35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

An Unsupervised Method for On-Chip Neural Spike Detection in Multi-Electrode Recording Systems

Jelena Dragas, David Jäckel, Felix Franke and Andreas Hierlemann, *Member, IEEE*

Abstract— **Emerging multi-electrode-based brain-machine interfaces (BMIs) and large multi-electrode arrays used in *in vitro* experiments, enable recording of single neuron's activity on multiple electrodes and allow for an in-depth investigation of neural preparations, even at a sub-cellular level. However, the use of these devices entails stringent area and power consumption constraints for the signal-processing hardware units. In addition, the high autonomy of these units and an ability to automatically adapt to changes in the recorded neural preparations is required. Implementing spike detection in close proximity to recording electrodes offers the advantage of reducing the transmission data bandwidth. By eliminating the need of transmitting the full, redundant recordings of neural activity and by transmitting only the spike waveforms or spike times, significant power savings can be achieved in the majority of cases. Here, we present a low-complexity, unsupervised, adaptable, real-time spike-detection method targeting multi-electrode recording devices and compare this method to other spike-detection methods with regard to complexity and performance.**

I. Introduction

Extracellular monitoring of the electrical activity of different neural structures has allowed for a better understanding of the behavior of neural cells and networks. The information is passed along neural networks in the form of electrical activity, i.e., voltage pulses, known as action potentials or spikes. When recording spikes with extracellular electrodes, their temporal and spatial characteristics, as well as the specific neuron they originated from can, to a certain degree, be determined by employing dedicated signal processing. This process is known as spike sorting. Different neural recording devices have been developed, most of which are able to sense the activity of a neuron solely on one single electrode. However, it has been shown that the recording of multi-unit activity on multiple electrodes strongly increases spike-sorting performance [1], [2]. Devices featuring the required small inter-electrode distances include needle electrodes and high-density microelectrode arrays (HDMEAs) [3] - [6].

Power storage capacity is very limited in implantable neural-recording devices, or so-called brain-machine interfaces (BMIs), and tight constraints are imposed on power consumption due to the general preference for wireless over wired communication between a BMI and the remainder of the system. The power constraints in these devices become even tighter with the ever-increasing number of readout channels. Considering that during most of the time neurons are inactive, and, that in most cases the vast majority of the recorded neural data represents background noise, extracting and sending out only spike waveforms can significantly decrease the necessary data transmission bandwidth. Therefore, performing spike classification immediately after the detection would further reduce system bandwidth. However, reliable spike-classification methods are highly complex and would entail massive power consumption (>3 times, [7]). On the other hand, reliable spike detection can be achieved with methods of comparably low complexity and an extraction of the spike waveforms allows for spike classification to be performed at the other end of the transmission line, on the host-machine. Moreover, when the identity of the spiking neurons is not needed, transmitting only the spike times ensures minimal power consumption.

Among the existing large electrode arrays that are used for *in vitro* experiments, the number of readout channels is rapidly increasing (e.g., 4096, [4]) in the attempt to simultaneously monitor the activity of a large number of cells. Here, a reliable spike detection performed on chip would ensure a decrease in data-handling complexity and system bandwidth, making the systems more flexible and easier to use.

As the recorded data are non-stationary (i.e., changes in the recording environment cause changes in the recorded signals), the detection method must be adapted to potential changes. Detection methods that rely on *a priori* knowledge of the spikes' average waveforms (templates), such as the template-matching filter-based, Integral Transform (IT), etc., are, therefore, not very well suited ([8]-[10]). Furthermore, adaptation should not be implemented on the host machine, since frequent communication with the host introduces additional strain on the system bandwidth. Similarly, algorithms that require off-line training, such as Principal Component (PC) based algorithms ([7]), are also unsuitable for implementation in such systems. Employing one of the wavelet transform (WT) based methods in spike detection ([11] - [13]) ensures a highly accurate unsupervised detection, but has proven to be non-optimal for on-chip implementation due to its computational complexity ([7]). Lower complexity, energy-based spike detectors have been considered in [8], [10], [14]-[16] mainly Nonlinear Energy Operator (NEO) and local energy measures, which provide good autonomy and accuracy, as well as a low complexity and hence are suitable for an on-chip implementation. Another low-complexity spike-detection method is absolute value thresholding, commonly used as a reference in spike-detection method performance estimation, however, in many cases, without taking into account the cost of an unsupervised estimation of its optimal threshold ([7], [10]).

Despite the large number of papers dealing with the issue of neural spike detection, some of which have been mentioned above, there exists only a small number of methods that exploit the specificities and advantages of multi-electrode data ([17], [18]). In [17] a lifting scheme for

J. Dragas, D. Jäckel, F. Franke and A. Hierlemann are with the Department of Biosystems Science and Engineering of ETH Zurich, in Basle, Switzerland. E-mail: jelena.dragas@bsse.ethz.ch

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

wavelet computation-based hardware has been presented, which is suitable for on-chip implementation. However, as the number of channels reaches the order of several hundred, sufficient data compression rates cannot be achieved.

In this paper we introduce a low complexity method that takes advantage of the nature of multi-electrode data in order to achieve good spike-detection performance and which is suitable for on-chip implementation.

II. Methods

A. Algorithm

Spike detection has proven to be challenging in recordings where the signal level captured by an electrode is low and cannot be clearly distinguished from the noise. Our method takes advantage of the fact that in the case of multi-electrode recordings, this weak activity of a single neuron is simultaneously visible on multiple electrodes. By summing up the signals recorded on the surrounding electrodes with the signal of each individual electrode ($x[n]$), the weak neuronal signal is often sufficiently amplified to exceed the detection threshold. The number of the summed electrode signals depends on the spatial electrode arrangement. An example of the original signal $x[n]$ and the locally summed signal that resulted from its summation with the surrounding electrode signals, $f[n]$, are shown in the left column in Fig. 1. The signals belonging to different electrodes are plotted on top of each other. The benefit of performing local summation of electrode signals is depicted on the plot in the right column of the same figure, showing a comparison of signal-to-noise ratios (SNRs) of spikes before and after the local summation. SNR of a spike, *s,* is defined as ratio between the peak value of the spike, ($|p_j^s|$), and the standard deviation of the noise signal on the electrode, *j,* on which the spike has the peak value: $SNR^s = |p_j^s|/\breve{\sigma}_j$. The increase of SNR in summed signal ensures a better spike-detection performance.

Further increase of SNR is achieved by a method based on the one presented in [16], applied to the local sums. The original method has been adapted for low-complexity hardware implementation by discarding the squaring and scaling operations and by transforming it in the form presented in (1). The parameter *N* is the length of a running window and depends on the characteristics of the spike waveforms that are present in the recording (e.g., frequency content). The $\bar{f}[n]$ factor represents the mean value of $f[n]$ within the window of *N* samples. The presented local energy measure (*LocEn*), unlike the commonly used energy-based method that simply averages the squared signal, reflects the signal frequency as well as its amplitude.

$$E_{LOC}[n] = \sum_{i=0}^{N-1} f^2[n-i] - N\bar{f}^2[n] \quad (1)$$

A threshold is then applied on this signal in a form of a scaled value of the mean of $E_{LOC}[n]$, as follows:

$$T[n] = C\frac{1}{M}\sum_{i=1}^{M} E_{LOC}[n-i]. \quad (2)$$

The size of the window *M* is on the order of several thousand samples, depending on the type of the neural preparation. The value of the scaling parameter *C* also needs to be chosen according to the type of the preparation in order to obtain optimum detection performance. An illustration of $E_{LOC}[n]$ is presented in Fig. 1.

Calculating the locally summed signal for each electrode ensures that, in the vast majority of cases, spiking of a neuron is detectable on several neighboring electrodes. This reduces the probability of false spike reports that can occur when certain electrodes are affected by noise (e.g. electrical) enough to cause a threshold crossing. To this end, the method includes the possibility of detecting only the events in which threshold crossing has occurred on two or more electrodes.

B. Test Data Sets

Testing and verification of the proposed spike-detection method has been conducted on a set of realistic simulated data. The multi-electrode data set has been introduced in [19] and consists of real noise signal, recorded on the device presented in [5]. Superimposed on the noise are spike trains, based on real mouse retinal ganglion cell spike templates. The test data consist of 20 different data sets, each containing 30s recordings of 28 neuronal cells positioned on top of an array, with 722 cells/mm^2 density. The array consists of 90 hexagonally arranged electrodes with 19 μm inter-electrode distance ([5]).

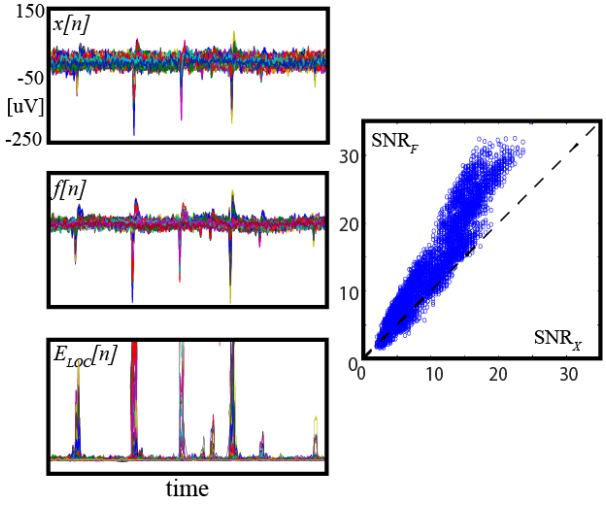

Figure 1. Left (top) Traces from single electrodes plotted on top of one another; different colors represent different electrodes. (middle) Locally summed signals. (bottom) Local energy measure (*LocEn*); Right: Comparison of spike SNRs before and after local signal summation.

C. Reference Spike-Detection Algorithms

The presented method has been compared to two spike-detection algorithms, the low-complexity of which meets the requirements of an on-chip implementation and that are based on function thresholding:

1) Absolute Value Thresholding (Abs) method consists of thresholding the absolute value of the signal $f[n]$. The optimal threshold has been chosen according to [20]:

$$T[n] = C_A\sigma_n[n], \qquad \sigma_n[n] = \text{median}\left\{\frac{|f[n]|}{0.6745}\right\}, \quad (3)$$

where C_A is a positive constant that depends on the type of experiment and $\sigma_n[n]$ is an estimate of the standard noise deviation.

2) Nonlinear Energy Operator (NEO), [15], according to the following equation:

$$\psi\{f[n]\} = f[n]^2 - f[n-\delta]\cdot f[n+\delta]\,, \qquad (4)$$

$\delta \in [1..10]$. Thresholding is performed using an integer multiple of the mean value of $\psi\{f[n]\}$.

D. Performance metrics

For the comparison of the spike-detection algorithms, we use the following definitions:

The *SNR* of neuron *i*: $SNR^i = \max(|t_j^i|)/\sigma_j$, where t_j^i is the template of the neuron i on the electrode j, on which it has the maximal peak, and σ_j is the standard deviation of the noise on that electrode.

Spike-detection error rate (e): $e = \frac{FP + FN}{TS}$, where *FP* stands for the overall number of false reports of spikes, *FN* for overall false negative reports and *TS* for the true number of spikes present in the test data set.

Probability of false alarm: $P_{FA} = FP \cdot L_S/TN$, where *TN* stands for the number of true negative samples, i.e., samples that do not contain spikes and L_s for the number of samples within a spike.

Probability of detection: $P_D = 1 - FN/TS$.

III. Evaluation Results

An example comparison of the thresholded functions corresponding to the three spike-detection algorithms considered in this paper is given in Fig. 2. The topmost plot depicts the absolute value of the multi-electrode test data and represents the function to be thresholded during the execution of the *Abs* algorithm. The signals on the individual electrodes are plotted on top of one another. The center plot corresponds to the *NEO*-processed signal according to (4). The bottom plot shows the *LocEn* signal according to (1). This example illustrates the benefit of applying the local energy measure *LocEn* to increase the signal to noise ratio of a signal, when compared to the other two methods.

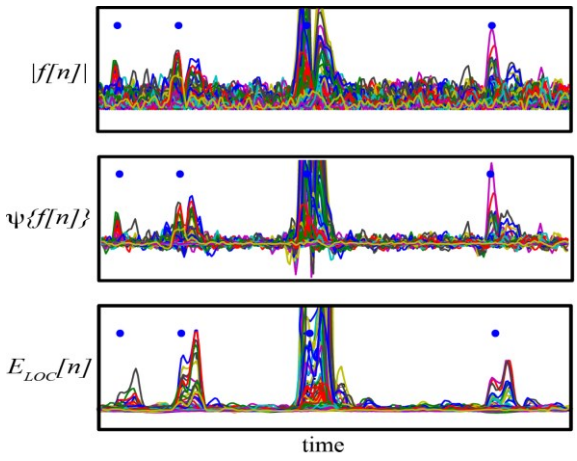

Figure 2. Comparison of three spike-detection algorithms on multi-electrode data: *Abs*, *NEO* and *LocEn* detection (from top to bottom). Dots mark the true positions of spikes within the depicted time interval. The higher SNR, as compared to the other two methods, indicates that the local energy measure-based detection is more suitable for thresholding.

A. Spike-Detection Performance

The parameters of the individual methods were varied in order to determine the trade-offs between the probability of detection (a measure of spike-detection reliability) and the probability of false alarms (a measure of bandwidth/power requirement overhead). The goal of the parameter space exploration was to ensure that only the points in the space for which P_{FA} is below 30% and P_D above 70% are considered in further investigations [7].

After opting for the parameters that ensure a maximal probability of detection and at the same time a minimal probability of false alarm, we investigated the dependency of the number of false negatives (normalized with respect to the true number of spikes) on the neuron's SNR (presented in Fig. 3). In the same figure we show the relation between the three algorithms in terms of number of false positives (normalized with respect to the true number of spikes). We see the advantage of the *LocEn* algorithm, which, for the similar rate of false negatives, yields a lower rate of false positives and thus imposes lower demands on system bandwidth and power consumption. Spike overlaps (defined as spikes for which negative peaks are less than 10 samples apart) were not considered in Fig. 3, since it is not clear to which of the overlapping neurons the detected overlap should be assigned when computing the false negative rate.

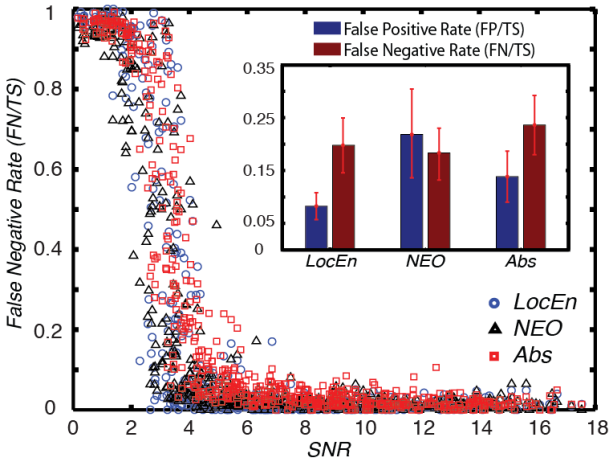

Figure 3. The false-negative rate dependence on the neuron's SNR. When set to achieve similar false negative rates, the algorithms show difference in overall false positive rates: *LocEn* provides the greatest savings in system bandwidth and power consumption.

The performance of each of the methods for the chosen trade-off between false reports and failure to detect spikes is expressed by the spike-detection error rate (e) in Table I. The results show that *LocEn* is more suitable to perform this trade-off, in comparison to the other two methods.

TABLE I. Error Rates of Different Methods in the Trade-off Between False Positive and False Negative Reports of Spikes

	LocEn	***NEO***	***Abs***
Error Rate (e)	31.89%	41.62%	39.91%

B. Hardware-Implementation Performance

From the hardware implementation point of view, three parameters are considered for each algorithm: computational logic and memory requirements, as well as the spike-detection latency; i.e., the number of sampling cycles that pass between the occurrence of a spike and its correct detection. Table II. shows the results of the comparison without taking into account hardware-resource sharing strategies and registers that are needed to store intermediate results and threshold values. Subtraction and comparison operations are counted as an addition, division is set equivalent to multiplication, and the squaring operation is given separately, since, if necessary, it can be implemented using look-up tables in order to save computational resources.

The mean of *LocEn* and *NEO* are calculated over windows of several hundred samples in length, such that the threshold values are constant within a running window and reflect the mean of the signal samples of the preceding window. This strategy ensures a low-complexity real-time implementation, as the history of the signal does not need to be stored, which would otherwise be necessary if a running window were considered instead of the one that precedes it. For estimating the median of the signal's absolute value, in order to calculate the threshold of the *Abs* algorithm, one of the available hardware-friendly algorithms, suitable for real-time implementation, has been considered [21].

TABLE II. COMPARISON OF HARDWARE REQUIREMENTS FOR DIFFERENT ALGORITHMS: COMPUTATION RESOURCES, MEMORY AND LATENCY, PER CHANNEL (NO HARDWARE-RESOURCE SHARING STRATEGY IMPLEMENTED)

Method	*Computation Resources*		*Memory [samples]*	*Latency [# sampl. Clk]*
	Preprocessing	*Threshold*		
Abs	2ADD	1MUL 5ADD	0	0
NEO(δ)	1SQ 1MUL 1ADD	1MUL 1ADD	δ	δ
LocEn	2SQ 2MUL 3ADD	1MUL 2ADD	N-1	0

IV. CONCLUSION

We have presented a spike-detection method, targeting multi-electrode recordings, that is suitable for on-chip implementation due to its ability to adapt in an unsupervised manner to changes in the recordings and its relatively low complexity. We have shown that it outperforms two commonly used low-complexity methods – Absolute Value Thresholding and Nonlinear Energy Operator in terms of spike-detection performance, while featuring reduced demands on system bandwidth and power consumption. The spike overlaps are detected as single spikes, and the separation of the overlapping spikes can be done in the spike classification stage on the host.

ACKNOWLEDGMENT

This work was financially supported by the European Community through the ERC Advanced Grant 267351, "NeuroCMOS".

REFERENCES

[1] 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," *J. of neuroscience methods,* 1995, pp. 43–54.

[2] F. Franke, D. Jaeckel, J. Dragas, J. Mueller, M. Radivojevic, D. Bakkum, A. Hierlemann, "High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments: an emerging technology to study neural plasticity," *Frontiers in Neural Circuits,* 2012.

[3] R. Eckhorn, and U. Thomas, "A new method for the insertion of multiple microprobes into neural and muscular tissue, including fiber electrodes, fine wires, needles and microsensors," *J. of neuroscience methods*, 1993, pp. 175–179.

[4] L. Berdondini*, et al.* "High-density electrode array for imaging in vitro electrophysiological activity," *Biosen. Bioelectron.* 21, pp. 167–174, 2005.

[5] U. Frey *et al,* "Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices," *Biosen. Bioelectron.* 24: 2191–2198, 2009.

[6] B. Eversmann, *et al.* "A 128 × 128 CMOS biosensor array for extracellular recording of neural activity," *Solid-State Circuits, IEEE J. of*, pp. 2306- 2317, Dec. 2003.

[7] S. Gibson, J.W. Judy, D. Markovic, "Technology-aware algorithm design for neural spike detection, feature extraction, and Dimensionality reduction," *Neural Systems and Rehabilitation Engineering, IEEE Trans. on*, pp.469-478, Oct. 2010.

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

[9] A. Zviagintsev, Y. Perelman, R. Ginosar, "A Low-Power Spike Detection and Alignment Algorithm," *Neural Engineering, 2005., 2nd International IEEE EMBS Conference on*, pp.317-320, March 2005.

[10] I. Obeid, P.D. Wolf, "Evaluation of spike-detection algorithms for a brain-machine interface application," *Biomedical Engineering, IEEE Trans. on*, pp. 905-911, June 2004.

[11] K.H. Kim, S.J Kim, "Wavelet-based action potential detector for the extracellular neural signal with low signal-to-noise ratio," *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002*, pp. 2016-2017, Oct. 2002.

[12] Z. Nenadic, J.W. Burdick, "Spike detection using the continuous wavelet transform," *Biomedical Engineering, IEEE Trans. on*, pp.74-87, Jan. 2005.

[13] R.J. Brychta, *et al.*, "Wavelet Methods for Spike Detection in Mouse Renal Sympathetic Nerve Activity," *Biomedical Engineering, IEEE Trans. on*, pp.82-93, Jan. 2007.

[14] I. Obeid, "Comparison of Spike Detectors based on Simultaneous Intracellular and Extracellular Recordings," *Neural Engineering, 2007. CNE '07. 3rd Int. IEEE/EMBS Conf. on*, pp.410-413, May 2007.

[15] J.F. Kaiser, "On a simple algorithm to calculate the `energy' of a signal," *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pp.381-384, Apr 1990.

[16] U. Rutishauser. E.M. Schuman, A.N. Mamelak, "Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo," *J. of Neuroscience Methods,* pp. 204-224, 2006.

[17] K.G. Oweiss, A. Mason, Y. Suhail, A.M. Kamboh, K.E. Thomson, "A Scalable Wavelet Transform VLSI Architecture for Real-Time Signal Processing in High-Density Intra-Cortical Implants," *Circuits and Systems I: Regular Papers, IEEE Trans. on*, pp.1266-1278, 2007.

[18] Y. Suhail, K.G. Oweiss, "Multiresolution Bayesian Detection of Multiunit Extracellular Spike Waveforms in Multichannel Neuronal Recordings," *IEEE-EMBS 2005. 27th Annual International Conference of the*, pp.141-144, Jan. 2006

[19] D. Jäckel, *et al.* "Applicability of independent component analysis on high-density microelectrode array recordings," *Neurophysiology, J. of*, pp. 334–48, 2012.

[20] D.L. Donoho, I.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika 81*, pp. 425–455, 1994.

[21] D. Feldman, Y. Shavitt, "An Optimal Median Calculation Algorithm for Estimating Internet Link Delays from Active Measurements," *End-to-End Monitoring Techniq. and Services, E2EMON, Workshop,* 2007.