A Low Power Seizure Detection Processor Based on Direct Use of Compressively-Sensed Data and Employing a Deterministic Random Matrix

Ali Jafari, Adam Page, Chris Sagedy, Emily Smith, Tinoosh Mohsenin
Department of Computer Science & Electrical Engineering
University of Maryland, Baltimore County

***Abstract*—This work presents a low power multi-channel seizure detection processor based on Compressive Sensing (CS) algorithm. Direct use of compressively-sensed data is proposed for feature extraction and classification in order to reduce computational and data transmission energy due to reduced number of input samples. To further reduce power consumption of the system, using a deterministic random matrix (DRM) is proposed instead of implementing a random number generator circuit (LFSR) for compressive sensing circuit. For feature extraction, simple features are used and for classification Logistic Regression (LR) is employed. Three different architectures are implemented in Virtex-5 FPGA and are compared against each other. For compression rates of 2-16x, the energy consumption of the proposed 22-channel seizure detection processor including CS, feature extractor and classifier is 2.36-0.38 μJ and the detector performance for sensitivity and specificity is 80.7-78.8% and 85.3-83.5%, respectively. The proposed system with 16x compression rate consumes 0.38 μJ which is 6 times lower than the system without using compressive sensing.**

***Index Terms*—Compressive Sensing, Seizure Detection, Logistic Regression, FPGA, Energy Efficiency, ASIC.**

I. Introduction

Epilepsy is a chronic neurological disorder characterized by recurrent epileptic seizures; a seizure in turn is a transient of symptoms due to abnormal excessive or synchronous neuronal activity in the brain [1]. Epilepsy is the third most common neurological disorder affecting more than 50 million people in the world. The fear of sudden unexplained death in epilepsy (SUDEP) is always a concern for patients with epilepsy; almost a quarter of deaths in people with epilepsy are thought to be attributable to SUDEP. Hence, continuous monitoring of seizure patients is a critical task. On the other hand, constant monitoring in the hospital is very expensive and also frustrating for patients. Thus, wearable in-home monitoring systems have been promising for seizure detection [2], [3].

There has been a vast improvement in wearable in-home monitoring systems for seizure detection, but power consumption is still a large concern because these systems are typically wireless and use a battery as a power supply. Hence, the high energy cost to transmit a bit of information and the radios limited bandwidth necessitate some types of data compression at the sensor node to reduce energy consumption and data throughput [4].

Fig. 1 shows different system block diagrams used at the sensor side of a bio-medical monitoring system. Model 1 shows a system which captures signal from analog front end (AFE) sensor and sends all raw data through a transmitter to a receiver side for processing, such as feature extraction and classification. The drawback of this type of system is high power consumption due to constantly transmitting raw data. Assuming AFE takes 256 samples 16 bits each, the transmitter needs to send 90,112 bit/s (for 22-channel). In Model 2, the system compresses the input signal (using CS) and sends the compressed data out. For example for a compression rate of 16×, the transmitter needs to send approximately 5,632 bit/s, constantly. Although Model 2 reduces the data transmission rate, this system is still not suitable for a wearable monitoring system due to the high power consumption, even by ignoring the fact that this system needs a reconstruction algorithm in receiver side [5]. Model 3 shows a system which does all the signal processing at the sensor node and only sends the final decision through the transmitter [6], [7]. This system consumes less power than Model 1 and 2 as it needs to send 1 bit/s and also the transmitter can be in sleep mode while processing.

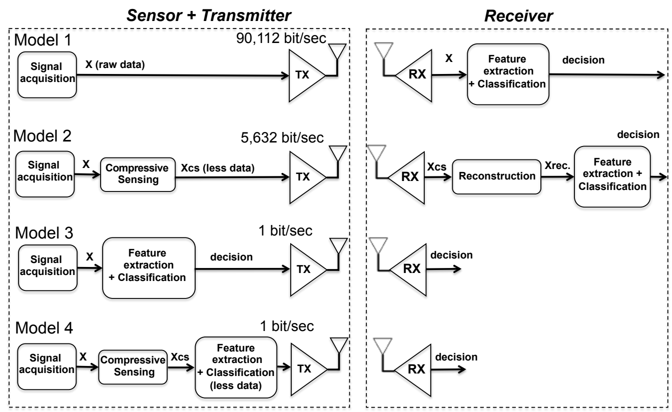


Fig. 1. Four different system Models which can be used in bio-sensor applications. Assuming the TX power scales with data rate.

In this paper, a new system as shown in Model 4 is proposed which compresses sensor data using CS method and performs all processing directly on the compressively-sensed data, locally at the sensor. The proposed approach has the following advantages which reduces the overall power consumption.

- All processing is done on sensor and only final decision is sent out (one bit), rather than sending all raw data.

- There is no need for any reconstruction algorithm which consumes a considerable amount of power. Thus, it is a more efficient system in comparison with Model 2.
- Using CS, less number of input samples (depending on compression rate) are used for feature extraction in comparison with system shown in Model 3.
- Employing a DRM eliminates the need for a LFSR which requires extra hardware and latency [4].

II. System Design

An overview of the proposed system architecture is discussed here. The system consists of a compressive sensing front-end, a feature extractor to produce features from raw data, and a machine learning classifier.

A. Compressive Sensing

CS provides a structure to acquire data at a rate proportional to information rate rather than rate required by the Nyquist rate [8]. In CS, an N-sample signal is multiplied by an $M \times N$ projection matrix to create an M-sample signal (with $M < N$).

$$Y = \Phi X \tag{1}$$

Where X is input signal, Y is compressed version of the input signal and Φ is sensing matrix. CS theory relies on two principles: (1) the N-sample signal is sparse in a secondary basis Ψ, meaning it has K non-zero entries and the remaining N–K entries are all zero. (2) Φ and Ψ are incoherent from each other and satisfy the restricted isometry property [9]. Random sensing matrices with sufficient sample size exhibit low coherence with any fixed basis. Traditionally, Gaussian random matrix or pseudo-random Bernoulli matrix have been employed to create a sensing matrix. However, using a deterministic random matrix is proposed in this work to decrease the hardware cost and consequently power consumption while retaining a good performance close to a non-deterministic random matrix. This matrix can be created by randomly choosing a subset of the rows of an identity matrix [10].

B. Feature Extraction

Using raw time series data as input to most classifiers results in low accuracy, making feature extraction a crucial task. A trained human can determine if a patient is having a seizure with almost perfect accuracy simply by looking at the EEG waveform, which implies that useful features provide information about the shape of the waveform. These features include area under the wave, normalized decay, line length, average peak amplitude, and average valley amplitude [6], [11]. The formulas for these five features are given in Table I.

C. Machine Learning Classifier

Logistic Regression is chosen as it was shown by authors that it has the minimum complexity and power consumption among other machine learning classifier algorithms [6].

LR is a discriminative probabilistic model that utilizes the logistic function to map real-valued inputs between 0 and 1 which can be interpreted as probabilities. Training is

TABLE I

Formulas for 5 features extracted from raw EEG signal in each window: Area under the wave, normalized decay, line length, avg peak amplitude, and avg valley amplitude. In this work a one second window (256 samples) is used. W = *window length*, x = *input*, P = # *peaks*, V = # *valleys*

Area Under Curve	Normalized Decay	
$A = \frac{1}{W}\sum_{i=0}^{W-1} x_i$	$D = \frac{1}{W-1}\lvert\sum_{i=0}^{W-2} I(x_{i+1} - x_i < 0) - 0.5\rvert$	
Line Length	**Avg Peak Amplitude**	**Avg Valley Amplitude**
$\ell = \sum_{i=1}^{W} -1\lvert x_i - x_{i-1}\rvert.$	$P_A = \log_{10}\left(\frac{1}{P}\sum_{i=0}^{P-1} x^2_{p(i)}\right)$	$V_A = \log_{10}\left(\frac{1}{V}\sum_{i=0}^{V-1} x^2_{v(i)}\right)$

an iterative process that finds maximum likelihood regression coefficients (weights). Samples are typically classified by selecting the label with the greater probability between $P(Y = 1|X)$ and $P(Y = 0|X)$ given in (2), where $w_{k,i}$ are the weights for each of the N features for the label, k, and X is the input vector:

$$P(Y = k|X) = \frac{1}{1 + exp\left(w_{k,0} + \sum_{i=1}^{N} w_{k,i} x_i\right)} \tag{2}$$

Also, probability ratios are used which significantly reduce computational complexity as shown in (3).

$$\frac{P(Y = 1|X)}{P(Y = 0|X)} \Rightarrow exp\left(w_0 + \sum_{i=1}^{N} w_i x_i\right) \Rightarrow \sum_{i=0}^{N} w_i x_i \tag{3}$$

D. Seizure Data and Performance Measurement

The data used to evaluate the performance of the proposed system consists of EEG recordings from 22 subjects (5 males, ages 3-22; and 17 females, ages 1.5-19) with intractable seizures collected at the Children's Hospital Boston [12]. In all, there was a total of 336 hours of data containing 184 seizure onsets. This data is used as the input to the proposed CS for different compression rates (CR), up to 16×. A window of 256 samples (one second per channel) is chosen for the input of CS. As it is mentioned in section II.A, a deterministic random matrix is used for sampling. Then, the compressed data are used to create 5 simple features for each second of EEG data per channel. A LR classifier uses these features to produce the decision per window.

Performance of the seizure detector is characterized in terms of sensitivity and specificity. Sensitivity refers to the percentage of seizures onsets identified. Specificity refers to the percentage of incorrectly detected seizures onsets.

To estimate the detector performance on the data from a patient, a leave-one-record-out cross-validation approach is used [13]. Let NNS denote the number of non-seizure records and NS denote the number of seizure records. To estimate the detector sensitivity, the detector is trained on NNS non-seizure records from a patient with $NNS = 37$ and on $NS - 1$ seizure records with $NS = 5$. The detector is then tasked with detecting the seizure in the withheld seizure record. This process is repeated NS times so that each seizure record is tested. For each round, the detected test seizure is recorded and sensitivity is measured. To estimate the detector specificity, the detector is trained on NS seizure records and on $NNS - 1$ non-seizure records. Then, the detector is tested on the withheld non-seizure record. This process is repeated NNS times and specificity is measured.

III. Results

A. Analysis of Seizure Detection Performance with direct processing on compressively sensed data

The proposed system could achieve a sensitivity of 80.8% and specificity of 86.5% for Nyquist-domain seizure data (without compression). Fig. 2. shows the seizure detector performance using compressively sensed data for compression rates from 2× up to 16×. As it can be seen from the figure, the degradation in sensitivity and specificity is approximately 2% and 3%, respectively, up to a compression rate of 16×.

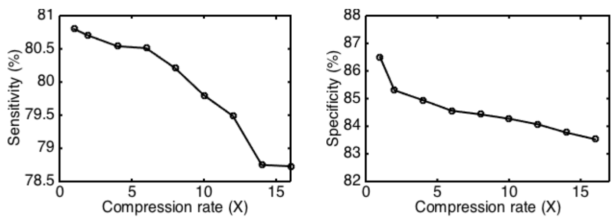


Fig. 2. Performance of the proposed seizure detection algorithm with direct use of compressively-sensed signals and employing a deterministic random matrix. Detector performance is well maintained for compression rates from 2× up to 16×.

B. FPGA Implementation

Fig. 3 shows the top-level block diagram of the proposed seizure detector hardware system including compressive sampler, feature extractor and classifier. CS samples the input data and gives the compressed data to the next stage which is Feature Extractor. The CS architecture using a deterministic random matrix is shown in Fig. 4(a). A block RAM is used to store one window of input signal (in this work 256 samples) for sampling purpose. Furthermore, a small ROM is used to store the locations of non-zero entries of deterministic random matrix. Indeed, in this method there is no need to save all zeros and ones in the ROM. Furthermore, to prove that CS architecture using deterministic random matrix consumes less power than previous works [4], [14], a CS module using LFSR circuit (to create random matrix) is implemented (Fig. 4(b)).

TABLE II
Comparison of dynamic energy consumption of CS-DRM and CS-LFSR modules shown in Fig. 4 which are implemented on Virtex-5 FPGA. Results are based on $CR = 2 - 16\times$.

Design	CS-DRM	CS-LFSR
Latency (Cycle)	256-32	98,304-12,228
Slice Count	21-12	36-34
Memory (KB)	5.12-4.2	4.1
Leakage Power (mW)	1042.5	1042.5
Dynamic energy (nJ)	16.15-1.87	8,326-1,073

Table II shows latency and dynamic energy consumption of DRM and LFSR. The latency of CS-DRM is 384 times less than using CS-LFSR and the energy consumption is 515-573 times smaller for compression rate of $CR = 2 - 16\times$. Hence,

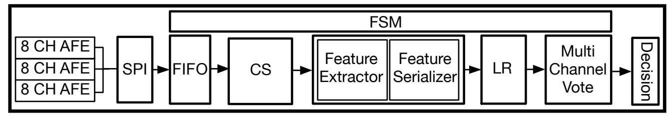
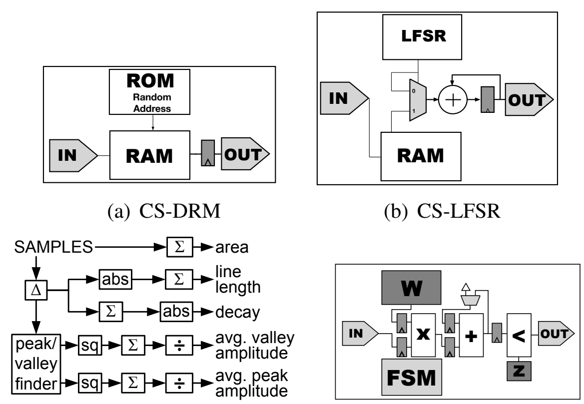


Fig. 3. Block diagram of the proposed seizure detection processor containing compressive sampler, feature extractor, classifier, multichannel vote, and IO interface. Note that SPI and AFE blocks are not implemented.

using the proposed low power CS-DRM reduces total energy consumption drastically.

The Feature Extractor block which is shown in Fig. 4(c) receives CS outputs serially and calculates five features in parallel. The serializer buffers the five features and sends them serially to the classifier block. A detailed architecture of the feature extraction process is shown in Fig. 4(c).

The LR classifier serially outputs one classification per channel to the Multi-Channel Vote block, which employs majority voting. The final decision is reported as a seizure if the number of channels that are classified as a seizure exceeds a predefined threshold. The FSM block controls the flow for all blocks. Details of the LR classifier used for the proposed processor is shown in Fig. 4(d). LR classifier has very straightforward implementations that follow the derivation given in (3). It contains just one lookup table to store the regression weights.



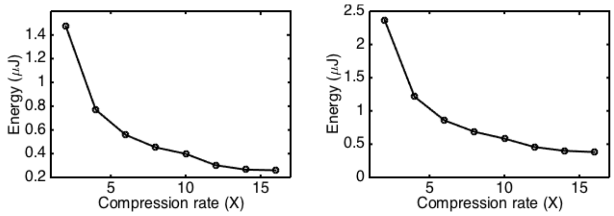
(a) CS-DRM (b) CS-LFSR

(c) Feature Extractor (d) LR classifier

Fig. 4. Block diagrams of CS, Feature Extractor and LR classifier.

Fig. 5 shows the effect of compression rate scaling on the feature extractor and total processor energy in the proposed system (Fig. 1, Model 4). As it can be seen from the figure, the energy decreases as the compression rate increases. However, the classifier energy doesn't change as the number of input samples to the classifier remains 5 samples for any compression rate.

Table III compares the performance of different seizure detection system Models shown in Fig. 1. All power results are for a complete system processing of one window EEG data (one second over 22 channels) and using SPBT2632C2A (Bluetooth module from STMicroelectronics) transmitter module (active and sleep modes energy are included). Also, note that the results for the energy of Model 4 in table III and



(a) Feature Extractor Energy (b) Total Processor Energy

Fig. 5. Feature Extractor and total processor energy of Model 4 for different compression rate up to 16×.

Fig. 5 are not similar, as the transmitter energy is added in table III which is not included in Fig 5. The results show that the proposed Model (Model 4) ranks best in dynamic energy consumption. Models 3, 2 and 1 rank second, third and fourth, respectively. Compared to Model 1 and 2, the proposed Model reduces the energy consumption by factor of 195× and 13×, respectively. Furthermore, Model 3 and 4 consume 183 μJ and 181 μJ respectively; however, as total energy of the processor is very small (as it can be seen from Fig. 5), the energy values for Model 3 and 4 in table III are dominated by transmitter energy. Hence, ignoring transmitter energy for Model 3 and 4 (they are almost same for Model 3 and 4), the proposed Model reduces energy consumption by factor of 6× for a compression rate of 16× compared to Model 3.

TABLE III

Comparison of power consumption for different seizure detection system Models shown in fig. 1 on Virtex-5 FPGA. The numbers are based on energy of CS, feature extractor, classifier and transmitter modules. Results for Model 2 and 4 are for *CR* = 16×. The transmitter power is based on Bluetooth SPBT2632C2A.

Design	Model1	Model2	Model3	Model4
Specificity (%)	86.5	86.5	86.5	83.5
Sensitivity (%)	80.8	80.8	80.8	78.8
Latency (Cycle)	*	704	5786	858
Slice Count	*	30	2583	2613
Memory (KB)	*	4.6	0.3	4.9
Leakage Power (mW)	*	1042.5	1042.87	1042.99
Dynamic energy (mJ)	35.13	2.366	0.183	0.181

IV. Discussion on ASIC implementation and comparison

In our previous work, a 22-channel seizure detection processor with simple feature extraction and LR classifier and without CS block was implemented in 65-nm CMOS technology, which occupied 0.008 mm^2 and consumed 19 nJ at 1 V [6]. From Table II, it is observed that the CS block overhead is very small, thus it is projected that ASIC implementation results of the proposed processor will follow the same trend as the FPGA implementation with overall smaller power consumption. The work presented in [14] also uses direct compressively-sensed data processing followed by FIR filters for feature extraction and SVM for classifier and is implemented in 0.13um CMOS, which consumes 7.13-0.11 μJ for the compression rate of 2-24 and 18 EEG channels.

V. Conclusion

This work presented a low power multi-channel seizure detection system based on direct use of compressively sensed EEG data. Using this approach not only eliminates reconstruction costs, but also it leads to computational energy savings due to the reduced number of input samples. Furthermore, using a deterministic random matrix is proposed to further reduce the power consumption of compressive sampler as there is no need to implement a random number generator circuit. The proposed system is fully implemented in Virtex-5 FPGA and allows transmission rate to be reduced from 90,112 bit/s and 5,632 bit/s to 1 bit/s in comparison with Model 1 and 2, respectively. Consequently, the dynamic energy reduced from 35.13 mJ and 2.366 mJ to 0.181 mJ (195 times smaller than Model 1, and 13 times smaller than Model 2). Furthermore, while transmission bit rate is same for Model 3 and 4, total dynamic energy for Model 4 is reduced by factor of 6× for a compression rate of 16× (without including transmitter energy). The proposed approach can be applied as a general framework to other low power bio-medical applications.

Acknowledgment

This material is based upon work partly supported by the National Science Foundation under Grant NO.1350035.

References

[1] G. Ullah and S. J. Schiff, "Models of epilepsy," *Scholarpedia*, vol. 4, no. 7, p. 1409, 2009.

[2] A. Page *et al.*, "An ultra low power feature extraction and classification system for wearable seizure detection," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, Aug 2015.

[3] D. Chandler Jr *et al.*, "Real-time multi-channel seizure detection and analysis hardware," in *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE*. IEEE, 2011, pp. 41–44.

[4] F. Chen *et al.*, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 3, pp. 744–756, 2012.

[5] A. Kulkarni and T. Mohsenin, "Accelerating compressive sensing reconstruction omp algorithm with cpu, gpu, fpga and domain specific many-core," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 970–973.

[6] A. Page *et al.*, "A flexible multichannel eeg feature extractor and classifier for seizure detection," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 62, no. 2, pp. 109–113, 2015.

[7] S. Viseh, M. Ghovanloo, and T. Mohsenin, "Toward an ultralow-power onboard processor for tongue drive system," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 62, no. 2, pp. 174–178, 2015.

[8] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[10] C. Moler, "magic reconstruction: Compressed sensing," *Mathworks News & Notes*, 2010.

[11] D. Wulsin *et al.*, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of neural engineering*, vol. 8, no. 3, p. 036015, 2011.

[12] A. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[13] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 975–982.

[14] M. Shoaib *et al.*, "A compressed-domain processor for seizure detection to simultaneously reduce computation and communication energy," in *Custom Integrated Circuits Conference (CICC), 2012 IEEE*, pp. 1–4.