BMC
Neuroscience

Scalable, Bayesian, multi-electrode spike sorting

Jason Prentice[1*], Jan Homann[1], Kristina Simmons[2], Gasper Tkacik[1], Philip Nelson[1], Vijay Balasubramanian[1,2]

Multi-electrode array technology provides an efficient means of recording from many neurons. However, as arrays become larger, a greater computational burden falls on the spike-sorting algorithm. We have developed a new method, that scales linearly with array size, for sorting multi-electrode signals from retinal ganglion cells. We believe that our techniques represent progress toward solving many of the source separation problems that will become ubiquitous as large multi-electrode arrays become commonplace.

The broad outline of our method is to identify spikes in the raw data, cluster a subset, generate template waveforms, then fit the templates to all the data using an iterative Bayesian algorithm. Spikes are identified as spatiotemporally connected patches of threshold-crossing voltage samples. The spike waveform is taken from a fixed neighborhood centered on the electrode having the peak voltage within each patch. This approach allows for segmentation of simultaneous, yet spatially separated, events and prevents the waveform dimensionality from increasing with array size.

Next we cluster a small subset of spikes. We use an existing algorithm, OPTICS, which orders the waveforms so that similar spikes are placed together. This linear ordering makes cluster boundaries easily distinguishable by the user. We have built a GUI in which the manual cluster cutting can be performed efficiently. For each cluster, we align the waveforms and take the median to get templates.

The primary obstacle in fitting templates to the data is the presence of overlapping spikes, which distort the observed waveforms. One established approach to the problem is to simply fit single templates, then subtract the best fit and iterate. However, the amplitude of the observed spike can differ substantially from the template, producing errors upon subtraction. We avoid this problem by allowing the amplitude of the template to vary; this is most naturally incorporated into a Bayesian framework. We model each waveform as a linear superposition of templates with Gaussian-distributed amplitudes, plus correlated Gaussian noise. We then seek the most probable template, spike time, and amplitude given the data. The spatial localization of spikes narrows down the list of candidate templates, speeding up the algorithm. The Gaussian amplitude prior allows the amplitudes to be marginalized analytically, avoiding an explicit sum.

We have tested the method on many data sets recorded with a dense 30-electrode array, under a variety of stimulus conditions. It always produces very low error rates. Tests with larger arrays, different species, and synthetic data are ongoing.

Author details
[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Institute of Neurological Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA.

* Correspondence: jprentic@sas.upenn.edu
[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA