

**DESIGN AND DEVELOPMENT OF OPTIMAL PATH TRACE BACK  
USING GRAPH THEORY TECHNIQUE FOR ACCELERATE DNA  
SEQUENCE ALIGNMENT ACCELERATOR**

**NOR SHUHaida BINTI OTHMAN**

**FACULTY OF ELECTRICAL ENGINEERING  
UNIVERSITI TEKNOLOGI MARA  
MALAYSIA  
MAY 2010**

## ACKNOWLEDGEMENT

Assalammualaikum,

In this valuable opportunity, I would like to thank Allah the Almighty because of his love for giving me chance to finish this thesis in order to accomplish my study. I wish to express my deepest gratitude and appreciation to my project supervisor, Mr. Syed Al Junid Al Mutalib for giving consistent guidance and opinion as well as prevision of his valuable time, encourage and patient for my research paper to ensure this project is successfully completed. Thanks to Mr. Fairol and Miss. Hanani for their willingness to evaluate my project presentation. Not forgetting to all my friends especially Mohd Hidayatullah Ibrahim and Nor Hafizah who always willing to lend a helping hand such as give some ideas and been helping a lot technically and mentally while I am doing my research study. We have successfully gone through together all the hard time and at last managed to accomplish this report study. Most of all, this project is especially dedicated to my beloved family who have always pray for my success, assisted me, giving me courage most, morale support as well as materially, love and understanding during the course of my study in UiTM. Not forgotten, a lots of thankful for those who involved in directly for their support, understanding, help and advice. Lastly, hope this report will satisfy all people and may Allah let you in his sight and bless you always.

Thank you.

Nor Shuhaida Binti Othman  
2006687944  
Faculty of Electrical Engineering  
Universiti Teknologi MARA  
40450 Shah Alam  
Selangor

## **ABSTRACT**

This paper presents the improvement of Smith-Waterman algorithm in comparing between two DNA sequence alignments. The existing Smith-Waterman algorithm had complex memory and low performance where the time also was consumed caused by high sensitivity of the large sequences long. This new technique is been developpe in order to overcome these problem. The Verilog HDL coding is been written using Quartus 2 version 9.1. The Smith-Waterman is been simplified into four modules which were initialization, score calculation, matrix filling and optimal path. The scope of the paper based on optimal path trace back using graph theory. The kind of graph theory that been used was indirected graph with linear algebra. This method had improved the accuracy of trace back and the initialization modules had achieved 93.75% reduction in memory space requirement.

## **TABLE OF CONTENTS**

<b>CONTENTS</b>	<b>PAGE</b>
<b>DECLARATION</b>	i
<b>ACKNOWLEDGEMENT</b>	ii
<b>ABSTRACT</b>	iii
<b>TABLE OF CONTENTS</b>	iv
<b>LIST OF TABLES</b>	vi
<b>LIST OF FIGURES</b>	vii
<b>ABBREVIATION</b>	viii
<b>CHAPTER 1 : INTRODUCTION</b>	
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Objectives	3
1.4 Scopes of the Project	3
<b>CHAPTER 2 : LITERATURE REVIEW</b>	
2.1 DNA	5
2.1.1 Introduction	5
2.1.2 The Function of DNA	6
2.1.3 Property of DNA	9
2.2 DNA Sequence Alignment	10
2.3 Smith-Waterman Algorithm	12
2.3.1 Overview	12
2.3.2 The History Of the Smith-Waterman Algorithm	13
2.3.3 Dynamic Program	13
2.3.4 Graph Theory	14
2.4 Target Board (FPGA)	14

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 BACKGROUND OF STUDY**

DNA Sequence Alignment is a very important area of study and research in the field of Bioinformatics. Bioinformatics is applied in the creation and maintenance of a database to store biological information at the beginning of the genomic revolution [1] such as nucleotide and amino acid sequences. The growth of this type of database involved the development of complex interfaces instead of design issues. It means researchers could both access existing data as well as propose new or revised data.

In general, this research is based on studies of ways to compare between two strings (DNA sequences alignment). This comparison is to verify the percentage of similarity between both sequences and made the conclusion based on the findings. Researcher have developed many techniques to analyze the sequences however the demand for susceptible and high performance DNA sequence alignment tools for DNA sequences analysis study and investigation keep increased year by year. Nowadays DNA researchers require tools to effectively and efficiently analyze and develop the existing technique. The processing of results need to be more efficient and accurately as demand. Intelligent bioinformatics and flexible methods are the most important factors for success in science these days.