

State-of-the-Art GPGPU Applications in Bioinformatics

Nikitas Papangelopoulos, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece

Dimitrios Vlachakis, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece

Arianna Filntisi, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece & School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

Paraskevas Fakourelis, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece

Louis Papageorgiou, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece

Vasileios Megalooikonomou, Computer Engineering and Informatics Department, School of Engineering, University of Patras, Patras, Greece

Sophia Kossida, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Athens, Greece

ABSTRACT

The exponential growth of available biological data in recent years coupled with their increasing complexity has made their analysis a computationally challenging process. Traditional central processing unit (CPUs) are reaching their limit in processing power and are not designed primarily for multithreaded applications. Graphics processing units (GPUs) on the other hand are affordable, scalable computer powerhouses that, thanks to the ever increasing demand for higher quality graphics, have yet to reach their limit. Typically high-end CPUs have 8-16 cores, whereas GPUs can have more than 2,500 cores. GPUs are also, by design, highly parallel, multicore and multithreaded, able of handling thousands of threads doing the same calculation on different subsets of a large data set. This ability is what makes them perfectly suited for biological analysis tasks. Lately this potential has been realized by many bioinformatics researches and a huge variety of tools and algorithms have been ported to GPUs, or designed from the ground up to maximize the usage of available cores. Here, we present a comprehensive review of available bioinformatics tools ranging from sequence and image analysis to protein structure prediction and systems biology that use NVIDIA Compute Unified Device Architecture (CUDA) general-purpose computing on graphics processing units (GPGPU) programming language.

Keywords: *Compute Unified Device Architecture (CUDA), Graphics Processing Unit (GPU), Image Analysis, Multithreaded, Next Generation Software (NGS), Protein-Protein Interactions Prediction, Protein Docking, Sequence Alignment*

DOI: 10.4018/ijssbt.2013100103

INTRODUCTION

Ever since the human genome project was completed in 2001 biological data availability has been steadily growing. Especially since the introduction of new technologies, most notably microarrays and next generation sequencing (NGS), this growth has become exponential. Other areas such as molecular modelling and medical imaging are also heavily contributing not only to the aforementioned exponential growth of the available data, but also to their algorithmic complexity. At the same time, with falling genome sequencing prices and newer, more efficient instruments, this trend will continue or even accelerate for the foreseeable future (Richter & Sexton, 2009). This huge amount of data combined with their increasingly intricate, interconnecting structures will eventually lead to their analysis becoming intractable by current personal computers.

This is especially true if one considers that in the last few years the increase of computing power is starting to fall behind Moore's law, with R&D focused more on energy efficiency for mobile applications and less on increasing CPU raw power. Even the introduction of multicore processors wasn't enough to reverse this trend since most applications are not optimized to use more than one core (multithreaded) and efficient utilization of multiple cores is still a topic of active research in computer science. What is more the higher end, server grade, CPUs are cost prohibitive to any medium sized research institute or biotech company.

Thus, it is becoming obvious that we are steadily moving towards a plateau in microprocessor capabilities and consequently a stalemate in our analysis capabilities of the unrelenting wave of complex biological data.

Enter General Purpose Graphics Processing Units (GPGPU) spearheaded by NVIDIA and its pivotal Compute Unified Device Architecture (CUDA) platform.

Advent of GPUs and CUDA Appearance Evolution

It all started in the mid 90's with the appearance of the first GPUs whose sole purpose was 3D graphics acceleration, namely for use in computer games. NVIDIA produced its first graphics accelerator in 1999. These units contained a single processing unit and were extremely limited in both rendering capabilities, raw processing power and were cumbersome to write code for. But soon, the market demand for high definition content and photorealistic graphics transformed these first attempts to highly parallel, multithreaded computing powerhouses with many cores and never seen before memory bandwidth. At the same time, NVIDIA coined the term GPUs and the first attempts to utilize the raw potential of these cards for a range of scientific applications other than graphic rendering begun to appear. This was the dawn of the era of General-Purpose computation on GPU, or GPGPU.

Unfortunately, programming on these platforms was still very challenging, requiring scientists to formulate their algorithms in terms of geometrical shapes such as triangles and polygons. At the same time, the available tools offered only limited accessibility to the full computing capabilities of the GPUs. NVIDIA was the first company to realize the potential of this technology and invested heavily in creating fully programmable GPUs and developing intuitive software and hardware tools. This effort culminated in November 2006 with the introduction of CUDA, a general purpose parallel computing platform and programming model, that comes with a software environment based on the C programming language.

GPU computing momentum is growing faster than ever before. With over 375 million CUDA-enabled GPUs installed in a wide range of computers, today CUDA is the dominant platform for GPGPU. It is the preferred method for affordable High Performance Computing

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/state-of-the-art-gpgpu-applications-in-bioinformatics/105596

Related Content

Improving Prediction Accuracy via Subspace Modeling in a Statistical Geometry Based Computational Protein Mutagenesis

Majid Masso (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 54-68).

www.irma-international.org/article/improving-prediction-accuracy-via-subspace/49549

Influence of Genomic and Other Biological Data Sets in the Understanding of Protein Structures, Functions and Interactions

N. Srinivasan, G. Agarwal, R. M. Bhaskara, R. Gadkari, O. Krishnadev, B. Lakshmi, S. Mahajan, S. Mohanty, R. Mudgal, R. Rakshambikai, S. Sandhya, G. Sudha, L. Swapna and N. Tyagi (2011). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 24-44).

www.irma-international.org/article/influence-genomic-other-biological-data/52769

Low Dose Pioglitazone Attenuates Oxidative Damage in Early Alzheimer's Disease by Binding mitoNEET: Transcriptome-To-Reactome™ Biosimulation of Neurons

Charles D. Hammack, George Perry, Richard G. LeBaron, Greg Villarealand Clyde F. Phelix (2015). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 24-45).

www.irma-international.org/article/low-dose-pioglitazone-attenuates-oxidative-damage-in-early-alzheimers-disease-by-binding-mitoneet/165548

In Silico Analysis of the CST6 Tumor Suppressor Gene

Athanasia Pavlopoulou and Georgios Tsaramiris (2013). *International Journal of Systems Biology and Biomedical Technologies* (pp. 42-58).

www.irma-international.org/article/in-silico-analysis-of-the-cst6-tumor-suppressor-gene/97741

Classification Tree

(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations* (pp. 53-67).

www.irma-international.org/chapter/classification-tree/53897