DESIGN AND ANALYSIS OF 4x4 SMITH WATERMAN'S BASED DNA ALIGNMENT ACCELERATOR USING MULTICORE ARCHITECTURE

NUR FARIZA BINTI MOHD NAJIB

FACULTY OF ELECTRICAL ENGINEERING
UNIVERSITI TEKNOLOGI MARA
MALAYSIA

ACKNOWLEGMENT

First and foremost I offer my sincerest gratitude to my supervisor, Mrs Syazilawati Mohamed, who has supported me throughout my thesis with her patience and effort and without her this thesis, would not have been completed or written. I am also deeply grateful to my supervisor, for her detailed and constructive comments, and for her important support all the way through this project. One simply could not wish for a better or friendlier supervisor. I am so grateful to all of people with whom I have had the pleasure to work during this project especially like to thank Mr Karimi Halim, for being my teacher and mentor, he has taught me more than I could ever give him credit for here. He has shown me, by his example, what a good ASIC designer should be. Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents, whose love and guidance are with me in whatever I pursue. They are the ultimate role models for me by providing unending inspiration. In my daily work I have been blessed with a friendly and cheerful group of fellow electronic students. Any kind support and guidance have been of great value in this study. Thank you.

ABSTRACT

This paper presents the design and analysis of 4x4 Smith Waterman's based DNA sequences alignment accelerator using multicore architecture. The first objective of this paper is to design 4 x 4 matrix filling module of DNA sequence alignment accelerator. The second objective is to combine matrix filling module with traceback and reconstruction module design by previous student in order to create single core. The third objective is to construct modules that consist of 2, 4 and 8 cores. The fourth objective is to simulate and verify the functionality of the new matrix filing, single and multicore modules on Xilinx FPGA design flow. The last objective is to verify, optimize and analyze the multicore modules using VCS, DC, ICC and PT. This paper focuses on the timing analysis, power and area of multicores architecture using Smith-Waterman algorithm. To achieve a higher performance with low latency and high throughput data has become a serious concern for today's DNA laboratory as the increase of number of DNA database all around the world. Researchers may have done different kind of architecture such as pipeline, vector, multicycle to boost up the speed performance of DNA sequences alignment. All design is written in Verilog language and the result is verified on Synopsys EDA tool which are Verilog Compiler Simulator (VCS), Design Compiler (DC) and ICCompiler (ICC).

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Bioinformatics is relates to the science fields such as biology, computer science and technology that merge into a single discipline. Within bioinformatics, there are also three importance sub-disciplines and they are the development of new algorithms and statistics with which to assess relationship among members of large data sets, the analysis and interpretation of various types of data including nucleotide and amino sequences and the development and implementation of tools that enables efficient access and management of different types of information. Bioinformatics is widely used in human genome research by developing computer database and algorithm to facilitate biological research. Genetic sequence searching has grown immensely throughout the years, is highly computationally intensive and requires high processing power. The Smith Waterman algorithm which adopts dynamic programming can be used to address the above issue as it is accurate and can be implement to be high in performance. However, it is inefficient referring to computational complexity, the need of memory and Smith Waterman analysis makes this algorithm difficult to be implemented. The Smith-Waterman algorithm is a well-known algorithm for performing local sequence alignment that is, for determining similar regions between two nucleotide or protein sequences.