FPGA Implementation of RLSE Algorithm for Multichannel Brain Imaging

Muhammad Shahid Nazir\textsuperscript{1a}, Haroon-Ur-Rasheed Khan\textsuperscript{1b}, Abubaker Akram\textsuperscript{1c}, Bhagesh Maheshwari\textsuperscript{1d}, Muhammad Aqil\textsuperscript{1e}

RECEIVED ON 21.02.2019, ACCEPTED ON 26.07.2019

ABSTRACT

This paper describes the implementation of a computationally efficient embedded system on an Field Programmable Gate Array (FPGA) platform for real-time brain activity estimation with multiple channels. The brain signals from multiple channels are considered as output of independent linear systems with unknown parameters representing the brain activity in corresponding channels. Multiple adaptive Recursive Least-Squares Estimation (RLSE) cores are implemented in FPGA to independently estimate the brain activity in each channel concurrently. The proposed RLSE-FPGA system provides dedicated (no time or resource sharing) and parallel processing environment. The universal asynchronous receiver transmitter core is also developed to communicate the measured and estimated parameters supported by storage facility programmed as shared memory. The computational precision is guaranteed by deploying a 32-bit floating point core for all the variables. The validation carried out by real Functional Near-Infrared Spectroscopy dataset and comparative analysis with the previously reported result, demonstrates the effectiveness of the proposed system. The computational cost endorses the effectiveness of concurrent processing of multiple channels’ data in a sample before the arrival of the next sample. The proposed methodology has potential in real-time medical, military and industrial applications.

Keywords: FPGA, Recursive Least Squares Estimation, Brain Imaging,

1. INTRODUCTION

Functional Near Infrared Spectroscopy (fNIRS) monitors the brain activity by measuring variation in the absorption of near infrared light in brain tissues. It is non-invasive technology that uses the variation in infrared light of wavelength 650–950 nm to monitor the brain activity [1]. The variation in absorption of infrared light is directly related with the concentration of Oxygenated Hemoglobin (HbO) and Deoxygenated Hemoglobin (HbR) in brain tissues. As subject is involved in performing cognitive tasks, the concentration of oxygenated and deoxygenated hemoglobin varies in brain tissues, these variations are then measured by fNIRS techniques. In recent years, many researchers have used fNIRS technique to investigate for biomedical and Brain-Computer Interface (BCI) applications [2-5]. Non-invasive feature of fNIRS, makes it attractive to utilize it for brain imaging (BI) and BCI. In reference to BCI techniques, fNIRS offers several benefits over other non-invasive technologies, like functional magnetic resonance imaging (fMRI) in terms of better temporal resolution, cost and computational efficiency. The fNIRS offers good spatial resolution and noise immunity against electromagnetic interference, when it is compared to Electroencephalography (EEG) technique.

Studies have been carried out to investigate for the
possibility of a fast fNIRS response [6, 7]. The averaging of multiple trials besides burdening the computational cost limits the use especially for an online environment. In recent years, works have been reported for recursive estimation of brain activity as coefficients of a linear system by the RLSE algorithm [2-4], and by a Kalman Filter (KF) [3, 8]. These methodologies, having potential of being utilized for real-time applications, were validated online in a time sharing environment by a sequential processing strategy of each channel. The online results provide a fast (compared to an offline) solution to the experimental investigation. But the outcome of each sample of multichannel brain data cannot be ensured before the occurrence of the next multichannel sample. As a consequence, the delay between the occurrence of a sample and its processing grows. Accumulation in this latency drags the online processing away from the real-time results. This is a swear issue for event related real-time studies, e.g., BCI and BI for rehabilitation and gaming, etc. A real-time solution would be having the potential for complex, multichannel, and multitasking BCI studies [9,10].

The RLSE is used as adaptive filter that is meant to estimate the model iteratively [2-4, 11]. It is advantageous over other iterative algorithms: E.g., fast convergence compared to least mean square filtering and computational efficiency over KF. The RLSE exhibits extremely fast convergence due its second-order nature and consistently performs to a higher accuracy than the other iterative algorithms [12].

Reprogrammable hardware attracted a growing interest of researchers for the obtainment of enormous gain in speed and saving of energy besides its capability to be reprogrammed to desired application or functionality requirements after deployment [13-16]. The ease of providing parallel processing environment makes FPGA a high performance computing hardware for real-time processing of multichannel (multi-input, multi-output) systems [17], [18]. An FPGA required minimum expertise of a Hardware Description Languages (HDL; e.g., Verilog or VHDL) to program it. Third party software (like MatLab) introduced HDL code generator to convert the scientific coding into an HDL code for FPGA deployment [19]. Recently, the FPGA vendors presented an open computing language (OpenCL) [20] to get two-fold benefits: (i) A generalized coding environment for different vendors' FPGAs, and (ii) Abstract (high-level) coding environment for an FPGA. Although the programming of an FPGA by this high level language/software significantly reduces the development time compared to the traditional low-level languages (e.g., Verilog or VHDL) but results in significantly under-utilizing the computing capabilities of the device [21]. Thus, a trade-off exist between ease of coding and optimized coding. Researchers are focusing on investigating appropriate optimizers with the high-level software to reduce this trade-off [21].

The current study is aimed to provide a hardware solution for the estimation of brain activity, modelled as coefficients of a linear recursive model, in real-time. For this purpose, the RLSE algorithm [2-4, 11] is implemented on a Field Programmable Gate Array (FPGA) platform. Its multiple cores are instantiated to process multiple channels’ data of a sample simultaneously. A communication core is implemented to receive the sampled data serially (with RS-232 protocol) and place it as global variables. Each RLSE core is signaled by the communication core once its liable data is received and tapped at the designated place. Upon completion of each RLSE process, the computed parameters are transmitted back in real-time for display purposes. The proposed FPGA based embedded environment provides the parallel processing of multiple channels in a sample well before the arrival of the new sample. Thus, ensures the actual real-time processing of brain signals. The computational efficiency of the proposed system is described with quantitative details. On the basis of which, the sampling frequency of the imaging modality is aided for real-time processing of multiple channels. The entire methodology is realized on an FPGA kit (Spartan-6 LX150T) and validated with an fNIRS dataset. A t-statistics is formed to signify the results. A t-score based interpolated brain activation map is draw over the range of the channels. The accuracy of the obtained results are compared with the previously reported offline and online results.

This paper is organized as follows: Materials and methods section provides the modelling and
FPGA Implementation of RLSE Algorithm for Multichannel Brain Imaging

estimation theory of the activity parameters. It further provides the FPGA implementation and computational efficiency for the targeting estimation theory. The subsequent section covers the Results and Discussion of proposed embedded system realized as octal-core on an FPGA platform, which follows the concluding remarks.

2. MATERIALS AND METHODS

Block diagram of the proposed RLSE-FPGA embedded system is shown in Fig. 1. The brain signals, modeled as the coefficients of a linear model, are measured in real-time and received into the platform along with the modeling regressors. The platform stores the data sample to a shared memory and inform the RLSE cores to start the estimation task. Upon completion, the RLSE cores intimate the transmitter to send the estimated parameters for display purpose. The descriptive detail of each block is provided in the following subsections.

2.1. Brain Activity Model

A linear brain activation model with coefficient representing the brain activity is considered as [2-4]

\[ y^1(k) = x_1(k)\beta_1^1(k) + x_2(k)\beta_2^1(k) + \ldots + x_p(k)\beta_p^1(k) + \epsilon^1(k), \]

where k is discrete sampling time instant, \( y^1(k) \) is the brain signal acquired from i-th channel (i = 1,2, ..., N) at k sampling instant, \( x_1(k), x_2(k), \ldots, x_p(k) \) are k-th samples of p regressors, \( \beta_1^1(k), \beta_2^1(k), \ldots, \beta_p^1(k) \) are the coefficients of corresponding regressors representing their strength in channel i at k sampling instance, and \( \epsilon^1(k) \) is the Gaussian noise with zero mean. Equation (1) can be rewritten in vector form as

\[ y^1(k) = X^T(k)\beta^1(k) + \epsilon^1(k). \]

The estimation method and its deploying hardware will decide the accuracy, precession, and speed of the modelled activity parameters.

2.2. RLSE Estimation of Activity Parameters

For each value of input and output at discrete time instant, RLSE recursively estimates the coefficients of an adaptive filter. In our case, the RLSE estimates the optimal values of \( \beta(k) \) of a measured channel by minimizing the following cost-function,

\[ V^i(\beta, k) = \frac{1}{2} \lambda^{k-1} \sum_{j=1}^{k} [\epsilon^i(j)]^2, \]

where \( \epsilon^i \) is the estimation error and \( \lambda \) is the forgetting factor whose value can be varied from 1 to 0 to rely on recent measurements more than the previous ones. The estimation error at time instance k is given by,

\[ \epsilon^i(k) = y^i(k) - X(k)^T\beta^i(k - 1). \]

Solving for minimizing the cost-function in (3), new set of equations are obtained as in [2-4, 11]

\[ K(k) = P(k - 1)X(k)(\lambda + X^T(k)P(k - 1)X(k))^{-1}, \]

\[ \beta^i(k) = \beta^i(k - 1) + K(k)e^i(k), \]

\[ P(k) = [P(k - 1) - K(k)X^T(k)P(k - 1)]/\lambda, \]

where \( K(k) \in \mathbb{R}^p \) is known as weighting vector and \( P(k) \in \mathbb{R}^{p \times p} \) is the recursive inverse of the input covariance matrix. The above mentioned RLSE equations (4)-(7) recursively estimate the optimal values of \( \beta(k) \) for a single channel. For multichannel estimation, multiple RLSE processes are needed to workout independently, as illustrated in Fig. 1. Thus, it becomes the responsibility of the implementing hardware to provide independent processing, precise
estimation, and fast (without latency: processing of the last sample before the arrival of the next sample) environment.

3. PROPOSED HARDWARE

Field programmable gate arrays are reprogrammable semiconductor devices based on configurable logic blocks. These logic blocks can be programmed in any configuration to perform desired arithmetic and logical operations through programmable interconnects. Nowadays, the Matlab based high level coding is advantageous for user prospective [19]. But the Matlab converted verilog/VHDL code is generalized in nature and is not optimal [21]. Such implementation uses more resources and introduces computational complexities as compare to the same algorithm implemented with low level Verilog/VHDL coding. Unlike microcontrollers and digital signal processors where the execution of the code is performed sequentially amongst the channels' estimation, FPGA provides the concurrent and order independent execution of different core instances. This approach drastically reduces the computational time.

The lower (near to machine) level coding of programmable interconnects of an FPGA is not easy especially for a real-time interfacing with high density data computations. But the FPGA platform is preferred owing to its concurrent, order independent, and fast performance. Furthermore, the coding based re-configurability is advantageous for further advancement of the design in future. A block diagram of the proposed FPGA based system is shown in Fig. 1. The fast processing speed of the FPGA along with the low level programing methodology reduces the computation time of an RLSE core. Multiple RLSE cores process multiple channels simultaneously and elevate the computational efficiency.

4. IMPLEMENTATION OF EMBEDDED RLSE FPGA

This section describes the hardware considerations of FPGA platform for the targeting blocks: a serial receiver core, N RLSE cores, and a serial transmitter core. The communication interface is defined for two single precision 32 bit data inputs $X(k) \in \mathbb{R}^p$ and $y(k) \in \mathbb{R}^{1 \times N}$ by RS232_RXD, and an output $\beta(k) \in \mathbb{R}^{p \times N}$ by RS232_TXD. Whereas RS232_RXD and RS232_TXD provides single bit serial receive and transmit interfaces, respectively, based on standard RS-232 protocol. The communication is facilitated by RESET and CLK as controlling inputs while LED[7:0] as status indicators. To make a serial interface compatible with universal asynchronous receiver transmitter (UART) protocol, a CLKX is opted for a desired baud-rate of 115200 bps. The receiver CLK16X is defined 16 times faster to avoid the bit error of the receiving data. Receiver module uar_top receives N channels' sampled dataset $y(k) \in \mathbb{R}^{1 \times N}$ followed by the regression vector $X(k) \in \mathbb{R}^p$ and generates start signals rls_starti for $i = 1$ to $N$. Upon receiving the start signals, the RLSE cores initiate the processing to compute activity parameters. Finally, these cores generate rdy2_i (for $i = 1$ to $N$) signals for the transmitter core uat_top to transmit the computed parameters $\beta(k) \in \mathbb{R}^{p \times N}$ for real-time presentation. The rdy_i (for $i = 1$ to $N$) signals, connected with LEDs, are used as test points to verify RLSE cores' computations. A typical realization of the proposed system for eight channel ($N=8$) is provided in Fig. 2, only selected elements are shown. An RS-232 based UART module is instantiated in top level module including two cores: (i) uar_top to receive the data inputs, and (ii) uat_top to transmit back the estimated parameters. The received data is stored in a memory, not shown in Fig. 2 for simplicity of illustration, and shared to all RLSE cores deployed in FPGA. Though a single RLSE core is shown for simplicity, the complete port list for multiple RLSE cores, working simultaneously, can easily be seen from Fig. 2.

5. COMPUTATIONAL EFFICIENCY

The computational efficiency of the proposed embedded system is assessed for a real-time sample. For a baud-rate of 115200bps, the bit communication time is 8.6805 µs and the transmission time for a single precision parameter (4 bytes $\times$ 10 bits/bytes = 40 bits) is 0.347 ms. A receiving sample of the dataset consists of $p$ parameters of regression vector $X(k)$ and one measured (HbO/HbR) parameter each for $N$ channels. Whereas, transmitting sample consists of $p$ estimated...
parameters, $\beta_1(k), \beta_2(k), \ldots, \beta_p(k)$, each for $N$ channels. Thus, total time consumed in communicating data of $N$ channels is summarized in Table 1.

6. FNIRS DATASET FOR VALIDATION

The validation of the proposed methodology is carried out by utilizing the block-design finger-tapping data made available by Ye et al. [22]. The fNIRS signals were measured from the left motor cortex at 24 locations. Fig. 3 shows the channels’ configuration. The behavioral protocol consisted of an initial 42 sec for signal equilibrium followed by ten repetitions of 21 sec right-index-finger-tapping alternated with 30 sec rest periods. The Oxymon MK III, Artnis instrument was used to acquire the fNIRS data with a sampling rate of 9.75Hz (sample time of 102.6 msec). The availability of the complete dataset provides controllability of the imitated sampling rate to verify the speed of the computations of the system under test.

7. RESULTS AND DISCUSSION

For convenient comparison with the previously reported results, four regressor functions ($p = 4$) have been used [2, 3]. To model brain activity for a cognitive task, $x_4$ is taken to be the Hemodynamic Response Function (HRF) of the experiment with $x_2$ and $x_3$ are the delay and dispersion derivatives of $x_1$, respectively, while $x_4$ is taken as unity to counter the offset.

7.1. Typical Realization of FPGA-RLSE System

The validation is carried out by realization the proposed FPGA-RLSE system on Spartan-6 LX150T FPGA development platform for eight channels ($N=8$). The eight channels are concurrently processed by eight RLSE cores at the available CLK speed of 100 MHz. Top level RTL schematic of the selected elements is shown in Fig. 2. At the beginning, the
receiver core receives the regression vector $X(k)$ which would be common to all the channels for model driven approach. Then receiver core receives the NIRS measurement $y(k)$ for channel 1 and generates rls_start1 while receiving the measurements of channel 2. This enables RLSE core 1 to start its processing promptly. Similarly, each of the eight RLSE cores is interrupted by its respective start signal once its dataset is being completed. Likewise, when the first RLSE core finishes its processing, it generates the ready signals: (i) rdy2_1 to send the computed parameters serially, and (ii) rdy_1 to indicate the task completion on LED0. Similarly, the other RLSE cores will respond to the transmitter as soon as their parameters are ready to be transferred. This approach reduces the computation time further even with a serial interface.

The stated hardware aspects described above are programmed on the FPGA platform by Verilog HDL. The device utilization summary of the platform is described in Fig. 4 as obtained in ISE Design Suit 14.2. The implemented system processes eight channels concurrently in 1.32 $\mu$s. It is worth mentioning that more than eight RLSE cores will take the same processing time because they will be working in parallel.

### Device Utilization Summary

<table>
<thead>
<tr>
<th>Slice Logic Utilization</th>
<th>Used</th>
<th>Available</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Slice Registers</td>
<td>19,602</td>
<td>184,304</td>
<td>10%</td>
</tr>
<tr>
<td>Number used as Flip Flops</td>
<td>19,530</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number used as Latches</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number used as Latch-thrus</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number used as AND/OR Logics</td>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Slice LUTs</td>
<td>51,729</td>
<td>92,152</td>
<td>56%</td>
</tr>
<tr>
<td>Number used as logic</td>
<td>51,355</td>
<td>92,152</td>
<td>55%</td>
</tr>
<tr>
<td>Number using O6 Output only</td>
<td>39,474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number using O5 Output only</td>
<td>695</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number using O5 and O6</td>
<td>11,186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number used as ROM</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number used as Memory</td>
<td>0</td>
<td>21,680</td>
<td>0%</td>
</tr>
<tr>
<td>Number used exclusively as route-thrus</td>
<td>374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number with same slice register load</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number with same slice carry load</td>
<td>309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number with other load</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of occupied slices</td>
<td>16,397</td>
<td>23,038</td>
<td>71%</td>
</tr>
<tr>
<td>Number of MUXCYs used</td>
<td>20,288</td>
<td>46,076</td>
<td>44%</td>
</tr>
<tr>
<td>Number of LUT Flip Flop pairs used</td>
<td>53,308</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number with an unused Flip Flop</td>
<td>35,796</td>
<td>53,308</td>
<td>67%</td>
</tr>
<tr>
<td>Number with an unused LUT</td>
<td>1,579</td>
<td>53,308</td>
<td>29%</td>
</tr>
<tr>
<td>Number of fully used LUT-FP pairs</td>
<td>15,933</td>
<td>53,308</td>
<td>29%</td>
</tr>
<tr>
<td>Number of Unique Control Sets</td>
<td>248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Slice Register Sites Lost to Control Set Restrictions</td>
<td>470</td>
<td>184,304</td>
<td>1%</td>
</tr>
<tr>
<td>Number of bounded IOBs</td>
<td>12</td>
<td>396</td>
<td>3%</td>
</tr>
<tr>
<td>Number of LOCed IOBs</td>
<td>9</td>
<td>12</td>
<td>75%</td>
</tr>
<tr>
<td>Number of RAMB16BWERs</td>
<td>0</td>
<td>268</td>
<td>0%</td>
</tr>
</tbody>
</table>

Fig. 4: Device Utilization Summary of Spartan-6 Lx150t for the fNIRS-RLSE System Implemented for 8 Channels

### 7.2 Real-Time fNIRS Estimation

A User-friendly graphical User Interface (GUI), shown in Fig. 5(a) is made on PC to effectively interface the fNIRS data in real-time with the proposed RLSE-FPGA system. Initially the Connect/Disconnect option checks the availability of the system for connection or disconnect the link. Once available, the
Start/Stop button initiates/inhibits the real-time process by sending/stopping the data. The data is sent sample-by-sample with a predefined sampling rate. The Iteration Number indicates the number of the under processing sample. After reception in PC, the estimated brain activity parameters, $\beta(k)$, of each channel is saved in the local drive besides plotting the channels' HRF coefficients $\beta_i(k)$ (for $i = 1$ to $8$) only along with the estimated outputs for quick interpretation of the results. The HRF regressor is plotted alongside for reference. The results of two distinguished channels are plotted in Fig. 5(b-c) for elaboration.

An interpolated $t$-statistics based brain activation map is drawn, illustrated in Fig. 6, by using the methodology of Aqil et al. [2,3]. Although the map is drawn over the range of 24 measuring channels for comparative analysis but the $t$-scores of only first 8 processed channels are utilized by keeping the remaining channels at non-active values. The brain-mapping template (left lateral view) was depicted using the open-source software NIRS-SPM that was made available by Ye et al. [22]. It is apparent from Fig. 5-6 that the brain activity parameters obtained in real-time are (i) in accordance with the offline results in Ye et al. [22] and (ii) consistent with the online results presented in Aqil et al. [2] with the same dataset (Comparative analysis may be carried out with other FPGA implementations related to the field: No such implementation within the targeted BCI field). Same results in comparison with a non-real-time but precision environment proves the attainment of desired accuracy of the manually implemented data format and their handling (mathematical operations and communication). The proposed FPGA-RLSE system offers beneficial BCI and BI applications in real-time including prompt medical diagnostics and therapeutics.

### 7.3. Computational Cost of Octal Core FPGA-RLSE

Considering the bit size of the parameters for octal core realization, the twelve parameters were received in the FPGA in 4.164 msec. The activity parameters were estimated by the octal RLSE cores concurrently in 1.32 $\mu$s. In return, the four activity parameters each for eight channels (4 $\times$ 8 $= 32$ parameters) were transmitted from the FPGA in 11.104 msec. Thus, the total time consumed to process a sample of 8 channels was 15.269 msec, which is much lower than the sampling time of a typical fNIRS datasets (in the order of 100 ms). Although, the proposed embedded system has faster sample processing rate than the usual sampling rate of an fNIRS modality, the sample processing rate can be made much faster by reducing the communication time by switching from serial (RS-232) to fast protocols (e.g., USB, Ethernet etc.). It is worth mentioning that the processing time of the proposed fNIRS-RLSE realization is very small as compared to the time consumed in communicating the parameters. The processing time still can be reduced further by realizing the proposed system on a fast clocking FPGA platform.

Currently, the FPGA platform transmits the estimated brain activity parameters to a PC for storage in a file besides real-time plotting. The real-time visualization of the results can be facilitated in FPGA by implementing a video graphic adapter (VGA) or high definition multimedia interface (HDMI) core on FPGA. Once introduced, the VGA/HDMI core can further be advanced to display brain activity on an anatomical brain template for effective brain imaging applications.

### 11. CONCLUSIONS

An embedded RLSE-FPGA system is provided for real-time brain activity estimation. Brain activity of multiple channels, coefficients of independent linear models, are estimated by concurrent RLSE cores implemented in an FPGA platform. The interfacing of the RLSE cores for (i) measuring and modeling inputs, and (ii) estimated activity outputs are facilitated by UART interface and FPGA memory. A 32-bit floating point core is deployed to provide computational accuracy. The real-time processing is ensured by assessing the computational cost of the proposed RLSE-FPGA system. The demonstration is carried out by an octal core realization at Spartan-6 LX150T platform validated by real fNIRS dataset. The proposed system has potential real-time BCI and BI applications.
ACKNOWLEDGMENT

A part of this project is supported by National ICT R&D fund by the Ministry of Information Technology, Government of Pakistan.

REFERENCES


