, MA, USA
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School, Massachusetts General Hospital, Charlestown, MA, USA
[4] Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA

Understanding how ensembles of neurons represent and transmit information in the patterns of their joint spiking activity is a fundamental question in computational neuroscience. At present, analyses of spiking activity from neuronal ensembles are limited because multivariate point process (MPP) models cannot represent simultaneous occurrences of spike events at an arbitrarily small time resolution. Solo recently reported a simultaneous-event multivariate point process (SEMPP) model to correct this key limitation. In this paper, we show how Solo's discrete-time formulation of the SEMPP model can be efficiently fit to ensemble neural spiking activity using a multinomial generalized linear model (mGLM). Unlike existing approximate procedures for fitting the discrete-time SEMPP model, the mGLM is an exact algorithm. The MPP time-rescaling theorem can be used to assess model goodness-of-fit. We also derive a new marked point-process (MkPP) representation of the SEMPP model that leads to new thinning and time-rescaling algorithms for simulating an SEMPP stochastic process. These algorithms are much simpler than multivariate extensions of algorithms for simulating a univariate point process, and could not be arrived at without the MkPP representation. We illustrate the versatility of the SEMPP model by analyzing neural spiking activity from pairs of simultaneously-recorded rat thalamic neurons stimulated by periodic whisker deflections, and by simulating SEMPP data. In the data analysis example, the SEMPP model demonstrates that whisker motion significantly modulates simultaneous spiking activity at the 1 ms time scale and that the stimulus effect is more than one order of magnitude greater for simultaneous activity compared with non-simultaneous activity. Together, the mGLM, the MPP time-rescaling theorem and the MkPP representation of the SEMPP model offer a theoretically sound, practical tool for measuring joint spiking propensity in a neuronal ensemble.

Keywords: multivariate point-process, simultaneous events, multinomial GLM, thalamic synchrony

INTRODUCTION

The study of how neurons represent and transmit information in the patterns of their ensemble spiking activity has been greatly facilitated in the last 17 years by the technical capability to measure simultaneously multiple single neuron activity using multiple electrode recording techniques (Wilson and McNaughton, 1993; Maynard et al., 1997). For this reason, the development of statistical methods to characterize ensemble spiking activity is an active field of neuroscience research (Brown et al., 2004; Grün, 2009).

Several histogram-based methods have been used to analyze ensemble neural spiking activity. These include the cross-correlogram (Brody, 1999), the cross-intensity function (Brillinger et al., 1976) and the joint peristimulus time histogram (Gerstein and Perkel, 1969). An appeal of these methods is that, as discrete approximations to well-known statistics used

to analyze continuous-valued processes, they are easy to understand and to compute. However, like all histogram methods, large sample theory is required to justify this approach (Klemelä, 2009). Furthermore, these methods only measure the association between pairs of neurons and the fundamental assumption of stationarity that underlies their construction can be hard to justify given that neural spiking activity is often highly plastic and adaptive.

Algorithms to detect precise patterns of spike timing are another method of measuring associations among neural spike trains (Grün et al., 2002; Gütig et al., 2002; Pipa and Grün, 2002; Grün, 2009). The appeal of these methods is that they offer a way to evaluate higher-order neural interactions in ensemble spiking activity beyond pairwise comparisons. These methods require the user to specify the complexity of the pattern to be analyzed and assume that, at least approximately, neural systems

use repeatedly the same patterns to represent and transmit information. Frequency domain estimates of coherence which also require a stationarity assumption (Brillinger, 1992; Mitra and Bokil, 2008) and gravity clustering, a high-dimensional graphical technique, Lindsey and Gerstein (2006) have also been used to analyze joint spiking activity. A common paradigm in neurophysiology is to apply a stimulus and to observe the response of the neural ensemble. The methods cited thus far have limited ability to relate the effects of the stimulus to the ensemble response.

Likelihood methods using either an information-geometric (Nakahara and Amari, 2002; Amari and Nakahara, 2006; Shimazaki et al., 2012) or a point-process (Ogata, 1981; Chornoboy et al., 1988; Okatan et al., 2005) representation provide an alternative parametric model-based approach to analyzing ensemble neural spiking activity. Likelihood methods can relate the ensemble activity to any relevant covariates. When the parametric model accurately describes the data, these analyses have important optimality properties. However, this approach has an important shortcoming that is especially relevant for analysis of coincident spiking activity. At an arbitrarily small time scale these methods do not allow simultaneous spiking (Ogata, 1981; Karr, 1991; Daley and Vere-Jones, 2003). Current analysis of spiking activity from neuronal ensembles circumvent this limitation by either assuming neurons are independent conditioned on history, or simply ignoring simultaneous events. Ventura et al. (2005) developed a likelihood procedure to overcome this limitation for analyzing a pair of neurons. In Kass et al. (2011), Kass et al. extend Ventura's approach to multiple neurons. Solo recently reported a simultaneous-event multivariate point process (SEMPP) model to correct this key limitation in general (Solo, 2007).

In this paper, we propose a multinomial generalized linear model (mGLM) of a discrete-time formulation of the SEMPP, which can be efficiently fit to ensemble neural spiking activity, and derive a new marked point-process (MkPP) representation of the SEMPP model. We focus on the important, non-trivial, algorithmic implications and properties of the mGLM and the MkPP representation. Existing procedures for fitting a GLM to discrete-time SEMPP data (Chornoboy et al., 1988; Okatan et al., 2005) are approximate ones which, conditioned on history, fit separate GLMs to each component of the SEMPP. In contrast, the mGLM algorithm is an exact procedure which fits a joint model to the discrete-time SEMPP data. We use the multivariate point process (MPP) time-rescaling theorem (Daley and Vere-Jones, 2003; Vere-Jones and Schoenberg, 2004) to assess model goodness-of-fit. The MkPP representation and its implications are not trivial consequences of the treatment in Solo (2007). In particular, it leads to new thinning and time-rescaling algorithms for efficiently simulating an SEMPP stochastic process. These algorithms are much simpler than those based on obvious extensions to SEMPPs of algorithms for simulating a univariate point-process (Brown et al., 2002). We illustrate the new SEMPP model representation by analyzing neural spiking activity from pairs of simultaneously-recorded rat thalamic neurons stimulated by periodic whisker deflections and by simulating SEMPP data.

METHODS

In this section, we develop the theory behind our algorithms for handling simultaneous events in MPP. We propose a statistical model, based on generalized linear models, to analyze simultaneous recordings from pairs of thalamic neurons. Unlike existing procedures, the algorithm for fitting this mGLM estimates the parameters of interest jointly, using the discrete-time multivariate SEMPP data. We also introduce a new MkPP representation of SEMPP data and highlight its implications for simulation of such data.

THEORY

Simultaneous-event multivariate point processes

In Solo (2007), Solo lays out a theoretical SEMPP model. In order to develop our algorithms, we review the essential features of Solo's model. We also show how one can derive the joint probability density function (PDF) of an SEMPP in discrete and continuous time using straightforward heuristic arguments.

We consider an observation interval $(0, T]$ and, for $t \in (0, T]$, let $N(t) = (N_1(t), N_2(t), \ldots, N_C(t))'$ be a C-variate point-process defined as $N_c(t) = \int_0^t dN_c(u)$, where $dN_c(t)$ is the indicator function which is 1 if there is an event at time t and 0 otherwise, for $c = 1, \ldots, C$. $N_c(t)$ counts the number of events for component c in the interval $(0, t]$. We assume that each component c has a conditional intensity function (CIF) defined as

$$\lambda_c(t|H_t) = \lim_{\Delta \to 0} \frac{P[N_c(t + \Delta) - N_c(t) = 1|H_t]}{\Delta}, \quad (1)$$

where H_t is the history of the C-variate point process up to time t. Let $dN(t) = (dN_1(t), dN_2(t), \ldots, dN_C(t))'$ be the vector of indicator functions $dN_c(t)$ at time t. We may treat $dN(t)$ as a C-bit binary number. Therefore, there are 2^C possible outcomes of $dN(t)$ at any t. C of these outcomes have only one non-zero bit [that is, only one event in one component of $dN(t)$] and $2^C - C - 1$ have two or more non-zero bits. That is, there is an event at time t in at least two components of $dN(t)$. The last outcome is $dN(t) = (0, \ldots, 0)'$.

We define $N(t)$ as a SEMPP if, at any time t, $dN(t)$ has at least two non-zero bits. That is, events are observed simultaneously in at least two of the components of $N(t)$. The special case in which, at any t, $dN(t)$ can only take as values one of the C outcomes for which only one of the bits of $dN(t)$ is non-zero is the MPP defined by Daley and Vere-Jones (2003). The joint probability density of $N(t)$ in this special case is given by the Jacod likelihood function (Ogata, 1981; Karr, 1991; Daley and Vere-Jones, 2003).

The disjoint and marked point process representations

To derive the joint probability density function of an SEMPP, we develop, in a fashion similar to Solo (2007), an alternative representation of $N(t)$. Let $M = 2^C$ be the number of possible outcomes of $dN(t)$ at t. We define a new $M - 1$-variate point process $N^*(t) = (N_1^*(t), N_2^*(t), \ldots, N_{M-1}^*(t))'$ of disjoint outcomes of $N(t)$. That is, each component of $N^*(t)$ is a counting process for one and only one of the $2^C - 1$ outcomes of $dN(t)$ (patterns of C bits) that have at least one non-zero bit. For any t, the vector $dN^*(t) = (dN_1^*(t), \ldots, dN_{M-1}^*(t))'$ is a $M - 1$-bit binary

number with at most one non-zero bit. The non-zero element of $dN^*(t)$ (if any) is an indicator of the pattern $dN(t)$ of C bits which occurs at t. $dN^*(t) = (0, \ldots, 0)'$ corresponds to $dN(t) = (0, \ldots, 0)'$. We define the CIF of $N_m^*(t)$ as

$$\lambda_m^*(t|H_t) = \lim_{\Delta \to 0} \frac{P\left[N_m^*(t + \Delta) - N_m^*(t) = 1 | H_t\right]}{\Delta}, \quad (2)$$

where the counting process is $N_m^*(t) = \int_0^t dN_m^*(u)$. We term $N^*(t)$ the *disjoint* process or representation.

One simple way to map from $dN(t)$ to $dN^*(t)$ is to treat the former as a C-bit binary number, reverse the order of its bits, and convert the resulting binary number to a decimal number. We use this decimal number as the index of the non-zero component of $dN^*(t)$. The inverse map proceeds by finding the index of the non-zero entry of $dN^*(t)$, expressing this index as a C-bit binary number, and reversing the order of the bits to obtain $dN(t)$. This one-to-one map is described in detail in Supplemental Material 1 for the arbitrary C-variate case. **Table 1** illustrates this one-to-one map for the case $C = 3$ and $M = 8$. In this example, $N(t)$ is related to $N^*(t)$ by

$$N_1(t) = N_1^*(t) + N_3^*(t) + N_5^*(t) + N_7^*(t) \quad (3)$$

$$N_2(t) = N_2^*(t) + N_3^*(t) + N_6^*(t) + N_7^*(t) \quad (4)$$

$$N_3(t) = N_4^*(t) + N_5^*(t) + N_6^*(t) + N_7^*(t). \quad (5)$$

The CIFs of $N(t)$ are related to those of $N^*(t)$ in a similar manner.

If we let $0 < t_1 < t_2 < \cdots < t_L \le T$ denote the times in the observation interval $(0, T]$ at which $dN(t)$ has at least one non-zero bit, then we can express the disjoint process $N^*(t)$ as a MkPP $\{(t_\ell, dN^*(t_\ell)\}_{\ell=1}^L$. At t_ℓ, at least one of the bits of $dN(t)$ is non-zero. The non-zero bit of $dN^*(t_\ell)$ then indicates, through the map described in Supplemental Material 1, exactly which of the $M-1$ patterns of C bits (outcomes of $dN(t)$ other than $(0, \ldots, 0)'$) occurred at t_ℓ. At any other t, $dN(t) = (0, \ldots, 0)'$. We term the unmarked process $\{t_\ell\}_{\ell=1}^L$ the ground point process (Daley and Vere-Jones, 2003) and denote by $dN_g(t)$ the indicator function that is 1 at t_ℓ, $\ell = 1, \ldots, L$ and zero at any other t. The ground point process defines the times of occurrence of *any* pattern of C bits (outcomes of $dN(t)$) that are not all zero. For each m, the times at which $dN_m^*(t)$ is non-zero define the times of occurrence of *one specific* pattern of C bits that are not all zero. It follows that the counting process and the CIF of the ground point process are

respectively

$$N_g(t) = \sum_{m=1}^{M-1} N_m^*(t) \quad (6)$$

$$\lambda_g^*(t|H_t) = \sum_{m=1}^{M-1} \lambda_m^*(t|H_t). \quad (7)$$

The probability of the marks is given by the multinomial probability mass function

$$P[dN_m^*(t) = 1 | dN_g(t) = 1, H_t] = \frac{\lambda_m^*(t|H_t)}{\lambda_g^*(t|H_t)}, \quad (8)$$

for $m = 1, \ldots, M-1$. The derivation of Equations (7, 8) are in Supplemental Material 1, on the Frontiers website. The MkPP representation provides an efficient description of $N(t)$. The probability of an event occurring in $(0, T]$ is governed by the CIF $\lambda_g^*(t|H_t)$ of the ground point process. When an event is observed in $dN_g(t)$, the marks are drawn from an $M-1$-dimensional history-dependent multinomial distribution (Equation 8) to produce the corresponding event in $N^*(t)$, or equivalently $N(t)$.

Three algorithmically-useful forms of the joint probability density function

We give three forms of the joint PDF of an SEMPP. The derivations for these PDFs follow from simple heuristic arguments, which we detail in Supplemental Material 1. Here, we focus on the algorithmic importance of these forms.

First, we give the joint PDF on an SEMPP in discrete-time, which is the basis for the mGLM algorithm used in **Data Analysis**. Then, we give the continuous PDF of the disjoint representation, from which a multivariate analog of the time-rescaling theorem and a Kolmogorov–Smirnov (KS) test for goodness-of-fit assessment can be deduced (Brown et al., 2002; Daley and Vere-Jones, 2003). Lastly, we express the continuous PDF in terms of the MkPP representation. This latter form leads to simple algorithms for simulating an SEMPP stochastic process.

We define the discrete-time representations of $N(t)$ and $N^*(t)$ in Supplemental Material 1. In this notation, ΔN^* is the $I \times M-1$ matrix of discretized outcomes of $dN^*(t)$ for the observation interval $(0, T]$. Each column ΔN_i^* of ΔN^*, where i is the discrete-time index, is a realization from a multinomial trial with M outcomes (roll of an M-sided die). The probability mass function of ΔN^* can be written as the product of conditional M-nomial trials:

$$P[\Delta N^*] = \prod_{i=1}^I \prod_{m=1}^{M-1} \left(\lambda_m^*[i|H_i]\Delta\right)^{\Delta N_{m,i}^*}$$

$$\cdot \left(1 - \lambda_g^*[i|H_i]\Delta\right)^{1 - \Delta N_{g,i}} + o(\Delta^L). \quad (9)$$

We can obtain the continuous-time joint PDF $p\left[N_{(0,T]}^*\right]$ of the disjoint process $N^*(t)$ by relating it to the discrete-time joint PDF and then taking limits. This leads to $p\left[N_{(0,T]}^*\right]$ being expressed

Table 1 | Map from $dN(t)$ to $dN^*(t)$, $C = 3$, $M = 8$.

$dN(t)$	m	$dN^*(t)$
(1,0,0)	1	(1,0,0,0,0,0,0)
(0,1,0)	2	(0,1,0,0,0,0,0)
(1,1,0)	3	(0,0,1,0,0,0,0)
(0,0,1)	4	(0,0,0,1,0,0,0)
(1,0,1)	5	(0,0,0,0,1,0,0)
(0,1,1)	6	(0,0,0,0,0,1,0)
(1,1,1)	7	(0,0,0,0,0,0,1)

as the product of $M-1$ continuous-time univariate point process PDFs

$$p\big[N^*_{(0,T]}\big] = \prod_{m=1}^{M-1} \exp\left\{\int_0^T \log\lambda^*_m(t|H_t)dN^*_m(t)\right\}$$
$$\cdot \exp\left\{-\int_0^T \lambda^*_m(t|H_t)dt\right\}. \quad (10)$$

The PDF $p\big[N^*_{(0,T]}\big]$ can also be written in terms of the MkPP representation as

$$p\big[N^*_{(0,T]}\big] = \prod_{\ell=1}^{L} \prod_{m=1}^{M-1} \left(\frac{\lambda^*_m(t_\ell|H_{t_\ell})}{\lambda^*_g(t_\ell|H_{t_\ell})}\right)^{dN^*_m(t_\ell)}$$
$$\cdot \lambda^*_g\big(t_\ell|H_{t_\ell}\big)^{dN_g(t_\ell)} \exp\left\{-\int_0^T \lambda^*_g(t|H_t)dt\right\}. \quad (11)$$

We use the discrete-time representation (Equation 9) to form the mGLM (Fahrmeir and Tutz, 2001) to analyze the joint spiking activity of a pair of thalamic neurons in **Data Analysis**. Existing procedures for fitting a GLM to discrete-time SEMPP data (Chornoboy et al., 1988; Okatan et al., 2005) are approximate ones which, conditioned on history, fit separate GLMs to each component of the disjoint representation of the SEMPP. In contrast, the mGLM algorithm is an exact procedure which jointly estimates the parameters of the model using the multivariate discrete-time SEMPP data.

The continuous-time PDF of the disjoint representation (Equation 10) is the form derived by Solo (2007). If we let $N^*(t)$ be the MPP defined by restricting $dN^*(t)$ to the C components which are indicators for the outcomes for which only one bit of $dN(t)$ is non-zero (that is, if we disregard simultaneous occurrence of events), then Equation (10) gives the joint PDF of the MPP defined by the Jacod likelihood which has no simultaneous events (Chornoboy et al., 1988; Karr, 1991; Okatan et al., 2005). The case $M = 2$ corresponds to the joint PDF of a univariate point process (Truccolo et al., 2005). $N^*(t)$ is an $M-1$-variate MPP because it is composed of the disjoint events of $dN(t)$. Therefore, by Proposition 7.4.VI in Daley and Vere-Jones (2003), it can be transformed into an $M-1$-variate point process with independent unit-rate Poisson processes as its components. This observation, applied to Equation (10), allows us to create a multivariate extension of the KS plots to assess goodness-of-fit of SEMPP models fit to simultaneous neural spiking activity.

Finally, Equation (11) makes explicit the formulation of the joint PDF of an SEMPP as an MkPP. The PDF of Equation (11) follows from the observation that, in any small time interval Δ, the SEMPP is a multinomial model with 2^C possible outcomes (Equation 9, $M = 2^C$). A multinomial model with 2^C outcomes can be written as the product of a binomial probability model and a conditional multinomial probability model with 2^C-1 outcomes. The binomial probability model defines the ground process and the conditional multinomial process defines the marked process. The CIF of the ground point process is given in Equation (7) and the mark process is defined by the

history-dependent $M-1$-dimensional multinomial distribution in Equation (8). The intuition behind the MkPP representation, as well as its implications, should not be dismissed as a trivial consequences of the treatment in Solo (2007). An important consequence of the MkPP representation is that it can be combined with either a thinning algorithm or the univariate time-rescaling theorem (Brown et al., 2002) to yield highly efficient simulation algorithms for SEMPP models. As we shall see, these algorithms are much simpler than those based on obvious extensions to SEMPPs of algorithms for simulating a univariate point-process (Brown et al., 2002). Indeed, the simulation algorithms which use the MkPP representation make simulation of SEMPP data almost as simple as simulation of a univariate point process.

Details of the derivations for Equations (9–11) are in Supplemental Material 1. The derivation for Equation (10) follows from simple heuristic arguments used previously in Truccolo et al. (2005) for univariate point processes and in Solo (2007) for SEMPPs. The key idea is to start with a discrete-time form, as in Equation (9), and then take limits.

Simulating simultaneous-event multivariate point processes

For univariate point processes, an important consequence of the time-rescaling theorem is that it leads to an algorithm for simulating univariate point-process data (Brown et al., 2002). The time-rescaling theorem for MPPs also leads to an algorithm for simulating SEMPP data, which we describe below. In the absence of history dependence of the CIFs, this algorithm is a trivial extension of the one for simulating a point process using the univariate time-rescaling theorem (Brown et al., 2002). For reasons detailed below, this algorithm, which uses the disjoint representation, becomes unwieldy in the more interesting case where the CIFs depend on the history of the SEMPP. In contrast, using the MkPP representation leads to two algorithms (also described below), which make the task of simulating SEMPP data almost as simple as simulating a univariate point process.

Algorithm 1 (Multivariate time-rescaling):

1. Set $t_0 = 0$, $\ell = 1$, $\ell_m = 1 \ \forall \ m \in \{1, \ldots, M-1\}$.
2. $\forall \ m$, draw τ_{ℓ_m} an exponential random variable with mean 1.
3. $\forall \ m$, find t_{ℓ_m} as the solution to:
   $\tau_{\ell_m} = \int_{t_{\ell_{m-1}}}^{t_{\ell_m}} \lambda^*_m(t|H_t)dt.$
   Let $m^+ = \arg\min_m t_{\ell_m}$, $t_\ell = t_{\ell_{m^+}}$.
4. If $t_\ell > T$, then stop the algorithm, else
5. If $m = m^+$, set $dN^*_{m^+}(t_\ell) = 1$, $\ell_m = \ell_m + 1$ and draw τ_{ℓ_m} an exponential random variable with mean 1.
6. If $m \neq m^+$, ℓ_m does not change, set
   $\tau_{\ell_m} = \tau_{\ell_m} - \int_{t_{\ell_{m-1}}}^{t_\ell} \lambda^*_m(t|H_t)dt,$
   $t_{\ell_{m-1}} = t_\ell,$
   $dN^*_m(t_\ell) = 0,$
7. $dN(t_\ell)$ is obtained from $dN^*(t_\ell)$ using the map described in Supplemental Material 1.
8. $\ell = \ell + 1$.
9. Go back to 3.

Although Algorithm 1 is not easy to parse, we describe the intuition behind it, which follows from the time-rescaling theorem for MPPs. Step 2 of the algorithm prescribes that we should first generate independent exponential random variables with mean 1, one for each component of the disjoint representation of the SEMPP. Then, for each component, we find in Step 3 the times at which the area under each CIF integrates to the value prescribed by the exponential draws. If the CIFs did not depend on history, it would be sufficient to repeat these two steps, until we've simulated the SEMPP data up to a given time T. In other words, without dependence of the CIFs on history, one can simulate an SEMPP by *separate* simulation of multiple univariate point processes. Intuitively, this is because the integration of the CIFs in Step 3 of Algorithm 1 become decoupled when there is no history dependence, that is, no component of the SEMPP needs to update its own history based on events that occur in the other components. However, in the more interesting case when the CIFs depend on history, Step 3 of Algorithm 1 becomes problematic: integration of the CIFs must be carried out *in parallel* because whenever an event occurs in one component, all processes must update their history accordingly. Moreover, only the process for which an event occurs must perform a new exponential 1 draw (Step 5). All other processes keep computing the area under their CIF until it reaches the value prescribed by their exponential 1 draw (Step 6).

Simply put, when the CIFs depend on history, Step 3 of Algorithm 1 leads to an undesirable amount of bookkeeping, which grows with the number of component of the SEMPP.

Algorithm 1 should be contrasted with one which uses the MkPP representation. In this algorithm, we simulate events from the ground process (which is a univariate point process) using the univariate time-rescaling theorem, e.g., as in Brown et al. (2002). When an event occurs in the ground process, we roll an $M-1$-sided die to decide to which component of the disjoint process this event should be assigned. The key feature of this algorithm, which makes it simpler and more elegant than Algorithm 1, is that we must perform only one uniform draw at each step. The components of the SEMPP only interact with each other, through the mark pmf, when an event occurs in the ground process. This makes all of the bookkeeping of Algorithm 1 unnecessary.

The following algorithm for simulating an SEMPP based on the MkPP representation is an extension of the thinning simulation algorithm for MPP models developed by Ogata (1981).

Algorithm 2 (Thinning): Suppose there exists λ such that $\lambda_g^*(t|H_t) \leq \lambda$ for all $t \in (0, T]$:

1. Simulate observations $0 < t_1 < t_2 \cdots < t_K \leq T$ from a Poisson point process with rate λ.
2. Set $k = 1$.
3. while $k \leq K$

    a. Draw u_k from the uniform distribution on (0,1)
    b. if $\frac{\lambda_g^*(t_k|H_{t_k})}{\lambda} \geq u_k$

        i. Draw m_k from the $(M-1)$-dimensional multinomial distribution with probabilities $\frac{\lambda_m^*(t_k|H_{t_k})}{\lambda_g^*(t_k|H_{t_k})}$, $m = \{1, \ldots, M-1\}$
        ii. set $dN_{m_k}^*(t_k) = 1$ and $dN_m^*(t_k) = 0$ for all $m \neq m_k$
    c. else, set $dN_m^*(t_k) = 0$ for all $m \in \{1, \ldots, M-1\}$
    d. $dN(t_k)$ is obtained from $dN^*(t_k)$ as in Supplemental Material 1
    e. $k = k + 1$.

An alternative form of Algorithm 2 is given in Supplemental Material 1. We can also simulate data from an SEMPP model using an algorithm based on the univariate time-rescaling theorem (Brown et al., 2002) as follows:

Algorithm 3 (Time-rescaling): Given an interval $(0, T]$

1. Set $t_0 = 0$ and $\ell = 1$.
2. Draw u_ℓ from the uniform distribution on (0,1).
3. Find t_ℓ as the solution to: $\log(u_\ell) = \int_{t_{\ell-1}}^{t_\ell} \lambda_g^*(t|H_t)dt$.
4. If $t_\ell > T$, then stop, else
5. Draw m_ℓ from the $(M$-1$)$-dimensional multinomial distribution with probabilities $\frac{\lambda_m^*(t_\ell|H_{t_\ell})}{\lambda_g^*(t_\ell|H_{t_\ell})}$, $m = \{1, \ldots, M-1\}$.
6. set $dN_{m_\ell}^*(t_\ell) = 1$ and $dN_m^*(t_\ell) = 0$ for all $m \neq m_\ell$.
7. $dN(t_\ell)$ is obtained from $dN^*(t_\ell)$ as in Supplemental Material 1.
8. $\ell = \ell + 1$.
9. Go back to 2.

DATA ANALYSIS

To illustrate our model, we analyze simultaneously-recorded spiking activity data from pairs of neurons in the rat thalamus. The experiments were previously described in detail in Temereanca et al. (2008).

Experiment

Simultaneous single unit activity from pairs of thalamic neurons was recorded with two electrodes placed in the same electro physiologically-identified barreloid of the rat ventral posteromedial nucleus. For the neurons analyzed here, spiking activity was recorded from the pairs in response to whisker deflections at 50 mm/s administered at 8 Hz for a period of 2000 ms. A delay period of 500 ms preceded and followed each stimulus period. For each neuronal pair, the responses were recorded across 50 trials. We divided the 50 trials into a training set and a test set by randomly choosing 1 of every sequence of 3 trials and assigning it to the training set (17 trials). The remaining trials were assigned to the test set (33 trials).

The standard raster plots show that the stimulus (**Figure 1A**) induces strong modulation of the neural spiking in the training set (**Figure 1B**) and in the test set (**Figure 1C**). Our approach suggests an alternative raster plot which shows more clearly the simultaneous spiking activity of the pairs. Our new raster plots of the three components of ΔN^* (**Figure 2**) show more clearly the effects of the stimulus (**Figure 2A**) on the training set (**Figure 2B**) and the test set (**Figure 2C**). The $\Delta N_{3,i}^*$ component
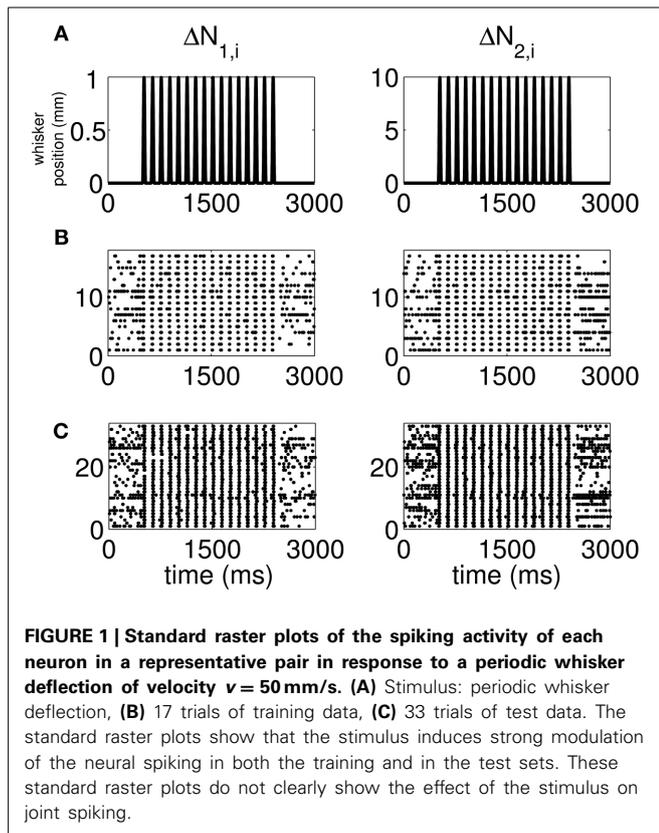
FIGURE 1 | Standard raster plots of the spiking activity of each neuron in a representative pair in response to a periodic whisker deflection of velocity $v = 50$ mm/s. (A) Stimulus: periodic whisker deflection, **(B)** 17 trials of training data, **(C)** 33 trials of test data. The standard raster plots show that the stimulus induces strong modulation of the neural spiking in both the training and in the test sets. These standard raster plots do not clearly show the effect of the stimulus on joint spiking.

of ΔN^* shows the joint spiking activity of the two neurons in response to the stimulus. This is the component that analyses based on the Jacod likelihood point-process models would simply ignore. For our analyses, we have $C = 2$, $M = 4$ and $\Delta = 1$ ms.

Statistical model

We assume that the data are a sample from a bivariate SEMPP, whose discrete-time likelihood can be written as a product of conditional four-nomial trials. If we let

$$\log\left[\frac{\lambda_m^*[i|H_i]\Delta}{1 - \lambda_g^*[i|H_i]\Delta}\right] = \beta_{m,0} + \sum_{j=0}^{J-1} \beta_{m,j}^{(0)} s_{i-j}$$

$$+ \sum_{c=1}^{2} \sum_{k=1}^{K_c} \beta_{m,k}^{(c)} \Delta N_{c,i-k} \quad (12)$$

then our model becomes an mGLM with four-nomial observations and logit link (Fahrmeir and Tutz, 2001). The model expresses the log odds of each outcome with respect to $\Delta N_{g,i} = (0, \ldots, 0)'$ outcome as the convolution of the stimulus s with a finite length kernel $\{\beta_{m,j}^{(0)}\}_{j=0}^{J-1}$, and the history of ΔN_1 and ΔN_2, respectively with finite length kernels $\{\beta_{m,k}^{(1)}\}_{k=1}^{K_1}$ and $\{\beta_{m,k}^{(2)}\}_{k=1}^{K_2}$. Estimation is performed by maximizing the discrete-time likelihood [Equation (9) re-arranged as in Equation (S28), Supplemental Material 2] of the data under the model in

Equation (12). We select J, K_1 and K_2 using Akaike's information criterion (AIC) and assess goodness-of-fit using the time-rescaling theorem to construct KS plots (Brown et al., 2002). In Supplemental Material 1, we state the multivariate time-rescaling theorem (Proposition 7.4.VI in Daley and Vere-Jones, 2003) and describe how to construct the KS plots we use to assess goodness of fit.

Assessing independence

It is not hard to show that the components of a bivariate Bernoulli random vector are independent if and only if they are uncorrelated. In each discrete-time bin, the model of Equation (12) results in an estimate of a *joint* pmf, conditioned on history. Therefore, we can assess the time-varying dependence between the neurons in a pair using the (conditional) covariance in each time bin.

Algorithms for fitting GLMs to SEMPP data

The goal of algorithms for fitting a model to discrete-time SEMPP data is to estimate the time and/history-dependent CIFs of the components of the SEMPP model. Conventional approaches are approximate ones which, given history, *separately* estimate the CIFs of each of the components of the SEMPP model. In the mGLM framework, one *jointly* estimates the CIFs of the components of the SEMPP model, thereby preserving the true discrete-time nature of the data.

More specifically, conventional approaches (Chornoboy et al., 1988; Okatan et al., 2005) fit a GLM to the discrete-time SEMPP data by first discretizing the continuous likelihood of Equation (10), to obtain the approximate discrete-time likelihood of Equation (S15), Supplemental Material 1. In effect, these approaches are similar to that of Berman and Turner (1992) in the univariate case. The resulting approximate discrete-time likelihood is the product of discrete-time univariate point process likelihoods. Then, using algorithms available in standard scientific computing packages, *separate* univariate Poisson GLMs are fit to the data.

In contrast, our mGLM algorithm uses the exact discrete-time likelihood of Equation (9), thus treating the multivariate SEMPP data as samples of conditional multinomial trials. The mGLM algorithm *jointly* fits multiple components to the data. This allows us to simultaneously estimate the time and/or history-dependent CIFs of the components of the SEMPP. This approach is different from conventional ones which assume that the components of the SEMPP are independent conditioned on history. The impact of independence conditoned on history is that it allows for fitting of *separate* univariate point-process models to the components of the SEMPP. The degree of conditional independence depends on how much history is allowed in process (e.g., 10, 20, 50).

Both approaches perform maximum likelihood estimation using Newton's method. One can recover the conventional approach of fitting separate GLMs to the components of the discrete-time SEMPP data by assuming that certain terms in the Hessian of the exact discrete-time likelihood are $o(\Delta)$. The Hessian of the exact discrete-time likelihood [after substituting Supplemental Material 2 Equation (S30) in (S28)] involves the multinomial covariance matrix, the entries of which are of
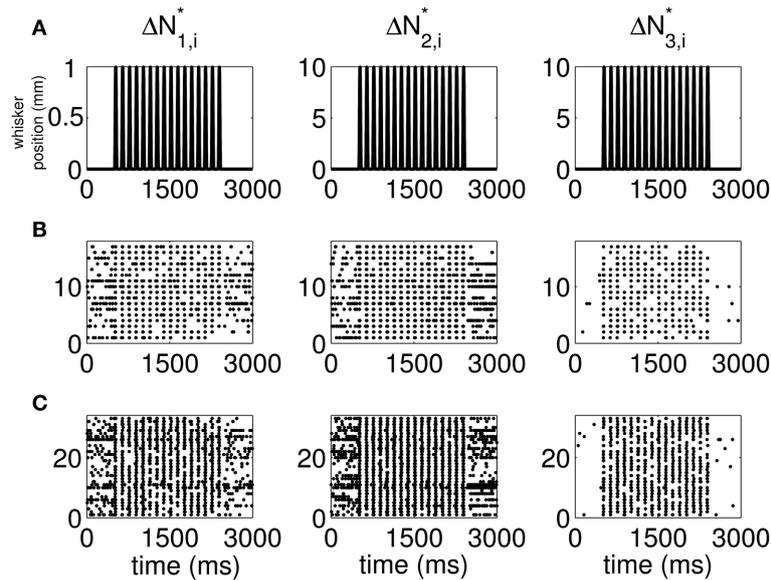
FIGURE 2 | New raster plots of non-simultaneous ("10" and "01," Columns 1 and 2) and simultaneous ("11," Column 3) spiking events for the neuron pair in Figure 1. Each column corresponds to one of the components of ΔN^*. **(A)** Stimulus, **(B)** 17 trials of training data, **(C)** 33 trials of test data. The new raster plots of the three components show clearly the effects of the stimulus on non-simultaneous and simultaneous spiking activity. The $\Delta N_{3,i}^*$ component of ΔN^* shows that the joint spiking activity of the two neurons in response to the stimulus is pronounced.

the form $\lambda_m^* \Delta \delta_{m,m'} - \lambda_m^* \Delta \cdot \lambda_{m'}^* \Delta$, where $\delta_{m,m'} = 1$, if $m = m'$ and 0 otherwise, $m, m' = 1, \ldots, M-1$. If we assume that the $\lambda_m^* \Delta \cdot \lambda_{m'}^* \Delta$ component is $o(\Delta)$, the Hessian of the mGLM becomes block-diagonal. Then, each block corresponds to the Hessian matrix for the parameters of *one* of the components of the SEMPP. The block-diagonal structure of the Hessian implies that the parameters for each component can be estimated separately, as conventional approaches do.

This approximate equivalence between the multinomial GLM and multiple separate univariate Poisson GLMs extends the equivalence developed in Truccolo et al. (2005) between univariate Poisson and Bernoulli GLMs.

As we'll see below, conventional approaches are simpler and typically faster than the mGLM, due to their parallel nature. However, they require the CIF of the ground process to be uniformly small. This is an assumption that may be plausible in some cases but is hard to justify for all data from neurophysiology experiments, particularly those that use explicit stimuli such as the data from the whisking experiment (Temereanca et al., 2008), a small subset of which we analyze below.

An alternate SEMPP model and algorithm

Kass et al. fit separate GLMs to each component of the discrete-time representation of $N(t)$. Then, they quantify the amount of excess simultaneous events that is observed beyond what is expected under independence (Kass et al., 2011). The mGLM algorithm results in a time-varying assessment of dependence, conditioned on history. In the case of Kass et al., the dependence structure is summarized using several scalar statistics. In that sense, the mGLM algorithm results in a finer characterization of the dependence structure of the components of the SEMPP.

In the case of a bivariate SEMPP, this dependence structure is fully-summarized by the covariance between its components. The mGLM algorithm gives a time-varying estimate of this covariance as function of the relevant covariates, as well as history.

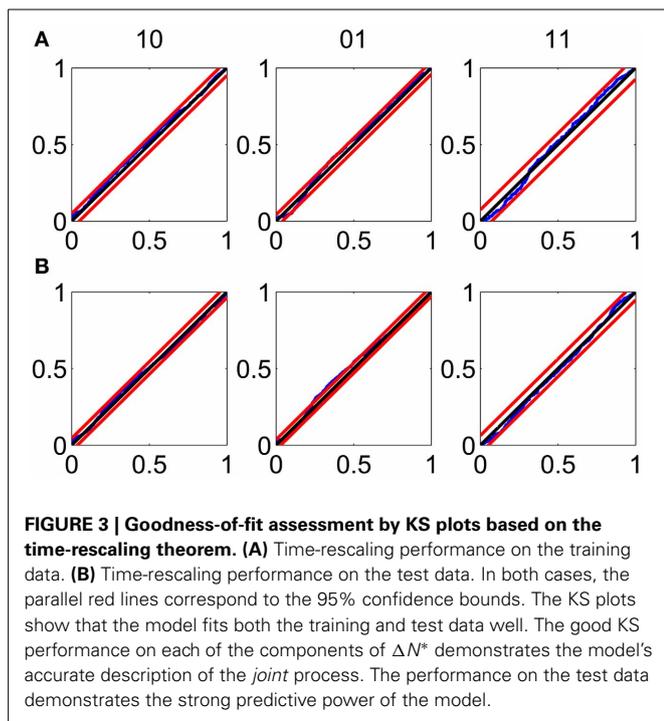## RESULTS

STIMULUS-INDUCED INCREASES IN JOINT FIRING

To select the optimal model order of our model (Equation 12), we considered values for J, K_1, and K_2 ranging from 2 to 50 ms, in 1 ms increments. We used the results of preliminary GLM analyses on each neuron separately to reduce the dimension of the search space. The optimal orders obtained were $J = 5$, $K_1 = 37$, and $K_2 = 14$.

We found that reducing J to a value as low as $J = 2$ did not affect the goodness-of-fit, as measured by the number of points outside of the 95% confidence bounds in the KS plots. Therefore, the results we report here are for $J = 2$, $K_1 = 37$, and $K_2 = 14$.

The KS plots show that the model fits both the training (**Figure 3A**) and test data (**Figure 3B**) well. The good KS performance on each of the components of ΔN^* demonstrates the model's accurate description of the *joint* process. The performance on the test data demonstrates the strong predictive power of the model.

The analysis indicates that there is a high propensity for simultaneous firing due to the stimulus. This is reflected by the stimulus modulation of the simultaneous-spiking event ("11") (**Figure 4A**, Column 2), which is defined by

$$sm_{11}[i] = \exp\left\{ \sum_{j=0}^{J-1} \beta_{3,j}^{(0)} s_{i-j} \right\} \qquad (13)$$

FIGURE 3 | Goodness-of-fit assessment by KS plots based on the time-rescaling theorem. (A) Time-rescaling performance on the training data. (B) Time-rescaling performance on the test data. In both cases, the parallel red lines correspond to the 95% confidence bounds. The KS plots show that the model fits both the training and test data well. The good KS performance on each of the components of ΔN^* demonstrates the model's accurate description of the *joint* process. The performance on the test data demonstrates the strong predictive power of the model.

FIGURE 4 | Stimulus modulation of non-simultaneous ("10" and "01") and simultaneous ("11") events, over a single stimulus period. The stimulus modulation of an event is the amount by which the stimulus increases the probability of that event. (A) Left: Stimulus modulation of the non-simultaneous events, $\left(\exp \left\{ \sum_{j=0}^{J-1} \beta_{m,j}^{(0)} s_{i-j} \right\}, m = 1, 2 \right)$. Right: Stimulus modulation of the simultaneous event, $\left(\exp \left\{ \sum_{j=0}^{J-1} \beta_{3,j}^{(0)} s_{i-j} \right\} \right)$. (B) Stimulus over a single period. The figure shows that the maximum amount by which the stimulus increases the probability of joint spiking is one order of magnitude greater than that by which the stimulus increases the probability of non-simultaneous spiking. Thus, simultaneous spiking can be attributed for the most part to the administration of the stimulus.

This is a unitless quantity which represents, at a ms time scale, the amount by which the stimulus increases the probability of a simultaneous event. Similarly, we can define the quantities

$$sm_{01}[i] = \exp \left\{ \sum_{j=0}^{J-1} \beta_{2,j}^{(0)} s_{i-j} \right\} \qquad (14)$$

and

$$sm_{10}[i] = \exp \left\{ \sum_{j=0}^{J-1} \beta_{1,j}^{(0)} s_{i-j} \right\}, \qquad (15)$$

which represent the stimulus modulation of the "01" and "10" events, respectively. As **Figure 4A**, Column 1 suggests, there is also a modulation of the spiking activity in response to the stimulus for the non-simultaneous events ("10" and "01"). However, the maximum stimulus effect for the simultaneous spiking (≈ 1200) is one order of magnitude stronger than the maximum effect for the non-simultaneous spiking (≈ 90).
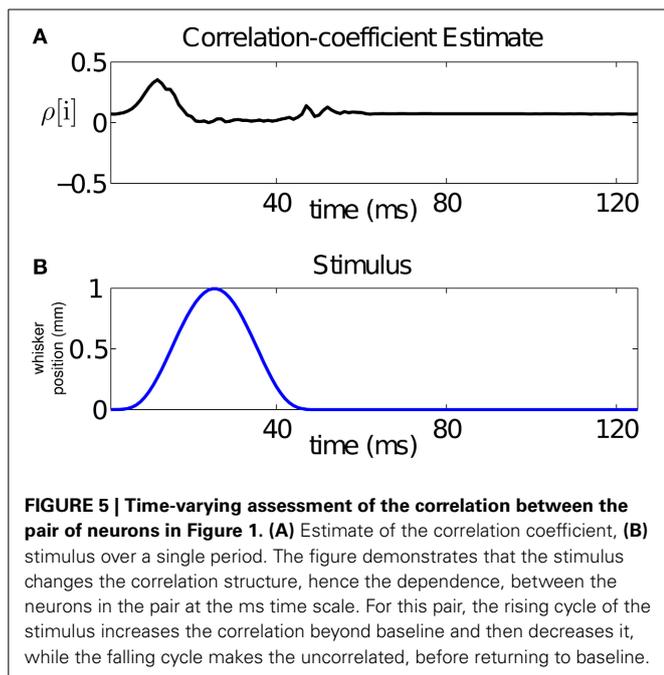
From our analysis, we can compute a time-varying estimate of the CIFs for each of the 17 trials used for training, and hence of the covariance between the neurons for each of these trials. For a bivariate binary process, the covariance fully summarizes the dependence between its components. For visualization purposes, we average the CIF estimates across trials and use these averages to obtain a trial-averaged covariance estimate. Using the trial-averaged covariance estimate, as well as the trial-averaged CIFs, we compute a trial-averaged correlation coefficient estimate as follows:
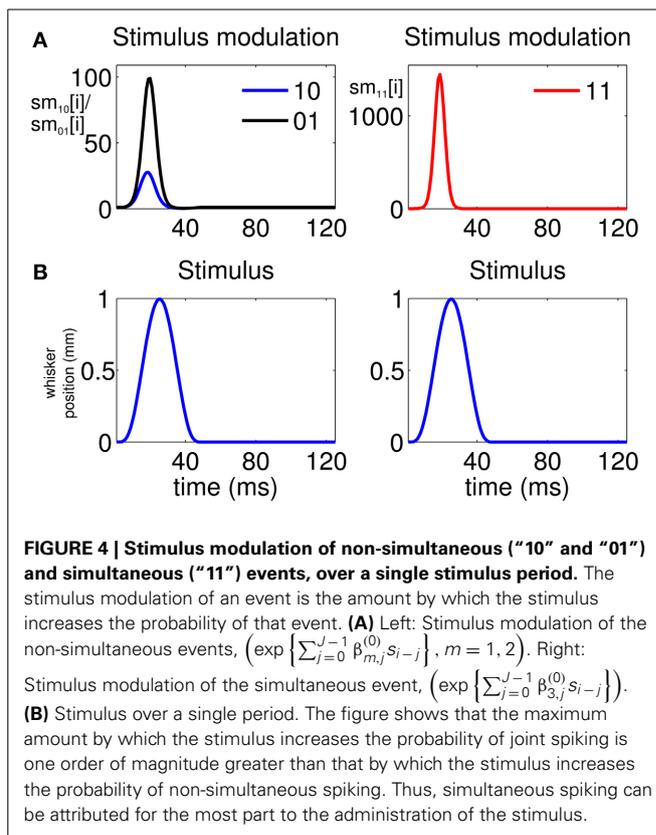
$$\rho[i] = \frac{\lambda_3^*[i|H_i]\Delta - \lambda_1[i|H_i]\Delta \cdot \lambda_2[i|H_i]\Delta}{\sqrt{\lambda_1[i|H_i]\Delta(1 - \lambda_1[i|H_i]\Delta)\lambda_2[i|H_i]\Delta(1 - \lambda_2[i|H_i]\Delta)}}. \qquad (16)$$

Since the stimulus is periodic in the interval between 500 and 2500 ms (with period 125 ms), we average our estimate of the across-trial correlation coefficient over the 16 stimulus cycles. The result is displayed in **Figure 5A**. The figure demonstrates that the stimulus changes the correlation structure, hence the dependence, between the neurons in the pair at the ms time scale. In particular, changes in the correlation coefficient mirror changes in the stimulus. For this pair, the rising cycle of the stimulus increases the correlation beyond baseline and then decreases it, while the falling cycle makes the neurons uncorrelated, before returning to baseline.

Despite the strong stimulus modulation of simultaneous firing, there is a smaller number of simultaneous occurrences than non-simultaneous occurrences (**Figure 2**). The intrinsic dynamics of Neuron 1 and Neuron 2 (Figure S8C, Columns 2 and 3) limit their high propensity to fire simultaneously. Indeed, in these supplemental figures (available in Supplemental Material 3 on the Frontiers website), the first few coefficients corresponding to the effect of Neurons 1 and 2 on the probability of simultaneous firing are negative for *both* neurons. This means that if both neurons have just fired, their probability of simultaneous firing within the next 1–3 ms decreases significantly. On the other hand, for non-simultaneous events, *only one* of the set of first few coefficients mentioned above is negative. This explains

FIGURE 5 | Time-varying assessment of the correlation between the pair of neurons in Figure 1. (A) Estimate of the correlation coefficient, **(B)** stimulus over a single period. The figure demonstrates that the stimulus changes the correlation structure, hence the dependence, between the neurons in the pair at the ms time scale. For this pair, the rising cycle of the stimulus increases the correlation beyond baseline and then decreases it, while the falling cycle makes the uncorrelated, before returning to baseline.

why we observe a greater number of non-simultaneous than simultaneous occurrences.

Although some of the 95% confidence intervals for the individual parameters (Figure S8) overlap with zero, these small coefficients contribute significantly to the descriptive (**Figure 3A**) and predictive (**Figure 3B**) power of the model. Results of analysis of a different pair of neurons are in Figures S9–S14 (Supplemental Material 3), which can be found on the Frontiers website.

SIMULATED NEURAL SPIKING DATA

We use the time-rescaling algorithm (Algorithm 3) to simulate simultaneous spiking activity from three thalamic neurons in response to periodic whisker deflections of velocity 50 mm/s. We model the CIFs of the neurons as in Equation (12) to simulate 33 trials of the experiment described above. For these simulations, we chose $J = 2$, $K_1 = 2$, $K_2 = 2$, and $K_3 = 2$. Figure S15 shows the standard raster plots of the simulated data. There is strong modulation of the activity of each of the neurons by the stimulus. **Figure 6** shows the raster plots of each of the seven disjoint components of ΔN^*. As the figure indicates, the parameters of the model were chosen so that the stimulus strongly modulates simultaneous occurrences from the pairs Neuron 1 and Neuron 2, Neuron 2 and Neuron 3, as well as simultaneous occurrences from the triple.

COMPARING mGLM AND SEPARATE UNIVARIATE GLMs

We used the data from (**Figure 1**) to compare the mGLM algorithm to the approximate algorithm which fits separate univariate Bernoulli GLMs to each of the components of the disjoint representation (**Figure 2**).

The mGLM, as well as algorithms for fitting univariate Bernoulli GLM, can be implemented using Newton's method.

Each Newton step corresponds to a high-structured linear system, which is typically solved using the QR-decomposition. Our experience with neural data sets shows that using linear conjugate gradient (CG) instead can significantly speed up computation (Komarek and Moor, 2005; Ba, 2011). This can be mainly attributed to the sparse patterns of design matrices that arise from neural data sets. Therefore, our comparisons use CG-based implementations of the algorithms for fitting the mGLM and the univariate Bernoulli GLM. The CG algorithm requires a tolerance to which the linear system is solved, as well as a maximum number of iterations to perform. We set the tolerance to 10^{-10}, and the number of CG iterations to the maximum possible, which is the dimension of the linear system. Each CG iteration is initialized using the parameters of the previous Newton update. We use the same values of J, K_1, and K_2 as in **Data Analysis** to fit the models.

Figure 7 plots the deviance of each of the algorithms as a function of time elapsed from the beginning of the algorithm to the end of a Newton iteration. For GLMs, the deviance generalizes the idea of mean-squared error for linear models: minimizing the deviance is equivalent to maximizing the likelihood. The algorithms which fits separate GLMs is parallel in nature. Therefore, to compute the time elapsed until the end of the kth Newton iteration, we use the maximum (across all separate components) of the time elapsed until the end of each component's kth iteration. The figure shows that the mGLM algorithm achieves a smaller value of the deviance (larger value of the likelihood), but is slower to converge. That is, the mGLM required $\approx$5 s to reach the value of the deviance that the algorithm which fits separate GLMs converged to in $\approx$2 s.

DISCUSSION
mGLM: A VERSATILE PARADIGM FOR ENCODING ANALYSES OF ENSEMBLE NEURAL ACTIVITY

We have presented a new algorithmic framework for the analysis of ensemble neural spiking activity based on the joint probability density of a MPP with simultaneous events. We have developed an equivalent MkPP representation of this joint density which, together with the previously-described continuous and discrete-time representations (Solo, 2007), provide a principled way to conduct model-based data analyses and simulation studies of ensemble neural activity that offer several advantages over current analyses approaches.

Likelihood-based analyses
The discrete-time SEMPP representation provides an efficient means of performing likelihood analyses of multiple simultaneously-recorded neurons using an mGLM algorithm for model fitting (Fahrmeir and Tutz, 2001) and the disjoint representation (Equation 10) (Solo, 2007) allows us to assess goodness-of-fit using the time-rescaling theorem. The mGLM algorithm is a multivariate extension of the GLM algorithm used to fit univariate point process models to single neural spike trains (Paninski, 2004; Truccolo et al., 2005). Unlike existing approaches to fitting a GLM to SEMPP data, the mGLM jointly estimates the CIFs of the component of the SEMPP, and achieves a larger value of the likelihood (**Figure 7**). The algorithm which fits separate components achieves coefficient estimates which are
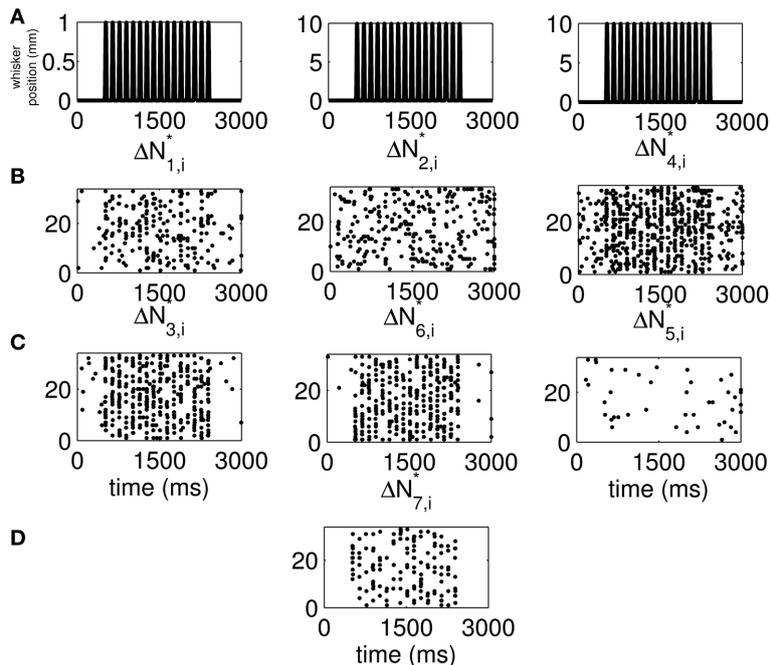
FIGURE 6 | New raster plots of non-simultaneous ("100," "010," and "001") and simultaneous ("110," "011," "101," and "111") spiking events for the three simulated neurons of in Figure S15. (A) Stimulus (same as in **Figures 1A**, **2A**). **(B)** Non-simultaneous events, from left to right, "100," "010," and "001." **(C)** Simultaneous events from pairs of neurons, from left to right, "110," "011," and "101." **(D)** Simultaneous event from the three neurons ("111"). The new raster plots of the three components show clearly the effects of the stimulus on non-simultaneous and simultaneous spiking. The $\Delta N_{4,i}^{*}$ and $\Delta N_{5,i}^{*}$ components of ΔN^{*} show that the joint spiking activity of the pairs consisting of Neurons 1 and 2 on the one hand, and Neurons 2 and 3 on the other hand is pronounced. The $\Delta N_{7,i}^{*}$ component of ΔN^{*} shows that the joint spiking activity of the three neurons is also pronounced. The information in these raster plots about the joint spiking activity of neurons could not be gathered from Figure S15.

very close to the ones of the mGLM algorithm, for the non-simultaneous components ("10" and "01"), but less so for the simultaneous component ("11" event) (data not shown). Our intuition tells us that the coefficients of the simultaneous "11" event are underestimated due to few number of events (**Figure 2**), coupled with the fact that the separate fitting procedure, unlike the mGLM algorithm, does not use second-order correlation information between its components.

Assessing goodness-of-fit

The goodness-of-fit analysis (**Figure 3**) uses the MPP time-rescaling theorem (Daley and Vere-Jones, 2003; Vere-Jones and Schoenberg, 2004) and is a multivariate extension of the one-dimensional time-rescaling theorem used to assess goodness-of-fit for univariate point process models (Brown et al., 2002). Once adequate goodness-of-fit is established, our inference framework is likelihood-based. Therefore, it uses the Fisher information to compute parameter standard errors and confidence intervals (Figures S8, S14), and established procedures such as AIC and BIC to perform model selection. Our approach also carries all the optimality properties of the likelihood framework.
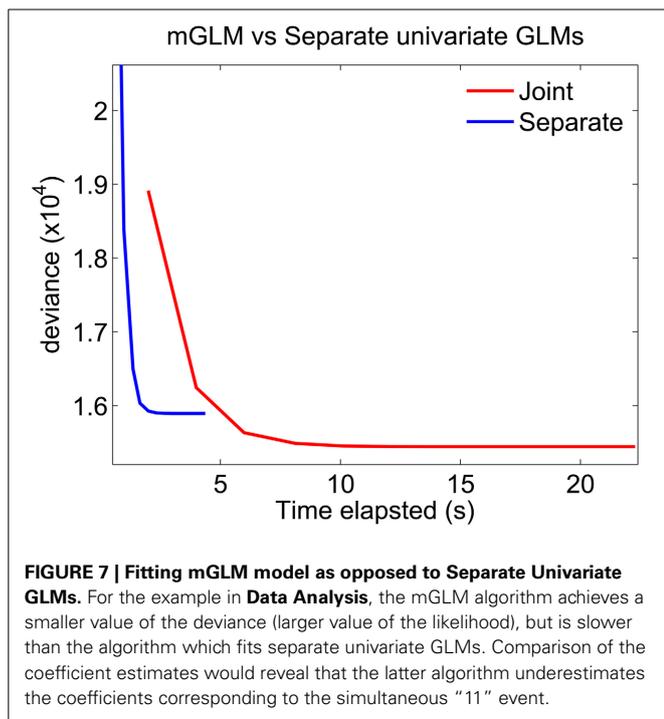
Analyzing an arbitrary number of neurons

Our paradigm applies to multiple neurons and is therefore more general than the common bivariate histogram methods (Gerstein and Perkel, 1969; Brillinger et al., 1976; Brody, 1999). It also obviates the stationarity assumption required by histogram methods (Gerstein and Perkel, 1969; Brillinger et al., 1976; Brody, 1999). Unlike the Jacod likelihood (Chornoboy et al., 1988; Karr, 1991; Okatan et al., 2005), the SEMPP model allows simultaneous events at an arbitrarily small time-scale. Although we only applied our techniques to simultaneous recordings from pairs of neurons, our approach characterizes in each small time interval of length Δ the probability of all $2^{C}-1$ patterns of spiking activity from C neurons. For this reason, our approach offers a key improvement over pattern analysis algorithms (Grün et al., 2002; Gütig et al., 2002; Pipa and Grün, 2002; Grün, 2009) because it analyzes in a likelihood framework all of the patterns defined by the recording resolution Δ.

Simulating ensemble neural activity

The MkPP representation also gave a new thinning and a new time-rescaling algorithm for simulating simultaneous neural spiking activity in continuous time (**Figures 6**, S15). This makes it possible to use the same modeling approach for data analysis and for simulation studies. The MkPP representation follows from the observation that, in any small time interval Δ, the SEMPP is a multinomial model with 2^{C} possible outcomes. A multinomial model with 2^{C} outcomes can be written as the product of a binomial probability model and a conditional multinomial probability model with $2^{C}-1$ outcomes. The binomial probability model defines the ground process and the conditional

FIGURE 7 | Fitting mGLM model as opposed to Separate Univariate GLMs. For the example in **Data Analysis**, the mGLM algorithm achieves a smaller value of the deviance (larger value of the likelihood), but is slower than the algorithm which fits separate univariate GLMs. Comparison of the coefficient estimates would reveal that the latter algorithm underestimates the coefficients corresponding to the simultaneous "11" event.

multinomial process defines the marked process. The simulation algorithms from the MkPP representation (Algorithms 2 and 3) are simpler than one based on the time-rescaling theorem for MPPs (Algorithm 1), which uses the disjoint representation.

JOINT ENCODING OF WHISKER MOTION BY TWO PAIRS OF THALAMIC NEURONS

Our analysis of the joint spiking activity of the pair of thalamic neurons in response to whisker stimulation reveals the types of new insights that could be learned from the SEMPP model. It suggests a new raster plot to assess simultaneity (**Figure 2**, Column 3). The key finding of our model analysis that is not apparent from this new raster plot is that the stimulus modulation of the joint spiking activity is more than one order of magnitude greater than its modulation of either of the unpaired spiking activity (**Figure 4**). These estimates of the stimulus modulation are model-based spike-triggered averages (Simoncelli et al., 2004). Our analysis also reveals that changes in the correlation structure, and hence the dependence structure, between the neurons in a pair mirror changes in the stimulus (**Figures 5**, S13). These brief analyses suggest that our approach could offer new insights into the importance of joint spiking activity for understanding representations of stimuli in ensemble neural spiking activity.

To our knowledge, these examples constitute the first practical demonstrations of the versatility of the SEMPP model for analysis and simulation of jointly-recorded neural spiking activity.

FUTURE DIRECTIONS AND IMPLICATIONS OF THE mGLM MODEL

The new SEMPP model and analysis suggests several important future extensions of this work.

Implications of mGLM for joint encoding of whisker motion by pairs of thalamic neurons

The pair of neurons analyzed here, as well as the pair in Figures S9–S14, are from a set of 17 pairs of thalamic neurons simultaneously recorded during stimulation by three different whisker velocities (Temereanca et al., 2008). In a future report, we will present our findings on the analysis of the complete data set using the SEMPP model fit with the mGLM algorithm.

Applications of mGLM to assessment of functional connectivity

By substituting the SEMPP likelihood for the Jacod likelihood, our methods suggest a new approach to analyzing functional connectivity in neuronal ensembles (Okatan et al., 2005), that may give a more accurate assessment of the extent to which it is modulated by the intrinsic dynamics of the neurons and/or external stimuli (Truccolo et al., 2009).

Decoding analyses of sensory representations and prediction of whisker motion

Thalamic neurons use their firing patterns to form a representation of the sensory information conveyed by whisker deflections (Temereanca et al., 2008). We constructed a highly accurate model of this sensory representation by applying the mGLM to simultaneous recordings from pairs of thalamic neurons in response to periodic whisker deflections. Encoding models fit using mGLM are arguably superior to existing approaches because of their ability to capture the effect of stimuli (e.g., whisker motion) on the simultaneous-spiking activity of neurons. As witnessed by our analyses of the thalamic data, this effect can be quite substantial. In such scenarios, when the effect of sensory stimuli on simultaneous spiking is pronounced, using the mGLM as an encoding model can result in superior decoding performance. For the thalamic data analyzed here, we believe that the ability to predict whisker motion using an encoding model fit using mGLM, and a carefully-designed decoding algorithm using the SEMPP likelihood (Ba, 2011), will be superior compared to the case when one uses the convential GLM (Paninski, 2004; Truccolo et al., 2005) for encoding and the Jacod likelihood (Eden et al., 2004) for decoding. Indeed, our SEMPP model also suggests a new approach to designing point-process filters for ensemble neural spike train decoding and adaptive filtering studies of neural plasticity (Brown et al., 1998, 2001). These new algorithms based on the disjoint representation make explicit use of simultaneous events. The computational complexity of fitting the SEMPP model increases exponentially with the number of neurons. Therefore, the design of more efficient model-fitting algorithms must be an important focus of future work.

Simultaneous spiking at smaller time scales

The impact of using smaller time scales (Δ), rather than the conventional 1 ms time scale, is an important question that we hope to address in future work. The mGLM and convential GLM (Paninski, 2004; Truccolo et al., 2005) are based on discrete-time approximations of the continuous-time likelihood of an SEMPP and univariate point-process, respectively. In recent work (Citi et al., 2013), we have introduced a novel discrete-time approximation to the continuous-time likelihood of

a point-process that is significantly more accurate than that used in the conventional GLM (Paninski, 2004; Truccolo et al., 2005). We have demonstrated that the approximation used in the conventional GLM can require time scales that are up to one order of magnituded smaller ($10\,\mu$s as opposed to 1 ms) to achieve the same goodness-of-fit as our new approximation (Citi et al., 2013). The improvement comes from explicitly accounting for the refractory period of neurons in the new approximation. It is not hard to extend this formulation to the mGLM. We believe this formulation provides a better ground for assessing the impact of smaller time scales on the mGLM analyses.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fncom. 2014.00006/abstract

REFERENCES

Amari, S., and Nakahara, H. (2006). Correlation and independence in the neural code. *Neural Comput.* 18, 1259–1267. doi: 10.1162/neco.2006.18. 6.1259

Ba, D. (2011). *Algorithms and Inference for Simultaneous-Event Multivariate Point-Process, with Applications to Neural Data*. Ph.D. thesis, Massachusetts Institute of Technology.

Berman, M., and Turner, T. (1992). Approximating point process likelihoods with GLIM. *Appl. Stat.* 41, 31–38. doi: 10.2307/2347614

Brillinger, D. R. (1992). Nerve cell spike train data analysis. *J. Amer. Stat. Assoc.* 87, 260–271. doi: 10.1080/01621459.1992.10475205

Brillinger, D. R., Bryant, H. L., and Segundo, J. P. (1976). Identification of synaptic interactions. *Biol. Cybern.* 22, 213–220. doi: 10.1007/BF00365087

Brody, C. D. (1999). Correlations without synchrony. *Neural Comput.* 11, 1537–1551. doi: 10.1162/089976699300016133

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

Brown, E. N., Barbierri, R., Ventura, V., Kass, R., and Frank, L. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computat.* 14, 325–346. doi: 10.1162/08997660252741149

Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461. doi: 10.1038/nn1228

Brown, E. N., Nguyen, D., Frank, L., Wilson, M., and Solo, V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *PNAS* 98, 12261–12266. doi: 10.1073/pnas.201409398

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* 59, 265–275. doi: 10.1007/BF00332915

Citi, L., Ba, D., Brown, E. N., and Barbieri, R. (2013). Likelihood methods for point processes with refractoriness. *Neural Comput.* 26, 237–263. doi: 10.1162/NECO_a_00548

Daley, D. J., and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, 2nd Edn., Vol. 1. New York, NY: Springer.

Eden, U. T., Frank, L. M., Solo, V., and Brown, E. N. (2004). Dynamic analyis of neural encoding by point process adaptive filtering. *Neural Comput. Lett.* 16, 971–998. doi: 10.1162/089976604773135069

Fahrmeir, L., and Tutz, G. (2001). *Multivariate Statistical Modeling Based on Generalized Linear Models*, 2nd Edn. New York, NY: Springer. doi: 10.1007/978-1-4757-3454-6

Gerstein, G. L., and Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 164, 828–830. doi: 10.1126/science.164.3881.828

Grün, S. (2009). Data-driven significance estimation of precise spike correlation. *J. Neurophysiol.* 101, 1126–1140. doi: 10.1152/jn.00093.2008

Grün, S., Diesmann, M., and Aersten, A. (2002). Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. *Neural Comput.* 14, 81–119. doi: 10.1162/089976602753284464

Gütig, R., Aersten, A., and Rotter, S. (2002). Statistical significance of coincident spikes: count-based versus rate-based statistics. *Neural Comput.* 14, 121–153. doi: 10.1162/089976602753284473

Karr, A. F. (1991). *Point Processes and Their Statistical Inference*. New York, NY: Dekker.

Kass, R., Kelly, R., and Loh, W. (2011). Assessment of synchrony in multiple neural spike trains using log linear point process models. *Annal. Appl. Stat.* 5, 1262–1292. doi: 10.1214/10-AOAS429

Klemelä, J. (2009). Multivariate histograms with data-dependent partitions. *Stat. Sin.* 19, 159–176. Available online at: http://www3.stat. sinica.edu.tw/statistica/j19n1/j19n18/j19n18.html

Komarek, P., and Moor, A. W. (2005). "Making logistic regression a core data mining tool with TR-IRLS," in *Proceedings of the 5th IEEE International Conference on Data Mining* (Houston, TX), 685–688. doi: 10.1109/ICDM. 2005.90

Lindsey, B. G., and Gerstein, G. L. (2006). Two enhancements of the gravity algorithm for multiple spike train analysis. *J. Neurosci. Methods* 150, 116–127. doi: 10.1016/j.jneumeth.2005.06.019

Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* 102, 228–239. doi: 10.1016/S0013-4694(96)95176-0

Mitra, P. P., and Bokil, H. S. (2008). *Observed Brain Dynamics*. New York, NY: Oxford University Press.

Nakahara, H., and Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Comput.* 14, 2269–2316. doi: 10.1162/08997660260293238

Ogata, Y. (1981). On lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* 27, 23–31. doi: 10.1109/TIT.1981.1056305

Okatan, M., Wilson, M. A., and Brown, E. N. (2005). Analyzing functional connectivity using a network likelihood model of neural ensemble spiking activity. *Neural Comput.* 17, 1927–1961. doi: 10.1162/0899766054322973

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Netw. Comput. Neural Syst.* 15, 243–262. doi: 10.1088/0954-898X/15/4/002

Pipa, G., and Grün, S. (2002). Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neuocomputing* 52–54, 31–37. doi: 10.1016/S0925-2312(02)00823-8

Shimazaki, H., Amari, S.-i., Brown, E. N., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput. Biol.* 8:e1002385. doi: 10.1371/journal.pcbi.1002385

Simoncelli, E., Paninski, L., Pillow, J., and Schwartz, O. (2004). "Characterization of neural responses with stochastic stimuli," in *The Cognitive Neurosciences*, ed M. Gazzaniga (Cambridge, MA: MIT Press), 327–338.

Solo, V. (2007). "Likelihood functions for multivariate point processes with coincidences," in *Proceedings of IEEE Conference on Decision and Control*, Vol. 3 (New Orleans, LA), 4245–4250. doi: 10.1109/CDC.2007.4434050

Temereanca, S., Brown, E. N., and Simons, D. J. (2008). Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *J. Neurosci.* 28, 11153–11164. doi: 10.1523/JNEUROSCI.1586-08.2008

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., and Brown, E. N. (2005). A point-process framework for relating neural spiking activity to spiking history, neural ensemble, and extrensic covariate effects. *J. Neurophysiol.* 93, 1074–1089. doi: 10.1152/jn.00697.2004

Truccolo, W., Hochberg, L. R., and Donoghue, J. P. (2009). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nat. Neurosci.* 13, 105–111. doi: 10.1038/nn.2455

Ventura, V., Cai, C., and Kass, R. E. (2005). Statistical assessment of time-varying dependency between two neurons. *J. Neurophysiol.* 94, 2940–2947. doi: 10.1152/jn.00644.2004

Vere-Jones, D., and Schoenberg, F. (2004). Rescaling marked point processes. *Aust. N. Z. J. Stat.* 46, 133–143. doi: 10.1111/j.1467-842X.2004.00319.x

Wilson, M. A., and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science* 261, 1055–1058. doi: 10.1126/science.8351520