Matched Subspace Detector Based Feature Extraction for Sorting of Multi-Sensor Action Potentials

Shun Chi Wu, A. Lee Swindlehurst and Zoran Nenadic

Abstract— This paper proposes a novel matched subspace detector (MSD) based algorithm for extracting discriminant features from multi-sensor measurements of extracellular action potentials (APs) to facilitate their subsequent separation according to the neuron of origin. The method does not require the construction of AP templates, and is therefore suitable for unsupervised AP sorting applications. In addition, detailed simulations show that the proposed algorithm outperforms existing single-sensor based feature extraction approaches.

I. INTRODUCTION

Sorting of extracellular action potentials (APs) refers to the process of detecting and classifying neural activity from extracellular recordings which are contaminated by background noise with unknown properties. It typically represents the first step in the analysis of extracellularly recorded multi-unit activity. In practice, manual sorting of APs in large volumes of experimental data may be prohibitively time-consuming, and so automated procedures for AP sorting have become essential. An automated AP sorting algorithm can roughly be divided into three steps: (1) AP detection, (2) feature extraction, and (3) AP clustering. In the feature extraction step, information is extracted that is useful for discriminating between the different AP clusters. This step is crucial since it removes redundant information in the input data so that clustering algorithms can work efficiently. A key question is what features are most beneficial for clustering. The three most common feature categories discussed in the literature are: (1) AP shape-related features [1], [2], such as AP height, width, peak-to-peak amplitude, inter-AP interval, and first-order derivative, (2) principal components [1], [3], [4] and (3) wavelet coefficients [5].

One common characteristic of the above mentioned features is that they only capture "temporal" information since they are obtained by processing single-sensor measurements. However, AP sorting based only on temporal features is challenging because neurons with similar geometries and about equal distances to the electrode can generate very similar AP waveforms and therefore similar features. To overcome the problem, multi-sensor probes (e.g. tetrodes) for extracellular recordings of neuronal activity have been suggested [4], [6], [7], [8]. The simplest way to use data from multi-sensor probes is to apply standard feature extraction techniques to all of the channels individually, and then combine all the extracted features as inputs for clustering. Other approaches like using neuron location estimates [8] or independent components [6] as feature vectors for clustering have also been reported. To locate a neuron with multi-sensor measurements, a "forward model" for describing how the AP propagates through the conductive neural media is necessary. Monopole models are the simplest and the most referred to [8], [9], but these models only provide a very simplified view of the energy propagation. Using an over-simplified model for localization based on such assumptions can lead to location estimates with high variance and in turn to poor clustering outcomes. Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents. However, a strong assumption in ICA is that the subcomponents are statistically independent [10], [6], which cannot be assumed in practice.

In this paper, we propose a feature extractor for revealing discriminant information from the data recorded by multi-sensor probes. Our technique is based on the concept of the matched subspace detector (MSD) [11], [12], which allows both spatial and spatio-temporal features to be extracted. Unlike algorithms based on estimating the neuron location, the spatial information is extracted without using a forward propagation model. Furthermore, there is no need to use AP templates, and therefore our method is suitable for unsupervised AP sorting.

II. STANDARD DATA MODEL AND ASSUMPTIONS

Assume that the APs in an AP sequence can be detected by existing approaches (e.g. [13]) and that AP bundles can be constructed based on the detection information. Here, an AP bundle is a multi-channel, time-aligned data stack extracted from measurements of all the channels that contains data corresponding to a single AP. The procedure for feature extraction based on MSD for a given AP bundle is depicted in Figure 1.a. Notice that MSD provides two kinds of features: (1) a spatial signature $\hat{a}$ which describes the spatial structure of the potential fields at the sensor locations and (2) a spatio-temporal feature vector $\hat{c}$ which is the coefficient vector of some orthonormal transform (e.g., the discrete wavelet or cosine transform, DWT or DCT) of the weighted combination of the data in all channels of the AP bundle. Feature reduction and clustering are then performed to obtain AP clusters based on $\hat{a}$ or $\hat{c}$ as shown in Figure 1.b. Finally, a suitably chosen criterion is evaluated to determine which type of clusters (spatial or spatio-temporal) will provide optimal discrimination.
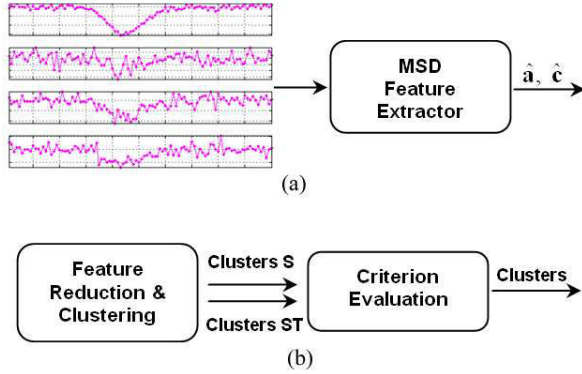
Fig. 1. Procedure for feature extraction and AP clustering based on MSD. In (a), the 4 traces of signals represent a simulated AP bundle acquired by a tetrode.

A. Standard Data Model

Following the mathematical model of array signal processing, an AP bundle is modeled as

$$\mathbf{X} = \mathbf{S} + \mathbf{W} = \mathbf{a}\mathbf{v}^T + \mathbf{W}, \qquad (1)$$

where $\mathbf{S} \in \mathbb{R}^{m \times n}$ are the noise-free multi-sensor measurements of the AP generated by the target neuron with m being the number of sensors and n being the number of time samples in the bundle. The $m \times 1$ vector $\mathbf{a}$ is the spatial signature of the target neuron, the $n \times 1$ vector $\mathbf{v}$ is the AP of the target neuron, and $\mathbf{W} \in \mathbb{R}^{m \times n}$ is composed of background neural and sensor noise. The model in (1) assumes that each sensor receives a scaled version of the same AP waveform $\mathbf{v}$, so that $\mathbf{S}$ is a rank-one matrix. This is only an approximation due to the distributed nature of the AP source, but is a reasonably accurate representation of the data. By vectorizing the data matrix, we have

$$
\begin{aligned}
\mathbf{x} &= \mathbf{s} + \mathbf{w} = \mathbf{v} \otimes \mathbf{a} + \mathbf{w} \\
&= \mathbf{\Phi}\mathbf{c} \otimes \mathbf{a} + \mathbf{w} = (\mathbf{\Phi} \otimes \mathbf{a})\mathbf{c} + \mathbf{w}, \qquad (2)
\end{aligned}
$$

where $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{w}$ are $mn \times 1$ vectors and $\otimes$ denotes the Kronecker product. The term $\mathbf{v} = \mathbf{\Phi}\mathbf{c}$ models the AP in the absence of any specific information about the waveform, with matrix $\mathbf{\Phi} \in \mathbb{R}^{n \times p}$ representing a chosen orthonormal basis and $\mathbf{c} \in \mathbb{R}^{p \times 1}$ representing the corresponding coefficient vector. Modeling the AP signal in this way not only provides the possibility of a compact representation for the AP but also eliminates the need for AP templates, which enables unsupervised spike sorting. Although $\mathbf{\Phi}$ can be any orthonormal basis, one with a compact support such as the wavelet basis [5] is preferred in general since APs are transient pulses.

B. Assumptions

The following assumptions are adopted in this paper: (A1) APs in an AP sequence can be detected using existing methods (e.g. [13]) and the corresponding AP bundles can be constructed. (A2) Each AP bundle contains an action potential from only one target neuron, i.e., there are no overlapping APs. (A3) AP bundles from two different neurons have distinct spatial signatures which are unique to the neuron locations.

III. MATCHED SUBSPACE DETECTOR (MSD) BASED FEATURE EXTRACTION

A. MSD Using Single-Sensor Measurements

The matched filter is a special case of MSD obtained when $m = 1$. For single-sensor measurements, using the matched filter for feature extraction leads to maximizing the following objective function [14]:

$$\hat{\mathbf{c}} = \arg\max_{\mathbf{c}} \; \|\mathbf{s}^T \mathbf{x}'\|^2 \quad \text{s.t. } \|\mathbf{c}\| = \|\mathbf{x}'\|, \qquad (3)$$

where $\mathbf{x}'$ is the whitened or interference-suppressed measurements which can be obtained for example using a filter based on an autoregressive moving average (ARMA) model [15]. In this special case with $m = 1$, the vector $\mathbf{a}$ in (2) reduces to a scalar gain factor that is absorbed in $\mathbf{c}$. The solution to (3) is obtained by choosing $\hat{\mathbf{c}} = \mathbf{\Phi}^T \mathbf{x}'$, which is simply an orthonormal transformation of the interference-suppressed measurements. If the Haar wavelet basis is used for $\mathbf{\Phi}$, $\hat{\mathbf{c}}$ will be exactly the same as the feature vector obtained in [5].

B. MSD Using Multi-Sensor Measurements

MSD is a natural extension of matched filtering when dealing with multi-sensor measurements ($m > 1$). MSD solves the following maximization problem [11], [12]:

$$
\begin{aligned}
\hat{\mathbf{a}}, \hat{\mathbf{c}} &= \arg\max_{\mathbf{a}, \mathbf{c}} \; \|\mathbf{s}^T \mathbf{P}_H \mathbf{x}\|^2 \\
&\text{s.t. } \|\mathbf{a}\| = 1 \;,\; \|\mathbf{c}\| = \|\mathbf{P}_H \mathbf{x}\| = \alpha,
\end{aligned}
\qquad (4)
$$

where the $mn \times mn$ matrix $\mathbf{P}_H$ is a projection matrix used for suppressing the spatially and temporally correlated background neural noise, and where the norms of $\mathbf{a}$ and $\mathbf{c}$ are constrained to avoid a trivial solution. The scaling for $\mathbf{c}$ is chosen to be data dependent, in order to preserve information about the AP amplitude. This is consistent with the solution to (3) in [5]. To construct $\mathbf{P}_H$, multi-sensor measurements of the background noise are necessary, which can be accomplished by recording signals when the neuron of interest is inactive. One could then set $\mathbf{P}_H = \mathbf{P}_T \otimes \mathbf{P}_S$, where $\mathbf{P}_T \in \mathbb{R}^{n \times n}$ and $\mathbf{P}_S \in \mathbb{R}^{m \times m}$ are projection matrices that are approximately orthogonal to the temporal and spatial subspaces, respectively, in which the interference is observed to lie [16].

Using this approach, the cost function in (4) can be rewritten as

$$\|\mathbf{s}^T \mathbf{P}_H \mathbf{x}\|^2 = \mathbf{c}^T \mathbf{Y}(\mathbf{a})\mathbf{Y}^T(\mathbf{a})\mathbf{c} \quad \text{s.t. } \|\mathbf{c}\| = \alpha, \qquad (5)$$

with $\mathbf{Y}(\mathbf{a}) = (\mathbf{\Phi}^T \mathbf{P}_T \otimes \mathbf{a}^T \mathbf{P}_S)\mathbf{x} \in \mathbb{R}^{p \times 1}$. It is easy to show that maximizing (5) can be achieved by choosing $\mathbf{c} = \alpha(\mathbf{Y}(\mathbf{a})/\|\mathbf{Y}(\mathbf{a})\|)$, and the spatial signature $\hat{\mathbf{a}}$ can be found through

$$
\begin{aligned}
\hat{\mathbf{a}} &= \arg\max_{\mathbf{a}} \; \|\mathbf{Y}(\mathbf{a})\|^2 = \arg\max_{\mathbf{a}} \; \|\mathbf{a}^T \mathbf{P}_S \mathbf{X} \mathbf{P}_T \mathbf{\Phi}\|^2 \\
&= \arg\max_{\mathbf{a}} \; \|\mathbf{a}^T \mathbf{P}_S \mathbf{X} \mathbf{P}_T\|^2 \quad \text{s.t. } \|\mathbf{a}\| = 1.
\end{aligned}
\qquad (6)
$$

The solution of (6) is thus obtained by letting $\hat{\mathbf{a}}$ be the left singular vector of $\mathbf{P}_S \mathbf{X} \mathbf{P}_T$ with the largest singular value. Finally, the estimate of the coefficient vector $\mathbf{c}$ is obtained by

$$\hat{\mathbf{c}} = \alpha \hat{\mathbf{c}}' / \|\hat{\mathbf{c}}'\|, \qquad (7)$$

with $\hat{\mathbf{c}}' = \mathbf{Y}(\hat{\mathbf{a}}) = \boldsymbol{\Phi}^T \mathbf{P}_T \mathbf{X}^T \mathbf{P}_S \hat{\mathbf{a}}$. Notice that $\mathbf{P}_T \mathbf{X}^T \mathbf{P}_S \hat{\mathbf{a}}$ is a weighted combination of interference-suppressed data within an AP bundle, and $\hat{\mathbf{c}}'$ is, for example, the DWT of $\mathbf{P}_T \mathbf{X}^T \mathbf{P}_S \hat{\mathbf{a}}$, if the DWT basis is used.

IV. FEATURE REDUCTION AND CLUSTERING

In many applications, the feature extraction process provides too much data to be efficiently handled by the subsequent clustering algorithm, and a further reduction in feature dimensionality is required even though the MSD feature extractor has discarded most of the redundant information in the measurements. Moreover, the MSD provides two kinds of features for further clustering, but it remains to be determined which feature vector ($\hat{\mathbf{a}}$ or $\hat{\mathbf{c}}$) is best for clustering. In this section, we address these two issues.

Assuming that there are N_d detected spike bundles whose features have been extracted using MSD, we have, by stacking the feature vectors, $\mathbf{A} \in \mathbb{R}^{m \times N_d} = [\hat{\mathbf{a}}_1 \dots \hat{\mathbf{a}}_{N_d}]$ and $\mathbf{C} \in \mathbb{R}^{p \times N_d} = [\hat{\mathbf{c}}_1 \dots \hat{\mathbf{c}}_{N_d}]$, which are referred to as the spatial and spatio-temporal feature matrices, respectively. The mean vector $\mathbf{m}_A \in \mathbb{R}^{m \times 1}$ of $\mathbf{A}$ is defined as $\mathbf{m}_A = \frac{1}{N_d} \sum_{j=1}^{N_d} \hat{\mathbf{a}}_j$. Similarly, for $\mathbf{C}$, we have the mean vector $\mathbf{m}_C \in \mathbb{R}^{p \times 1}$ which can be calculated in the same fashion as $\mathbf{m}_A$. The variance vector $\mathrm{Var}_A \in \mathbb{R}^{m \times 1}$ of $\mathbf{A}$ is defined as $\mathrm{Var}_A = diag\{\frac{1}{N_d} \sum_{j=1}^{N_d} (\hat{\mathbf{a}}_j - \mathbf{m}_A)(\hat{\mathbf{a}}_j - \mathbf{m}_A)^T\}$ and $diag\{\}$ is used to extract the matrix diagonal. The number of spatio-temporal features can be reduced by choosing the features that correspond to the largest N_f components of $\mathbf{m}_C$, similar to the approach used in [5]. This means that the most "energetic" features are selected for clustering. For the spatial features, we choose those corresponding to the largest N_f components of Var_A since features are more spread in these coordinates. AP clustering is then performed based on these reduced feature sets.

For a set of classes $\Omega = \{\omega_1, \omega_2, \cdots, \omega_C\}$ clustered based on feature vectors $\mathbf{f}_i \in \mathbb{R}^{N_f \times 1}$ ($i = 1, \cdots, N_d$), with class ω_j containing n_j feature vectors from the jth class and $\sum_{j=1}^{C} n_j = N_d$, a scatter separability criterion can be calculated through [17]

$$J = tr(\mathbf{S}_M^{-1} \mathbf{S}_B), \qquad (8)$$

where $\mathbf{S}_M = \frac{1}{N_d} \sum_{i=1}^{N_d} (\mathbf{f}_i - \mathbf{m})(\mathbf{f}_i - \mathbf{m})^T$ and $\mathbf{S}_B = \sum_{j=1}^{C} p_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T$ are mixture and between-class scatter matrices with $\mathbf{m}_j = \frac{1}{n_j} \sum_{\mathbf{f}_i \in \omega_j} \mathbf{f}_i$, $\mathbf{m} = \frac{1}{N_d} \sum_{i=1}^{N_d} \mathbf{f}_i$ and $p_j = \frac{n_j}{N_d}$. The J quantity in (8) is a well-known measure of the separability among classes and is widely used as a criterion for feature selection. Based on the reduced dimension $\mathbf{A}$ or $\mathbf{C}$, there will be two different scatter separability criteria obtained, and thus we simply choose the clusters that have larger J values as the final clusters. Note that clustering using spatial and spatio-temporal features separately may not be optimal in general; we have chosen this method due to its simplicity, but better performance may be obtained with a combined approach.

V. SIMULATION

A. Simulation Setup

A simulation setup similar to that used in [5] was adopted in this paper. Two identical cell models were placed in parallel to mimic the parallel organization of the cells in cortical columns. In a local coordinate system, the somata of these cells were centered at [0, 0, 0] μm and [50, 0, 0] μm, respectively. To account for complex nonlinearities in the kinetics of voltage-gated channels, the simulations were carried out with a variable step size in NEURON [18], [5], ranging between 0.02 ms and 1.0 ms and the data sampling rate was set to 20 kHz. Two sampling directions mimicking the process of recording from a sulcus and regular cortical layers with several individual sampling tracks along each of these two directions were simulated as in [5]. Tracks in the vertical and horizontal directions (hereafter called V and H tracks, respectively) started in a plane located at $z = 100$ and $y = 80$ with ranges of electrode movement $z \in [100, -100]$ and $y \in [80, -120]$, respectively. The tetrode geometry specifications were adopted from a commercial supplier [19], and the distance between sensors was between 20 and 40 μm. Moreover, only white noise was added to the source signals to generate the noisy measurements. In general, neural noise may be spatially correlated and nonstationary. The variance σ^2 of the white noise was specified by a given signal-to-noise ratio (SNR), defined as $(\|\mathbf{S}_1\|_F^2 + \|\mathbf{S}_2\|_F^2)/(2\|\mathbf{W}\|_F^2)$ with $\mathbf{S}_i$ representing the noise-free multi-sensor measurements of the AP generated by the ith neuron. To obtain statistically significant results, 100 trials were conducted for each neuron at each position along the sampling tracks.

B. Results and Discussion

To evaluate the performance of the proposed feature extractor, the K-means clustering algorithm was used and features obtained by the Haar wavelet transform of the tetrode tip measurements (for mimicking one-channel data) were used for comparison. The number of features N_f used in the simulation was set to 2. Figure 2 shows scatter plots of the AP data taken at V-track-5 position 7 (tip position
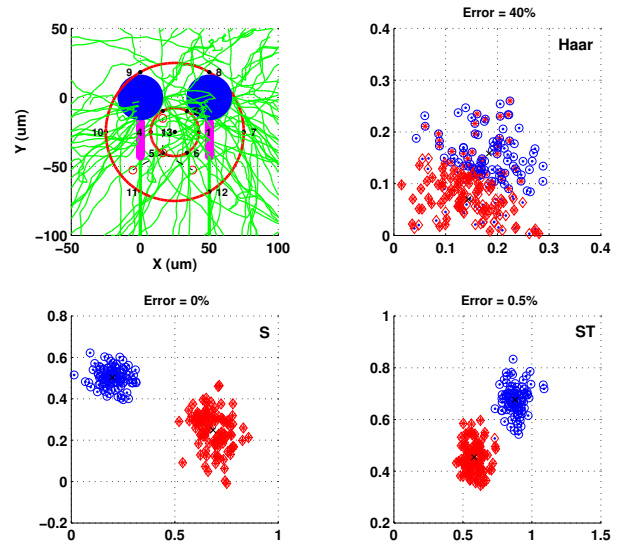

Fig. 2. Top left: a close-up of the two somata (blue) showing 13 vertical sampling tracks. Axon hillocks/initial segments and dendrites are in magenta and green, respectively. The tetrode axis lies along the V-track-5 in this simulation. Scatter plots of spike data taken at V-track-5 position 7 (tip position at [16.25, -40.16, -50.00]) using Harr (top right) and MSD features (bottom left and right). "S" denotes spatial, "ST" denotes spatio-temporal.

at [16.25, -40.16, -50.00]) based on spatial, spatio-temporal and Haar features with SNR = 0 dB. As shown in these figures, the MSD features are more discriminative than the Haar features, and achieve a classification error less than 0.5%. Figures 3 and 4 show plots of the classification errors at all sampling positions for the V-tracks and H-tracks respectively. The SNR was 0 dB in these cases as well. As shown in these two figures, MSD outperforms the existing one-channel Haar wavelet transform approach regardless of which track is considered. This is not surprising since the MSD method uses space and time features for clustering compared to only time features used in the one-channel Haar wavelet transform approach. For V-tracks at z larger than 50 μm, the classification errors of the MSD features become large since in this region, the tetrode is far away from the somata and hillocks, and the received signals are mainly due to the surrounding dendrites. Similar trends are observed in the case single-sensor measurements. However, the simulations demonstrate the advantage of using multi-electrode measurements in this application.

VI. CONCLUSIONS AND FUTURE WORK

This paper provides a new approach for using multi-sensor measurements in AP sorting applications, based on a natural extension of the work in [5]. The algorithm does not require an approximate forward propagation model and is suitable for unsupervised implementation. Our simulations demonstrate that the proposed spatial or spatial-temporal MSD feature extractor yields features that are more discriminatory, thereby facilitating better clustering results than the existing single-electrode approach. The proposed MSD approach could also be used for neuron localization. Upon clustering, the mean spatial signature of a target neuron can be calculated and its null space can be used to perform localization with standard algorithms, such as MUSIC [20] or other approaches.

Fig. 3. Classification errors for V-tracks. Different symbols mean different tracks. The red line represents the central track that lies exactly in between the two neurons (i.e. track 13 shown in Figure 2).
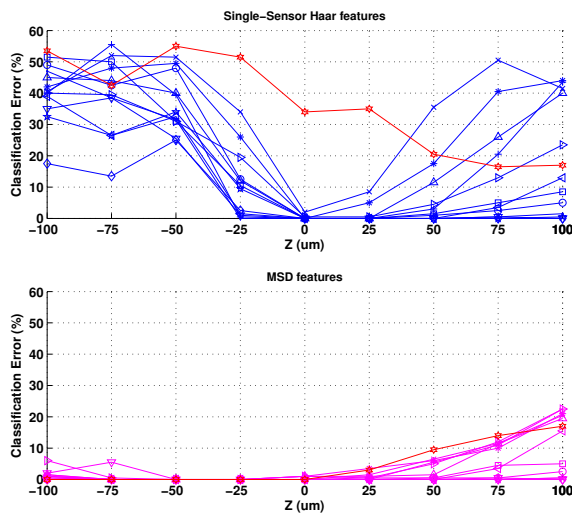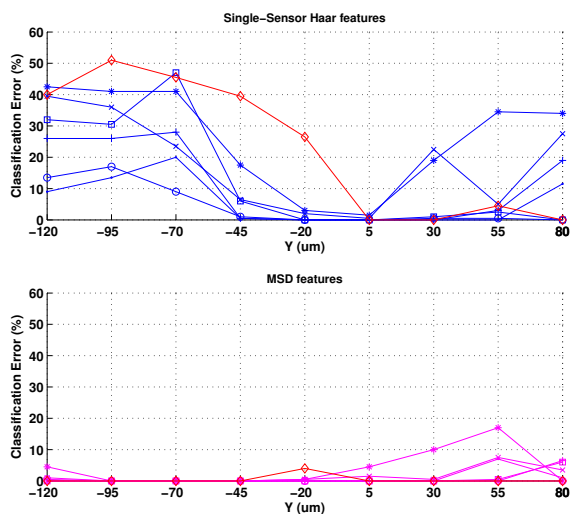

Fig. 4. Classification errors for H-tracks. Different symbols mean different tracks. The red line represents the central track that lies exactly in between the two neurons.

REFERENCES

[1] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network: Computation Neural Syst.*, vol. 9, pp. R53–R78, 1998.

[2] Z. Yang, Q. Zhao, and W. Liu, "Improving spike separation using waveform derivatives," *J. Neural Eng.*, vol. 6, p. 046006, 2009.

[3] C. M. Gray, P. E. Maldonado, M. Wilson, and B. McNaughton, "Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex," *Journal of Neuroscience Methods*, vol. 63, pp. 43–54, 1995.

[4] K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki, "Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements," *Journal of Neurophysiology*, vol. 84, pp. 401–414, 2000.

[5] Z. Nenadic and J. W. Burdick, "A control algorithm for autonomous optimization of extracellular recordings," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 941–955, 2006.

[6] S. Takahashi, Y. Anzai, and Y. Sakurai, "A new approach to spike sorting for multi-neuronal activities recorded with a tetrode: how ICA can be practical," *Neurosci. Res.*, vol. 46, pp. 265–272, 2003.

[7] A. A. Emondi, S. P. Rebrik, A. V. Kurgansky, and K. D. Miller, "Tracking neurons recorded from tetrodes across time," *J. Neurosci. Meth.*, vol. 135, pp. 95–105, 2004.

[8] M. I. Chelaru and M. S. Jog, "Spike source localization with tetrodes," *J. Neurosci. Meth.*, vol. 142, pp. 305–315, 2005.

[9] C. W. Lee, H. Dang, and Z. Nenadic, "An efficient algorithm for current source localization with tetrodes," *Proc. IEEE EMBS*, pp. 1282–1285, 2007.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2001.

[11] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2157, 1994.

[12] P. Parker and A. Swindlehurst, "Space-time autoregressive filtering for matched subspace STAP," *IEEE Trans. Aero. and Elec. Sys.*, vol. 39, pp. 510–520, 2003.

[13] Z. Nenadic and J. W. Burdick, "Spike detection using the continuous wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 74–87, 2005.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998.

[15] I. Bankman, K. Johnson, and W. Schneider, "Optimal detection, classification, and superposition resolution in neural waveform recordings," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 836–841, 1993.

[16] S. C. Wu, A. L. Swindlehurst, and Y. C. Yao, "Direct interference suppression in EEG/MEG dipole source localization," *Proc. IEEE ICASSP*, pp. 574–577, 2010.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second, Ed. Academic Press, New York, 1990.

[18] N. T. Carnevale and M. L. Hines, *The NEURON Book*. Cambridge University Press, 2006.

[19] Tetrode datasheet. http://www.thomasrecording.com.

[20] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenn. and Propagat.*, vol. 34, pp. 276–280, 1986.