The University of York   Department of Computer Science

Submitted in part fulfilment for the degree of MEng.

Parallel Programming Tools for Exploring Immune System Development

Oliver Binns

28th January 2018

Supervisor: Dr. Kieran Alden, Dr. Fiona Polack

Number of words = -1, as counted by wc -w.
This includes the body of the report only.

Abstract

More powerful computers are paving the way for realistic simulations of previously underexplored biological systems. As advancements in computing tend increasingly towards parallelism and distributed systems, these ever more complex simulations must take full advantage of this change in order to be computed in a reasonable time. In this project, we will explore some of the current technologies for doing so and propose new tools for the future.

Contents

Contents

List of Figures

List of Tables

1 Introduction

1.1 Project Overview

This project has been carried out with regard to the set of ethical guidelines set out by the University of York. This project does not involve human participants, so guidelines on informed consent and confidentiality will be met. The simulation of the biological model is for the purpose of developing understanding of applying GPGPU methods to an agent based model. It will not be used to publish results, in its current form and does not simulate a biological process.

TALK ABOUT MODELING AND THE DESIRE FOR Faster AND More Accessible Modelling

1.2 Ethics

No confidential medical data or personal information has been used during the course of the project development. This project has involved no animal or human participation.

1.3 Report Structure

..?

2 Background

2.1 Simulation

Simulations are used in a wide range of disciplines on applications such as .. nuclear weapons, medical.., video games, product development. Within scientific discovery, simulations may be used alongside existing real-world testing or an alternative to it. Often real-world testing may not be feasible for a number of reasons including time, cost, danger (nuclear weapons) or scale. Moral or legal aspects may also play a significant limitation here.

There are some significant limitations as to where simulations can be used.

Bugs

Skills Shortage Significant skills shortages in Computer Science, across the world, significantly limit the possibility for cross-disciplinary work to utilise

Insignificant Domain Knowledge

Compute Power

2.2 Biological Simulations

Simulations have even been proposed as a method for exploring a potential set of first principles and mathematics that are specific to biology which could even constitute a new subject- theoretical biology[1].

10

2.2.1 Existing Simulations

2.2.2 Cell Dynamics

2.2.3 PPSim

As these simulations becoming more and more prevalent: tools need developing to allow less technical users to create these simulations easier and run faster...

This project focuses on simulation as a tool for exploring biological systems at cell level. It uses the existing simulation of Peyer's Patch[2] and attempts to use parallel computer architectures in order to speed this simulation up. Finally, I will propose a new tool, which builds on existing work in order to make this power available to non-technical users?

3 Improving Simulations

3.1 Ease of Creation

Ease of creation (and maintenance) is an important feature for future simulation particularly due to the aforementioned computer science skills shortage (Section 2.1).

3.1.1 Flexible Modelling

Flexible Modelling tools could be a good avenue for [3]

3.2 Speed Up

3.2.1 Machine Learning

A solution to the speed problem that has been proposed recently is to use machine learning on a small set of results to produce This has problems in that... Standard Machine Learning issues? Bias? Overfitting?

3.2.2 Parallelism

Parallelism fundamentally changes the game and allows computers to keep following Moore's law even has engineers are struggling to make transisters ever smaller and smaller.[citation needed]

As modern computers tend further towards parallelism to keep providing the speed-ups that have been inherent in the industry over recent years, new parallel algorithms need developing in order to take full advantage of the computing power available.

We are building on a previous project[4] which layed the groudwork for this. This previous project justified the difference between CPU and GPU parallelism and outlines the justification for using GPUs- simplisitically put, this is due to the significantly greater speed ups that can be achieved.

3.2.3 Understanding GPGPUs and CUDA

3.2.4 Agent Based Modelling

Mention Flame (traditional) is a start..

3.2.5 ABM for GPUs

3.2.6 FlameGPU

4 The Experiment

4.1 The Domain Model

The domain model is taken from an existing simulation...

4.2 The Platform Model

OPTIONS:
   Custom Code?[4] not well tested less easy to update to support new tools and hardware, new CUDA GPUs
   model based -> FlameGPU[5]

4.2.1 Testing

Talk about how the model was tested to ensure correctness
missing link in model from Kieran's paper

5 Results and Evaluation

5.1 Findings

[6] could be important for evaluating the performance of FlameGPU
against original PPSim

5.2 Conclusion

5.3 Further Work

5.3.1 Hardware Availability

5.3.2 Software Generalisibility

Bibliography

[1]   D. Noble, 'The rise of computational biology', *Nature Reviews Molecular Cell Biology*, vol. 3, pp. 459–463, 2002. DOI: 10.1038/nrm810.

[2]   K. J. Alden, 'Simulation and statistical techniques to explore lymphoid tissue organogenesis', PhD thesis, University of York, Heslington, York, 2012.

[3]   R. F. Paige, A. Zolotas and D. Kolovos, 'The changing face of model-driven engineering', in *Present and Ulterior Software Engineering*, M. Mazzara and B. Meyer, Eds. Cham: Springer International Publishing, 2017, pp. 103–118. DOI: 10.1007/978-3-319-67425-4_7. [Online]. Available: https://doi.org/10.1007/978-3-319-67425-4_7.

[4]   P. Drew, 'Parallel programming tools for exploring immune system development', Master's thesis, University of York, Heslington, York, 2017.

[5]   P. Richmond, D. Walker, S. Coakley and D. Romano, 'High performance cellular level agent-based simulation with flame for the gpu', *Briefings in Bioinformatics*, vol. 11, no. 3, pp. 334–347, 2010. DOI: 10.1093/bib/bbp073.

[6]   A. Arcuri and L. Briand, 'A practical guide for using statistical tests to assess randomized algorithms in software engineering', in *Proceedings of the 33rd International Conference on Software Engineering*, ser. ICSE '11, Waikiki, Honolulu, HI, USA: ACM, 2011, pp. 1–10. DOI: 10.1145/1985793.1985795. [Online]. Available: http://doi.acm.org/10.1145/1985793.1985795.

6 Appendix

6.1 Simulation Parameters

6.2 Cell Data Structures