33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Implantable Neural Spike Detection using Lifting-Based Stationary Wavelet Transform

Yuning Yang, Andrew J. Mason, *Senior Member, IEEE*

***Abstract*—Spike detection from high data rate neural recordings is desired to ease the bandwidth bottleneck of bio-telemetry. An appropriate spike detection method should be able to detect spikes under low signal-to-noise ratio (SNR) while meeting the power and area constraints of implantation. This paper introduces a spike detection system utilizing lifting-based stationary wavelet transform (SWT) that decomposes neural signals into 2 levels using 'symmlet2' wavelet basis. This approach enables accurate spike detection down to an SNR of only 2. The lifting-based SWT architecture permits a hardware implementation consuming only 6.6 μW power and 0.07mm^2 area for 32 channels with 3.2 MHz master clock.**

I. Introduction

THE development of microelectrode arrays allows neuroscientists to record up to 100 extracellular neural signals simultaneously [1]. Wireless transmission of multichannel neural signals without loss of information requires tens of Mbps in bandwidth. However, reliable bio-telemetry techniques typically achieve bandwidths of less than 10 Mbps, which presents a bottleneck to data rate of neural signals. Furthermore, transmission at high data rate consumes significant power that must be managed to avoid damage to living tissue.

For scientific research and neuroprosthetic applications, the neurologically relevant information is contained within action potentials or spikes fired by neurons. It has been shown that the firing rate of spikes is typically between 10-100Hz [2], indicating that spikes alone provide a sparse representation of neural signals. A potential solution to ease transmission bandwidth is to perform online spike detection and transmit only spike information instead of raw data. For example, at a data sampling rate of 25 Ksamples/sec per channel with 10-bit resolution, spike extraction would reduce the bandwidth more than 90%. However, the recorded neural signals consist of spikes and background noise. Experimental observations show that the signal-to-noise ratio (SNR) varies among channels in a probe and changes day by day [3], where the SNR is defined as

$$\text{SNR} = \frac{\text{peak to peak amplitude}}{2\times\text{standard deviation of noise}} \qquad (1)$$

In practice, sometimes less than 50% of the total recording sites are considered to have good SNR (SNR > 4) [3]. As a result, it is necessary to develop a spike detection algorithm that is efficient under low SNR.

An evaluation of computationally efficient spike detection algorithms has been reported recently [2]. Absolute threshold and non-linear energy operator (NEO) are two computationally efficient methods that have been implemented in integrated circuits [4-5]; however, the performance of both methods degrades severely as SNR becomes poor. In contrast, template matching [2] is very effective but is too computationally intense for implantable circuit implementation. Similarly, spike detection utilizing stationary wavelet transform (SWT) has demonstrated good performance [6], but the algorithm is not suitable for realization within an implantable circuit.

This paper introduces a hardware-efficient spike detection system utilizing a lifting-based architecture for SWT. The presented SWT-based spike detection system outperforms all reported implant-capable methods, performs exceptionally well at low SNR, and can be implemented within the power and area constraints of implantation.

II. Wavelet Transforms for Spike Detection

A. DWT Versus SWT

Wavelet transform (WT) has been applied to biomedical signals for denoising, compression and detection. WT provides a large degree of freedom to study signals by choosing different wavelet bases and decomposing signals into different frequency sub-bands. Wavelet-based spike detection applies different threshold values to each level of the detail coefficients. Coefficients with an absolute value greater than the threshold are thought to indicate spikes. SWT is an implementation of WT that overcomes the shift variance issue in discrete wavelet transform (DWT) at the cost of increased computational complexity. The choices of wavelet basis and decomposition levels have a great impact on both the quality of results and the efficiency of hardware implementation.

Our previous work shows that DWT using 'symmlet4' basis is a hardware efficient way to compress signals to ease the bandwidth bottleneck of bio-telemetry [10]. The compressed wavelet coefficients can be transmitted and reconstructed in computers with little loss of information. However, DWT suffers a key limitation in terms of shift variance. Because coefficients are decimated by two at each level, it is possible that the odd decimation coefficients of spikes have a large distinction from even coefficients causing shift variance An intuitive way to observe the effect of shift variance on spike detection is illustrated in Fig. 1. The signal is decomposed into 2 levels with odd and even decimation shown separately. The spike can be observed clearly in odd decimation while it is difficult to see in either level of the

even decimation. Because of shift variance, it is equally possible that spikes could be missed by observing either odd or even samples.

For a typical 25 KHz sampling rate, the sampling point can be a random value between two samples range from 0 to 40 µS. A good performance detector should be robust to any time shift (phase delay) in this range. Simulated neural data was used to test the detection accuracy for DWT and SWT as the sampling point varies across the sample interval. Here, detection accuracy is defined as

$$\text{Accuracy} = \frac{\#\text{ of true detection}}{\#\text{ of total true spikes} + \#\text{ of false positives}} \quad (2)$$

The data set was composed of 150 hundred thousand samples containing around 300 spikes under SNR equal to 3, where the SNR is defined in (1). Both DWT and SWT decomposed signals into 2 levels while using optimal bases: 'symmlet4' for DWT and 'symmlet2' for SWT. Fig. 2 shows the detection accuracy of SWT is very high and almost constant over the range of time shift, which is expected because SWT is undecimated at each level and the coefficients are shift invariant. However, the performance of DWT changes as a function of time shift, and the difference between the best and worst case is as large as 30%. In practical applications, only the worst case result can be expected, and thus SWT significantly outperforms DWT in real time spike detection.

B. Choice of Levels

The SWT decomposition process can be expressed as

$$\begin{aligned} a_{j+1}(n) &= \sum_k h_j(n-k)a_j(k) \\ d_{j+1}(n) &= \sum_k g_j(n-k)a_j(k) \\ h_{j+1} &= upsample\ (h_j) \\ g_{j+1} &= upsample\ (g_j) \end{aligned} \quad (3)$$

where at each level the approximation coefficients a_j are convolved with a half band low pass filter h_j generating the approximation coefficients a_{j+1} and a half band high pass filter g_j generating the detail coefficients d_{j+1} for the next level. The low and high pass filters are upsampled at each level. (3) describes the SWT cutting the approximation coefficients into two half bands at each decomposition.

Most wavelet methods for spike detection decompose neural signals into more than 4 levels [6-7]. However, high level detail coefficients correspond to low frequency signal bands. A recent study shows that the noise associated with recorded neural signal exhibits a frequency dependency that can be approximated as $1/f^{\alpha}$ [8]. Fig. 3. plots a noise power spectrum of data recorded from a rat motor cortex [3]. The noise power of 40 dB attenuates at frequencies above 5 KHz. Spikes are considered as instantaneous energy changes at high frequency, and to detect spikes it is preferable to discriminate spikes from noise within the high frequency band. Considering the relative frequency responses of background noise and spikes, it is ineffective to use information from higher decomposition levels. For a typical neural signal digitized at 25 Ksamples/sec, the wavelet

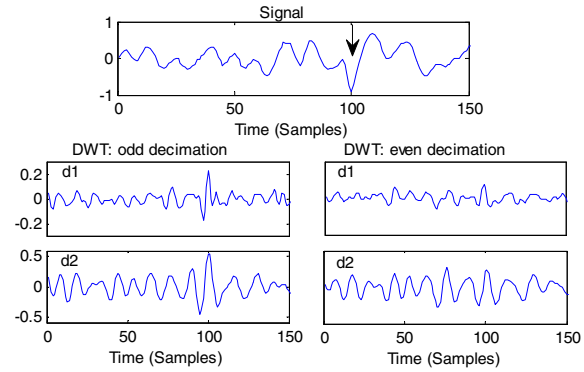

Fig. 1. Odd and even decimation of neural signals into 2 levels using DWT with 'symmlet4' basis. Top shows the original signal with spike indicated.

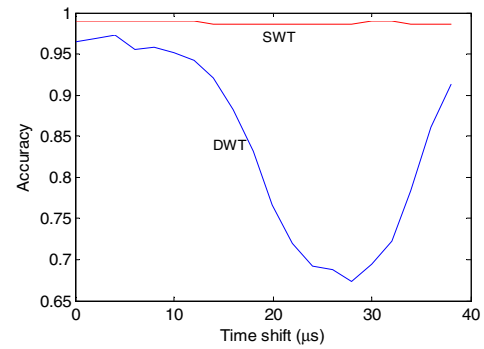

Fig. 2. Detection accuracy of DWT and SWT against time shift for SNR=3 under 25 KHz sampling rate.

decomposition of two levels of detail coefficients covers a frequency band from 3.125 KHz to 12.5 KHz. In this case, wavelet decomposition coefficients beyond level 2 provide negligible information for detecting spikes.

III. Hardware Design

A. Lifting Architecture and Wavelet Basis

The lifting scheme for computing DWT has been well developed. It provides fast computation and efficient hardware realization. Fig. 4 (a) describes the lifting architecture concept for DWT where coefficients are predicted and updated by factorizing polyphase matrix *E(z)* of low and high pass filters into N steps to obtain Laurent polynomials *P(z)* and *U(z)* as

$$E_{DWT}(z) = \prod_i \left\{ \begin{bmatrix} 1 & U_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ P_i(z) & 1 \end{bmatrix} \right\} \quad (4)$$

The lifting scheme of SWT is based on the DWT scheme without splitting data into odd and even samples [9]. However, the delay of each sample has to be doubled as shown in Fig. 2 (b). The lifting process of SWT can be expressed as

$$E_{SWT}(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-2} \end{bmatrix} \prod_i \left\{ \begin{bmatrix} 1 & U_i(z^2) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ P_i(z^2) & 1 \end{bmatrix} \right\} \begin{bmatrix} 1 & 0 \\ 0 & z^2 \end{bmatrix} \quad (5)$$

Because the filters have to be upsampled by 2 at each level as expressed in (3), the lifting process also needs to be changed accordingly at each level as

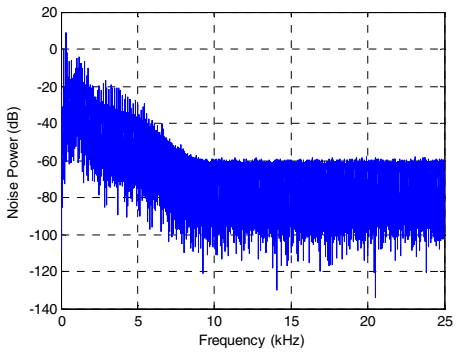

Fig. 3. Noise power spectrum recorded from a rat motor cortex.

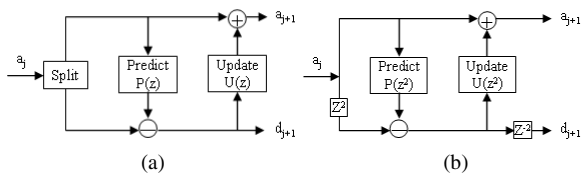

Fig.4. (a) Lifting scheme of DWT. (b) Lifting scheme of SWT.

$$E_{SWT\,j+1}(z) = E_{SWT\,j}(z^2) \tag{6}$$

The delay elements are realized in hardware as registers. From (5) and (6), the number of registers required at each level is equal to $(N+1)\times L$, where N is the length of wavelet filters and L is the level index. Registers consume considerable area in integrated circuits; thus an implantable design for multichannel applications requires the filter length to be short. Although 'haar' wavelet has the shortest length, its shape has poor similarity to neural spikes and thus it is not suitable for spike detection. To address this design tradeoff, the 'symmlet2' basis was chosen; it has a reasonable filter length of four while requiring less than half of the hardware resources as other bases. As shown in Section IV, 'symmlet2' only slightly degrades detection accuracy while providing a significant benefit in terms of hardware efficiency.

For the chosen 2 levels of decomposition and the 'symmlet2' basis, the lifting steps of SWT are

$$\begin{aligned}
Q_1(n-1) &= x_1(n) + B_0 x_1(n-1)\\
a_1(n-3) &= x_1(n-3) + B_1 Q_1(n-1) + B_2 Q_1(n-3)\\
d_1(n-1) &= Q_1(n-1) + B_3 a_1(n-3)\\
Q_2(n-2) &= x_2(n) + B_0 x_2(n-2)\\
a_2(n-6) &= x_2(n-6) + B_1 Q_2(n-2) + B_2 Q_2(n-6)\\
d_2(n-2) &= Q_2(n-2) + B_3 a_2(n-6)
\end{aligned} \tag{7}$$

where x is input signal, Q is an intermediate value, B_0-B_3 are constant coefficients and the subscripts of x, Q, a, d represent the level. (7) implies that 15 delay samples need to be stored to complete a 2 level calculation for one input sample, including 3 for x, 2 for Q_1, 6 for a_1 and 4 for Q_1. The signal flow diagram of the lifting steps is described in Fig. 5.

To detect spikes, each detail level coefficient, d_1 and d_2, is compared to a carefully chosen threshold value, and a coefficient greater than its threshold value is considered to indicate a spike. Threshold values are set to a scaled version of the standard deviation of d_1 and d_2 separately based on the low probability that background noise exceeds this threshold [4].

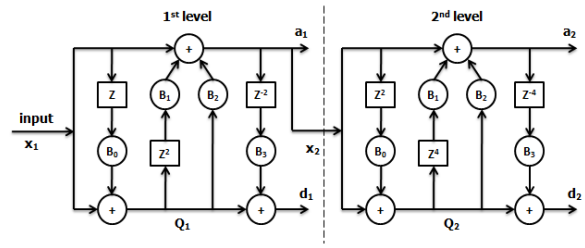

Fig. 5. Data flow diagram of SWT with two level decomposition using symmlet2 wavelet basis.

B. System design

To implement the data flow of Fig. 3 in hardware, the SWT system shown in Fig. 6 (a) was developed based on our previous design of an area-power efficient VLSI architecture for lifting based DWT [10]. The computation core (CC) is used to calculate the lifting steps in (7). The CC can be implemented using one multiplier and one adder as shown in Fig. 6 (b) to sequentially evaluate each expression in (7). Two CCs are required to calculate 2 levels decomposition in a pipeline. The channel/level memory stores 15 delay elements for each channel. The controller manages the computation sequence for the CC and works with the address generator to controls memory access. The system requires four clock cycles to complete computation of one sample per channel. Thus the master clock should be at least $4\times f_s\times N_{ch}$, where f_s is the sampling rate and N_{ch} is the number of channels.

IV. Results

Synthetic neural signals were generated using a neural signal simulator with ten real recorded spike pattern templates for which the recording method is described in [11]. Fig. 7 shows ten channels used for simulation, each composed of three different spike templates. Spike detection accuracy of the new SWT-based method was compared with absolute threshold (AT), NEO and DWT methods under different SNRs. As shown in Fig. 8, SWT achieves above 90% accuracy even at SNR=2. The performance of NEO is 15% less than SWT at low SNRs. Due to its shift variance, DWT cannot successfully detect all spikes even at high SNRs.

For 32 channels with a 3.2 MHz clock frequency, the SWT architecture described in Section III was implemented in CMOS using *Verilog* mapped to a 130nm standard cell library. Table I lists the simulated power and layout area for

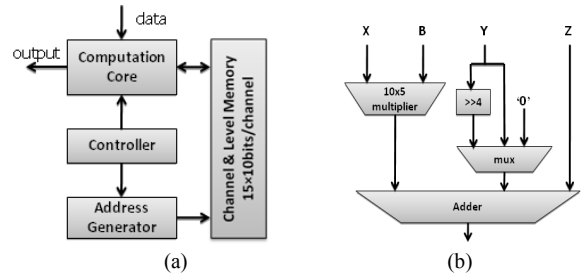

Fig. 6. (a) System architecture of SWT. (b) CC architecture.

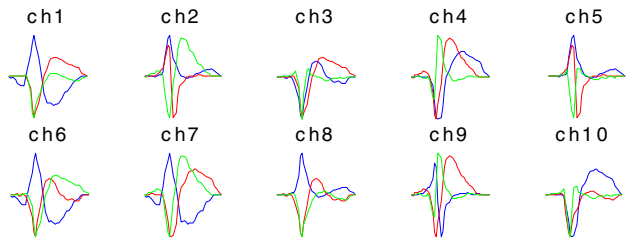

Fig. 7. Ten spike channels containing three spikes in each channel. All spikes are from one of ten spike templates.

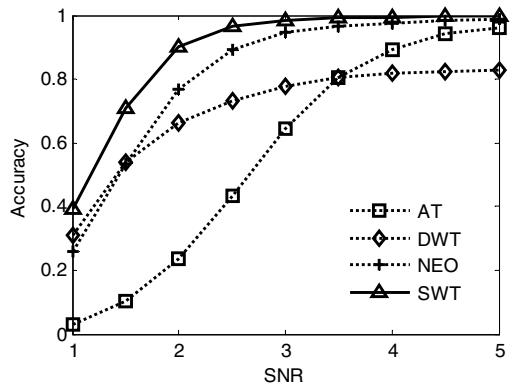

Fig. 8. Detection accuracy vs. SNR for AT, NEO, DWT and SWT.

major blocks and the fully functional complete system. The CC dissipates 70% of total power while memory occupies 85% of total area. The whole SWT block consumes only 6.6 μW and 0.07mm^2 and thus could easily be implemented within a neural recording implant.

Our previous work using SWT for spike detection with the 'symmlet4' basis demonstrated good detection accuracy [12]. Fig. 9 (a) compares both bases in terms of detection accuracy. Although 'symmlet4' performs slightly better than 'symmlet2', its implementation requires twice as many registers for storing the delay samples and has a major impact on required area. To illustrate the overall performance comparison, a normalized figure of merit (FOM) described by

$$\text{FOM} = \frac{\text{detection accuracy}}{\text{power} \times \text{area}} \tag{8}$$

is plotted in Fig. 9 (b) and shows 'symmlet2' is significantly better than 'symmlet4' for all SNRs.

V. Conclusion

A neural spike detection method based on SWT was presented along with an efficient VLSI implementation utilizing the lifting scheme. Optimized for neural data noise characteristic and implantable hardware requirements, the SWT spike detector realizes 2 level decomposition and uses the 'symmlet2' basis. The new system shows the best reported spike detection performance even for neural signals with low SNR. The power and area efficient design make it suitable for implantation within a wireless neural recording array.

Table I
Power and Area Values for 32 Channels at 3.2 MHz Master Clock

	Power		Area	
	(μW)	*percentage*	*(μm^2)*	*percentage*
Computation Core	4.35	66%	9362	14%
Memory	1.02	15%	57277	83.7%
Controller	0.59	9%	230	0.3%
Address Generator	0.63	10%	1586	2%
Total	6.59	100%	68455	100%

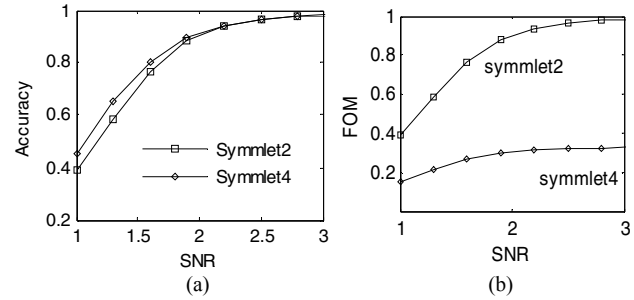

Fig. 9 Comparison of (a) detection accuracy (b) and FOM between'symmlet2' and 'symmlet4' basis for SWT.

References

[1] A.C. Hoogewerf and K. D. Wise, "A three-dimensional microelectrode array fix chronic neural recording," *IEEE Trans. Biomed. Eng.,* vol. 41, pp. 1136-1146, 1994.

[2] I. Obeid and P. D. Wolf, "Evaluation of spike-detection algorithms for a brain-machine interface application," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 905-911, 2004.

[3] K. Ludwig, J. Uram, J Yang, D. Martin, D. Kipke, "Chronic neural recordings using silicon microelectrode arrays electrochemically deposited with a poly(3,4-ethylenedioxythiophene) (PEDOT) film," *J. Neural Eng*, pp 59-70, March 2006.

[4] R. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," *IEEE Eng. in Medicine and Biology Conf.*, pp. 3325-3328, 2003.

[5] B. Gosselin, M. Sawan, "A ultra low-power CMOS automatic action potential detector," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 4, pp 346-353, Aug 2009.

[6] K. Kim and S. Kim, "A wavelet-based method for action potential detection from extracellular neural signal recording with low signal-to-noise ratio," *IEEE Trans. Biomed. Eng.,* vol. 50, no. 8, pp. 999-l011, Aug. 2003.

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

[8] Z. Yang, Q. Zhao, E. Keefer and W. Liu, "Noise Characterization, Modeling, and Reduction for In Vivo Neural Recording," *Advances in Neural Information Processing Systems*, pp 2160-2168, 2010.

[9] C. S. Lee, C. K. Lee and K. Y. Yoo, "New lifting based for undecimated wavelet transform," *Electronics Letters*, vol. 36, pp. 1894-1895, 2000.

[10] A. M. Kamboh, M. Raetz, K. G. Oweiss, A. Mason, "Area-power efficient VLSI implementation of multichannel DWT for data compression in implantable neuroprosthetics," *IEEE Trans. Biomed. Circuits and Systems*, vol. 1, no. 2, pp. 128-135, Jun. 2007.

[11] M. Aghagolzadeh, K. Oweiss, "Compressed and Distributed Sensing of Neuronal Activity for Real Time Spike Train Decoding," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, pp. 416-427, Apr. 2009.

[12] Y. Yang, A. Kamboh, A. J. Mason, "Adaptive Threshold Spike Detection using Stationary Wavelet Transform for Neural Recording Implants," *IEEE Biomed. Circuits Systems Conf.*, Cyprus, Nov. 2010.