

# Chapter 33

## FPGA–Based Accelerators for Bioinformatics Applications

**Alba Cristina Magalhaes Alves de Melo**  
*University of Brasilia, Brazil*

**Nahri Moreano**  
*Federal University of Mato Grosso do Sul, Brazil*

### ABSTRACT

*The recent and astonishing advances in Molecular Biology, which led to the sequencing of an unprecedented number of genomes, including the human, would not have been possible without the help of Bioinformatics. Bioinformatics can be defined as a research area where computational tools and algorithms are developed to help biologists in the task of understanding the organisms. Some Bioinformatics applications, such as pairwise and sequence-profile comparison, require a huge amount of computing power and, therefore, are excellent candidates to run in FPGA platforms. This chapter discusses in detail several recent proposals on FPGA-based accelerators for these two Bioinformatics applications, highlighting the similarities and differences among them. At the end of the chapter, research tendencies and open questions are presented.*

### 1. INTRODUCTION

Bioinformatics is an interdisciplinary field that involves computer science, biology, mathematics and statistics (Mount 2004). Its main goal is to analyze biological sequence data and genome content in order to obtain the function/structure of the sequences as well as evolutionary information.

Once a new biological sequence is discovered, its functional/structural characteristics must be established. In order to do that, the newly discovered sequence is compared against the sequences that compose biological databases, in search of similarities. Sequence comparison is, therefore, one of the most basic operations in Bioinformatics. A sequence can be compared to another sequence (pairwise comparison), to a profile that describes a family of sequences (sequence-profile compari-

DOI: 10.4018/978-1-4666-3604-0.ch033

son) or to a set of sequences (multiple sequence comparison). The most accurate algorithms to execute pairwise and sequence-profile comparisons are usually based on dynamic programming, with quadratic time and space complexity. This can easily lead to extremely high execution times and huge memory requirements, since biological databases are growing exponentially and biological sequences usually are extremely long.

Parallel processing can be used to produce results faster, reducing significantly the time needed to obtain results with dynamic programming-based algorithms. However, using parallel processing in Bioinformatics is not straightforward since the problems are often solved by complex methods, with a great amount of data dependencies. Moreover, it has been observed that, for some problems that involve really huge data sets, software-only parallel processing techniques are not able to reduce the execution time to reasonable limits. For instance, in order to use the dynamic programming exact method to compare the human chromosome 21 and the ape chromosome 22, with sizes 46 Million Base Pairs (MBP) and 34 MBP, respectively, it is estimated that a cluster with 2048 processors would take approximately one day (Batista et al., 2008). This shows clearly that new approaches and technologies must be used to further reduce such high execution times.

Designing specific hardware for such algorithms is a very attractive alternative to software-only solutions. Since the hardware will be tailored to execute a specific algorithm, drastic performance improvements can be achieved. This observation was made by many researchers in Bioinformatics in the 1990s, that led to a great variety of hardware-based solutions. Due to its flexibility, FPGAs (Field Programmable Gate Arrays) are the natural choice for these designs. In the last years, FPGAs have intensively been used to build specific circuits, especially targeted to accelerate compute-intensive Bioinformatics applications.

The goal of this chapter is to discuss in detail and compare the recent advances in FPGA-based accelerators for some classes of Bioinformatics applications. The problems discussed are two types of Biological Sequence Comparison: pairwise sequence comparison and sequence-profile comparison. The first one is widely used all over the world as a first step in the solution of complex problems such as the determination of the evolutionary history of the species. The second one is extremely important since it is used to decide if a recently sequenced protein belongs or not to a particular protein family/superfamily. For both problems, many FPGA-based accelerators have been proposed that obtained impressive speedups over the sequential and software-only parallel implementations.

The structure of this chapter is the following. In Section 2, we will introduce the Biological Sequence Comparison problem and present the most widely used algorithms to solve it. In Section 3, we discuss several state-of-the-art FPGA accelerators that tackle the pairwise sequence comparison problem. In Section 4, several state-of-the-art FPGA-based accelerators for the sequence-profile comparison problem are discussed. Finally, Section 5 concludes the chapter, presenting the future tendencies in this research area.

## 2. PAIRWISE COMPARISON AND SEQUENCE-PROFILE ALIGNMENT

### 2.1. Alignment and Score

A biological sequence is a single and continuous molecule of nucleic acid or protein. It is represented by a linear list of residues, which are nucleotide bases (for DNA sequences, for instance) or amino acids (for protein sequences). To compare biological sequences, we need to find the best alignment between them, which is to place one sequence above the other making clear the correspondence

29 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/chapter/fpga-based-accelerators-bioinformatics-applications/76086](http://www.igi-global.com/chapter/fpga-based-accelerators-bioinformatics-applications/76086)

## Related Content

---

### Political Discourse as Sliding Mode Manifestations

Ekaterina Yuryevna Aleshina (2022). *International Journal of Applied Research in Bioinformatics* (pp. 1-10).

[www.irma-international.org/article/political-discourse-as-sliding-mode-manifestations/282690](http://www.irma-international.org/article/political-discourse-as-sliding-mode-manifestations/282690)

### Hybrid High-Performance Computing Algorithm for Gene Regulatory Network

Dina Elsayad, Safawat Hamad, Howida Abd-Alfatah Shedeed and Mohamed Fahmy Tolba (2024).

*Research Anthology on Bioinformatics, Genomics, and Computational Biology* (pp. 968-982).

[www.irma-international.org/chapter/hybrid-high-performance-computing-algorithm/342558](http://www.irma-international.org/chapter/hybrid-high-performance-computing-algorithm/342558)

### Gene Expression Data Sets

(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations* (pp. 6-9).

[www.irma-international.org/chapter/gene-expression-data-sets/53893](http://www.irma-international.org/chapter/gene-expression-data-sets/53893)

### Medical Informatics and Bioinformatics

Julio Cesar Facelli, John F. Hurdle and Joyce A. Mitchell (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 195-221).

[www.irma-international.org/chapter/medical-informatics-bioinformatics/76063](http://www.irma-international.org/chapter/medical-informatics-bioinformatics/76063)

### Discriminative Subgraph Mining for Protein Classification

Ning Jin, Calvin Young and Wei Wang (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 36-52).

[www.irma-international.org/article/discriminative-subgraph-mining-protein-classification/47095](http://www.irma-international.org/article/discriminative-subgraph-mining-protein-classification/47095)