ADAPTIVE NEURAL MATCHING ONLINE SPIKE SORTING VLSI CHIP DESIGN FOR WIRELESS BCI IMPLANTS

Zaghloul Saad Zaghloul, Magdy Bayoumi Center for Advanced Computer Studies (CACS) University of Louisiana at Lafayette LA, USA

ABSTRACT

Controlling the surrounding world by just the power of our thoughts has always seemed to be just a fictional dream. With recent advancements in technology and research, this dream has become a reality for some through the use of a Brain Computer/Machine Interface (BCI/BMI). One of the most important goals of BCI is to enable handicap people to control artificial limbs. Some research proposed wireless implants that do not require chronic wound in the skull. However, the communications consume a high bandwidth and power that exceeds the allowed limits, 8-10mW. This study proposes and implements a modified version of realtime spike sorting for wireless BCI [4] that simplifies and uses less computation via an adaptive neural-structure; which makes it simpler, faster and power and area efficient. The system was implemented, and simulated using Modalism and Cadence, with ideal case and worst case accuracy of 100% and 91.7%, respectively. Also, the chip layout of 0.704mm2, with power consumption of 4.7mW and was synthesized on 45nm technology using Synopsys.

Index Terms— BCI/BMI, VLSI, layout, Adaptive, Spike Sorting

1. INTRODUCTION

BCI/BMI is a system that provides communication between a living brain and a machine through translating neural electro-activity into computer commands. One of the most important goals of BCI research is to enable handicap people to control artificial limbs. The most common BCI sensors come mainly in two types: a wearable cap of sensor (e.g. 10-20 ISO), or an implant through an open wound. Both of these methods are non-practical for normal human daily use, or increase the infection hazard, respectively.

Wireless implants [1] do not require chronic wounds in the subject's skull. However, the transmission of the every single recorded neuron activity signal will consume a high bandwidth. In addition, the increased power consumption conflicts with the 8-10mW power limitations of the implant [2-3][7]. Thus, by adding the spike sorting circuit inside the implant directly after EEG signal acquisition will dramatically reduce the amount of data and the required bandwidth towards a more implementable BCI implant. In [4], a real-time spike sorting VLSI architecture for wireless implants was proposed to reduce complexity, bandwidth, and power consumption (see Figure 1).



Figure.1 The block diagram of the proposed Wireless BCI Sensor (spike sorting module: Neural Fingerprinting).

In this study, the authors modified, simplified the complexity, and added an adaptive matching behavior to neural-based real-time spike sorting architecture. Additionally, the circuit design, via Circuit Schematics and a Verilog code, was used for the simulation using ModelSim. Finally, the chip layout was synthesized on 45 nm technology using Synopsys[®]. The chip has an area of 0.704mm2, and consumes 4.7mW of power from a 1.1V voltage source.

2. BACKGROUND

2.1. Wireless BCI implants

Several wireless implants were proposed as a solution to the sensor placement, wiring, and usability issues [1]. It provides an acceptable SNR (~ 8.4 dB), and does not need an open wound to communicate with the controlled machine or device. It is based on high density microelectrode arrays of size 1cm2 that can record a group of 500 single neurons activity on the cortex of the brain, but it still needs a high bandwidth of RF to communicate with the controlled machine.

2.2. Power Consumption and Bandwidth limits

The bandwidth in the wireless communication plays a very important role due to the power and area of the implant. The input consists of a high amount of neural data coming from the microelectrode array. Research in [5] showed that 100electrode arrays sampled at 25 kHz per channel yield to 30 Mbps data rate. Such a bandwidth is huge for the basic wireless circuit that is needed to be fit into the implant. Therefore, this is one of the main limitations for wireless BCI implants from being used in the prosthetic limb applications.

2.4. Motivation and Approach

In this study, the main goal is to design neural spike sorting chip for a wireless BCI implant with an adaptive behavior that can be placed over the microelectrode array inside the patient's skull without using the regular EEG spike shape based sorting, clustering and feature extraction steps that requires a lot of calculations and can provide unsupervised, real-time neural spike sorting. The main challenge is to make the chip have:

- a low bandwidth (by adding a neuron classifier after the EEG signal acquisition and before the wireless transmission)
- a small area(by simplifying the large arithmetic units and reducing the chip layout routing)
- low power (through replacing the long instruction with parallel operation circuit of the control unit).

The next sections will briefly explain the related works followed by the proposed system architecture design. Then, the implementation, the simulation, and the results will be discussed, and, finally, the conclusions.

3. RELATED WORK

A VLSI friendly architecture [4] was proposed consisting of five units: EEG Sensors and input amplifiers, analog to digital converts (ADC), Digital Wavelet Transform (DWT), Neural Fingerprint and the output RF Module. The Neural Fingerprint Unit consists of: (1) squaring unit, (2) division unit, (3) comparator unit, and (4) a Memory.

This architecture reduces the signal transmission bandwidth by reducing the amount of data that needs to be transmitted, and, as a result, reducing the power consumption. The study in [5] describes the main implementation issues. Because the chip will be implanted in a living subject body, (1) the circuit area should be relatively small (~ 1cm2), (2) the energy dissipation should be low to not increase the temperature of the tissue ($\leq 1 \text{ C}^\circ$), and (3) the power consumption must be kept low (8~10mW).

The first successful system-on-chip (SoC) implantation was presented in [8] based on an ultra-wideband (UWB) environment which uses FIR filters and a tri-core CORDIC processor over a 64-channel microelectrode array. However, in this paper, we consider a spike sorting chip that is independent of the pre-conditioning filter systems. To our knowledge, the only related VLSI Implementation for a BCI neural unsupervised spike sorting chip was proposed in [9], and represented a multi-channel, online, unsupervised clustering spike-sorting DSP. A brilliant idea of adding spike ID instead of the value was also introduced, and the chip design was based on a two-stage implementation of an online clustering algorithm, a noisetolerant distance metric, and a selectively clocked, high-V register bank, Figure.2.



Figure.2 The two stages online neural spike sorting multi-channel cluster mapping algorithm with parallel channel identification block diagram. This system uses 16 input channels implemented in a 65nm technology, and has a power dissipation of 75μ W at a supply voltage of 270mV.

However, to reduce the system complexity and increase the processing speed of the chip to make it more feasible for implants, in this paper, we use a different approach to the reduce the arithmetical operations that are required by the mathematical model by using Neural network structure model approach with adaptive machine learning method. This method is based on considering the overall snapshot of the neural active area (the major firing spike recorded on a specific channel and the affected surrounding channels) instead of calculating each and every single neural spike.

4. NEURAL FINGERPRINT

The main idea of the neural fingerprint [4] is to capture the signature of each neural spike, and benefit from the neural echo effect captured via the adjacent channels. Thus, the system does not only consider a single neural spike activity on its major channel, but also its echo across all other affected surrounding channels.

Thus, given a time event t(i), *n* neurons, ch(i, 1) is channel *i* neuron *l*, and m channels, we search for the maximum of all coefficients over all the channels, where the maximum of channel *i* is defined as p(i). We define the pivot to be the maximum of all p(i)'s as follows:

$$p(i) = max \{ Ch(i, 1), ..., Ch(i, 1) \}$$
(1)

$$P = max \{p(1), ..., p(m)\}$$
 (2)

A fingerprint is constructed for each neuron j which have the highest firing value at Ch(j,m) in the form of a pair of vectors: channel fingerprint Chn.Fpt vector and nodal fingerprint *Nod.Fpt* vector as follows:

$$Chn.Fpt = (p(1)/P ... p(m)/P)$$
 (3)

$$Nod.Fpt = (Ch(j,1)/P \dots Ch(j,m)/P)$$
(4)

After the first fingerprint is calculated, it will be stored in a fingerprint lookup table (signature table) with an associated neural ID. Then, for each new EEG we calculate the corresponding neural fingerprint and search for a match. If a match is found, the system will mark the identified neuron as "detected" and sent its ID to the output; otherwise, the new fingerprint will be considered as a new neuron, and will be appended to the signature table with a new neural ID.

5. THE PROPOSED DESIGN

The proposed design consists of four main units: (1) Pivot Finder, (2) channel's and neutral's Fingerprint Generator, (3) Adaptive Matching Unit (AMU), and (4) a SRAM for the Neural Signature Table (NST) in addition to the global system controller. The block diagram is shown in Fingure.3.

Similar to [4], Figure.1, the inputs to the Neural Fingerprint Unit are coming from the Discrete Wavelet Transform (DWT) from a standard AMI 3Metal 2Poly $0.5\mu m$ CMOS technology with 4 channels EEG, where each DWT coefficients is represented in 10-bits [6] [10] independent of the Analog front-end or signal pre-conditioning phase or filter that could be added. This also helps the system scalability due to the modular design.

Unlike the previously proposed systems, our design reduces the complexity of the system by:

- replacing the squaring of the signed values with a two's complement unit which requires only one clock cycle,
- eliminating the divider (the most complex arithmetic unit) by just saving the divisor and the dividend values,
- using a fixed, two point calculation which still produces the required degree of accuracy.



Figure.3 The spike neural identification system architecture and block diagram

Additionally, an adaptive neuron matching unit was added to the system to increase the correctness of the neuron matching procedure within a higher SNR environment regarding Type I and Type II errors. The algorithm checks each newly generated fingerprint, and tries to matches it with a corresponding entry of all the NST. It will be discussed in the next subsections. The rest of this section explains the process steps and the associated implementation enhancements of the system.

5.1. Pivot Finding Unit

The Pivot Finding unit is the first step of the process. Thus, in order to increase the system speed and reduce the area, the comparator unit was replaced by a bitwise parallel comparison circuit for each of the four inputs. The bitwise parallel comparison circuit executes the Pivot Finding of the channel in O(1), as shown in Figure.4.



Figure.4 The Pivot Finding: a) max find unit, and b) parallel comparator basic building block.

5.2. Adaptive Matching Unit

Unlike other studies [11], here we drop the assumption of the Gaussian distribution of the neurons, which makes it closer to the real situation, but we still assume that each neuron has a unique firing action potential pattern, which is a generally agreed upon assumption. However, the neurons do not fire the exact waveform at every simulation event, as during the neural reading there are sometimes several small variations for the same group of neurons. It differs based on a number of factors: EEG noise, BCI artifacts, relaxation level on the human subject, environmental condition, and time when the signal was captured but still has a shared entropy. This means that the exact matching of the EEG values would result in large amounts of Type I errors.

Thus, an AMU addition was proposed to solve the neural match verities problem. The AMU is based on a parameterized Mean Square Error method, where we allow a Fuzzy Logic style matching behavior. Let us define x(i) to be a single field in the neuron fingerprint table stored in the

SRAM, and y(i) to be a neuron fingerprint value that needs to be matched. We define $\alpha(p)$ as the matching sensitivity for the pivot *P* of the fingerprint record, which is based on the principle of locality, such that:

$$x(i) = y(i) \pm \alpha(P) \tag{5}$$

Note that the equation is linear to keep the simplicity of the hardware implementation, and the value of $\alpha(P)$ was calculated by trial and error through a supervised learning phase during the simulation where an initial lookup table was constructed for the values of $\alpha(P)$, Also, a feedback $\underbrace{\underline{\xi}}_{0.4}^{0.6}$ connection provides the adaptive algorithm with a reward $\underbrace{\underline{\xi}}_{0.4}^{0.6}$ for every correctly matched neuron by increasing the weight of the $\alpha(P)$ for that specific neuron.

5.3. The Neural Signature Table

The neural signature table stores the neural fingerprints and the associated neural IDs. It is designed as a chip memory with width of (n+1)*m and depth of $n*\alpha(max)$, where n is the number of individual neurons, m is the number of channels, and α is the sensitivity. In our design, 30Kb of memory was allocated for 500 different neurons per a microelectrode array [1]. The memory word structure is shown in Figure.5.

Channel Finger Print				Local Finger Print				D	Pivot
CFP1	CFP2	CFP3	CFP4	LFP1	LFP2	LFP3	LFT4	Ν	Р

```
Figure.5 The SRAM field of memory word
```

However, these enhancements cost an increase in the SRAM size by 180% comparing to [4], and reduce both the clock cycle by 5x times and the required area of the arithmetic units by orders of magnitude. Therefore, makes this design still meets the total area limitations.

6. SIMULATION AND RESULTS

The Neural Fingerprint System was implemented in Verilog, and simulated using ModelSim software. The used sets of input data from the DWT unit and the SNR of the output is computed based on the technique described in [12] and used in [4] with real sample EEG data.

Seven different data sets were used with noise levels of: ideal case, and (4, 5, 6, 7, 10, 15) dB. The first set was used to get the a(P): P initial lookup table only during the design phase. The rest of sets were used for testing. The results show that, in the ideal case, the system perform at 100% accuracy, and with the noisy sets, the system gets a worst case accuracy of 91.7% which is still acceptable with the BCI performance limits (>90%).

The results were compared to previous work. We used the pair of (False Positive Rates (FPR), True Positive Rates (TPR)) [4] based on the Euclidean Distance, and we compared it with the ideal case (0, 1).

We calculated the range of the operation between the best and worst performance for our work as: Best False Positive (BFP), Worst False Positive (WFP), Best True Positive (BTP) and Worst True Positive (WFP). The best and worst performance for our work is: Best Proposed False Positive (BPFP), Worst Proposed False Positive (WPFP), as Best Proposed True Positive (BPTP) and Worst Proposed True Positive (WPFP). The result is shown in Figure.6.



Figure.6 Sample Neural Fingerprint Spike Sorting simulation

From the previous graph, we can see that the proposed design not only fits the overall performance region using less calculations but also has a noticeably (the solid lines) better performance in the best case for both FPR and TPR. Also, it has an almost equal performance for the worst case due to the adaptive behavior.



Figure.7 Neural Fingerprint Spike Sorting Chip Layout

The layout of the system was generated in 45 nm technology using Synopsys[®], which occupies area of 0.704 mm2 with 1109 cells, by consuming 4.7346mW of power from a 1.1 V source. The chip layout is shown in Figure.7 to demonstrate the relative area between the main building blocks.

7. CONCLUSION

We modified, reduced the complexity, and implemented an adaptive real-time BCI Neural Spike Sorting chip layout that meets the power, size, bandwidth, and accuracy constraints of the BCI wireless implants. The speed and power optimization come against the size, but still in the acceptable region. The adaptive system behavior increases the system accuracy with simplified design than the previous systems. However, by reducing the memory size and technology will help the realization of the wireless BCI implants for artificial limb control.

8. ACKNOWLEDGMENT

We would like to thank Nasirian Nasim, for helping with Cadence software, Amy Kern and Andy Andersen for a great help and support.

9. REFERENCES

[1] P. Mohseni, K. Najafi, S. Eliades, X. Wang, "Wireless multichannel bio-potential recording using an integrated FM telemetry circuit," IEEE Transaction on Neural Systems and Rehabilitation Engineering, vol. 13, No. 3, pp. 263–271, 2005.

[2] P. Merolla et al, "A million spiking-neuron integrated circuit with a scalable communication network and interface," AAAS, J. Neuroscience 345, pp. 668-672, ISSN 0036-8075, 2014.

[3] W. Biederman, D. Yeager, N. Narevsky, A. Koralek, "A Fully-Integrated, Miniaturized (0.125 mm^2) 10.5 μ W Wireless Neural Sensor," Berkele, USA. IEEE Journal of Solid-State Circuits, vol. 48, No. 4, pp. 960–970, 2013.

[4] F. Abu-Nimeh, M. Aghagolzadeh, K. Oweiss, "VLSI-friendly algorithm for real-time spike sorting in Brain Machine Interface applications," IEEE, BioCAS, 2008.

[5] A. Kamboh, M. Raetz, K. Oweiss, A. Mason, "Area-Power Efficient VLSI Implementation of Multichannel DWT for Data Compression in Implantable Neuroprosthetics," IEEE Transactions on Biomedical Circuits and Systems, vol. 1, No. 2, 2007.

[6] H. Cecotti, A. Graser, "Convolutional Neural Networks for P300 Detection with Application to BCI," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, No. 3, 2011.

[7] K. Mahajan, M. Vargantwar, S. Rajput, "Classification of EEG using PCA, ICA and Neural Network," International J. IJEAT ISSN: 2249–8958, vol. 1, No. 1, 2011.

[8] K. Abdelhalim, H. Jafari, L. Kokarovtseva, J. Velazquez, R Genov. "64-Channel UWB Wireless Neural Vector Analyzer SOC With a Closed-Loop PhaseSynchrony-Triggered Neurostimulator". IEEE Journal of Solid-State Circuits, vol. 48, No. 10, 2013.

[9] V. Karkare, S. Gibson, D. Marković, "A 75-μW, 16-Channel Neural Spike-Sorting Processor With Unsupervised Clustering", IEEE Journal of Soild-State Circuts, vol. 48, No. 9, 2013.

[10] K. Oweiss, A. Mason, Y. Suhail, A. Kamboh, Kyle E. Thomson, "A Scalable Wavelet Transform VLSI Architecture for Real-Time Signal Processing in High-Density Intra-Cortical Implants," IEEE Transaction on Circuits and Systems, vol. 54, No. 6, 2007.

[11] L. Miao, J. Zhang, C. Chakrabarti, A. Papandreou-Suppappola, N. Kovvali, "Real-time closed-loop tracking of an unknown number of neural sources using probability hypothesis density particle filtering," IEEE, SiPS, pp. 367-372, 2011.

[12] K. Oweiss, "A systems approach for data compression and latency reduction in cortically controlled brain machine interfaces," IEEE Trans. on Bio-Medical Engineering, 2006.