Online Bayesian Inference for a Latent Marked Poisson Process: Applications in Spike Sorting

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 Introduction

1 paragraph on the following:

1. opening
2. challenge
3. approach
4. resolution

2 Model

Our data is a time-series of multielectrode recordings $\mathbf{X} \equiv (\mathbf{x}_1, \cdots, \mathbf{x}_T)$, and consists of T recordings from C channels. The set of recording times lie on regular grid with interval length Δ, while $\mathbf{x}_t \in \mathbb{R}^C$ for all t. This time-series of electrical activity is driven by an unknown number of neurons and we want to... outline scientific goals. We let the number of neurons be unbounded, though only a few of the infinite neurons dominate. These neurons contribute the majority of the activity in any finite interval of time; however, as time passes, the total number of observed neurons increases (Justify?). The neurons themselves emit continuous-time voltage traces, with the outputs of all neurons superimposed and discretely sampled to produce the recordings $\mathbf{X}$. At a high level, we model the output of each neuron as a series of idealized spikes smoothed with appropriate kernels (the latter determines the shape of each action potential). We describe this in detail, starting first with the model for a single channel recording $X \equiv (x_1, \cdots, x_T)$.

2.1 Modelling a single neuron output

We model the spiking activity of each neuron as stationary and memoryless, with its set of spike times distributed as a homogeneous Poisson process. Comment on refractoriness or leave for discussion/future work. Similarly on the generalization to inhomogeneity? The neurons themselves are heterogeneous, with r_i the (unknown) firing rate for neuron i. Call the ordered set of spike times of the ith neuron E_i; then the time between successive elements of E_i is exponentially distributed with mean $1/r_i$. We write this as

$$E_i \sim \text{PoissProc}(r_i) \tag{1}$$

The actual electrical output of a neuron is not a binary event; instead each spiking event is a smooth voltage perturbation about a resting state. This perturbation forms the shape of the spike (without any

loss of generality, we set the resting state to zero). (figure? better biological description? comment on how we preprocess the data to get zero mean?). While the spike shape varies across neurons as well as across different spikes of the same neuron, each neuron has its own characteristic distribution over shapes. Figure? We let $\theta \in \Theta$ parametrize this distribution, and whenever neuron i emits a spike, we draw a voltage trace independently from the corresponding distribution. This is then offset to the time of the spike, and the complete output of the neuron is the superposition of all these spike waveforms. (Figure?) More concretely, we model each spike shape as a linear combination of a dictionary of K basis functions $A \equiv (A_1(t), \cdots, A_K(t))$, shared across all neurons. For the ith neuron, the jth spike $e_{ij} \in E_i$, is associated with a random K-dimensional weight vector $\tilde{y}_{ij}$, and the shape of this spike is given by the weighted sum $\sum_{k=1}^{K} \tilde{y}_{ijk} A_k(t)$. We let $\tilde{y}_{ij}$ be Gaussian distributed, with $\theta_i \equiv (\mu_i, \Sigma_i)$ determining its mean and variance. Then, at any time t, the output of neuron i is

$$x_i(t) = \sum_{j=1}^{|E_i|} \sum_{k=1}^{K} \tilde{y}_{ijk} A_k(t - e_{ij}) \tag{2}$$

The total signal recorded $x(t)$ at any electrode is the superposition of the outputs of all neurons. Define $E = \cup_{i=1}^{\infty} E_i$ as the (ordered) superposition of the spike times of all neurons. Furthermore, let $n(j)$ be the neuron to which the jth element of E belongs, and let $p(j)$ index the position of the jth spike of E in the spike train $E_{n(i)}$ of neuron $n(i)$ (so that $e_j = e_{n(j)p(j)}$). Then, we have that

$$x(t) = \sum_{i=1}^{\infty} x_i(t) = \sum_{j=1}^{|E|} \sum_{k=1}^{K} y_{jk} A_k(t - e_j) \tag{3}$$

where

$$y_j \equiv \tilde{y}_{n(j)p(j)} \sim N(\mu_{n(j)}, \sigma_{n(j)}) \tag{4}$$

From the superposition property of the Poisson process [1], the overall spiking activity E is a Poisson process with rate $R = \sum_{i=1}^{\infty} r_i$. The signal $x(t)$ is a functional of a marked Poisson process, where the jth event is labelled by the neuron to which it is assigned $(n(j))$, and the shape of its spike waveform (y_j). From the properties of the Poisson process, we have that the marks $n(j)$ are i.i.d. distributed with $P(n(j) = i) = \frac{r_i}{R}$. Given $n(j)$, y_j is distributed as in equation 4.

2.2 Completely random measures (CRMs)

In this work, we take a nonparametric approach, letting the number of neurons be unbounded (so that $n(i) \in \{1, 2, \cdots\}$). Since only a finite number of spikes are observed in any finite interval, the total rate R must also be finite; moreover, as we described earlier, we want this to be dominated by a few r_i. A natural framework that captures these modelling requirements is that of completely random measures [2]. Completely random measures are stochastic processes that form flexible and convenient priors over infinite dimensional objects like probability distributions, hazard functions etc. These have been well studied in the Bayesian nonparametrics and machine learning community, and there exists a wealth of literature on their theoretical properties, as well as on computational approaches to posterior inference.

Recall that each neuron is characterized by a pair (r_i, θ_i); the former characterizes the distribution over spike times, and the latter over spike shapes. We map the infinite collection of pairs $\{(r_i, \theta_i)\}$ to an atomic measure on Θ:

$$R(\mathrm{d}\theta) = \sum_{i=1}^{\infty} r_i \delta_{\theta_i} \tag{5}$$

For any subset Θ of Θ, the measure $R(\Theta)$ equals $\sum_{\{i:\theta_i \in \Theta\}} r_i$. We allow $R(\cdot)$ to be random, modelling it as a realization of a completely random measure. Such a random measure has the property that for any two disjoint subsets Θ_1 and $\Theta_2 \in \Theta$, the measures $R(\Theta_1)$ and $R(\Theta_2)$ are independent. This distribution over measures is induced by a distribution over the infinite sequence of weights (the r_i's), and a distribution over the sequence of their locations (the θ_i's). For a CRM,

the weights r_i form the jumps of a Lévy process [3], and their distribution is characterized by a Lévy intensity $\rho(r)$. The locations θ_i are drawn i.i.d. from a base probability measure $H(\theta)$; we let this be the conjugate normal-Wishart distribution. As it typical, we assume these to be independent (though this is not necessary). if there's space, I can elaborate on the construction of the CRM from its Levy measure, though this is not necessary

The CRM we choose is the Gamma process (ΓP); this has Lévy intensity $\rho(r) = r^{-1}\exp(-r\alpha)$. The Gamma process has the convenient property that the total mass $R \equiv R(\Theta) = \sum_{i=1}^{\infty} r_i$ is Gamma distributed (and thus conjugate to the Poisson process prior on E). The Gamma process is also closely connected with the Dirichlet process [4], which will prove useful later on. Other choices of the Lévy intensity can capture greater uncertainty in the number of neurons active in any finite interval, power-law behaviour etc. In any case, our overall model is then:

$$R(\mathrm{d}\theta) \sim \Gamma\mathrm{P}(\alpha, H(\theta)) \tag{6}$$

$$E_i \sim \text{PoissProc}(r_i) \quad i \text{ in } 1, 2, \cdots \tag{7}$$

$$\tilde{y}_{ij} \sim N(\mu_i, \Sigma_i) \quad i, j \text{ in } 1, 2, \cdots \tag{8}$$

$$x_i(t) = \sum_{j=1}^{|E_i|} \tilde{y}_{ij} A_j(t - e_{ij}) \tag{9}$$

$$X = \sum_{i=1}^{\infty} x_i \tag{10}$$

It will be more convenient from the point of inference to work with the marked Poisson process representation of equations 3 and 4. The superposition process E is a rate R Poisson process, and under a Gamma process prior, R has a conjugate Gamma distribution with shape and scale parameters 1 and α respectively. As we saw, the labels $n(\cdot)$ assigning events to neurons are drawn i.i.d. from a normalized Gamma process $G(\mathrm{d}\theta)$:

$$G(\mathrm{d}\theta) = \frac{r_j}{R} \tag{11}$$

$G(\mathrm{d}\theta)$ is a random probability measure that belongs to a class called a normalized random measures [5]; for the Gamma process, this is a draw from the Dirichlet process. For the j spike, given its neuron assignment $n(i)$, its shape vector is drawn from a normal distribution with parameters $(\mu_{n(j)}, \Sigma_{n(j)})$. Thus the weight vectors are distributed according to a Dirichlet process mixture model, with the neurons forming clusters. This insight allows us to marginalize out the infinite-dimensional rate vector R, and assign spikes to neurons via the Chinese restaurant process (CRP). Under the CRP, the jth spike is assigned to one of the earlier neurons with probability proportional to the number of earlier spikes assigned to that neuron. It is assigned to a new neuron with probability α. Unlike most applications with observe the outputs of a CRP, we observe a functional of it. Furthermore, (again, for the Gamma process), the random probability measure G is independent of the total mass $R(\Theta)$. We thus have the following model equivalent to the one above:

$$R \sim \text{Gamma}(1, \alpha) \tag{12}$$

$$G(\mathrm{d}\theta) \sim \text{DP}(\alpha) \tag{13}$$

$$E \sim \text{PoissProc}(R) \tag{14}$$

$$n(j) \sim G, \quad j = 1, \cdots, |E| \tag{15}$$

$$y_e \sim N(\mu_{n(e)}, \Sigma_{n(e)}), \quad j = 1, \cdots, |E| \tag{16}$$

$$x(t) = \sum_{j=1}^{|E|} \sum_{k=1}^{K} y_{jk} A_k(t - e_j) \tag{17}$$

2.3 A discrete-time approximation

In the previous paragraphs, we described a continuous-time voltage output by a neuron. Our data on the other hand consists of recordings at a discrete set of times. While it is possible to make inferences about the continuous-time process that underlies these discrete recordings, in this work we restrict ourselves to discrete-time inferences. Towards this, we start by providing a discrete-time

approximation to the model above. This follows easily from the marked Poisson process characterization of the model. Recall first the Bernoulli approximation to the Poisson process: a sample from a Poisson process with rate R can be approximated by discretizing time at a granularity Δ, and assigning each interval an event independently with probability $R\Delta$. This approximation becomes exact as Δ tends to 0.

This suggests the following approximation at a time resolution Δ. Draw the random Poisson process rate R drawn from a Gamma$(1, \alpha)$ distribution. Simultaneously, draw a random probability measure G from a Dirichlet process. Assign an event to an interval independently with probability $R\Delta$, and to each event, assign a random mark drawn from the DP. Given the marks, we can evaluate the recordings at each time.

2.4 Noise and nonstationarity

The signal recorded by an electrode is the neuron output corrupted by noise, we model this noise as independent of the signal, additive and Gaussian. However, rather than modelling the noise as independent across time bins, we model it as a first-order autoregressive process. This can capture effects like the movement of electrodes during the experiment. Furthermore, rather than keeping the cluster parameters fixed, we model these as AR processes as well, capturing the evolution of the neuron shape with time.

2.5 Modelling multielectrode recordings

3 Inference

We perform inference in an online manner [6]. As observations arrive, our inference algorithm decides whether or not a new spike is present, which neuron (cluster) to assign that spike to, as well as the shape of the spike waveform. On the other hand, our algorithm maintains a posterior distribution over the cluster parameters that characterize the distribution over shapes. Having identified the location and shape of spikes from earlier times, we subtract these from the observations treat the residual as an observation from a DP mixture model. The cluster assignment of earlier spikes determines the seating arrangement of customers in the Chinese restaurant associated with the DP. Given the corresponding distribution over parameters, $p(\theta)$, we decide whether there is an underlying spike, which cluster it is assigned to, and what the shape of that spike is. We simultaneously update the distribution over parameters of clusters. Assume each spike waveform spans W time intervals. Define the residual at time t as $X_t - \sum_{i=1}^{W} A$. At time t, let y_t represent the shape of the action potential. Letting $\tilde{x}_t$ be the observation at time t, we have

$$z_t \sim Bern(p) \tag{18}$$

if $z_t == 1$

$$\gamma_t|\gamma_{1:t-1} \sim CRP \tag{19}$$

$$\theta_t|\gamma_t = i \sim \mathcal{N}(\theta_{t-1}, \Sigma) \tag{20}$$

References

[1] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[2] J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

[3] Ken ti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1990.

[4] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

[5] L.F. James, A. Lijoi, and I. Pruenster. Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.*, 36:76–97, 2009.

[6] L. Wang and D.B. Dunson. Fast bayesian inference in dirichlet process mixture models. *Journal of Computational & Graphical Statistics*, 2009.

[7] D A Henze, Z Borhegyi, J Csicsvari, A Mamiya, K D Harris, and G Buzsaki. Intracellular feautures predicted by extracellular recordings in the hippocampus in Vivo. *J. Neurophysiology*, 2000.

[8] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 1998.

[9] Jan Gasthaus, Frank Wood, Dilan Gorur, and Yee Whye Teh. Dependent dirichlet process spike sorting. *Advances in neural information processing systems*, 21:497–504, 2009.

[10] Jason S Prentice, Jan Homann, Kristina D Simmons, Gašper Tkačik, Vijay Balasubramanian, and Philip C Nelson. Fast, scalable, Bayesian spike identification for multi-electrode arrays. *PloS one*, 6(7):e19884, January 2011.

4 Results

4.1 Performance on partial ground truth data

Experiments were performed on a subset of the publicly available[1] dataset described in [7]. We used the dataset d533101 that consisted of an extracellular tetrode and a single intracellular electrode. The recording was made simultaneously on all electrodes and was set up such cell with the intracellular electrode was also recorded on the extracellular array implanted in the hippocampus of a rat.

The intracellular recording is relatively noiseless and gives nearly certain firing times of the intracellular neuron. The extracellular recording contains the spike waveforms from the intracellular neuron as well as an unknown number of additional neurons. The data is a 4-minute recording at a 10kHz sampling rate and preprocessed with a high-pass filter at 800 Hz.

Each algorithm gives a clustering of the detected spikes. In this dataset, we only have a partial ground truth, so we are only able to analyze whether a spike comes from the intracellular (IC) recording or not. In these experiments, we define a detected spike to be an IC spike if the IC recording has a spike within .5ms of the detected spike. We define the cluster with the greatest number of intracellular spikes as a the "IC cluster." In order to run replicates of the data to get uncertainty estimates, we split the data into random 2 minute segments. The results on the IC cluster are shown in Figure 1a (plot will be changed to reflect this). The single-channel experiments were all run on channel 2 (the results were nearly identical for all channels). The spike detections for the offline methods were given by using a threshold at 3 times the noise standard deviation [8]. (Need to do more comparisons and descriptions) and windowed at a size $P = 30$ (check if P matches with Vinayak notation). The action potential window was set at 30, and PCA was used to reduce the space to K=3 for the experiments. The results were very similar for K=2, K=3, and K=4. There were 3742 spikes detected with the threshold, and 753 of those corresponded to the intracellular spikes. The PCA learned from the thresholded spikes was used as an input to the online algorithms.

The online algorithms were all run with weakly informative parameters (add parameters once I get vinayak's notation). The parameters were insensitive to minor changes. Running time in unoptimized MATLAB code for 4 minutes of data was 31s was a single channel and 3 minutes for all 4 channels on a 3.2GHzIntel Core i5 machine with 6GB of memory. We show the inferred y_k fordetected spikes in the 2-PCA space in Figure 1b.

Sometimes, we will not have a good estimate of the low-dimensional space to use. However, since spikes are very similar it is possible to use a pre-learned dictionary from a similar dataset. To show robustness with the choice of similar dictionaries, we used a dictionary learned offline on a similar tetrode implanted in the hippocampus of a different rat. The results were similar and are shown in Figure 1a (will be added).

[1] http://crcns.org/data-sets/hc/hc-1/

4.2 Time-Varying Waveform Adaptation

As as been demonstrated previously (cite paninski), the waveform shape of a neuron may change over time. The mean waveform over time for the intracellular neuron is shown in Figure 2a. For the IC cluster, it is of interest to investigate the properties of the true positives, the false positives, as well as the IC spikes that are missed by the algorithm. Errorbar plots for these groups are shown in Figure 2b; this figure demonstrates the great similarity of the true positives and the false positives, while the missed positives (spike waveforms not detected or not in the IC cluster) have high variability and a different mean shape.

4.3 Overlapping Spike Detection

These algorithms are capable of detecting overlapping spikes as we go. We find that we detect 124 overlapping spikes out of 3593 total detected spikes.

4.4 Multi-Electrode Array

In the tetrode case the waveform undoubtedly appears on all channels at once; it is possible to concatenate the channels to jointly process the data [9]. When the action potential will only appear on a subset of channels it is nice to allow the action potential to vary in a low-dimensional subset in each of the channels instead of a low-dimensional subset over all the channels. [10]

We use processed data from novel NeuroNexus devices implanted in the rat motor cortex. The data was recorded at 32.5kHz in freely-moving rats, and the data was first processed by high-pass filtering at 800Hz. The first device we consider is a set of 3 channels of data shown in Figure 4a. The neighboring electrode sites in these devices have 30 μm between electrode edges and 60 μm between electrode centers. These devices are close enough that a locally-firing neuron could appear on multiple electrode sites (cite that paper on action potential overlap), and neighboring channels warrant joint processing. The second, 8-channel device is shown in Figure **??**, and has similar properties to the first device.

The top 3 clusters found in the first 10 minutes of data on the 3-channel device are shown in Figures 3b, 3c, and 3d. The waveform in channel 3 is very similar for the waveforms in Figure 3b and 3c, and would be difficult to separate if we were analyzing each channel individually; we gain the ability to distinguish the waveforms by looking at all the channels simultaneously. The top 3 clusters found in the 8-channel device can be found in Figures 4b, 4c, and 4d.

perhaps add comments about time-evolution and the false positives we avoid by using multi-channel analysis

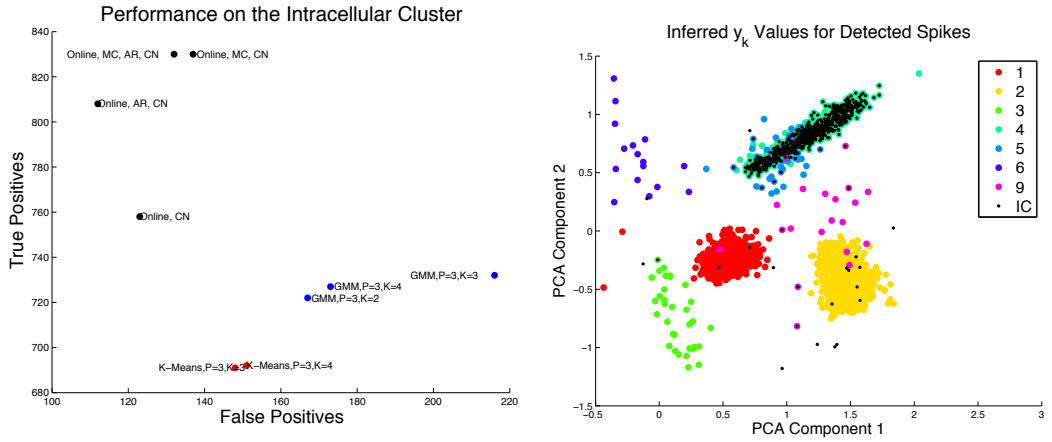
5 Discussion

1 paragraph on:

1. summary
2. relate work - including Franke and Pillow
3. future work - including adaptive stimulus design & decoding

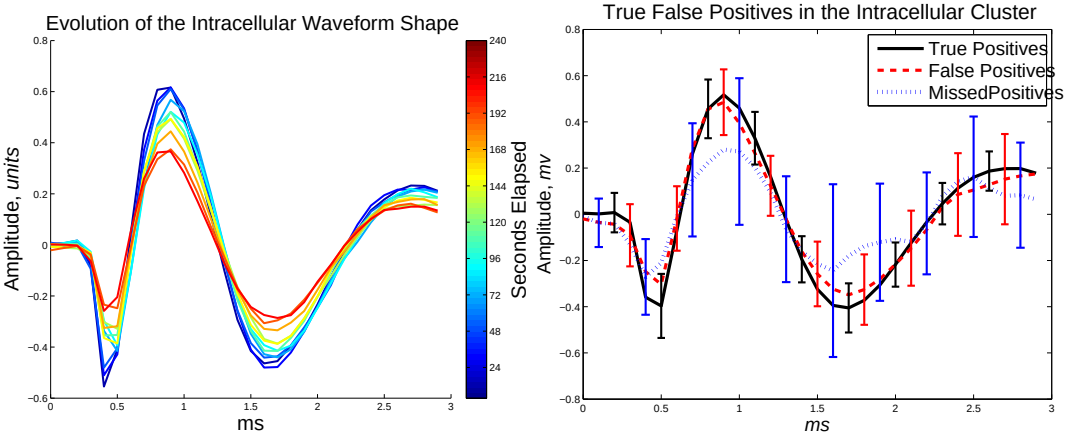
References

[1] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.



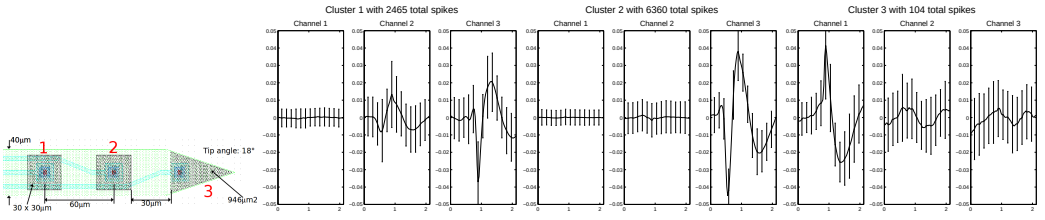
(a) (b)

Figure 1: Results on the d533101 dataset. (a) This shows the number of true positives versus the number of false positives in the intracellular cluster. The methods that combine the clustering and the detection have similar numbers of false positives but greater numbers of true positives. (b) Inferred y_t for detected spikes from the online Kalman filter method.



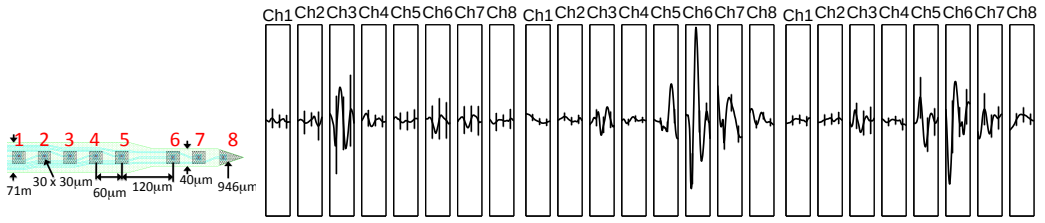
(a) (b)

Figure 2: (a) Mean IC waveforms over time. Each colored line represents the mean of the waveform averaged over 24s and the color gives the elapsed time. This neuron decreases in amplitude over the period of the recording. (b) Errorbar plots of the true positives and the false positives in the IC cluster. While the false positives have slightly more variability, the mean shape for the false positives and the true positives is nearly identical. The true misses have a significantly lower amplitude as well as high variability



(a) (b) (c) (d)

Figure 3: (a)3 electrode device showing local proximity of electrodes with channel indexes in large, red numbers. (b,c,d) Top 3 most prevalent waveforms. Note in (a) we have a waveform that appears on both channel 2 and channel 3, wheras in (b) the waveform only appears in channel 3. If only channel 3 was used, it would be difficult to separate the waveform in (a) and (b).



(a) (b) (c) (d)

Figure 4: (a)8 electrode device showing local proximity of electrodes with channel indexes in large, red numbers. (b,c,d) Top 3 most prevalent waveforms.

[2] J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

[3] Ken ti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1990.

[4] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

[5] L.F. James, A. Lijoi, and I. Pruenster. Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.*, 36:76–97, 2009.

[6] L. Wang and D.B. Dunson. Fast bayesian inference in dirichlet process mixture models. *Journal of Computational & Graphical Statistics*, 2009.

[7] D A Henze, Z Borhegyi, J Csicsvari, A Mamiya, K D Harris, and G Buzsaki. Intracellular features predicted by extracellular recordings in the hippocampus in Vivo. *J. Neurophysiology*, 2000.

[8] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 1998.

[9] Jan Gasthaus, Frank Wood, Dilan Gorur, and Yee Whye Teh. Dependent dirichlet process spike sorting. *Advances in neural information processing systems*, 21:497–504, 2009.

[10] Jason S Prentice, Jan Homann, Kristina D Simmons, Gašper Tkačik, Vijay Balasubramanian, and Philip C Nelson. Fast, scalable, Bayesian spike identification for multi-electrode arrays. *PloS one*, 6(7):e19884, January 2011.