

A Resource-Optimized VLSI Implementation of a Patient-Specific Seizure Detection Algorithm on a Custom-Made 2.2 cm² Wireless Device for Ambulatory Epilepsy Diagnostics

Tianyu Zhan, Syyeda Zainab Fatmi, Sam Guraya, Hossein Kassiri

Abstract—A patient-specific epilepsy diagnostic solution in the form of a wireless wearable ambulatory device is presented. First, the design, VLSI implementation, and experimental validation of a resource-optimized machine learning algorithm for epilepsy seizure detection are described. Next, the development of a mini-PCB that integrates a low-power wireless data transceiver and a programmable processor for hosting the seizure detection algorithm is discussed. The algorithm uses only EEG signals from the frontal lobe electrodes while yielding a seizure detection sensitivity and specificity competitive to the standard full EEG systems. The experimental validation of the algorithm VLSI implementation proves the possibility of conducting accurate seizure detection using quickly-mountable dry-electrode headsets without the need for uncomfortable/painful through-hair electrodes or adhesive gels. Details of design and optimization of the algorithm, the VLSI implementation, and the mini-PCB development are presented and resource optimization techniques are discussed. The optimized implementation is uploaded on a low-power Microsemi Igloo FPGA, requires 1237 logic elements, consumes 110 μ W dynamic power, and yields a minimum detection latency of 10.2 μ s. The measurement results from the FPGA implementation on data from 23 patients (198 seizures in total) shows a seizure detection sensitivity and specificity of 92.5% and 80.1%, respectively. Comparison to the state of the art is presented from system integration, the VLSI implementation, and the wireless communication perspectives.

I. INTRODUCTION

Recurrent and unpredictable interruptions of normal brain function due to transient occurrence of symptoms of abnormal excessive or synchronous neuronal activity in the brain is commonly known as seizures. A neurological disorder characterized by an enduring predisposition to generate seizures is called epilepsy [1]. It is the most common chronic brain disorder and currently affects more than 50 million people of all ages worldwide.

There are three main categories of therapy known for patients with epilepsy: pharmacological solutions, brain surgery, and artificial electrical brain stimulation. An estimated 65-to-70% of the patients could be seizure-free if properly diagnosed and receive the right dosage of anti-epileptic drugs in a timely manner [2]. Nearly half of the remaining one-third of the patients who are refractory to the currently-available pharmacological treatments are accepted as candidates for

This work was supported by Natural Sciences and Engineering Research Council of Canada and Canadian Microelectronics Corporation (CMC).

H. Kassiri, S. Z. Fatmi, and S. Guraya are with the Department of Electrical Engineering and Computer Science, York University, Toronto, ON M3J 1P3, Canada (email: Kassiri@yorku.ca).

T. Zhan is with the EE Department, Beijing Institute of Technology, China.

brain resection surgery. For the approximately 15% remaining patients, the only available option is electrical stimulation of neuronal systems to disrupt the synchronous neuronal activity in the brain by injecting electrical pulse trains in close proximity of the seizure origin in the brain [3].

Despite their fundamental differences in their seizure control mechanisms, all three treatment options require monitoring the patient's brain activity with high temporal resolution. For patients under drug treatment, both timing and dosage of drug release must be customized for each patient depending on the severity, frequency, and type of their seizures [4]. Currently, the drug dosage is decided by the physicians based on the patient's diary and their answers to a few behavioural questions. Such practice fails to take into account seizures that happen during the sleep, as well as the severity of an episode, especially for absence (i.e., non-convulsive) seizures, which amount to approximately one-third of all events [5]. It also completely puts the burden of seizure logging on the patients, making it completely useless for a large number of infant patients. Therefore, a more patient-independent objective seizure monitoring method seems to be required for maximizing the efficacy of pharma-based epilepsy therapies.

For patients selected as candidates for surgery, the seizure focus must be precisely identified through a multi-modal imaging process collectively known as pre-surgical monitoring. Currently, various methods such as sleep deprivation and photic stimulation are used to induce seizures while the patient is tethered to electroencephalography (EEG) recording setup for continuous monitoring. Given that for some patients seizures might happen every few days, such tests impose a very high level discomfort and the recorded seizures are triggered in an unnatural way, making the extracted information less reliable. Moreover, the very limited number of such facilities and the required trained personnel operating them has resulted in wait times as long as a year [6].

Similar to drugs and surgery, the success rate of the brain neuro-stimulators heavily depends on real-time monitoring of neuronal activities [25], [26], [29], [30]. Modulation of the brain activity through electrical stimulation was conventionally done in an open-loop/blind fashion, i.e., injecting electrical pulses to the brain periodically without any prior knowledge of the brain state. In recent years, a few companies such as Neuropace and Medtronic have introduced closed-loop neuro-stimulators where stimulation is only triggered upon detection of an abnormal activity in the brain [3]. However, the best commercially-available solution today can only render 13% of patients seizure free, partly due to the tight power/area budget

of implantable devices that strictly limits the computational complexity of the embedded signal processing algorithms [7].

Given the superiority of surface EEG recording compared to brain imaging technologies such as MRI in terms of temporal resolution, and its advantage over implantable solutions in terms of invasiveness, spatial coverage, and computational power, it seems to be the optimal choice of technology for long-term monitoring of brain neuronal activities for patients with epilepsy, independent of their treatment choice. With the right implementation, a wearable EEG recording/processing solution could (a) provide objective logging of seizure origin, severity, and frequency in patients under a drug-based treatment plan, (b) allow pre-surgical ambulatory EEG monitoring of patients who are considered as candidates for brain resection surgery in a comfortable and well-paced manner, and (c) offers the means for the physician to monitor the efficacy of an implantable neuromodulation device and to continuously improve it over time by adjusting the stimulation parameters.

Motivated by the above, the development of such small-form-factor wearable wireless devices has been investigated over the past decade. Such a solution is envisioned to enable EEG recording and analysis with medical-grade quality, while minimizing the time and staff/equipment resources required. Beside meeting electrical design specifications such as wireless connectivity, long battery life, and high signal-to-noise ratio (SNR), one of the most primitive yet crucial requirements for an ambulatory EEG monitoring solution is the practicality of its use while the patients do their normal day-to-day activities. This simple criterion disqualifies several of commercially-available wearable solutions due to their large form factor or heavy weight for everyday use [8], [9]. Reviewing these devices, their undesirable physical attributes seems to be mainly rooted in the use of bulky through-hair dry contact electrodes such as [11] and [13]. In addition to bulkiness, these electrodes also cause discomfort, and in some cases pain, when used for a long time.

It should also be noted that for the envisaged wearable solution, a monolithic integrated circuit implementation, such as those presented in [27], [39], is not necessarily desired over a discrete implementation. This is simply due to the large physical distance between the electrodes in surface EEG recording compared to implantable devices used for electrocorticography and intracranial EEG. For a wearable EEG device, using a monolithic solution that integrates all the recording and processing circuitry in a mm-scale IC results in the weak EEG signals traveling up to 10 cm to the chip before being amplified, which could easily lead to being contaminated/drowned with noise/interference. To address this issue, we have used a design based on active electrodes that amplify and digitize the EEG signals at each electrode site before sending them to the central DSP unit for processing.

Unlike through-hair electrodes, flat dry electrodes can be integrated into the solution with minimum to no increase in the system size [10], [15]. This has enabled the implementation of wearable solutions in the form of a light comfortable headbands such as [16] and [10]. However, use of these electrodes comes with two main disadvantages: (a) motion artifacts: with no conductive adhesive gel used to support

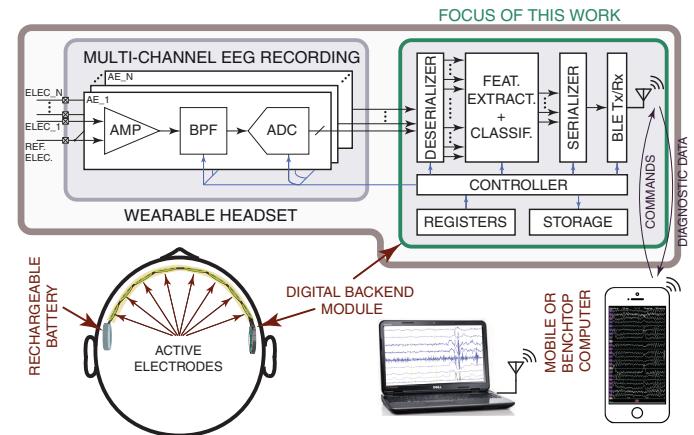


Fig. 1: Top: Top-level block diagram of a multi-channel wireless wearable surface EEG recording and processing device. Bottom: Envisioned use of the device as a wearable headset.

the electrode-skin contact, dry electrodes are generally prone to artifacts induced by electrode motions; (b) limited spatial coverage: with no mechanism to penetrate through hair, flat dry electrodes are only suitable for EEG recording from hairless skin (e.g., forehead), preventing them from conducting a full (e.g., 10-20 standard) EEG recording. To address the first issue, we have previously shown analog/mixed-signal methods to detect and compensate motion artifacts in [10]. One of our goals in this work is to investigate the possibility and the efficacy of detection of a neurological event (i.e., an epilepsy seizure) despite the spatial-coverage limitations imposed by use of dry flat electrodes.

Equally important to the efficacy of EEG recording, the wearable solution must also be capable of neural signal processing and wireless communication. The power consumption of a high-throughput wireless transmitter capable of sending out raw recorded EEG signals from all channels in real time can easily drain the system battery in a short period. By integrating a signal processing module for seizure detection, the data needed to be wirelessly communicated will be limited to key information such as seizure frequency or duration. However, while the hardware implementation of seizure detection algorithms results in avoiding the high-throughput power-hungry wireless transmitters, the detection accuracy of these algorithms is heavily dependent on the available computational resources, which are very limited for an embedded processor compared to a mobile or desktop computer [12]. Another goal of this work is to approach the seizure detection algorithm development from a hardware and application perspective. We start with considering the known limitations in terms of power consumption and computational resources (e.g., logic elements count in an FPGA) while taking into account the upper limit for targeted detection latency, and try to optimize the seizure detection performance within those constraints.

In this work, first, we present the software implementation, optimization, and performance characterization of a data-driven algorithm for patient-specific seizure detection using only EEG signals from frontal and front-temporal lobes. This will allow the hardware implementation of the algorithm to

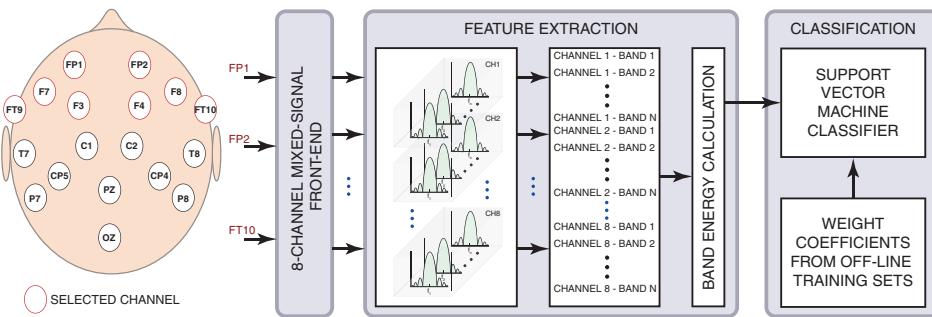


Fig. 2: Simplified block diagram of the signal path from signal acquisition at the surface EEG electrodes to producing seizure detection results at the output of the classification module. The frontal and front-temporal electrodes that are used are highlighted.

be a wearable solution that uses flat dry electrodes, yielding a minimal-form-factor light-weight headband-style wearable device. Next, we present an energy- and resource-optimized VLSI implementation of the data-driven seizure detection algorithm and validate its efficacy experimentally. Last, we present the design and implementation of a miniaturized $13 \times 17 \text{ mm}^2$ PCB that integrates (a) the FPGA that hosts the seizure detection algorithm, and (b) a Bluetooth BLE 5.0 wireless transceiver with embedded antenna for data/command communication. We also discuss how the presented PCB is interfaced with an 8-channel EEG recording front-end device we have previously reported for ambulatory EEG monitoring [10]. This paper extends on an earlier report of the principle and demonstration in [14], and offers a more detailed analysis of the algorithm design, its VLSI implementation, and most importantly the miniaturized hardware realization of the presented diagnostic device.

The rest of the paper is organized as follows. Section II presents the overall system architecture of the presented diagnostic device and the way it is interfaced with a multi-channel front-end device. Section III discusses design and optimization of the data-driven seizure detection algorithm. Section IV presents the resource-optimized energy-efficient VLSI implementation of the algorithm as well as the design and development of a mini-PCB enabling the signal processing and wireless communication features. Section V presents simulation, hardware synthesis, and experimentally-measured results illustrating the performance of different aspects of the presented work and compares them to the state of the art both at the system level and VLSI-implementation level.

II. SYSTEM ARCHITECTURE

Figure 1(top) shows the top-level block diagram of the envisioned wearable device. As shown, the device is comprised of two main modules, the multi-channel analog recording front-end and the digital backend. The analog front-end module is responsible for sensing the EEG signals, which includes amplification, signal conditioning (e.g., filtering, artifact removal, etc.), and multi-bit quantization before passing the digitized data to the backend module for further signal processing. We have previously reported an example 8-channel design of such analog front-end module where the recording channels were implemented as active electrodes on a fully-flexible polyimide substrate [10]. As shown in Figure 1(bottom) the flexible substrate is designed to be connected to a rechargeable battery

on one side and a backend board on the other side. In this work, our focus will be on the backend board as all the signal processing including the seizure detection will be done by the circuitry on this board.

As shown, the backend board receives the digitized EEG signals from the active electrodes, extract various features from the signals, and performs seizure/non-seizure classification. The low-power ADC used in each active electrode (TI-ADS1018IDGSR) can sample the signals with a maximum of 3.3 kSps. In this work, we set the sampling frequency at 512 Hz and used the ADCs single-shot mode that powers down the chip between two conversions, resulting in a significant power saving. The ADCs nominal resolution is listed as 12 bits, but our measurement results showed an ENOB of 9.9 to 10.2 bits. We considered the effective resolution of EEG signals to be approximately 10 bits for our seizure detection signal processing. The FPGA on the backend board generates a configuration command vector that is sent serially (using SPI protocol) to control the operation of the 8-channel front-end module. This vector includes, (i) a 2-bit signal for the ADC mode (single shot or continuous) selection and start/stop time, (ii) a 3-bit multiplexer selector for the ADC input selection, (iii) A 3-bit command for ADC input voltage range selection, (iv) A 3-bit command for setting the sampling rate, (v) a 4-bit command for ADC chip selection, and disabling additional feature, (vi) a 2-bit command going to the amplification section of the active electrodes to provide timing signals for the motion artifact removal module.

The serialized digital data from each active-electrode channel is sent to the FPGA on the backend for signal processing. Once the FPGA selects one of the ADCs by enabling its "chip select" signal, the ADC starts a new conversion and sends back a "data ready" flag upon completion. Next, the ADC receives data from that channel for 32 clock cycles. The first 16 clock cycles are used to receive data from that channel and the second 16 clock cycles are used to send the configuration commands of that ADC for the next conversion. At the end of 32 cycles, the ADC goes to a "power down" mode and FPGA selects the next channel.

In addition to managing the interface with the front-end module, the FPGA is also (and perhaps, mainly) used for digital signal processing and managing the wireless transmission of the recorded signals, which is the main focus of the presented work. This includes, various power-efficient low-pass and band-pass filtering, feature extraction, and machine-

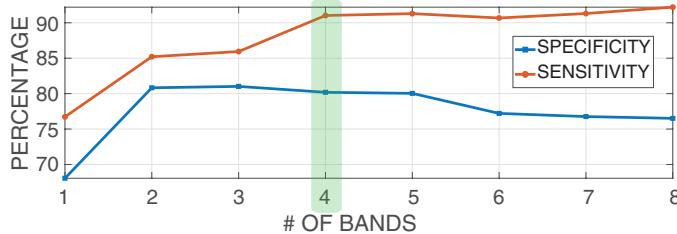


Fig. 3: MATLAB simulation results showing detection sensitivity and specificity vs the number of bands used for spectral energy calculation.

learning based classification. As for the wireless transmission, once an event of interest (e.g., a seizure onset) is detected, the key information related to that event (e.g., start and stop time, etc.) are wirelessly communicated to a hand-held or benchtop computer and/or gets stored on the on-board memory. The wireless link is also used to receive commands to the described control unit.

III. METHODOLOGY: SW IMPLEMENTATION

Figure 2 shows the end-to-end block diagram of the EEG signal path, and different functions applied on them on their way from the recording electrodes to back-end classifiers. As shown, following amplification, signal preconditioning, and quantization in the front-end, the digitized signals are fed to the backend module where the signal processing is conducted. Since the focus of this work is on the backend part, the digitized neural signals are assumed to be given as inputs to the presented system. For testing and verification of the backend module, a pre-recorded surface EEG dataset (CHB-MIT Scalp EEG Database [17], [18]) is used that includes 969 hours of scalp EEG recording with 198 seizures from 23 patients. There is a good level of diversity both in terms of patients (male and female, ages 10 to 22) and the seizure types (clonic, tonic, atonic). The dataset is labeled by an epilepsy expert and the start and stop time of the seizures are known. It must be mentioned that the seizures' origin on the brain varies among patients and there is no guarantee that it'll be close to the frontal electrodes that we intend to use, which will make the dataset unbiased to our method.

As illustrated in Figure 2, only a subset of electrodes that are located on the frontal part of the scalp covering frontal lobe and parts of temporal lobes are used. This will allow for the use of flat dry electrodes (instead of spring-loaded bulky through-hair electrodes) yielding a minimal form factor for the device. The signal processing is divided into two major steps of feature extraction and classification. There are several features reported in the literature that have been used (alone or combined) for seizure detection. This list includes signal's amplitude, phase, phase synchronization, spectral band energy, entropy, etc. [18], [21]–[23]. Based on the (a) overall performance of each algorithm (i.e., sensitivity, detection accuracy, and specificity) in detecting seizures using only frontal-lobe EEG, and (b) the computational resources required for their implementation, we decided to use the spectral band energy as the key feature for classification.

As shown in Figure 2, following digitization, the

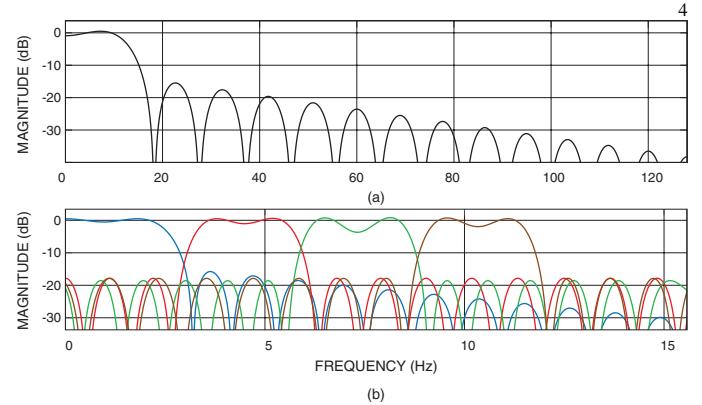


Fig. 4: Magnitude response of the implemented (a) anti-aliasing LPF, and (b) four band-selecting BPFs.

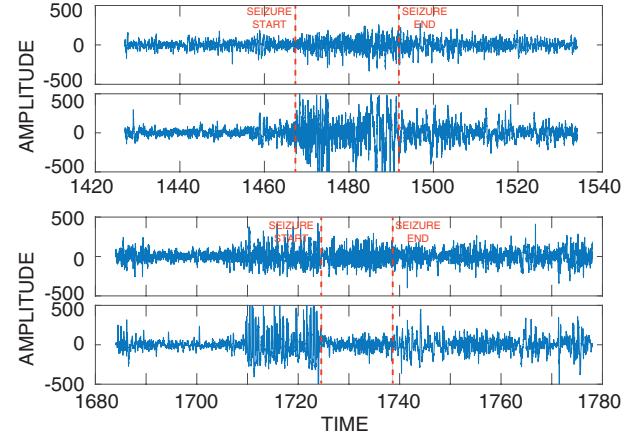


Fig. 5: The significant variation in time-domain manifestation of seizures for different recording sites and different patients.

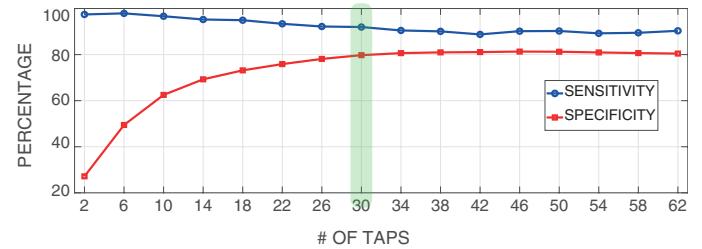


Fig. 6: MATLAB simulation results showing detection sensitivity and specificity vs the number of taps used for the FIR BPFs.

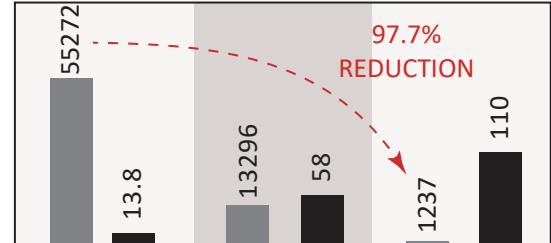


Fig. 7: Power and LE consumption of the presented seizure detection algorithm using three different FIR filter implementation options.

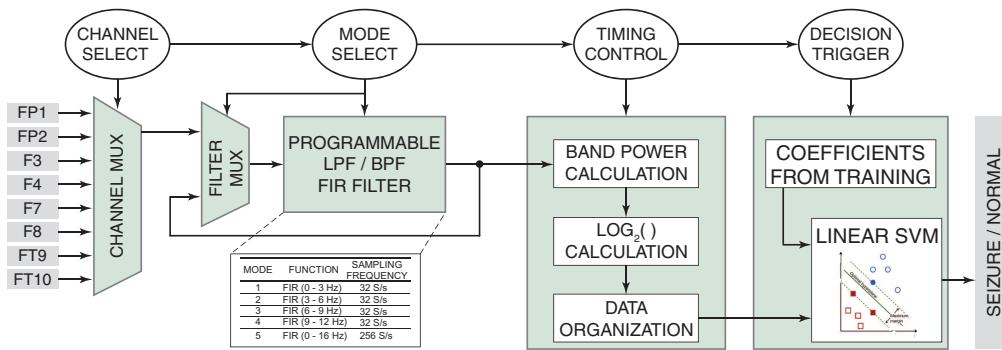


Fig. 8: Top-level block diagram of the VLSI implementation of the SVM-based seizure detection algorithm on the FPGA, illustrating the resource sharing and timing control employed.

recorded signal from each electrode is passed through multiple bandpass filters to extract the spectral information. The output of the filters are used to calculate the energy of the signal in each band. For the case where EEG is recorded from N electrodes and each signal is passed through M bands, this process results in a matrix of $M \times N$ elements (features), each representing the energy of an electrode signal in a particular frequency band.

It is commonly-accepted that the most important range of frequency for seizure detection using surface EEG is from DC to 30 Hz [18], [29]. It is also shown that when this band is divided to multiple sub-bands, the changes in signal energy in each sub-band could be used as an indicator to differentiate seizure and non-seizure states [18]. To maximize the efficacy of this method, we looked into different possibilities for the optimal number of sub-bands (i.e., 30Hz divided by the width of each band). Such optimization requires definition of quantitative performance measures such as sensitivity (true positive rate) and specificity (true negative rate). Here, we have used the following definitions for these measures [19]:

- True Positive (TP): The number of EEG epochs that contain an ictal event that are detected as a seizure by the algorithm.
- False Positive (FP): The number of EEG epochs during interictal period but are identified as a seizure event by the algorithm.
- True Negative (TN): The number of EEG epochs during interictal period that are identified as a non-seizure event by the algorithm.
- False Negative (FN): The number of EEG epochs that contain an ictal event but are identified as a non-seizure event by the algorithm.
- Sensitivity: The ratio of TP to TP+FN.
- Specificity: The ratio of TN to TN+FP.

The seizure detection algorithm was trained and tested for different number/width of frequency sub-bands and the performance in terms of the measures defined above was evaluated. Prior to deciding on the count and width of sub-bands, the spectral energy of different frequencies within the 0-30 Hz EEG bands (Delta, Theta, Alpha, and Beta) during ictal and inter-ictal periods were compared, and their impact on the performance metrics were evaluated. Our simulation results

indicated that while the band energy for 0-12Hz changes significantly from inter-ictal to ictal periods, no meaningful changes happen for upper-Alpha (12-16Hz) and Beta (16-31Hz) bands. Therefore, we decided to limit the signal processing to 0-12 Hz frequencies and focus the sub-band optimization to this range. Figure 3 shows the sensitivity and specificity of the algorithm versus the number of sub-bands used. As shown, seizure detection sensitivity seems to be monotonically improving by the number of bands used. However, while too few bands will result in a poor performance both in terms of sensitivity and specificity, increasing the number of bands (i.e., making the width of each band narrower) will degrade the specificity of detection. Based on these results, we decided to go with four sub-bands.

The anti-aliasing filter is an equi-ripple 30-tap direct-form low-pass FIR filter with sampling frequency of 256Hz and bandwidth of 16Hz. The band-selecting filters are all constrained equi-ripple 30-tap direct-form band-pass FIR filters with sampling frequency of 32Hz and various bandwidths of DC-3Hz, 3-6Hz, 6-9Hz, and 9-12Hz. The number of taps are optimized for both performance and required logic elements for their hardware implementation, as described in section IV. Fig. 4(a) and (b) show the magnitude response of the anti-aliasing and the band-selecting filters, respectively.

Following feature extraction, classification is performed. Conventionally, this is done using deterministic classifiers with thresholds values that are needed to be set manually. Figure 5 demonstrate seizure events from two different patients and from different recording channels. As shown, the way seizure is manifested varies from patient to patient, which calls for a data-driven algorithm that is adaptive to each patient. To address this issue, machine learning algorithms have been utilized in a few of recently-reported works [26], [27]. Similar to feature extraction, the choice of machine learning classification algorithms heavily depends on (a) overall performance in terms of seizure detection, and (b) resource (i.e., computational elements and power consumption) requirement. It has been shown that while computationally-expensive deep learning algorithms (e.g., recurrent convolutional neural networks) can achieve unparalleled performance in cross-patient seizure detection (i.e., an algorithm that is indifferent to patient-to-patient variations), for patient-specific detection, their performance is fairly comparable to significantly simpler algorithms such as support vector machines (SVMs) [28]. In addition, authors

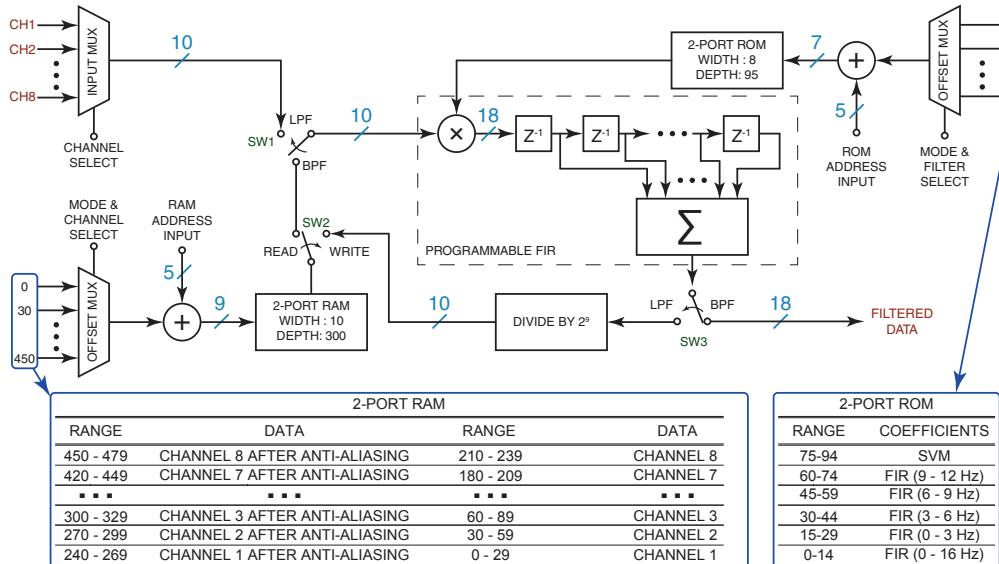


Fig. 9: Logic diagram for a reconfigurable multichannel BP FIR and anti-aliasing LP filter.

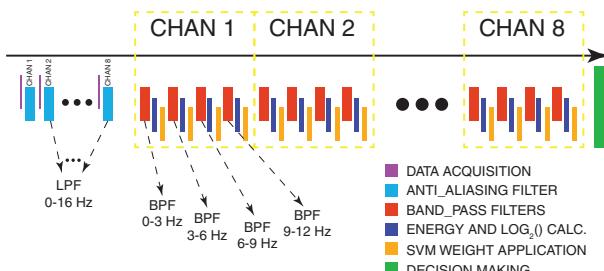


Fig. 10: Timing sequence of different tasks performed in the backend FIR filter and SVM classifier.

in [31] have compared tens of seizure detection algorithms and have concluded that if the feature extraction is done properly (e.g., correct feature selection), then the classifier computational complexity plays a minor role in overall seizure detection performance.

In this work, we used an SVM classifier for patient-specific seizure detection. The classifier receives the aforementioned feature matrix as the input and finds a multi-dimensional maximum-margin hyperplane that separates seizures and non-seizures data. 60% of the dataset is used for training, 20% is used for cross validation, and 20% is used for testing. The SVM uses a linear kernel and receives a 32-element (8 channels \times 4 sub-bands) feature matrix as its input.

The SVM training is done offline while the classification is performed online (on the FPGA). For the training, the algorithm is first normalized. Since the scale of various features used are different, this will ensure that the differences in one feature is not overshadowed by another one, simply due to the former's smaller scale. Next the search for optimal hyper-parameters is conducted. A logarithmic approach is taken to ensure that the entire range of values for regularization parameters are covered in a reasonable time and without sacrificing accuracy. The SVM was trained using different combinations of such hyper-parameters and the best model

was used. A weight term was also introduced and optimized to compensate for the highly-imbalanced nature of the dataset (seizure episodes being less than 1% of the dataset). It helps to make the model more sensitive to seizures as a false negative (i.e., missing a seizure) is significantly worse than a false positive (false alarm) for this application.

IV. HARDWARE AND VLSI IMPLEMENTATION

VLSI implementation of the described seizure detection algorithm, in the form presented in Figure 2, will require 55,272 logic elements (LEs). This number is far beyond the number of LEs available in a low-power FPGA that is suitable for battery-powered medical devices. In our case, we used a Microsemi AGL250 which has a total of 6144 LEs. This motivated investigating the major resource consumers and proposing ideas for a compact VLSI implementation.

For band-pass filtering, finite impulse response (FIR) filters are preferred over infinite impulse response (IIR) ones as they provide a linear phase response and their transfer function is more robust to process, voltage, and temperature variations. However, compared to IIR, FIR implementation requires significantly higher amount of resources on chip. In fact, our synthesis report shows that the majority of LEs used for the algorithm implementation are dedicated to the 32 (8 channels \times 4 bands) FIR filters used for band selection. The resource consumption of this block could be reduced by (a) lowering the number of taps in cost of losing filter's spectral selectivity, and (b) computational resource sharing among filters in cost of increasing the signal processing latency.

To find the optimal number of taps, we looked into seizure detection sensitivity and specificity as the ultimate performance measures of the signal processing module. The algorithm was trained and tested for various number of taps ranging from 2 to 64 and the seizure detection sensitivity and specificity was measured for each value. Based on Figure 6, 30 taps seems to be the optimal point for overall performance of the algorithm. In addition to optimizing the resource consumption of a single

filter, the number of filters should be reduced as well. Given the programmability of FIR filters, all 4 BPFs and the anti-aliasing LPF of each channel could be realized using one physical implementation. Since the signals in all 8 channels must be filtered to the same 4 frequency bands, all the 40 filters (32 BPF + 8 LPFs) could be realized by time-multiplexing a single implementation and five sets of filter coefficients stored in the on-board memory. Even within one filter, the multi-bit multiplier, which is the most resource-consuming block, could be shared between taps to reduce the number of required logic gates. The resource sharing comes at the cost of either higher dynamic power consumption due to increasing the clock frequency, or increasing the computation delay.

In this work, our approach is to (a) reduce the number of required logic elements to a number lower than the maximum available on the FPGA, (b) find an optimal solution for the trade-off between the power consumption and the latency of the algorithm by decreasing the delay as much as the power consumption budget allows. In this work, we decided to keep the dynamic power consumption of the signal processing module around $100\mu\text{W}$ so that it remains significantly lower than the power consumption of the front-end recording module and the wireless transmitter (i.e., more than 1mW), hence, its effect on battery lifetime stays negligible. We also targeted a maximum classification latency of half a second, based on what is commonly-accepted in the literature [27]. Starting with these numbers as limitations, we tried to use time-division multiplexing and other techniques described below to minimize the number of logic elements. Figure 7 shows three different FIR filter implementation options ((1) dedicated filter for each band and each channel, (2) shared filter for all bands and all channels, (3) shared multiplier for all taps of a filter shared between all bands and all channels.) and compares the seizure detection algorithm with these options in terms of resource and power consumption. The power numbers does not scale up as fast as the resource consumption is reduced as we have allowed the classification latency to be higher than its absolute minimum. While the FPGA used allows us to operate at a maximum frequency of 250MHz yielding a delay of only $10.2\mu\text{s}$, we chose to set our nominal frequency to a significantly lower value (5.2kHz) and keep the power consumption at $110\mu\text{W}$ by allowing the delay to increase to 492ms , which is still below the permitted limit.

Figure 8 shows the top-level block diagram of the system with the programmable FIR filter, time-multiplexed for different transfer functions. Implementation of programmable FIR BPF and anti-aliasing filters requires the storage of data and filtering coefficients. In this work we used the non-volatile FlashROM memory of the FPGA to store the coefficient. It results in a significant LE saving at the cost of a slight increase in the read/write latency. As shown in Figure 9, to retrieve different sets of channel-specific FIR coefficients with only a constant range of FlashROM input address (0 - 15), a channel-specific address offset can be selectively added to each input address by an offset mux in order to facilitate coefficient retrieval in a shared FIR structure. The same idea could be also applied to FIR input data storage and retrieval inside RAM. Since the FIR needs to process signals from eight channels, 32

input data from each channel needs to be stored and accessed efficiently in one RAM block on the FPGA. To store or retrieve different sets of channel-specific FIR input data with only a constant range of RAM input address (0 - 30), a channel specific RAM offset address is selectively added to each input address. As a result, with minimal additional LE cost, this address offset methods easily converts a single channel FIR to a configurable multichannel FIR filter using multiplexers.

First, SW1 is switched to the LPF mode and the EEG data samples from the first channel are read and fed to the programmable FIR structure shown in the dashed box. The FIR is programmed to operate as an anti-aliasing filter and the proper coefficients are loaded from ROM prior to reading the input data. In this phase, SW3 is in the LPF mode and the SW2 is in the write mode so that the output of the FIR filter is sent back and stored in the RAM in the space dedicated for channel 1. This procedure is repeated 8 times while SW1, 2, and 3 maintain their position and the only thing changing is the address in RAM where the alias-free data is written.

Once the filtered data from all 8 channels are written in RAM, SW1 and SW2 switch to the BPF and READ modes, respectively, and SW3 is switched to BPF mode. The controller provides the proper RAM and ROM address offset to read the right channel and filter data, respectively. For each set of data read from RAM, 4 different set of coefficients are loaded onto the filter from ROM and the band-pass filtered data for 4 different bands are generated at the output. This is repeated for another 8 times until all 32 (8 channels \times 4 bands) filtered data samples are generated. For each of these 32 operations, while the filtering is being done for the i^{th} sample, the energy of the filtered $i - 1^{\text{th}}$ data sample is calculated and the proper SVM weights are applied to it and sent to the classifier. Once all 32 features are ready, the classifier makes a seizure/non-seizure decision. Fig. 10 summarizes the above timing sequence description.

Figure 11 depicts a simplified block diagram of the digital/wireless backend board and the way it is interfaced with the analog front-end and the wireless receiving stationary module. As shown, the digitized EEG signals are first fed to the signal processing module for seizure detection. The result of the seizure detection block is used to further optimize the data acquisitions (e.g., channel omission, sampling rate variation, etc.) as well as serialized and fed to the on-board wireless transceiver microcontroller unit using a serial peripheral interface (SPI) protocol. The microcontroller receives and organizes the raw data from the FPGA and prepares it for transmission. In this work, we have used the Rigado BMD-350 module that integrates both the transmitter Bluetooth chip and its planar antenna. Table I compares this transceiver module with the most popular wireless module for medical applications used in the literature. On the receiver side, a development board from Rigado is used to receive and decoded the data and send it to the USB port of a bench-top computer using a universal asynchronous receiver-transmitter interface.

Figure 12 shows the top and bottom view of a $13\text{mm} \times 17\text{mm}$ PCB implementation of the presented backend module. The board is populated with the low-power Microsemi FPGA as the main processor to conduct seizure detection as well as

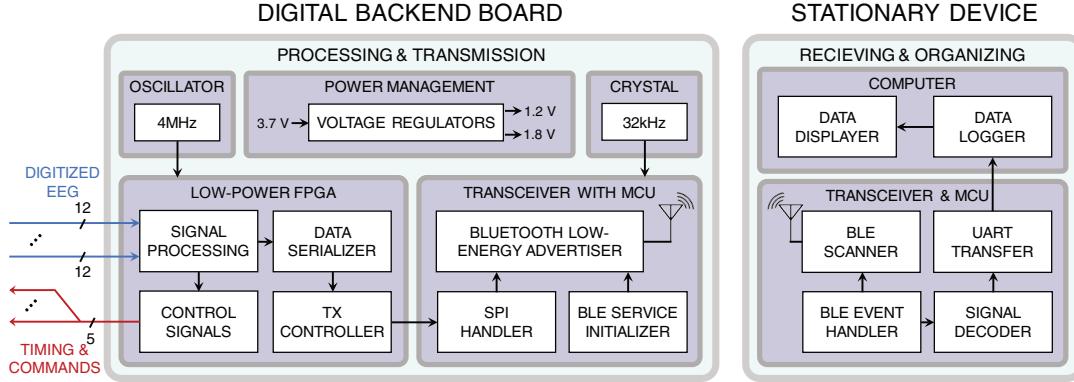


Fig. 11: Block diagram of the backend mini-PCB with focus on the signal path through the wireless communication channel.

TABLE I:
Comparison of the off-the-shelf available Bluetooth microchips/modules for medical sensory applications.

| Transceiver | NINA-B112 | BL652 | ZL70103 | CC2640R2F | BMD-350 |
|--------------------------|-----------|----------|----------|-----------|-------------|
| Energy Efficiency (nJ/b) | [32] | [33] | [34] | [35] | [36] |
| I_{DC} at 0dBm (mA) | 9.19 | 9.19 | 15.19 | 11.23 | 12.25 |
| VDD (V) | 5.3 | 5.3 | 5.3 | 6.1 | 7.1 |
| Duplex | 1.7-3.6 | 1.7-3.6 | 2.8-3.5 | 1.8-3.6 | 1.7-3.6 |
| Physical Size (mm^3) | 14×10×4 | 14×10×2 | 6×5×2 | 7×7×2 | 8.7×6.4×1.5 |
| Antenna | Internal | Internal | External | External | Internal |
| Max DR (Mbps) | 1 | 1 | 0.8 | 1 | 2 |

N/R: Not reported

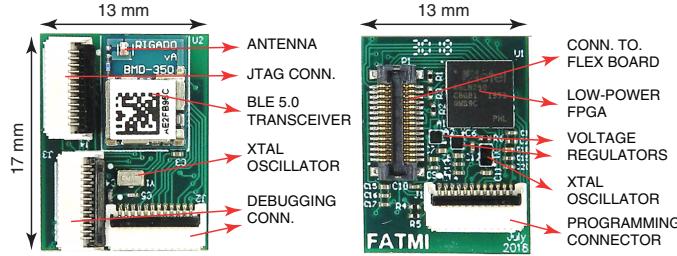


Fig. 12: Top and bottom views of the presented backend mini-PCB with major components annotated.

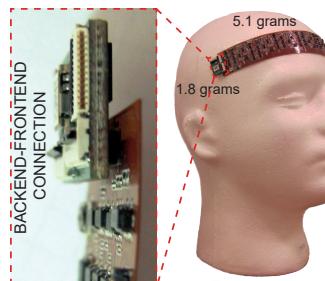


Fig. 13: Mechanical/electrical connection between the flexible front-end recording board [10] and the backend digital board.

controlling both the backend and the front-end modules, a Bluetooth low energy (BLE) 5.0 transceiver module with built-in antenna for wireless communication (Rigado BMD350 built based on the nRF52832 SoC from Nordic Semiconductor). Figure 12 also shows other components on the backend PCB including voltage regulators, connectors, crystal oscillators and

TABLE II: List of the off-the-shelf components used in the backend digital mini-PCB

| Component | Company | Model |
|--------------------|-------------------|--------------------|
| Regulator-3.3 | Analog device | ADP121-ACBZ33R7 |
| Regulator-1.2 | Analog device | ADP121-ACBZ12R7 |
| Adj Regulator | Linear Technology | LT3020EMSxPBF |
| BLE Tx | Rigado Inc | BMD-350-A-R |
| FPGA | Microsemi | AGLN250V2-CSG81 |
| XTAL 32kHz | ECS Inc | ECS-.327-6-12-C-TR |
| Oscillator 4MHz | SiTime | 8021AI-J4-XXS-4.0E |
| Vertical Connector | Panasonic | AXT430124 |
| Vertical Connector | Panasonic | AXT430324 |
| FFC Connector | Molex | 5034801200 |

peripheral passive components to ensure that the board can operate as a self-contained module and only require a 3.7V supply voltage. Figure 13 shows how the presented 1.8-gram PCB is connected to a previously-reported flexible front-end board using a pair of Panasonic connectors that provide both electrical connection and mechanical support [10]. The full list of components used in this implementation is provided in Table II.

V. MEASUREMENT RESULTS

The described resource-optimized VLSI implementation of the seizure detection algorithm was uploaded onto the on-board low-power Microsemi FPGA (AGL250). CHB-MIT dataset signals from the 8 frontal electrodes indicated in Figure 2 were fed to the FPGA in the form of 10-bit digital signals for each channel to conduct real-time seizure detection. Our synthesis

TABLE III: Percentage breakdown of LE Usage for algorithm's VLSI implementation on the Microsemi FPGA

| Module | Total LE Consumption | % of Available Resources |
|--------------------|----------------------|--------------------------|
| Signal Acquisition | 94 | 1.53% |
| FIR Filter | 590 | 9.6% |
| Energy Calculation | 225 | 3.67% |
| Classification & | 259 | 4.21% |
| Control Logic | 69 | 1.12% |
| Total | 1237 | 20.13% |

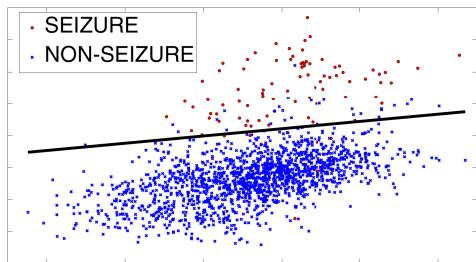


Fig. 14: SVM sample classification results for 1-hour data from a single patient.

and simulation results show that the algorithm requires 2563 clock cycles to generate an output, which translates into a minimum latency of $10.2 \mu\text{s}$ for a maximum clk frequency of 250MHz, or a 492ms delay for the nominal 5.2kHz frequency.

Table III lists all the major modules of the code implemented on the FPGA, the total number of logic elements they require, and the percentage of the available resources on the FPGA they use. Figure 16(a) and (b) show the resource and power breakdown of the major modules of the VLSI implementation on the FPGA. Thanks to the resource optimization techniques described above, the entire signal processing (feature extraction + classification) only takes slightly more than 20% of resources

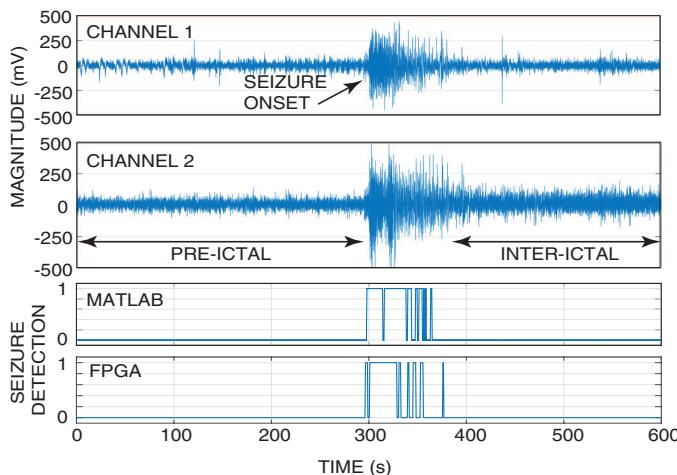


Fig. 15: Sample EEG recording from two channels and the seizure detection results of the MATLAB and FPGA implementation.

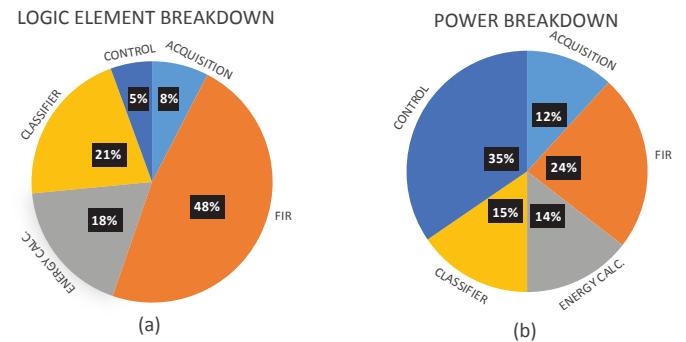


Fig. 16: LE and power breakdown of the major modules of the VLSI implementation of the seizure detection algorithm on the FPGA.

TABLE IV: Comparison with the FPGA-based implementations of surface-EEG-based seizure detection

| Ref. | [37] ISCAS'18 | [38] TCAS-II'15 | This work |
|---------------------------------|------------------|--------------------|----------------------|
| Seizure Recording | Focal standard | Focal Standard | General Frontal Only |
| # of Channels | 6 | 6 | 8 |
| Algorithm | SVM | MCMC | SVM |
| # of LEs | 3629 | 1198 | 1237 |
| Memory (kB) | 892 | 8.3 | 5.56 |
| Max Frequency (MHz) | 150 | 200 | 250 |
| Latency (cycle) | 91928 | 2192 | 2563 |
| Latency (μs) | 600 | 11 | 10.2 |
| Nominal Frequency (kHz) | 184 | 4.37 | 5.2 |
| Dynamic Power (μW) | 131.83 | 3.37 | 110 |
| Sensitivity (%) | 95.39 | 92.88 | 92.51 |
| Specificity (%) | N/R | N/R | 80.1 |

N/R: Not reported

on the FPGA, which suggests that the implementation could be further optimized by increasing this number to near 100% capacity (i.e., a hybrid serial/parallel filter implementation), which allows for reducing the clock frequency, and therefore, the dynamic power consumption. We did not pursue further optimization as the FPGA remaining resources were required for other signal processing and control purposes.

Figure 14 shows a sample output of the SVM classifier for the data from one of the 23 patients. Figure 15 shows the sample recordings of the EEG signals during a seizure episode from two channels, as well as the seizure detection results for both MATLAB and FPGA implementation. The algorithm was trained using 60% of the available data from all 23 patients of CHB-MIT dataset, 20% was used for cross validation, and 20% was used for testing. Our experimental results shows an average seizure detection sensitivity (window) and specificity of 92.5% and 80.1%, respectively, based on the definitions described in Section III and in [19]. It should be noted that these numbers could vary (slightly, as the EEG spatial resolution is approximately 1-10 cm [20]) as the electrode

TABLE V: Comparison with the state-of-the-art hardware implementations of seizure detection algorithms.

| | JSSC'13 [39] | JSSC'13 [40] | JSSC'14 [26] | TBCAS'11 [41] | JSSC'16 [24] | TBCAS'16 [27] | THIS WORK |
|---------------------------|-----------------|-----------------|-----------------|------------------|-----------------|------------------|-----------|
| # of channels | 8 | 1 | 8 | 1 | 64 | 8 | 8 |
| Readout Circuit | Yes | No | Yes | Yes | Yes | Yes | Yes |
| Wireless Data | No | Yes | Yes | No | Yes | No | Yes |
| Transmission DR (Mbps) | N/A | N/R | 4 | N/A | 1 | N/A | 2 |
| Embedded Classification | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient Specificity | Yes | Yes | Yes | Yes | No | Yes | Yes |
| Detection Rate (%) | 84.4 | 100 | 92 | 100 | 92 | 91 | 92.5 |
| False Positive (per hour) | 5.1 | 1.2 | - | - | - | 0.27 | - |
| Specificity (%) | - | - | - | N/R | 93 | - | 80.1 |

N/R: Not reported

placement of the presented device is different from the selected channels of the standard 10-20 system.

Table IV compares the presented implementation with state of the art FPGA-based seizure detection algorithm implementations in terms of channel count, latency, power consumption, and seizure detection performance, among other parameters. Table V compares the entire presented wireless EEG monitoring and processing solution to the state of the art in terms of system-level features such as wireless connectivity, having integrated readout circuit, etc.

VI. CONCLUSION

A wireless 1.8-gram self-contained mini-PCB hosting a resource-optimized VLSI architecture for patient-specific seizure detection using frontal-lobe EEG was presented. The seizure detection algorithm was developed and optimized to detect seizures using only frontal-lobe EEG signals. The VLSI implementation of the algorithm was developed to meet power, resource (logic element count) and detection latency requirements at the same time. Despite the use of a subset of standard 10-20 electrodes, the seizure detection results are competitive to the state of the art, promising the possibility of conducting accurate seizure detection using quickly-mountable dry-electrode ambulatory EEG headsets that can only provide reliable recording from hairless parts of the scalp. The FPGA implementation of the algorithm yields a dynamic power consumption of $110\mu\text{W}$ and a nominal and minimum detection latency of 492ms and $10.2\mu\text{s}$, respectively. The measurement results from the FPGA implementation shows a seizure detection sensitivity and specificity of 92.5% and 80.1%, respectively.

REFERENCES

- [1] R. Fisher, W. Van Emde Boas, W. Blume, Ch. Elger, P. Genton, Ph. Lee, and J. Engel Jr. "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470-472, 2005.
- [2] M. T. Salam, et al., "Rapid Brief Feedback Intracerebral Stimulation Based on Real-time Desynchronization Detection Preceding Seizures Stops the Generation of Convulsive Paroxysms," Vol. 56, No. 8, pp. 1227-1238, *Epilepsia*, 2015.
- [3] H. Kassiri, S. Tonekaboni, M. T. Salam, N. Soltani, K. Abdellahim, J. L. Perez Velazquez, R. Genov, "Closed-Loop Neurostimulators: A Survey and a Seizure-Predicting Design Example for Intractable Epilepsy Treatment," *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 11, No. 5, pp. 1026-1040, Oct. 2017.
- [4] B. Schelter, J. Timmer, and A. Schulze-Bonhage, Eds., "Seizure Prediction in Epilepsy," Weinheim: Wiley, 2008.
- [5] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with support vector machines," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464-473, Mar. 2011.
- [6] [Online]. Available: <https://lfpres.com/news/local-news/patients-wait-as-long-as-a-year-for-epilepsy-test>
- [7] G. K. Bergey et al., "Long-term treatment with responsive brain stimulation in adults with refractory partial seizures," *Neurology*, vol. 84, no. 8, pp. 810-817, Feb. 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4339127/>
- [8] J. Xu, et.al, "A Wearable 8-Channel Active-Electrode EEG/ETI Acquisition System for Body Area Networks," *IEEE J. Solid-State Circuits*, vol.49, no.9, pp.2005-2016, Sept. 2014.
- [9] [online] Available: <http://cognionics.com/index.php/54-products/headsets>
- [10] A. Dabbaghian, et al., "A 9.2-gram Fully-Flexible Wireless Dry-Electrode Headband for Non-Contact Artifact-Resilient EEG Monitoring and Programmable Diagnostics," In *Circuits and Systems (ISCAS), 2019 IEEE International Symposium on*, pp. 1-4. IEEE, 2019.
- [11] [Online]. Available: <http://www.gtec.at/Products/Electrodes-and-Sensors/g.SAHARA-Specs-Features>
- [12] M. T. Salam, H. Kassiri, N. Soltani, H. He, J. L. Velazquez, and R. Genov, "Tradeoffs between wireless communication and computation in closedloop implantable devices," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2016, pp. 1838-1841
- [13] [Online]. Available: <https://www.cognionics.net/flex-sensors>
- [14] T. Zhan, et al., "A Resource-Optimized VLSI Architecture for Patient-Specific Seizure Detection Using Frontal-Lobe EEG," In *Circuits and Systems (ISCAS), 2019 IEEE International Symposium on*, pp. 1-4. IEEE, 2019.
- [15] [Online]. Available: <https://www.emotiv.com/insight/>
- [16] [Online]. Available: <https://choosemuse.com/muse-2/>
- [17] A. Goldberger et al., "PhysioBank, physiotoolkit, and physionet: Components of a new research resource for complex

- physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-220, 2000.
- [18] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2009.
- [19] H. Kassiri et al., "Electronic Sleep Stage Classifiers: A Survey and VLSI Design Methodology," *IEEE Transactions on Biomedical Circuits and Systems*, 2016.
- [20] S. Ha et al., "Silicon-Integrated High-Density Electrocortical Interfaces," *Proc. IEEE*, vol. 105, no. 1, pp. 11-33, Jan. 2017.
- [21] N. Moghim, et al., "Predicting epileptic seizures in advance," *PLoS one* 9.6 (2014): e99334.
- [22] Y. Liu, et al., "Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG," *IEEE transactions on neural systems and rehabilitation engineering* 20, no. 6 (2012): 749-755.
- [23] K. Samiee, et al., "Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction." *Knowledge-Based Systems* 118 (2017): 228-240.
- [24] H. Kassiri, A. Bagheri, N. Soltani, K. Abdelhalim, H. Jafari, M. T. Salam, J. L. Perez Velazquez and R. Genov, "Battery-Less Tri-Band-Radio Neuro-Monitor and Responsive Neuro-Stimulator for Diagnostics and Treatment of Neurological Disorders," *IEEE Journal of Solid-State Circuits*, Vol. 51, No. 5, pp. 1274-1289, May 2016.
- [25] H. Kassiri, et al., "All-Wireless 64-channel 0.013mm²/ch Closed-Loop Neurostimulator with Rail-to-Rail DC Offset Removal," *IEEE International Solid-State Circuits Conference*, Feb. 2017, pp. 452-454.
- [26] W.-M. Chen, et al., "A fully integrated 8-channel closed-loop neural prosthetic SoC for real-time epileptic seizure control," *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2013, pp. 286-287.
- [27] M. Altaf and J. Yoo, "A 1.83 μ J/classification, 8-channel patient-specific epileptic seizure classification SoC using non-linear support vector machine," *IEEE Trans. Biomed. Circ. Syst.*, vol. 10, no. 1, pp. 49-60, Feb. 2016.
- [28] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthcare (MLHC)*, Los Angeles, CA, USA, Jul. 2016, pp. 178-190.
- [29] H. Kassiri, M. T. Salam, M. R. Pazhouhandeh, N. Soltani, J. L. Perez Velazquez, P. L. Carlen, R. Genov, "Rail-to-Rail-Input Dual-Radio 64-channel Closed-Loop Neurostimulator," *IEEE Journal of Solid-State Circuits*, Vol. 52, No. 11, pp. 2793-2810, Oct. 2017.
- [30] H. Kassiri, et al., "Battery-less Modular Responsive Neurostimulator for Prediction and Abortion of Epileptic Seizures," *IEEE International Symposium on Circuits and Systems (ISCAS)* 2016.
- [31] K. Gadhoumi et al., "Seizure prediction for therapeutic devices: A review," *J. Neurosci. Methods*, vol. 260, pp. 270-282, 2016.
- [32] [Online]. Available: <https://www.u-blox.com/en/product/nina-b1-series>
- [33] [Online]. Available: <https://www.lairdconnect.com/wireless-modules/bluetooth-modules/bluetooth-5-modules/bl652-series-bluetooth-v5-nfc>
- [34] [Online]. Available: <https://www.microsemi.com/product-directory/implantable-medical-transceivers/3915-zl70103>
- [35] [Online]. Available: <http://www.ti.com/product/CC2640R2F>
- [36] [Online]. Available: <https://www.rigado.com/products/modules/bmd-350/>
- [37] 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.
- [38] L. Marni, et al., "A Real-Time Wearable FPGA-based Seizure Detection Processor Using MCMC," In *Circuits and Systems (ISCAS), 2018 IEEE International Symposium on*, pp. 1-4. IEEE, 2018.
- [39] J. Yoo, L. Yan, D. El-Damak, M. A. Altaf, A. Shoeb, and A. P. Chandrakasan, "An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 214-228, Jan. 2013.
- [40] 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 J. Solid-State Circuits*, vol. 48, no. 7, pp. 1625-1636, Jul. 2013.
- [41] M. Salam, M. Sawan, and D. Nguyen, "A novel low-power-implantable epileptic seizure-onset detector," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 6, pp. 568-578, Dec. 2011.



Tianyu Zhan received her B.Eng. degree in Electronic Science and Technology at Beijing Institute of Technology in 2019. She is currently an MS student in Electrical and Computer Engineering at Georgia Institute of Technology. As a MITACS Globalink Intern, she worked in the Integrated Circuits and Systems lab in summer 2018 at York University and was the lead person in charge of the development of the resource-optimized VLSI implementation for seizure detection algorithm.



Syyeda Zainab Fatmi received a B.Eng. degree in Computer Engineering from York University in 2019. She is currently an M.A.Sc student in the Department of Electrical Engineering and Computer Science at York University, studying verification techniques for probabilistic systems. As an undergraduate research assistant, she worked in the Integrated Circuits and Systems lab during the summer of 2018 and was the lead person in charge of the development of a miniaturized custom-designed module for digital bio-signal processing and wireless communication. She is the recipient of the York University Presidents scholarship, HSBC scholarship, CGS-M scholarship, and the NSERC USRA award in 2018 and 2019. She is also the co-winner of the Students' Choice award for the poster presentation in the undergraduate research conference at York University in 2018.



Sam Guraya is currently in his 4th year of Computer Engineering in the Department of Electrical Engineering and Computer Science at York University. In the summer 2018, he worked as a research assistant in the Integrated Circuits and Systems Lab where he helped develop models for seizure detection. He is the recipient of the York University LURA award in summer 2017 and NSERC USRA in summer 2018. He is also a co-winner of the oral presentation awards in the undergraduate research conference at York University in 2018.



Hossein Kassiri (S'10-M'16) received the B.Sc. degree from University of Tehran, Iran, the M.A.Sc. degree from McMaster University, and the PhD degree from the University of Toronto, all in electrical and computer engineering. He is currently an Assistant Professor in the Department of Electrical Engineering and Computer Science at York University, where he is the director of the Integrated Circuits and Systems Lab as well as the Center for Microelectronics Prototyping and Test (MPT). His research interests lies primarily in the area of design

and development of wireless and battery-less multi-modal neural interfacing systems and their application in monitoring and treatment of neurological disorders.

Since January 2019, he has been an Associate Editor of the IEEE Transaction on Biomedical Circuits and Systems (TBioCAS). He also holds the position of CTO at BrainCom Inc., a company he co-founded in September 2015, specialized in implantable brain-computer interfaces. Dr. Kassiri is the recipient of IEEE ISSCC 2017 Jack Kilby Award for Outstanding Student Paper, IEEE ISCAS best paper award (biocas track) (2016), Ontario Brain Institute (OBI) entrepreneurship award in 2015, Heffernan commercialization award in 2014, and the CMC Brian L. Barge award for excellence in microsystems integration in 2012.