Detection of bursts in extracellular spike trains using hidden semi-Markov point process models

Surya Tokdar
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
Email: stokdar@stat.cmu.edu
Phone: (412) 268 3556
Fax: (412) 268 7828

Peiyi Xi
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

Ryan C. Kelly
Department of Computer Science
Center for the Neural Basis of Cognition
Carnegie Mellon University
Pittsburgh, PA 15213

Robert E. Kass
Department of Statistics
Center for the Neural Basis of Cognition
Carnegie Mellon University
Pittsburgh, PA 15213

May 20, 2009

Abstract

Neurons *in vitro* and *in vivo* have epochs of bursting or "up state" activity during which firing rates are dramatically elevated. Various methods of detecting bursts in extracellular spike trains have appeared in the literature, the most widely used apparently being Poisson Surprise (PS). A natural description of the phenomenon assumes (1) there are two hidden states, which we label "burst" and "non-burst," (2) the neuron evolves stochastically, switching at random between these two states, and (3) within each state the spike train follows a time-homogeneous point process. If in (2) the transitions from non-burst to burst and burst to non-burst states are memoryless, this becomes a hidden Markov model (HMM). For HMMs, the state transitions follow exponential distributions,

and are highly irregular. Because observed bursting may in some cases be fairly regular—exhibiting inter-burst intervals with small variation—we relaxed this assumption. When more general probability distributions are used to describe the state transitions the two-state point process model becomes a hidden semi-Markov model (HSMM). We developed an efficient Bayesian computational scheme to fit HSMMs to spike train data. Numerical simulations indicate the method can perform well, sometimes yielding very different results than those based on PS.

Keywords: Hidden Markov Model, Markov-Modulated Poisson Process, Poisson Switching Model, Poisson Surprise, Rank Surprise, Up State

1 Introduction

Extracellularly-recorded spike trains often contain clusters of several spikes, separated by unusually small inter-spike intervals (ISIs). Such clusters may represent sudden epochs of elevated firing rate due to a neuron's intrinsic dynamics, a response to bistable network behavior, or oscillations traveling through a region of the brain (Koch, 2004; Doiron et al., 2003; Cooper et al., 2005; Izhikevich et al., 2003; Wilson and Cowan, 1972). They are called "bursts" or "up states" depending on the context. Analysis of bursting or up state data requires identification of when the bursts occur, their number, and their duration. From extracellular measurements alone, however, there is an immediate problem of definition: when the underlying voltage fluctuations are not observed, it is unclear what should constitute a burst. In this paper we discuss a conceptually simple approach: at each time point t we assume there is—for a given neuron—an unobserved dichotomous state, which we label "burst" or "non-burst". The statistical problem then becomes one of identifying the hidden states, which may be accomplished using maximum likelihood or Bayesian methods. This approach seems quite natural, and accords well with theoretical conceptions of both intrinsic bursting and up/down networks. We ignore any distinction here between bursts and up states, because the statistical detection methods will be the same regardless of the physiological situation, and we use the term "burst" throughout.

Some authors have devised extracellular burst-detection methods based on interspike interval (ISI) length, often with a criterion for a minimal number of spikes with small ISIs (Lo et al., 1991; Martinson et al., 1997; Corner et al., 2002; Turnbull et al., 2005; Tam, 2002). A conceptual difficulty for such methods, illustrated in Figure 1, is that spike trains generated from pure renewal processes—such as those produced by simple integrate-and-fire models—can exhibit clusters of neurons that have the appearance of bursts. In some circumstances these "null cases" might be ruled out by substantive considerations. In most others, one needs a careful, model-based calibration to determine whether or not a cluster of short ISIs represents a burst.

A very popular burst-detection method, called Poisson Surprise (PS) (Legendy and Salcman, 1985), offers such a calibration based on a probabilistic model on the ISIs. With PS, one calculates a surprise value S to measure how unlikely it is that a cluster with n spikes in a time interval T, would occur by chance. The method performs its chance calculation under the assumption that the ISIs are independent realizations from an exponential density. This amounts to assuming the neuron to spike according to a time homogeneous Poisson process. It is, however, known that spike trains often exhibit distinctly non-Poisson behavior (Koch, 2004; Gourevitch and Eggermont, 2007). To avoid the Poisson assumption Gourevitch and Eggermont (2007) proposed a rank surprise (RS) index, which again computes a surprise value but instead uses the distribution

Figure 1: Spike train simulated from inverse Gaussian renewal model. By chance, spikes tend to form clusters, resembling bursts. The inverse Gaussian mean was 30.8 spikes per second with shape parameter 19.3, which were maximum likelihood estimates from an inverse Gaussian fit to the retinal ganglion data analyzed below.

of ISI ranks to perform the chance calculation.

Both PS and RS calibrate clusters of ISIs under the null model that all ISIs in the spike train are generated from a pure renewal process. An alternative is to specify a model that directly accommodates the existence of both bursting and non-bursting phases via a hidden binary state. We have developed point process models in which the firing rate is one of two values according to whether a binary hidden state is bursting or non-bursting. With an appropriate probability model on the hidden state, Bayes's rule may be used to compute—for every time t—the conditional probability that the hidden state was bursting at t given the entire spike train recorded from the neuron. The time intervals where this probability exceeds a certain cutoff (such as 0.5) are then declared to be bursts.

The simplest two-state model for neural spike trains is a switching Poisson (SP) process (or Markov-modulated Poisson process) (Scott 1999; Abeles et al., 1995) in which spiking activity follows two homogeneous Poisson processes, one for each state, and the state transitions from non-bursting to bursting and bursting to non-bursting occur according to a Markov chain. Such hidden Markov models (HMMs) (Baum and Petrie, 1966; Rabiner 1989) have two potential restrictions. First, the neural activity within bursting and non-bursting states continues to be considered Poisson, which may well be inaccurate. Second, as a Markov model the transitions are assumed memoryless. This means that the inter-burst intervals (and inter-non-burst intervals) follow exponential distributions, which are maximally irregular (they are distributions that maximize entropy subject to being positive with a fixed expectation). Thus, spike trains that exhibit regular bursting activity (such as roughly oscillatory bursting) will be poorly fit by HMMs, and this may cause sub-optimal behavior of the detection algorithm. We relaxed both the within-state Poisson assumption and the between-state exponential assumption by implementing a switching gamma process model in which the state transitions were also governed by gamma distributions. The latter formulation makes this a hidden semi-Markov model (HSMM). The parameters of the gamma distribution, and therefore the ISI distributions in both bursting and non-bursting states, are learned from the data, and the

bursting and non-bursting state transitions are estimated. The purpose of this article is to describe our HSMM implementation and study its effectiveness.

An appealing feature of HMMs is computational tractability, most often via an expectation-maximization algorithm known as the Baum-Welch algorithm. This algorithm uses a fast forward-backward recursion to perform maximum likelihood estimation of model parameters and conditional probability evaluation of the hidden state given the estimated parameters. Chib (1996) developed a variation of this to construct a Markov Chain Monte Carlo (MCMC) algorithm for a Bayesian treatment of this model. In the Bayesian setting, the conditional probability evaluation requires an additional integration over the model parameters with respect to their joint posterior distribution given the observed spike train. Chib (1996) noted that the Markov chain structure of the HMM model allows efficient sampling from the posterior distribution through a Gibbs sampler. The computational approach we have developed applies Gibbs sampling to HSMMs by expanding the state space so that the HSMM takes a Markovian form. In Section 2 we provide details of our implementation; in Section 3 we give results from a small simulation study, comparing our HSMM to PS and RS, and also to a point-process HMM; in Section 4 we apply the HSMM to a data set analyzed previously by several other authors; and in Section 5 we discuss the results.

2 Methods

2.1 Hidden Binary Model

We denote the hidden binary state of the neuron at time t by $C(t)$ with $C(t) = 1$ coding a bursting state, $C(t) = 0$ a non-bursting state. We suppose the observation time interval is $[0, T]$ and assume that on this interval $C(t)$ can have only finitely many transitions between its two states. The waiting time from one transition to the next will be called an inter-transition interval (ITI). The ITIs are assumed independent. Those corresponding to the bursting state ($C(t) = 1$) are from a density $f_1^{\mathrm{ITI}}(\cdot)$ and those from a non-bursting state are from a density $f_0^{\mathrm{ITI}}(\cdot)$.

Within each state, we consider the neural spike train to be governed by a renewal process with an inter-spike interval (ISI) probability density f_1 or f_0 depending on the state level. Such a conception may be technically inaccurate because transitions may occur in the midst of an ISI. However, for simplicity we assume that every ISI is regulated by the state of the neuron at the completion of the previous spike. (As we mention in the discussion, we also investigated an alternative method that did not make this assumption, but it was much more cumbersome and gave similar results in the cases we examined.) More formally,

letting y_i be the i-th ISI and $\tau_i = \sum_{j=1}^{i} y_j$ denote the time of the i-th spike, the first ISI Y_1 is assumed to follow $f^{\text{ISI}}_{C(0)}$ and the subsequent ISI's are conditionally modeled as:

$$y_i \mid (y_1, \cdots, y_{i-1}, C[0, \tau_{i-1}]) \sim f^{\text{ISI}}_{C(\tau_{i-1})} \tag{1}$$

with $C[0,T] = \{C(t) : 0 \leq t \leq T\}$. Each of the four densities f^{ITI}_s, f^{ISI}_s, $s \in \{0,1\}$ is assumed known only up to finitely many parameters, all of which are collected together into a vector θ. In addition to the hidden binary burst states, the vector θ must be learned from the data.

2.2 Burst Detection

Note that the hidden process $C[0,T]$ completely determines the bursting and non-bursting states of the neuron. After observing the spike train, statistical inference about $C[0,T]$ can be drawn from its posterior distribution given $y_{1:n} = (y_1, \cdots, y_n)$ (or equivalently given $\tau_{1:n} = (\tau_1, \cdots, \tau_n)$). Identification of whether the neuron is bursting at a time point $t \in [0,T]$ is based on the posterior probability $\Pr(C(t) = 1 \mid y_{1:n})$. Under the interpretation of this posterior probability as a reasonable degree of belief as to whether the neuron is bursting at time t, the most intuitive cutoff value would be .5, i.e., the neuron would be determined to be bursting whenever the reasonable degree of belief favored bursting rather than non-bursting. Using this cutoff, we would identify the neuron as bursting at time t if $\Pr(C(t) = 1 \mid y_{1:n}) > .5$. Other cutoff values could be used and, in fact, we do use other cutoffs in examining properties of the procedure, below. These posterior probabilities were computed by using an efficient Markov chain Monte Carlo (MCMC) method. Our assumption that the distribution of y_i depends only on the state $s_i = C(\tau_{i-1})$ ensures that the posterior distribution of $C[0,T]$ factors as

$$p(C[0,T] \mid y_{1:n}) = p(C[0,T] \mid s_{1:n}) p(s_{1:n} \mid y_{1:n})$$

where $s_{1:n} = (s_1, \cdots, s_n)$. For this reason, we focused our MCMC on sampling only from $p(s_{1:n} \mid y_{1:n})$.

We designed our MCMC to produce a sample of $(s_{1:n}, \theta)$ from their joint posterior distribution given $y_{1:n}$. Our MCMC algorithm alternates between updates of θ and of $s_{1:n}$. We update θ by using a Multiple-Try Metropolis move (Liu *et al.* 2001) that leaves the conditional posterior density $p(\theta \mid s_{1:n}, y_{1:n})$ invariant. This density is easy to compute (up to a normalizing constant), see expression (2) below. We use a Gibbs update for $s_{1:n}$, which is essentially a random draw of $s_{1:n}$ from the conditional posterior $p(s_{1:n} \mid y_{1:n}, \theta)$. Below we describe this update in more detail.

2.3 Sampling from $p(s_{1:n} \mid y_{1:n}, \theta)$

Given θ, the pair $(s_{1:n}, y_{1:n})$ defines a hidden Markov model when f_1^{ITI} and f_0^{ITI} are exponential densities. This results in the property that $C(t, T]$ is conditionally independent of $C[0, t)$ given $C(t)$ for any $0 \leq t < T$, which ensures that the s_i's form a Markov chain. For such models where the s_i's live on a finite state space, Chib (1996) developed an efficient algorithm to sample from the posterior distribution of $s_{1:n}$ given $Y_{1:n}$. He introduced a Gibbs sampler that sampled all of $s_{1:n}$ in a single draw, which made the procedure much faster than Metropolis-Hastings samplers where the hidden states were updated one at a time.

Chib's method, however, relies on the Markov property of $s_{1:n}$ and thus does not apply to the case where f_1^{ITI} and f_0^{ITI} are not exponential densities. We have introduced a suitable variable augmentation which induces the required Markov property on an extended state space. The basic idea is to archive with every spike time τ_i the time since the last transition. Although this time cannot be exactly determined from $y_{1:n}$ and $s_{1:n}$, it can be bracketed by $(r_i - y_i, r_i)$ where

$$r_i = y_{i-m_i+1} + \cdots + y_i$$

with $m_i = \min\{j > 0 : s_{i-j} \neq s_j\}$. If we restrict $C(t)$ from having more than one transition between two successive spikes, then

$$p(s_{1:n}, y_{1:n} \mid \theta) = p(s_1) f_{s_1}^{\text{ISI}}(y_1) \prod_{j=2}^{n} \text{Bernoulli}(I(s_i = s_{i-1}) \mid \phi_{s_{i-1}, r_{i-1}, y_{i-1}}) \quad (2)$$

where $\text{Bernoulli}(x \mid p) = p^x (1-p)^{1-x}$ denotes a Bernoulli pdf with probability p, $I(\cdot)$ denotes the indicator function and $\phi_{s,r,y} = (1 - F_s^{\text{ITI}}(r))/(1 - F_s^{\text{ITI}}(r-y))$ is the probability that an ITI in state s will stretch beyond r given that it is already larger than $r - y$; here F_0^{ITI} and F_1^{ITI} are the cumulative distribution functions corresponding to f_0^{ITI} and f_1^{ITI}. With these definitions (s_i, r_i, y_i) forms a Markov chain with the distribution of $s_1 = C(0)$ unspecified.

We now present details of our adaptation of Chib's method to sample from $p(s_{1:n}, r_{1:n} \mid y_{1:n}, \theta)$ where $r_{1:n} = (r_1, \cdots, r_n)$. In the following we suppress θ from the notations, because all computations are done for θ fixed at its current value in the MCMC. The following notations and derivations closely follow the constructions given in Chib (1996). For any vector $x_{1:n} = (x_1, \cdots, x_n)$, let $x_{i:j}$ denote the sub-vector $(x_i, x_{i+1}, \cdots, x_j)$, $1 \leq i \leq j \leq n$. Notice that

$$p(s_{1:n}, r_{1:n} \mid y_{1:n}) = p(s_n, r_n \mid y_{1:n}) \prod_{i=1}^{n-1} p(s_i, r_i \mid s_{i+1:n}, r_{i+1:n}, y_{1:n})$$

with

$$\begin{aligned} p(s_i, r_i \mid s_{i+1:n}, r_{i+1:n}, y_{1:n}) &\propto p(y_{i+1:n}, s_{i+1:n}, r_{i+1:n} \mid s_i, r_i, y_{1:i}) p(s_i, r_i \mid y_{1:i}) \\ &\propto p(y_{i+1}, s_{i+1}, r_{i+1} \mid s_i, r_i) p(s_i, r_i \mid y_{1:i}). \end{aligned}$$

Because $r_{1:i}$ is always completely determined by $y_{1:i}$ and $s_{1:i}$, the first probability distribution above can be written as

$$p(y_{i+1}, s_{i+1}, r_{i+1} \mid s_i, r_i) \quad = \\ I(r_{i+1} = I(s_{i+1} = s_i)r_i + y_{i+1})p(s_{i+1}, y_{i+1} \mid s_i, r_i).$$

Therefore, once the pdfs $p(s_i, r_i \mid y_{1:i})$ are known, $(s_{1:n}, r_{1:n})$ can be easily sampled from $p(s_{1:n}, r_{1:n} \mid y_{1:n})$ by sequentially sampling (s_n, r_n), (s_{n-1}, r_{n-1}) and so on. Below we outline how these pdfs can be computed in a recursive manner—this again follows the strategy in Chib (1996) but with some important differences.

Let $l_{ij} = y_i + y_{i-1} + \cdots + y_{i-j+1}$, $i = 1, \cdots, n$ and $j = 1, \cdots, i$. It is clear that the i-th pdf $g_i(s_i, r_i) = p(s_i, r_i \mid y_{1:i})$ is to be evaluated only at $(s_i, r_i) \in \{0, 1\} \times \{l_{ij};\ j = 1, 2, \cdots, i\}$. Suppose these evaluations have been done for a given i. Then the next pdf $g_{i+1}(s_{i+1}, r_{i+1})$ can be evaluated at the desired values via the following two steps:

1. *Prediction:*

$$\begin{aligned}
&p(s_{i+1} = 1, r_{i+1} = l_{i+1,1} \mid y_{1:i}) \\
&\quad = \sum_{j=1}^{i} g_i(s_i = 0, r_i = l_{ij})\text{Bernoulli}(1 \mid \phi(0, l_{ij}))f(y_{i+1} \mid \lambda(1, 0, l_{ij}, \theta)) \\
&p(s_{i+1} = 0, r_{i+1} = l_{i+1,1} \mid y_{1:i}) \\
&\quad = \sum_{j=1}^{i} g_i(s_i = 1, r_i = l_{ij})\text{Bernoulli}(0 \mid \phi(1, l_{ij}))f(y_{i+1} \mid \lambda(0, 1, l_{ij}, \theta)) \\
&p(s_{i+1} = 1, r_{i+1} = l_{i+1,j} \mid y_{1:i}) \\
&\quad = g_i(s_i = 1, r_i = l_{i,j-1})\text{Bernoulli}(1 \mid \phi(1, l_{i,j-1}))f(y_{i+1} \mid \lambda(1, 1, l_{i,j-1})) \\
&p(s_{i+1} = 0, r_{i+1} = l_{i+1,j} \mid y_{1:i}) \\
&\quad = g_i(s_i = 0, r_i = l_{i,j-1})\text{Bernoulli}(0 \mid \phi(0, l_{i,j-1}))f(y_{i+1} \mid \lambda(0, 0, l_{i,j-1}))
\end{aligned}$$

2. *Update:*

$$g_{i+1}(s_{i+1}, r_{i+1}) = \frac{p(s_{i+1}, r_{i+1} \mid y_{1:i})}{c_{i+1}}$$

where

$$c_{i+1} = \sum_{k=0}^{1} \sum_{j=1}^{i+1} p(s_{i+1} = k, r_{i+1} = l_{i+1,j} \mid y_{1:i}).$$

The algorithm described above demands $O(n^2)$ flops and storage. This can be reduced to $O(n)$ by splitting the spike train into contiguous segments and

updating the states of the ISI's within each segment together. Choosing these segments to be of length $O(w)$, the entire train can be updated with only $O(nw)$ flops and storage. In our examples we chose the segment length randomly from a discrete uniform distribution on the integers in [5; 20]. The segments are created and processed from right to left until the whole train is covered. Notice that choosing the window length as large as n would be practically infeasible except for very small spike trains. On the other hand choosing the window too short would resemble the less inefficient one-state-at-a-time update.

3 Simulation Study

After implementing the HSMM described above we assessed its performance, comparing it to HMM, PS and RS. For our comparisons we used spike trains simulated from 5 distinct processes, which we call settings, chosen to combine realistic ISI distributions together with features that might pose a challenge to the methods. We then evaluated the methods based on estimated number of bursts and ROC curves.

Each setting corresponded to a model that was likely to produce clusters of small ISIs purely by chance in addition to, and in one case dominating, a hidden binary process $C(t)$ having moderate regularity in its transitions. The first setting was the "null" setting illustrated in Figure 1; the second and third settings were inverse Gaussian and gamma switching processes with the ISI distributions either largely overlapping (setting 2) or clearly separated (setting 3) under the bursting and non-bursting states; the fourth setting produced non-bursting state ISIs from a mixture model, which is different than the HSMM and might confuse a burst detection algorithm; the fifth setting used an exponential distribution for the down-state durations, which makes the inter-burst durations maximally irregular and state identification more difficult to detect. More specifically, the processes we considered were as follows:

1. (**Null.**) Here we set $C(t) = 0$ for all $t \in [0, T]$. We generated spike trains from a pure renewal process with ISI distribution f_0^{ISI} given by an inverse Gaussian with shape 19.33 and mean 30.76, as in Figure 1.

2. (**IGovlp.**) The hidden state $C(t)$ follows a switching gamma process with $f_1^{\text{ITI}} = \text{Gamma}(10, 10/(25\text{ms}))$ and $f_0^{\text{ITI}} = \text{Gamma}(10, 10/(200\text{ms}))$. We chose the average duration of 25ms for an bursting state and 200ms for a non-bursting state according to the HSMM fit to the goldfish data. The bursting state ISIs were simulated from $f_1^{\text{ISI}} = \text{Gamma}(20, 20/(7\text{ms}))$. Our fit to the retinal ganglion data (described below) provided the choice of 7ms as the average bursting state length. We generated the non-bursting state ISIs according to the inverse Gaussian distribution of the Null process described above.

3. (**IGsep.**) Same as IGovlp but we instead generated the down state ISIs from an inverse Gaussian distribution with shape 150 and mean 50ms. This choice ensured that the two ISI distributions were well separated—the shortest non-bursting state ISIs were likely to be considerably larger than most bursting state ISIs.

4. (**Gmix.**) Same as IGovlp but we instead generated the down state ISIs from the mixture $(2/3)\text{Gamma}(10, 10/(10\text{ms}))+(1/3)\text{Gamma}(10, 10/(75\text{ms}))$. The first component of the mixture had a substantial overlap with f_1^{ISI}.

5. (**IGirr.**) Same as the IGsep setting but we instead generated the non-bursting state ITI of $C(t)$ from $f_0^{\text{ITI}} = \text{Exponential}(1/(200\text{ms}))$ distribution. Thus $C(t)$ was memoryless during the down state.

In all cases the spike trains were 10s in duration, with about 500 spikes per record on average.

We implemented HSMM burst detection with a switching gamma model given by

$$\begin{aligned} f_0^{\text{ISI}} &= \text{Gamma}(\alpha_0, \alpha_0/\mu_0) \\ f_1^{\text{ISI}} &= \text{Gamma}(\alpha_1, \alpha_1/\mu_1) \\ f_0^{\text{ITI}} &= \text{Gamma}(15, 15/\lambda_0) \\ f_0^{\text{ITI}} &= \text{Gamma}(15, 15/\lambda_1) \end{aligned}$$

with $\theta = (\alpha_0, \alpha_1, \mu_0, \mu_1, \lambda_0, \lambda_1)$. We modeled the ISI shapes α_0 and α_1 with a log-normal prior: $\log \alpha_i \sim \text{Normal}(\log(10), 1^2)$. Similarly, we modeled the ISI means (in ms) as $\log \mu_i \sim \text{Normal}(\log(20), 2^2)$ and the ITI means (in ms) as $\log \lambda_i \sim \text{Normal}(\log(100), 4^2)$. We first obtained an MCMC estimate to the burst probability $p_i = \Pr(s_i = 1 \mid y_{1:n})$ for each ISI. We labeled each ISI with $s_i = 1$ if p_i was at least as large as a chosen cutoff level, and 0 otherwise. We considered each contiguous string of states with $s_i = 1$ to be a burst. An estimate of the time the neuron spent in the bursting state is the sum of the ISIs with $s_i = 1$. Cutoffs used were $0.00, 0.01, \cdots, 1.00, 1.01$.

We implemented HMM burst detection by modifying the HSMM algorithm described above so that $f_i^{\text{ITI}} = \text{Exponential}(1/\lambda_i)$, $i = 0, 1$. Note that this HMM is more general than the switching Poisson process model because it allows non-Poisson firing within burst and non-burst periods. We implemented the PS and RS methods via the exhaustive surprise maximization (EMS) search technique of Gourevitch and Eggermont (2007) with a chosen surprise cutoff $-\log(\alpha)$ with α in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 05\}$. To maintain parity between all four methods, we did not count the bursts made of a single ISI and truncated f_0^{ISI} at the 75-th percentile of the observed ISI values. Both these limits are hard coded in the EMS implementation of Gourevitch and Eggermont (2007).

To evaluate burst count accuracy of the 4 methods we assessed the root mean squared error (RMSE). Each method requires a choice of cutoff, and each method will falsely identify some non-burst clusters of spikes as bursts. To make the methods comparable, we started with the null model, where there are no bursts, and picked cutoff values that produced roughly the same number of (false) bursts for each method; we then assessed the ability of the methods to track burst counts for the remaining 4 settings, where bursts were truly present—this is analogous to the standard statistical practice of fixing type I error and then examining power. Results are given in Table 1. In each of the non-null settings, the RMSE for HSMM was much smaller than those for PS and RS. HSMM and HMM had similar results except in the cases IGovlp and Gmix, where the RMSE for HMM was about 3 times larger than that for HSMM. Figure 3 shows a visual summary of burst detection by all four methods on a simulated spike train.

Comparisons of the type given in Table 1 are important but, because the cutoffs were fixed according to null-setting performance, they are only part of the story. We examined all methods across a range a cutoffs using ROC curves. For each method, for a set of cuttoffs, we evaluated both sensitivity (proportion of time in bursting state correctly identified) and specificity (proportion of time in non-bursting state correctly identified). For every method, as the cutoff is increased bursts become harder to detect, so that the specificity increases and the sensitivity decreases. The ROC curve plots sensitivity on the y-axis and $1-$ specificity on the x-axis. An optimal ROC curve begins at the origin, hugs the y-axis up to (0,1), and then moves to the point (1,1) along the line $y = 1$. ROC curves for the 4 methods in the 4 non-null settings are displayed in Figure 2.

We draw three general conclusions from the ROC analysis. First, HSMM performs as well as the other methods, and for the IGovlp and Gmix settings it performs better; the curves for HSMM were generally to the left of the others, meaning that HSMM was much more specific, spending significantly less time falsely identifying bursting states; and the curves for HSMM were also generally above those for the other methods, meaning that HSMM was generally spending more of the time correctly identifying bursting states. Second, in some cases PS and RS with particular cutoffs performed well, with RS outperforming PS in the IGovlp and Gmix settings, where the 4 methods clearly differed. However, it should be noted that a key feature of these curves is their upper left-most point, which is obtained for a particular "optimal" cutoff, and that cutoffs near this optimal cutoff have generally good sensitivity and specificity. Such cutoffs that allowed PS and RS to perform well in some settings produced unreliable results in other simulation settings. In the Null case we found that the ability of any method to handle false burstiness depends crucially on the cutoffs. Thus, in particular, in the last case, IGirr, which had memoryless switching from non-bursting to bursting state, all methods performed well according to the curves in Figure 2; but the cutoff values for which RS and PS attained the good sensitivity and specificity in IGirr were much smaller than those that would

Setting	HSMM	HMM	PS	RS
Null (0)	15	14	14	14
IGovlp (38)	1	3	10	14
IGsep (38)	0	1	4	16
Gmix (38)	8	23	8	15
IGirr (81)	5	3	106	88

Table 1: Root mean squared error (RMSE) of estimated burst counts according to 5 distinct model settings (first column). One hundred simulated spike trains were used for every setting. For all settings the average true burst counts are given in parentheses. To ensure the methods were comparable under the null setting, we used probability cutoffs of 0.5 for HSMM and 0.8 for HMM, and surprise cutoffs of $-\log(0.01)$ for PS and $-\log(0.05)$ for RS.

typically provide good performance in the Null setting. Third, in Gmix, which we designed to contradict the assumptions of the HSMM (the HSMM could not possibly provide fit the bimodal f_0^{ISI} because it is based on a unimodal gamma density), the HSMM performed extremely well.

One additional point worth mentioning is that the HMM usually produced a nice fit to the ISI histogram for the spike trains generated under Gmix. The reason is that the non-burst ISI distribution from the first component of the mixture was similar to the burst ISI distribution—and the HMM identified both kinds of ISIs as burst ISIs. Consequently, it did a poor job of burst detection but that was not apparent from its fit to the ISI histogram. In general, goodness of fit can not be judged solely from a fit to the ISI histogram.

4 Data Analysis

To illustrate our proposed method on real data we use a spike train recorded from a goldfish retinal ganglion cell neuron in vitro (Brown *et al.* 2004; Levine 1991; Iyengar and Liao 1997). The data include 971 spikes recorded over about 30 seconds. The plot of the spikes from this neuron in Figure 4 shows some apparent clusters of spikes with shorter intervals, and these clusters are separated by spikes with longer intervals between them.

We analyzed this spike train with HSMM, HMM, PS, RS as well as with the switching Poisson (SP) model which is a special case of HMM with the shape parameters of the ISI densities fixed at 1. The cutoffs were probability .5 for HSMM, HMM, and SP, and surprise $-\log .01$ for PS and $-\log 0.05$ for RS. Based on the simulations reported above, we would expect HSMM to be the most accurate, and the question remains whether the methods are appreciably different for this data set. HSMM found 127 bursts, whereas PS and

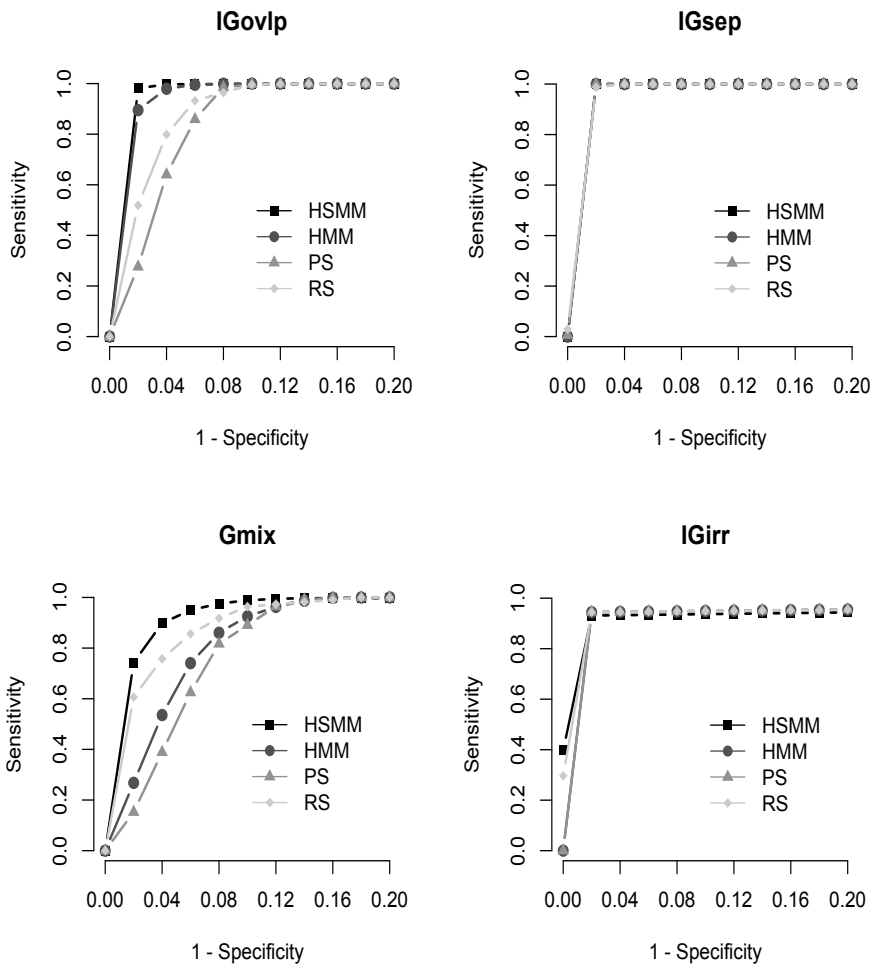

Figure 2: ROC curves for HSMM, HMM, PS and RS approximated from 100 simulated spike trains, for each of the 4 non-null simulation settings. Each (x, y) point on a curve corresponds to sensitivity and $1-$ specificity for a particular cutoff value.

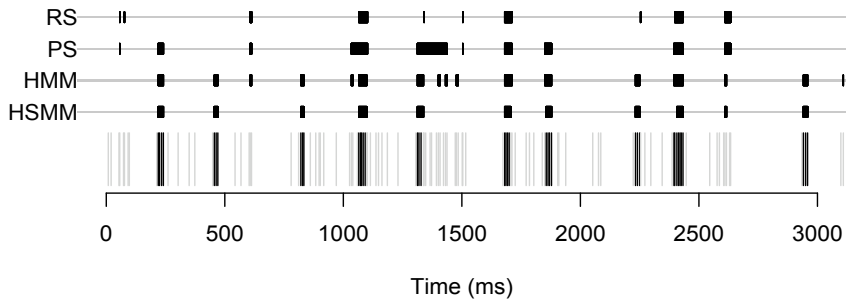

Figure 3: Burst detection with HSMM, HMM, PS and RS on a spike train simulated from a two-state model (IGovlp, see text; only the first 3s of a total of 10s are shown here). We generated ISIs in the bursting state from a gamma distribution with shape 20 and mean 7ms and in the non-bursting state from a 2:1 mixture of two gamma distributions with means 10ms and 75ms each with shape of 10. The gamma distributions controlling the durations of the bursting and non-bursting states each had shape 10 and with mean 25ms and 200ms respectively. The vertical bars show the spike times, color coded according to the hidden state of the preceding ISI: bursting state ISIs are marked as black, the non-bursting states one are marked as grey. The horizontal lines above the spikes represent (from bottom to top) burst detection with HSMM (probability cutoff: 0.5), HMM (probability cutoff: 0.5), PS (surprise cutoff: $-\log(0.01)$) and RS (surprise cutoff: $-\log(0.05)$). The thick dark strips on each line denotes the time intervals identified as bursts. Note that the HMM cutoff here was smaller than that in Table 1; this cutoff produced 33 false bursts for the Null simulation in Table 1.

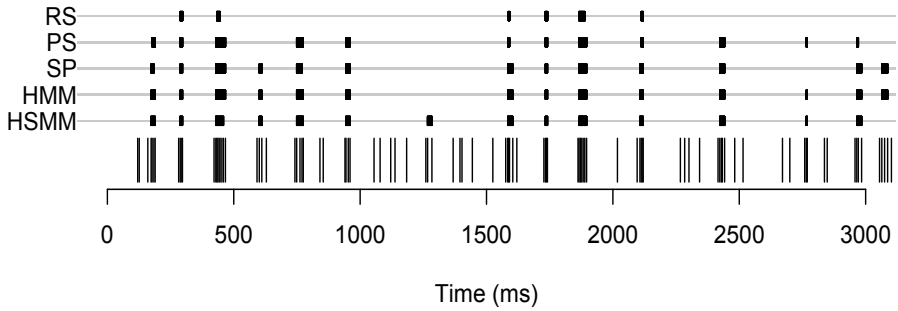

Figure 4: First 3 seconds of 971 spikes recorded over about 30 seconds in vitro, from a goldfish retinal ganglion cell neuron. It is obvious that groups of spikes with shorter ISIs are separated by individual spikes with longer ISIs.

Method	HSMM	HMM	SP	PS	RS
Number of bursts	127	133	122	91	51
Time (ms) spent in up state	3240	3482	3651	2273	902
(percentage)	(11%)	(12%)	(12%)	(8%)	(3%)
Average up state firing rate (Hz)	145	148	138	173	227
Average burst length (ms)	26	26	30	25	18

Table 2: Goldfish retinal ganglion cell data: summary of estimated bursting activity for each of 5 methods.

RS found many fewer—the latter turned out to be extremely conservative by comparison—and by several other burstiness measures the three hidden-state models gave similar results, but the surprise methods made the neuron appear less bursty. We checked the fit of HSMM, HMM and SP with a P-P plot. The P-P plot uses a basic result about the probability integral transform, which is that when a random variable X follows a theoretical probability distribution having a cumulative distribution function $F(x)$ the probability integral transformed random variable $F(X)$ is uniformly distribution on the interval (0,1). This implies that when a theoretical model describes the variation in a variable well, a plot of the probability integral transform of the ordered observations (fitted cumulants) against the corresponding probabilities for a uniform distribution (uniform cumulants) should fall close to the line $y = x$ (see also Brown *et al.* 2001). As shown in Figure 5, the HSMM appeared to give the best fit among these three hidden-state methods.

5 Discussion

PS is a fast and simple method of detecting bursts in extracellular spike trains, and RS is a useful modification of it. Because the goal is to identify a pair of unknown states (burst vs. non-burst), we developed algorithms for fitting point-process HSMMs that are based on hidden binary states. We wished to find out whether building a statistical model from this simple intuition would lead to improved burst detection results. The HSMM is more attractive, intuitively, than the hidden Markov switching Poisson process model both because neural spike trains often exhibit non-Poisson spiking behavior, and because the switching process may be different than the maximally irregular switching assumed by the HMM. HSMM code, written by the first author, is available on the web site of the last author.

Our results support the notion that PS can be an effective method when bursting states are well discriminated from non-bursting states, or when an experimenter has confidence in the choice of the PS cutoff value. Our results also illustrate the additional boost in performance that RS can provide in comparison

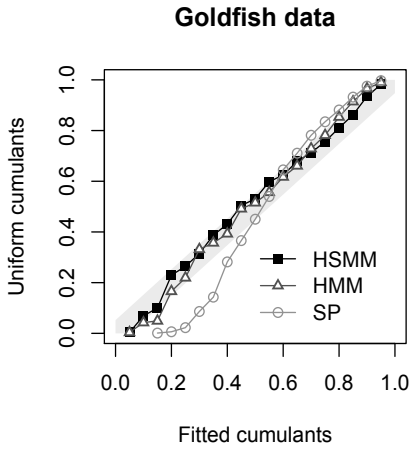

Figure 5: P-P plot (also called a KS plot) for HSMM, HMM and SP fits to goldfish data. A good fit should produce a plot along the diagonal line. The light grey band in the background represents the $\pm 1.63/\sqrt{n}$ pointwise band corresponding to the Kolmogorov-Smirnov criterion. HSMM fits reasonably well, and HMM is nearly as good, but the SP model fits poorly.

with PS. On the other hand, for cases in which the bursts are not clearly isolated, or the trade-off between identifying too many bursts or too few bursts is unclear, hidden state models are likely to be preferable. Overall, our simulation results indicated that HSMM performs as well as other methods and in some cases performs much better.

When we applied the three hidden-state and two surprise-based methods to the goldfish retinal ganglion spike train we found that the three hidden-state methods produced similar results according to several measures of burstiness. Consistently with some of our simulation results, this suggests that for many situations hidden Markov models should perform well. An additional idea is to (i) transform all ISIs by taking logarithms and then (ii) apply standard (off the shelf) two-state HMM software. We tried this, too, for our simulated data and found the results to be nearly the same as those for the HMM model. Thus, we expect this relatively easy method to be useful in many situations. In more difficult scenarios where there may be some subtlety in discriminating bursting and non-bursting states, we recommend the HSMM.

As we showed in detail, our implementation takes advantage of the semi-Markovian structure of HSMMs by extending the Gibbs sampling method of Chib (1996). It assumes that each ISI may be assigned to either the bursting state or the non-bursting state, and does not allow mid-ISI transitions. We

also explored an alternative method that does allow mid-ISI transitions, along the general lines used by Scott (1999) for switching Poisson process models. We found that in practice this alternative approach did not produce greatly different results, and we therefore preferred to present the more computationally efficient method. For convenience we used gamma distributions to describe both ISIs and ITIs. We would expect that in most cases, replacing gammas with alternative two-parameter families, or introducing more general families, would not have a large impact on results. However, the algorithm we described was formulated to allow such further generality. It is also possible that faster methods based on EM-type algorithms may be possible based on the same general idea of exploiting the Markovian structure inherent in these HSMMs. This is a topic for future research.

References

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Thishby, N., Vaadia, E. (1995), Cortical activity flips among quasi stationary states, *Proc. Nat. Acad. Sci.*, 92: 8616-8620.

Baum, L. E. and Petrie, T. (1966), Statistical Inference for probabilistic functions of finite state Markov chains, *Ann. Math. Stat.*, vol. 37, pp. 1554–1563.

Brown, E. N., Barbieri, R., Eden, U. T., and Frank, L. M. (2004), Likelihood Methods for Neural Spike Train Data Analysis, in *Computational Neuroscience: A Comprehensive Approach*, edited by Jianfeng Feng, CRC Press, pp.253-286.

Brown, B.N., Barbieri, R, Ventura V., Kass, R.E., and Frank, L.M. (2002), The time -rescaling theorem and its applications to neural spike train data analysis, *Neural Computation*, 14: 325–346.

Chen, Z., Vijayan, S., Barbieri, R., Wilson, M.A., Brown, E. N. (2008), Discrete- and continuous-time probabilistic models and inference algorithms for neuronal decoding of up and down states, submitted.

Chib, S. (1996), Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics*, 75: 79-97.

Cooper, D.C., Chung, S., Spruston, N. (2005), Output-Mode Transitions Are Controlled by Prolonged Inactivation of Sodium Channels in Pyramidal Neurons of Subiculum, *PLoS Biology*, Vol. 3, pp. 1123-1129.

Corner, M.A., van Pelt, J., Wolters, P.S., Baker, R.E., and Nuytinck, R.H. (2002), Physiological effects of sustained blockade of excitatory synaptic trans-

mission on spontaneously active developing neuronal networks-an inquiry into the reciprocal linkage between intrinsic biorhythms and neuroplasticity in early ontogeny, *Neurosci Biobehav Rev*, 26: 127-185.

Doiron, B., Chacron, M.J., Maler, L., Longtin, A., Bastian, J. (2003), Inhibitory feedback required for network oscillatory responses to communication but not prey stimuli. *Nature*, 421:539-543.

Gerkin, R.C., Yassin, K., Clem, R.L., Shruti, S., Kass, R.E., Barth, A.L. (2008), Seizure-dependent changes in cortical output: evaluation using an adaptive threshold for the detection of state changes in intracellular recordings, submitted to *Journal of Computational Neuronscience*.

Gourevitch, B., Eggermont, J. J. (2007), A nonparametric approach for detection of bursts in spike trains, *Journal of Neuroscience Methods*, 160:349-358.

Iyengar, S., and Liao, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cyber.*, 77, 289-295.

Izhikevich, E.M., Desai, N.S., Walcott, E.C., Hoppensteadt, F.C. (2003), Bursts as a unit of neural information: selective communication via resonance. *Trends in Neuroscience*, 26:161-167.

Koch, C., Biophysics of Computation: Information Processing in Single Neurons, Oxford University Press: New York, 2004.

Legendy, C. R. and Salcman, M. (1985), Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons, *Journal of Neurophysiology*, 53: 926-939.

Levine, M.W. (1991). The distribution of intervals between neural impulses in the maintained discharges of retinal ganglion cells. *Biol. Cybern.*, 65: 459-467.

Liu, J. S., Liang, F. and Wong. W. H. (2000). The use of multiple-try method and local optimization in Metropolis sampling, *Journal of the American Statistical Association*, 96(454): 561-573.

Lo, F.-S., Lu, S.-M., and Sherman, S. M. (1991), Intracellular and extracellular in vivo recording of different response modes for relay cells of the cat's lateral geniculate nucleus. *Exp. Brain* Rex 83: 3 17-328.

Martinson, J., Webster, H., Myasnikov, A., and Dykes, W. (1997), Recognition of temporaly structured activity in spontaneously discharging neurons in the somatosensory cortex in waking cats, *Brain Res* 750:129-140.

Rabiner, L.R. (1989), A Tutorial on hidden Markov modelss and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77 (2), pp. 257-286.

Scott, S. L.(1999), Bayesian Analysis of a Two-State Markov Modulated Poisson Process, *Journal of Computational and Graphical Statistics*, 8: 662-670.

Tam, D. C. (2002). An alternate burst analysis for detecting intra-burst firings based on inter-burst periods, *Neurocomputing*, 44-46 (C): 1155-1159.

Turnbull, L., Gross, G.. W. (2005), The string method of burst identification in neuronal spike trains, *Journal of Neuroscience Methods*, 145: 23-35.

Wilson, H.R., Cowan, J.D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons, *Biophys. J.*, 12:1-24.