ARTICLE TYPE

Hardware implementation of IIR anti-notch filter for Exon region identification in Eukaryotic genes

Vikas Pathak*[1,2] | Satyasai Jagannath Nanda[1] | Amit Mahesh Joshi[1] | Sitanshu Shekhar Sahu[3]

[1]Malaviya National Institute of Technology, J. L. N. Road, Jaipur-302017, Rajasthan, India

[2]Swami Keshvanand Institute of Technology, Jagatpura, Jaipur-302017, Rajasthan, India

[3]Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India

Correspondence

*Vikas Pathak, 2013rec9567@mnit.ac.in

Present Address

Ph. D. Research Scholar, MNIT, Jaipur (Raj.), India

Summary

The discrimination of exons from introns in the DNA sequence of a Eukaryotic gene is important to understand the functionality of protein formation inside a living organism. Several signal processing techniques involving transforms and filtering have been used to identify the exon regions by exploring the periodicity-3 property. Fast processing of massive DNA sequences is desirable to detect the disease-causing mechanism, which is helpful to prepare individual-centric drugs. In this manuscript, a hardware implementation is carried out for the direct form-II structure of the IIR anti-notch filter to achieve the fast processing of the DNA sequence. Implementation result on Zynq-series (Zybo board) FPGA reveals that the proposed implementation is capable to identify the exon regions of five benchmark Eukaryotic genes. The FPGA implementation has achieved a maximum clock frequency of 34.629 MHz, which is further improved to 54.41 MHz using the retiming concept. Compared to MATLAB 2014a simulation the proposed FPGA implementation has achieved similar accuracy with 39 to 43 times faster computing time for the five benchmark datasets. Further, an ASIC implementation is carried out in the CADENCE RTL compiler tool with GPDK 90nm technology, due to which the hardware anti-notch filter is 120 to 133 times faster compared to its MATLAB counterpart while maintaining the comparable accuracy.

KEYWORDS:

Period-3 property, FPGA implementation, ASIC design, Anti-notch IIR Digital Filter, Eukaryotic genes, Retiming, Exon region identification.

1 | INTRODUCTION

In the last decade, the area of Genomic Signal processing (GSP)[1] has gained a lot of attention due to the multi-disciplinary researcher's involvement in determining the solution of complex problems (ex. Identification of Hereditary diseases, classification of varieties of diseases based on genetic information, to reduce the time frame involved in the innovation of new drugs, the design of individual-centric drugs and much more narrated in the literature[2]). The research in this field is relatively young and needs exploration as it covers several applications in diversified areas of medicine, bioinformatics, and agriculture. Determining significant information-carrying regions (exons) in a Eukaryotic gene is very important. The Eukaryotic gene comprises of both exons and introns (non-coding regions). The exons are the protein-coding regions responsible for the generation of proteins from DNA (Deoxyribonucleic acid) sequence. A DNA sequence consists of genes (functional regions) and inter-genetic spaces.

[0]**Abbreviations:** ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

VLSI implementation of anti-notch lattice structure for identification of exon regions in Eukaryotic genes

Vikas Pathak[1,2] ✉, Satyasai Jagannath Nanda[2], Amit Mahesh Joshi[2], Sitanshu Sekhar Sahu[3]

[1]Department of Electronics and Communication Engineering, Swami Keshvanand Institute of Technology, Jaipur-302017, India
[2]Department of Electronics and Communication Engineering, Malaviya National Institute of Technology, Jaipur-302017, India
[3]Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra-835215, India
✉ E-mail: 2013rec9567@mnit.ac.in

Abstract: In a Eukaryotic gene, identification of exon regions is crucial for protein formation. The periodic-3 property of exon regions has been used for its identification. An anti-notch infinite impulse response (IIR) filter is mostly employed to recognise this periodic-3 property. The lattice structure realisation of anti-notch IIR filter requires less hardware over direct from-II structures. In this study, a hardware implementation of IIR anti-notch filter lattice structure is carried out on Zynq-series (Zybo board) field programmable gate array (FPGA). The performance of hardware design has been improved using techniques like retiming, pipelining and unfolding and finally assessed on various Eukaryotic genes. The hardware implementation reduces the time frame to analyse the DNA sequence of Eukaryotic genes for protein formation, which plays a significant role in detecting individual diseases from genetic reports. Here, the performance evaluation is carried out in MATLAB simulation environment and the results are found similar. Application-specific integrated circuit (ASIC) implementation of the anti-notch filter lattice structure is also carried out on CADENCE-RTL compiler. It is observed that the FPGA implementation is 31 to 34 times faster and ASIC implementation is 58 to 64 times faster compared to the results generated by MATLAB platform with similar prediction accuracy.

1 Introduction

DNA molecules store digital information that constitutes the genetic blueprint of any living organism [1]. A single DNA sequence is a biomolecule, which consists of four linked smaller components (Adenine (A), Thynine (T), Cytosine (C) and Guanine (G)) termed as nucleotides. The DNA sequence is divided into two parts: Genes (which are the functional regions and responsible for protein formation from DNA sequence) and Intergenic spaces (regulatory regions). Proteins are responsible for the functionality of any living organism, which is also a biomolecule consisting of 20 – linked, smaller components called amino acids. The amino acids are made up of triplets (called as codons) of 4-possible DNA nucleotides (A, T, C & G), whose total combinations $= 4^3 = 64$ possible codons of DNA. Any protein has a particular genetic code, which maps each of the 64-possible codons of DNA character into one of the 20-possible amino acids [2]. Genes are of two types: Eukaryote (cells that have nucleolus) and Prokaryote (have no nucleus). The Eukaryote genes consist of two regions: Exons (protein coding regions responsible for protein formation from a gene of any DNA sequence) and Introns (non-coding regions). The Prokaryote genes do not have any such classification (they possess only exons-regions). In every Eukaryote gene, the Introns need to be identified and eliminated before actual protein coding (or beginning of the synthesis process). Thus there is a need to identify the exons and intron regions accurately for production of proteins from DNA sequence. The processing and analysis of DNA sequence data is helpful for discovering families of genes or gene products that can be used to classify disease, thereby leading to molecular-based diagnosis and prognosis [3].

The Exon regions reflect a period-3 property due to the existence of codon structure in protein synthesis process [4]. Several Digital signal Processing (DSP) methods have been reported by researchers to explore this period-3 property, such as: sliding DFT (STDFT) [5–10], digital filtering [11–15], continuous wavelet transform [16–18], discrete wavelet transform [19, 20] and S-transform [21]. In this paper, the focus is given on digital filtering method.

Vaidyanathan and Yoon [11] used second-order digital filters like anti-notch and multi-stage filters for protein-coding region identification. Ramachandran et al. [12] implemented the narrow band-pass digital filter and low pass filter for exon region detection and reported lower computational time compared to STDFT [10] based approach. An improved comb filter-based approach is reported in [13] for detection of Exon region with better prediction accuracy and less computational complexity. For the same problem, Hota and Srivastava [14] had used three anti-notch IIR filters: conjugate suppression anti-notch filter (suppresses the conjugate frequency component), anti-notch filter followed by moving average filter (decrease the background noise) and harmonic suppression anti-notch filter (removes the harmonic frequency component). This model improved the prediction accuracy of proteins Exon coding regions identification. Recently, the modified conjugate suppression anti-notch IIR filter and linear predictive coding model are combined by Farsani et al. [15] to propose an efficient algorithm for improving the performance of conventional Goertzel algorithm in determining the protein-coding regions.

Lattice filter structures are generally used to implement the finite and infinite impulse response (FIR and IIR) filters. They have several advantages over direct form structures, such as [22]: lesser sensitivity to coefficients quantisation and roundoff errors; architecture is modular in nature (i.e. an extra stage can be easily added if the order of the filter is increased [23]); in the hardware implementation, there is a requirement of the lower number of multipliers. The lattice structures are also effective as predictors/ identifiers (as the structure simultaneously generates the forward and backward prediction errors). Vaidyanathan and Yoon [11] have also shown the cascaded lattice structures [24, 25] of IIR anti-notch digital filters. Designing of lattice structures (with minimum multipliers) of notch filter from all-pass digital filters is explained by Regalia et al. [26].

In the literature, this Exon region identification problem has been simulated with MATLAB software by many authors, but sufficient work has not been done in its hardware implementation. The hardware design is necessary to identify the gene