DESIGN AND ANALYSIS OF DNA SEQUENCE ALIGNMENT MODULE USING SMITH-WATERMAN SCORING PATTERNS

This project report is presented in partial fulfilment for the award of the

B.Eng. (Hons) Electronic

WAN ABDUL QAYYUM MOH SALLEH

FACULTY OF ELECTRICAL ENGINEERING

UNIVERSITI TEKNOLOGI MARA

40450 SHAH ALAM, SELANGOR

ACKNOWLEDGEMENT

Firstly, I would like to thank Allah the Almighty for His continuous blessing. I would like to express my deepest appreciation to all those who helped me in conducting my Final Year Project and provided me the possibility to complete this report. It would not have been possible without the kind of support and help of many individuals.

I am highly indebted to my supervisor, Mr. Abdul Karimi Halim for his guidance and constant supervision as well as for providing necessary informations regarding the project and also for his supports in completing the project.

A special thanks to all my friends and classmates, mainly Mohd Mustakim Subri, Nur Amalina Abu Seman and Muhsin whose words and suggestions have generated ideas to improve my report. Special thanks and best regards given to them for acknowledging their supports to me in successfully completing my Bachelors Engineering (Hons) Electronics.

Also thanks to my family and friends who tried their best to give their supports to me by giving me a lot of encouragement to keep up with this project or by supporting financially in order to complete my Final Year Project.

Last but not least, I would like to apologize to any individuals whose names I have failed to mentioned but had helped me through the project since I am a poor servant of Allah.

ABSTRACT

This project is to design and analyze the DNA sequence alignment module using Smith-Waterman scoring patterns. The main objective is to create SW pattern in matrix using Excel based equation. The second objective is to construct the module for SW sequence alignment using the matrix pattern. One of the major problems rose from DNA sequence alignment is the speed and accuracy. The computational demand needed to explore for analyze the data faster and accurate in these databases is quickly becoming a great concern. This system will be using Verilog language and optimize it using pattern recognition in order to overcome the accuracy issues. The system optimizes the aligning DNA fragment using Smith-Waterman algorithm with a pattern recognition algorithm. FPGA design, synthesis and simulation are done using the Xilinx ISE software and obtain the RTL schematic as well as the waveform of the module. Synopsys tools are used for analysis the project in ASIC design. VCS for verification the design for further process in ASIC design tools are used. Design Compiler is for resynthesize and remodelling the design with setup constraints. Static timing analysis using Prime Time is for advance timing analysis. Based on the result obtained, the minimum frequency that is suitable for this project is 50 MHz. The average power consumption for normal compile high compile and ultra-compile are 53 uW, 52.8 uW and 49.8 uW. The average Cell area for normal compile high compile and ultra-compile are 18700 um^2, 18280 um^2 and 9532 um^2.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Some organisms have multiple copies of chromosome and other kind of acid. The secret within DNA still a mystery among scientist. Most of the researchers believe that DNA contains a numerous of information such as life span, genetic disease, characteristic and many more. DNA sequencing is a method or process in determining the precise order of nucleotides in DNA strand. There are various methods for DNA sequence alignment and due to these advent rapids methods, it accelerated biological and medical research and discovery. Most of the problems in sequencing the DNA sequence alignment are large execution time. Current methods have reduced the time taken to sequence but it is still lack in accuracy. Hence, Smith-Waterman algorithm will be used to improve and optimize speed.

1.2 PROBLEM STATEMENTS

Since the discovery of DNA by Francis Harry Compton Crick and James D. Watson, scientists and researchers have sought to measure or best way in alignment and DNA sequence.

i. Problems for the DNA sequence alignment is speed. Various methods have been introduced and many algorithms have been designed to unlock and solve this problem. Genome data base is increasing and this makes the process of calculating the alignment of DNA sequences slower.