Low-Power Hardware Implementation of Movement Decoding for Brain Computer Interface with Reduced-Resolution Discrete Cosine Transform

Minho Won, Hassan Albalawi, Xin Li, *Senior Member*, *IEEE*, Donald E. Thomas, *Fellow*, *IEEE*

***Abstract*—This paper describes a low-power hardware implementation for movement decoding of brain computer interface. Our proposed hardware design is facilitated by two novel ideas: (i) an efficient feature extraction method based on reduced-resolution discrete cosine transform (DCT), and (ii) a new hardware architecture of dual look-up table to perform discrete cosine transform without explicit multiplication. The proposed hardware implementation has been validated for movement decoding of electrocorticography (ECoG) signal by using a Xilinx FPGA Zynq-7000 board. It achieves more than 56× energy reduction over a reference design using band-pass filters for feature extraction.**

I. Introduction

Over the past several decades, brain computer interface (BCI) has been considered as a promising communication technique for patients with neuromuscular impairments. For instance, neural prosthesis provides a direct control pathway from brain to external prosthesis for paralyzed patients. It can offer substantially improved quality of life to these patients. To create a neural prosthesis, we must appropriately measure the brain signals and then accurately decode the movement information from the measured signals [1]-[6].

A variety of signal processing algorithms have been proposed for movement decoding in the literature. Most of these algorithms first extract the important features to compactly represent the information carried by the brain signals. Next, the extracted features are provided to a classification and/or regression engine to decode the movement information of interest.

While most movement decoding algorithms in the literature are implemented with software on microprocessors, there is a strong need to migrate these algorithms to hardware in order to reduce the power consumption for practical BCI applications. Recently, significant efforts have been made to develop efficient hardware implementations for brain signal processing of epileptic seizure detection [7]-[10]. In this paper, we aim to extend these research works on hardware design and build a low-power hardware platform for movement decoding of BCI. Unlike the conventional feature extraction that relies on spectral density estimation [6] or band-pass filters [7], we adopt discrete cosine transform (DCT) to reliably extract BCI features with low power consumption. By taking advantage of the periodicity of cosine functions with reduced resolution, we are able to minimize the total number of multiplications required for feature extraction and, hence, substantially reduce the power consumption for movement decoding.

Moreover, we propose a new hardware architecture of dual look-up table (LUT) to implement the reduced-resolution DCT without explicit multiplication. By exploiting the fact that brain signal modulation often occurs within the same frequency bands across different channels [6], the proposed dual LUT can re-use the same set of cosine functions for feature extraction of multiple channels, thereby further reducing the power consumption.

Our proposed hardware implementation is applicable to a number of different BCI systems based on electroencephalography (EEG), electrocorticography (ECoG), etc. It has been validated by using a Xilinx FPGA Zynq-7000 board for a set of ECoG measurement data. As will be demonstrated by the experiment results in Section IV, our proposed hardware design achieves more than 56× energy reduction over a reference design using band-pass filters for feature extraction.

The reminder of this paper is organized as follows. In Section II, we propose the DCT-based feature extraction for movement decoding, and then describe our low-power hardware implementation in Section III. Experimental results are presented in Section IV to demonstrate the superior performance of the proposed hardware design over the reference design. Finally, we conclude in Section V.

II. Reduced-Resolution Discrete Cosine Transform for Feature Extraction

A. DCT-based Feature Extraction

Conventionally, a set of band-pass filters are used to extract the important features (i.e., the spectral densities within different frequency bands) of brain signals [7]. Filtering a given signal in the time domain involves a convolution between the signal samples and the filter coefficients. To complete such a convolution operation, a large number of multiplications and summations are needed. These multiplications are particularly expensive for hardware implementation and often dominate the overall power consumption. For this reason, various techniques have been explored in the literature to perform power-efficient multiplications, e.g., by using LUTs [7].

Instead of relying on band-pass filters, we propose to use DCT for feature extraction in this paper. Given a discrete-time

This work has been supported in part by the National Science Foundation under contract CCF–1314876.

M. Won, H. Albalawi, X. Li and D. E. Thomas are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (email: minowon@cmu.edu, halbalaw@cmu.edu, xinli@cmu.edu, dthomas@cmu.edu).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

signal $\{x_n, n = 0, 1, 2, \cdots, N-1\}$, DCT calculates the following coefficient [11]:

$$X_m = \sum_{n=0}^{N-1} x_n \cdot \cos\left[\frac{\pi}{N} \cdot m \cdot \left(n + \frac{1}{2}\right)\right]. \quad (1)$$

In (1), X_m quantitatively indicates the signal energy at a particular frequency. Hence, the DCT coefficient X_m is highly correlated to the spectral density calculated by a band-pass filter and it can be used to represent the important feature required for movement decoding.

Unlike the conventional filter-based approach that uses convolution to extract each feature, our DCT-based approach only needs to calculate the inner product of two vectors (i.e., the signal samples and the cosine function values). Hence, the number of multiplications is substantially reduced. In practice, multiple DCT coefficients should be calculated for several frequencies over a number of channels, resulting in a set of DCT-based features for decoding.

B. Reduced-resolution DCT

For low-power hardware implementation, floating-point arithmetic is overly expensive. Instead, fixed-point arithmetic must be used to minimize the power consumption. This has a two-fold meaning. First, the brain signals are quantized, e.g., by using an analog-to-digital converter (ADC) for the analog front-end. Second, but more importantly, the cosine function in (1) should also be quantized. As will be demonstrated by the experimental results in Section IV, the brain signals and the cosine functions can be quantized to 8 bits and 11 levels (i.e., less than 4 bits) respectively for our ECoG-based BCI without substantially sacrificing the decoding accuracy.

Note that the cosine function in (1) is periodic. Once the cosine function is quantized (e.g., to 11 levels only), its values become identical at a large number of time points. Hence, we can partition the signal samples $\{x_n, n = 0, 1, 2, \cdots, N-1\}$ into several groups where the quantized cosine function shares the same value within the same group. Next, instead of directly calculating N multiplications for the DCT in (1), we first sum the signal samples from the same group and then the partial sum is multiplied by the quantized cosine function.

For illustration purposes, we assume that the signal samples $\{x_n, n = 0, 1, 2, \cdots, N-1\}$ are partitioned into K groups when calculating the DCT-based feature X_m defined in (1):

$$\left\{x_{m,k,0}, x_{m,k,1}, x_{m,k,2}, \cdots\right\}, \quad (2)$$

where $x_{m,k,i}$ represents the i-th signal sample in the k-th group. The feature X_m in (1) is then calculated via two steps. First, the partial sum $s_{m,k}$ is calculated:

$$s_{m,k} = x_{m,k,0} + x_{m,k,1} + x_{m,k,2} + \cdots. \quad (3)$$

Next, all partial sums $\{s_{m,0}, s_{m,1}, s_{m,2}, \cdots\}$ are linearly combined to obtain the feature X_m:

$$, X_m = c_{m,0} \cdot s_{m,0} + c_{m,1} \cdot s_{m,1} + c_{m,2} \cdot s_{m,2} + \cdots, \quad (4)$$

where $c_{m,k}$ represents the quantized value of the cosine function corresponding to the k-th group.

The aforementioned approach is referred to as *reduced-resolution DCT* in this paper. It significantly reduces the number of the multiplications and, hence, the power consumption for feature extraction. In addition, our proposed feature extraction can be efficiently implemented with a LUT-based architecture for additional power reduction, as will be discussed in detail in the next section.

III. Low-Power Hardware Architecture for Movement Decoding of Brain Computer Interface

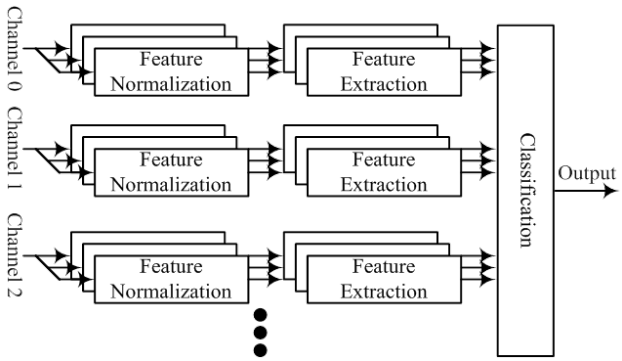

Figure 1. A simplified block diagram is shown for the proposed hardware implementation.

Figure 1 shows a simplified block diagram for the proposed hardware implementation. It consists of three major components: (i) feature normalization, (ii) feature extraction, and (iii) classification. In what follows, we will describe the implementation details of these three components with emphasis on power reduction.

A. Feature Normalization with Bit-shift Operation

The spectral densities of brain signals vary from subject to subject, from channel to channel and from frequency to frequency. For instance, it has been reported in the literature that the ECoG signal in high-gamma band (70~110 Hz) is highly correlated with movement, but its spectral density is substantially lower than that of the low-frequency signal [6]. Hence, the DCT-based features defined in (1) may substantially vary over a large range and representing these features by fixed-point arithmetic requires a large wordlength (i.e., a large number of bits).

In order to minimize the wordlength and, consequently, the power consumption for fixed-point computation, we must appropriately normalize the brain signal from each channel before calculating the DCT-based features. In our BCI application, the normalization factor is determined by estimating the range of each feature based on the training data. In addition, we constrain the normalization factor to be a power of two, i.e., 2^k where k is an integer. As such, the normalization can be performed by simply bit-shifting the signal, instead of dividing each brain signal by the corresponding normalization factor. It, in turn, reduces the power consumption required for feature normalization.

B. Feature Extraction with Dual Look-up Table

As previously stated, feature extraction dominates the overall power consumption of movement decoding in most cases. Therefore, building a power-efficient feature extraction engine is of great importance. Figure 2 shows the simplified block diagram for our proposed feature extraction engine. It is implemented with two LUTs to minimize the power

consumption.

1) LUT: Group Index

The first LUT defines the grouping structure for the signal samples $\{x_n, n = 0, 1, 2, \cdots, N-1\}$. For each sample x_n, its group index is stored in the LUT. Based on the group index, x_n is passed to one of the accumulators $\{s_0, s_1, s_2, \cdots\}$. The partial sums in (3) are then calculated at these accumulators.

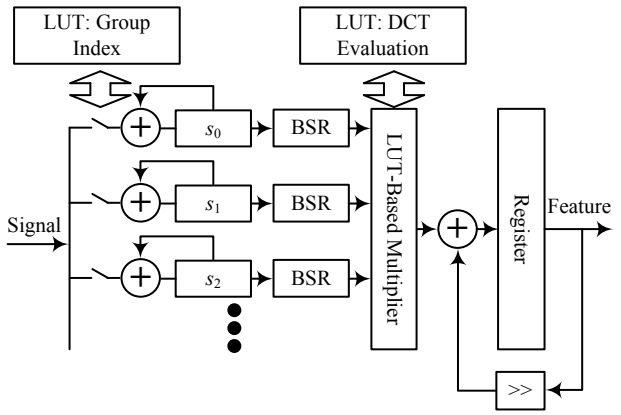

Figure 2. A simplified block diagram is shown for the proposed feature extraction engine with dual LUT.

2) LUT: DCT Evaluation

The second LUT "implicitly" multiplies the partial sum of each group with the quantized cosine function. Similar to the idea presented in [7], we first convert the partial sums $\{s_0, s_1, s_2, \cdots\}$ to a set of serial bits by using bit shift registers (BSRs). Next, these partial sums are implicitly multiplied by the quantized cosine function one bit by one bit using the LUT, instead of performing explicit multiplications. More details of the aforementioned LUT-based multiplication can be found in [7]. Finally, the weighted partial sums are added together to calculate the DCT-based feature in (4) and the final result is stored in the register shown in Figure 2.

It is worth mentioning that since brain signal modulation often occurs within the same frequency bands across multiple channels [6], the DCT-based features should be extracted at the same frequencies from all the channels. Hence, the two LUTs in Figure 2 can be shared for different channels, thereby reducing the hardware complexity.

C. Classification Engine

Once all DCT-based features are extracted for multiple channels, they are further combined to decode the movement information. In this paper, we focus on the problem of directional decoding based on a linear classifier. Namely, all features should be linearly combined to determine the movement direction of interest. Here, a variety of linear classification algorithms [12] (e.g., linear discriminant analysis, support vector machine, etc.) can be used, where the classifier training is performed offline. The on-chip classification engine performs the multiply-and-accumulate operations to determine the final output (i.e., the movement direction) from all DCT-based features.

IV. Experimental Results

The proposed hardware platform based on reduced-resolution DCT is implemented with a Xilinx FPGA Zynq-7000 board. For testing and comparison purposes, we further implement a reference design that extracts the BCI features by using a set of band-pass filters. In this section, we demonstrate the superior performance of our proposed hardware implementation over the reference design.

A. Movement Decoding Accuracy

To compare the decoding accuracy for DCT- and filter-based designs, we consider the ECoG data set collected from a human subject with tetraplegia due to spinal cord injury [6]. The ECoG signals are recorded with a high-density 32-electrode grid over the hand and arm area of the left sensorimotor cortex. The sampling frequency is 1.2 kHz. The human subject is able to voluntarily activate his sensorimotor cortex using attempted movements.

Our objective is to decode the binary movement direction (i.e., left or right) from a single trial that is 300 ms in length. The ECoG data set contains 70 trials for each movement direction (i.e., 140 trials in total). For both DCT- and filter-based methods, 7 important channels with 6 features per channel (i.e., 42 features in total) are selected based on the Fisher criterion [12]. Two linear classifiers are trained by using the DCT- and filter-based features respectively. When implementing both classifiers with hardware, we use 8-bit fixed-point arithmetic. For our proposed reduced-resolution DCT, the cosine function in (1) is quantized to 11 levels (i.e., less than 4 bits).

Table 1. Movement decoding accuracy for ECoG-based BCI

	DCT-based (Proposed)	Filter-based (Reference)
Fixed-point	82.9%	75.0%
Floating-point	75.7%	71.4%

Table 1 summarizes the decoding accuracy estimated by 5-fold cross-validation for the aforementioned ECoG data set. For comparison purposes, we list the decoding accuracy for both fixed- and floating-point arithmetic. Studying Table 1 reveals two important observations. First, the fixed-point implementation is even slightly more accurate than the floating-point implementation, probably because the quantization posed by fixed-point arithmetic partially removes the random noises of the ECoG signals. This observation demonstrates the feasibility of aggressively reducing the resolution of fixed-point arithmetic without surrendering any decoding accuracy.

Second, the DCT-based design is slightly more accurate than the filter-based design. It, in turn, implies that the proposed DCT-based features accurately capture the information that is required for movement decoding.

B. Power Consumption

We estimate the power and energy consumption for both DCT- and filter-based designs by using Xilinx Power Analyzer [13], where the clock frequency is set to 0.5 MHz. The ECoG data set described in Section IV.A is used as the input to both hardware designs for movement decoding. Table 2 summarizes the results. Note that the proposed DCT-based

design achieves more than 56× energy reduction over the filter-based design. Such a significant energy reduction comes from three major sources: (i) replacing the convolution operation of filter by the inner product operation of DCT, (ii) calculating the partial sums for reduced-resolution DCT to minimize the number of multiplications, and (iii) implementing the reduced-resolution DCT by LUT without explicit multiplication.

Table 3 further shows the power consumption for different functional blocks of the DCT-based design. Note that feature extraction dominates the overall power consumption for our proposed hardware implementation. Hence, additional efforts should be pursued to further reduce the power consumption of feature extraction in our future research.

Table 2. Power and energy consumption per decoding operation

	DCT-based (Proposed)	Filter-based (Reference)
Power (mW)	0.72	3.8
Runtime (ms)	1.094	11.71
Energy (μJ)	0.787	44.5

Table 3. Power consumption of different functional blocks for DCT-based design

Feature Normalization (μW)	25.2
Feature Extraction (μW)	690.2
Classification (μW)	2.6

C. FPGA Board-level Validation

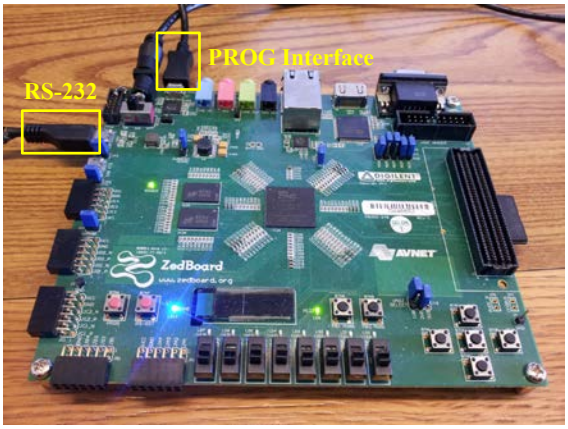

Figure 3. A Xilinx FPGA Zynq-7000 board is used to validate the proposed DCT-based hardware design for movement decoding.

To validate the proposed DCT-based hardware design on the Xilinx Zynq-7000 board, we first load our hardware design to the FPGA chip through the programming interface. Next, the ECoG data set is copied to an SD card that is connected to the Zynq-7000 board. When running the movement decoding flow, a single trial of the ECoG signals is first loaded to the SRAM block inside the FPGA chip. Next, these signals are passed to the functional blocks of feature normalization, feature extraction and classification for decoding. The decoding results are read back to an external computer through an RS-232 serial port on the Zynq-7000 board so that we can verify the decoding accuracy. Figure 3 shows a photograph of the Xilinx FPGA Zynq-7000 board where the RS-232 port and the programming interface are both highlighted.

V. Conclusion

In this paper, we develop a novel hardware implementation for movement decoding of BCI. Our proposed design minimizes the power consumption via two complementary avenues. First, a novel feature extraction method is proposed to reduce the computational complexity by using reduced-resolution DCT. Second, a new hardware architecture of dual LUT is developed to implement the reduced-resolution DCT without explicit multiplication. Our proposed DCT-based design has been validated by using a Xilinx FPGA Zynq-7000 board. The experimental results demonstrate that it achieves more than 56× energy reduction over a reference design using band-pass filters for feature extraction. As an important aspect of our future research, we will further design an application-specific integrated circuit (ASIC) to implement the proposed hardware design for movement decoding and integrate the IC into a practical BCI system.

Acknowledgment

The authors wish to thank Dr. Wei Wang from University of Pittsburgh for sharing his valuable experience and measurement data of ECoG-based BCI.

References

[1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of The IEEE*, vol. 89, no. 7, pp. 1123-1134, Jul. 2001.

[2] E. Leuthardt, et al., "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 1, no. 2, pp. 63-71, Jun. 2004.

[3] M. Lebedev and M. Nicolelis, "Brain-machine-interface: past, present and future," *Trends in Neurosciences*, vol. 29, no.9, pp.536-546, Sep. 2006.

[4] A. Schwartz, et al., "Brain-controlled interfaces: movement restoration with neural prosthetics," *Neuron*, vol. 52, no. 1, pp. 205-220, Oct. 2006.

[5] J. Donoghue, "Bridging the brain to the world: a perspective on neural interface systems," *Neuron*, vol. 60, no. 3, pp.511-521, Nov. 2008.

[6] W. Wang, et al., "An electrocorticographic brain interface in an individual with tetraplegia," *PLoS One*, vol. 8, no. 2, pp. 1-8, Feb. 2013.

[7] J. Yoo, et al., "An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 214-228, Jan. 2013.

[8] M. Altaf, et al., "A 1.83μJ/classification nonlinear support-vector-machine-based patient-specific seizure classification SoC," *International Solid-State Circuits Conference*, pp. 100-101, Feb. 2013.

[9] W. Chen, et al., "A fully integrated 8-channel closed-loop neural-prosthetic SoC for real-time epileptic seizure control," *International Solid-State Circuits Conference*, pp.286-287, Feb. 2013.

[10] K. Lee and N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 7, pp. 1625-1637, Jul. 2013.

[11] N. Ahmed, et al., "Discrete cosine transform," *IEEE Trans. on Computers*, vol. 23, no.1, pp. 90-93, Jan. 1974.

[12] C. Bishop, *Pattern Recognition and Machine Learning*, Prentice Hall, 2007.

[13] *Xilinx XPower Estimator User Guide*, http://www.xilinx.com/support/documentation/user_guides/ug440.pdf.