Ain Shams University
Faculty of Engineering
Computer and Systems Engineering

On the Hardware Acceleration of Bioinformatics Algorithms

A Thesis submitted in partial fulfilment of the requirements of the degree of
Master of Science in Electrical Engineering
(Computer and Systems Engineering)

by

Reem Mohamed Khairy Mostafa AbdelFattah
Bachelor of Science in Electrical Engineering
(Computer and Systems Engineering)
Faculty of Engineering, Ain Shams University, 2012

Supervised By

Prof. Dr. Mohamed Watheq Ali Kamel El-Kharashi

Dr. Mona Mohamed Hassan Safar

Cairo - (2018)

Ain Shams University
Faculty of Engineering
Computer and Systems Engineering

Examiners Committee

Name: Reem Mohamed Khairy Mostafa AbdelFattah

Thesis: On the Hardware Acceleration of Bioinformatics Algorithms

Degree: Master of Science in Electrical Engineering
(Computer and Systems Engineering)

Name and affiliation	Signature
Prof. Dr. Hassen Taher Dorrah Prof. at Electrical Engineering Faculty of Engineering, Cairo University	…………
Prof. Dr. Ayman Mohamed Mohamed Hassan Wahba Prof. at Computer and Systems Engineering Faculty of Engineering, Ain Shams University	…………
Prof. Dr. Mohamed Watheq Ali Kamel El-Kharashi Prof. at Computer and Systems Engineering Faculty of Engineering, Ain Shams University	…………

Date: 17 February 2018

Statement

This thesis is submitted as a partial fulfilment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain shams University.

The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Reem Mohamed Khairy Mostafa AbdelFattah

Signature

…………..……….

Date: 17 February 2018

Researcher Data

Name: Reem Mohamed Khairy Mostafa AbdelFattah

Date of birth: 01/07/1990

Place of birth: Cairo, Egypt

Last academic degree: Bachelor of Science

Field of specialization: Electrical Engineering

University issued the degree: Ain Shams University

Date of issued degree: 2012

Current job: Senior Software Development Engineer at Mentor, a Siemens business

Abstract

Scanning genomic sequence databases is a very common task in bioinformatics to lookup for similarities between the query sequence and particular database sequences. The speedup of these searches became a must to keep up with the rapid growth in gene banks; where every year their size is scaled by a factor of 1.5 to 2. The DNA Basic local alignment search tool (BLASTN) is one of the most widely used tool in analysis of DNA sequences. This tool uses heuristics approach to accelerate the time consuming search algorithm. Although BLASTN is a highly optimized algorithm, the growth of biological databases outpaces its speed improvement.

There are several acceleration approaches adopted in BLASTN. Some of these approaches try to accelerate BLASTN in software using more efficient algorithms or heuristics. Other approaches use a hybrid software/hardware architectures. Some of these approaches design hardware accelerators for BLASTN. Since the implementation of BLASTN can benefit from the acceleration through fine-grained and coarse-grained parallelism, the acceleration using hardware is an appealing approach. Similar to any hardware design process, the design process of the BLASTN accelerator is laborious and error prone. Moreover, the algorithm complexity causes the design process to be more complicated. Therefore, there is a rising need to find a more convenient tool for complex systems design.

Raising the abstraction level in the design process became very important in order to cope with the increasing system design complexities. High level synthesis (HLS) introduces such high abstraction level that is essential for efficient modern system design in different domains. The capabilities of HLS help designers produce efficient system in terms of performance and cost by simply writing its functionalities using high level languages, such as C, C++, or SystemC.

The main aim of this research is developing a hardware accelerator for BLASTN algorithm using a more efficient designing tool. Therefore, we present an HLS solution to accelerate the BLASTN algorithm. Our solution implements the BLASTN stages using C++ as a high level language. The HLS tool automatically transforms the implementation into register transfer level

(RTL). The hardware implementation is generated without any knowledge of the RTL design in general and without going through the tedious design process. Also, we implemented an optimized version from our HLS solution using the pipeline directive provided by the HLS tool.

Our experiments show that the use of HLS to directly implement BLASTN is very efficient and improves both the latency and throughput of the algorithm. The HLS implementation achieves a latency and throughput speedup around 20x. Further acceleration is achieved by applying the pipeline directive to optimize the HLS implementation, which yields a latency and throughput speedup around 5x. A comparison versus software implementation shows that an overall latency and throughput speedup of 100x is achieved. We augmented our results with a throughput comparison against similar and previous hardware-accelerated algorithms. Our HLS BLASTN achieves an average throughput speedup of 70x over the NCBI implementation. Furthermore, our HLS BLASTN outperforms the Mercury BLASTN implementation by 11x.

Summary

Alignment is a very common operation in bio-sequence analysis applications. The target of the alignment operation is finding similarities between query and database sequences. The Basic Local Alignment Search Tool for DNA (BLASTN) is a fundamental bioinformatics algorithm used in bio-sequence analysis applications. This algorithm is highly optimized to accelerate the search operations. However, the exponential increase in size of gene banks outpaces the speed improvement of BLASTN. There is a rising need to accelerate BLASTN algorithm in order to accommodate the rapid growth in the size of bio-sequence databases.

Several acceleration approaches have been done in order to enhance the performance of BLASTN algorithm. Among these approaches, the hardware accelerators achieve the best performance. However, the design process of hardware accelerator for such complex algorithm as BLASTN requires hard-ware expert designer. To simplify the design process for all designers, it is required to increase the abstraction level. High Level Synthesis (HLS) represents this required level of abstraction for efficient modern systems design. HLS allows designers to write the algorithm behavior in high level languages without worrying about the implementations details, such as clock, processes, and technology. Then, the HLS tool automatically transforms these high level specifications into efficient RTL design.

We present in this thesis a HLS solution to design FPGA-based accelerator for the BLASTN algorithm. Our solution implements the BLASTN stages using a high level language. The HLS tool automatically transforms the implementation into register transfer level (RTL). The hardware implementation is generated without any knowledge of the RTL design in general and without going through the tedious design process.

The thesis contains seven chapters. A brief description about the contents of each chapter is presented hereinafter.

Chapter 1 introduces the thesis. The background and the motivations of the proposed solution are highlighted. Also, we summarize the contributions of this research.

Chapter 2 presents the implementation details of BLASTN, a fundamental bioinformatics algorithm.

Chapter 3 reviews several acceleration attempts of the BLASTN algorithm on different platforms.

Chapter 4 introduces the proposed design methodology adopted in designing hardware-accelerated BLASTN. This chapter discusses the advantages and design flow of HLS design methodology.

Chapter 5 presents the implementation details of our proposed HLS solution. We present two implementations in C++. The first one is a direct implementation of BLASTN algorithm. The second implementation is optimized using HLS directives.

Chapter 6 discusses the experimental results of implementing BLASTN algorithm using HLS. This chapter shows the latency and throughput speedup achieved from the direct and optimized HLS implementations of BLASTN algorithm. Also, it compares these results with software implementation. A comparison versus previous hardware acceleration attempts is listed.

Chapter 7 summarizes the thesis and provides our contributions. Also, an overview on the future work is introduced.

Keywords:

Bioinformatics, BLASTN, Bloom filter, Hardware acceleration, High Level Synthesis, Reconfigurable computing

Acknowledgment

I dreamed a lot about the moment, where writing the acknowledgment of my thesis, and I worked so hard to reach it. First and last, all thanks Allah for giving me the strength and ability to fulfil my dream. May Allah accept this work from me as “science that benefits”.

I would like to express my sincere gratitude to Prof. Dr. Mohamed Watheq El-Kharashi for his continuous support, encouragement, and patience. It is an honor to work under his supervision.

Also, I would like to recognize the continuous support and motivation of Dr. Mona Safar. Thanks for being patient and kind in all the phases of thesis work. All her valuable comments helped me improving my work.

My kind parents, I cannot get enough thanking both of you for being so supportive, materially and morally. Without your prayers, I would never have reached what I am today.

My lovely little sister, thanks for helping me in formatting some sections of this thesis and for your willingness to do anything just to support me.

My dear husband, I would never thank you enough for your endless support, motivation and love. Also, I would like to apologize for making this whole year like hell for you and thank you for bearing me.

Unfortunately, this page is not enough to thank all people to whom I am really grateful. All thanks to everyone helped me by a simple advice, motivated me to continue, checked on me, and understood my situation in the last period.

Reem Mohamed Khairy Mostafa AbdelFattah
Computer and Systems Engineering
Faculty of Engineering
Ain Shams University
Cairo, Egypt

February 2018

Table of Contents