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

A Multichannel Integrated Circuit for Neural Spike Detection Based on EC-PC Threshold Estimation

Tong Wu and Zhi Yang*

***Abstract*— In extracellular neural recording experiments, spike detection is an important step for information decoding of neuronal activities. An ASIC implementation of detection algorithms can provide substantial data-rate reduction and facilitate wireless operations. In this paper, we present a 16-channel neural spike detection ASIC. The chip takes raw data as inputs, and outputs three data streams simultaneously: field potentials down sampled at 1.25 KHz, band-pass filtered neural data, and spiking probability maps sampled at 40 KHz. The functionality and the performance of the chip have been verified in both *in-vivo* and benchtop experiments. Fabricated in a 0.13 μm CMOS process, the chip has a peak power dissipation of 85 μW per channel and achieves a data-rate reduction of 98.44%.**

I. Introduction

Spike detection is to differentiate extracellular neural spikes from background noise. Its motivation is twofold: to extract neural spikes for data analysis and closed-loop execution, and to compress neural data and facilitate wireless operations. Many algorithms have been reported where there are three evaluation criteria of corresponding hardware. First, detectors should be suitable for online implementation and not requiring significant computational resources or storage. Second, detectors should be nonparametric and unsupervised to avoid frequent parameter tuning. Third, detectors are preferred to consistently perform well with different recording preparations and experiment protocols.

Several spike detection hardware circuits have been reported [1], [2], [3], [4] to meet some of the requirements. In [1], Rizk presented a FPGA implementation based on an absolute-value thresholding algorithm. This detector is efficient and easy to implement, yet its performance is not satisfactory at moderate or low SNRs and very sensitive to thresholds. An energy-based detector called nonlinear energy operator (NEO) was implemented in a multichannel neural spike-sorting DSP [4]. NEO is meant to boost the differentiation between signals and noise, assuming signals are transient and not correlated with noise. However, neural noise tends to be nonstationary and unstable, resulting in compromised detection performance of NEO. To the best knowledge of the authors, no detection methods other than absolute-value thresholding and energy-based detectors have been implemented in integrated neural recording hardware. Given the unsatisfactory performance of these two groups of detectors, there is a need to have efficient and reliable detection algorithm for implementation [5].

In this paper, we report a detection algorithm followed by its ASIC implementation. In comparison with other detectors and unsolved challenges on efficiency, parameter tuning, and reliability, our detector has the following features. First, through online and iterative learning, the required on-chip storage has been reduced, enabling online and area-efficient hardware design. Second, all parameters are estimated from raw data and adaptively updated, avoiding frequent parameter tuning. Third, the detector is biophysically plausible and featuring fast training within 2.5 sec. As a result, it works reliably with different preparations, a wide range of SNRs, and nonstationary data characteristics. The ASIC has three output streams, which are field potentials, spikes data, and spiking probability maps. From the chip inputs to outputs, 16-channel raw data are compressed from 10.24 Mbps to 160 Kbps, achieving a more than 90% data-rate reduction, which is feasible for reliable wireless transmission .

The rest of the paper is organized as follows. Section II describes our detection algorithm. Section III presents the chip architecture, design trade-offs, and circuit implementation details. In Section IV, system prototype is presented with experiment results. Section V concludes the paper.

II. EC-PC Spike Detection Algorithm

The algorithm is outlined below [6] and its flowchart is given in Fig. 1.

EC-PC Spike Detection Algorithm

Input: Digitized neural data $V(m\Delta T)$, m is the sampling index and ΔT is the sampling interval.

Output: Probability map $p_s(m\Delta T)$ to indicate spike presence. For example, $p_s(m\Delta T)$=1 means the sample is 100% a spike.

- Band-pass filter $V(m\Delta T)$ into $V_{bpf}(m\Delta T)$.
- Transform $V_{bpf}(m\Delta T)$ into Hilbert space as $HV(m\Delta T)$, and form analytic signal $V_{st}(m\Delta T) = V_{bpf}(m\Delta T) + \mathrm{i}HV(m\Delta T)$.
- Define $Z(m\Delta T)$ the power of $V_{st}(m\Delta T)$ and estimate its probability density function $f(Z)$ through histograms.
- Decompose $f(Z)$ into two components, $\widetilde{f}_n(Z) = e^{-\lambda_1 Z}$ (EC) and $\widetilde{f}_d(Z) = Z^{-\lambda_2}$ (PC).
- Calculate $p_s(m\Delta T) = \widetilde{f}_d(Z)/(\widetilde{f}_d(Z) + \widetilde{f}_n(Z))$.

According to [6], recorded neural data are a combination of two components which are noise (exponential component, EC) and detectable spikes (polynomial component, PC) in the Hilbert space. A spiking probability map can be estimated from EC and PC and used for detecting spikes. This detector is nonparametric and self-adaptive. It also has lower

Tong Wu and Zhi Yang* are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077, Singapore. e-mail: {elewut, eleyangz}@nus.edu.sg. *Asterisk indicates corresponding author*

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

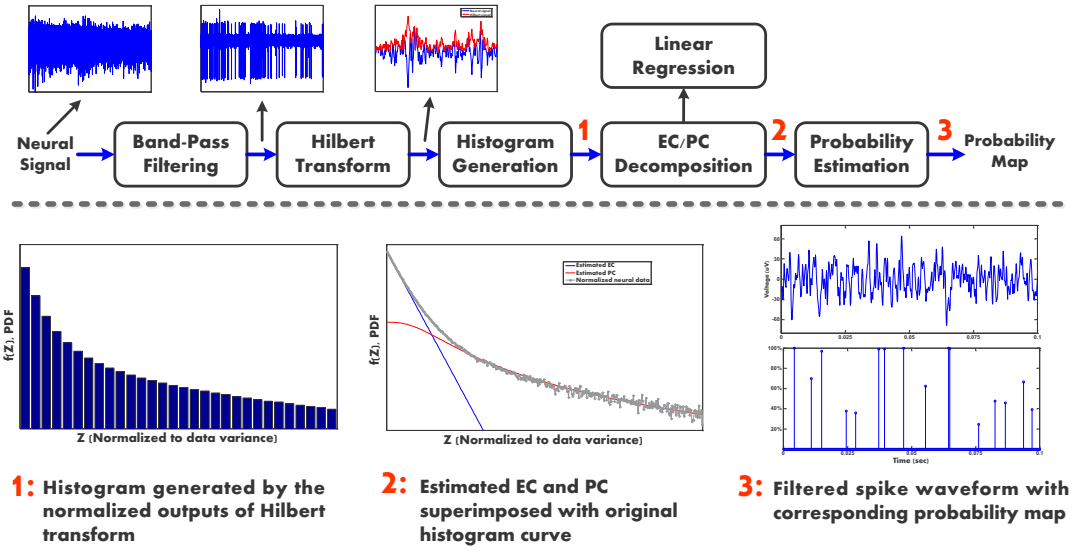


Fig. 1. Flowchart of the proposed detector. Plots 1, 2 and 3 illustrate the histogram, EC-PC decomposition and spiking probability maps.

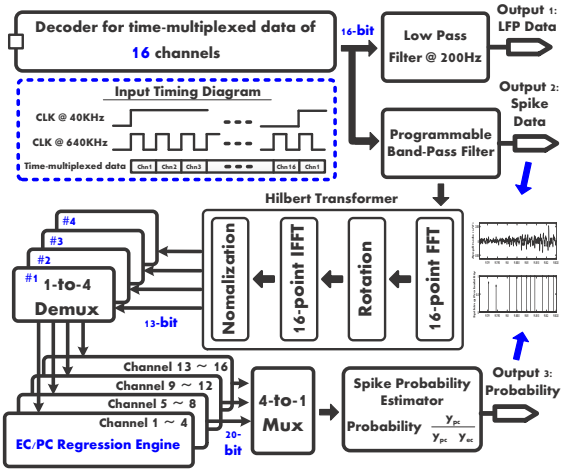


Fig. 2. Architecture of the neural spike detection ASIC.

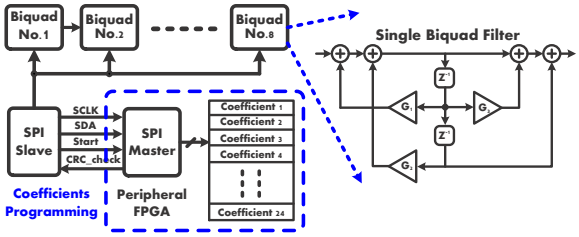


Fig. 3. Structure of the programmable band-pass filter.

TABLE I

HARDWARE RESOURCES SUMMARY OF MAIN BLOCKS.

	Combinational	Sequential	Total count
Low-Pass Filter	9.78K	1.47K	11.25K (5.04%)
Band-Pass Filter	71.07K	6.72K	77.80K (34.84%)
Hilbert Transform	18.09K	5.94K	24.03K (10.76%)
Regression Engine	45.26K	7.35K	52.61K (23.56%)
Probability Gen	56.64K	0.96K	57.60K (25.80%)
Sum	200.84K	22.44K	223.28K (100%)

computational complexity compared with template matching or wavelet-based methods.

III. CIRCUIT IMPLEMENTATION

A. Architecture Design

Fig. 2 illustrates the system architecture where individual blocks are correlated with the main steps in the algorithmic flow. The ASIC receives raw neural data and outputs three data streams: field potentials, spike data, and spiking probability maps. To facilitate an efficient implementation, interleaved architecture has been adopted, which allows different processing channels to share most common combinational circuits, thus reducing hardware cost. For a quantitative measure of hardware savings, combinational and sequential logic for the main blocks obtained from synthesis results are listed in Table I.

Clearly, combinational circuits consumed the most area. By sharing combinational circuits through interleaving, a cost reduction of 82.03% has been achieved.

B. Programmable Band-Pass Filter

A 16-order infinite-impulse response (IIR) filter with tunable corner frequencies is shown in Fig. 3, which consists of eight digital biquad filters cascaded in series. The programmability is supported by a serial peripheral interface (SPI). Cyclic redundancy check (CRC) is incorporated to enhance the data transmission reliability. The coefficient downloading through SPI and simultaneous CRC are coordinated by a on-chip finite-state machine (FSM). Compared with [7], where the parameter adjustment is realized by altering the number of serially connected pseudo-resistors, the proposed mechanism gives a much higher flexibility. Throughout the available spectral bands, the out-of-band rejection is over 64 dB and the in-band ripples are less than 0.08 dB.

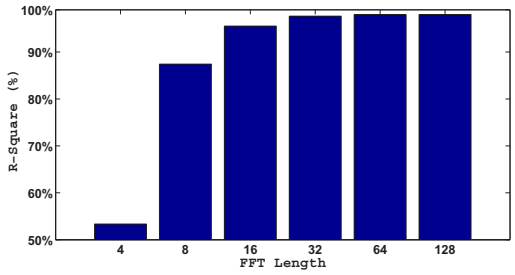


Fig. 4. Comparison of different lengths of Hilbert transform. The histogram generated by Hilbert transform on a 100 sec long neural data is used as the ground truth. Ideally, the length of Hilbert transform equals the length of the data sequence. Different choices of the length from 4-point to 128-point are used to generated histograms, and their similarities to the ideal histogram are evaluated using R^2.

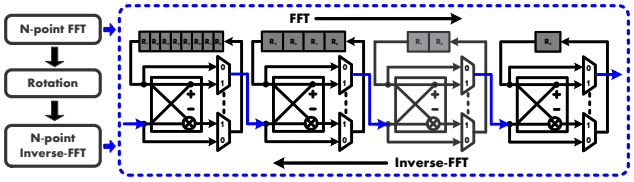


Fig. 5. R2SDF structure of 16-point Hilbert transform.

C. Hilbert Transform

The motivations to perform Hilbert transform are twofold. (1) Extracellular spikes may have significant variations in shape and require multiple detection windows, while only one window is needed after Hilbert transform [8]. (2) Neural data have more compact representations in Hilbert space, which simplifies the EC-PC decomposition.

In our ASIC, Hilbert transform is based on fast-Fourier transform (FFT) and inverse-FFT. According to the evaluation given in Fig. 4, lengths greater than 16 can only achieve less than 3% improvement of accuracy at the cost of linearly increased storage requirement. Therefore, 16-point is selected in this design. An efficient structure with low computational requirement and moderate processing delay called Radix-2 single delay feedback (R2SDF) [9] is used to implement the Hilbert transform, as shown in Fig. 5.

D. EC-PC Decomposition

To ensure adequate training accuracy, the word-length of the bins in histograms for EC and PC estimation are 14-bit and 10-bit, leading to 752 bytes in total for 16 channels. To reduce hardware cost, one histogram is shared by 4 channels sequentially, achieving a 4X reduction in required storage for histograms. Simulation results given in Fig. 6 shown that a 2.5 sec training period for switching histograms among channels can roughly yield a reliable estimation of neuron firing-rates. By setting the training period to be 2.5 sec, coefficients of each channel are updated every 10 sec based on the 2.5 sec training. As shown in Fig. 7, the EC and PC bins are time-multiplexed and processed by the curve fitting units which simultaneously perform two first-order linear regression tasks in the linear-log axis and log-log axis, respectively. After each period, one regression engine is switched to the next channel and builds another histogram in place. The switching is scheduled by a control unit coordinating all regression engines. At the end of each training period, the exponential and polynomial curve fitting units are activated and to estimate the coefficients within 0.75 ms. The estimated coefficients of one channel will remain constant until the regression engine is switched back.

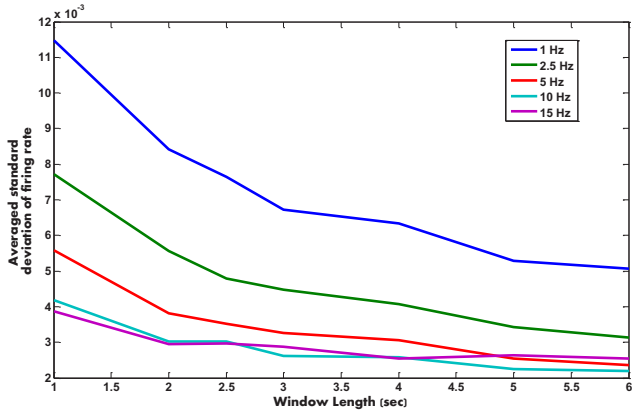


Fig. 6. Evaluation of the training periods for regression. 100 data sequences for each firing rate of the five: 1 Hz, 2.5 Hz, 5 Hz, 10 Hz and 15 Hz. One sequence has 100 segments, each with the same length as the testing training period. Standard deviation of the inferred firing rates from 100 segments are averaged across the 100 sequences, corresponding to one original firing rate. The x-axis: the training periods. The y-axis: the averaged standard deviations for different firing rates.

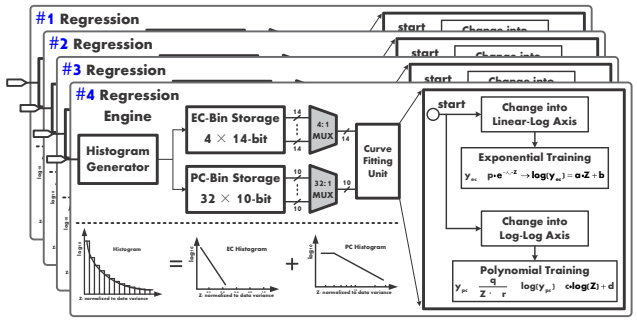


Fig. 7. Architecture of multichannel EC-PC linear regression engine.

IV. Prototyping and Measurements

A. Experiment Setup

As shown in Fig. 8, the chip is packaged in a small printed circuit board (PCB) with a size of 1.9 cm × 1.5 cm, connected to a NeuroNexus microelectrode array. A credit card size board (5.4 cm × 7.5 cm) including a FPGA, SRAMs, level shifters, power managements and interfaces is used as an evaluation board to provide a complete testing benchtop that requires only one USB cable as power and data link. This benchtop can transmit 15 Mbps data bi-directionally.

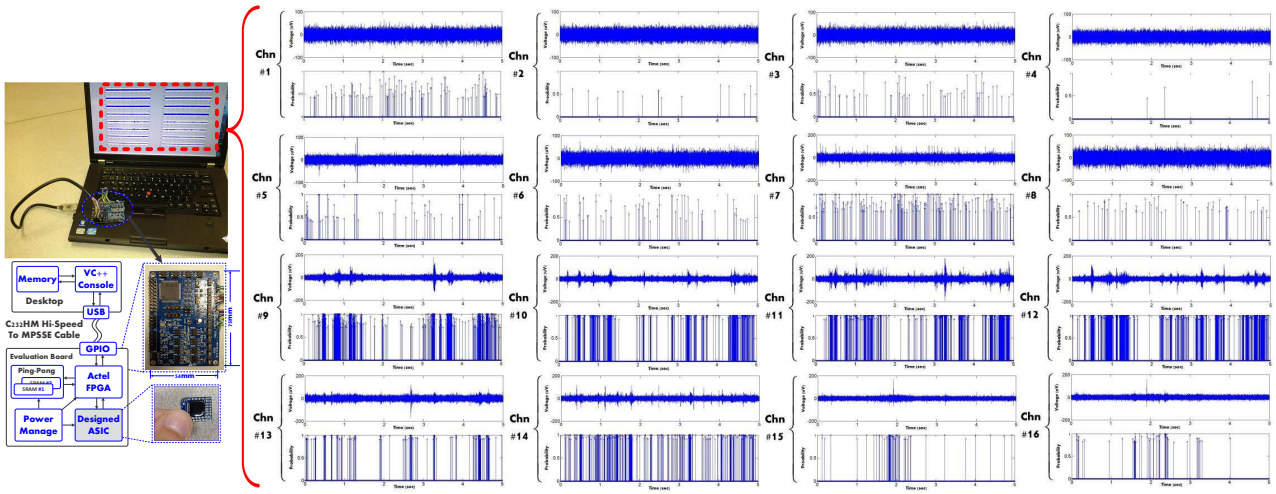
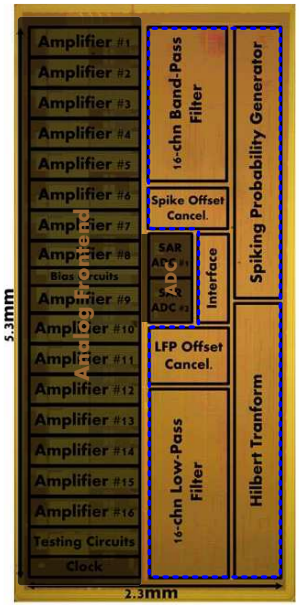


Fig. 8. 16-channel outputs of the ASIC. For each channel, the top is the band-pass filtered neural data (300 Hz - 8 KHz) and the bottom is the corresponding spiking activity map.



Interface & Decoder & Clock	
Power/Area	23.1% and 17.9%
Low-pass filter	
Type	34-order FIR
Astop/Apass	40dB / 0.05dB
Power/Area	8.5% and 10.6%
Band-pass filter	
Type	16-order IIR
Astop/Apass	64dB / 0.08dB
Power/Area	16.6% and 18.2%
Hilbert transform	
Storage	0.37 Kb
Power/Area	26.8% and 19.7%
Regression engine	
Storage	1.504 Kb
Training period	2.5 sec
Power/Area	12.0% and 13.7%
Probability generator	
Power/Area	12.9% and 19.8%

Fig. 9. Chip micrograph and measured circuit specifications.

B. Testing Results

A demonstration of the ASIC to output spike signals and probability maps for 16 channels is shown in Fig. 8. The 16 testing sequences cover a wide range of spiking activities with different SNRs and firing-rates. The outputs are encoded to enhance transmission reliability with an effective data-rate of 10.24 Mbps. The chip microphoto and measured specifications are given in Fig. 9. The core area of the ASIC is 6.71 mm^2. The ASIC consumes 85 μW per channel when its functions are fully activated.

V. Conclusion

In this paper, a 16-channel spike detection ASIC chip is presented. The chip is capable of outputting 16-channel field potentials, spikes and probability maps simultaneously and has achieved over 98% data-rate reduction to facilitate wireless operation. By interleaving across 16 channels, more than 80% hardware cost has been reduced, making the chip suitable for implantable applications. Testing prototypes have also been developed to facilitate the operations of the ASIC in neural recording experiments.

Acknowledgement

This work is supported by Singapore A*STAR and MOE grants R-263-000-699-305, R-263-000-A32-305, R-263-000-A29-133, and R-263-000-619-133.

References

[1] M. Rizk, I. Obeid, S. Callender and P. D. Wolf, A single-chip signal processing and telemetry engine for an implantable 96-channel neural data acquisition system, J. Neural Eng., vol. 4, no. 3, pp. 309–321, 2007

[2] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger and F. Solzbacher, A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System, IEEE J. Solid-State Circuits, vol. 42, no. 1, pp. 123–133, 2007

[3] B. Gosselin, A. E. Ayoub, J. Roy, M. Sawan, F. Lepore, A. Chaudhuri and D. Guitton, A Mixed-Signal Multichip Neural Recording Interface With Bandwidth Reduction, IEEE Trans. Biomed. Circuits Syst., vol. 3, no. 3, pp. 129–141, 2009

[4] V. Karkare, S. Gibson and D. Marković, A 130-μW, 64-Channel Neural Spike-Sorting DSP Chip, IEEE J. Solid-State Circuits, vol. 46, no. 5, pp. 1214–1222, 2011

[5] G. Yuan, Z. Yuanjin, D. Shengxi, T. Wei-Da, A. Chyuen-Wei, J. Minkyu and H. Chun-Huat, Low-Power Ultrawideband Wireless Telemetry Transceiver for Medical Sensor Applications, IEEE Trans. Biomedical Eng., vol. 58, no. 3, pp. 768–772, 2011

[6] Z. Yang, W. Liu, M. R. Keshtkaran, Y. Zhou, J. Xu, V. Pikov, C. Guan and Y. Lian, A new EC-PC threshold estimation method for *in vivo* neural spike detection, J. Neural Eng., vol. 9, no. 4, 2012

[7] A. Rodriguez-Perez, J. Ruiz-Amaya, M. Delgado-Restituto and A. Rodriguez-Vazquez, A Low-Power Programmable Neural Spike Detection Channel With Embedded Calibration and Data Compression, IEEE Trans. Biomed. Circuits Syst., vol. 6, no. 2, pp. 87–100, 2012

[8] Y. Zhou, T. Wu and Z. Yang, A Novel EC-PC Spike Detection Method for Extracellular Neural Recording, *in submission to* IEEE Trans. Biomedical Eng.

[9] E. H. Wold and A. M. Despain, Pipeline and parallel-pipeline FFT processors for VLSI implementation, IEEE Trans. Comput., vol. C-33, no. 5, pp. 414–426, 1984