DESIGN AND ANALYSIS OF HIGH PERFORMANCE MATRIX FILLING FOR DNA SEQUENCE ALIGNMENT ACCELERATOR USING ASIC DESIGN FLOW

This thesis is presented in partial fulfillment for the award of the

Bachelor of Electrical Engineering (Hons)

UNIVERSITI TEKNOLOGI MARA

MALAYSIA

NURZAIMA BT MAHMOD

2007270622

Faculty of Electrical Engineering

UNIVERSITI TEKNOLOGI MARA

40450 SHAH ALAM, SELANGOR

ACKNOWLEDGEMENTS

All praise is to Allah S.W.T, The Most Gracious and Most Merciful who has given me the strength, ability and patience to complete this project. I would like to thank many people who made my life at Universiti Teknologi MARA so memorable and continuously encouraged me to complete my study toward my Bachelor's Degree.

First and foremost, I would like to express my sincere gratitude and appreciation to my project supervisor, Puan Norhazlin Binti Khairudin, my co-supervisor Mr. Abdul Karimi Halim and other contributors for their guidance, support and helpful advices to develop the Design and Analysis of High Performance Matrix Filling for DNA Sequence Alignment Accelerator Using ASIC Design Flow. Also, thanks to all my fellow colleagues for their invaluable support and motivation given either directly or indirectly towards the completion of this project. Special thanks to the staff of Faculty of Electrical Engineering for their assistance for the use of the facilities at the laboratory.

Last but not least, my deepest appreciation goes to my beloved family, especially my father and mother who are my greatest source of inspiration, for their moral support, motivation and understanding towards the accomplishment of this project.

.

ABSTRACT

This report presents the design and analysis high performance matrix filling for DNA sequence alignment accelerator using ASIC design flow. The objective of this paper is to design and analysis matrix module of DNA sequence alignment accelerator using clock cycle to get high performance. The scope of this paper is to optimize the DNA sequences alignment on the matrix filling module by implementing a parallel method of the SmithWaterman algorithm. This method provides magnificent speed up over than traditional sequential implementation methods while it sensitivity detection is still remained. To optimize the performance of the algorithm by exploiting parallelism in the design several techniques have been developed. In the advanced engineering technology, the massive parallelism can be implemented by using the Field Programmable Logic Array (FPGA) techniques. The design was developed in Verilog HDL coding and synthesis by using LINUX tools. From the LINUX tools, the optimum combination of parameters is manipulated to produce the most energy efficient IC. The design produces an ASIC that can work at 5ns until 10ns clock period and range of ICC time between 0.63ns until 1.67ns. The area of this design is 10304.358um^2.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Biology is in the middle of a major paradigm shift driven by computing technology. A new hybrid field (partly molecular biology and partly computer science) began to emerge was called computational molecular biology [1]. In most common terms sequence alignment may be defined as an arrangement of two or more DNA sequences to highlight the regions of their similarity. This in turn indicates the genetic relatedness between the organisms. Deoxyribonucleic acid (DNA) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses [2]. The main role of DNA molecules is the long-term storage of information.

The problem of sequence similarity is recurrent in the fields of Genetics and Bioinformatics. When a new gene is discovered its role and function may be inferred by its similarity to known sequences. In genetics the focus is usually on sequences of nucleic acids drawn from the set of four possibilities Adenine (’A’), Guanine (’G’), Cytosine (’C’) and Thymine (’T’) for DNA [3]. A typical problem will search for the best match of a query string inside a much larger database string. Methods for solving this problem with various additional boundary conditions are well known and drawn from the class of dynamic programming methods. Smith-Waterman is a dynamic programming