A Novel Algorithm for Automated Cell Finding

Jose L. Vega and Stuart N. Baker
Institute of Neuroscience, Newcastle University, UK. Email: jose.vega@ncl.ac.uk ; stuart.baker@ncl.ac.uk

Introduction

Simultaneous multi-channel single unit recordings are a key tool to investigate neural function. We have developed a simple algorithm to automate the process of finding extracellular single units in real time using a multi-channel microelectrode drive. The algorithm uses brief (1s) periods of recording to assess recording quality. The microelectrode is advanced until well isolated units are detected.

The autonomous algorithm consists of four steps:

1. Detection and alignment of spike waveforms using thresholding;
2. Dimensional reduction of the differences between noisy spike waveforms using Principal Component Analysis (PCA);
3. Determining the number of different single-units (Bayesian model selection);
4. Finally, a qualitative measure of unit separation is estimated, also using a Bayesian approach.

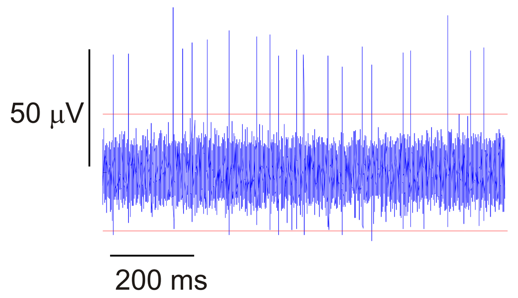


Fig. 1. Simulated signal (1 s) and detection of spikes using thresholding.

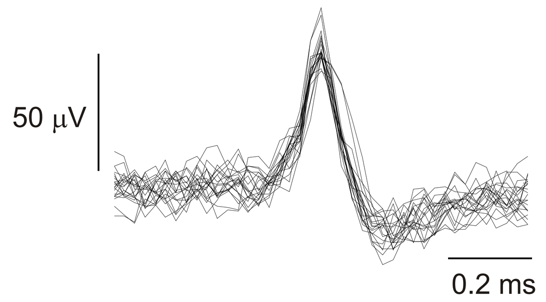


Fig. 2. Alignment of spike waveforms.

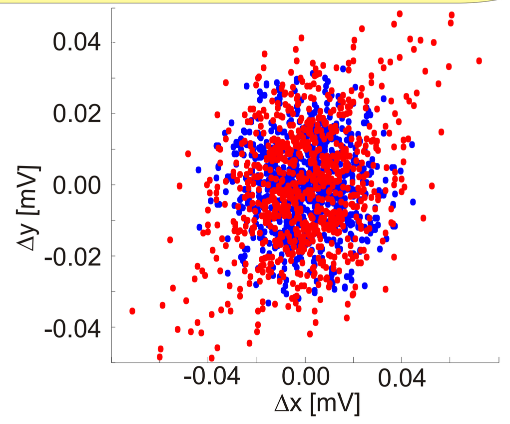


Fig. 3. Pairwise differences of: spike waveforms (red points) and noise (blue points) compressed to two dimensions using PCA.

Pairwise differences of the spike waveforms

Data in Fig. 3 are obtained by all the differences of the spike waveforms:

$$\vec{\Delta}_{ij}(t) = \vec{W}_i(t) - \vec{W}_j(t) = \vec{S}_i(t) + \vec{N}_i(t) - (\vec{S}_j(t) + \vec{N}_j(t))$$

$$\vec{\Delta}^{noise}_{ij}(t) = \vec{W}'_i(t) - \vec{W}'_j(t) = \vec{N}'_i(t) - \vec{N}'_j(t)$$

where $\vec{W}_i(t)$ is the i-th recorded spike waveform comprising the underlying spike shape $\vec{S}_i(t)$ plus noise $\vec{N}_i(t)$, and $\vec{W}'_i(t)$ is a signal vector with no spikes.

If the spikes come from one cell (Fig. 2):

$$\vec{S}_i(t) = \vec{S}_j(t)$$

$$\vec{\Delta}_{ij}(t) = \vec{N}_i(t) - \vec{N}_j(t) \ .$$

In Fig. 3, the differences of the noise spike waveforms $\vec{\Delta}_{ij}(t)$ and the differences of the noise $\vec{\Delta}^{noise}_{ij}(t)$ exhibit the same distribution.

If the spikes come from two (Fig. 4) and three (Fig. 5) different cells, $\vec{\Delta}_{ij}(t)$ will show multiple clusters:

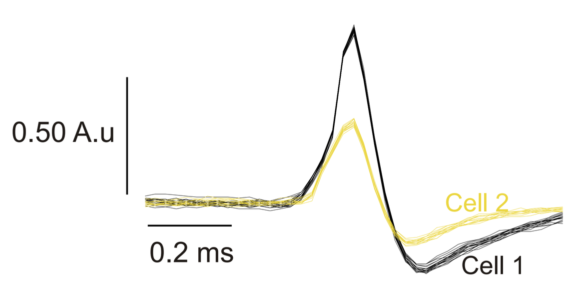


Fig. 4a. Spike waveforms from two different neurons.

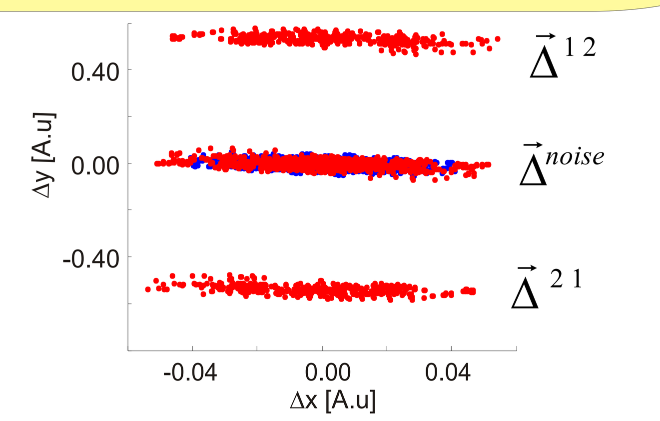


Fig. 4b. Pairwise differences between the spike waveforms (red points) and noise (blue points) from Fig. 4a.

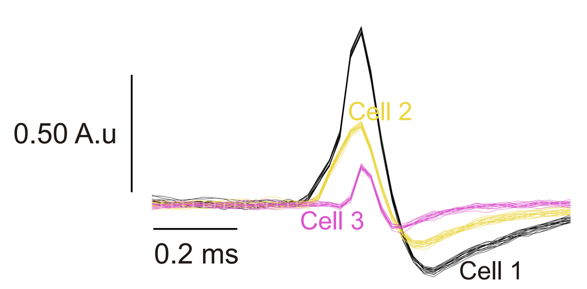


Fig. 5a. Spike waveforms from three different neurons.

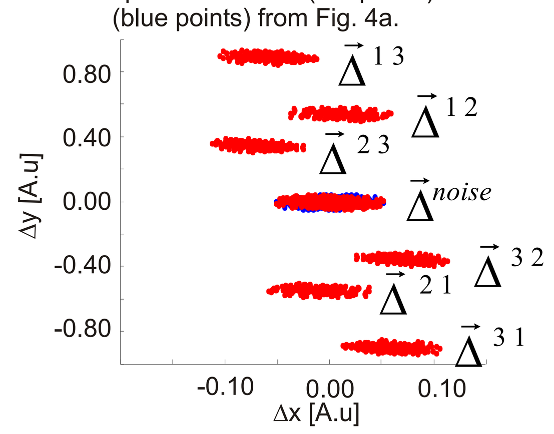


Fig. 5b. Pairwise differences between the spike waveforms (red points) and noise (blue points) from Fig. 5a.

Bayesian Model Selection (BMS)

We are interested to know which model M_k, with k = 1, 2 or 3 different spikes has maximum likelihood given $\vec{\Delta}_{ij}(t)$ = data (red cluster points in Fig. 3, 4b and 5b).

We calculate the posterior probability $P(M_k|data)$ using a Bayesian Model Selection (BMS) based on Bayes' theorem:

$$P(M_k|data) = \frac{P(M_k)P(data|M_k)}{P(data)}, \text{ assuming } \frac{P(M_k)}{P(data)} = const.$$

where the const. is determined by the normalization constraint:

$$\sum_{k=1}^{3} P(M_k|data) = 1 \ .$$

$P(data|M_k)$ is the likelihood function of the model M_k given by:

$$P(data|M_k) = \int_{\vec{\theta}_k} P(\vec{\theta}_k|M_k)P(data|\vec{\theta}_k, M_k)d\vec{\theta}_k$$

where $P(\vec{\theta}_k|M_k)$ and $P(data|\vec{\theta}_k, M_k)$ are the prior probability and likelihood function of the parameter vector $\vec{\theta}_k$ respectively.

We assume a uniform $P(\vec{\theta}_k|M_k) = \frac{1}{\Delta\vec{\theta}_k}$, and $P(data|\vec{\theta}_k, M_k)$ is determinated by:

$$P(data|\vec{\theta}_k, M_k) = \prod_{u=1}^{U} P(data_u = (\Delta x, \Delta y)_u \,|\vec{\theta}_k, M_k),$$ where U is the number of two-dimensional points, and $P(data_u|\vec{\theta}_k, M_k)$ obey a probability density function depending of the model M_1, M_2, or M_3 respectively:

$$P(data_u = (\Delta x, \Delta y)_u \,|\vec{\theta}_1, M_1) = G(data_u, \vec{\theta}_1), \quad with \quad \vec{\theta}_1 = (\vec{\mu} = \vec{\mu}^{noise}, \vec{\sigma} = \vec{\sigma}^{noise}) \quad (no \quad free \quad parameters)$$

$$P(data_u = (\Delta x, \Delta y)_u \,|\vec{\theta}_2, M_2) = \lambda G(data_u, \vec{\theta}_1) + \lambda_{12} G(data_u, \vec{\mu} + \vec{\mu}_{12}, \vec{\sigma}) + \lambda_{12} G(data_u, \vec{\mu} - \vec{\mu}_{12}, \vec{\sigma}), \quad \lambda = 1 - 2\lambda_{12} \quad and \quad \vec{\theta}_2 = (\vec{\theta}_1, \lambda_{12}, \vec{\mu}_{12}) \quad (free \quad parameters = 3)$$

$$P(data_u = (\Delta x, \Delta y)_u \,|\vec{\theta}_3, M_3) = \lambda G(data_u, \vec{\theta}_1) + \lambda_{12} G(data_u, \vec{\mu} + \vec{\mu}_{12}, \vec{\sigma}) + \lambda_{12} G(data_u, \vec{\mu} - \vec{\mu}_{12}, \vec{\sigma}) + \lambda_{13} G(data_u, \vec{\mu} + \vec{\mu}_{13}, \vec{\sigma}) + \lambda_{13} G(data_u, \vec{\mu} - \vec{\mu}_{13}, \vec{\sigma}) \ +$$
$$\lambda_{23} G(data_u, \vec{\mu} + \vec{\mu}_{23}, \vec{\sigma}) + \lambda_{23} G(data_u, \vec{\mu} - \vec{\mu}_{23}, \vec{\sigma}), \quad \lambda = 1 - 2\lambda_{12} - 2\lambda_{13} - 2\lambda_{23} \ with \quad 0 \le \lambda_{12} \le \frac{1}{2} \quad 0 \le \lambda_{13} \le \frac{1}{2} - \lambda_{12} \quad 0 \le \lambda_{23} \le \frac{1}{2} - \lambda_{12} - \lambda_{13} \quad and \quad \vec{\theta}_3 = (\vec{\theta}_2, \vec{\mu}_{13}, \vec{\mu}_{23}) \quad (free \quad parameters = 7).$$

We assume $G(data_u, \vec{\theta}_1)$ obey the same multivariate Gaussian distribution of $data^{noise} = \vec{\Delta}^{noise}_{ij}(t)$, Fig. 3:

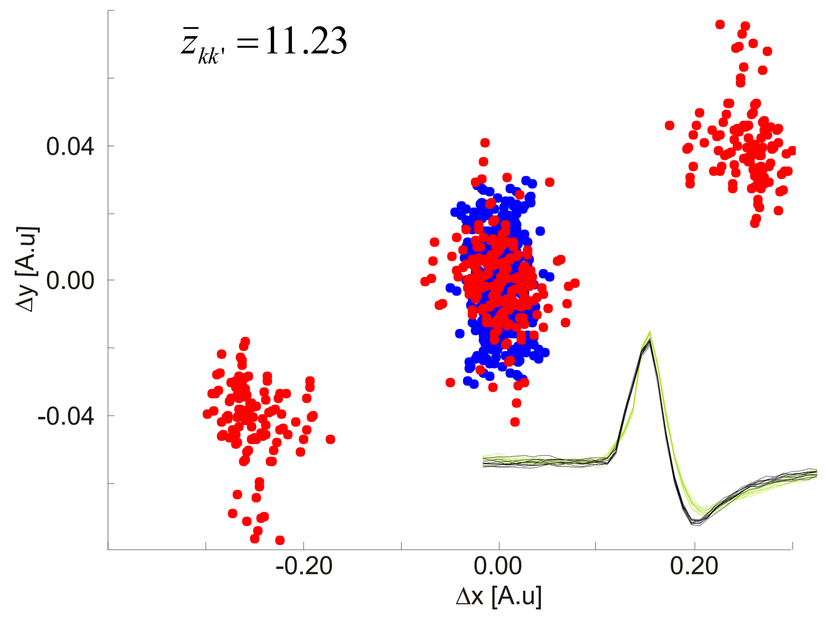
$$G(\vec{\Delta}^{noise}_u, \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|}\exp\left(-\frac{1}{2}(\vec{\Delta}^{noise}_u - \vec{\mu})^T \Sigma^{-1} (\vec{\Delta}^{noise}_u - \vec{\mu})\right)$$

, where d = 2 dimensions and Σ is the covariance matrix (2x2).

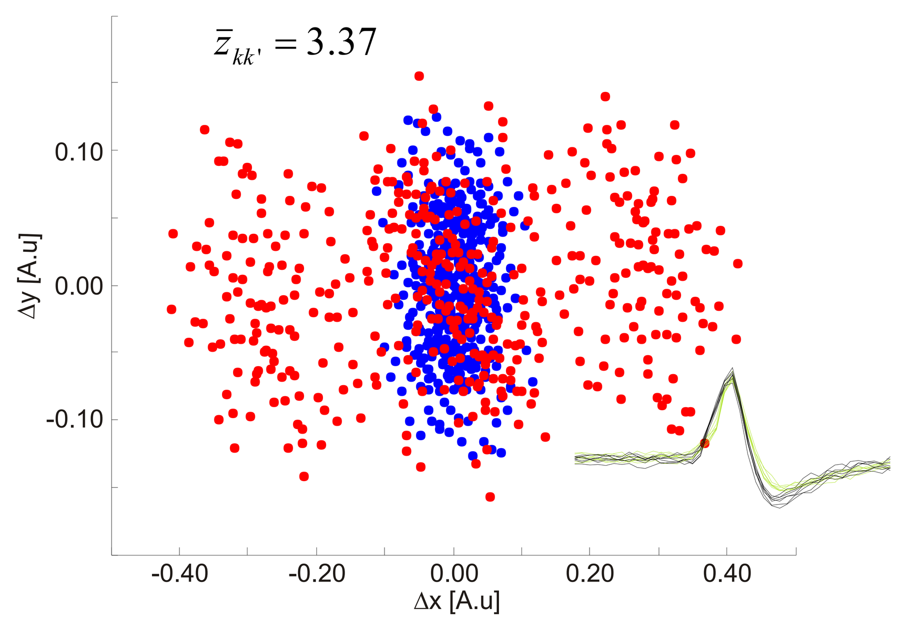
Separation Measure

To decide if two or three spikes are cleanly discriminable or not, we use a separation measure:

$$\bar{z}_{kk'} = \int_{\vec{\theta}_k} z_{kk'} P(\vec{\theta}_k|M_k)P(data|\vec{\theta}_k, M_k)d\vec{\theta}$$, where $$z_{kk'} = \frac{|\vec{\mu}_{kk'}|^2}{\vec{\sigma}_{Noise}\vec{\mu}_{kk'}}$$

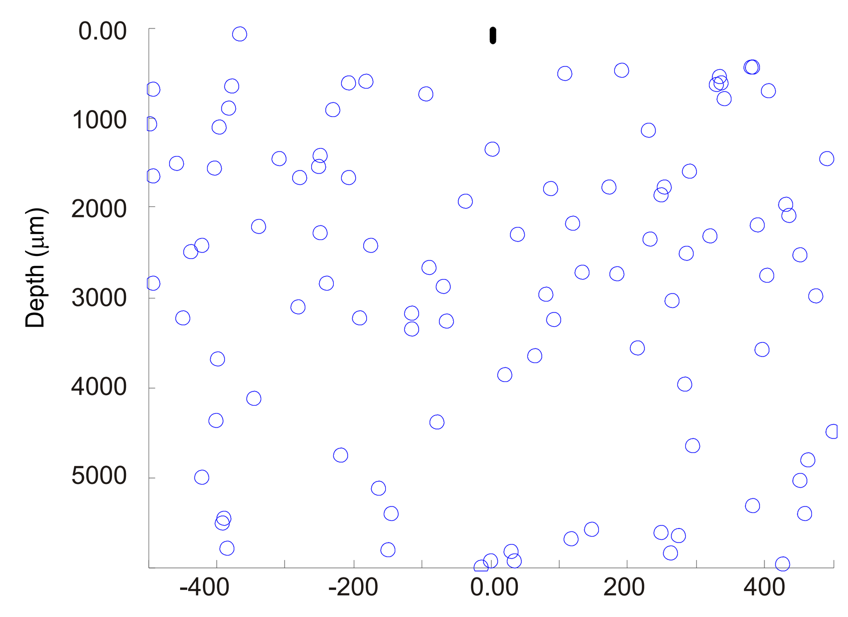


Pairwise differences of the spike waveforms showing well separated clusters at a $\bar{z}_{kk'} > 10$.

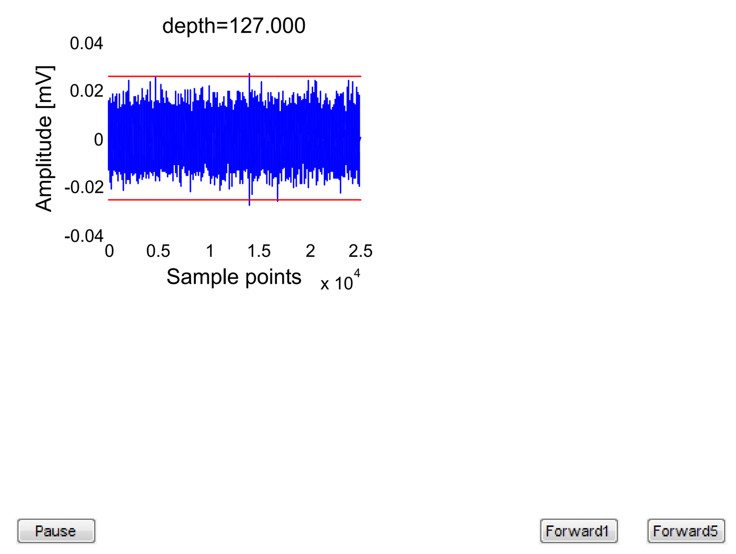


Pairwise differences of the spike waveforms showing a poor cluster separation at a $\bar{z}_{kk'} < 5$.

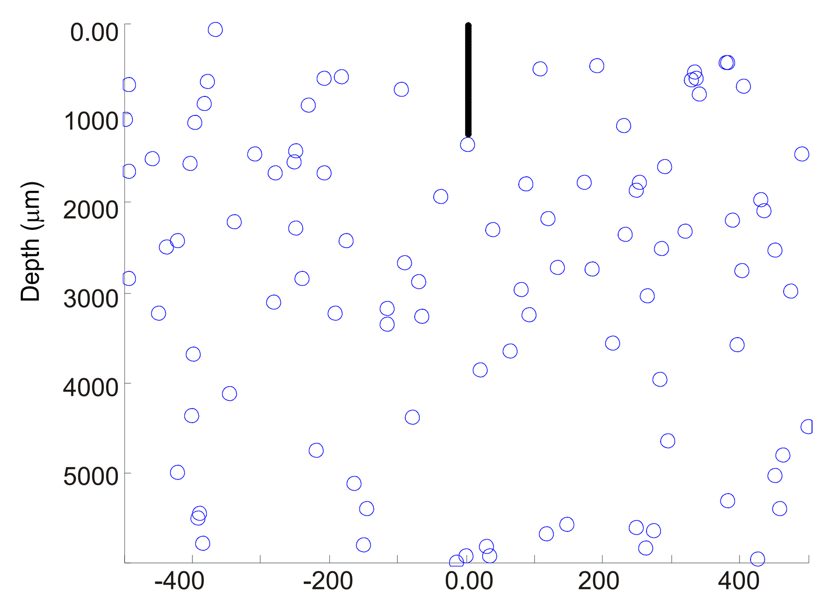
Simulation of Cell Finding



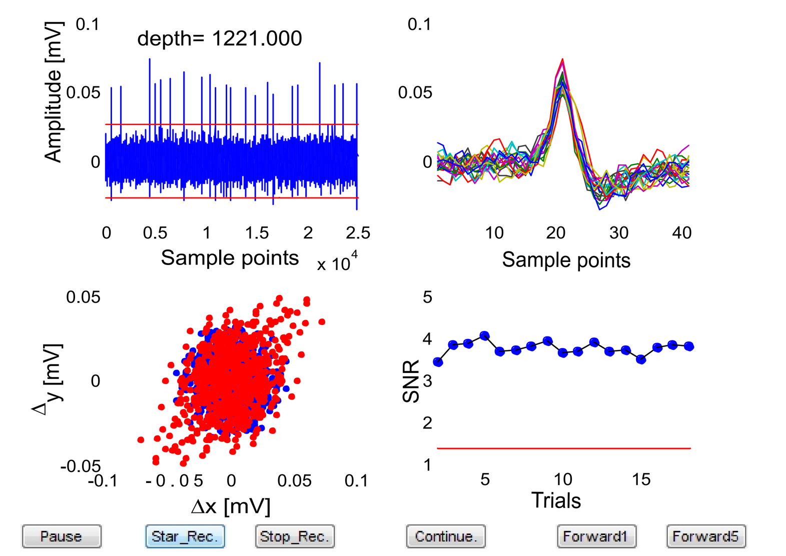
Simulation with 100 neurons. We can see how the electrode (black line) at depth = 127 μm is far away from neurons.



At this depth, the electrode does not record any single-unit acivity and the software decided to move the electrode down.

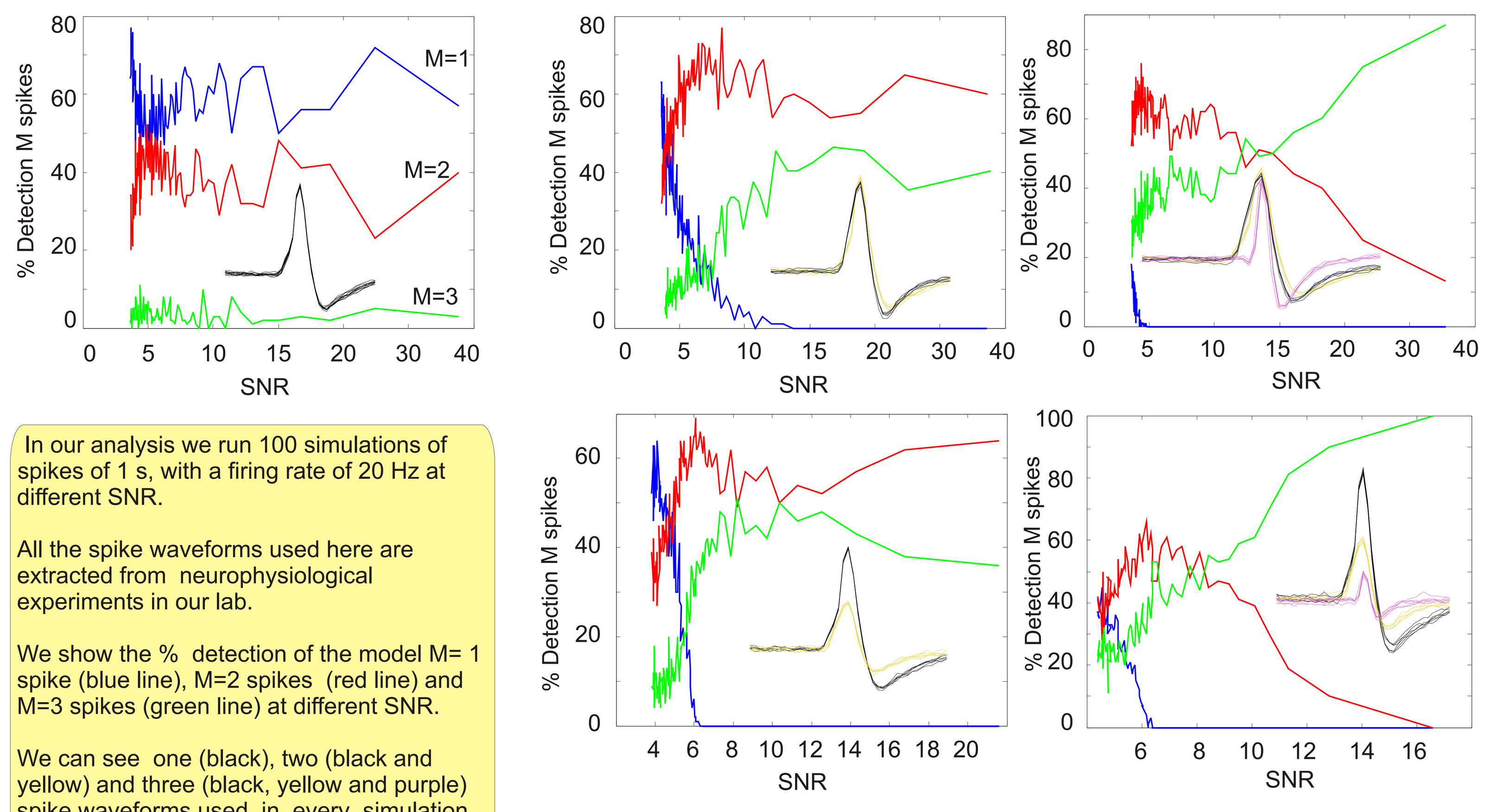


At a depth of 1221 μm the electrode is very close to one neuron.



Single unit activity is detected and the software stops electrode movement. The Signal-Noise-Ratio (SNR) right bottom plots the rise above the SNR threshold (redline), and the blue circles show detection of one spike.

Parametric Analysis



In our analysis we run 100 simulations of spikes of 1 s, with a firing rate of 20 Hz at different SNR.

All the spike waveforms used here are extracted from neurophysiological experiments in our lab.

We show the % detection of the model M= 1 spike (blue line), M=2 spikes (red line) and M=3 spikes (green line) at different SNR.

We can see one (black), two (black and yellow) and three (black, yellow and purple) spike waveforms used in every simulation.

The spike waveforms have different shape and same (top) and different (bottom) size.

Conclusions

The algorithm differs from more usual approaches to unit clustering, in that it does not attempt to assign cluster identities to individual spikes. Consequently, the algorithm is effective on only short records, and fast enough to implement in real time.