Implementation of first order statistical processor on FPGA for feature extraction

Sugondo Hadiyoso¹, Ahmad Zaky Ramdani¹, Indrarini Dyah Irawati¹, Inung Wijayanto²
¹School of Applied Science, Telkom University, Bandung, Indonesia
²School of Electrical Engineering, Telkom University, Bandung, Indonesia

ABSTRACT
Statistical calculations on signals commonly used in feature extraction. In software processing, statistical computation is an easy task. However, providing a computer requires high costs for simple statistical processing. Another consideration is the need for implementation with real-time and portable processing. Therefore, an alternative device is needed, one of which is the field programmable gate array (FPGA). FPGA is a logic circuit board that can be reconfigured according to computing needs. FPGA can also be used as a prototyping of electronic chips. However, implementing statistical formulas in FPGA is interesting in developing its architecture. Therefore, this research proposes a logic circuit design that can be used for first-order statistical calculations. Statistical parameters include the mean, variance, standard deviation, skewness, and kurtosis. The validation test was performed on the electrocardiogram (ECG) signal series and compared with manual calculations. Validation shows that the mean and variance has very high accuracy with an average error of less than 0.06%.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Sugondo Hadiyoso
School of Applied Science, Telkom University
Bandung, Indonesia
Email: sugondo@telkomuniversity.ac.id

1. INTRODUCTION
The computer process for manipulating and analyzing signals is known as digital signal processing [1]. Basic statistical formulas are often used in digital signal processing to obtain some features [2]. Basic or first order statistical calculations include: mean, variance, standard deviation, skewness, and kurtosis [3]. All of these processes have a high correlation with features in pattern recognition based on features to recognize certain signals. Basic statistical calculations are, of course, very easy on computer applications [4]. Many signal processing applications are used in everyday life such as digital cameras, radar signal detection [5], video processing, processing of various sensor arrays, so the method of processing and transmitting data must be efficient [6].

Statistical computation-based feature extraction has been commonly used in biomedical signal classification. Performance evaluation with several classifier methods shows high accuracy. Statistical computations on the case classification of electrocardiogram (ECG) signals are reported in [7]–[9]. Extraction of statistical features on cases of sleep apnea detection based on ECG signals reported in [10]. Statistical parameters are also used in ECG biometrics as reported in [11]. Other studies on the characterization of EEG signals also use statistical computations [12]–[15]. From previous studies related to the use of statistical methods for feature extraction, it shows that this method is capable of producing high performance. Nevertheless, these studies are applied to computers with large resources. Some real-time
applications require devices that are low cost, easy to move and lead to wearable devices. System on chip (SoC) is an attractive alternative to be developed in digital signal processing applications [16]. However, the implementation calculation on SoC is difficult and encounters many obstacles in its design [17]. SoCs can be developed using a field programmable gate array (FPGA) by designing a logic circuit that performs statistical functions.

Several studies have developed statistical calculations in FPGA devices. An implementation of 2D convolution and media access control (MAC) using Xilinx Vertex FPGA was utilized for the processing of diverse image processing tasks [18]. FPGA demonstrates its capability to serve as an effective platform for speech processing tasks. By implementing a dual-core architecture on FPGA, the computing speed of empirical mode decomposition (EMD) is accelerated, resulting in enhanced efficiency for robust speech recognition [19]. Implementation of the function of calculating variance values in image fusion using FPGA devices [20]. Another study of circuit implementation to calculate the variance and average is reported in [3]. Hardware implementation is one of the accelerated solutions using the FPGA presented to process the variance and averaging frameworks. Other studies have designed logic circuits to calculate the mean reported in [20]. This approach provides a much shorter processing time compared to microcomputers or microcontrollers [21]. This FPGA technology-based approach provides fast, compact, low power consumption for computing and has the ability to execute multiple tasks in parallel [22], [23].

FPGAs have demonstrated their capability not only in fault analysis and circuitry but also in performing complex logic tasks such as neural network implementation and biomedical signal processing. In the fault analysis of multi-level inverters, FPGAs enable the incorporation of decision tree machine learning algorithms to analyze the inverter switches efficiently [24]. Moreover, FPGAs prove their suitability for implementing compact neural networks that replace extensive code in higher-level languages for estimating thermodynamic properties and their derivatives in real-time applications. This allows for efficient computation and storage, crucial for applications like model predictive control and monitoring of power plants and industrial processes [25]. Additionally, FPGAs excel in real-time acquisition and processing of biomedical signals, as demonstrated in the proposed platform for acquiring and processing electroencephalographic (EEG) signals. By combining the parallelism and speed capabilities of FPGAs with the simplicity of a general-purpose processor on a single chip, FPGA-based systems enable real-time operation and high-level task solving, making them ideal for brain-computer interfaces and other biomedical applications [26]. The versatility and flexibility of FPGAs showcase their ability to handle complex logic tasks, making them a valuable tool in various domains such as neural networks and biomedical signal processing.

Studies that are closely related to statistical calculations in FPGA only focus on the mean and variance in digital image processing. On the other hand, a design for digital signal processing is also required. Therefore, this research proposes a logic circuit design that can be used for first-order statistical calculations. The calculated statistical parameters include the mean, variance, standard deviation, skewness, and kurtosis. The validation test was carried out on the ECG signal series. There has been a study proposing a new diagnostic algorithm to accurately detect cardiac disorders at an early stage with an FPGA based design using DE1_SoC by Terasic, which is equipped with a Cyclone V 5CSEMA5F31C6 [27]. In research [28], [29] the ECG signal is modeled by FPGA module, DAC AD9767 14-bit which is observed in real-time with performance based on MSE parameters.

The research proposal aims to design a logic circuit for calculating statistical parameters, including the mean, variance, standard deviation, skewness, and kurtosis on an FPGA board. This system can be used for feature extraction of biomedical signals. We designed this system to process ECG signals in real time. The purpose of this research is described as follows: i) design of logic circuits for basic statistical computations, ii) implementation on FPGA board, and iii) use this system for statistical feature extraction of biosignals.

2. ARCHITECTURAL DESIGN

In the proposed design as shown in Figure 1, this accelerator will be placed as a separate layer after the interface at the hardware level dedicated to incoming data flow. The proposed design is optimized but not limited for use in cases where ECG or EEG signal inputs where data periodically comes continuously. The concept implemented in the proposed design is statistical calculations running as a background process that is separate from the work of the main processor. When the accelerator is in active condition, every incoming data will automatically be included in the calculation. The main processor will be able to get the calculation result by accessing the accelerator address through the advanced extensible interface (AXI) connection.

The first process that will be passed by each incoming data is the buffer process and data accumulation for all incoming data. In addition, there is a counter that will count the amount of data. The mean is calculated as the result of dividing the total accumulated data by the total as expressed in (1).
Mean = $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$ \hspace{1cm} (1)

The results of this mean calculation will then be used again in calculating the variance. The target application of this design is to enumerate continuous data streams without being limited to a predetermined amount of data. Sample variance commonly calculates as average distance sample to mean value. This calculation approach required iterative addition of subtracting the result between all received value and its mean, followed by division with the number of received data.

This approach will be given so much problem in FPGA implementation because the mean value will be kept updated for each new data that is entered, thus the distance calculation must be done repeatedly which means the data must be accommodated first. The idea of this implementation is to provide flexibility in the size of the data to be calculated for the statistical component, so that the strategy adopted is implementation without having to accommodate the data. Thus, to implement the algorithm, the statistical parameters should be calculated as running mean and running variance. We can interpret the running variance itself as the average distance between each data to the mean. After going through a little elaboration, we get a formula that the running variance can be calculated as the average of squared data subtracted by the squared mean. After this variance value is obtained, the value will be calculated by its square root so that we can get the standard deviation value as shown in (2)-(7).

$$Variance = \sigma_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$ \hspace{1cm} (2)

$$Variance = \sigma_x^2 = \frac{\sum_{i=1}^{N} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{N}$$ \hspace{1cm} (3)

$$Variance = \sigma_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \frac{2\sum_{i=1}^{N} x_i}{N} + \frac{N\bar{x}^2}{N}$$ \hspace{1cm} (4)

$$Variance = \sigma_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - 2\bar{x}^2 + \bar{x}^2$$ \hspace{1cm} (5)

$$Variance = \sigma_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2$$ \hspace{1cm} (6)

$$Standard\,\,Deviation = \sigma_x = \sqrt{Variance}$$ \hspace{1cm} (7)

With the same approach, the memoryless implementation of kurtosis and skewness is carried out using a formula that does not require reading back data. Kurtosis and skewness with the most general definitions require seeing the distribution of the entire data and in their calculations several algorithms
Implementation of first order statistical processor on FPGA for feature extraction

(Sugondo Hadiyoso)

require a mean or modulus value for their calculations. Of course, it requires storing the entire data value. Something that will be impractical to be implemented in the memoryless system. Thus, in this implementation the kurtosis and skewness formulas used for performance parameter that can see in the (8) and (9).

\[
Kurtosis = \alpha_4 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{\sigma^4} \tag{8}
\]

\[
Skewness = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^3}{(N-1)\sigma^3} \tag{9}
\]

3. HARDWARE IMPLEMENTATION

The implementation employed in this proposed system utilizes fixed-point arithmetic, as it offers a notable advantage over floating-point implementation on FPGA due to its lower complexity. This choice is made considering the potential quantization errors associated with fixed-point usage [29]–[31]. In order to mitigate the impact of such errors, various techniques are being developed, one of which involves the establishment of a dynamically adjustable quantization range [32], [33] more deep study focusing on the error model of the algorithm has been done [34]. Within this implementation, a similar approach is adopted, whereby the range is dynamically determined, such as by employing varying data widths that adapt to the specific operation stages.

At the level of computational operations, the implementation in this paper is based on extensive research and references in the field of computational operations implementation. For multiplication, the chosen algorithm is the Booth algorithm, with a fixed-width data extension [35]. This selection is supported by prior studies that have demonstrated its effectiveness in achieving accurate results. Furthermore, for square root calculations, the implementation builds upon the insights and findings presented by Putra [36]. These studies have provided valuable guidance in devising an efficient and reliable approach to square root computation within the proposed system.

The hardware is designed to calculate the statistical parameters of the ECG signal series. The ECG signal in the form of a series of decimal numbers is sent serially using the universal asynchronous receiver-transmitter (UART) protocol to the FPGA. The data received by the FPGA is then calculated for the mean, variance, standard deviation, skewness, and kurtosis. The logic block for calculating the mean, variance, and standard deviation is presented in Figure 2. The mean is calculated from the average value by dividing the accumulated results of each data entry by the total amount of data. Meanwhile, for the variance and standard deviation, the calculation of the mean is basically used plus subtractor operations and quadratic functions.

![Figure 2. Mean, variance, and standard deviation processing block](image-url)

Based on the mean, variance and standard deviation values that have been obtained in the previous process, the kurtosis and skewness calculations are carried out according to the formula described in the previous section. The implementation for this kurtosis and skewness calculation block, as was done in the
previous design, is divided into divider, multiplier, accumulator, and subtractor sub-blocks. The design logic for calculating kurtosis and skewness can be seen in Figure 3.

Figure 3. System block for kurtosis and skewness calculation

The design also employs the delay sub-block which functions to accommodate the data value that was entered last time. To perform kurtosis and skewness calculations, variance and standard deviation values are required with a long calculation process. The delay block is also used to ensure that incoming data is not overwritten by new data as long as the variance and standard deviation values are still being calculated.

Table 1 presents the processing time required for each sub-block and each operation to complete the calculation. The longest calculation is for the kurtosis and skewness calculations, each of which requires 187 clock cycles, but part of this period is the initial delay to wait for the values of other components in the form of variance and standard deviation, namely for 87 and 102 clock cycles. As for the periodic calculation process or the latency itself, there is a maximum of 100 clock cycles, from this figure it can be concluded that the fastest new data flow that can still be processed by this design is 100 times the clock period.

Table 1. Required processing time for each operation

<table>
<thead>
<tr>
<th>Operation</th>
<th>Initial</th>
<th>Latency</th>
<th>Total</th>
<th>Period (200 MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divisor (48:32 bit)</td>
<td>0</td>
<td>54</td>
<td>54</td>
<td>270 ns</td>
</tr>
<tr>
<td>Multiplication (16x16 bit)</td>
<td>0</td>
<td>19</td>
<td>19</td>
<td>95 ns</td>
</tr>
<tr>
<td>Multiplication (32x32 bit)</td>
<td>0</td>
<td>35</td>
<td>35</td>
<td>175 ns</td>
</tr>
<tr>
<td>Square root (48 bit)</td>
<td>0</td>
<td>15</td>
<td>15</td>
<td>75 ns</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>52</td>
<td>52</td>
<td>110 ns</td>
</tr>
<tr>
<td>Variance</td>
<td>0</td>
<td>87</td>
<td>87</td>
<td>435 ns</td>
</tr>
<tr>
<td>Variance</td>
<td>0</td>
<td>87</td>
<td>87</td>
<td>435 ns</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>87</td>
<td>15</td>
<td>102</td>
<td>75 ns</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>87</td>
<td>100</td>
<td>187</td>
<td>500 ns</td>
</tr>
<tr>
<td>Skewness</td>
<td>102</td>
<td>85</td>
<td>187</td>
<td>425 ns</td>
</tr>
</tbody>
</table>

Hardware design is described using the very high speed integrated circuit (VHSIC) hardware description language (VHDL). The width of the data used as input is 16 bits in the fix point system, the resolution of the digital ECG data itself is actually only 12 bits or has a value of -2048 to 2048. As for the accumulator, it has a data width of 32 bits which means it has a maximum range of 4,294,967,295 so it will not experienced an overflow of up to 2 million more samples.

In Figure 4, the results of the simulations performed to verify the results of calculations performed by the logic block are presented. The ECG data tested has a length of 2,048 data samples. The excluded mean, variance, and standard deviation values are values with a multiplier of 1 while for kurtosis and skewness are values that have been multiplied by 65,535 and 256 with the aim of avoiding reduced accuracy due to rounding. As described in Table 1, with a clock frequency of 200 MHz, the overall value will be obtained at 500 ns after the last data is obtained.

According to the findings presented in Table 1, it is evident that the most time-consuming component within the proposed design is associated with the computation of kurtosis, demanding a
processing time of 500 nanoseconds. This implies that the implemented system is well-equipped to facilitate real-time calculations, accommodating data rates of up to 2 mega-samples per second. The highest frequency rate typically associated with biosignals is in the range of kilohertz (kHz). Biosignals, which include various physiological signals such as EEG, ECG, electromyography (EMG), and others, generally have frequency components that fall within the kHz range. Consequently, the adoption of the proposed design appears highly conducive for the execution of more intricate feature extraction processes.

4. RESULTS AND DISCUSSION

In this section, validation is carried out by comparing the calculation results of the proposed design and application tools. Another analysis is the use of resources and computation time on FPGA. Figure 5 shows the implementation of the calculation on the Zynq-7000 FPGA board. The ECG signal becomes system input which is processed in real-time.

To be able to carry out the calculation process, there are additional components that are integrated with the design in the form of a serial interface and a buffer so that the FPGA board can receive ECG data to be calculated from a PC. The results of the calculations themselves are seen using the integrated logic analyzer (ILA) which is embedded in the FPGA chip itself. The calculation results obtained are then matched with the results calculated using python as a verification of the accuracy of the calculation.

The results of the timing analysis on the design with a clock speed of 200 MHz, obtained a worst negative slack of 0.229 ns. This means that with this clock speed the block logic is still very flexible to run because with a target period clock of 5 ns the required process delay is still very small. A more complete time analysis report can be seen in Figure 6.
Based on the generated footprint, the proposed design has a relatively small size in terms of logic resources, which is still below 4% for the logic resource. The resource logic circuit can be implemented on the XCZ030SBG which is the target board of this system. Complete results of the required logic resources can be seen in Figure 7.

Figure 6. Timing analysis report on Vivado

Figure 7. Required logic resource for the proposed architecture on XCZ030SBG device

The performance of calculation accuracy for each statistical variable can be seen in Table 2. The mean and variance values have very high accuracy with the calculation difference between the block logic output and the python software calculations on average below 0.06%. Whereas the standard deviation has a slightly higher value, this is because there is a square root calculation block which has the potential to result in rounding of the fractional value of the calculation results. In skewness and kurtosis the ratio error value is slightly higher, this is reasonable because it has a longer calculation path so it is very susceptible to rounding, considering that this implementation is done at a fixed point.

Table 2. The error ratio of the calculation results of each statistical component

<table>
<thead>
<tr>
<th>Signal</th>
<th>Mean (%)</th>
<th>Variance (%)</th>
<th>Standard deviation (%)</th>
<th>Skewness (%)</th>
<th>Kurtosis (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVC1</td>
<td>0.0764</td>
<td>0.0048</td>
<td>2.4729</td>
<td>0.0017</td>
<td>0.0141</td>
</tr>
<tr>
<td>PVC2</td>
<td>0.0697</td>
<td>0.0127</td>
<td>1.0564</td>
<td>0.8350</td>
<td>4.9174</td>
</tr>
<tr>
<td>PVC3</td>
<td>0.0760</td>
<td>0.0131</td>
<td>0.0065</td>
<td>0.3131</td>
<td>1.8769</td>
</tr>
<tr>
<td>PVC4</td>
<td>0.0105</td>
<td>0.0329</td>
<td>0.4605</td>
<td>0.9990</td>
<td>0.9802</td>
</tr>
<tr>
<td>PVC5</td>
<td>0.0288</td>
<td>0.0088</td>
<td>0.0116</td>
<td>0.3773</td>
<td>1.1052</td>
</tr>
<tr>
<td>PVC6</td>
<td>0.1004</td>
<td>0.0106</td>
<td>0.0098</td>
<td>0.1238</td>
<td>0.0021</td>
</tr>
<tr>
<td>PVC7</td>
<td>0.0658</td>
<td>0.0021</td>
<td>2.2269</td>
<td>3.2938</td>
<td>7.5669</td>
</tr>
<tr>
<td>PVC8</td>
<td>0.0282</td>
<td>0.0008</td>
<td>0.7445</td>
<td>2.4123</td>
<td>3.8631</td>
</tr>
<tr>
<td>PVC9</td>
<td>0.0523</td>
<td>0.0177</td>
<td>0.9019</td>
<td>2.5813</td>
<td>1.7055</td>
</tr>
<tr>
<td>PVC10</td>
<td>0.0562</td>
<td>0.0189</td>
<td>0.8562</td>
<td>0.7903</td>
<td>1.2981</td>
</tr>
<tr>
<td>Average</td>
<td>0.05643</td>
<td>0.01224</td>
<td>0.88372</td>
<td>1.17276</td>
<td>2.33295</td>
</tr>
</tbody>
</table>

In comparison to previous studies [37] the implementation presented in this research demonstrates significantly improved accuracy in calculating the mean and variance. However, it should be noted that the accuracy of the standard deviation is relatively lower. The standard deviation is derived from the square root...
of the variance, indicating that the observed decrease in accuracy may be attributed to potential quantization errors during the square root operation. To mitigate such errors, it is recommended to utilize a wider data width prior to performing the square root operation, thereby enhancing accuracy and minimizing the impact of quantization errors.

5. CONCLUSION

In this research, a logic circuit architecture has been designed and simulated for computing statistical parameters namely mean, variance, standard deviation, skewness and kurtosis using VHDL. The developed architecture was then implemented on the Zynq 7000 FPGA board with a fixed-point 16-bit input data width configuration. The synthesis results obtained through Vivado showed that the number of logic sources used was 2675 LUTs and 2614 FFs, which is less than 4% of the total logic sources available in the XCZ030SBG used in this implementation. The results of the timing analysis on the design with a clock speed of 200 MHz, the worst negative slack is 0.229 ns. This design is then tested to calculate the characteristics of the ECG signals. The calculation results which are obtained is then matched with the results calculated using python as verification of the accuracy of calculations. Validation revealed that the mean and variance exhibited very high accuracy with an average error of less than 0.06%. Meanwhile, the standard deviation had a slightly higher error value due to a square root calculation block that had the potential to round the calculated fractional values, resulting in average errors of 1.173% and 2.333%, respectively.

The developed architecture has also been tested in real time. As an additional analysis, system testing will also be carried out in classifying ECG signal. It is expected that the developed architecture can be used for real-time feature extraction of signals originating from bio-sensors. The implementation is performed in a fixed-point numerical format, which represents a constraint open to future enhancements. Utilizing fixed-point notation not only possesses the potential to compromise the precision of computations but also harbors the possibility of data overflow occurrences, particularly when continuous data streams accumulate over extended periods. Hence, further development in the form of implementing a floating-point system within the existing framework warrants serious consideration.

REFERENCES


BIOGRAPHIES OF AUTHORS

Sugondo Hadiyoso received the doctor in electrical engineering from Institut Teknologi Bandung, Indonesia in March 2023. Since 2010, he joined as a lecturer in the Diploma of Telecommunication Technology, School of Applied Science, Telkom University. He is also currently members of Biomedical Instrumentation Research Group in Telkom University. His research interests are wireless sensor network, embedded system, logic design on FPGA, and biomedical engineering. The focus of his doctoral research is signal processing and analysis of EEG signals. He can be contacted at email: sugondo@telkomuniversity.ac.id.
Ahmad Zaky Ramdani is an experienced professional in microelectronics and telecommunication engineering. With a master's degree from ITB in 2015 and a bachelor's degree from Telkom University in 2011, he has dedicated over a decade to research and development. His focus areas include FPGA design, digital signal processing, and algorithm implementation. His work has left a mark in diverse industries such as telecommunications, defence, oil and gas, and biomedical instrumentation. Since 2016, he joined as a visiting lecturer in the Diploma of Telecommunication Technology, School of Applied Science, Telkom University. He can be contacted at email: ahmadzakyramdani@gmail.com.

Indrarini Dyah Irawati obtained B.E. and M.E. degree in electrical engineering at Telkom University, Bandung, Indonesia, and doctoral degree in the School of Electrical and Information Engineering, Institute of Technology Bandung. She joined School of Applied Science, Telkom University as an instructor (2007-2013), assistant professor (2014-2018), associate professor (2019-present). Her main research interests are in the areas of compressive sensing, watermarking, signal processing, computer network, and IoT. She can be contacted at email: indrarini@telkomuniversity.ac.id.

Inung Wijayanto is a lecturer in School of Electrical Engineering at Telkom University, Bandung, Indonesia. He received the B.S. and M.S. degree from Telecommunication Engineering, Institut Teknologi Telkom Bandung, Indonesia, in 2008 and 2011 respectively. Currently he is pursuing his Ph.D. degree in Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interest includes, image processing, and biomedical signal analysis. He can be contacted at email: iwijayanto@telkomuniversity.ac.id.