

Genetic sequence alignment acceleration using a FPGA based platform

What methods can be used to accelerate the Smith-Waterman algorithm for genetic sequence alignment with an FPGA equipped platform?

Robin NOLLET

Supervisors: Ing. Václav Šimek
: Ing. Jonas Lannoo

Master Thesis to obtain the degree of
Master of Science in Engineering Technology:
Electronical engineering

Academic Year 2019 - 2020



©Copyright KU Leuven

Without written permission of the supervisor(s) and the author(s) it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilise parts of this publication should be addressed to KU Leuven Campus Brugge, Spoorwegstraat 12, B-8000 Brugge, +32 50 66 48 00 or via e-mail iiw.brugge@kuleuven.be.

Acknowledgements

Het voorwoord vul je persoonlijk in met een appreciatie of dankbetuiging aan de mensen die je hebben bijgestaan tijdens het verwezenlijken van je masterproef en je hebben gesteund tijdens je studie.

Summary

De (korte) samenvatting, toegankelijk voor een breed publiek, wordt in het Nederlands geschreven en bevat **maximum 3500 tekens**. Deze samenvatting moet ook verplicht opgeladen worden in KU Lokaal.

Abstract

Het extended abstract of de wetenschappelijke samenvatting wordt in het Engels geschreven en bevat **500 tot 1.500 woorden**. Dit abstract moet **niet** in KU Loket opgeladen worden (vanwege de beperkte beschikbare ruimte daar).

Keywords: Voeg een vijftal keywords in (bv: Latex-template, thesis, ...)

Contents

Acknowledgements	iii
Summary	iv
Abstract	v
Table of contents	vi
List of figures	vii
List of tables	viii
List of symbols	ix
List of abbreviations	x
1 Introduction	1
2 Introduction to bioinformatics and embedded systems	2
2.1 Introduction to bioinformatics	2
2.1.1 Biology and DNA	2
2.1.2 sequencing and the need for Bioinformatics	2
2.1.3 DNA sequence aligning	2
2.1.4 Clinical applications	2
2.2 Platforms for sequence alignment algorithms	2
2.2.1 CPU	2
2.2.2 GPU	2
2.2.3 FPGA	2
2.2.4 ASIC	2
2.3 Problem definition	2

3	Methods for genetic sequence alignment	3
3.1	Local VS global alignment	3
3.2	commonly used algorithms	3
3.2.1	Dynamic programming algorithms	3
3.2.2	Heuristic algorithms	3
3.3	algorithm selection	3
3.3.1	Smith Waterman	3
4	Reference mapping accelerated	4
4.1	problems with de direct approach	4
4.2	acceleration techniques	4
5	System implementation for reference genome mapping	5
6	implementation results and speedup	6
7	Conclusion and future research	7
A	Uitleg over de appendices	9

List of Figures

List of Tables

List of symbols

Maak een lijst van de gebruikte symbolen. Geef het symbool, naam en eenheid. Gebruik steeds SI-eenheden en gebruik de symbolen en namen zoals deze voorkomen in de hedendaagse literatuur en normen. De symbolen worden alfabetisch gerangschikt in opeenvolgende lijsten: kleine letters, hoofdletters, Griekse kleine letters, Griekse hoofdletters. Onderstaande tabel geeft het format dat kan ingevuld en uitgebreid worden. Wanneer het symbool een eerste maal in de tekst of in een formule wordt gebruikt, moet het symbool verklaard worden. Verwijder deze tekst wanneer je je thesis maakt.

b	Breedte	$[mm]$
A	Oppervlakte van de dwarsdoorsnede	$[mm^2]$
c	Lichtsnelheid	$[m/s]$

List of abbreviations

Chapter 1

Introduction

blabla blabla tekst tekst.

Chapter 2

Introduction to bioinformatics and embedded systems

2.1 Introduction to bioinformatics

2.1.1 Biology and DNA

2.1.2 Sequencing and the need for Bioinformatics

2.1.3 DNA sequence aligning

2.1.4 Clinical applications

2.2 Platforms for sequence alignment algorithms

2.2.1 CPU

2.2.2 GPU

2.2.3 FPGA

2.2.4 ASIC

2.3 Problem definition

Chapter 3

Methods for genetic sequence alignment

3.1 Local VS global alignment

3.2 commonly used algorithms

3.2.1 Dynamic programming algorithms

3.2.1.1 Needleman-Wunsch

3.2.1.2 Smith-Waterman

3.2.2 Heuristic algorithms

3.2.2.1 FASTA

3.2.2.2 BLAST

3.3 algorithm selection

3.3.1 Smith Waterman

Chapter 4

Reference mapping accelerated

4.1 problems with the direct approach

4.2 acceleration techniques

Chapter 5

System implementation for reference genome mapping

...

Chapter 6

implementation results and speedup

...

Chapter 7

Conclusion and future research

...

Bibliography

- Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Khalek, S. A., Abdelalim, A., Abdinov, O., Aben, R., Abi, B., Abolins, M., et al. (2012). Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29.
- Cottrell, J. A., Hughes, T. J., and Bazilevs, Y. (2009). *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons.
- Hughes, T. J., Cottrell, J. A., and Bazilevs, Y. (2005). Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement. *Computer methods in applied mechanics and engineering*, 194(39):4135–4195.

Appendix A

Uitleg over de appendices

Bijlagen worden bij voorkeur enkel elektronisch ter beschikking gesteld. Indien essentieel kunnen in overleg met de promotor bijlagen in de scriptie opgenomen worden of als apart boekdeel voorzien worden.

Er wordt wel steeds een lijst met vermelding van alle bijlagen opgenomen in de scriptie. Bijlagen worden genummerd met een drukletter A, B, C,...

Voorbeelden van bijlagen:

Bijlage A: Detailtekeningen van de proefopstelling

Bijlage B: Meetgegevens (op USB)

FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN
CAMPUS BRUGGE
Spoorwegstraat 12
8200 BRUGGE, België
tel. + 32 50 66 48 00
iiw.brugge@kuleuven.be
www.iw.kuleuven.be

