

A LOW-POWER ANALOG SPIKE DETECTOR FOR EXTRACELLULAR NEURAL RECORDINGS

Christy L. Rogers and John G. Harris

Department of Electrical and Computer Engineering
University of Florida, Gainesville, FL 32611, USA
[christy, harris]@cnel.ufl.edu

ABSTRACT

This paper discusses a low-power spike detection circuit, which reduces bandwidth from neural recordings by only outputting a short pulse at each neural spike time. Communication bandwidth is dramatically reduced to the number of spikes. The principal idea is to use two low pass filters, one with a higher cutoff frequency to remove high frequency noise and the other with a lower cutoff frequency to create a local average. When the difference between the signal and the local average rises above a threshold a spike is detected. The circuit uses subthreshold CMOS to keep the power consumption low enough for integration of many channels in an implanted device. This spike detection method shows promising results towards a robust and unsupervised algorithm that is lower power and more compact than existing spike detection methods.

1. INTRODUCTION

Brain-machine interfaces (BMI) [1] and other applications require ultra-low power neural instrumentation electronics, such as spike detection circuitry, to be implanted within the body. Low power is necessary due to the difficulty of charging or changing implanted batteries. No external wires should pass through the skin due to the risk of infection, thus the required implantation. Implantation also requires a robust spike detection method and circuit implementation. BMI systems place strong constraints on the wireless transmission because hundreds of channels are recorded.

This paper specifically refers to spike detection for a BMI device that could ultimately allow a quadriplegic to control a robotic arm with his or her brain signals. BMI devices currently record from over a hundred channels and are working towards adding more. Since spikes are sparse in neural data, only transmitting the time of the detected spike, and not the entire signal, could significantly reduce the bandwidth of transmission but would not allow for spike sorting. All BMI systems currently use a human tuned and computationally intensive spike sorting process, which recovers several individual neural signals from each electrode

at the cost of additional power consumption, increased size, and a larger communication bandwidth. Recent results suggest that the spike sorting step may be eliminated without severe degradation of BMI performance [1] thus lending credibility to solely using spike detection.

Robustness is necessary for long-term implantation because of varying noise sources, SNR fluctuations, and DC drift. The electrodes drift over time, which causes the distance from each neuron to each electrode to change. These fluctuations affect the SNR strength because the closer a neuron is to the electrode the stronger its signal. Noise is also affected because distant neurons are lumped together to form noise, correlated with the signal. DC drift occurs as the electrode moves through the brain tissue over time. An implanted circuit must be robust enough to operate over these wide ranges of fluctuations.

Spike detection is a classical problem in neuroscience, with many proposed algorithms in the literature. Popular spike detection methods include amplitude thresholding, wavelets, matched filters, and template matching. Currently, there is no consensus in the community as to the best approach to spike detection, particularly for robust, unsupervised, and computationally simple methods.

Lewicki [2] provides a thorough review of most of these spike detection methods. Each of the proposed detection techniques has shortcomings. The simplest method, amplitude thresholding, quickly begins to fail as SNR drops and is not robust to DC drift. Wavelets have been used for off-line analysis but are presently too computationally expensive for real-time implantable devices. Template matching and matched filtering are two of the most accurate spike detection methods, but they require intensive computation. They also require supervision because they lack robustness toward any change of shape in neural signals that will occur as the electrode drifts over time.

This paper describes an analog spike detection method that is low power and robust. The circuitry is described in Section 2. *Matlab* simulations in Section 3 use recorded neural data to evaluate the performance of this spike detection method.

2. CIRCUITRY

The basic operating principal for the analog spike detector is inspired by the auditory onset detection scheme of Smith [3]. Thus, it will be referred to as the onset detection method. The difference of two low-pass filters is used to enhance the spike. One filter has a higher cutoff frequency to remove high frequency noise and the other has a lower cutoff frequency to create a local average. When the difference between the signal and the local average rises above a threshold, a spike is detected. This method is robust to changes in the noise level as well as DC offsets, both of which are common for long term neural recordings. The basic circuit blocks are shown in Fig. 1.

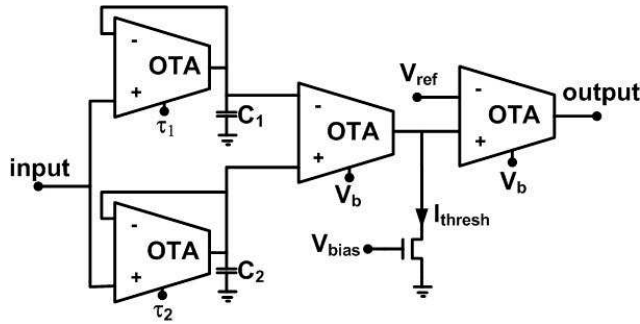


Figure 1: Circuit Diagram

An operational transconductance amplifier (OTA) is configured as a follower integrator for the first-order low-pass filters. The OTAs are run in the subthreshold region to reduce power [4][5]. The τ bias voltages are set off chip to enable adjustment of the cutoff frequencies after fabrication. The desired cut-off frequencies for the two filters were found to be 1.4kHz and 5.3kHz from *Matlab* simulations described in Section 3.2. With the transconductance amplifiers' bias voltages set for a g_m of 150nA/V , $C_1 = 22.5\text{pF}$ and $C_2 = 4.9\text{pF}$. The threshold is set with V_{bias} but requires an autonomous method for multiple channels.

Low power consumption is critical because the circuit must be implanted. *Cadence SpectreS* simulations show the circuit consumes an average of $1\mu\text{W}$ of power.

3. SIMULATED AND MEASURED RESULTS

3.1. Data

High SNR neural recordings, sampled at 20kHz , were used to increase the confidence of the ground truth spikes times determined from the data set. Then, white Gaussian noise was added to give the detection problem a more realistic SNR level. A slowly varying 1Hz , 10mV amplitude sinusoid was also added to the signal to simulate the slowly

varying DC offset. Fig. 2 (a) shows the original neural data waveform and (b) shows the 0dB SNR waveform with an offset. SNR was calculated as the average spike magnitude divided by the average noise magnitude.

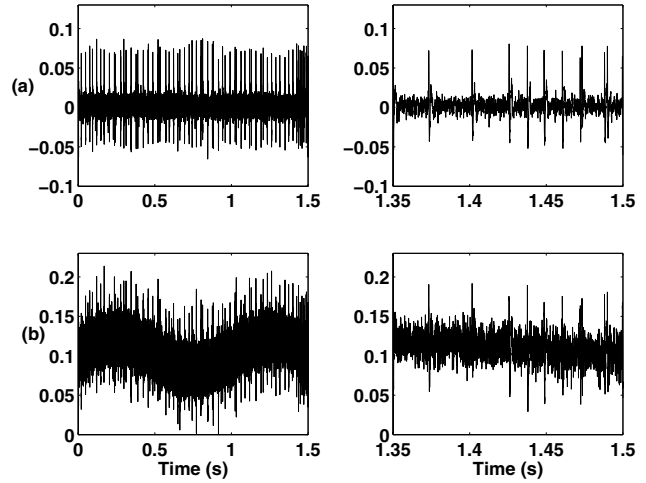


Figure 2: (a) original waveform (b) 0dB SNR waveform with offset. Column two is zoomed in from column one.

3.2. Matlab Simulations

Receiver operating characteristic (ROC) curves are typically used to quantify the performance of detection algorithms [6]. There is always a trade-off between the optimal detection of all the spikes and the erroneous detection of noise as a spike. This detection problem also requires spike time estimation. A detection was considered correct if it occurred within $300\mu\text{s}$ of the actual spike time. Fig. 3 shows that as a larger percentage of spikes are detected more noise will be falsely detected as a spike (also known as a false alarm). The ratio of correct detections to incorrect detections can be set to the desired operating point on the ROC curve by choosing the corresponding threshold level.

To determine the desired circuit cut-off frequencies, ROC curves were constructed from nested cut-off frequency iterations. The circuit's cut-off frequencies were chosen with the minimum number of false alarms at 90% correct detection, 1.4kHz and 5.3kHz .

The onset detection method was compared to the threshold method at 0dB SNR with the results shown in Fig. 3. For comparison purposes the two methods were examined at their 90% correct detection operating point. The onset method outperformed the amplitude threshold method by over 30dB in terms of false alarm rate. Because spikes are sparse in neural data the probability of a false alarm needs to be a fraction of a percent not to swamp the number of correct detections.

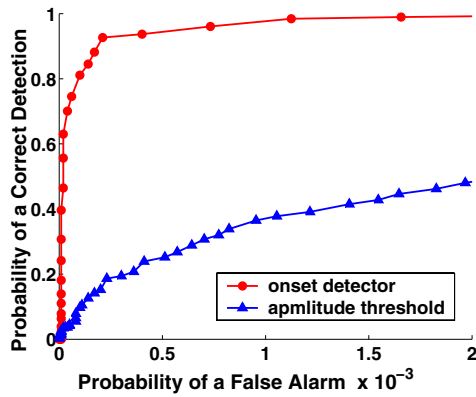


Figure 3: ROC Curves, 0dB SNR

The data used has an average spiking rate of 76Hz so during one second of data at 90% correct detections there should be 68 correct detections out of 76. At 0dB the onset method had a 2×10^{-4} probability of a false alarm, 4 false detections per second. The incorrect detection probability for the onset detector was 6%. For the amplitude threshold method there were 230 false detections per second, so its incorrect detection percentage is much greater at 76%.

The onset detection method consistently outperformed the amplitude threshold method over varying SNR values. Once the SNR became too low, -2dB, neither method performed well. Here the onset method degraded to 25% incorrect detections and the amplitude threshold method was extremely poor at 82% incorrect detections.

Second-order filters were simulated for the onset spike detection circuit but their performance over first-order filters was negligible. Since they require additional chip area and power without noticeable performance improvement, they were not investigated further.

Analysis of the *Matlab* simulation results showed that at 90% correct detection almost all of the false alarms came from noise riding on the second peak of the action potential. Spike-like noise over other parts of the signal was only detected 2% of the time a false alarm occurred. The major portion of false alarms could be reduced by blinding the detector for a short period after it detects a spike. The trade-off to this would be the detector losing resolution between spikes. The amount of time the detector is blinded equals the minimum time required between spikes for detection.

3.3. Chip Results

The onset spike detector chip was fabricated using AMI $0.5\mu\text{m}$ CMOS technology. The chip was $1.5\text{mm} \times 1.5\text{mm}$ with $253\mu\text{m} \times 223\mu\text{m}$ of circuit area.

Due to the difficulty of testing the chip with real neural data, the chip was tested with two basic signal generators to

crudely approximate neural data. A square wave was used to mimic the spike and a high frequency sine wave was used to simulate noise on the signal. The input signal was based on three characteristics of real neurons: spike width, time between spikes, and amplitude.

Neural spike widths vary between 0.3ms to 3ms depending on the species, the brain area, and whether the recording is near the axon or the dendrite. Because the circuit detects the onset of the spike, the effective spike width is the width of the first rise in the action potential. With infinite SNR, this would mean approximately half the action potential time, but as SNR degrades it reduces. The chips functionality was tested with pulse widths of 100 – 400 μs .

The second signal characteristic is the time between spikes. Individual neurons have a refractory period, which sets a minimum time between spikes. Without spike sorting multiple neurons can contribute to one waveform and the refractory period is not a determining factor in the minimum time between spikes. It is then optimal to detect a spike as close the previous one as possible since superimposed spikes can not be discriminated between. The filter speeds and the time to charge the load capacitance are the two factors which determine the minimum detectable time between spikes for the circuit.

Amplitude is the third characteristic of the input signal. Extracellular neural signals have peak-to-peak amplitudes of $50\mu\text{V} - 500\mu\text{V}$. This small signal must first be amplified to give a larger voltage swing for the analog spike detection circuit to be more accurate. Today, low noise, low power neural amplifiers can achieve a gain of up to 100, so the input signal amplitude ranges between 5 and 50mV [7].

The result of a 35mV square wave with a 125 μs pulse width at 25% duty cycle combined with a 15mV high frequency sine wave (to mimic neural noise) is shown in Fig. 4 as the bottom waveform. It shows that 10 μs after the input spikes the output goes high for a short period.

The chip was tested over a wide range of input signal characteristics loosely patterned after neural data. The threshold voltage allows the chip to be adjusted to change the false alarm penalty, and correspondingly its probability of correct detection, in accordance with its ROC curve. The chip detected the crude neural signal representation down to about 25mV with 15mV of noise. The minimum detectable amplitude is not a fixed value because it depends on the SNR and the duration of the spike. Jitter measurements were not performed because the output jitter is negligible with the standard 100ms bins used to decrease the sparseness of neural data before further processing.

These initial chip testing results are in no way exhaustive and the crude signal approximations used for these tests is not an adequate performance measure. Currently, work to obtain a neural signal simulator to more accurately quantify the chip's performance is being pursued.

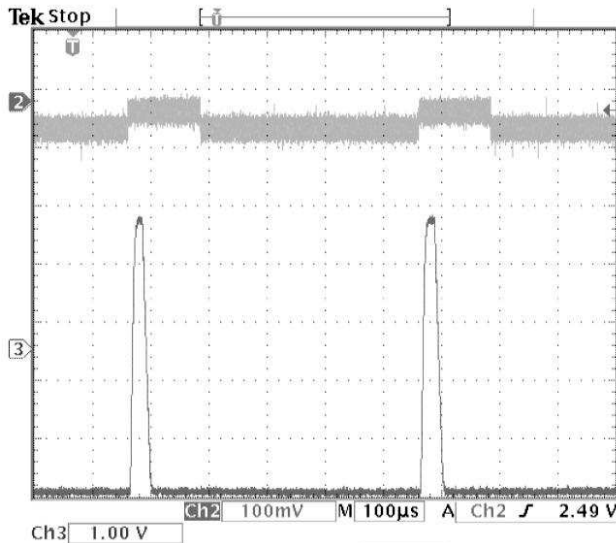


Figure 4: Chip results (bottom signal) for input (top signal).

4. CONCLUSION

The onset spike detector is an excellent compromise between power, transmission bandwidth, area, and robustness. Cadence *SpectreS* simulations showed the chip only consumes $1\mu W$ of power. Since the amplitude threshold is a subset of the onset method's circuitry, it will consume less power. However, its lack of robustness to low SNR and slowly varying DC offsets hinders its performance for BMI devices. *Matlab* simulations with real neural recordings showed that with 90% correct detections at 5dB SNR the onset method outperformed the amplitude thresholding method with 1% incorrect detections versus the 71% incorrect detections respectively. This performance continued for lower SNR such as 0db with the onset detector having only a 6% incorrect detection rate while that of the amplitude threshold method was 76%. Though the results of computationally intensive methods such as matched filtering and template matching were not examined, they are sure to provide better results given enough information is known about the signal. However, even if enough information about the signals were known their power consumption and required supervision to adjust parameters as spike shapes and noise change over time prohibits implantation.

All spike detectors offer a tremendous reduction in wireless transmission bandwidth. For the presented data (76Hz average spiking rate) the onset detector would asynchronously transmit 76 spikes per second. The traditional pulse communication system necessary for spike sorting needs 120kbps assuming a Nyquist sampling rate of 10Hz and 12 bit data. The bandwidth reduction provides a strong motivation to consider dropping spike-sorting for applica-

tions whose performance will not suffer as a result.

The best spike detection method will be the one that still allows the patients to learn to control their neural prostheses while consuming the least power. These trade-offs will be addressed when spike detection methods are tested with real data from BMI type experiments. More extensive data sets are also needed to better examine the robustness of the onset spike detector under different real world noise conditions with a variety of spike shapes from different patients. The onset spike detector has demonstrated a promising compromise between performance and power consumption when implantation is necessary.

5. ACKNOWLEDGMENTS

This material is based on work supported under a NSF Graduate Research Fellowship. Any opinions, finding, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF. This research is also supported by DARPA sponsor grant #N66001-02-C-8022.

6. REFERENCES

- [1] J. Carmena, M. Lebedev, R. Crist, J. O'Doherty, D. Santucci, D. Dimitrov, P. Patil, C. Henriquez, and M. Nicolelis, "Learning to control a brain-machine interface for reaching and grasping by primates," *PLoS Biology*, vol. 1, no. 2, pp. 193–208, Nov. 2003.
- [2] M. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Computation and Neural Systems*, vol. 9, pp. R53–R78, 1998.
- [3] L. Smith, "Using an onset-based representation for sound segmentation," in *Int'l. Conf. on Neural Networks and their Applications*, Marseilles, France, Dec. 1995.
- [4] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, 1989.
- [5] Liu, Kramer, Indiveri, Delbruck, and Douglas, *Analog VLSI: Circuits and Principles*, MIT Press, 2002.
- [6] R. Hippenstiel, *Detection Theory Applications and Digital Signal Processing*, CRC Press, 2002.
- [7] D. Chen, J. G. Harris, and J. C. Principe, "A bio-amplifier with pulse output," in *Int'l. Conf. IEEE Engineering in Medicine and Biology Society*, San Francisco, California, 2004.