UNIVERSITI TEKNOLOGI MARA

EFFICIENT USAGE STRATEGY OF LIMITED SHARED MEMORY IN GRAPHICAL PROCESSING UNIT (GPU) FOR ACCELERATE DNA SEQUENCE ALIGNMENT

AHMAD HASIF BIN AZMAN

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science
(Electrical Engineering)

College of Engineering

September 2023

ABSTRACT

DNA sequence alignment is expected to help uncover important information about the human body, disease, genetics and other biological relationships when discovered. In addition, intensive efforts have been made to improve the performance of sequence alignment through hardware-based acceleration using the Graphical Processing Unit (GPU) accelerator. This implementation is becoming increasingly popular due to the flexibility of the accelerator design, parallel computational solutions and the ability to simultaneously increase the performance of the alignment. The performance of the DNA sequence alignment system is highly dependent on the algorithm, GPU designed architecture and accelerator performance. In this study, the focus is on utilizing the memory capabilities of GPUs to accelerate the Smith-Waterman algorithm has been proposed. Three new approaches based on global memory, shared memory and a combination of global and shared memory are used in this design. Moreover, the execution time proves that the design is able to speed up the computational process by about 90% compared to the Central Processing Unit (CPU). Again, the result proves that the acceleration of the GPU is able to speed up the processing of the DNA sequence alignment without affecting the result. Finally, the results obtained have shown that the proposed system offers better performance and design than previous work on accelerating SWA DNA sequence alignment using GPU accelerators.

ACKNOWLEDGEMENT

First and for the most, I wish to thank God for giving me the opportunity to embark on my MSc and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Assoc Prof. Ts. Ir. Dr. Syed Abdul Mutalib Al Junid kind guidance and support throughout the work. I would also like to thank Assoc Prof. Ts. Dr. Abdul Hadi bin Abdul Razak for his comments and suggestions on some of the ideas of my research work as co-supervisor.

Last but not least, my deepest appreciation goes to my family, which has always encouraged me with invaluable love. Without their encouragement and support, I would not be able to complete and accomplish this work.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

1.1   Research Background

In the last decade, molecular biology research in the field of deoxyribonucleic acid (DNA) has taken another leap forward, as the application of DNA can be used to reveal information about living organisms. In terms of health, the focus on genetic testing can provide data on genetic and other diseases at the cellular level [1]. Recently, several research have also been conducted using DNA to find out the roots of our ancestry [2]–[4]. While a person's outward appearance can be used as an indicator of race, digging deeper into DNA can change the genetic outcome for the appearance of a particular human race [5]. There is also a collaboration between Microsoft and Washington University using DNA molecules as data storage, showing the importance of DNA applications in the near future [6].

Nowadays, molecular biology research on DNA sequences is one of the most important studies [7], [8]and scientists estimate that nearly 150 species become extinct every day due to pollution and greenhouse effect [9]. In order to prevent species extinction, the Earth Bio genome Project was launched as a countermeasure. Within the framework of this project, the DNA sequences of 1.5 million species around the world are being mapped [10].

Basically, DNA is located in the nucleus or nuclear DNA in a person's body and is stored as a code consisting of chemical bases or nucleotides. There are several tools you can use to analyse the DNA sequence, including the Alignment Tool. In addition, the Alignment Tool analyses the properties of DNA molecules containing adenine (A), cytosine (C), guanine (G) and thymine (T) [11], [12], and by combining these bases, DNA strands are formed.

In the history of molecular biology, intensive research in the field of DNA sequences brought another breakthrough when it was discovered that DNA molecules contain various vital information of living organisms. Since this discovery, DNA