Computationally Efficient Feature Denoising Filter and Selection of Optimal Features for Noise Insensitive Spike Sorting

Yuning Yang[a], Samuel Boling[a], Amir Eftekhar[b], Sivylla E. Paraskevopoulou[b], Timothy G. Constandinou[b], Andrew J. Mason[a]

[a] Department of Electrical and Computer Engineering, Michigan State University, USA

[b] Department of Electrical and Electronic Engineering, Imperial College London, UK

***Abstract*— Feature extraction is a critical step in real-time spike sorting after a spike is detected. Features should be informative and noise insensitive for high classification accuracy. This paper describes a new feature extraction method that utilizes a feature denoising filter to improve noise immunity while preserving spike information. Six features were extracted from filtered spikes, including a newly developed feature, and a separability index was applied to select optimal features. Using a set of the three highest-performing features, which includes the new feature, this method can achieve spike classification error as low as 5% for the worst case noise level of 0.2. The computational complexity is only 11% of principle component analysis method and it only costs nine registers per channel.**

I. Introduction

A study of demographic and injury trends has shown that millions of humans suffer severe body paralysis and cannot move their limbs [1]. Recent experiments on neuroprosthetic control by individuals with tetraplegia have demonstrated that brain machine interfaces (BMI) can provide a solution to restore motor function [2]. BMI devices utilize microelectrode arrays (MEA) for extracelluar recordings of neural signals from individual neurons. Because a single electrode usually receives neural signals from multiple neurons and BMI depends on single-unit activity inputs, real-time spike sorting hardware is mandatory for neuronal signal processing. With the advent of high density MEAs, future BMI devices could incorporate a large number of electrodes, on the order of one thousand [3, 4]. To support implantable BMI applications, the algorithm development and hardware implementation of spike sorting methods must be both power and area efficient.

One of the important steps in spike sorting is feature extraction (FE), which occurs after a spike is detected and extracts the most informative features from detected spikes that are then fed into a clustering block for spike identification. On one hand, hardware efficiency requires that FE algorithms be computationally simple for the purpose of low power consumption. On the other hand, the number of features extracted should be as small as possible to ease the design complexity and thus the area of the following clustering block. Traditional methods like principle component analysis (PCA) [5] and discrete wavelet transforms (DWT) [6] demand significant multiplicative operations, and DWT produces a large number of features. Recent studies on hardware-efficient FE algorithms like integral transforms (IT) [7], zero crossing features (ZCFs) [8], discrete derivatives (DD) [9], and first and second derivative extrema (FSDE) features [10], use only additive operations to reduce the computational complexity by at least 95% and generate no more than four features. These methods aim to maximize between-class variances in features, but they neglect within-class variances due to the effect of noise. Neural signals have been shown to suffer high background noise during daily recordings [11]. The existence of noise on features degenerates the separability between different types of spikes. Thus, it is desirable to develop a feature extraction approach that is both insensitive to noise and hardware efficient.

In this paper, a new FE method is presented that utilizes a feature denoising (FD) filter to improve noise immunity of extracted features. Fig. 1 shows the dataflow path of the new FE method. Detected spikes are processed by an FD filter before features are extracted. The FD filter is designed based on an analysis of the power spectrum of neural spikes and background noise. A set of features are then extracted from the filtered spike samples, including extrema, peak-to-peak amplitudes, ZCFs, and a new feature called integration of repolarization (IR). An optimal feature vector is selected from the feature set according to a separability criteria. Finally, the performance of this FE method is compared with other methods in terms of both clustering accuracy using the K-means algorithm and computational complexity as measured in hardware resource counts.

II. Design of feature denoising filter

A. Simulated Datasets

Simulated neural signals sampled at 24KHz will be used for analysis in this paper [6]. They include four datasets with each dataset containing three different spike shapes at four different noise levels from 0.05 to 0.2. The authors label their datasets as Easy1, Easy2, Difficult1 and Difficult2. The noise level is defined as the standard deviation of background noise normalized with respect to the spike peak. Overlapping spikes are not considered in this paper.

B. Spectral Analysis of Spikes and Noise

Usually, a white noise distribution is assumed for noisy

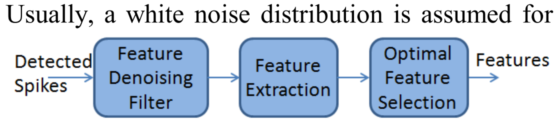

Fig. 1. Procedure of feature extraction method.

signals. However, neural signals exhibit a similar $1/f$ power spectrum in both spikes and noise [6]. A simple derivative method is used in [12] to emphasize high-frequency spike information. To better distinguish between spike information and noise, signal to noise ratio (SNR) is examined in the frequency domain.. This SNR is defined as

$$SNR(\omega) = 10\log\frac{P_{spike}(\omega)}{P_{noise}(\omega)} \qquad (1)$$

where $P(\omega)$ is the power spectrum of analyzed signals and is defined as square of Fourier transform normalized by the length of signals. The power spectra of mean spike templates in the datasets are calculated and compared with the noise spectrum. Fig. 2 shows SNR for all the datasets with three templates in each dataset under a high noise level of 0.2. The spikes can be easily distinguished in the frequency range from 3KHz to 7KHz. In the lower and higher frequency bands, the noise has a significant impact on spikes and the spectra of spikes are overlapped. This implies that a FD filter should be designed by emphasizing the power spectrum in this frequency range.

C. Design of FD Filter

An ideal FD filter should have passband from 3KHz to 7KHz and stopband in other regions. A trade-off exists between the accuracy of frequency response and the number of filter taps, since longer filters require more registers for hardware implementation. In spike sorting DSP, registers dominate a huge fraction of the circuit area. As the number of channels is scaled up to one thousand in the future, minimization of area for each channel is a critical concern in the design.

Assume the FD filter has an impulse response given by

$$h(n) = \sum_{i=1}^{N} c_i \delta(n-i), \qquad (2)$$

where N is the length of filter $h(n)$ and c_i are the filter coefficients. Its Fourier transform can be expressed as

$$H(\omega) = \sum_{n=0}^{N-1} h(n) e^{-j\omega n}, \qquad (3)$$

To have a desired frequency response, $h(n)$ should at least meet the following constraints:

$$\begin{aligned} H(0) &= 0 \\ H(\pi) &= 0 \\ |H(\omega_c)| &= 1 \end{aligned} \qquad (4)$$

where ω_c is the center frequency. To obtain the coefficients c_i, the following optimization problem should be solved under (4) [13]:

$$H_{opt}(\omega) = \arg\min_{H} \left\| G(\omega)[1 - H(\omega)] \right\|^2 \qquad (5)$$

where $G(\omega)$ is a weighting function and is chosen to have a higher weight near the center frequency than in low or high frequency regions. Fig. 3a compares the length of the filter against the width of its transition band. The length of filter has to be five at least, which results in the maximum transition bandwidth of 0.5. Here six was chosen to be the length of filter with transition bandwidth of 0.49 for the purpose of providing

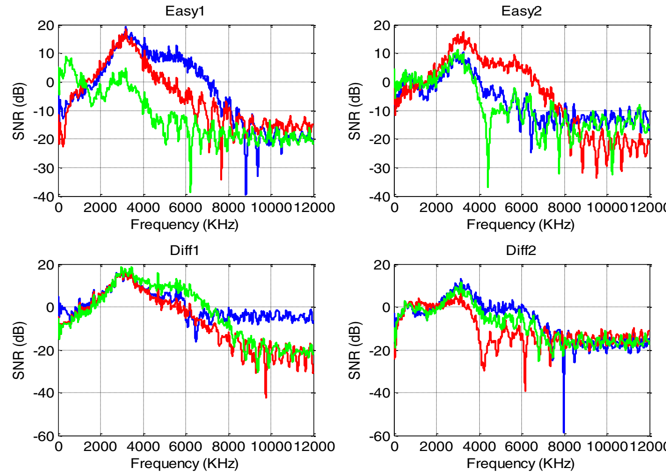

Fig. 2. SNR of three spikes in the frequency domain for different datasets.

better attenuation at stopband. In this case, $G(\omega)$ is defined as $1/2-|1/2-\omega|$ as shown in Fig. 3(b) in order to ensure a symmetric frequency response with ω_c near 1/2. Using the optimization tool box in Matlab, $h(n)$ is found to be:

$$h(n) = [\ 0.5\ -0.5\ -1\ 1\ 0.5\ -0.5\ ] \qquad (6)$$

Fig. 3c shows the frequency response of $h(n)$ and illustrates that (3) is met. It can be seen from (6) that convolution with such a filter only requires operations of shift and addition, and thus can be easily implemented in hardware.

III. Feature Extraction and Evaluation

A. Feature Sets

After spikes are processed by FD filter, features are extracted in the filtered spike samples to reduce the noise effect. Traditionally, maximum, minimum and peak to peak (PtP) amplitude are used to separate different spikes [5, 10]. These features are considered here for filtered spikes. ZCFs are also considered using filtered spike samples, but the positions of zeros are determined from the original spikes.

One important characteristic of the shape of a spike is its

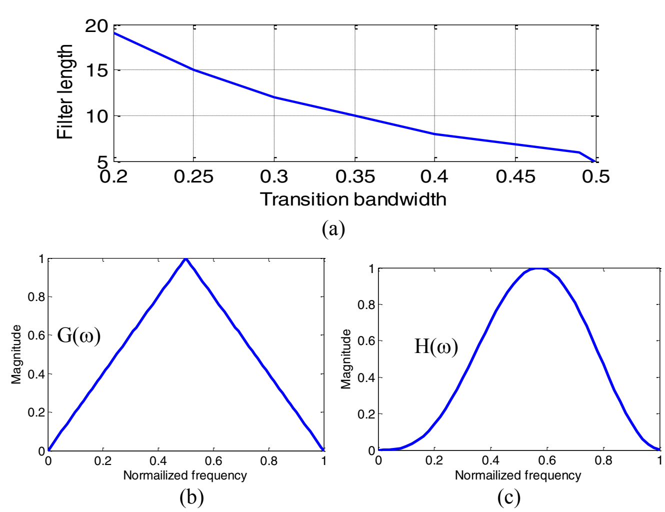

Fig. 3. (a) The length of the filter as a function of transition bandwidth. (b) Weight function of $G(\omega)$. (c) Frequency response of the FD filter.

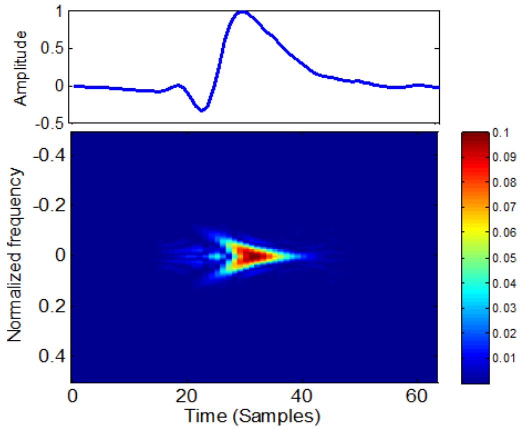

Fig. 4. An example Wigner distribution (bottom) of a spike (top).

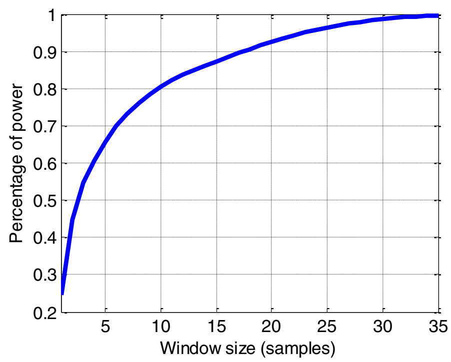

Fig. 5. The ratio of power in the window size to the total power after a spike's maximum amplitude.

repolarization phase. Discrete wavelet transforms have been used to obtain this localized time and frequency information [14]. Fig. 4 shows the Wigner distribution of a spike template whose power spectrum is described in the time-frequency domain. It can be seen that the repolarization phase occupies a significant amount of power while there is negligible power in the spike's refractory period. Integration of filtered samples in this repolarization phase could provide useful information for discriminating spike classes, This phase starts from the point where the maximum amplitude in an original spike is detected (explain this when you first use the term, earlier in the paragraph). This new feature, named as integral of repolarization (*IR)* can be expressed as

$$IR = \sum_{i=I_{\max}}^{M} x_{filtered}(i) \tag{7}$$

where $x_{filtered}$ represents a denoised spike and I_{max} indicates the position of maximum amplitude of an original spike. M is the window size of integration beginning from I_{max}. Fig. 5 shows the averaged ratio of power in the window size M to the total power after I_{max} calculated from all the spike templates in these datasets. Thus, M is set to be ten because 80% of power is included in this window.

B. Feature Evaluation

Fig. 6 illustrates all the features extracted for evaluation and selection. Usually, each feature has its own specific separability on certain spike types. A combination of the features should be selected in order to achieve high separability among different types of spike. The combination can be a subset or all of the features. Generally more features yield better distinction between spike shapes [5]. It is necessary to examine whether a subset of features could provide performance comparable with that of the entire set.

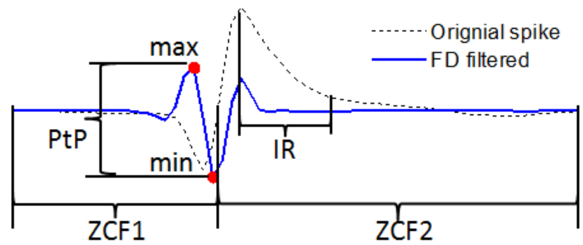

Fig. 6. Illustration of all the extracted features from FD filtered spike.

A separability index (SI) is introduced to evaluate a feature vector [15]:

$$SI = \frac{\sum_{k=1}^{K} [f(t_k) + f(t_k') + 1] \bmod 2}{K} \tag{8}$$

Here, t represents the whole data set. t_k' is the nearest neighbor of t_k using Euclidean metric. K is the total number of points in the data set. f is a binary target function which is equal to +1 if t_k belongs to class i and -1 if t_k belongs to class j. *mod2* repents modulo two operation. Thus, SI is zero if two classes are fully overlapped and one if they are totally separable. The permutations of the feature subsets are evaluated and compared with all of the features at a noise level of 0.2 in table I. Table I lists the best subsets of two, three, four and five features against the performance of all features. The best subset of two features still has a limitation on a particular dataset but can achieve high SI on others. By adding one more feature, the best three features can improve the separability by 11% on the dataset that is predicted poorly by the best two. Although best set of 5 and 6 features provide the best performance, they are only slightly better than the best set of three. Thus, the best three feature subset is used to represent a spike for clustering. The selected features are maximum, minimum, and integral of repolarization.

IV. RESULTS

K-means clustering with the maximum number of iterations limited to 20 was used to test the spike classification performance, and the results of our method using denoising filter are compared with PCA, FSDE and DD methods in Fig. 7. Because there are three features used in this work, the same

Table I COMPARISON OF SI ON BEST FEATURES COMBINATIONS

	Easy1	Easy2	Difficult1	Difficult2
Best 2 features	1	0.79	1	0.99
Best 3 features	1	0.90	1	0.99
Best 4 features	1	0.90	1	0.99
Best 5 features	1	0.92	1	0.99
All 6 features	1	0.92	1	0.99

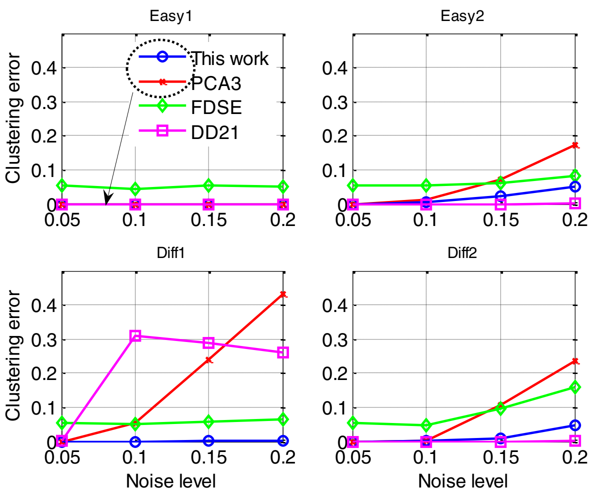

Fig. 7. Comparison of spike classification error using K-means clustering algorithm between this work, PCA FDSE and DD.

number of features of PCA and FSDE are used for fair comparison. For DD, 21 coefficients were used based on the Lilliefors test [9]. Fig. 7 shows that our new method outperforms PCA and FSDE; it can achieve less than 0.4% clustering error when the noise level is less than 0.1. The worst classification error is less than 5% when the noise level is as high as 0.2. Compared to PCA, both have low error for low noise levels, but PCA tends to increase the error as noise goes high. Compared with FSDE, both are less affected by noise variance, but FSDE cannot achieve error free classification under low noise levels. Although DD performs slightly better classification accuracy for most datasets than this work, it has poor performance on the dataset difficult2. Furthermore, both PCA and DD methods require training periods to obtain PC vectors and DD coefficients. In sum, these results demonstrate that the features extracted from FD filtered spikes can provide high classification accuracy and are very robust to noise.

Table II lists the computational complexity and the required registers for each compared method where the number of samples in a spike is 64 in each dataset. Each multiplication is assumed to be equal to 10 additions [7]. PCA is both computationally heavy and registers hungry. DD also requires a significant number of registers. This work requires only 11% of the computations and 9% of the registers required for PCA. Although the computation load is higher than FSDE, actually, the hardware implementation only takes two more adders than FSDE. In summary, the new FE method is hardware-efficient for implantable application.

V. Conclusion

This paper presented a new FE method that uses a FD filter to minimize the noise effect on features. A set of features including a new IR feature were then extracted from the filtered spikes. A separability index was utilized to select and evaluate features for different datasets. Based on the analysis, three of the features were selected for spike classification. The results show that features from FD filtered spikes outperform other methods and can maintain the classification error as low as 5% under worst-case noise levels. Results also show that

Table II COMPARISON OF COMPLEXITY AND REQUIRED REGISTERS FOR DIFFERENT METHODS

	Addition	Multiplication	Complexity	Registers
This work	330	0	330	9
PCA	256	256	2816	64
FDSE	192	0	192	6
DD	181	0	181	21

computational complexity is only 11% of PCA while requiring only nine registers per channel. Overall, these results demonstrate a useful new FE method for high accuracy and low-resource spike sorting suitable for future high-density neural recording systems.

References

[1] A. I. Nobunaga, B. K. Go, and R. B. Karunas, "Recent demographic and injury trends in people served by the model spinal cord injury care systems," *Archives of Physical Medicine and Rehabilitation,* vol. 80, pp. 1372-1382, 1999.

[2] J. L. Collinger, et al, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet,* vol. 381, pp. 557-564, 2013.

[3] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," *Nat Neurosci,* vol. 14, pp. 139-142, 2011.

[4] B. Johnson, S. T. Peace, A. Wang, T. A. Cleland, and A. Molnar, "A 768-Channel CMOS Microelectrode Array With Angle Sensitive Pixels for Neuronal Recording," *Sensors Journal, IEEE,* vol. 13, pp. 3211-3218, 2013.

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

[6] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Computation,* vol. 16, pp. 1661-1687, Aug 2004.

[7] A. Zviagintsev, Y. Perelman, and R. Ginosar, "Low-Power Architectures for Spike Sorting," in *2nd International IEEE EMBS Conference on*, 2005, pp. 162-165.

[8] A. M. Kamboh and A. J. Mason, "Computationally Efficient Neural Feature Extraction for Spike Sorting in Implantable High-Density Recording Systems," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 21, pp. 1-9, 2013.

[9] V. Karkare, S. Gibson, and D. Markovic, "A 130uW, 64-Channel Neural Spike-Sorting DSP Chip," *Solid-State Circuits, IEEE Journal of,* vol. 46, pp. 1214-1222, 2011.

[10] S. E. Paraskevopoulou, D. Y. Barsakcioglu, M. R. Saberi, A. Eftekhar, and T. G. Constandinou, "Feature extraction using first and second derivative extrema (FSDE) for real-time and hardware-efficient spike sorting," *Journal of Neuroscience Methods,* vol. 215, pp. 29-37, 2013.

[11] K. A. Ludwig, J. D. Uram, J. Y. Yang, D. C. Martin, and D. R. Kipke, "Chronic neural recordings using silicon microelectrode arrays electrochemically deposited with a poly(3,4-ethylenedioxythiophene) (PEDOT) film," *Journal of Neural Engineering,* vol. 3, pp. 59-70, Mar 2006.

[12] Z. Yang, C. Tung-Chien, and L. Wentai, "A neuron signature based spike feature extraction algorithm for on-chip implementation," *Engineering in Medicine and Biology Society, IEEE*, 2008, pp. 1716-1719.

[13] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *Image Processing, IEEE Transactions on,* vol. 14, pp. 360-369, 2005.

[14] J. C. Letelier and P. P. Weber, "Spike sorting based on discrete wavelet transform coefficients," *Journal of Neuroscience Methods,* vol. 101, pp. 93-106, 2000.

[15] C. Thornton, "Separability is a Learner's Best Friend," in *4th Neural Computation and Psychology Workshop, London,* , 1998, pp. 40-46.