

**FEDERAL UNIVERSITY OF PELOTAS**  
**Technology Development Center**  
**Postgraduate Program in Computing**



Master's Thesis

**An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for  
ECG Data Compression**

**Arthur Cardozo Godinho**

Pelotas, 2024

**Arthur Cardozo Godinho**

**An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for  
ECG Data Compression**

Master's Thesis presented to the Postgraduate  
Program in Computing of Technology Develop-  
ment Center of the Federal University of Pelotas,  
in partial fulfillment of the requirements for the  
degree of Msc in Computer Science.

Advisor: Prof. Dr. Rafael Iankowski Soares  
Coadvisor: Prof. Dr. Eduardo Antonio Cesar da Costa

Pelotas, 2024

Universidade Federal de Pelotas / Sistema de Bibliotecas  
Catalogação da Publicação

G585u Godinho, Arthur Cardozo

*An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for ECG Data Compression* [recurso eletrônico] / Arthur Cardozo Godinho ; Rafael Iankowski Soares, orientador ; Eduardo Antonio Cesar da Costa, coorientador. — Pelotas, 2024.

87 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2024.

1. ECG. 2. Wavelet. 3. Compression. I. Soares, Rafael Iankowski, orient. II. Costa, Eduardo Antonio Cesar da, coorient. III. Título.

CDD 005

Elaborada por Simone Godinho Maisonave CRB: 10/1733

**Arthur Cardozo Godinho**

**An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for  
ECG Data Compression**

Master's Thesis approved in partial fulfillment of the requirements for the degree of Msc in Computer Science, Postgraduate Program in Computing, Technology Development Center, Federal University of Pelotas.

**Date:** 26th March 2024

**Reviewers:**

Prof. Dr. Rafael Iankowski Soares (orientador)

Doutor em Ciência da Computação pela Pontifícia Universidade Católica do Rio Grande do Sul.

Prof. Dr. Eduardo Antônio César da Costa (coorientador)

Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul.

Prof. Dr. Sérgio José de Melo Almeida

Doutor em Engenharia Elétrica pela Universidade Federal de Santa Catarina.

Prof. Dr. Leomar Soares da Rosa Jr.

Doutor em Microeletrônica pela Universidade Federal do Rio Grande do Sul.

## **ACKNOWLEDGMENT**

First of all, I would like to thank God for offering me the gift of life and giving me so many blessings.

I would like to give a very special thank you to PhD student Morgana da Rosa who made an invaluable contribution to the preparation of the work, helping in the preparation of practically all areas of the work and considering that without her help the creation of this work would not have been possible.

Also giving special thanks to professors Rafael Soares and Eduardo da Costa for all the guidance, directions and corrections that were extremely important in the construction of the work.

And finally, I would like to thank all my family and friends who supported me in every possible way and were also very important in the success of this work.

*I only know that I know nothing.*

— SÓCRATES

## RESUMO

GODINHO, Arthur Cardozo. **An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for ECG Data Compression**. Advisor: Rafael Iankowski Soares. 2024. 88 f. Dissertação (Mestrado em Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2024.

Este trabalho propõe um esquema de compressão VLSI altamente eficiente baseado em DHWT (transformada wavelet Haar discreta) para permitir uma melhor transmissão e armazenamento, especialmente em ambientes com recursos limitados, em especial para a realização do processamento de sinais de (eletrocardiograma). Este trabalho apresenta arquiteturas de hardware podadas, aproximadas e truncadas (PDHWT, AxDHWT e TDHWT) até o nível 5 de compressão. Dentre as diversas soluções desenvolvidas, a solução mais eficiente em termos de consumo energético e área ocupada foi aquela que utiliza todas as técnicas de compressão simultaneamente (poda, aproximação e truncagem). Com relação a esta solução, é possível afirmar que a abordagem PDHWT gera maior redução no consumo de energia ao remover os componentes desnecessários. A abordagem AxDHWT garante maior desempenho ao remover os componentes mais custosos. E por fim, a solução TDHWT reduz o número de bits da entrada, reduzindo assim a complexidade geral da arquitetura. Considerando a solução mais eficiente, foi obtido uma taxa de compressão de 0,03125 (1/32) e um PRD (percent root difference)  $< 1,54$ . A implementação da arquitetura VLSI foi baseada na tecnologia CMOS de 65 nm, com valores de frequência fixos. O trabalho desenvolvido apresentou como grande diferencial o consumo energético extremamente baixo sendo este de apenas  $0,84 \mu W$ , sendo este o menor consumo entre todas as soluções comparadas, advindas da literatura recente para compressão de sinais de ECG.

Palavras-chave: Compressão de dados. Transformadas. Hardware. Eletrocardiografia. Eficiência Energética.

## ABSTRACT

GODINHO, Arthur Cardozo. **An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for ECG Data Compression**. Advisor: Rafael Iankowski Soares. 2024. 88 f. Dissertation (Masters in Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2024.

This work proposes a highly efficient compression scheme utilizing a VLSI-based Discrete Haar Wavelet Transform (DHWT) architecture. This scheme aims to facilitate improved transmission and storage, particularly in resource-constrained environments for electrocardiogram (ECG) signal processing. The work presents pruned, approximated, and truncated DHWT architectures (denoted as PDHWT, AxDHWT, and TDHWT, respectively) up to level 5, targeting ultra-high energy efficiency in ECG data compression. Among the developed solutions, the most energy-efficient and area-optimized approach employs all three compression techniques simultaneously (pruning, approximation, and truncation). In this combined approach, the PDHWT technique achieves significant energy savings by eliminating redundant components. The AxDHWT approach enhances performance by removing the most computationally expensive elements. Finally, TDHWT reduces the number of input bits, leading to a lower-complexity architecture. This combined solution achieves a compression ratio of  $0.03125 \left(\frac{1}{32}\right)$  and a percent root difference (PRD) less than 2.08. The VLSI architecture implementation leverages 65 nm CMOS technology with maximum frequency of 1.1 GHz. A key differentiator of this work is the remarkably low energy consumption of  $0.6 \mu W$ , which surpasses all existing ECG signal compression solutions reported in recent literature.

Keywords: Data compression. Transform. Hardware. Electrocardiography. Energy efficiency.



## LIST OF FIGURES

Figure 1	The ECG Signal (SEIDEL et al., 2021) . . . . .	17
Figure 2	ASIC-based synthesis methodology diagram (DA ROSA et al., 2023).	24
Figure 3	ECG wearable devices: a) Wearable 12-lead ECG Smart-Vest System(LIU et al., 2019). b) Rigid-flex design of the sensor patch for health monitoring(WU et al., 2020). c) Sensor patch is attached to a subject's chest using bio-compatible tapes(WU et al., 2020). . . . .	25
Figure 4	Representation of wavelet transforms. . . . .	27
Figure 5	Five levels cascade filter-bank structure for Haar wavelet transform (DING; ZILIC, 2016). . . . .	28
Figure 6	Graphical representation of the signal concatenations. . . . .	35
Figure 7	Complete visualization of the QRS complex with intervals (JINDAL et al., 2018) . . . . .	36
Figure 8	Block scheme of Pan-Tompkins based detector (GALA; BARABAS; KRAJNAK, 2020) . . . . .	37
Figure 9	Algorithm processes of Pan-Tompkins: a) Signal entry, b) Signal resulting from adaptive filter and adaptive thresholds, c) Resulting signal from derivation, d) Squared signal, e) Result of the signal at the output of the window integrator and the adaptive thresholds of the signal, f) Out of the R peak detection algorithm (COSTA, 2021) . . . .	38
Figure 10	Blocks M1 - a) and M2 - b) (ROSA et al., 2021). . . . .	43
Figure 11	Block M3 (ROSA et al., 2021). . . . .	43
Figure 12	Haar wavelet transform composed of five M1 blocks. . . . .	45
Figure 13	The SCM block (SEIDEL et al., 2021). . . . .	46
Figure 14	Haar wavelet transform composed of two M2 and one M1 blocks. . .	46
Figure 15	Haar wavelet transform composed of one M2 and one M3 blocks. .	47
Figure 16	Pruned Haar wavelet transform composed of five M1 blocks. . . . .	48
Figure 17	Pruned Haar wavelet transform composed of two M2 and one M1 blocks. . . . .	48
Figure 18	Pruned Haar wavelet transform composed of one M2 and one M3 blocks. . . . .	48
Figure 19	Approximate and pruned Haar wavelet transform composed of five M1 blocks. . . . .	49
Figure 20	Approximate and pruned Haar wavelet transform composed of two M2 and one M1 blocks. . . . .	50
Figure 21	Approximate and pruned Haar wavelet transform composed of one M2 and one M3 blocks . . . . .	50

Figure 22	Block diagram exemplifying the bit truncation (LEE; KIM, 2023) . . .	51
Figure 23	Block diagrams presenting the bit truncation types: a) input number with zero truncation, b) input number with nonzero truncation, c) random number with zero truncation (LEE; KIM, 2023). . . . .	52
Figure 24	Accuracy comparison in all compression levels for both NSR and PVC signals . . . . .	59
Figure 25	Step by step execution of the peak detection algorithm in a sample of the NSR signal . . . . .	59
Figure 26	Step by step execution of the peak detection algorithm in a sample of the PVC signal . . . . .	60
Figure 27	Results of applying the peak detection algorithm to the NSR signal at the five compression levels . . . . .	60
Figure 28	Results of applying the peak detection algorithm to the PVC signal at the five compression levels . . . . .	61
Figure 29	MAE values considering the NSR and PVC signals at each compression level . . . . .	61
Figure 30	Evolution of NCC along decomposition levels of NSR and PVC signals	62
Figure 31	PRD comparison in all compression levels for NSR and PVC signals	63
Figure 32	RMSE comparison in all compression levels for NSR and PVC signals . . . . .	63
Figure 33	SNR comparison in all compression levels for NSR and PVC signals	64
Figure 34	SSIM comparison in all compression levels for NSR and PVC signals	64
Figure 35	Truncating full PVC signals . . . . .	66
Figure 36	Truncating PVC signals with 2,000 samples . . . . .	66
Figure 37	Relationship between accuracy, decomposition levels and approximate number of bits . . . . .	68
Figure 38	Relationship between MAE, decomposition levels and approximate number of bits . . . . .	69
Figure 39	Relationship between NCC, decomposition levels and approximate number of bits . . . . .	69
Figure 40	Relationship between PRD, decomposition levels and approximate number of bits . . . . .	70
Figure 41	Relationship between RMSE, decomposition levels and approximate number of bits . . . . .	71
Figure 42	Relationship between SNR, decomposition levels and approximate number of bits . . . . .	71
Figure 43	Relationship between SSIM, decomposition levels and approximate number of bits . . . . .	72

## LIST OF TABLES

Table 1	Comparison between the main related works and the work developed	54
Table 2	Comparison of synthesis results between the developed architectures with a period of 909 ms. . . . .	56
Table 3	Comparison of synthesis results between the developed architectures with a period of 7812500 ms . . . . .	56
Table 4	Table containing accuracy metrics for the NRS signal . . . . .	57
Table 5	Table containing accuracy metrics for the PVC signal . . . . .	58
Table 6	Accuracy metrics for 1-bit truncation . . . . .	67
Table 7	Accuracy metrics for 4-bit truncation . . . . .	67
Table 8	Accuracy metrics for 6-bit truncation . . . . .	67
Table 9	Accuracy metrics for 8-bit truncation . . . . .	68
Table 10	Comparison of synthesis results between different truncation levels for the M2M2M1 architecture . . . . .	72
Table 11	Comparison between the main developed solutions and the baseline	73
Table 12	Comparison with other ECG data compression architectures. . . . .	73

## LIST OF ABBREVIATIONS AND ACRONYMS

ECG	electrocardiogram
CVDs	Cardiovascular diseases
WHO	World Health Organization
DHWT	Discrete Haar Wavelet Transform
IoT	Internet of Things
VLSI	Very Large Scale Integration
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
MOS	Metal-Oxide Semiconductor
RTL	Register Transfer Level
SDF	Standard Delay Format
CPD	Critical Path Delay
VHDL	Very High-Speed Integrated Circuits Hardware Description Language
VCD	Value Change Dump
TCF	Toggle Count Format
SDF	Standard Delay Format
PLE	Physical Layout Estimator
LEF	Library Exchange Format
AALO	Ant Lion Optimizer
NILM	Non-intrusive load monitoring
IDWT	inverse discrete wavelet transform
AxC	Approximate Computing
KG	Kalman Gain
CMA	Carry-Maskable Adder
NSR	Normal Sinus Rhythm
PVCs	Premature Ventricular Contractions
MAE	Mean Absolute Error

NC	Normalized Cross-correlation
PRD	Percentage Root-mean-square difference
RMSE	Root-mean-square Deviation
SNR	Signal-to-noise ratio
SSIM	Structural similarity index measure
ODHWT	Original Haar Wavelet Transform
PDHWT	Pruned Haar Wavelet Transform
AxDHWT	Approximate and Pruned Haar Wavelet Transform
SCM	single constant multiplications
TPDHWT	Truncated and Pruned Discrete Haar Wavelet Transform
TAXDHWT	Truncated and Approximated Discrete Haar Wavelet Transform
EMG	Electromyogram
EEG	Electroencephalography
AI	artificial intelligence

# CONTENTS

<b>1</b>	<b>INTRODUCTION</b>	<b>16</b>
1.1	The Challenge	16
1.2	The Problem Definition and Solution	18
1.3	The Contributions of this work	19
1.4	Outline	20
<b>2</b>	<b>BACKGROUND OF HARDWARE DESIGN CONCEPTS AND TECHNIQUES</b>	<b>21</b>
2.1	Power Dissipation	21
2.1.1	Power dissipation background	21
2.1.2	Synthesis flow for power estimation	23
2.1.3	Energy consumption overview in biomedical area	24
2.2	Wavelet Transform	26
2.2.1	Wavelet transform background	26
2.2.2	Wavelet transform overview	29
2.3	AxC: Approximate Computing	30
2.3.1	AxC Background	31
2.3.2	AxC Overview	32
2.3.3	AxC: Approximate computing in compression data	33
2.4	Methodology	34
2.4.1	Utilized database	34
2.4.2	Metrics for Accuracy Evaluation	35
2.5	Chapter conclusion	41
<b>3</b>	<b>DISCRETE WAVELET TRANSFORM ARCHITECTURE PROPOSALS</b>	<b>42</b>
3.1	Developed Architectures	42
3.1.1	ODHWT: Original Haar Wavelet Transform	45
3.1.2	PDHWT: Pruned Haar Wavelet Transform	46
3.1.3	AxDHWT: Approximate and Pruned Haar Wavelet Transform	48
3.1.4	Truncation	50
3.2	Related Work: Compression Techniques for Hardware Applications	52
3.3	Chapter Conclusion	53
<b>4</b>	<b>RESULTS AND ANALYSIS</b>	<b>55</b>
4.1	Synthesis Results	55
4.2	Accuracy Results	57
4.2.1	Summary of accuracy results	57
4.2.2	Results with truncation	65
4.3	Result comparisons with works from the literature	73

<b>5</b>	<b>CONCLUSION . . . . .</b>	<b>76</b>
<b>5.1</b>	<b>Possibilities for future work . . . . .</b>	<b>77</b>
	<b>REFERENCES . . . . .</b>	<b>78</b>

# 1 INTRODUCTION

This chapter introduces the work by presenting the problem, the primary challenges to overcome, and a brief presentation of the solution developed. It also offers a brief explanation regarding the composition of the ECG signal, its wave structure, its main characteristics, and finally, it presents the main contributions of the work.

## 1.1 The Challenge

Cardiovascular diseases (CVDs) are the leading cause of death in the world according to the World Health Organization (WHO) (CHIANG et al., 2019). Sudden cardiac death (SCD), mainly caused by arrhythmia, is defined as death due to cardiovascular causes in a patient with or without known preexisting heart disease, in whom the mode and time of death are unexpected. It is well known that SCD is the manifestation of a fatal heart rhythm disorder, such as ventricular tachycardia (VT) and ventricular fibrillation (VF) (LAI et al., 2019). Those diseases need to be treated instantly, and the diagnosis needs to be quick to preserve life. Electrocardiography is one of the most used techniques to ensure the diagnosis of these diseases in a non-invasive way. This procedure records the electrical changes on the skin due to the heart's electrical activity. The data provided by this procedure is known as an electrocardiogram (ECG), which can be utilized for various diagnostic purposes, among them are the detection and help to diagnose not only cardiovascular diseases but also other problems that indirectly affect the heart activity (BURGUERA, 2019).

The Electrocardiogram is a graphical of minute electrical samples from cardiac muscles (HAMZA; RIJAB; HUSSIEN, 2021). In a long-range watching of cardiac patients, record samples of ECG data show a big capacity from memory for storage or a significant bandwidth of communication channels in case of sending ECG data (CHANDRA'; SHARMA; SINGH, 2021). Therefore, ECG data collection must be carried out constantly to monitor patients health conditions efficiently. This operation generates a large amount of data, which needs to be handled efficiently and safely to enable constant monitoring. Another major challenge faced in the persistent collection of medical signals,



especially in the case of ECG signals, is energy consumption. Since the equipment for constant monitoring needs to be powered using batteries, the equipment that collects these signals needs to be energetically efficient to guarantee greater autonomy (ZHANG et al., 2018).

Given the problems above, it is possible to conclude that the transfer and manipulation of raw ECG data is unsuitable for most use cases, especially for continuous monitoring, due to the large consumption of computational resources, energy consumption being the most critical factor. Therefore, the solution used by most works in the literature is signal compressing. Compression is achieved by eliminating the redundancies from data, reducing storage requirements, and minimizing the redundant data points found in the ECG data while protecting the essential diagnosis features of the original data (SINGH PAL et al., 2022).

There are two main types of compression, i.e., the lossy compression and the lossless compression. The former approach presumes that no losses (distortions) are introduced and, thus, all valuable information is preserved. Unfortunately, in this case, the compression rate usually is small (KRIVENKO et al., 2018). On the other hand, lossy compression appears as an option, as it allows much higher compression rates and can achieve the requirements generating practical compression algorithms for fast signal transmission (ABDULBAQI et al., 2018).

The ECG signal wave comprises the P and T waves and the QRS complex. This wave results from contracting and expanding a heart muscle, as seen in Figure 1. The depolarization of atrial muscles produces a P wave caused by contraction of the atria. At the same time, the QRS complex describes the effect of the atrium re-polarization and the essentially concurrently depolarization of ventricles due to atrium contracting. The ventricular re-polarization generates a T wave, which causes profit to the resting part of the ventricular mass (HAMZA; RIJAB; HUSSIEN, 2021).

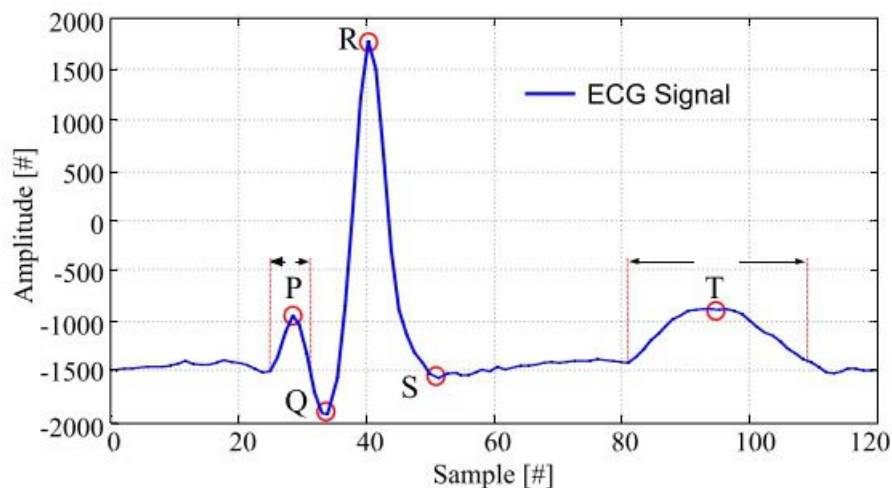


Figure 1 – The ECG Signal (SEIDEL et al., 2021)

In summary, collecting ECG signals in real-time and with sufficient quality is necessary to make a correct medical diagnosis possible. Therefore, working with data directly is not viable, requiring techniques such as compression to manipulate them. The proposed solution will be presented below, along with its general details and implications.

## 1.2 The Problem Definition and Solution

Regarding methods for performing signal compression, a significant improvement has been accomplished in transform-based techniques, where signals are transformed to another domain. The most popular transforms are the Fourier transform, discrete cosine transform (DCT), and discrete Wavelet transform (DWT), commonly used in different compression algorithms (SINGH PAL et al., 2022). Among these methods, DCT compression offers a significant advantage: the ability to analyze a signal through a two-dimensional time vs. frequency function with a variable resolution. Among the families of Wavelet transform, the Haar stands out for its mathematical simplicity and low computational effort required to perform its calculations (SEIDEL et al., 2020).

This way implies that, at the same time, a high reduction in computational resources is necessary to enable the real-time application of the solution. It is also essential to ensure that the QRS complex remains unchanged during the compression process, allowing accurate extraction of information from the signal.

Therefore, the primary objective of the work is to develop an architecture capable of compressing ECG signals to allow great energy and area savings. We use the Discrete Haar Wavelet Transform (DHWT) and the three main strategies from the literature for hardware optimization to perform the compression. As secondary goals, we developed several sub-architectures that, while failing to achieve the most excellent efficiency when it comes to the consumption of hardware resources, are solutions with good trade-off potential between signal quality and savings. Still, as a secondary objective, we also compare the different forms of hardware implementation of the DWT transform, highlighting the best solution and its characteristics.

The pruning was the first strategy to optimize the already computationally simple Haar transform. The pruning primary goal is to reduce the hardware based on custom acceleration algorithms to maintain the required mathematical property (SEIDEL et al., 2021). The pruning can be carried out in several different ways, whether to reduce the size of the training models in IA applications (XIONG et al., 2023), to compress neural networks (XUANHUI; JUAN; QUAN, 2021), or to reduce the overall project cost via removal of unused components (SEIDEL et al., 2021). This last case was the pruning applied to the transform Haar, removing unnecessary components for generating compression. This way guarantees a gain in energy consumption and hardware

space without losing compression quality. One can classify compression with pruning as lossless compression, with the difference that, unlike most lossless compression techniques, in this case, there is a significant real gain in terms of the total project cost.

The pruning strategy allowed us to obtain a considerable performance gain. However, there are other techniques capable of expanding the gains already obtained. Hardware approximation stands out among the leading existing solutions for optimization at the hardware level. Approximate computing is one of the effective methods used to optimize circuit design. One uses approximate computing by maintaining the desired level of accuracy in error-resilient applications where approximate results are optimal alternatives to exact results (BARAATI et al., 2022). As the lossy compression approach, like approximate hardware components, uses inexact approximations, understanding the impact of data compression techniques in terms of bit and error rate could be of significance in making ECG records usable in remote clinical healthcare (MOON et al., 2020) (BURGUERA, 2019).

Considering this scenario, the approach will punctually act on the most expensive hardware component, and its replacement/removal would bring the greatest advantage to the hardware in terms of performance. This component is the multiplier once multipliers are one of the most widely used arithmetic blocks in digital circuits. Also, multipliers have high power consumption and hardware area occupation (BARAATI et al., 2022). Considering the application of the Haar transform, the multiplier is an essential logic component responsible for most of the overall power consumption in the signal processing (RAMASAMY; NARMADHA; DEIVASIGAMANI, 2019). Therefore, its optimization is necessary for developing an efficient compression system.

In addition to the optimization strategies mentioned above, we used bit truncation to achieve the maximum possible performance. Since, at the moment, the most inherently adopted approximation in practice for significant savings in power is the reduction of bit-widths by truncation (SEIDEL et al., 2021). Bit truncation has made it possible to achieve a higher degree of energy consumption reduction and highly competitive results compared to other solutions in the recent literature.

We promoted direct comparisons with other solutions in the literature to validate the developed solution, especially with the current state-of-the-art. Gold standard error metrics performed the comparisons, allowing a complete analysis of the quality of the generated signal and an evaluation of the consistency of the essential features of the ECG signal (DA ROSA et al., 2022a).

### **1.3 The Contributions of this work**

The novel contributions of this work are as follows:

1. It presents a Discrete Haar Wavelet Transform (DHWT) matrix approximation

capable of satisfying the quality of service in the application and producing a multiplier-less hardware architecture.

2. It offers pruning in the approximate DHWT matrix, reducing the DHWT hardware architecture to just seven parallel additions.
3. It demonstrates that the developed solution can support truncation without losing the signal's essential characteristics and generating significant distortions, resulting in considerable energy savings.
4. It promotes comparisons of the developed solutions with the state-of-the-art, showcasing the trade-offs made during the project's development while emphasizing its primary strengths.
5. It allows a discussion regarding the efficiency in the detection of the main properties of the ECG signal after compression, outlining the circuit space and power consumption advantages, guaranteeing the procession of ECG signals with exceptional quality.

## **1.4 Outline**

The work is organized as follows: Chapter 2 presents the design of the hardware to be developed, as well as detailed explanations regarding both the construction of the hardware and the synthesis methodology. Chapter 3 presents the complete solution and all developed versions in greater details. Chapter 4 brings the obtained results and their comparisons with other relevant works from the literature and between the implemented architectures. Finally, Chapter 5 contains the conclusion and possibilities for future works and improvements.

## **2 BACKGROUND OF HARDWARE DESIGN CONCEPTS AND TECHNIQUES**

This chapter is divided into five subchapter. First, an overview of power dissipation and energy consumption is given, highlighting its main problems, and opportunities for optimization. The second subchapter discusses the conceptualization of approximate computing and its main applications, focusing on data compression. The third subchapter presents the methods used: the synthesis method, which is responsible for describing the main tools and techniques used, as well as the results in terms of energy and area savings. The other method presented is the accuracy method, which provides more detailed information on how the signal quality assessment was carried out and the minimum conditions for acceptance. Soon after, The fourth subchapter presents wavelet transform theory and its applications, especially in signal compression and optimization. The fifth and final subchapter presents the topic's conclusion and briefly summarizes the main points covered in the chapter.

### **2.1 Power Dissipation**

In recent literature, many works have reduced energy consumption as their central focus, especially regarding equipment with continuous use or IoT (Internet of Things) devices. In the era of the IoT, where continuous monitoring of signals, carried out by wearable devices, becomes increasingly a reality, the concern with reducing energy consumption becomes immense (ROSA et al., 2021). This section presents the main concepts of energy consumption and overviews some works that treat energy consumption reduction in the biomedical area.

#### **2.1.1 Power dissipation background**

Power dissipation is one of the major concerns when designing digital circuits. The increase of portable devices requires the design of VLSI (Very Large Scale Integration) circuits from the point of view of reducing power dissipation (Rabaey (1996)). Excessive power dissipation integrated circuits make their use in portable systems infeasible.

ble and cause overheating that degrades performance and shortens their life (Najm (1994)). This section discusses the most influential parameters of power dissipation in CMOS (Complementary Metal-Oxide Semiconductor) circuits.

The main reason for the popularity of CMOS logic is that conventional logic gates (*AND*, *OR*, *NAND*, *NOR*, *NOT*, *XOR*) do not have static power dissipation unless their outputs switch between logic levels (Weste; Eshraghian (1994)). However, power dissipation is generated in the transistors for each output of a CMOS gate whose logic value changes (Martin (2000)). The original reason for this power dissipation is the movement of electric charges that charge and discharge external load capacitances and internal parasitic capacitances.

There are three primary sources of power dissipation in CMOS circuits (Weste; Eshraghian (1994)): i) power dissipation caused by leakage currents and sub-threshold currents, ii) short-circuit power dissipation, which occurs due to the direct current flow from the power supply to the ground during the switching process in a transistor gate, and iii) dynamic power dissipation, which is the result of loading and unloading the capacitances of the circuits.

Static dissipation arises from leakage currents in metal-oxide-semiconductor field-effect transistors (MOSFETs) and from parasitic elements created by interconnecting with the substrate. Ideally, the static power dissipation of a CMOS circuit is zero. However, with new technologies below 90nm, this powerful component has become quite significant. There are always leakage currents that can occur when one transistor is off, and another active transistor charges the drain concerning the substrate potential (Chandrakasan; Sheng; Brodersen (1992)) (up/down).

This fraction of power dissipation becomes increasingly more prominent in the total power dissipation, especially when the semiconductor process technology reaches values below 100nm (Kim; Blaauw; Mudge (2003)). Studies show that in the case of an inverter circuit 70nm technology subjected to simulations at 125°C, leakage currents can contribute about 49% to the total power consumption of the circuit (Kim; Roy (2002)).

Equation 1 shows the static power as a function of leakage current and supply voltage. On the other hand, Equation 2 shows the leakage current given by the reverse diode characteristic equation, where  $I_S$  is the reverse saturation current,  $V$  is the voltage across the diode, and  $V_T = \frac{KT}{q}$  is the thermal load.

$$P_{esttica} = I_{leakage} \cdot V_{dd} \quad (1)$$

$$I_{leakage} = I_S (e^{\frac{V}{V_T}} - 1) \quad (2)$$

According to Chandrakasan; Brodersen (1995), consumption of static power by the subthreshold current due to the diffusion of carriers between source and drain ( $V_{ds}$ )

occurs when the gate-source voltage ( $V_{gs}$ ) exceeds the weak inversion point but still has a value below the threshold voltage ( $V_t$ ). In this regime, the MOS (Metal-Oxide Semiconductor) transistor behaves similarly to a bipolar transistor, and the subthreshold current is exponentially dependent on the gate-source voltage ( $V_{gs}$ ). Equation 3 shows the subthreshold current, where  $V_T$  is the thermo-electric voltage and  $Ke$  is a technology-dependent parameter. In the equation,  $n$  corresponds to:  $n = 1 + \frac{\omega t_{ox}}{D}$ , where  $t_{ox}$  is the thickness of the gate oxide,  $D$  is the channel depletion width, and  $\omega$  corresponds to the equation  $\omega = \frac{\epsilon_{si}}{\epsilon_{ox}}$ , where  $\epsilon_{si}$  is the permittivity of the insulating material and  $\epsilon_{ox}$  is the permittivity of the oxide.

$$I_{ds} = Ke^{\frac{V_{gs}-V_t}{nV_T}} (1 - e^{-\frac{V_{ds}}{V_T}}) \quad (3)$$

Due to switching operations, dynamic power dissipation comes from the charging and discharging of circuit capacitances. The dynamic switching component of power dissipation ( $P_{din}$ ) at a node at the output of a gate charged by a capacitor  $C_L$  is given by the Equation 4, where  $A$  is the output node activity measured in events/second for a full charge/discharge,  $f$  is the switching frequency, and  $V_{dd}$  is the supply voltage.

$$P_{din} = \frac{1}{2} C_L A V_{dd}^2 = \frac{1}{2} a f C_L V_{dd}^2 \quad (4)$$

### 2.1.2 Synthesis flow for power estimation

Figure 2 shows the current industrial methodology for determining power dissipation. First, the register transfer level (RTL) description is synthesized using a commercial synthesis tool. We use Cadence Genus<sup>TM</sup> Synthesis tool in this work. Secondly, the tool generates the circuit netlist in Verilog, the SDF (Standard Delay Format) file, which records the specific delays for gates and nets, as well as cell area, power dissipation, and critical path delay (CPD) reports. Then the simulation tool (such as Cadence Genus<sup>TM</sup>) simulates the generated netlist with the testbench files (in VHDL or Verilog or SystemVerilog language) and feeds the circuit with input data to generate the dump file, which is either in VCD (Value Change Dump) or TCF (Toggle Count Format) files. The input data used as stimuli is extracted from the target application software and stored in text files. Using these text files, the simulation tool obtains these values and runs the testbench with the gate-level netlist files and the SDF delay files to generate accurate switching activities. At the end of the simulation, a dump file is created containing the switching activities of all nodes in the circuit's netlist. Finally, the synthesis tool is run a second time with the same parameters. During this process, the dump file is passed to the software to generate accurate power dissipation reports. Modern synthesis tools support estimating the drawbacks of gate connections in terms of circuit area, power dissipation and delay. The Cadence Genus logic synthesis tool provides physical logic synthesis in Physical Layout Estimator (PLE) mode. This tool

estimates the length of networks and considers the impact of load capacity on power dissipation, using an estimate of layout routing. Such analysis requires the inclusion of Library Exchange Format (LEF) files, which primarily contain the physical layout information of the library. The LEF macro contains the capacity of the inner library cell, and the Tech LEF contains the process metal capacity for estimating interconnect capacity. In addition, standard cell libraries usually provide an additional file, the capacity table, which describes the technology capacities in more detail.

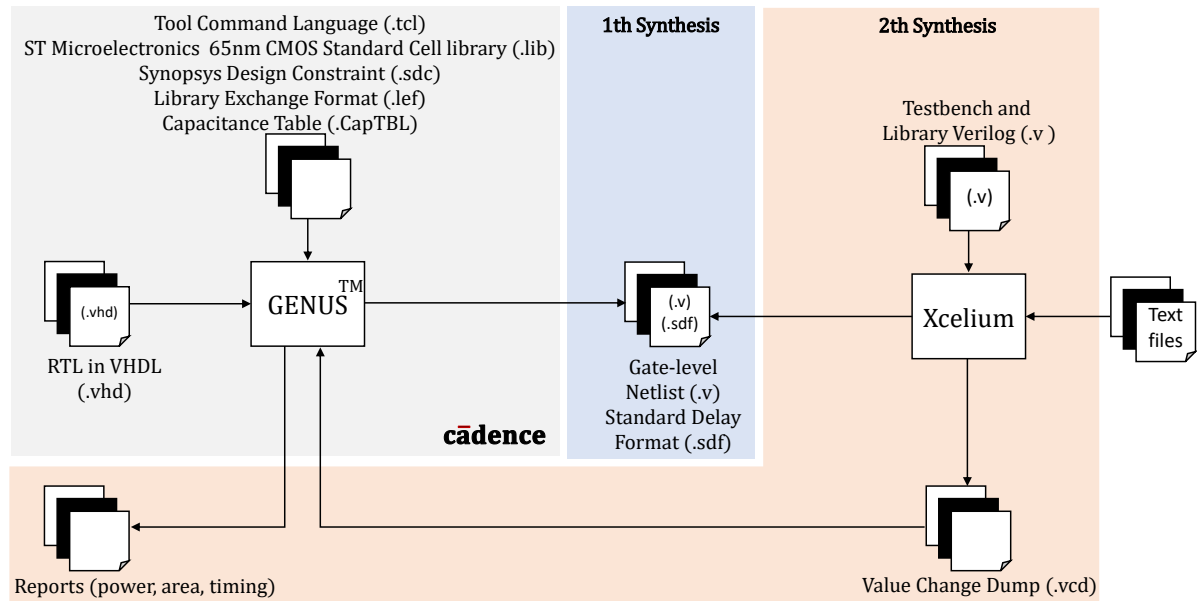


Figure 2 – ASIC-based synthesis methodology diagram (DA ROSA et al., 2023).

### 2.1.3 Energy consumption overview in biomedical area

The work proposed by (AHMED et al., 2022) highlights the limited energy storage capacity in IoT and mobile devices in general, primarily because these devices must be powered by batteries, resulting in a restricted amount of energy available for their operation. This work also presents some solutions for reducing consumption in image processing. The leading solution proposed is using the adaptive Haar transform for compression. In (VAANANEN; HAMALAINEN, 2021), it emphasizes the need to develop solutions capable of operating independently for long periods and presents energy consumption as the main limiting factor for this functionality. The work in question proposes reducing consumption in LoRa sensor networks. The work presented by (XI-ONG et al., 2023) uses a discrete wavelet transform (DWT) to compress test signals and facilitate learning AI techniques on broad data sets to facilitate management and predict consumption in smart electrical grids.

Focusing more on treating medical signs, the work developed by (FATHI et al., 2022) proposes an algorithm to reduce the energy consumption in Remote Healthcare Monitoring Systems. The algorithm uses the discrete Krawtchouk moments and the AALO



(Ant Lion Optimizer) to compress the ECG signal and extract its features with a significant reduction in the hardware resources used and, consequently, in energy consumption. The works (SEIDEL et al., 2021), (ROSA et al., 2021), and (SEIDEL et al., 2020) use the wavelet transform together with other methods of optimization like hardware pruning and approximate computing concepts, to reduce the power consumption in the context of ECG signal collection and transmission.

As we have already seen, energy consumption is a significant challenge that we have to face. The need to use ever more efficient and, above all, more energy-efficient devices is on the rise. With the advent of IoT technologies and the proliferation of edge computing, the need to use ever more sophisticated devices with more powerful hardware that must also have low consumption is increasing, as they are dependent on very restrictive energy sources with low autonomy, such as batteries (ZHOU et al., 2019). These high development restrictions make the constant evolution of the technologies used in these devices essential to meet the demand for new and more efficient devices.

One of the most critical areas where computing devices, and more recently IoT devices, are widely used is the medical area. In this area, wearable devices have been increasingly used, obtaining favorable results for improving patients' quality of life and data acquisition (CHEN et al., 2020). These devices detect various signals, such as the oxygen content of the blood, blood pressure, ECG, etc. Among the main monitored signals, the ECG signal has stood out due to its particular need for continuous monitoring since this signal is vital for detecting diseases with high fatality rates, such as arrhythmia and heart attack (SEIDEL et al., 2021). Figure 3 presents wearable applications for collecting ECG signals remotely and continuously.

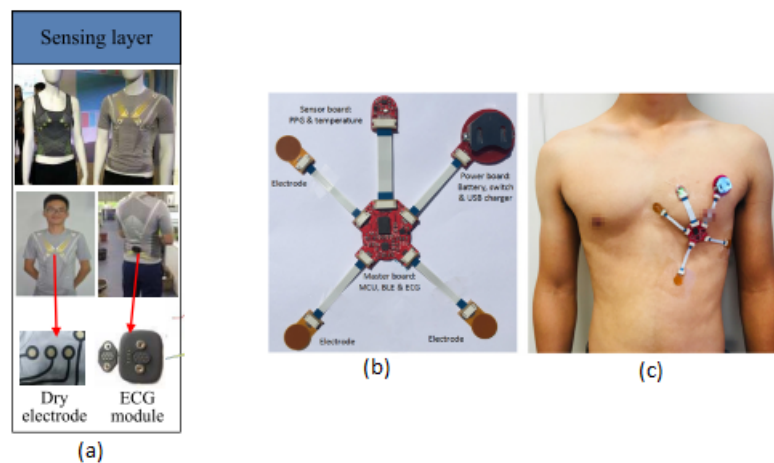


Figure 3 – ECG wearable devices: a) Wearable 12-lead ECG Smart-Vest System(LIU et al., 2019). b) Rigid-flex design of the sensor patch for health monitoring(WU et al., 2020). c) Sensor patch is attached to a subject's chest using bio-compatible tapes(WU et al., 2020).

The continuous acquisition of this signal over long periods is very costly in energy

consumption, making it necessary to develop solutions to minimize this consumption. A widely used form of compression is the application of compression of the signal obtained to alleviate communication and reduce consumption in the transmission/storage of collected data (SEIDEL et al., 2021). This dissertation follows this trend, focusing on reducing consumption through data compression to facilitate transportation and storage. We use the discrete Haar Wavelet transform for this purpose.

## 2.2 Wavelet Transform

This section presents a background of wavelet transform with concepts and characteristics and offers an overview of works from the literature exploring wavelet transform in many applications.

### 2.2.1 Wavelet transform background

The wavelet transform is nothing but a short waveform with finite duration with zero average value. Compared to the sine function, the wavelets range between  $-\infty$  to  $+\infty$ . Among the existing types of wavelet transforms, the discrete wavelet transform (DWT) is the most common technique used for general compression, including image and 1-D signal compression (KANAGARAJ; MUNEESWARAN, 2020). One of the most essential properties of wavelet transforms is their ability to capture information in both the time and frequency domains. Because of this property, these transforms are becoming increasingly critical in processing biomedical signals, where high precision and high optimization values are required (SEIDEL et al., 2021). Among the many medical signals, wavelet transforms have been widely used for analysis, noise removal, and compression of respiratory signals (ROSA et al., 2021).

Wavelet transforms are divided into families. There are many families of wavelet transforms that are used in different fields. Figure 4, created with the Matlab tool, shows a graphical representation of the waveforms from some of the main families of transforms. These are the families: Haar, Coiflets, Daubechies and Symlets. This figure shows its base waveforms in an amplitude x time scheme.

Figures 4 b) and c) shows the Coiflets and Symlets, that are wavelets developed to have zero moments, both in scale and in the mother wavelet. As for the coiflet, the main indicative feature of coiflet wavelet is to have the highest number of vanishing moments for both scaling and wavelet function for given support width (ROHIMA; BARKAH AK-BAR, 2020). The Symlet wavelet have high symmetry with the waveforms presented by the ECG signal, which makes it more favorable for denoising and compression of this type of signal (VIJAYAKUMARI; DEVI; MATHI, 2016).

In Figure 4 c), we have the Daubechies wavelet, this is one of the most widespread wavelets for signal analysis. Has as main characteristics the compact support, that is,

it completely cancels itself out of a finite interval of time, and is normally not differentiable, implying a function that does not present smoothness (OLIVEIRA, 2007). The wavelets of the Daubechies family also have great similarity to those presented by ECG signals, in order to facilitate their use in compression (BALACHANDRAN; GANESAN; SUMESH, 2014).

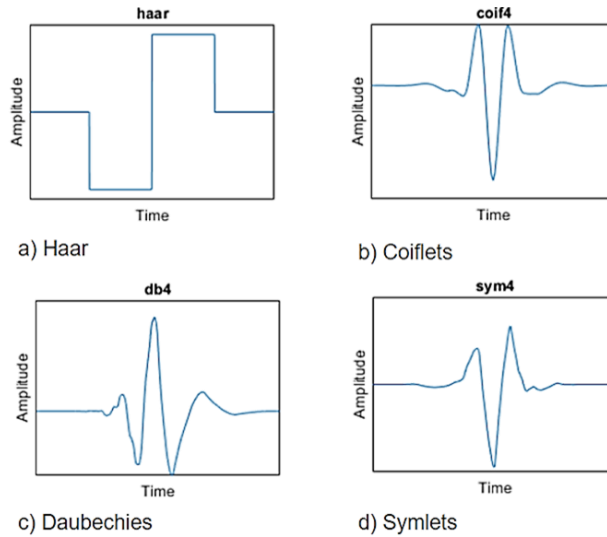


Figure 4 – Representation of wavelet transforms.

Among the transforms presented, the Haar wavelet, Figure 4 a), stands out especially for data compression, due to its great simplicity and computational efficiency (SEIDEL et al., 2020). Although its signal does not have great similarities with the ECG signal, unlike the other wavelets presented, due to its versatility and low complexity, this transform can be used in the compression of ECG signals.

#### 2.2.1.1 Haar wavelet transform

The Haar wavelet was first introduced by Alfred Haar (1909). It is the most straightforward wavelet. Discretely, Haar wavelets are called Haar transform (KANAGARAJ; MUNESWARAN, 2020). In addition to being the simplest transform, the Haar transform is a compactly supported orthogonal wavelet basis function (XIONG et al., 2023). The main advantages presented by the Haar transform are the following:

1. Computational Efficiency: The Haar transform requires minimal mathematical operations, primarily additions and subtractions. This inherent simplicity makes it computationally efficient and fast to implement, especially on resource-constrained hardware (KANAGARAJ; MUNESWARAN, 2020).
2. Potential for High Compression: The Haar transform's ability to decompose signals into localized averages and differences allows for efficient data compression. By discarding less significant details, the transform can achieve high compression

ratios while maintaining a reasonable level of signal fidelity, potentially leading to higher PSNR (Peak Signal-to-Noise Ratio) values (XIONG et al., 2023).

3. Scalable Decomposition: The Haar transform exhibits a recursive property. This means the decomposition process can be applied iteratively, progressively extracting finer details from the signal. This recursive nature allows for a flexible trade-off between compression ratio and signal fidelity (KANAGARAJ; MUNESWARAN, 2020).

The Haar wavelet, can be represented according to Equation (5).

$$\varphi(t) = \begin{cases} 1, 0 \leq t \leq \frac{1}{2} \\ -1, \frac{1}{2} \leq t \leq 1 \\ 0, \text{others.} \end{cases} \quad (5)$$

The Haar wavelet transform process can be visualized from Mallat's Algorithm as a cascade filter-bank structure seen in Figure 5. In this Algorithm, the signal decomposition is performed using high-pass and low-pass filters, which amounts to performing difference and average calculations, respectively, on adjacent data points. From this, the high-pass filter coefficients (cDx) offer a view of the changes occurring in the data, and the low-pass filter coefficients (cAx) give an approximation view (DING; ZILIC, 2016).

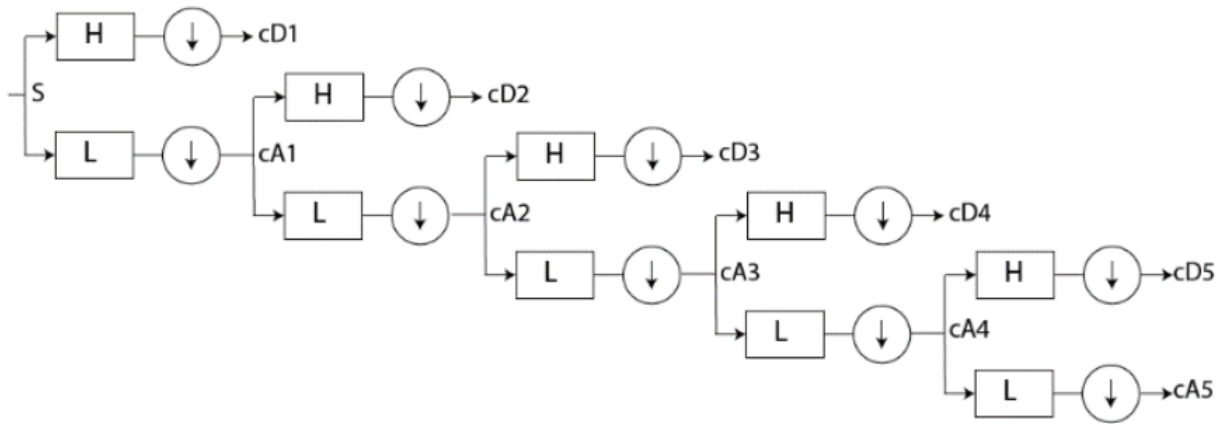


Figure 5 – Five levels cascade filter-bank structure for Haar wavelet transform (DING; ZILIC, 2016).

Thanks to the mathematical simplicity of the Haar wavelet, the Haar wavelet transform is particularly efficient in compressing signals and reducing energy loss (ROSA et al., 2021). In addition to significant energy savings and computational simplicity, the Haar wavelet exhibited superior performance when compressing ECG signal data, compared to other transforms (GODINHO et al., 2023). For this reason, the Haar transform has been widely used for data compression, and it is especially suitable to detect

the ECG R-peaks in low-power embedded systems with a very low power dissipation and general system cost (SEIDEL et al., 2021).

Many relevant works in the recent literature use wavelet transform for data compression, such as the work developed by (ROSA et al., 2021), which uses the Haar wavelet to compress respiratory signals in a resource-constrained environment. The work (XIONG et al., 2023) brings us an application of the Haar transform to compose a Non-intrusive load monitoring (NILM) to optimize energy consumption monitoring. With (KANAGARAJ; MUNEESSWARAN, 2020), the authors tried and achieved good results when applying the Haar transform for image compression in environments with significant hardware restrictions. However, the use case of the Haar wavelet transform that has gained greater prominence and highly positive results has been its application for compression and treatment of ECG signals in constrained systems. The works (GODINHO et al., 2023), (SEIDEL et al., 2020), (DING; ZILIC, 2016) and (SEIDEL et al., 2021) use the transform for this purpose. In this way, using this transformation to optimize ECG data compression is feasible, highly efficient, and competitive, as the presented solutions achieve equivalent/superior results compared to other techniques in the literature.

### **2.2.2 Wavelet transform overview**

Discrete Wavelet transform (DWT) is widely used in the literature in several applications, such as image and signal processing, biomedical, and cryptography. Over the years, several works have focused on proposing efficient hardware solutions for this transform. In the work (JANA et al., 2021), an area-efficient VLSI architecture for 2-dimensional discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT) have been proposed to reduce the hardware space occupation and energy consumption. In (RANA; HASAN; SINHA SHUVA, 2022), discrete wavelet transform and discrete cosine transform were used to watermark images and protect them from copyright. The work presented by (ELHOSENY et al., 2018) shows a model developed through integrating either 2-D discrete wavelet transform 1 level (2D-DWT-1L) or 2-D discrete wavelet transform 2 level (2D-DWT-2L) steganography technique with a proposed hybrid encryption scheme. This scheme aims to ensure an exchange of signals between medical devices connected in IoT networks, with high efficiency and security. The paper (YU; HOU; LI, 2018) utilizes DWT to operate simultaneously in both the time and frequency domains and with advanced Deep Neural Networks techniques to develop a false data injection attack detector capable of capturing spatial and temporal inconsistencies generated by the attacks.

The work (DING; LIU; YANG, 2021) utilizes a lifting wavelet transform to reduce the redundancy in the analyzed periodic wavelet data and uses the Deflate algorithm to compress the waveform data losslessly. The results presented a compression ratio

close to 4 in a real-time test environment. In (CONTRERAS-VALDES et al., 2019), a data compression algorithm based on the DWT is implemented to compress the information generated by electrical consumption monitors. With the work (KUMARI; RAJALAKSHMI, 2016), it is possible to observe an application of DWT to compress ECG signals to facilitate communication. The work in question achieves a compression ratio of 9.34:1, a result superior to other works in the literature. To carry out the tests, the database MIT-BIH arrhythmia database was utilized. The work presented by (IFTODE; FOSALAU, 2020) builds a system to collect, compress, and denoise ECG signals using an embedded system to collect and transmit the signal in real-time. They compared results with three transforms regarding compression performance, with the best performance presented by a Biorthogonal wavelet with three decomposition levels.

Among the various existing wavelet families, the Haar wavelet stands out, being widely used in a series of works presenting highly favorable results regarding energy savings and reducing hardware space. As far as this transform is concerned, one can cite a whole series of works with the most diverse applications that make use of this technique, such as presented by (KANAGARAJ; MUNEESWARAN, 2020), (applied to image compression), (DING; ZILIC, 2016) (sensors for collecting medical signals), (SEIDEL et al., 2021) and (DA ROSA et al., 2022a) (reduce general cost via signal compression using truncation and pruning).

The application of wavelet transformations has dramatically increased in signal processing. The discrete wavelet transform (DWT) has vast multimedia signal processing and numerical analysis applications. In the past few decades, various researchers extensively researched to develop suitable DWT architectures (JANA et al., 2021). The wavelet transform is widely used in feature extraction in fault diagnosis tasks as a mathematical tool to transform time series to another feature space (SHAO et al., 2019). Given its vast application in several areas, the low-power hardware solutions for DWT have been proposed in the literature, as in (GUPTHA; RAO; SELVARA, 2023) that offers a low-power DWT architecture for a pervasive biomedical image processing application using a multiplier-less structure. This dissertation also proposes a low-power DWT solution for ECG signal compression, particularly exploring its Haar family. Beyond the multiplier-less approach, the work also explores pruning and truncation based on approximate computing capabilities.

## **2.3 AxC: Approximate Computing**

AxC leverages the intrinsic error resilience of applications to inaccuracy in their inner calculations to achieve a required trade-off between efficiency, performance, and power demand and acceptable error of returned results. In particular, approximate re-

sults are hard to distinguish from perfect results for signal, audio, image, video processing, data mining, and information retrieval. Suitable solutions come from approximate arithmetic operators, implemented both at the hardware (Hegde; Shanbhag (1999); Palem et al. (2009); Mohapatra et al. (2011); Gupta et al. (2011); Kulkarni; Gupta; Ercegovic (2011); Chippa et al. (2010, 2011); Dutt et al. (2017); Kundi et al. (2020); Rosa et al. (2023); Da rosa et al. (2022)) and software (Chippa et al. (2011); Esmailzadeh et al. (2012); Cong; Gururaj (2011)) levels. The AxC paradigm emerged as a key alternative for trading off accuracy and energy efficiency. Error-tolerant applications, such as multimedia and signal processing, can process the information with lower-than-standard accuracy at the circuit level while still fulfilling a good and acceptable service quality at the application level. The automatic compression of the ECG signal is the essential step preceding ECG processing and analysis. The DHWT is a low-complexity pre-processing filter suitable for compression of the ECG signal in embedded systems like wearable devices, which are incredibly energy-constrained. This work presents an approximate DHWT hardware architecture for ECG processing at very high energy efficiency.

### **2.3.1 AxC Background**

The optimization of hardware architectures for the Haar Discrete Wavelet Transform (HDWT) has been an area of intensive research aiming to enhance energy efficiency and area utilization in integrated circuits without compromising the accuracy of signal processing operations, particularly in detecting R-peaks in electrocardiogram (ECG) signals. Among the most prominent techniques used to achieve this optimization are pruning, truncation, and approximation.

The pruning technique reduces the number of arithmetic operations required in a DHWT architecture by eliminating redundant calculations that do not directly contribute to the ultimate goal, such as the compression of ECG signals. It is achieved by removing unnecessary operations and logical paths, retaining only the essential elements for accurate compression of ECG signals. Pruning the architecture reduces arithmetic complexity, saving energy and area without compromising signal processing quality.

The truncation technique aims to reduce the precision of arithmetic operations, especially those involving multiplication and addition, by decreasing the number of bits used to represent the data. By truncating the input and output values of the DHWT, it is possible to reduce the complexity of calculations, energy consumption, and area occupied by the hardware. However, ensuring that truncation does not compromise signal processing quality is crucial, especially in the accurate compression of ECG signals.

The approximation technique seeks to simplify the matrix operations of the DHWT by modifying the coefficients of the transformation matrices to values that can be more efficiently represented in hardware. It may involve rounding coefficients and simplifying

operations, such as replacing constant values with simpler ones. By performing approximations on the transformation matrices of the DHWT, it is possible to reduce the complexity of arithmetic operations, saving energy and area in the hardware without compromising the accuracy of signal processing.

The mentioned strategies will be detailed in the next section when we show the developed architectures employing these prominent approaches.

### 2.3.2 AxC Overview

Approximate computing (AxC) is an emerging trend in digital design (ESPOSITO et al., 2018). The recent literature has addressed this topic as the work developed by (ESPOSITO et al., 2018) proposes novel approximate compressors and algorithms to exploit them to design efficient approximate multipliers. The work aims to develop circuits with better power and speed for target precision. In (ZANANDREA; MEINHARDT, 2022), the main goal is to evaluate multiplier circuits to explore approximate computing approaches for power-efficient scenarios, with its main target on machine learning and IoT applications. The exploration of the Kalman Gain (KG) equation is the focus and way of implementing hardware approximation in the work (PEREIRA et al., 2022). It explores approximate arithmetic units to reduce power dissipation and area in the VLSI design of the KG block, maintaining an appropriate precision level for the target application.

Many works use approximate computing techniques applied directly to medical signals. It is the case with the work (NAJAFI et al., 2023), despite the difficulties in using approximate computing in critical health monitoring applications. The work shows that by employing approximate computing properly, even an increase in the final accuracy is attainable, thanks to the regularization that approximate processing introduces. The work presented by (SATO; HAMA, 2023) proposes an error correction technique utilizing the Carry-Maskable Adder (CMA) that aims to enable approximate adders in scenarios where high precision is required. The developed algorithm acts on ECG signal processing so that approximate components become viable in processing this signal.

The work developed by (P. et al., 2023) aims to approximate an energy-efficient HDWT hardware architecture for ECG processing through a clock gating mechanism and pruning techniques. With (KANANI; BHATTACHARJYA; BANERJEE, 2021), the authors focused on developing a method that enables approximate computing for efficient biomedical wearable computing at the edge. The project's main objective is to approximate the adders used in ECG signal compression to reduce the power consumption and the chip area utilization. In (DA ROSA et al., 2022a), the approximate level-3 Haar wavelet transform is utilized for ECG compression to increase the hardware design efficiency and obtain a higher R-peak detection rate, which translates into



a more efficient and reliable system, capable of being used in critical medical systems and continuous monitoring. One of the most used forms of AxC is its use directly on data compression. We will detail this topic in the subsection presented below.

### **2.3.3 AxC: Approximate computing in compression data**

Approximate computing (AxC) in data compression emerged as an appealing alternative for building smaller energy-efficient circuits. AxC tries to trade tolerable hardware errors for higher circuit design efficiency in energy consumption, processing speed, and circuit area (PEREIRA et al., 2022). AxC is a prominent approach in digital systems design by relaxing the requirement of exact computation to substantially improve power, performance, and area (ZANANDREA; MEINHARDT, 2022). AxC can be used successfully in various fields, even in areas where the susceptibility to error is low and high precision is required. The primary examples of uses for compression are multimedia processing, data mining and recognition, machine learning (ESPOSITO et al., 2018), and the processing of linear signals.

AxC is being addressed at different levels of abstraction, such as physical properties of the transistors, logical functions of arithmetic operators, and variations in the architecture of computer systems, having obtained area reductions and energy savings for the implemented circuits (ZANANDREA; MEINHARDT, 2022). One of the most important applications of approximate computing to optimize hardware usage is in arithmetic operations. Notably, using approximate arithmetic units is among the main successful strategies for low-power design VLSI circuits (PEREIRA et al., 2022).

Multipliers are one of the most widely used arithmetic blocks in digital circuits. Also, multipliers have high power consumption (SEIDEL et al., 2021). Multipliers are fundamental subsystems for microprocessors, digital signal processors, and embedded systems with applications ranging from filtering to convolutional neural networks. Unfortunately, multipliers comprise complex logic designs and constitute one of the most energy-hungry digital blocks. Therefore, approximate multiplier design has become an important research subject in recent years (ESPOSITO et al., 2018) (BARAATI et al., 2022).

Recently, approximate transforms have been emerging as an alternative to reduce the complexity of its hardware. However, most of the literature's transform approximations have been proposed for multimedia processing, needing more ECG processing solutions (ZANANDREA; MEINHARDT, 2022). Currently in the literature, to the best of the author's knowledge, the works that obtained the viable results are only works developed by (SEIDEL et al., 2021) and (ROSA et al., 2021), which focus directly on the use of approximate computing, aiming at compressing ECG signals.

When applying AxC in VLSI design, the designer must use adequate metrics to evaluate the robustness of the obtained results. Beyond the proper metrics, this work

uses a rigorous synthesis methodology to show the benefits of our solutions and to provide a fair comparison with the state-of-the-art. Next, we present the methodology adopted in this work, showing the database used and the primary metrics for accuracy evaluation.

## 2.4 Methodology

The methodology used to develop the work is divided into the utilized data, synthesis, accuracy, and error measurement methods. This section discusses the utilized data first and then the metrics for accuracy evaluation.

### 2.4.1 Utilized database

The signals used for testing were taken from the Physio-Net website, more specifically from the MIT-BIH Arrhythmia Database (IFTODE; FOSALAU, 2020). The MIT-BIH is a database created at Boston's Beth Israel Hospital with the support of MIT. The database started as a set of standard test materials for evaluating arrhythmia detectors and has been used for basic research into cardiac dynamics at more than 500 sites worldwide. This database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979 (GOLDBERGER et al., 2000).

The database comprises 23 recordings from a pool of 4000 24-hour ambulatory ECG recordings gathered from a diverse group of inpatients and outpatients, randomly selected at Beth Israel Hospital in Boston. The remaining 25 recordings were selected from the same set to include less common but clinically significant arrhythmias that would not be well-represented in a small random sample. The recordings were digitized at 360 samples per second (360 Hz) per channel with 11-bit resolution over a 10 mV range (GOLDBERGER et al., 2000).

Two concatenations with different types of signals were selected. One contains only signals of the Normal Sinus Rhythm (NSR) type, which consists of normal ECG signals, i.e., without anomalies caused by diseases or other factors. The other concatenation of signals chosen is composed of signals with Premature Ventricular Contractions (PVCs). PVCs represent anomalous conditions where extra beats begin to occur in one of the two lower pumping chambers of the heart, disrupting the correct beating rhythm. Signals with this condition have very different waveforms, making their compression difficult and ideal signals to test in the proposed compression model.

Figure 23 shows the signals utilized in the tests of the solution. Figure 23 a) presents the graphical representation of the complete NSR signal, and b) the complete representation of the PVC signal, c) and d) show parts of the signals NSR and PVC respectively with 2,000 samples, and e) and f) show parts of both NSR and ECG sig-

nals with 5,000 samples, in the same way. These last four examples aim to facilitate visualization of the waveform presented by the signals used.

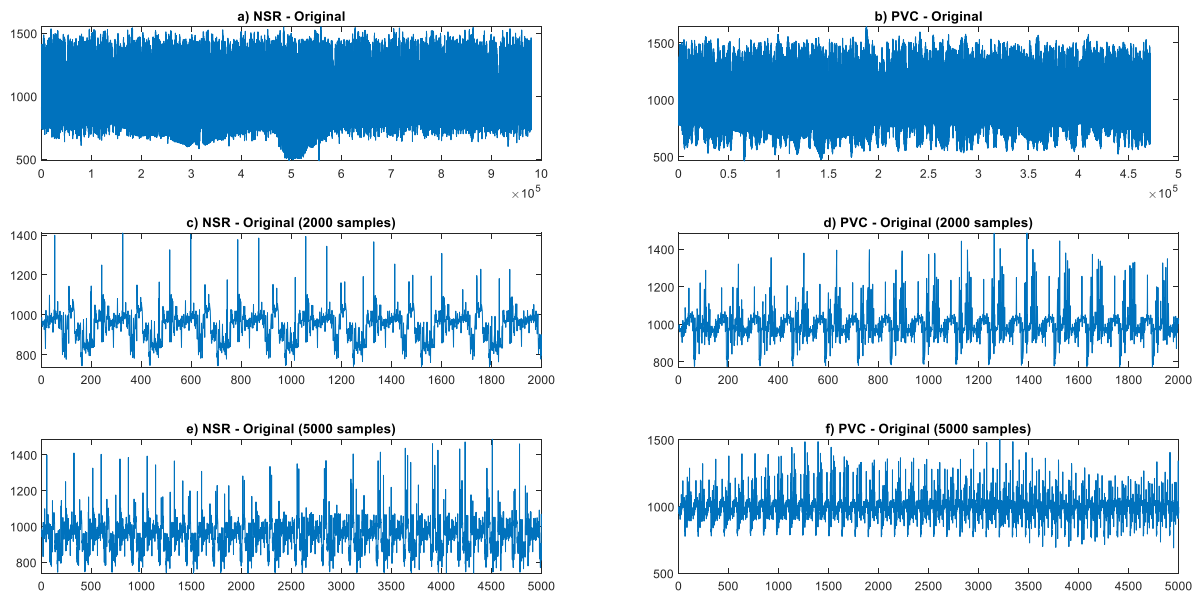


Figure 6 – Graphical representation of the signal concatenations.

#### 2.4.2 Metrics for Accuracy Evaluation

Applications in the medical field require a high level of precision. Small changes in data can significantly impact obtaining accurate diagnoses (CHIANG et al., 2019) and thus cause failures with severe consequences, compromising the integrity of the healthcare system. Therefore, we used several error analysis metrics to evaluate the compression to ensure that the minimal changes did not affect the original signal in a way that could compromise the diagnosis. Based on the multiple works in the literature analyzed, we applied the following error metrics:

1. Peak detection
2. Accuracy
3. Mean Absolute Error (MAE)
4. Normalized Cross-correlation (NC)
5. Percentage Root-mean-square difference (PRD)
6. Root-mean-square Deviation (RMSE)
7. Signal-to-noise ratio (SNR)
8. Structural similarity index measure (SSIM)

### 2.4.2.1 Peak detection

Peak detection is one of the main ways to ensure that signal characteristics are extracted robustly, guaranteeing signal analysis with the required high precision (IFTODE; FOSALAU, 2020). Correct detection of the QRS complex is essential for an accurate medical diagnosis (HAMZA; RIJAB; HUSSIEN, 2021). In this way, we have to automatically detect the R-peaks in the ECG signal, which is the essential step preceding ECG processing and analysis in industrial-strength biomedical applications (SEIDEL et al., 2021). Figure 7 shows the general waveform of the ECG signal with its main components and their expected ranges.

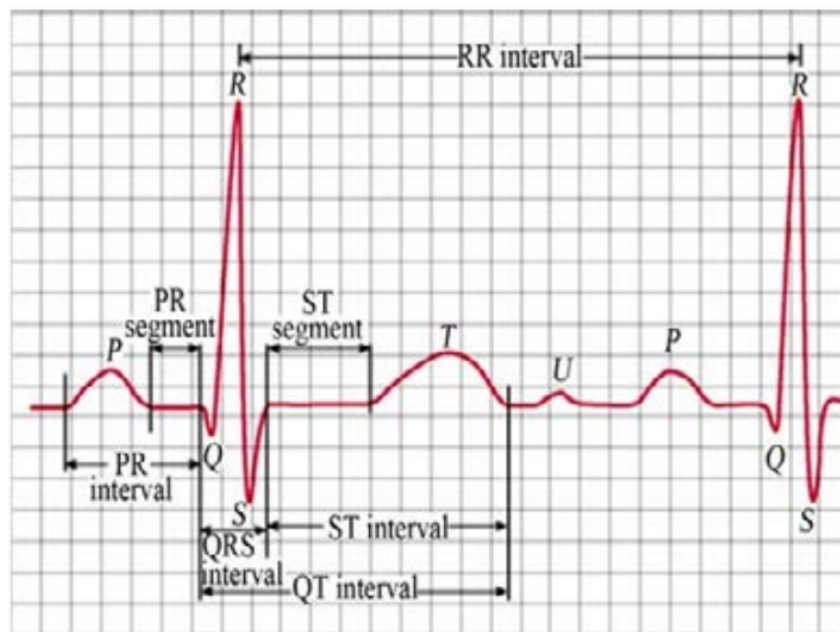


Figure 7 – Complete visualization of the QRS complex with intervals (JINDAL et al., 2018)

The QRS complex is the most striking waveform, and it serves as the basis for the automated determination of other ECG characteristics (WU et al., 2020). Many algorithms are used to detect the QRS complex correctly. The Pan-Tompkins (P&Q) algorithm has been identified as the most straightforward and efficient algorithm for embedded devices due to its low computational cost and competitive performance. It allows reliable peak detection while presenting lower energy consumption than other existing solutions (CLARK et al., 2018). This algorithm can be considered straightforward, since its implementation was carried out individually using a single module to complete the peak detection process and generation of results.

The algorithm was offered by Jiapu Pan and Willis J. Tompkins in 1985. QRS complex is detected by the digital study of amplitude, width, and slope of ECG signal, which includes three stages: Digital filtering linearly, Non-linear conversion, and algorithm for decision making. The linear process constitutes a bandpass filter, moving window integrator, and derivative. Non-linear conversions are performed by squaring the am-

plitude of the signal. Adaptive thresholding and discrimination of the T-wave technique are parts of the decision rule algorithm. Detection sensitivity escalates due to low thresholding after filtering (JINDAL et al., 2018). The block scheme of Pan-Tompkins can be observed in Figure 8.

In this case, the bandpass filter allows only frequencies in a particular range where there is a more significant possibility of locating peaks, given the behavior presented by the signal while attenuating frequencies outside it. In turn, the derivative filter computes the derivatives of the post-integrated data. The main function of this filter is to directly attenuate the noise generated by previous operations and the original signal. The derivation process performs this reduction.

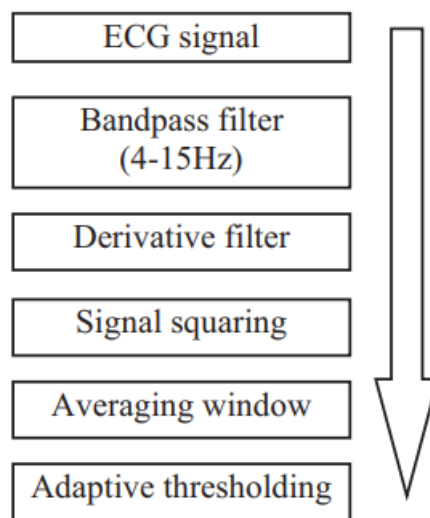


Figure 8 – Block scheme of Pan-Tompkins based detector (GALA; BARABAS; KRAJNAK, 2020)

The work developed by (COSTA, 2021) describes the Pan-Tompkins algorithm as one of the main algorithms for detecting R peaks. Figure 9 shows the signs at the end of each one of the phases of the algorithm. First, the signal passes through a digital bandpass filter to attenuate the noise. The desirable band to maximize the energy of the QRS complex is 5 to 12 Hz (Figure 9 b). After the filtering phase, the five-point differentiation is executed (Figure 9 c), exposing information about the QRS complex's slope. After that, the sign is squared (Figure 9 d). This operation intensifies the slope of the response curve in derivative frequency. It helps restrict the false positives created by the T waves with more energy than the usual waves. A movable window integrator, with a width of 15ms, is used to produce a signal that includes information about both the slope and the width of the QRS complex (Figure 9 e). To detect the R peak, a system of adaptive thresholds is used in the output signals of the bandpass filter and the moving window integrator (Figure 9 e). The window integrator identifies the local maximum of the R peak, and the exact value of peak R is determined at the output of the bandpass filter.

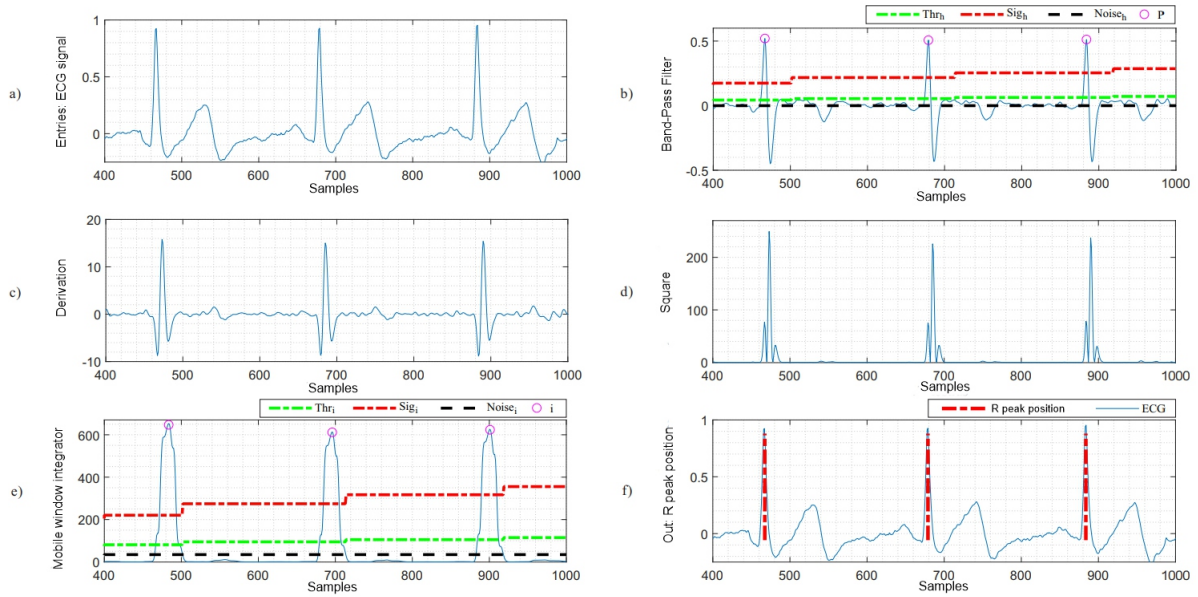


Figure 9 – Algorithm processes of Pan-Tompkins: a) Signal entry, b) Signal resulting from adaptive filter and adaptive thresholds, c) Resulting signal from derivation, d) Squared signal, e) Result of the signal at the output of the window integrator and the adaptive thresholds of the signal, f) Out of the R peak detection algorithm (COSTA, 2021)

#### 2.4.2.2 Accuracy

The direct calculation of accuracy is one of the most traditional methods to measure the correctness of a signal compared to its original after it has changed, be it through a compression process, the insertion of noise, or any other change to the original signal.

The accuracy indicator is the ratio between the number of correctly classified samples and all the test samples. Its mathematical expression is defined in (6), created by (BUY YA et al., 2023), where TP and TN stand for true positive and true negative, respectively, and are responsible for the correct detection of the signal. FP and FN stand out for false positive and false negative, respectively, being responsible for the incorrect classification of the signal (BUY YA et al., 2023).

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (6)$$

The definition presented above is the traditional definition of accuracy calculation commonly used in classifying AI models. In the experiments, we adapted a version of the algorithm for 2D ECG signals, which standardizes hits and misses.

#### 2.4.2.3 MAE: Mean Absolute Error

The average absolute error is the predictive value of the average absolute value of deviation from the true value. Since the deviation is absolute, the positive and negative offset will not be present. Therefore, the average absolute error is not sensitive to the abnormal value and reflects the actual situation of the error of the prediction value.

Equation (7) shows the calculation method of mean absolute error, in which  $n$  represents the number of samples,  $Y_i$  represents the value of the correct compression, and  $M$  represents the compressed value (ZHANG et al., 2019).

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - m|) \quad (7)$$

The smaller the values of MAE, the better the overall quality of the compression, and the closer the generated signal is to the original one (XIONG et al., 2023). This way, we obtain the lowest possible MAE values to guarantee more remarkable similarity with the original signal.

#### 2.4.2.4 NCC: Normalized Cross-correlation

Analysis of electrocardiogram using cross-correlation is a system that takes advantage of a feature that the ECG signal of a person under similar physical conditions at different timings shows very minute variations in its anatomy. Thus, this concept can be used for human authentication and identification purposes (VERULKAR; AMBALKAR, 2016). The Normalized Cross-correlation is given by Equation (8) presented by (CHANDRA'; SHARMA; SINGH, 2021), where  $y(n)$  is the mean value of the reconstructed signal. It is the arithmetical quantity of the degree to which deviations to the value of one variable predict variation to the value of another (CHANDRA'; SHARMA; SINGH, 2021). The NC is defined so that the closer its value is to 1, the greater the signal similarity. The minimum acceptable correlation coefficient by the literature is 0,978 for data from human subjects (FRASER et al., 2013). .

$$y_{xy} = \frac{\sum_{n=1}^N (x(n) - \bar{x}(n))(y(n) - \bar{y}(n))}{\sqrt{\sum_{n=1}^N (x(n) - \bar{x}(n))^2 (y(n) - \bar{y}(n))^2}} \quad (8)$$

#### 2.4.2.5 PRD: Percentage Root-mean-square difference

PRD (Percent of Root-Mean-Square Difference) is the ratio of the signal power to the original signal power. The PRD measures distortion. It also gives an average distortion in the reconstructed signal (ABDULBAQI et al., 2018). Equation (9) represents the PRD, presented by (FATHI et al., 2022), where  $s(x)$  is the original signal, and  $S(x)$  is the reconstructed signal.

$$PRD = \sqrt{\frac{\sum (s(x) - S(x))^2}{\sum s(x)^2}} \times 100 \quad (9)$$

A lower PRD represents a better quality of the reconstructed signal (CHIANG et al., 2019), and a PRD of zero denotes a lossless compression (BURGUERA, 2019). According to the literature, an acceptable PRD for medical applications would be less than

5% since, under that value, the signal has a good reconstruction quality (SINGH PAL et al., 2022). The distortion level measured by the PRD is one of the most used metrics when evaluating distortions applied to an ECG signal, being used by a wide range of works in the literature.

#### 2.4.2.6 RMSE: Root-mean-square Deviation

The RMSE determines the variance between the output predicted by the model and the actual output. Therefore, a small value of RMSE corresponds to a more minor difference and better performance (CHIANG et al., 2019). Equation (10) defines RMSE.

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{n=1}^N (x_i - \hat{x}_i)^2} \quad (10)$$

The RMSE approach is one of the most accurate performance metrics to estimate the system's error rate (PATIL; PAWAR, 2022).

#### 2.4.2.7 SNR: Signal-to-noise ratio

The Signal-to-noise ratio(SNR) represents the dimensionless ratio of the signal power to the noise power. Thus, SNR quantifies the amount of noise present in the signal concerning the signal of interest(DAS; CHAKRABORTY, 2017). SNR measures the amount of noise energy added into a signal due to compression and decompression procedures (FATHI et al., 2022). Equation (11) presents the SNR representation.

$$SNR = 10 \times \log \left( \frac{\sum (s(x) - s(\bar{x}))^2}{\sum (s(x) - S(x))^2} \right) \quad (11)$$

The greater the SNR is, the better the performance of the performed compression, generating a signal with lower error compared to the original signal and more resemblance to clean signals (CHIANG et al., 2019).

#### 2.4.2.8 SSIM: Structural similarity index measure

The structural similarity measure (SSIM) has been widely used in the image processing field and demonstrated to be superior to other distance metrics, including straightforward Euclidian metric. This metric distinguishes between the structures of the images and obtains the similarities between each pair of the predetermined moving local windows between two images. The total similarity is calculated based on the average of all the SSIM values over all the local windows in the entire image (SHAH-RIARI et al., 2018). This way, the metric distinguishes between images corresponding to ECG signals and compares the original and compressed signals.



SSIM is a technique for evaluating the quality and structure of two images to determine their similarity. SSIM measures the quality of image degradation with structural detail (OKTIVASARI et al., 2022). It incorporates images luminance, contrast, and structure when quantifying the distance between them. Given two images  $X$  and  $Y$ , where  $x_i$  and  $y_i$  are the  $i$ th local rectangular window, and  $M$  is the total number of windows in an image (SHAHRIARI et al., 2018), the SSIM between every two local windows is given by Equation (12).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12)$$

Analyzing the SSIM, we have that an SSIM value closer to one means that the generated image is more similar to the original one. In contrast, a value closer to zero means that the generated image is more different from the original. Because the ECG signals contain semantic information, it is essential to retain their structure. Therefore, the performance of a generated ECG image in retaining its structure is measured concerning the SSIM value (LEE et al., 2020).

## 2.5 Chapter conclusion

This chapter introduced all the main concepts applied in the project's development. At first, it briefly discusses the power consumption concepts and energy consumption of portable devices and devices with limited computing resources, reinforcing the importance of achieving the most significant possible savings to ensure the viability of the proposed solution. The next section carried out an overview of Wavelet transforms and their application in compression and optimization, with particular attention to the Haar transform, as it is used as the basis for the development of the proposed ECG data compression. The following section offered a contextualization regarding approximate computing, especially when referring to data compression. Then, the final section described the methodology used in the work, highlighting the database used in this work and the primary metrics for accuracy evaluation.

### **3 DISCRETE WAVELET TRANSFORM ARCHITECTURE PROPOSALS**

This chapter is divided into two main parts, showing the proposed solutions with their respective details and a general survey of works in the literature aimed at compressing high-performance ECG data. The developed architectures are described in detail, presenting their characteristics and complete descriptions. The chapter also offers details about the usage of approximate computing techniques in the architectures. Finally, a table compares the primary characteristics of the solutions in the literature and the work developed in this dissertation.

#### **3.1 Developed Architectures**

As previously presented, wavelet transforms comprise decomposition levels. As the levels advance, signal compression increases according to the definition of the mother Haar wavelet, represented by Equation (5). At the same time as this increase in compression is obtained, in the same way, as we advance the decomposition levels, we have a worsening signal quality (since we have a progressive reduction in its resolution).

The work developed by (SEIDEL et al., 2021) constructed a level 4 Haar wavelet for ECG data compression. With this architecture, the authors achieved highly positive results when they compared the implemented solution with the state-of-the-art in ECG signal compression. Considering the gains obtained, it opens a gap to explore even more advanced compression levels. Therefore, we chose to reduce the compression level 4, already consolidated in the literature, to study level 5.

There are many ways to construct a Haar wavelet transform. The direct construction of the transformation is not computationally efficient. Therefore, other methods are used to perform this construction. A very efficient form of construction widely used in the literature is the construction using lower-level transformation blocks. This technique connects these blocks in series to create transforms with higher levels of compression, mitigating the exponential increase in the complexity of the developed architecture.

Figure 10 presents two examples of blocks. In Figure 10 a), we can observe an M1 (level 1 Haar wavelet transform) block, while in Figure 10 b), we have an M2 (level 2 Haar wavelet transform) block, both complete according to the literature. Figure 11 presents the M3 (level 3 Haar wavelet transform) block. Due to their low complexity, those are the most used components in the composition of higher-level architectures.

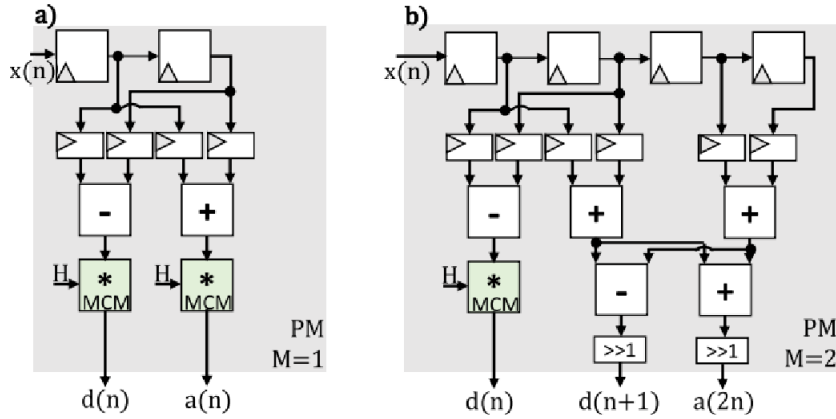


Figure 10 – Blocks M1 - a) and M2 - b) (ROSA et al., 2021).

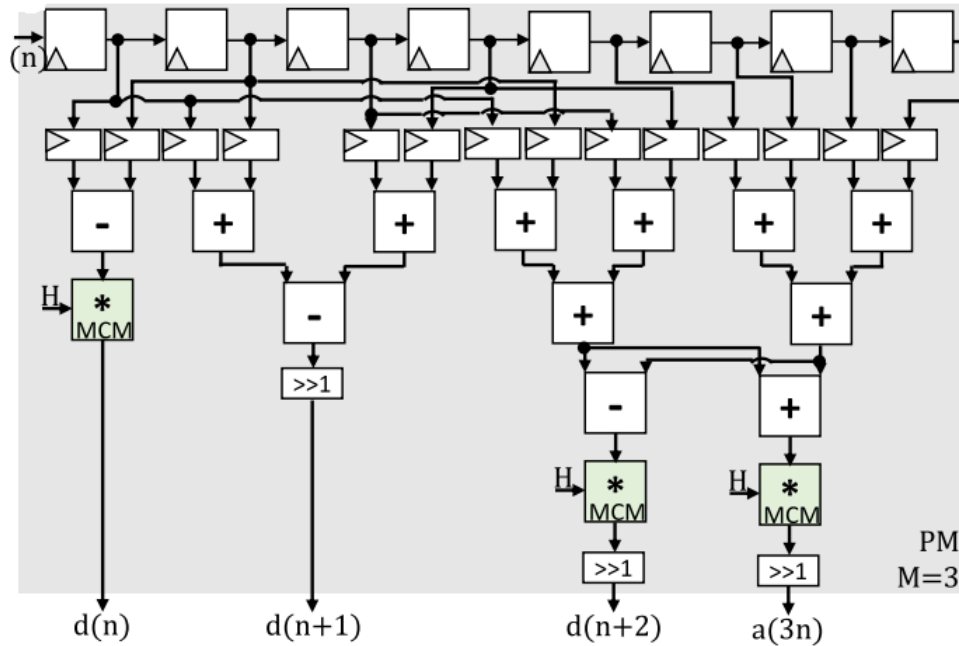


Figure 11 – Block M3 (ROSA et al., 2021).

One of the most common ways to represent wavelet transform blocks is in matrix format. In this way, it is possible to closely observe the coefficients of the blocks that make up the main transformations and to understand mathematically the impact of each change in the architecture, especially the removal of components due to pruning and approximation. Below we have, in matrices H1 (13), H2 (14), and H3 (15), the matrix representation of the three blocks used as a basis for building the models.

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (13)$$

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & \sqrt{-2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & \sqrt{-2} \end{bmatrix}. \quad (14)$$

$$H_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ \sqrt{2} & \sqrt{-2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} & \sqrt{-2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{-2} \end{bmatrix}. \quad (15)$$

Many different configurations can be used to compose a level 5 Haar wavelet transform using lower-level blocks. We performed the steps below to select the most efficient configurations for carrying out the implementation.

1. Initially, we searched recent literature to locate the state-of-the-art regarding using high compression levels with the Haar transform, using the works (K; A; J, 2023; SEIDEL et al., 2020; ROSA et al., 2021) as a basis.
2. Afterwards, several promising architectures were implemented in VHDL, according to literature sources. We implemented the architectures with the block combinations M5, M2M2M1, M3M1M1, 5M1, M2M1M1M1, and M2M3 since they have the higher potential to be the most efficient.
3. These architectures were subjected to the co-simulation process using Matlab, Modelsim, and QuestaSim software
4. The architectures were then tested using signals from the MIT-BIH ECG database (COMPUTATIONAL PHYSIOLOGY, 1999), which contains real ECG signal data for use in various experiments.
5. The results were subjected to the error metrics ACC, NCC, RMSE, SSIM, SNR, and PRD. The tests were realized in the Matlab software.
6. Afterwards, the architectures went through the synthesis process using the Cadence Genus software, and information regarding the critical path, energy consumption, and occupied area was extracted.

7. We compared the most relevant application characteristics (ECG data compression).
8. We selected the best architectures for more complete experimentation, i.e., M2M2M1, 5M1, and M2M3.

Next, we present the three architectures developed in this dissertation. The architectures are called ODHWT (Original Haar Wavelet Transform), PDHWT (Pruned Haar Wavelet Transform), and AxDHWT (Approximate and Pruned Haar Wavelet Transform). We also present the truncation strategy used in the architectures.

### 3.1.1 ODHWT: Original Haar Wavelet Transform

As previously mentioned, the first implementation was the original Haar wavelet transform (ODHWT), according to the literature and in complete form. We implemented four architectures to realize this transformation, three efficiently implemented by connecting lower-level blocks in series to create a higher level. The traditional implementation of the Haar M5 block was only carried out for comparison purposes, as this form of implementation is inefficient regarding both energy consumption and the area occupied. The coefficients referring to the blocks used in the implementations can be observed in the Equations: (13), H2 (14), and H3 (15).

The first architecture implemented comprises five M1 blocks. It can be seen in Figure 12. This architecture consists of ten registers, five subtractors, five adders, and ten multiple constant multiplication blocks. The multiplier block is the most computationally expensive component of the entire architecture (DA ROSA et al., 2022b). This block has the function of multiplying the wavelet signal by the Haar wavelet constant, equal to  $1/\sqrt{2}$ .

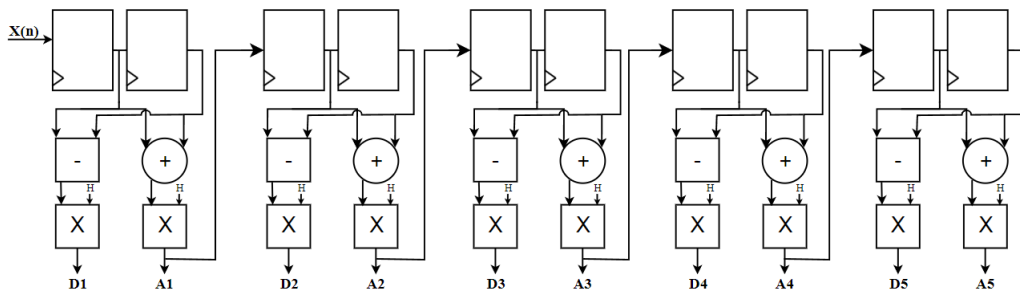


Figure 12 – Haar wavelet transform composed of five M1 blocks.

The construction of multiplier blocks considered approximate computing concepts to optimize the overall performance of the architecture. Therefore, we generated the optimized single constant multiplications (SCM) for the H coefficient using the Hcub algorithm by the Spiral tool (SEIDEL et al., 2021). This implementation only uses the approximated operation of tree adders and four shifters. According to (SEIDEL et al.,

2021), this implementation considerably reduces the architecture cost while maintaining excellent precision when performing multiplication by the wavelet coefficient. The architecture of the SCM block can be seen in Figure 13. This implementation aims to perform better using only the most straightforward component in its composition (M1).

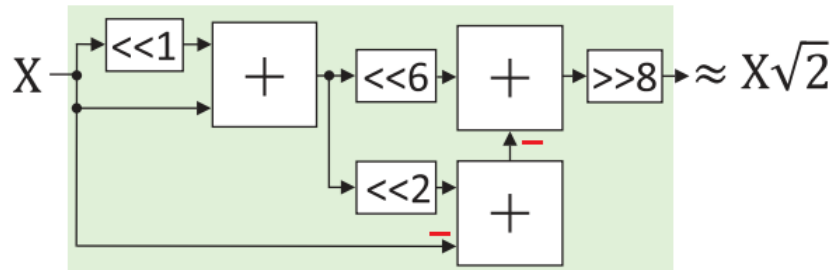


Figure 13 – The SCM block (SEIDEL et al., 2021).

The other two efficiently implemented architectures can be seen in Figures 14 and 15. In Figure 14, it is possible to observe the architecture implementation, which uses two M2 blocks and one M1 block in series. This implementation uses ten registers, nine adders, five subtractors, six multipliers, and four shifts. This implementation seeks a more fundamental approach, aiming to reduce the number of multipliers without using components with a self-level of complexity (M3 or higher).

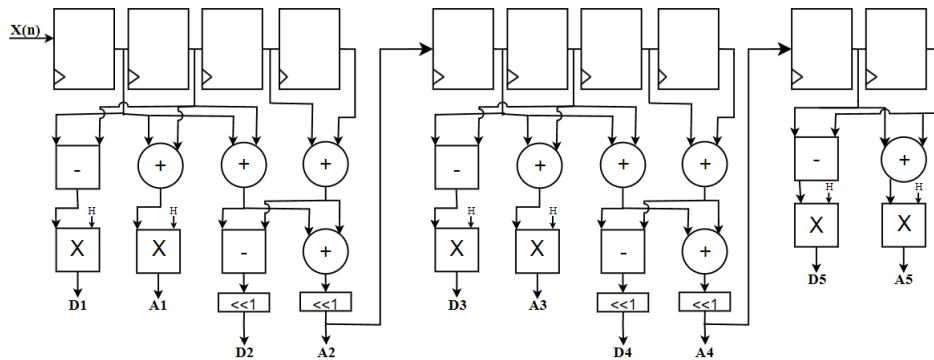


Figure 14 – Haar wavelet transform composed of two M2 and one M1 blocks.

The architecture in Figure 15 uses an M2 block in series with an M3 block. This architecture comprises 12 registers, 15 adders, five subtractors, six multipliers, and six shifts. This implementation uses a more complex block (M3) to mitigate the critical path while maintaining the multiplier components.

### 3.1.2 PDHWT: Pruned Haar Wavelet Transform

The first solution implemented to reduce the computational cost of the architecture in question uses the pruning technique, as we have already seen. This technique removes all unnecessary elements to generate valuable outputs for the desired application. This technique can be applied directly to the Original architecture since compression is

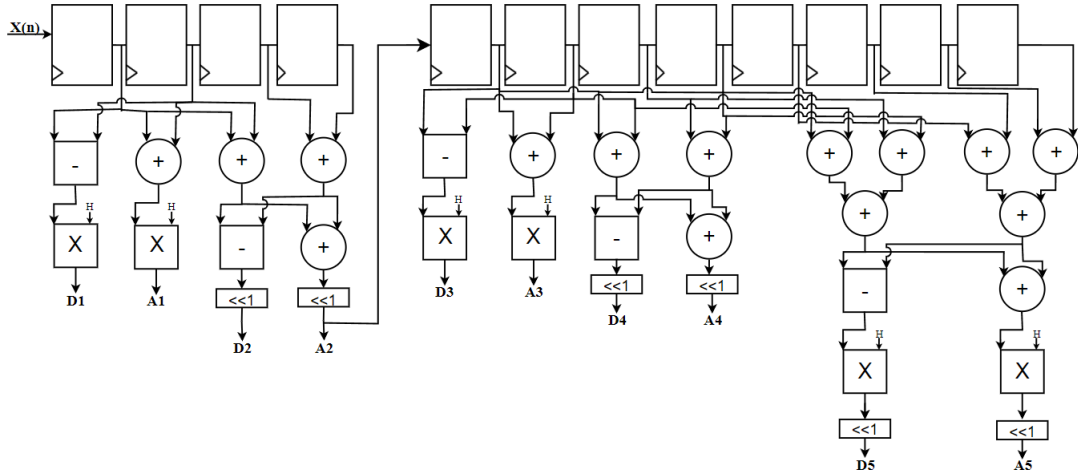


Figure 15 – Haar wavelet transform composed of one M2 and one M3 blocks.

given only by the approximation coefficients. One can eliminate the detail coefficients and all components used exclusively for their generation without affecting the quality of the target signal.

Therefore, removing the detailed components significantly simplifies the coefficients and the developed architectures. The resulting coefficient matrices will be presented below after the pruning process, where  $H_1$  (16),  $H_2$  (17), and  $H_3$  (18) are the matrix representation of the three blocks used as a basis for building the models with pruning.

$$H_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}. \quad (16)$$

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ \sqrt{2} & \sqrt{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (17)$$

$$H_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sqrt{2} & \sqrt{-2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (18)$$

Figure 16 shows the architecture composed of five M1 blocks with prune, Figure 17 shows an architecture consisting of two M2 and one M1 blocks, and Figure 18 finally shows the architecture composed of one M2 and one M3 blocks.

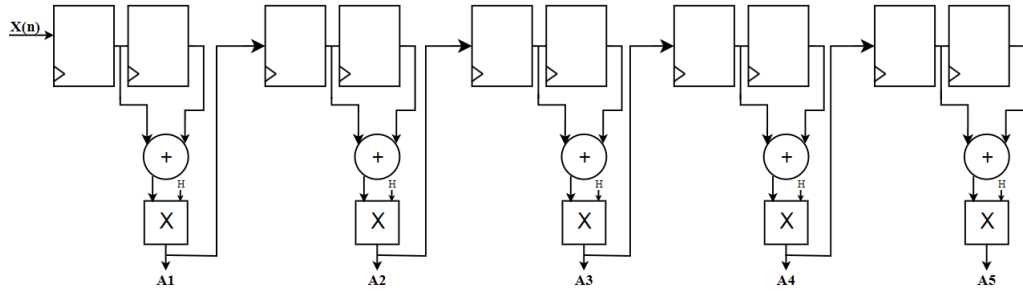


Figure 16 – Pruned Haar wavelet transform composed of five M1 blocks.

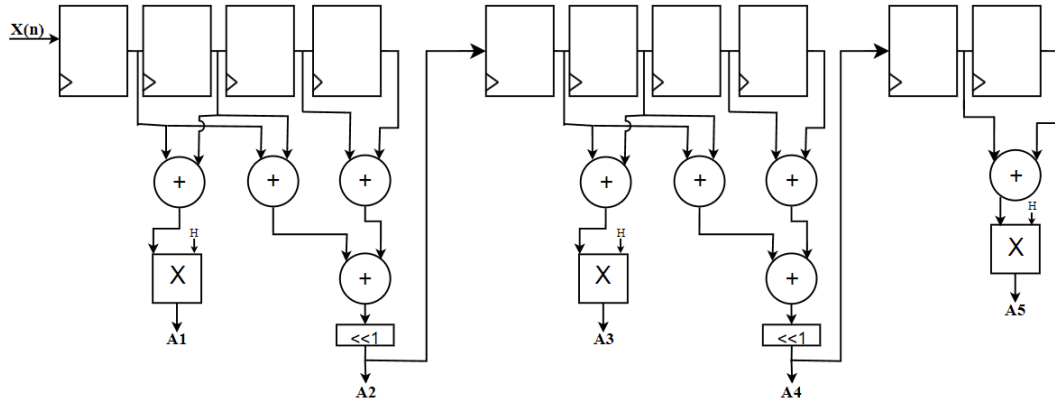


Figure 17 – Pruned Haar wavelet transform composed of two M2 and one M1 blocks.

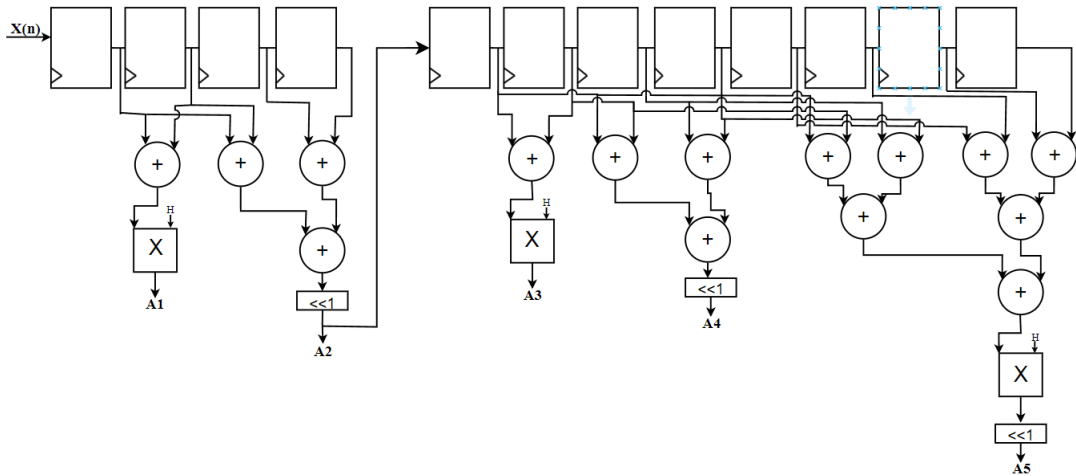


Figure 18 – Pruned Haar wavelet transform composed of one M2 and one M3 blocks.

### 3.1.3 AxDHWT: Approximate and Pruned Haar Wavelet Transform

The pruning strategy significantly reduced the total computational cost without losses. We use hardware approximation techniques as viable solutions to obtain an even more significant reduction. Multipliers are the most expensive components of the architecture and are responsible for performing multiplication by the wavelet constant. They are the main target for optimization. This way, the approach removes the multiplier blocks, which allows a massive gain in performance but, at the same time, negatively impacts the signal generated at the end of the compression process.

After removing the multipliers, the matrix coefficients have their terms  $\sqrt{2}$  and  $\sqrt{-2}$



simplified to 1 and -1 since multiplication by the wavelet constant will no longer be performed. The matrices of the coefficients after the removal of the multipliers are in the matrices: H1 (19), H2 (20), and H3 (21).

$$H_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad (19)$$

$$H_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (20)$$

$$H_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (21)$$

As far as the architectures developed with this approach are concerned, they have a structure very similar to that produced by pruning alone. The only significant difference is the removal of the multiplier circuits. Observing the architectures developed using the approximation technique and pruning in the figures below is possible to identify a structure similar to that presented by pruning, but only without the multiplier blocks. Figure 19 shows the architecture consisting of five M1 blocks, Figure 20 shows the architecture consisting of two M2 blocks and one M1. Finally, Figure 21 shows the architecture consisting of one M2 block and one M3.

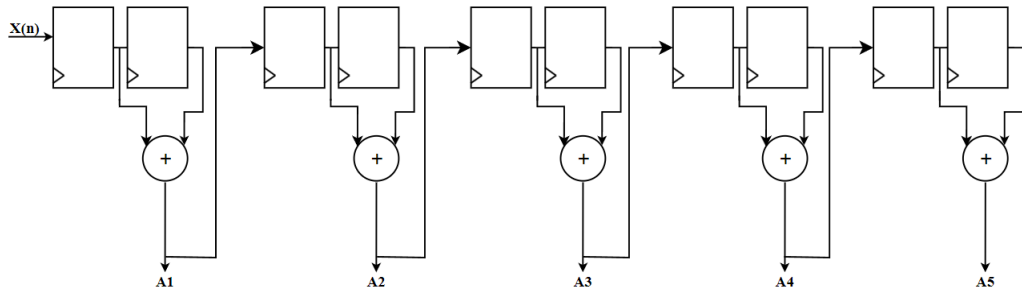


Figure 19 – Approximate and pruned Haar wavelet transform composed of five M1 blocks.

The next subsection discusses the implementation of the third hardware-level optimization technique, truncation.

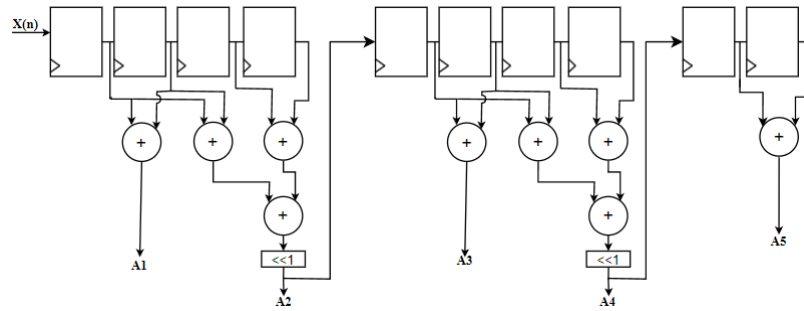


Figure 20 – Approximate and pruned Haar wavelet transform composed of two M2 and one M1 blocks.

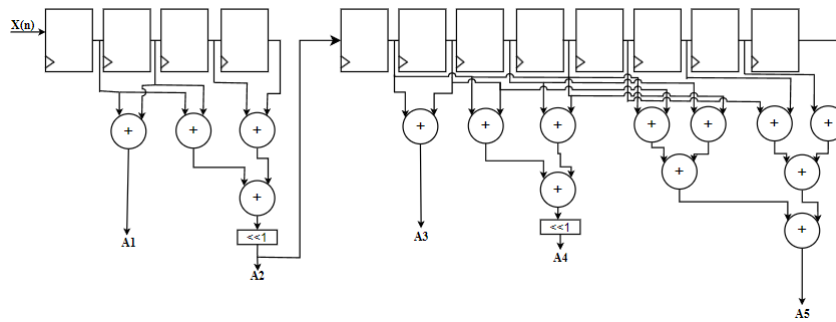


Figure 21 – Approximate and pruned Haar wavelet transform composed of one M2 and one M3 blocks

### 3.1.4 Truncation

We used the truncation technique to achieve an even more significant improvement regarding performance and a reduction in overall cost greater than other works in the literature. There are two ways to implement the truncation technique directly on the hardware. They are the truncation of the input bits, reducing the bits of the input signal, and truncation in the hardware to reduce the number of bits in the structures used for compression, eliminating less significant bits in the process.

We tested both techniques in the present work. By using them, we achieved a considerable reduction in the cost of the system via an increase in operating frequency and a reduction in components used in compression. The techniques are used in the entire circuitry so that designing it to operate with a certain width works appropriately for a specific application, even introducing errors in the magnitude of the processed results (SEIDEL et al., 2021).

The ECG signals extracted from the database are signals originally quantized into 11 bits. In this way, experiments with all architectures were developed using the two techniques previously presented, based on 11-bit architectures for greater precision in signal compression. After extracting the maximum gain by using these techniques, truncation tests began.

Initially, tests were conducted with bit truncation, with the zero truncation type, which means the removal of the less significant bits of the signal as it is considered the best

way to achieve a large gain when it comes to reducing energy consumption by reducing the least significant bits of the signal (DATTA; MOHANTA; MISHRA, 2019). Tests were carried out with different signals and using the three most efficient architectures implemented (architectures with approximation and pruning). The tests reduced the number of input bits, ranging from one to nine. With a reduction of up to five bits, acceptable values were obtained for most of the error metrics used, making signal compression more efficient among the proposed solutions, considering some insignificant losses in signal quality. With more significant reductions in the number of bits, the error becomes critical and directly impacts obtaining the characteristics of the signal, thus making its use unfeasible. The work developed by (LEE; KIM, 2023) uses the bit truncation technique to develop an architecture optimized for stochastic computing to reduce stochastic number generator hardware complexity. This work presents a graphical representation of the bit truncation and its types. That representation can be seen in Figure 22.

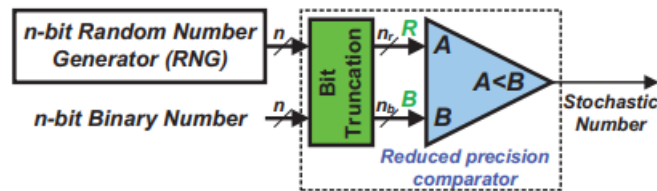


Figure 22 – Block diagram exemplifying the bit truncation (LEE; KIM, 2023)

Figure 23 shows the different types of bit truncation and how it performs the truncation process in each case. In Figure 23 a) we have an example of how the bit truncation process was used in the present work. Since truncation was used, which consists of removing the least significant bits from the input and replacing them with zero. As for the adaptation of the architecture, the inputs have been reduced to accommodate the new quantities that are passed through the input and feed the architecture.

We conducted tests using truncation directly in the hardware to use architectures built using fewer bits than the input. The tests conducted using this technique yielded unsatisfactory results due to the high errors obtained, even with a minimal reduction of 1 bit in the hardware. Consequently, the quality of the final signal was not satisfactory, rendering it impractical to implement in signals that demand high precision, which is the case of the ECG signal. Given the insufficiency of the results, we discarded this truncation model for use in subsequent tests.

Thus, the architectures developed in the tests with truncation are: The Truncated and Pruned Discrete Haar Wavelet Transform (TPDHWT), referring to the architecture using input bit truncation and pruning and Truncated and Approximated Discrete Haar Wavelet Transform (TAXDHWT) which uses not only truncation and pruning but also hardware approximation to generate maximum resource savings.

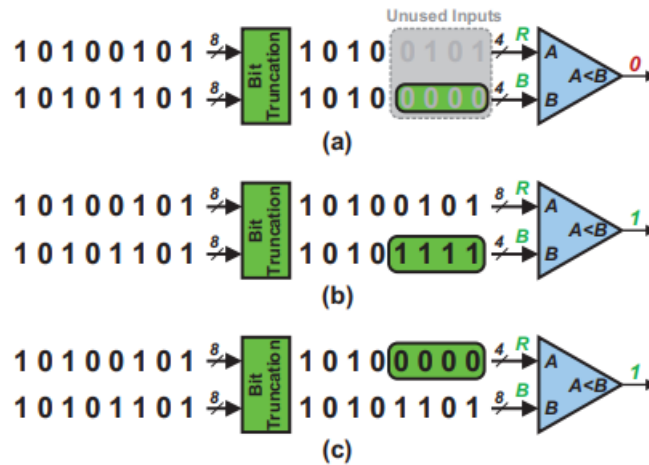


Figure 23 – Block diagrams presenting the bit truncation types: a) input number with zero truncation, b) input number with nonzero truncation, c) random number with zero truncation (LEE; KIM, 2023).

### 3.2 Related Work: Compression Techniques for Hardware Applications

In recent literature, many works have focused on using wavelet transforms to improve performance in general, especially when it comes to optimizing systems with considerable computational constraints and embedded systems (ZHOU et al., 2019; WU et al., 2020; LIU et al., 2019; P. et al., 2023; ROSA et al., 2021). One of the most important approaches used in the literature and real applications is optimization directly on the hardware, as this optimization reflects one of the most critical requirements, energy consumption. They can significantly reduce consumption, often with little or no loss in the quality of the information processed (JANA et al., 2021; SEIDEL et al., 2021).

Considering the abovementioned situation, three main ways to optimize hardware for general applications exist. The first way, and one of the most commonly used ways, is to use approximate computing techniques to reduce complexity or remove more expensive components. This technique takes advantage of the fact that many signals do not need to maintain their perfect shape and accept minor distortions without compromising their functionality. It is possible to achieve significant performance gains in exchange for a controlled loss of quality. In this way, the approach achieves its performance gain by balancing the trade-off between quality vs resource consumption (KANANI; BHATTACHARJYA; BANERJEE, 2021; PEREIRA et al., 2022; ESPOSITO et al., 2018; BARAATI et al., 2022).

The pruning strategy is another form of compression used in some instances to gain performance without losing quality. Pruning consists of removing components, or parts of the hardware, that are unnecessary to generate the desired result. This technique is widely used in neural networks to improve the network and discard incorrect nodes

(XUANHUI; JUAN; QUAN, 2021; ZHU et al., 2020; RATHI; PANDA; ROY, 2019; LIU et al., 2017). Now, it is being applied to hardware and obtaining very positive results, for example, in the case of (NAJAFI et al., 2023) to remove partial products of lesser significance in arrhythmia detection and to remove unnecessary coefficients generated from wavelet transforms like those presented in (DA ROSA et al., 2022a).

The third way of hardware compression is by performing truncation. This technique reduces the number of bits in the signal, thus reducing its computational complexity and facilitating processing. Like the approximation, this technique reduces the number of bits in the signal, thereby reducing the complexity of the calculation and promoting processing. Several examples of work in the literature apply this technique as (ESPOSITO et al., 2018; NAJAFI et al., 2023; SEIDEL et al., 2021).

Some projects use the three techniques simultaneously to extract the maximum possible efficiency from the developed hardware. These works are (SEIDEL et al., 2021; DA ROSA et al., 2022a), where it was possible to achieve considerable gains in performance and image quality, as this approach allows the use of a two-dimensional compromise. This trade-off consists of balancing not just two but four factors, allowing for a much more precise trade-off that allocates the most significant amount of resources to the most essential features at the expense of features of lesser importance.

Considering the scenario described above, the three compression techniques are used simultaneously in this work to achieve the best possible result in compressing medical ECG signals without losing information necessary for medical diagnosis. Next, we describe the built architectures, results obtained, and appropriate comparisons with other existing solutions.

Table 1 compares the main characteristics of the most relevant works in recent literature with the developed work.

### 3.3 Chapter Conclusion

This chapter presented the main architectures developed in this work. It shows one original architecture for the Haar wavelet transform (ODHWT). It also offers two other architectures using pruning (PDHWT) and approximate and pruning strategies (AxDHWT). The chapter also details the truncation strategy used in the proposed architectures. Besides, the chapter offered a discussion about related work using compression techniques, showing a table with the main differences between relevant works from the literature and our proposed solutions. The next chapter discusses the main results obtained in this work.

Table 1 – Comparison between the main related works and the work developed

Related work	A	B	C	D	E	F	G	H
(AHMED et al., 2022)	Adaptive Haar Wavelet (AHW)	✓	✓	×	×	×	×	×
(FATHI et al., 2022)	Krawtchouk Moments + Ant Lion Optimizer (AALO)	×	✓	×	×	×	×	×
(NAJAFI et al., 2023)	CNN	×	×	×	×	✓	×	×
(P. et al., 2023)	HDWT	✓	✓	✓	×	✓	×	✓
(KANANI; BHATTACHARJYA; BANERJEE, 2021)	N-point FIR Filter	✓	✓	×	×	✓	✓	×
(JANA et al., 2021)	DWT and IDWT	✓	✓	×	×	×	×	×
(KUMARI; RAJALAKSHMI, 2016)	HDWT	×	✓	✓	✓	×	×	×
(IFTODE; FOSALAU, 2020)	Undecimated wavelet transform (UWT)	×	✓	✓	✓	×	×	×
(YIN et al., 2021)	Adaptive Derivative	✓	✓	×	×	×	×	×
(WANG et al., 2019)	ADMM	✓	✓	×	×	✓	×	×
(CHEN; CHOU; WU, 2021)	FISTA	✓	✓	×	×	×	×	×
(HAMZA; RIJAB; HUSSIEN, 2021)	DWT (Haar, Bio. and Db)	×	✓	✓	✓	×	×	✓
(DING; ZILIC, 2016)	HWT	×	✓	✓	✓	×	×	×
(CHUMA et al., 2017)	HDWT (level 5)	✓	×	✓	✓	×	×	×
(CHEN; CHEN, 2015)	DWT	✓	×	✓	✓	×	×	×
(SEIDEL et al., 2021)	HDWT	✓	✓	✓	✓	✓	✓	✓
(ROSA et al., 2021)	HDWT (level 5)	✓	✓	✓	✓	×	×	×
(SEIDEL et al., 2020)	HDWT	✓	✓	✓	✓	×	×	×
Our Work	HDWT	✓	✓	✓	✓	✓	✓	✓

(A) Utilized solution. (B) Power results. (C) Data-driven methodology. (D) Usage of wavelet transform for ECG compression. (E) Exploration of multiple levels of wavelet transform compression. (F) Exploration of approximation techniques. (G) Realizes a study case of ECG pike detection (H) Uses truncation techniques for optimization.

## 4 RESULTS AND ANALYSIS

This chapter presents an analysis and discussion of the results obtained in this work. It commences by examining the synthesis results, focusing on the achieved energy and area savings. Subsequently, the chapter presents the accuracy and precision of the signals generated by each compression architecture, employing a comprehensive set of metrics.

### 4.1 Synthesis Results

This section presents the results concerning the total solution cost, encompassing occupied area, operating frequency, and energy consumption. These metrics are crucial for evaluating system optimization and enabling comparisons with existing literature to identify the most computationally efficient solution.

Evaluation of optimization is paramount in resource-constrained systems, where computational resources are limited, and maximizing optimization without signal distortion is critical. In mobile systems, often battery-powered, energy consumption is a critical factor for solution viability. For continuous ECG signal acquisition, energy consumption becomes a limiting factor, and reducing it translates to enhanced system durability and autonomy.

The developed architectures underwent synthesis to extract key hardware characteristics (energy consumption and occupied area). This process identifies the most efficient architecture based on the evaluated metrics. Each architecture underwent two syntheses to assess the impact, particularly on energy consumption, caused by varying the frequency (and consequently, the period). Table 2 presents the synthesis results obtained at the maximum operating frequency (period of 909 ps, corresponding to a frequency of 1.10 GHz). The synthesis employed the low-power ST 65 nm commercial standard cell library with a 1.25V supply voltage.

Table 3 compares the solutions developed at a frequency of 125KHz, generated using a period of 7812500 ps.

An analysis of the generated tables across all evaluated metrics reveals that so-

Table 2 – Comparison of synthesis results between the developed architectures with a period of 909 ms.

Architecture	Implementation	Leak [ $\mu W$ ]	Internal [ $\mu W$ ]	Switch [ $\mu W$ ]	Total [ $\mu W$ ]	Cells [K Gates]	Cell Area [ $\mu m^2$ ]	Net. Area [ $\mu m^2$ ]	Total Area [ $\mu m^2$ ]
Haar5 - M2M2M1	PDHWT	1.167	49.534	58.554	109.255	275	1363.960	557.171	1921.131
Haar5 - M2M2M1	AxDHWT	0.872	26.720	21.509	49.101	176	894.920	308.885	1203.805
Haar5 - 5M1	PDHWT	0.854	39.525	50.838	91.217	250	1022.320	472.261	1494.581
Haar5 - 5M1	AxDHWT	0.777	30.826	42.994	74.597	225	932.880	427.787	1360.667
Haar5 - M2M3	PDHWT	3.572	277.195	609.64	890.407	821	3189.680	1168.184	4385.864
Haar5 - M2M3	AxDHWT	2.281	92.14	381.296	475.717	522	3797.56	1471.044	5268.604
Haar5 - M5	PDHWT	7.848	1085.12	1556.46	2649.428	1644	9196.200	3447.367	12643.567
Haar5 - M5	AxDHWT	6.740	316.996	312.160	635.896	1319	7623.200	2649.260	10272.460

Cells highlighted in green ● and red ● represent the best and worst results, respectively.

Table 3 – Comparison of synthesis results between the developed architectures with a period of 7812500 ms

Architecture	Implementation	Leak [ $\mu W$ ]	Internal [ $\mu W$ ]	Switch [ $\mu W$ ]	Total [ $\mu W$ ]	Cells [K Gates]	Cell Area [ $\mu m^2$ ]	Net. Area [ $\mu m^2$ ]	Total Area [ $\mu m^2$ ]
Haar5 - M2M2M1	PDHWT	1.167	0.0057	0.0068	1.1795	275	1363.960	557.171	1921.131
Haar5 - M2M2M1	AxDHWT	0.842	0.0031	0.0025	0.8476	176	894.920	308.885	1203.805
Haar5 - 5M1	PDHWT	0.854	0.0045	0.0059	0.8644	250	1022.320	472.261	1494.581
Haar5 - 5M1	AxDHWT	0.769	0.0032	0.0049	0.8212	225	932.880	427.787	1360.667
Haar5 - M2M3	PDHWT	3.581	0.0322	0.0442	3.6574	821	3189.680	1168.184	4385.864
Haar5 - M2M3	AxDHWT	2.543	0.0107	0.0091	2.6538	522	3797.56	1471.044	5268.604
Haar5 - M5	PDHWT	7.848	0.1262	0.1810	8.1552	1644	9196.200	3447.367	12643.567
Haar5 - M5	AxDHWT	6.740	0.0368	0.0363	6.8131	1319	7623.200	2649.260	10272.460

Cells highlighted in green ● and red ● represent the best and worst results, respectively.

lutions incorporating approximation achieved superior results compared to architectures employing only pruning for optimization within the same construction. In terms of energy consumption, a significant increase is observed at the maximum frequency, exceeding the average consumption at lower frequencies by a factor of 100. However, the most energy-efficient solutions utilize five M1 blocks (5M1), two M2 blocks, and one M1 block (M2M2M1 configuration). Among these top performers, the M2M2M1 configuration exhibits a slight reduction of 25.496  $\mu W$  at the maximum frequency. At higher frequencies, the 5M1 architecture demonstrates the most favorable results, achieving a consumption reduction of 0.626  $\mu W$  compared to the second-most efficient M2M2M1 configuration.

As anticipated, the solution employing a single M5 block (M5) directly implemented based on traditional wavelet mathematics exhibits the worst performance. Regarding occupied area, solutions 5M1 and M2M2M1 again demonstrate a significant advantage. The M2M2M1 architecture achieves a marginal area advantage of 156.86  $\mu m^2$  compared to the second-most efficient solution. The M5 architecture exhibits the worst performance in terms of area, with a total area approximately 8.5 times larger than the M2M2M1 solution. This result aligns with the increased hardware complexity inherent to the M5 solution.

In terms of energy consumption, M2M2M1 was the most efficient solution at maximum frequency. However, based on the number of components that make up the solution, the 5M1 solution would be the most efficient in both cases. This difference is due to the fact that the critical path of the M2M2M1 solution is much smaller, resulting in lower internal energy consumption and, in particular, lower switching requirements. In relation to the smaller area of the M2M2M1 solutions compared to the 5M1 solutions. We believe that this is due to the internal optimizations performed by the synthesis tool, as these optimizations tend to provide more meaningful results for solutions with



smaller critical paths, thus reducing the overall cost of the solution.

## 4.2 Accuracy Results

This section presents a comparative analysis of the results obtained using the previously defined error metrics for the developed architectures. The aim is to evaluate the implemented architectures based on the quality of the generated signal. A comprehensive set of metrics is employed to assess various aspects of the reconstructed signal, thereby identifying the configurations that achieve optimal quality within the tested architectures.

The accuracy analysis focuses on verifying if the signals generated through the optimized compression processes satisfy minimal quality requirements. This ensures that the original signal remains undisturbed and usable for its intended purpose. This analysis is particularly crucial for medical signals like ECGs due to the stringent quality requirements. Even minor distortions or alterations in such signals can significantly impact information extraction, rendering them unsuitable for medical diagnoses.

The subsections below present the results for each analyzed error metric. The tests utilize the M2M2M1 architecture, identified as the most energy-efficient and area-optimized solution, across all its compression levels.

### 4.2.1 Summary of accuracy results

Tables 4 and 5 present the results for all accuracy metrics applied to each decomposition level developed using the M2M2M1 architecture as the baseline. Table 4 details the results obtained after subjecting the NRS signal to compression by the developed architecture. The error metrics are employed to compare these compressed signal results with those obtained through exact compression performed by the Matlab/Simulink software. Table 5 follows the same format, but utilizes the more complex PVC signal as the basis for comparison.

Table 4 – Table containing accuracy metrics for the NRS signal

Level/metric	Accuracy	SSIM	SNR	MAE	NCC	RMSE	PRD
Aprox. lv1	0.9982	0.5046	0.1336	42.9582	0.9989	75.8596	0.6419
Aprox. lv2	0.9970	0.4401	0.0564	69.0830	0.9982	95.1245	1.0649
Aprox. lv3	0.9941	0.3242	0.5894	189.3083	0.9194	150.1251	1.5422
Aprox. lv4	0.9919	0.3135	0.0901	184.7260	0.8842	223.1258	1.6001
Aprox. lv5	0.9757	0.3377	0.5287	411.7375	0.8312	392.2132	4.4501
Prune lv1	0.9997	0.5041	0.7854	40.9962	0.9995	74.3200	0.6747
Prune lv2	0.9971	0.4401	0.7590	69.0830	0.9989	95.0010	0.9286
Prune lv3	0.9945	0.3225	0.6887	125.3836	0.9276	149.1263	1.5132
Prune lv4	0.9933	0.3135	0.5977	184.7260	0.8891	223.0954	1.6466
Prune lv5	0.9792	0.3360	0.3258	328.6539	0.8354	402.2356	4.1208

Table 5 – Table containing accuracy metrics for the PVC signal

Level/metric	Accuracy	SSIM	SNR	MAE	NCC	RMSE	PRD
Aprox. lv1	0.9967	0.4400	0.0577	59.1011	0.9975	850.2500	0.6719
Aprox. lv2	0.9961	0.4432	0.0071	78.6420	0.9956	1357.9635	1.1349
Aprox. lv3	0.9941	0.2600	0.1125	153.7234	0.9875	1484.6992	1.7422
Aprox. lv4	0.9933	0.2631	0.0012	216.2800	0.9832	2999.1156	1.7001
Aprox. lv5	0.9792	0.4800	0.0541	303.8553	0.8837	3612.5631	4.5301
Prune lv1	0.9974	0.4409	0.0036	67.4410	0.9975	1103.1266	0.6757
Prune lv2	0.9965	0.4456	0.0071	78.6420	0.9956	1354.1443	0.9349
Prune lv3	0.9944	0.2606	0.0145	149.8548	0.9875	1698.6523	1.4632
Prune lv4	0.9938	0.2652	0.0012	216.2800	0.9832	3000.0010	1.6566
Prune lv5	0.9793	0.4836	0.0253	304.1353	0.8837	4212.8651	4.2508

Tables 4 and 5 solely present the results pertaining to signal quality metrics. Consequently, these tables offer an incomplete view of the solution's overall quality. In general, solutions achieving the best error metric values often exhibit lower compression ratios, resulting in a diminished gain in energy savings. Therefore, identifying the optimal solution necessitates a trade-off between energy efficiency and signal fidelity, which can vary depending on the specific application's constraints.

The following subsections provide a concise explanation and graphical representation of the results for each metric displayed in the aforementioned tables. The peak count metric is the only exception, as it will be addressed in greater detail in a dedicated subsection later.

#### 4.2.1.1 Accuracy

With respect to the accuracy metric, Figure 24 presents a comparative analysis of the accuracy calculations for the three developed architectures across all five decomposition levels. Figure 24 a) depicts the comparison for the NRS signal, while Figure 24 b) presents the comparison for the PVC signal.

Since the hardware responsible for compression remains unchanged, the original and pruned architectures exhibit identical accuracy. The approximate architecture demonstrates lower accuracy due to the removal of the multiplier component. When comparing accuracy across different compression levels, a trend emerges: lower accuracy values are associated with higher compression levels. This is expected, as the fifth level achieves the highest compression ratio but has a limited number of bits available for signal reconstruction, leading to reduced fidelity.

#### 4.2.1.2 Peak detection

As previously mentioned, peak detection was implemented using the Pan-Tompkins algorithm. Figures 25 (NSR) and 26 (PVC) illustrate the steps involved in peak detection according to this algorithm. Figures 25 and 26 a) depict the original raw signal

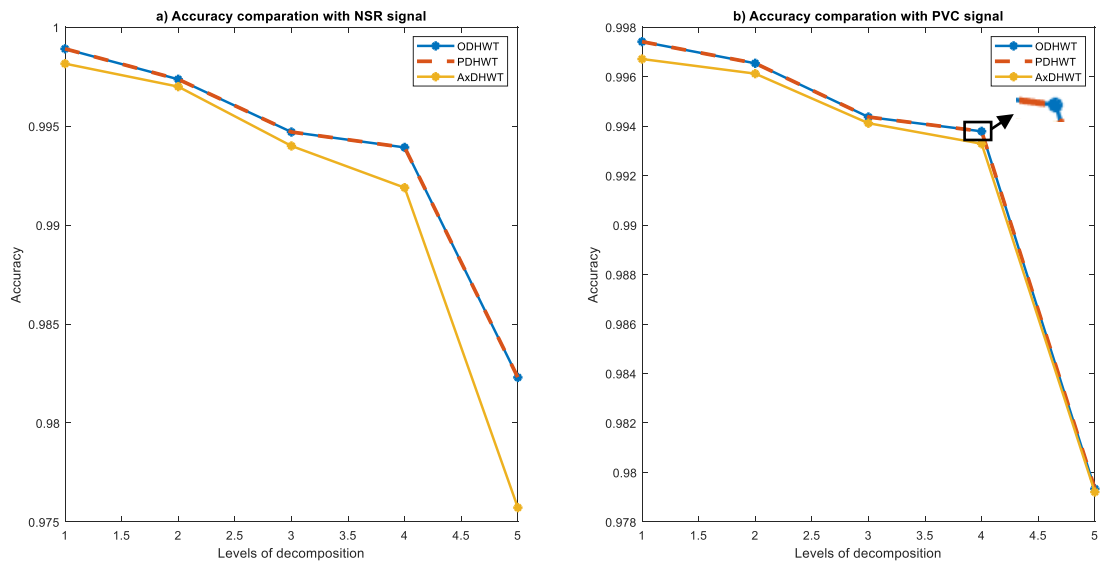


Figure 24 – Accuracy comparison in all compression levels for both NSR and PVC signals

obtained from the database. Figures 25 and 26 b) show the signal after passing through a first band-pass filter. Subsequently, the signal is processed by a derivative filter, as shown in Figures 25 and 26 c). The signal is then squared (Figures 25 and 26 d)), with the final results for a non-compression scheme presented in Figures 25 and 26 e). To enhance visualization, the images in the figure in question were assembled on the basis of a segment of the original signal containing 2000 samples.

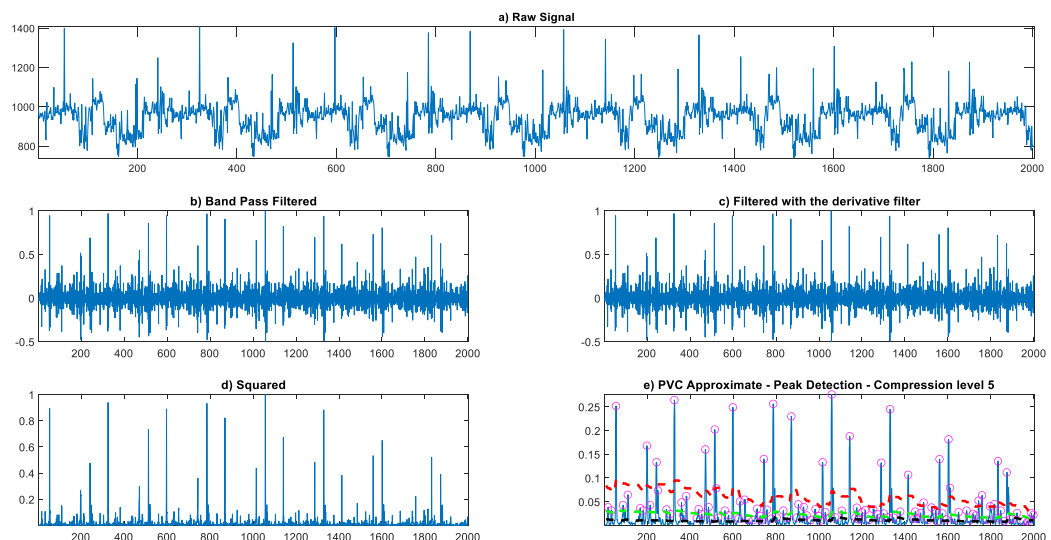


Figure 25 – Step by step execution of the peak detection algorithm in a sample of the NSR signal

Figure 27 shows the results of applying the peak detection algorithm to the 2,000 samples NSR signal in all its decomposition levels and both pruned and approximated architectures. These results consist of the QRS signal complexes, shown in Figures

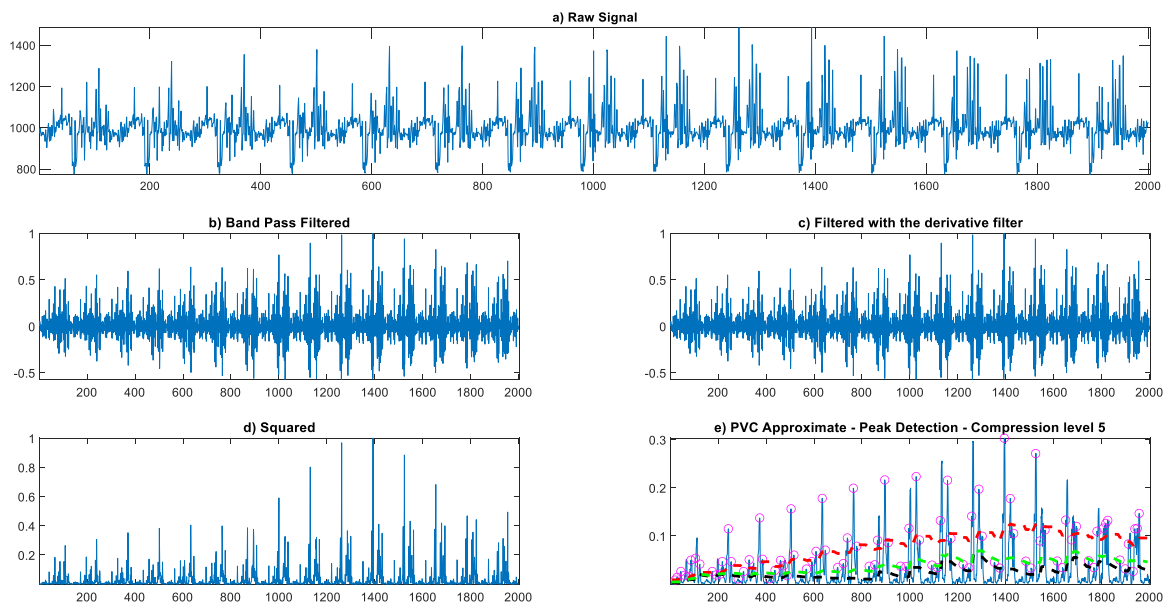


Figure 26 – Step by step execution of the peak detection algorithm in a sample of the PVC signal

27 a) and b) and the pulse train shown in Figure 27 c). Figure 28 presents the same analysis but the peak detection algorithm for the PVC signal.

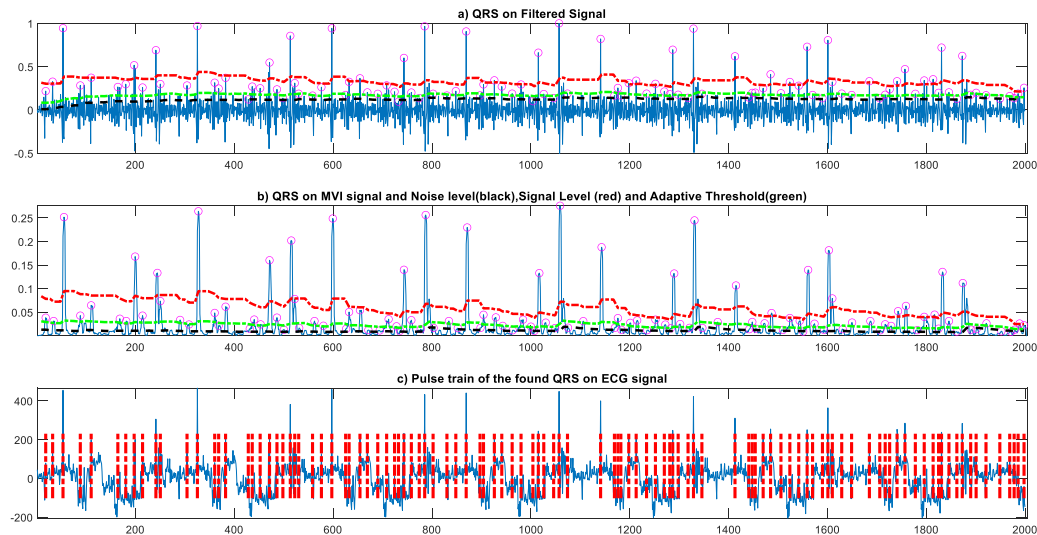


Figure 27 – Results of applying the peak detection algorithm to the NSR signal at the five compression levels

It is noteworthy that the peak detection algorithm achieved highly precise results for the first four decomposition levels in both the pruned and approximated architectures. However, peak detection accuracy was significantly compromised at the fifth level. This is attributed to the substantial reduction in the number of bits, resulting in a deformation of the post-compression waveform. The approximate solution, due to the absence of the multiplier circuit, exhibited a more pronounced loss in peak detection accuracy.

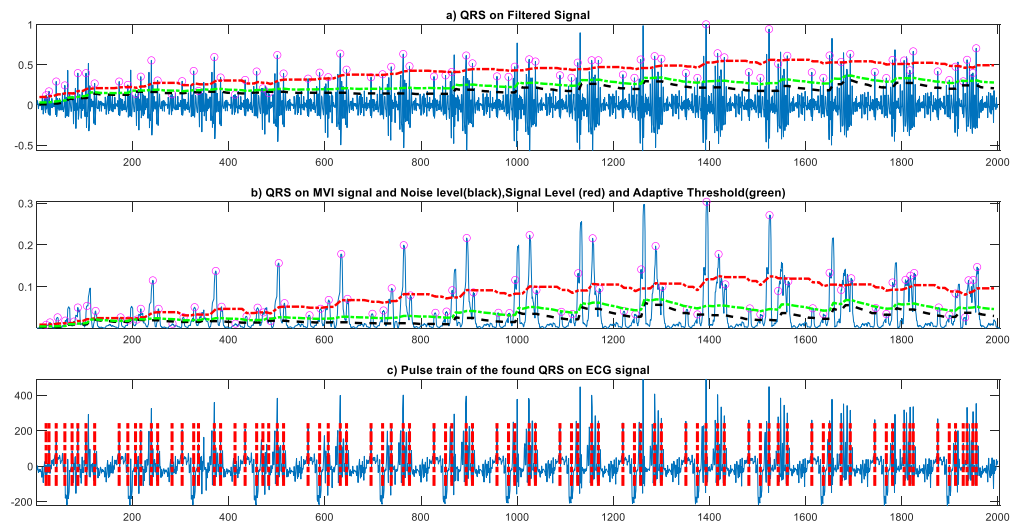


Figure 28 – Results of applying the peak detection algorithm to the PVC signal at the five compression levels

#### 4.2.1.3 Mean Absolute Error (MAE)

Figures 29 a) and b) present the MAE for the NSR and PVC signals, respectively.

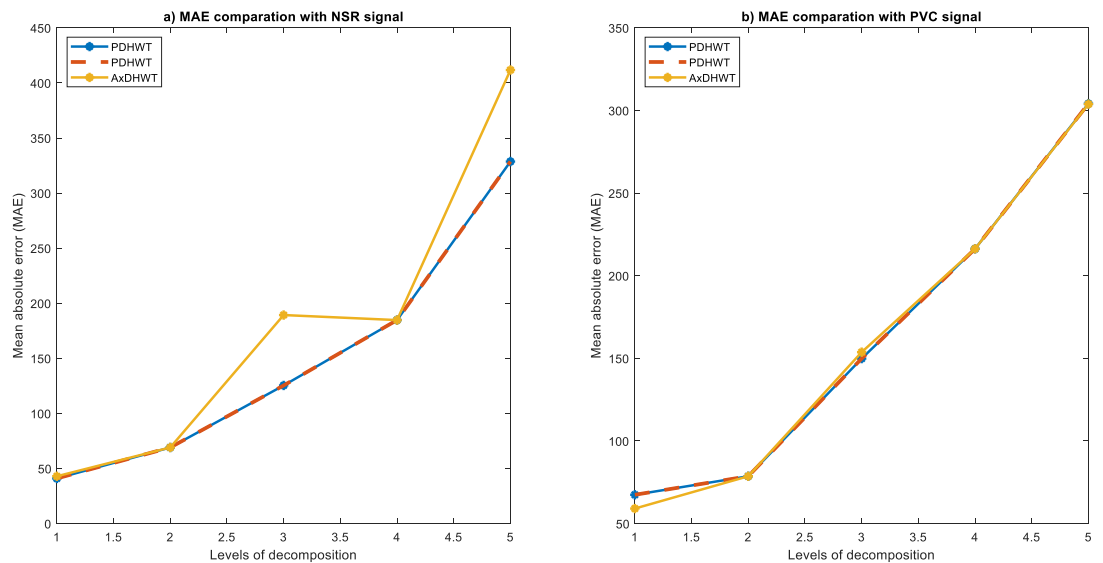


Figure 29 – MAE values considering the NSR and PVC signals at each compression level

A noteworthy observation is that the MAE error exhibits a proportional increase with each compression level, reaching its peak values at the fifth level of approximation. While both graphs display a similar trend, the overall error for the PVC signal is demonstrably larger.

#### 4.2.1.4 Normalized Cross-correlation (NCC)

Figures 30 a) and b) show the behavior presented by the NCC as the level of decomposition increases for both signals tested.

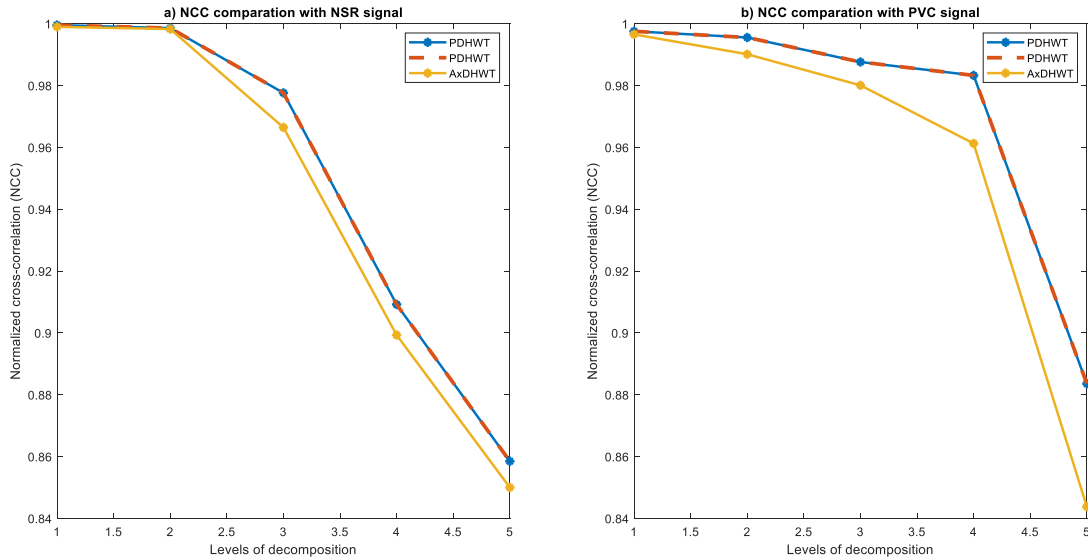


Figure 30 – Evolution of NCC along decomposition levels of NSR and PVC signals

The NCC presents a gradual reduction as decomposition levels increase. The NCC remained above 0.9 in both tested signals, considering up to the third compression level. The solutions with approximation present slightly lower values than those with only pruning.

#### 4.2.1.5 Percentage Root-mean-square difference (PRD)

In Figures 31 a) and b), it is possible to observe the values obtained with the PRD calculation at the tested decomposition levels. Those figures represent the results of the tests carried out with the NSR signal a) and the PVC signal b).

The PRD metric exhibits a behavior consistent with the observations from previous metrics. Similar to the MAE, the approximate solution demonstrates a marginally higher PRD across all levels. However, this difference is negligible in practical applications, and the substantial performance gain achieved through the removal of multipliers outweighs this minor trade-off.

#### 4.2.1.6 Root mean-square error (RMSE)

Figures 32 a) and b) compare the RMSE values at all tested compression levels. Those figures refer to the NSR signal a) and the PVC signal b).

The RMSE metric exhibits an approximately linear increase, with lower compression levels corresponding to lower RMSE values and higher compression levels resulting in higher values. This behavior is consistent for both signals; however, the PVC

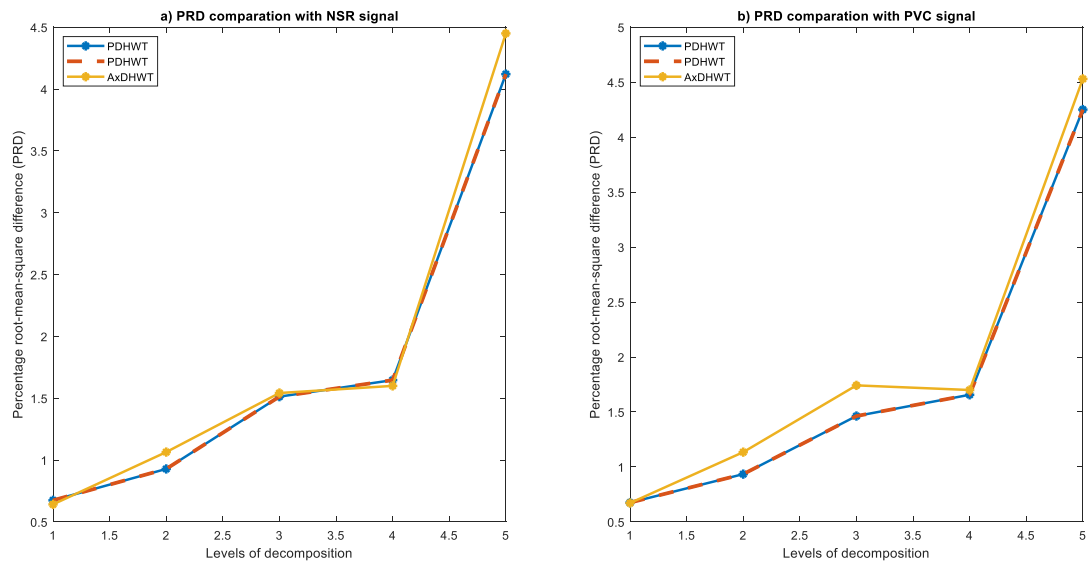


Figure 31 – PRD comparison in all compression levels for NSR and PVC signals

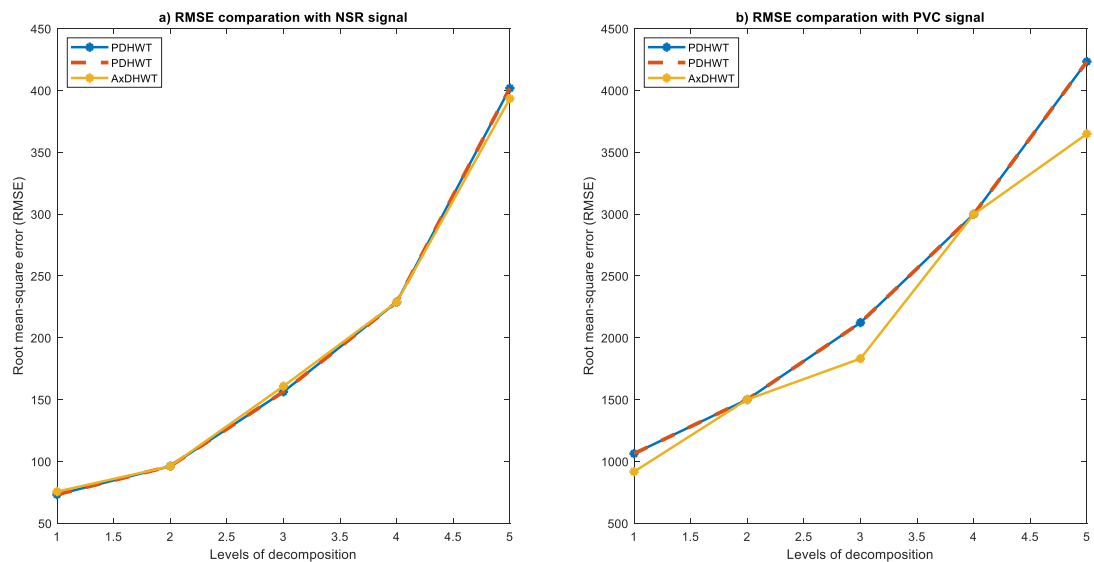


Figure 32 – RMSE comparison in all compression levels for NSR and PVC signals

signal inherently possesses greater instability, leading to significantly higher RMSE values. Notably, the architecture without approximation achieves marginally lower RMSE values in both scenarios. Nevertheless, this difference is negligible and can be disregarded without compromising quality.

#### 4.2.1.7 Signal-to-noise ratio (SNR)

Figures 33 a) and b) compare the SNR values at all tested compression levels. Those figures refer to the NSR signal a) and the PVC signal b).

Consistent with the behavior observed in other metrics, the SNR metric reveals a similar trend. Solutions employing pruning exhibit marginally superior values compared

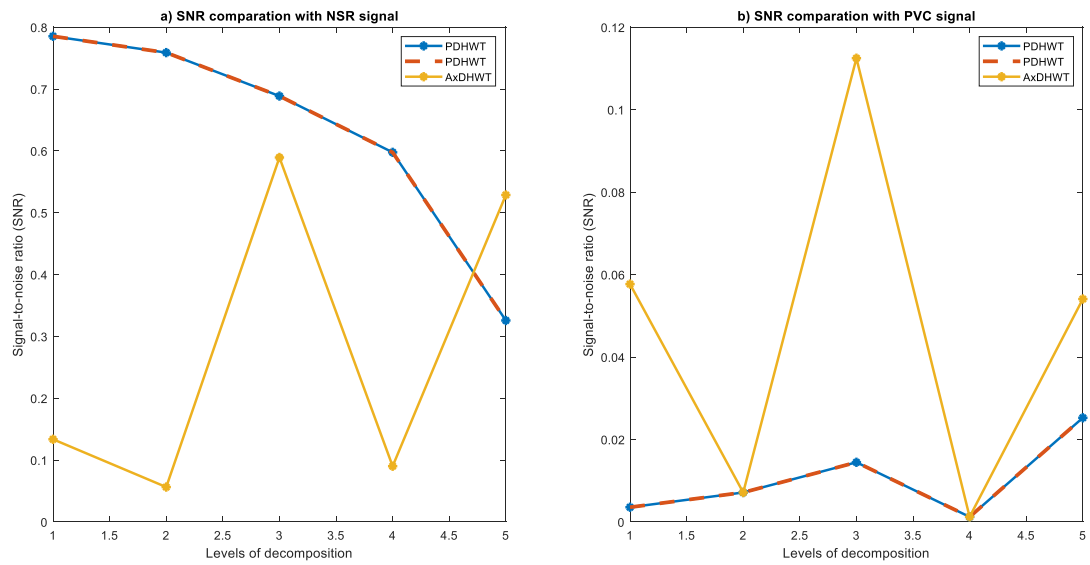


Figure 33 – SNR comparison in all compression levels for NSR and PVC signals

to those utilizing approximation. In terms of signal types, the PVC signal demonstrates significantly lower SNR values. Notably, the fifth level of approximation presents a reversal in SNR values, with approximated signals exhibiting lower values than pruned signals. However, this slight difference has a negligible impact on the overall compression effectiveness.

#### 4.2.1.8 Structural similarity index measure (SSIM)

Figures 34 a) and b) compare the SSIM values at all tested compression levels. Figure 34 presents waveforms for both NSR signal a) and the PVC signal b).

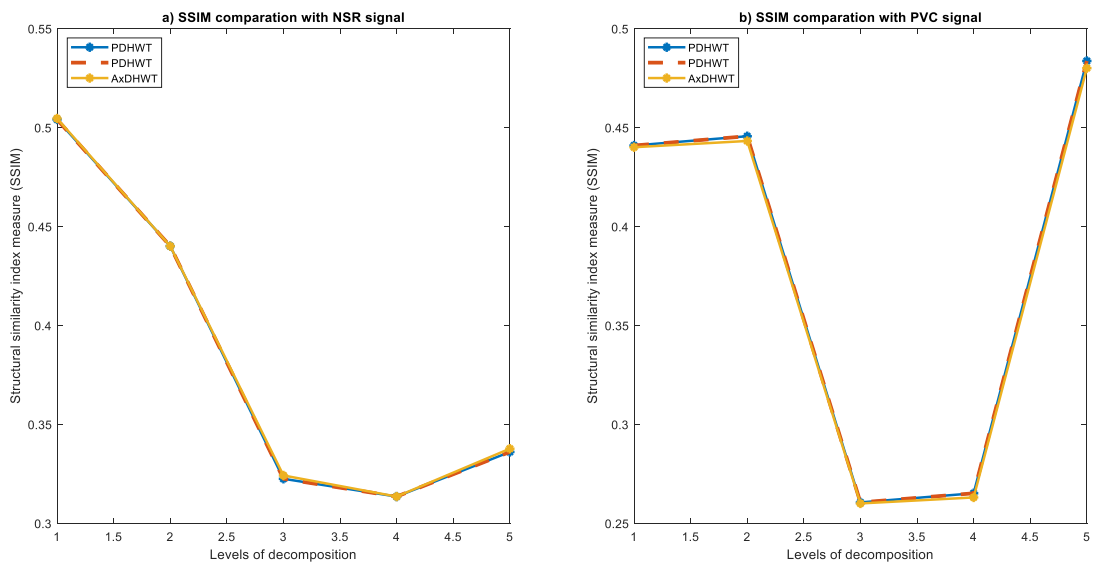


Figure 34 – SSIM comparison in all compression levels for NSR and PVC signals



The SSIM metric exhibits a notable reduction in value across the first three compression levels. This is followed by a stabilization of the metric value for both signals when transitioning from the third to the fourth level. However, an atypical behavior is observed in the PVC signal at the fifth level of decomposition, resulting in a significant increase in SSIM. The NSR signal also demonstrates a slight increase in SSIM at the fifth level.

#### 4.2.2 Results with truncation

This subsection discusses the results regarding bit truncation performed at the signal input. Based on an 11-bit input signal, experiments consider truncating 1, 4, 6, and 8 bits. These experiments evaluate the extent to which it is possible to reduce the input bits without losing the main properties of the original signal.

As truncation experiments are considerably time-consuming, we chose to perform them only with the PVC signal since it is the signal that presented the worst overall result in the accuracy metrics analyzed. Hence, this signal will offer greater resistance to bit removal. We present the results for both architectures developed with input bit truncation, i.e., Truncated and Pruned Discrete Haar Wavelet Transform and (TPDHWT) Truncated and Approximated Discrete Haar Wavelet Transform (TAXDHWT).

Figures 35 and 36 show a representation of the PVC signals one by one in each of their input bit reductions. The signals representation are as follows in Figures 35 and 36: a) represents 1-bit truncation, b) 2-bit, c) 3-bit, d) 4-bit, e) 5-bit, f) 6-bit, g) 7-bit and h) the most significant reduction tested of 8-bit. A reduction in the scale of the signal is visible as a reduction in the number of bits for its representation. In Figure 35, we visualize the complete signal after undergoing reductions in the number of bits. In contrast, Figure 36 shows a bit reduction on just 2,000 samples.

Tables 6, 7, 8, and 9 summarize the accuracy results obtained after applying the same error metrics used in the original signal tests to the signals subjected to varying bit truncation levels.

It is crucial to emphasize that, similar to the previous tables, these error metric tables do not represent the final solution quality. While lower error metric values are generally achieved with fewer input bit truncations, this translates to less hardware resource savings. Therefore, the optimal solution typically depends on the application's acceptable error tolerance in exchange for maximizing energy and space efficiency.

Table 6 presents the values for 1-bit truncation. Truncation of the smallest number of bits offers a low reduction in energy consumption despite having the best values in terms of quality. Based on the seven quality metrics analyzed, it is possible to observe a slight loss of quality when compared directly with the original solution.

Table 7 showcases the metrics obtained with 4-bit truncation. This truncation presents an intermediate between energy savings and signal quality, giving up a little more

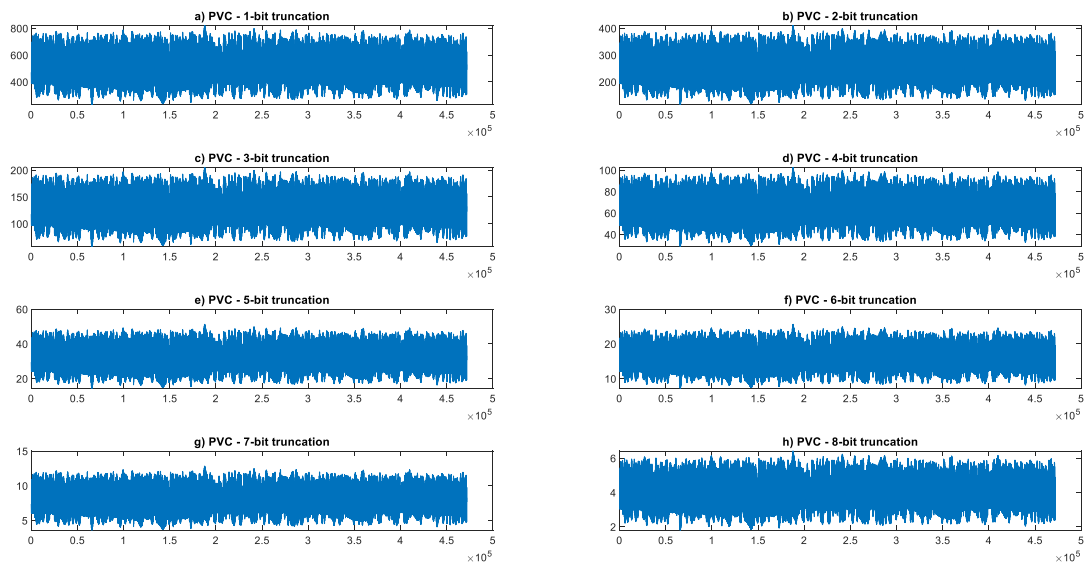


Figure 35 – Truncating full PVC signals

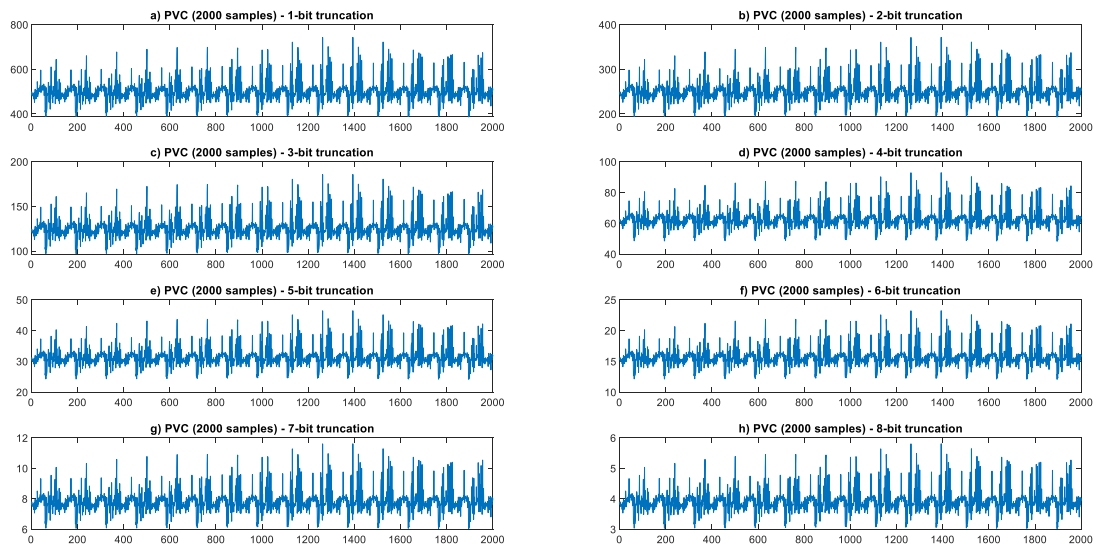


Figure 36 – Truncating PVC signals with 2,000 samples

quality to guarantee more energy savings, compared to the previous solution, which removes only 1 bit from the input.

Table 8 displays the results for 6-bit truncation. This truncation presents a greater focus on saving resources, giving up even more signal quality, and maintaining its values in the error metrics analyzed within the limits established by the literature.

Table 9 presents the results for the most significant truncation, 8-bit truncation. Using 8-bit truncation, we focus on energy savings, keeping signal quality at a minimum acceptable in several analyzed metrics to achieve the best possible energy and area savings among all the solutions developed.

Table 6 – Accuracy metrics for 1-bit truncation

level/metric	Accuracy	SSIM	SNR	MAE	NC	RMSE	PRD
Aprox. lv1	0.9940	0.8966	0.0418	8.44	0.9975	11.63	0.82
Aprox. lv2	0.9965	0.8611	0.0085	7.14	0.9958	17.72	0.89
Aprox. lv3	0.9938	0.9593	0.0176	17.87	0.9874	41.69	1.48
Aprox. lv4	0.9957	0.9808	0.0053	17.81	0.9923	46.13	1.16
Aprox. lv5	0.9809	0.7764	0.0306	109.76	0.8997	223.63	3.99
Prune lv1	0.9952	0.8594	0.0331	6.90	0.9975	10.94	0.78
Prune lv2	0.9965	0.8611	0.0085	7.14	0.9958	17.72	0.89
Prune lv3	0.9934	0.9534	0.0212	19.07	0.9973	41.83	1.48
Prune lv4	0.9957	0.9808	0.0053	17.81	0.9923	46.13	1.16
Prune lv5	0.9819	0.7760	0.0062	104.03	0.8997	223.30	3.87

Table 7 – Accuracy metrics for 4-bit truncation

level/metric	Accuracy	SSIM	SNR	MAE	NC	RMSE	PRD
Aprox. lv1	0.9888	0.6458	0.0912	15.80	0.9969	18.28	1.29
Aprox. lv2	0.9952	0.7398	0.0061	9.65	0.9953	18.52	0.93
Aprox. lv3	0.9906	0.8737	0.0511	26.66	0.9869	45.13	1.6
Aprox. lv4	0.9882	0.9424	0.0817	46.62	0.9917	61.06	1.54
Aprox. lv5	0.9624	0.7615	0.2609	213.11	0.8991	286.19	5.02
Prune lv1	0.9900	0.5192	0.0231	14.12	0.9923	18.02	1.28
Prune lv2	0.9897	0.7398	0.0803	20.47	0.9953	25.94	1.31
Prune lv3	0.9890	0.8352	0.0675	31.50	0.9851	51.33	1.81
Prune lv4	0.9882	0.9424	0.0817	46.62	0.9917	61.06	1.54
Prune lv5	0.9803	0.7589	0.0448	113.59	0.8988	230.79	4.1

Table 8 – Accuracy metrics for 6-bit truncation

level/metric	Accuracy	SSIM	SNR	MAE	NC	RMSE	PRD
Aprox. lv1	0.9844	0.2839	0.0233	21.64	0.9827	26.24	1.86
Aprox. lv2	0.9814	0.4428	0.1115	36.69	0.9816	44.93	2.24
Aprox. lv3	0.9827	0.6295	0.0478	48.56	0.9748	62.24	2.2
Aprox. lv4	0.9602	0.7782	0.3410	156.25	0.9776	172.18	4.41
Aprox. lv5	0.9691	0.6801	0.1505	175.70	0.8871	266.33	4.7
Prune lv1	0.9671	0.2190	0.0636	47.02	0.9341	61.36	4.34
Prune lv2	0.9774	0.4431	0.667	32.58	0.9816	39.83	1.99
Prune lv3	0.9746	0.5821	0.0029	72.05	0.9525	92.95	3.3
Prune lv4	0.9638	0.7766	0.0420	63.84	0.9776	82.54	2.08
Prune lv5	0.9660	0.6622	0.1136	171.25	0.8814	268.66	4.75

The following figures present a series of comparisons between the various accuracy metrics analyzed and the combined effects of decomposition level and input bit truncation. Each figure includes two panels: the first representing the approximate architecture and the second representing the architecture that only employs pruning.

Table 9 – Accuracy metrics for 8-bit truncation

level/metric	Accuracy	SSIM	SNR	MAE	NC	RMSE	PRD
Aprox. lv1	0.9152	0.1165	0.3498	132.34	0.7870	143.89	9.39
Aprox. lv2	0.9232	0.2109	0.1374	153.31	0.7784	187.83	11.69
Aprox. lv3	0.9290	0.3328	0.6232	230.29	0.7764	340.29	10.06
Aprox. lv4	0.9167	0.4748	0.3723	294.60	0.7725	407.74	10.45
Aprox. lv5	0.9081	0.4376	0.2209	465.30	0.7013	593.45	21.55
Prune lv1	0.9016	0.0724	0.6471	277.56	0.7112	288.79	9.97
Prune lv2	0.9106	0.2142	0.3226	176.49	0.7784	194.34	17.08
Prune lv3	0.8459	0.2824	0.4712	434.83	0.7023	464.48	9.17
Prune lv4	0.8192	0.4844	0.1009	321.0	0.7725	361.72	11.26
Prune lv5	0.8060	0.4174	0.2103	479.31	0.6725	638.39	143.89

From Figure 37 to Figure 43, we comprehensively compare the error metrics across different compression and truncation levels. The compared metrics include accuracy (in Figure 37), MAE (in Figure 38), NCC (in Figure 39), PRD (in Figure 40), RMSE (in Figure 41), SNR (in Figure 42), and SSIM (in Figure 43).

Figure 37 reveals highly consistent accuracy results, with acceptable values observed for input truncations up to 4 bits. A more significant loss in accuracy is only observed at the fifth level of decomposition for the approximate transform. Figure 37 a) depicts the accuracy values for the approximate solution, while Figure 37 b) presents the accuracy values for the pruned solution.

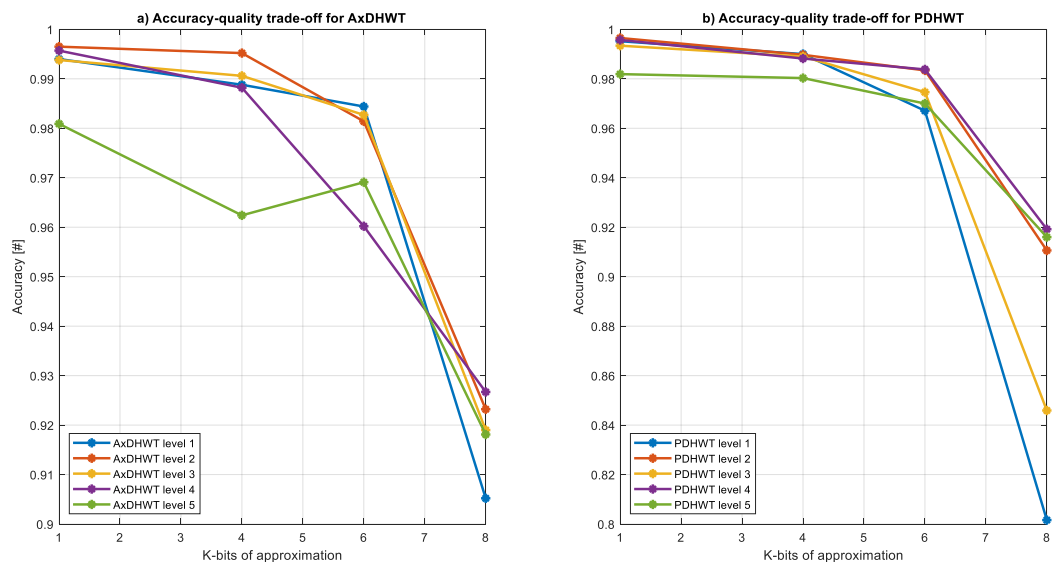


Figure 37 – Relationship between accuracy, decomposition levels and approximate number of bits

Figure 38 compares the MAE values obtained across different truncation levels. For the architecture employing only pruning (Figure 38 b), acceptable values are achieved up to 6-bit truncation of the input signal and a decomposition level of 4. In contrast,

the approximate transform (Figure 38 a) exhibits positive results with truncation up to 6 bits, but only for the first three decomposition levels. At level 4, a significant increase in error is observed starting from 4-bit truncation. Notably, the fifth level of decomposition presents high error values even with the minimum truncation of 1 bit.

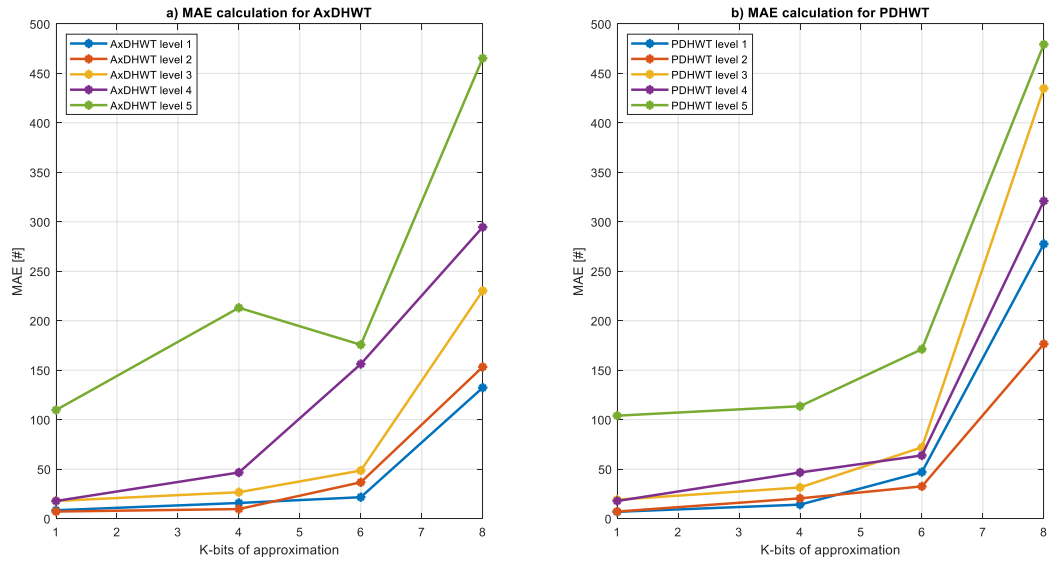


Figure 38 – Relationship between MAE, decomposition levels and approximate number of bits

The NCC metric, presented in Figures 39 a) and b), reveals that acceptable values are achievable for both the approximate and pruned architectures with input truncation up to 4 bits and a decomposition level of 4 or lower. Consistent with previous observations, the fifth level of decomposition exhibits unsatisfactory results, even with minimal truncation of 1 bit.

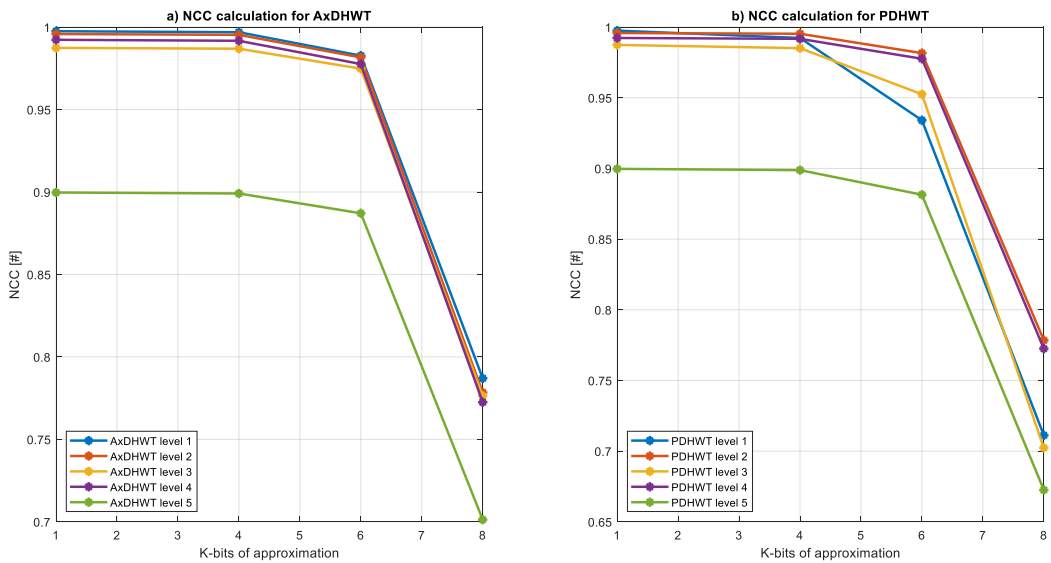


Figure 39 – Relationship between NCC, decomposition levels and approximate number of bits

Figure 40 depicts remarkably low PRD values observed for both the approximate architecture (in Figure 40 a) and the pruned architecture (in Figure 40 b) across all decomposition levels. It is achievable with input bit truncation up to 6 bits. However, removing a larger number of bits leads to significant errors at all decomposition levels, particularly for the approximate architecture and the fifth level of the pruned architecture.

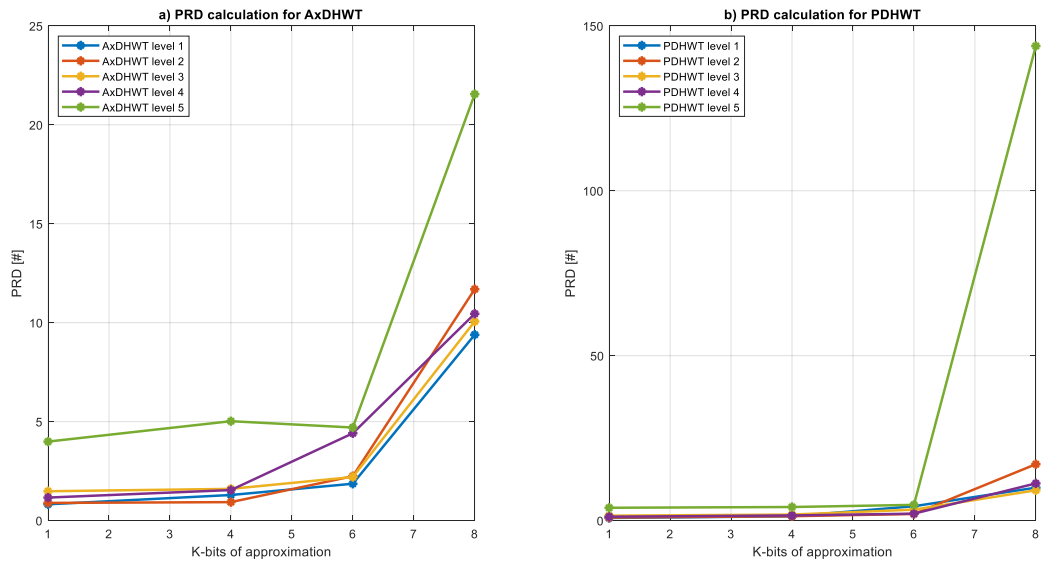


Figure 40 – Relationship between PRD, decomposition levels and approximate number of bits

Figure 41 compares RMSE values across input truncation levels. The pruned architecture (Figure 41 b) exhibits acceptable error values when removing up to 6 bits from the input signal and employing a maximum decomposition level of 4. Conversely, the approximate architecture (Figure 41 a) achieves acceptable error values with a maximum truncation of 4 bits and compression up to level 4. Consistent with prior observations, the fifth level of decomposition in both architectures displays high error values, even with minimal truncation of 1 bit.

The SNR metric, presented in Figure 42, exhibits limitations in generating highly precise results for truncated signals. However, we can still make some observations. The pruned architecture (Figure 42 b) maintains relatively low SNR values when removing up to 5 bits from the input signal. Conversely, the approximate architecture (Figure 42 a) only achieves it for minimal truncation of 1 bit.

Like the SNR metric, the SSIM metric presented in Figure 43 demonstrates limited suitability for evaluating one-dimensional signals with truncated input. However, for both the approximate architecture (Figure 43 a) and the pruned architecture (Figure 43 b), acceptable and comparable SSIM values are achievable with input truncation up to 2 bits. When removing more bits, the lowest SSIM values are observed at the lowest compression levels (particularly level 1) in both architectures.

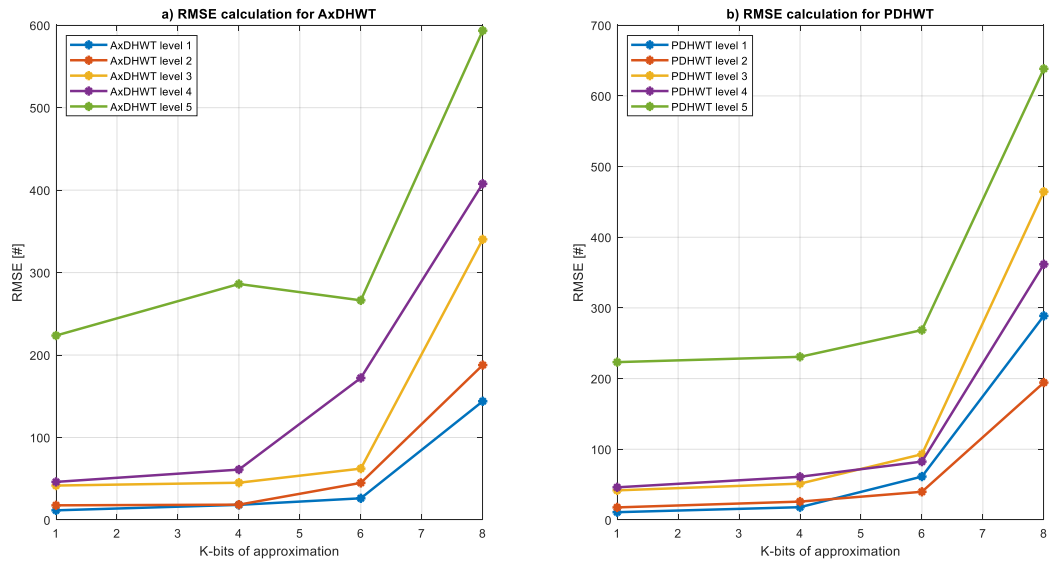


Figure 41 – Relationship between RMSE, decomposition levels and approximate number of bits

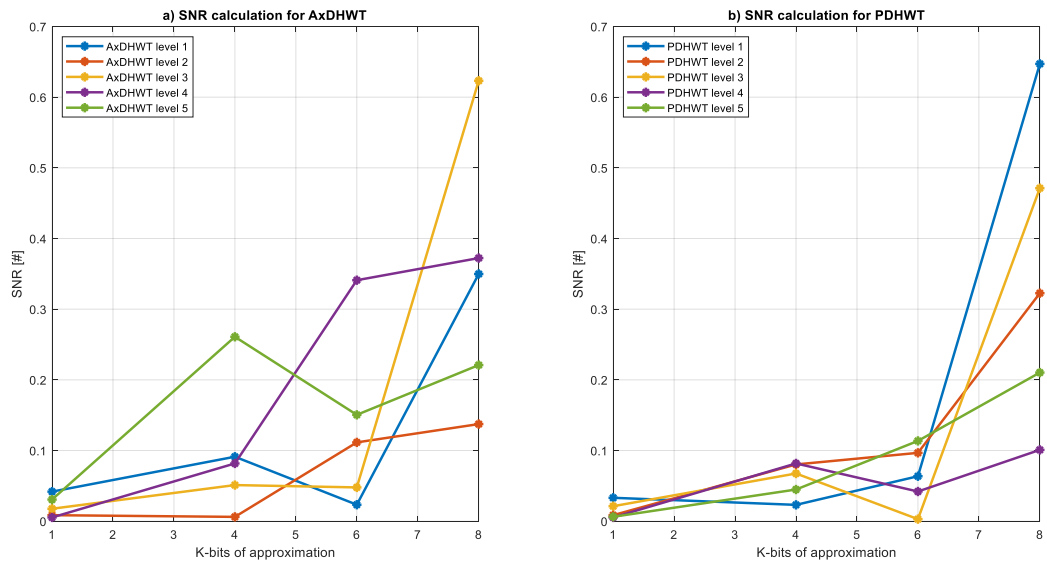


Figure 42 – Relationship between SNR, decomposition levels and approximate number of bits

A comprehensive analysis of all metrics across both architectures and compression levels has revealed several key findings. The results indicate a highly effective solution when considering input truncation of up to 4 bits (4 bit truncation) and acceptable results, for most metrics with up to 6-bit (6 bit truncation) both at a maximum decomposition level of 4. Error metrics consistently point to minimal information loss and preservation of overall signal quality. It is important to note that compressions performed at the fifth level of decomposition, even with minimal truncation (1 bit), yielded unsatisfactory results in most metrics. It suggests that the fifth level is unsuitable for employing the bit truncation technique at the input.

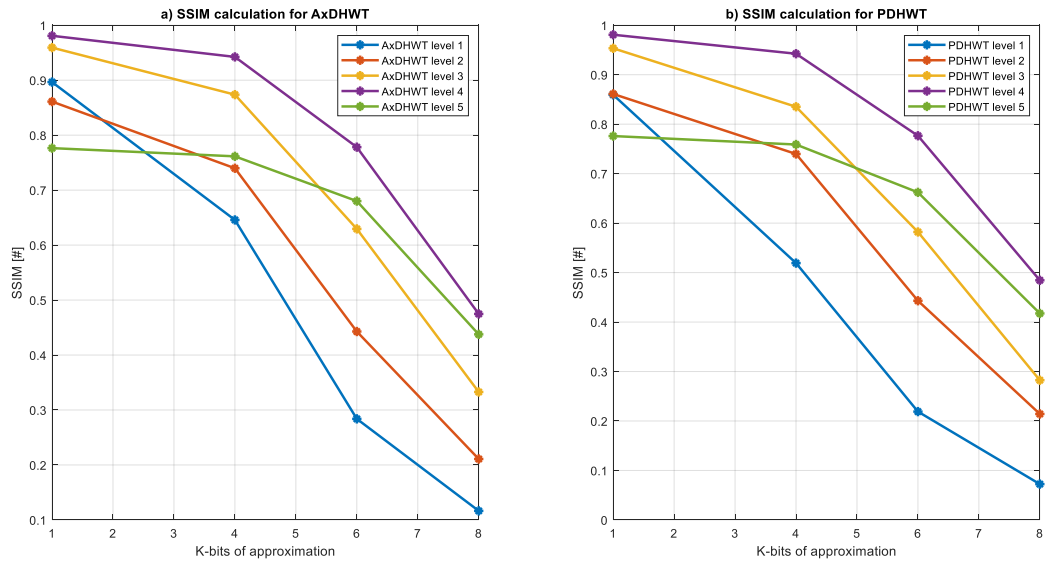


Figure 43 – Relationship between SSIM, decomposition levels and approximate number of bits

#### 4.2.2.1 Comparison between solutions with truncation

Table 10 compares various configurations for the most efficient architecture, M2M2M1. These configurations incorporate input bit truncation ranging from 0 to 6 bits. We chose this truncation range due to a significant degradation in signal quality observed with larger truncations. The comparison focuses on three key aspects: energy savings, cell count, and total area occupancy. For each truncation level, two architectures are presented: one utilizing only truncation and pruning and another employing all three compression techniques (truncation, pruning, and approximation). Finally, the baseline architecture, representing the original anatomized version, is included for reference. We can quantify the gains achieved through optimization by directly comparing the baseline with the optimized solutions.

Table 10 – Comparison of synthesis results between different truncation levels for the M2M2M1 architecture

Architecture	Implementation	Leak [ $\mu W$ ]	Internal [ $\mu W$ ]	Switch [ $\mu W$ ]	Total [ $\mu W$ ]	Cells [K Gates]	Cell Area [ $\mu m^2$ ]	Net. Area [ $\mu m^2$ ]	Total Area [ $\mu m^2$ ]
M2M2M1 - 11 bits	TPDHWT (0 bits trunc)	1.167	0.0057	0.0068	1.1795	275	1363.960	557.171	1921.131
M2M2M1 - 11 bits	TAxDHWT (0 bits trunc)	0.842	0.0031	0.0025	0.8476	176	894.920	308.885	1203.805
M2M2M1 - 10 bits	TPDHWT (1 bits trunc)	1.607	0.0079	0.0056	1.6211	204	1629.680	459.654	2089.334
M2M2M1 - 10 bits	TAxDHWT (1 bits trunc)	1.214	0.0061	0.0045	1.2257	173	1356.160	303.731	1659.891
M2M2M1 - 9 bits	TPDHWT (2 bits trunc)	1.469	0.0071	0.0051	1.4810	191	1465.360	421.030	1886.390
M2M2M1 - 9 bits	TAxDHWT (2 bits trunc)	1.108	0.0055	0.0040	1.1177	159	1237.600	278.028	1515.628
M2M2M1 - 8 bits	TPDHWT (3 bits trunc)	1.327	0.0062	0.0044	1.3380	169	1307.800	370.426	1678.226
M2M2M1 - 8 bits	TAxDHWT (3 bits trunc)	1.001	0.0049	0.0036	1.0099	145	1119.040	252.325	1371.365
M2M2M1 - 7 bits	TPDHWT (4 bits trunc)	1.189	0.0054	0.0038	1.1987	164	1239.160	355.485	1594.645
M2M2M1 - 7 bits	TAxDHWT (4 bits trunc)	0.6842	0.0035	0.0023	0.6900	102	778.440	168.461	946.901
M2M2M1 - 6 bits	TPDHWT (5 bits trunc)	0.968	0.0045	0.0030	0.9755	148	1180.201	309.850	1490.051
M2M2M1 - 6 bits	TAxDHWT (5 bits trunc)	0.618	0.0029	0.0022	0.6231	91	699.966	151.658	851.624
M2M2M1 - 5 bits	TPDHWT (6 bits trunc)	0.912	0.0037	0.0025	0.9192	134	962.520	279.730	1242.522
M2M2M1 - 5 bits	TAxDHWT (6 bits trunc)	0.595	0.0027	0.0021	0.6009	83	633.880	137.325	771.205
Baseline	Original	4.371	0.0266	0.0272	4.4250	358	1891.240	767.320	2658.560

Analysis of the developed solutions reveals a progressive decrease in energy consumption and occupied area by reducing the number of bits used in the architecture. Consequently, the solution employing 6-bit truncation in conjunction with approximation achieves the most significant reduction in both characteristics analyzed.






Furthermore, it is noteworthy that approximate architectures exhibit a general cost reduction comparable to that achieved with 2-bit truncation. It highlights the crucial role of approximation in facilitating compression and enabling a broader range of cost-quality trade-offs. Notably, the synthesis process to obtain these values was conducted at 125 kHz, with a period of 7812500 ps.

Table 11 compares the key performance indicators for the principal ECG data compression architectures developed in this study. It compares the baseline architecture, the AxDHWT solution (with pruning and approximation), and the TAxDHWT solution (with pruning, approximation, and 6-bit truncation). This table demonstrates the advantages of the solution incorporating truncation across all analyzed aspects.

Table 11 – Comparison between the main developed solutions and the baseline

Architecture	Energy Consumption [ $\mu W$ ]	Energy Savings [%]	Cells [K Gates]	Total Area [ $\mu m^2$ ]	Area Savings [%]
Baseline	4.4250	0	358	2658.560	0
AxDHWT	0.8212	522.06	176	1203.805	220.85
TAxDHWT (6-bit trunc)	0.6009	736.40	83	771.205	344.73



Cells highlighted in green , yellow  and red  represent the best, the intermediary and worst results, respectively.

### 4.3 Result comparisons with works from the literature

Table 12 compares the leading performance indicators with state-of-the-art implementations of ECG data compression.

Table 12 – Comparison with other ECG data compression architectures.

Related Work	Design Type	Process and Voltage	Algorithm	MCR	ACC	PRD	Area	Clock Freq	Power
(YIN et al., 2021)	ASIC	65 nm @ 0.9 V	Adaptive Derivative	17:1	0.975	<2.75	1.77 $mm^2$	0.128 MHz	0.00263 mW
(WANG et al., 2019)	ASIC	40 nm @ 0.6 V	ADMM	0.35	N/A	N/A	2.51 $mm^2$	265 MHz	9.1 mW to 81.5 mW
(CHEN; CHOU; WU, 2021)	ASIC	40 nm @ 0.9 V	FISTA	0.25	0.955	<9	0.239 $mm^2$	90 MHz	13.89 mW to 53.8 mW
(P. et al., 2023)	ASIC	40 nm @ 2.5 V	Clock Gating	N/A	N/A	N/A	2.25 $mm^2$	0.01 Mhz	22 mW to 45 mW
(ROSA et al., 2021)	ASIC	65 nm @ 1.25 V	DWT	0.03125	N/A	N/A	8.287 $\mu m^2$	500 Mhz	37.82 $\mu W$
(SEIDEL et al., 2021)	ASIC	65 nm @ 1.0 V	DWT	0.0625	0.9968	N/A	0.041 $mm^2$	360 Hz	144.8 $\mu W$
(HAMZA; RIJAB; HUSSEIN, 2021)	ASIC	N/A	DWT	16.33:1	N/A	<7	N/A	360 Hz	N/A
(DING; ZILIC, 2016)	ASIC	N/A	HWT	21.38:1	N/A	<2.54	N/A	500 Hz	N/A
(CHUMA et al., 2017)	FPGA	65 nm @ N/A	HDWT	0.03125	N/A	N/A	N/A	295.6 MHz	N/A
(CHEN; CHEN, 2015)	ASIC	40 nm @ 1.2 V	1D IDWT	N/A	N/A	<4.22	25,082 $\mu m^2$	200 MHz	17400 $\mu W$
(GODINHO et al., 2023)	ASIC	65 nm @ 1.25 V	PDHWT	0.125	N/A	<1.08	801.25 $\mu m^2$	125 KHz	0.83 $\mu W$
Our Work <sup>1</sup>	ASIC	65 nm @ 1.25 V	PDHWT	0.03125	0.9793	<4.25	1921.131 $\mu m^2$	125 kHz	1.18 $\mu W$
Our Work <sup>2</sup>	ASIC	65 nm @ 1.25 V	TPDHWT	0.0625	0.9606	2.08	1242.522 $\mu m^2$	125 kHz	0.91 $\mu W$
Our Work <sup>3</sup>	ASIC	65 nm @ 1.25 V	AxDHWT	0.03125	0.9792	<4.53	1203.805 $\mu m^2$	125 kHz	0.82 $\mu W$
Our Work <sup>4</sup>	ASIC	65 nm @ 1.25 V	TAxDHWT	0.0625	0.9602	<4.41	771.205 $\mu m^2$	125 kHz	0.60 $\mu W$

Cells highlighted in green  and red  represent the best and worst results, respectively. 1 - Pruned solution.

2 - Pruned solution with 6-bit truncation of input bits. 3 - Approximate solution. 4 - Approximate solution with truncation of input bits. N/A - Not Available. MCR - Minimum compression ratio.

Our proposed work demonstrates exceptional competitiveness compared to the most relevant recent publications with a similar focus on optimizing ECG signal transmission and manipulation. Consistent with most of these works, we chose an ASIC design approach. We opted to develop the solution using a 65nm CMOS cell library technology at a nominal voltage of 1.25V, as it represents the highest performing readily available cell library for direct experimentation. Notably, the work achieves the highest compression ratio at the fifth compression level compared to other architectures. It is crucial to acknowledge that the fifth level compression was only evaluated for architectures without truncation since we did not obtain satisfactory results with input bit truncation at this level.

A key takeaway is the trade-off inherent in high-compression solutions. Solutions pursuing extremely high compression (fifth level) prioritize compression over signal fidelity, resulting in lower accuracy and PRD values than lower-compression works. Conversely, when analyzing the solution with input bit truncation and a fourth-level compression rate, we observe highly competitive accuracy values and the best PRD values among the studied works. Regarding area, the work is well-positioned with a significantly smaller footprint than the worst-case scenarios in other solutions, presenting very acceptable and practical values.

The most significant advantage over other works is the remarkably lower energy consumption. The developed architecture exhibits the most critical energy savings across all metrics, boasting a substantial margin of advantage compared to existing solutions. The energy consumption is a mere  $0.60 \mu W$ , significantly outperforming the other developed solutions.

A metric-by-metric analysis confirms the advantageous positioning of the developed work compared to other highly relevant works. We will now delve into the error and consumption metrics to solidify this position further.

Accuracy-wise, our work exhibits good positioning compared to the literature. It does not achieve the absolute best accuracy values when directly compared, presenting a negative difference of 1.75% against the solution without truncation and a difference of 3.06% compared to the solution employing truncation. This disparity is justifiable, considering that the former utilizes a compression ratio of 1/32 (fifth level), leading to a more pronounced reduction in accuracy. The work in the literature with the highest accuracy results utilizes a compression ratio of 1/16 (fourth level), resulting in a more minor impact. When directly comparing the accuracy metrics obtained at both level 4 approximations, the difference is minimized to a minimal value of approximately 0.03% compared to the solution without truncation. The solution with truncation exhibits worse results due to the reduction of input bits, which leads to a significant loss of accuracy but yields substantial energy savings. Thus, validating the chosen trade-off.

Regarding the PRD, the solution developed with truncation achieves the lowest value. This solution surpasses the lowest value reported in the literature by approximately 20%, consequently being the best PRD value obtained for ECG signal compression at scales exceeding 1/8. We achieved this positive result with a compression ratio of 1/16 and bit truncation. As the solutions without truncation are based on a compression of 1/32, they incur a greater loss in PRD, leading to higher values.

The area occupied by the most efficient solution (TAXDHWT) positions itself consistently as the third-best solution among those analyzed. It possesses an area roughly 93.06 times larger than the smallest area reported in the literature and 3.02 times larger than the second smallest area. These area discrepancies are primarily attributed to the utilization of a higher number of compression levels, which necessitates a larger

footprint in exchange for enhanced compression and energy savings. The significant difference compared to the largest architecture also stems from discrepancies in the provided descriptions of the generated architectures and their focus on area optimization.

As previously highlighted, the work in question stands out for its exceptional energy efficiency. The combination of bit truncation with the two other employed techniques facilitates the realization of maximum savings in the ECG signal compression process. The developed system exhibits a reduction of 63.03 times in energy consumption for a compression ratio of 1/16 with bit truncation and a saving of 32.05 times when considering the solution without input bit truncation at the fifth compression level. When comparing the developed solutions, the solution with bit truncation delivers a reduction in consumption of approximately 50% when directly compared to the solution without truncation and without approximation (only pruning).

The NCC metric, widely employed for evaluating the quality of signal compression, particularly for medical signals demanding high precision, is not utilized by the primary works comparable to the developed solution. According to (FRASER et al., 2013), the minimum acceptable NCC value, for critical applications, reported in the literature is 0.978. Based on this value, architectures developed up to level 4 of decomposition exhibit values exceeding this established minimum, thereby providing acceptable outcomes for data compression.

The remaining metrics, MAE, RMSE, SNR, and SSIM, have limited applicability in evaluating one-dimensional medical signals but can still provide supplementary insights into signal quality. According to (XIONG et al., 2023), MAE values below 300 are considered acceptable, and all developed architectures satisfy this criterion. Regarding normalized RMSE, values consistently below 0.75 indicate acceptable signal quality, as supported by (CHIANG et al., 2019). For SNR, research by (ZHANG et al., 2017) suggests a favorable compression range of 0 to 40 dB. All SNR values obtained in our tests fall within this range, indicating successful compression. However, the SSIM metric presents a unique scenario. Unlike other metrics where a value closer to 1 signifies better quality, SSIM values closer to 0 indicate more significant errors. While minimum acceptable values are established around 0.90 in (LEE et al., 2020), it is important to note that SSIM is primarily used for image quality assessment in the studies referenced. In our work, SSIM exhibited an inverse relationship with other metrics, approaching 0 at lower compression levels and increasing with higher compression and bit reduction. This behavior underscores the limited applicability of SSIM for evaluating one-dimensional signals like ECG data.

## 5 CONCLUSION

Considering the great need to compress medical signals, especially ECG signals, this work presents several architectures with extremely low energy consumption to compress ECG signals, thus facilitating their acquisition and manipulation in real time and with the lowest possible use of computational resources.

We developed solutions using the discrete Haar wavelet transform (DHW) as a basis, applying three forms of optimization to this base transform, i.e., pruning, hardware approximation, and input bit truncation. With pruning, we removed unnecessary components, generating an architecture with only the essentials to carry out the signal compression process, allowing significant savings in occupied area and energy consumption.

We adopted a hardwired approach to achieve a more significant gain in energy consumption. It occurred by removing the most energetically costly component of the architecture, the single constant multiplier block. With this removal, at the cost of small losses in precision and accuracy metrics, an architecture capable of performing compression at an even lower cost was obtained.

To achieve maximum savings on hardware resources, in particular energy savings, bit truncation of the input was performed, obtaining excellent results considering a truncation of up to 6 bits at the input and the fourth compression level. The solution with the highest energy efficiency was developed, even among the best existing in current literature, using these three techniques together.

With bit truncation, the most energy efficient solution architecture obtained a maximum compression of the fourth level of Haar wavelet decomposition of 1/16. A worst-case accuracy of 0.9602. A PRD of a maximum of 4.41, a total area of 771.205  $\mu\text{m}^2$ , and the lowest energy consumption of 0.60  $\mu\text{W}$ . Considering the frequency of 1.25 kHz for greater energy savings. These results place the work developed among the leading ECG data compression systems regarding energy savings.

In addition to the most efficient solution, we generated other partial architectures despite not generating the greatest resource savings as the highlighted architecture. In most cases, they have higher values in accuracy metrics, making them ideal for use

in systems requiring greater data precision and not necessarily needing the most significant possible energy savings. From this, we can conclude that, in addition to the architecture with the highest possible efficiency, other architectures have been developed as by-products that are widely used in applications where the trade-off is more towards data precision at the expense of low energy consumption and area losses. This proposed design exploration space expands the application possibilities of the developed architectures.

## 5.1 Possibilities for future work

According to the excellent results obtained in this dissertation, we can highlight other possibilities for future works. These main possibilities for future works include:

1. To apply the developed architectures for compression of other one-dimensional medical signals, such as Electromyogram (EMG) and Electroencephalography (EEG).
2. To use compression techniques (pruning, approximation, and truncation) on a broader range of applications: Two-dimensional applications (images) and different types of one-dimensional signals.
3. To use the architecture developed for compressing other non-medical signals with lower precision requirements and more straightforward reconstruction. It is done to allow greater bit truncation or use a more significant number of decomposition levels, generating greater compression and consequent energy savings.
4. To use artificial intelligence (AI) techniques to help reconstruct the original signal and repair possible losses arising from compression.
5. Use of approximate adders in the developed solution to achieve greater energy savings.
6. Perform in-depth evaluations of the reliability of the developed system, especially with respect to the influence of noise and errors generated during the execution of the compression process, taking into account the wide use of approximation techniques

## REFERENCES

- ABDULBAQI, A. S.; SAIF, S. A.-D. M. N.; FALATH, F. M. M.; NAWAR, N. A. I. A Proposed Technique Based on Wavelet Transform for Electrocardiogram Signal Compression. In: ANNUAL INTERNATIONAL CONFERENCE ON INFORMATION AND SCIENCES (AICIS), 2018., 2018, USA. **Proceedings...** IEEE, 2018. p.229–234.
- AHMED, A. J.; HAMDI, M. M.; MUSTAFA, A. S.; RASHID, S. A. WSN Application Based on Image Compression Using AHAAR Wavelet Transform. In: INTERNATIONAL CONGRESS ON HUMAN-COMPUTER INTERACTION, OPTIMIZATION AND ROBOTIC APPLICATIONS (HORA), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–4.
- BALACHANDRAN, A.; GANESAN, M.; SUMESH, E. P. Daubechies algorithm for highly accurate ECG feature extraction. In: INTERNATIONAL CONFERENCE ON GREEN COMPUTING COMMUNICATION AND ELECTRICAL ENGINEERING (ICGCC-CEE), 2014., 2014. **Proceedings...** [S.l.: s.n.], 2014. p.1–5.
- BARAATI, F.; NASAB, M. T.; GHADERI, R.; JAFARI, K. FinFET-based Low-Power Approximate Multiplier for Neural Network Hardware Accelerator. In: IRANIAN INTERNATIONAL CONFERENCE ON MICROELECTRONICS (IICM), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.17–20.
- BURGUERA, A. Fast QRS Detection and ECG Compression Based on Signal Structural Analysis. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], v.23, n.1, p.123–131, 2019.
- BUYYA, A. et al. Arrhythmias Classification by using STFT-based Spectrograms, Transfer Learning and Concatenation of features. In: INTERNATIONAL CONFERENCE FOR EMERGING TECHNOLOGY (INCET), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.1–5.
- CHANDRA', S.; SHARMA, A.; SINGH, G. A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies. **IRBM**, [S.l.], v.42, n.4, p.227–244, 2021.

CHANDRAKASAN, A.; BRODERSEN, R. Low Power Digital CMOS Design. In: DORDRECHT: KLUWER ACADEMIC, 1995. **Proceedings...** [S.l.: s.n.], 1995.

CHANDRAKASAN, A.; SHENG, S.; BRODERSEN, R. Low-power cmos digital design. In: IEICE TRANSACTIONS ON ELECTRONICS, THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS, 1992. **Proceedings...** [S.l.: s.n.], 1992.

CHEN, K.-C.; CHOU, C.-Y.; WU, A.-Y. A Tri-Mode Compressed Analytics Engine for Low-Power AF Detection With On-Demand EKG Reconstruction. **IEEE Journal of Solid-State Circuits**, [S.l.], v.56, n.5, p.1608–1617, 2021.

CHEN, S.-W.; CHEN, Y.-H. Hardware Design and Implementation of a Wavelet De-Noising Procedure for Medical Signal Preprocessing. In: 2015. **Proceedings...** [S.l.: s.n.], 2015. v.10, n.10, p.19.

CHEN, Y. et al. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. **IEEE Intelligent Systems**, [S.l.], v.35, n.4, p.83–93, 2020.

CHIANG, H.-T. et al. Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders. **IEEE Access**, [S.l.], v.7, p.60806–60813, 2019.

CHIPPA, V. K. et al. Scalable effort hardware design: Exploiting algorithmic resilience for energy efficiency. In: DESIGN AUTOMATION CONFERENCE, 47., 2010. **Proceedings...** [S.l.: s.n.], 2010. p.555–560.

CHIPPA, V.; RAGHUNATHAN, A.; ROY, K.; CHAKRADHAR, S. Dynamic effort scaling: Managing the quality-efficiency tradeoff. In: DESIGN AUTOMATION CONFERENCE, 48., 2011. **Proceedings...** [S.l.: s.n.], 2011. p.603–608.

CHUMA, E. L.; MELONI, L. G. P.; IANO, Y.; BRAVO-ROGER, L. L. FPGA implementation of a de-noising using Haar level 5 wavelet transform. In: OF THE 48TH , 2017. **Proceedings...** [S.l.: s.n.], 2017. v.35, p.4.

CLARK, N. et al. A wearable ECG monitoring system for real-time arrhythmia detection. In: IEEE 61ST INTERNATIONAL MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS (MWSCAS), 2018., 2018. **Proceedings...** [S.l.: s.n.], 2018. p.787–790.

COMPUTATIONAL PHYSIOLOGY, M. L. for. **PhysioNet Database**. <https://physionet.org/about/database/>.

CONG, J.; GURURAJ, K. Assuring application-level correctness against soft errors. In: IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN (ICCAD), 2011., 2011. **Proceedings...** [S.l.: s.n.], 2011. p.150–157.

CONTRERAS-VALDES, A.; AMEZQUITA-SANCHEZ, J. P.; GRANADOS-LIEBERMAN, D.; VALTIERRA-RODRIGUEZ, M. Data compression based on discrete Wavelet transform and fault detection of short-circuit faults in transformers. In: IEEE INTERNATIONAL AUTUMN MEETING ON POWER, ELECTRONICS AND COMPUTING (ROPEC), 2019., 2019. **Proceedings...** [S.l.: s.n.], 2019. p.1–6.

COSTA, P. Ücker Leleu da. **Arquiteturas de Hardware Baseadas em Filtros Adaptativos para a Extração do Sinal de FECG**. 2021. 87p. Trabalho de Conclusão (Mestrado em engenharia eletrônica e computação) — Centro de Ciências Sociais e Tecnológicas, Universidade Católica de Pelotas, Pelotas.

DA ROSA, M. M. A.; DA COSTA, E. A. C.; SOARES, R. I.; BAMPI, S. Exploring Security Threats by Hardware-Faults in Approximate Arithmetic Computing. In: IEEE INTERREGIONAL NEWCAS CONFERENCE (NEWCAS), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.1–5.

DA ROSA, M. M. A. et al. A Multiplier-Less Level-3 Haar Wavelet Transform Approximation Requiring Five Additions Only. In: IEEE 15TH DALLAS CIRCUIT AND SYSTEM CONFERENCE (DCAS), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–6.

DA ROSA, M. M. A. et al. AxRSU: Approximate Radix-4 Squarer Unit. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1655–1659.

DA ROSA, M. M. A. et al. A Multiplier-Less Level-3 Haar Wavelet Transform Approximation Requiring Five Additions Only. In: IEEE 15TH DALLAS CIRCUIT AND SYSTEM CONFERENCE (DCAS), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–6.

DAS, N.; CHAKRABORTY, M. Performance analysis of FIR and IIR filters for ECG signal denoising based on SNR. In: THIRD INTERNATIONAL CONFERENCE ON RESEARCH IN COMPUTATIONAL INTELLIGENCE AND COMMUNICATION NETWORKS (ICRCICN), 2017., 2017. **Proceedings...** [S.l.: s.n.], 2017. p.90–97.

DATTA, A.; MOHANTA, A.; MISHRA, B. Approximation and Bit Truncation Techniques in Hardware for Edge Detection. In: INTERNATIONAL SYMPOSIUM ON EMBEDDED COMPUTING AND SYSTEM DESIGN (ISED), 2019., 2019. **Proceedings...** [S.l.: s.n.], 2019. p.1–5.

DING, H.; LIU, H.; YANG, F. A Real-Time Compression Method of Power System Waveform Data Based on 2-D Lifting Wavelet Transform and Deflate Algorithm. In: ASIA ENERGY AND ELECTRICAL ENGINEERING SYMPOSIUM (AEEES), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.626–630.



DING, Y. S.; ZILIC, Z. ECG compression for mobile sensor platforms. In: IEEE 13TH INTERNATIONAL CONFERENCE ON WEARABLE AND IMPLANTABLE BODY SENSOR NETWORKS (BSN), 2016., 2016. **Proceedings...** [S.l.: s.n.], 2016. p.99–104.

DUTT, S. et al. Approxhash: delay, power and area optimized approximate hash functions for cryptography applications. In: INTERNATIONAL CONFERENCE ON SECURITY OF INFORMATION AND NETWORKS, 10., 2017. **Proceedings...** [S.l.: s.n.], 2017. p.291–294.

ELHOSENY, M. et al. Secure Medical Data Transmission Model for IoT-Based Healthcare Systems. **IEEE Access**, [S.l.], v.6, p.20596–20608, 2018.

ESMAEILZADEH, H.; SAMPSON, A.; CEZE, L.; BURGER, D. Architecture support for disciplined approximate programming. In: ARCHITECTURAL SUPPORT FOR PROGRAMMING LANGUAGES AND OPERATING SYSTEMS, 2012. **Proceedings...** [S.l.: s.n.], 2012. p.301–312.

ESPOSITO, D. et al. Approximate Multipliers Based on New Approximate Compressors. **IEEE Transactions on Circuits and Systems I: Regular Papers**, [S.l.], v.65, n.12, p.4169–4182, 2018.

FATHI, I. S.; MAKHLOUF, M. A. A.; OSMAN, E.; AHMED, M. A. An Energy-Efficient Compression Algorithm of ECG Signals in Remote Healthcare Monitoring Systems. **IEEE Access**, [S.l.], v.10, p.39129–39144, 2022.

FRASER, G. D.; CHAN, A. D. C.; GREEN, J. R.; MACISAAC, D. T. Biosignal quality analysis of surface EMG using a correlation coefficient test for normality. In: IEEE INTERNATIONAL SYMPOSIUM ON MEDICAL MEASUREMENTS AND APPLICATIONS (MEMEA), 2013., 2013. **Proceedings...** [S.l.: s.n.], 2013. p.196–200.

GALA, M.; BARABAS, J.; KRAJNAK, J. Robust QRS complex detector algorithm based on modified Pan-Tompkins method and wavelet transform. In: INTERNATIONAL CONFERENCE ON TELECOMMUNICATIONS AND SIGNAL PROCESSING (TSP), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.633–636.

GODINHO, A. C. et al. An Ultra Low-Energy VLSI Approximate Discrete Haar Wavelet Transform for ECG Data Compression. **IEEE Transactions on Circuits and Systems**, [S.l.], p.4, 2023.

GOLDBERGER, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. In: 2020 43RD , 2000. **Proceedings...** [S.l.: s.n.], 2000. v.101, n.23, p.215–220.

GUPTA, V. et al. IMPACT: IMPrecise adders for low-power approximate computing. In: IEEE/ACM INTERNATIONAL SYMPOSIUM ON LOW POWER ELECTRONICS AND DESIGN, 2011. **Proceedings...** [S.l.: s.n.], 2011. p.409–414.

GUPTHA, M.; RAO, S.; SELVARA, R. An Efficient Discrete Wavelet Transform Architecture with Low Power and Multiplier-Less Structure for Pervasive Biomedical Image Processing Application. **EAI Endorsed Transactions on Pervasive Health and Technology Research Article**, [S.l.], v.9, p.1–8, 2023.

HAMZA, R. K.; RIJAB, K. S.; HUSSIEN, M. A. The ECG data Compression by Discrete Wavelet Transform and Huffman Encoding. In: INTERNATIONAL CONFERENCE ON CONTEMPORARY INFORMATION TECHNOLOGY AND MATHEMATICS (ICCITM), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.75–81.

HEGDE, R.; SHANBHAG, N. R. Energy-efficient signal processing via algorithmic noise-tolerance. In: LOW POWER ELECTRONICS AND DESIGN, 1999., 1999. **Proceedings...** [S.l.: s.n.], 1999. p.30–35.

IFTODE, I. A.; FOSALAU, C. Wavelet-based Techniques Applied to Digital Processing of ECG Signals. In: INTERNATIONAL CONFERENCE AND EXPOSITION ON ELECTRICAL AND POWER ENGINEERING (EPE), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.419–424.

JANA, J. et al. An Area Efficient VLSI Architecture for 1-D and 2-D Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT). In: DEVICES FOR INTEGRATED CIRCUIT (DEVIC), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.378–382.

JINDAL, B.; SAUDAGAR; EKTA; DEVI, R. Matlab Based GUI for ECG Arrhythmia Detection Using Pan-Tompkin Algorithm. In: FIFTH INTERNATIONAL CONFERENCE ON PARALLEL, DISTRIBUTED AND GRID COMPUTING (PDGC), 2018., 2018. **Proceedings...** [S.l.: s.n.], 2018. p.754–759.

K, R.; A, D. V. N.; J, J. A. Decryption of Encrypted Fractional Wavelet Transformed Image Using Its Inverse. In: SECOND INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS, INFORMATION AND COMMUNICATION TECHNOLOGIES (ICEEICT), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.1–7.

KANAGARAJ, H.; MUNESWARAN, V. Image compression using HAAR discrete wavelet transform. In: INTERNATIONAL CONFERENCE ON DEVICES, CIRCUITS AND SYSTEMS (ICDCS), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.271–274.

KANANI, A.; BHATTACHARJYA, R.; BANERJEE, D. S. ApproxBioWear: Approximating Additions for Efficient Biomedical Wearable Computing at the Edge. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE BIOLOGY SOCIETY (EMBC), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.7566–7569.

KIM, C.; ROY, K. Dynamic VTH Scaling Scheme for Active Leakage Power Reduction. In: DESIGN, AUTOMATION AND TEST IN EUROPE CONFERENCE AND EXHIBITION, PROCEEDINGS, PARIS – FRANCE, 2002. **Proceedings...** [S.l.: s.n.], 2002.

KIM, N.; BLAAUW, D.; MUDGE, T. Leakage Power Optimization Techniques for Ultra Deep Sub-Micron Multi-Level Caches. In: INTERNATIONAL CONFERENCE ON COMPUTER AIDED DESIGN ICCAD, 2003. **Proceedings...** [S.l.: s.n.], 2003.

KRIVENKO, S.; LUKIN, V.; KRYLOVA, O.; SHUTKO, V. Visually Lossless Compression of Retina Images. In: IEEE 38TH INTERNATIONAL CONFERENCE ON ELECTRONICS AND NANOTECHNOLOGY (ELNANO), 2018., 2018. **Proceedings...** [S.l.: s.n.], 2018. p.255–260.

KULKARNI, P.; GUPTA, P.; ERCEGOVAC, M. Trading accuracy for power with an underdesigned multiplier architecture. In: INTERNATIONAL CONFERENCE ON VLSI DESIGN, 2011., 2011. **Proceedings...** [S.l.: s.n.], 2011. p.346–351.

KUMARI, R. S. S.; RAJALAKSHMI, E. Development of wavelet based signal compression algorithm using adaptive coder in FPGA. In: ONLINE INTERNATIONAL CONFERENCE ON GREEN ENGINEERING AND TECHNOLOGIES (IC-GET), 2016., 2016. **Proceedings...** [S.l.: s.n.], 2016. p.1–6.

KUNDI, D. E. S. et al. AxMM: Area and power efficient approximate modular multiplier for R-LWE cryptosystem. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.1–5.

LAI, D. et al. An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers. **IEEE Access**, [S.l.], v.7, p.94701–94716, 2019.

LEE, D.; KIM, Y. Towards Quantized Stochastic Computing by Leveraging Reduced Precision Binary Numbers through Bit Truncation. In: IEEE 41ST INTERNATIONAL CONFERENCE ON COMPUTER DESIGN (ICCD), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.419–422.

LEE, J.; OH, K.; KIM, B.; YOO, S. K. Synthesis of Electrocardiogram V-Lead Signals From Limb-Lead Measurement Using R-Peak Aligned Generative Adversarial Network.

**IEEE Journal of Biomedical and Health Informatics**, [S.l.], v.24, n.5, p.1265–1275, 2020.

LIU, C. et al. Signal Quality Assessment and Lightweight QRS Detection for Wearable ECG SmartVest System. **IEEE Internet of Things Journal**, [S.l.], v.6, n.2, p.1363–1374, 2019.

LIU, Z. et al. Learning Efficient Convolutional Networks through Network Slimming. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2017., 2017. **Proceedings...** [S.l.: s.n.], 2017. p.2755–2763.

MARTIN, K. Digital Integrated Circuit Design. In: NEW YORK, USA, OXFORD UNIVERSITY PRESS, 2000. **Proceedings...** [S.l.: s.n.], 2000.

MOHAPATRA, D.; CHIPPA, V.; RAGHUNATHAN, A.; ROY, K. Design of voltage scalable metafunctions for multimedia, recognition and mining applications. In: DATE, 2011. **Proceedings...** [S.l.: s.n.], 2011. p.950–955.

MOON, A.; WOO SON, S.; JUNG, J.; JEONG SONG, Y. Understanding Bit-Error Trade-off of Transform-based Lossy Compression on Electrocardiogram Signals. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.3494–3499.

NAJAFI, A. et al. Approximate Computing in Critical Applications: ECG Arrhythmia Classification. In: INTERNATIONAL CONFERENCE ON MODERN CIRCUITS AND SYSTEMS TECHNOLOGIES (MOCAS), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.1–4.

NAJM, F. A survey of power estimation techniques in VLSI circuits. In: IEEE TRANSACTIONS ON LARGE SCALE INTEGRATION (VLSI) SYSTEM, USA, 1994. **Proceedings...** [S.l.: s.n.], 1994.

OKTIVASARI, P. et al. Analysis of ECG Image File Encryption using ECDH and AES-GCM Algorithm. In: INTERNATIONAL WORKSHOP ON BIG DATA AND INFORMATION SECURITY (IWBIS), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.75–80.

OLIVEIRA, H. de. **Análise de Fourier e Wavelets (Sinais Estacionários e não Estacionários)**. [S.l.: s.n.], 2007.

P., R.; V, S.; M, R. S.; S, N. A Low Power Clock Gated Approximate Pruned and Truncated HDWT for Power-Efficient ECG Signal Processing. In: INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING AND ELECTRICAL CIRCUITS AND ELECTRONICS (ICDCECE), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.1–5.

PALEM, K. V. et al. Sustaining moore's law in embedded computing through probabilistic and approximate design: retrospects and prospects. In: COMPILERS, ARCHITECTURE, AND SYNTHESIS FOR EMBEDDED SYSTEMS, 2009., 2009. **Proceedings...** [S.l.: s.n.], 2009. p.1–10.

PATIL, V. K.; PAWAR, V. R. How can Emotions be Classified with ECG Sensors, AI techniques and IoT Setup? In: INTERNATIONAL CONFERENCE ON SIGNAL AND INFORMATION PROCESSING (ICONSIP), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–6.

PEREIRA, P. T. L. et al. Exploring Approximate Arithmetic Units for a Power-Efficient Kalman Gain VLSI Design. In: IEEE INTERNATIONAL CONFERENCE ON ELECTRONICS, CIRCUITS AND SYSTEMS (ICECS), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–4.

RABAEY, J. Digital Integrated Circuits - A Design perspective. In: USA, PRENTICE HALL, 1996. **Proceedings...** [S.l.: s.n.], 1996.

RAMASAMY, M.; NARMADHA, G.; DEIVASIGAMANI, S. Carry based approximate full adder for low power approximate computing. In: INTERNATIONAL CONFERENCE ON SMART COMPUTING COMMUNICATIONS (ICSCC), 2019., 2019. **Proceedings...** [S.l.: s.n.], 2019. p.1–4.

RANA, M. S.; HASAN, M. M.; SINHA SHUVA, S. K. Digital Watermarking Image Using Discrete Wavelet Transform and Discrete Cosine Transform with Noise Identification. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNOLOGIES (CONIT), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–4.

RATHI, N.; PANDA, P.; ROY, K. STDP-Based Pruning of Connections and Weight Quantization in Spiking Neural Networks for Energy-Efficient Recognition. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems**, [S.l.], v.38, n.4, p.668–677, 2019.

ROHIMA; BARKAH AKBAR, M. Wavelet Analysis and Comparison from Coiflet Family on Image Compression. In: INTERNATIONAL CONFERENCE ON CYBER AND IT SERVICE MANAGEMENT (CITSM), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.1–5.

ROSA, M. M. A. d. et al. AxPPA: Approximate Parallel Prefix Adders. **IEEE Transactions on Very Large Scale Integration (VLSI) Systems**, [S.l.], v.31, n.1, p.17–28, 2023.

ROSA, M. M. da et al. An Energy-Efficient Haar Wavelet Transform Architecture for Respiratory Signal Processing. **IEEE Transactions on Circuits and Systems II: Express Briefs**, [S.l.], v.68, n.2, p.597–601, 2021.

SATO, T.; HAMA, H. Evaluating Sign Error Correction for Approximate Adders Employing ECG Signal Processing. In: INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATICS (EECSI), 2023., 2023. **Proceedings...** [S.l.: s.n.], 2023. p.43–48.

SEIDEL, H. B. et al. Approximate Pruned and Truncated Haar Discrete Wavelet Transform VLSI Hardware for Energy-Efficient ECG Signal Processing. **IEEE Transactions on Circuits and Systems I: Regular Papers**, [S.l.], v.68, n.5, p.1814–1826, 2021.

SEIDEL, H. et al. Energy-Efficient Haar Transform Architectures Using Efficient Addition Schemes. In: IEEE 11TH LATIN AMERICAN SYMPOSIUM ON CIRCUITS SYSTEMS (LASCAS), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p.1–4.

SHAHRIARI, Y. et al. Electrocardiogram Signal Quality Assessment Based on Structural Image Similarity Metric. **IEEE Transactions on Biomedical Engineering**, [S.l.], v.65, n.4, p.745–753, 2018.

SHAO, S.; MCALEER, S.; YAN, R.; BALDI, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. **IEEE Transactions on Industrial Informatics**, [S.l.], v.15, n.4, p.2446–2455, 2019.

SINGH PAL, H.; KUMAR, A.; VISHWAKARMA, A.; BALYAN, L. K. A Hybrid 2D ECG Compression Algorithm using DCT and Embedded Zero Tree Wavelet. In: IEEE 6TH CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY (CICT), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–5.

VAANANEN, O.; HAMALAINEN, T. LoRa-Based Sensor Node Energy Consumption with Data Compression. In: IEEE INTERNATIONAL WORKSHOP ON METROLOGY FOR INDUSTRY 4.0 IOT (METROIND4.0IOT), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.6–11.

VERULKAR, N. M.; AMBALKAR, R. R. Discriminant analysis of Electrocardiogram (ECG) signal using cross correlation. In: INTERNATIONAL CONFERENCE ON AUTOMATIC CONTROL AND DYNAMIC OPTIMIZATION TECHNIQUES (ICACDOT), 2016., 2016. **Proceedings...** [S.l.: s.n.], 2016. p.523–526.

VIJAYAKUMARI, B.; DEVI, J. G.; MATHI, M. I. Analysis of noise removal in ECG signal using symlet wavelet. In: INTERNATIONAL CONFERENCE ON COMPUTING TECHNOLOGIES AND INTELLIGENT DATA ENGINEERING (ICCTIDE'16), 2016., 2016. **Proceedings...** [S.l.: s.n.], 2016. p.1–6.

WANG, Y.-Z.; WANG, Y.-P.; WU, Y.-C.; YANG, C.-H. A 12.6 mW, 573–2901 kS/s Reconfigurable Processor for Reconstruction of Compressively Sensed Physiological Signals. **IEEE Journal of Solid-State Circuits**, [S.l.], v.54, n.10, p.2907–2916, 2019.

WESTE, N.; ESHRAGHIAN, K. Principles of CMOS VLSI Design. In: BOSTON: ADDISON-WESLEY, 2ND.ED., 1994. **Proceedings...** [S.l.: s.n.], 1994.

WU, T. et al. A Rigid-Flex Wearable Health Monitoring Sensor Patch for IoT-Connected Healthcare Applications. **IEEE Internet of Things Journal**, [S.l.], v.7, n.8, p.6932–6945, 2020.

XIONG, C. et al. An Improved Sequence-to-Point Learning for Non-Intrusive Load Monitoring Based on Discrete Wavelet Transform. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.72, p.1–16, 2023.

XUANHUI, D.; JUAN, C.; QUAN, W. Network Pruning Based On Architecture Search and Intermediate Representation. In: INTERNATIONAL COMPUTER CONFERENCE ON WAVELET ACTIVE MEDIA TECHNOLOGY AND INFORMATION PROCESSING (ICCWAMTIP), 2021., 2021. **Proceedings...** [S.l.: s.n.], 2021. p.230–233.

YIN, Y. et al. A 2.63 W ECG Processor With Adaptive Arrhythmia Detection and Data Compression for Implantable Cardiac Monitoring Device. **IEEE Transactions on Biomedical Circuits and Systems**, [S.l.], v.15, n.4, p.777–790, 2021.

YU, J. J. Q.; HOU, Y.; LI, V. O. K. Online False Data Injection Attack Detection With Wavelet Transform and Deep Neural Networks. **IEEE Transactions on Industrial Informatics**, [S.l.], v.14, n.7, p.3271–3280, 2018.

ZANANDREA, V.; MEINHARDT, C. Exploring Approximate Computing Approaches to Design Power-efficient Multipliers. In: IFIP/IEEE 30TH INTERNATIONAL CONFERENCE ON VERY LARGE SCALE INTEGRATION (VLSI-SOC), 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.1–2.

ZHANG, B. et al. Health Data Driven on Continuous Blood Pressure Prediction Based on Gradient Boosting Decision Tree Algorithm. **IEEE Access**, [S.l.], v.7, p.32423–32433, 2019.

ZHANG, B.; ZHAO, J.; CHEN, X.; WU, J. ECG data compression using a neural network model based on multi-objective optimization. **PloS one**, [S.l.], v.12, n.10, p.e0182500, 2017.

ZHANG, G. et al. Hardware Implementation for an Improved Full-Pixel Search Algorithm Based on Normalized Cross Correlation Method. **Electronics**, [S.l.], v.7, n.12, 2018.

ZHOU, Z. et al. Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. **Proceedings of the IEEE**, USA, v.107, n.8, p.1738–1762, 2019.

ZHU, C. et al. An Efficient Hardware Accelerator for Structured Sparse Convolutional Neural Networks on FPGAs. **IEEE Transactions on Very Large Scale Integration (VLSI) Systems**, USA, v.28, n.9, p.1953–1965, 2020.