


# An OpenCL-based parallel acceleration of a Sobel edge detection algorithm Using Intel FPGA technology

Abedalmuhdi Almomany<sup>a</sup> , Ahmad M. Al-Omari<sup>b</sup>, Amin Jarrah<sup>a</sup>, Mohammed Tawalbeh<sup>c</sup>, Amin Alqudah<sup>a</sup>

<sup>a</sup> Department of Computer Engineering, Yarmouk University, Irbid, Jordan

<sup>b</sup> Department of Biomedical systems and Bioinformatics Engineering, Yarmouk University, Irbid, Jordan

<sup>c</sup> Information Technology & Communications Center, Jordan University of Science and Technology, Irbid, Jordan

---

## ABSTRACT

This paper examines the feasibility of using commercial out-of-the-box reconfigurable field programmable gate array (FPGA) technology and the open computing language (OpenCL) framework to create an efficient Sobel edge-detection implementation, which is considered a fundamental aspect of image and video processing. This implementation enhances speedup and energy consumption attributes when compared to general single-core processors. We created the proposed approach at a high level of abstraction and executed it on a high commodity Intel FPGA platform (an Intel De5-net device was used). This approach was designed in a manner that allows the high-level compiler/synthesis tool to manipulate a task-parallelism model. The most promising FPGA and conventional implementations were compared to their single-core CPU software equivalents. For these comparisons, local-memory, pipelining, loop unrolling, vectorization, internal channel mechanisms, and memory coalescing were manipulated to provide a much more effective hardware design. The run-time and power consumption attributes were estimated for each implementation, resulting in up to 37-fold improvement of the execution/transfer time and up to a 53-fold improvement in energy consumption when compared to a specific single-core CPU-based implementation.

**Keywords:** FPGA, reconfigurable computing, parallel processing, edge detection, OpenCL, image processing, integrated circuits

**Categories:** • Hardware ~ Reconfigurable logic applications

## Email:

Abedalmuhdi Almomany [emomani@yu.edu.jo](mailto:emomani@yu.edu.jo) (CORRESPONDING),  
Ahmad M. Al-Omari [aomari@yu.edu.jo](mailto:aomari@yu.edu.jo),  
Amin Jarrah [amin.jarrah@yu.edu.jo](mailto:amin.jarrah@yu.edu.jo),  
Mohammed Tawalbeh [mt@just.edu.jo](mailto:mt@just.edu.jo),  
Amin Alqudah [amin.alqudah@yu.edu.jo](mailto:amin.alqudah@yu.edu.jo)

## Article history:

Received: 25 Sep 2019  
Accepted: 11 Mar 2020  
Available online: 20 Jul 2020

---

Almomany, A., Al-Omari, A.M., Jarrah, A., Tawalbeh, M., and Alqudah, A. (2020). An OpenCL-based parallel acceleration of a Sobel edge detection algorithm Using Intel FPGA technology. *South African Computer Journal* 32(1), 3–26. <https://doi.org/10.18489/sacj.v32i1.749>

Copyright © the author(s); published under a [Creative Commons NonCommercial 4.0 License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/). SACJ is a publication of the South African Institute of Computer Scientists and Information Technologists. ISSN 1015-7999 (print) ISSN 2313-7835 (online).

# Discovering regulatory motifs of genetic networks using the indexing-tree based algorithm: a parallel implementation

Regulatory motifs of genetic networks

Abedalmuhdi Almomany

*Department of Computer Engineering, Yarmouk University, Irbid, Jordan*

Ahmad M. Al-Omari

*Department of Biomedical Systems and Bioinformatics Engineering, Yarmouk University, Irbid, Jordan*

Amin Jarrah

*Department of Computer Engineering, Yarmouk University, Irbid, Jordan, and*

Mohammad Tawalbeh

*Information Technology Center, Jordan University of Science and Technology, Irbid, Jordan*

Received 19 February 2020  
Revised 30 April 2020  
27 May 2020  
Accepted 27 May 2020

## Abstract

**Purpose** – The problem of motif discovery has become a significant challenge in the era of big data where there are hundreds of genomes requiring annotations. The importance of motifs has led many researchers to develop different tools and algorithms for finding them. The purpose of this paper is to propose a new algorithm to increase the speed and accuracy of the motif discovering process, which is the main drawback of motif discovery algorithms.

**Design/methodology/approach** – All motifs are sorted in a tree-based indexing structure where each motif is created from a combination of nucleotides: ‘A’, ‘C’, ‘T’ and ‘G’. The full motif can be discovered by extending the search around 4-mer nucleotides in both directions, left and right. Resultant motifs would be identical or degenerated with various lengths.

**Findings** – The developed implementation discovers conserved string motifs in DNA without having prior information about the motifs. Even for a large data set that contains millions of nucleotides and thousands of very long sequences, the entire process is completed in a few seconds.

**Originality/value** – Experimental results demonstrate the efficiency of the proposed implementation; as for a real-sequence of 1,270,000 nucleotides spread into 2,000 samples, it takes 5.9s to complete the overall discovering process when the code ran on an Intel Core i7-6700 @ 3.4 GHz machine and 26.7 s when running on an Intel Xeon x5670 @ 2.93 GHz machine. In addition, the authors have improved computational performance by parallelizing the implementation to run on multi-core machines using the OpenMP framework. The speedup achieved by parallelizing the implementation is scalable and proportional to the number of processors with a high efficiency that is close to 100%.

**Keywords** Optimization, OpenMP, Parallel processing, Genetic network, Multi-core, Regulation motif

**Paper type** Research paper



This research was supported by the Deanship of Scientific Research at Yarmouk University under Grant number: 26/2018.

## Finding Regulatory Motifs of Genetic Networks Using Cut-Sort Algorithm

Ahmad M. Al-Omari<sup>1a</sup>, Mohammed H. Tawalbeh<sup>2b</sup>, Abedalmuhdi M. Almomany<sup>3c</sup>

<sup>1</sup>Department of Biomedical Systems and Informatics Engineering, Yarmouk University, Irbid, Jordan

<sup>2</sup>Information Technology & Communications Center, Jordan University of Science and Technology, Irbid, Jordan

<sup>3</sup>Department of Computer Engineering, Yarmouk University, Irbid, Jordan

<sup>a</sup>e-mail: aomari@yu.edu.jo

<sup>b</sup>e-mail: mt@just.edu.jo

<sup>c</sup>e-mail: emomani@yu.edu.jo

*Received: December 29, 2018*

*Accepted: February 23, 2019*

**Abstract**— Understanding the targets of regulatory genes has become a challenging problem for bioinformaticians and biologists in systems biology. The main issue in solving this challenge consists in finding motifs that are finding short, recurring patterns in DNA or in amino-acid sequences that presumably have a regulatory function. A motif is considered a signature for a protein family binding to sequence motifs in the genome. The major challenge in finding motifs arises from the fact that most of the time the motifs are not well conserved. To discover such degenerate motifs, aligning multiple sequence motifs becomes a challenge. Usually, a motif discovery algorithm uses some prior information about the motifs to be discovered. In this paper, we present a novel algorithm for finding conserved sequence motifs in DNA without having a priori knowledge about the motifs. However, the algorithm can be used for motifs sequence both in DNA and in proteins. Our algorithm mainly depends on cutting sequences that have conserved motifs into equal fragments, sorting the fragments and then extending in both fragment directions. The algorithm runs in a very short time period. It takes 5.5 seconds for a real data sequence with length  $N = 28,000$  nucleotides to find its identical, degenerate, long and short motifs; it can be easily parallelized by implementing it on General Purpose Graphical Processing Units. The algorithm guarantees to find any globally optimal solution within a short time even for sequences with very long motifs.

**Keywords**— Bioinformatics, Genetic network, GPGPU, Regulation motif, Systems biology.

### I. INTRODUCTION

Motif prediction represents one of the most active research areas in bioinformatics since the late 1980's [1]. Most DNA signals allow some minor variations in their sequence. The problem is central to biological systems because the identification of these motifs allows the construction of connections between genes and regulators through identification of DNA-protein interactions [2]-[4]. Thus, the signal will actually consist of several different possible words, often closely related. Such a collection of words is a repeated pattern which is called a motif that appears as a conserved sequence of nucleotides in a DNA or a conserved sequence of amino acids in a protein. This conserved pattern is usually taken to be a cis-regulatory sequence [3]. For example, finding motifs based on the prediction of a "conserved" sequence in the promoter regions of genes can help us discover genes that are possibly co-regulated by the same transcription factor(s) [3]. In general, these motifs are well defined, i.e., TATA box in the promotor area. Discovering gene regulation mechanisms helps us to understand the development, functioning and evolution of an organism. The importance of motifs has motivated many researchers to develop different tools and algorithms for finding them. Consensus [5] is a greedy multiple alignment that provided a strategy to assess the statistical significance of a given information content score based on large deviation statistic; MEME [6] uses statistical modeling techniques, such as expectation maximization, to automatically choose the best width, number of occurrences, and description for each motif from input of a



## Recognition and Retrieval of Mathematical Expressions from Arabic Documents

Emad Al-Shawakfa<sup>1</sup>, Mohammed Tawalbeh<sup>2</sup>

<sup>1</sup>Information Systems Department, Faculty of IT, Yarmouk University, Jordan, shawakfa@yu.edu.jo

<sup>2</sup>Information Systems Department, Faculty of IT, Yarmouk University, Jordan, mt@just.edu.jo

### ABSTRACT

Recently, researchers have strived to enhance existing search tools to enable the retrieval of a diversity of data types to help users in finding what they are looking for from the Web, however, very limited research is applicable to Arabic content.

In this research, we introduce a rule-based approach to search for mathematical expressions written in Arabic and/or English from Arabic stored documents. A set of normalization as well as math expressions' equivalence rules were built to enhance the capabilities of a math search engine. Furthermore, rules for structural search to enable search for sub-expressions were also built. An indexing mechanism and a mapping between Arabic and English expressions during the search process was also produced. The approach was applied using a set of forty queries; written in Arabic and/or English, was applied on a manually collected dataset of 100 documents which has produced an overall accuracy of 75%.

**Key words:** Math Information Retrieval, Arabic Math Expression, Text-Based Search, Normalized Math Tree.

### 1. INTRODUCTION

Search engines were developed to help in locating data of various types, including text, images, audio, and video over the Internet. However, domain specific documents may contain special forms and characters such as: mathematical expressions (ME), drawings, charts, tables, and diagrams that are more structured and difficult to retrieve using traditional search engines. Furthermore, such types of data cannot be effectively indexed and/or queried via conventional search engines. As mentioned by [1], navigation by search engines are either static or dynamic.

According to [2], web content can be categorized into two major types: structured and unstructured. As defined in [3], structured information is information that is ordered in a particular way and combined with extra features to describe it and to construct the relationship between its contents; a

information. On the other hand, unstructured information is defined as information that exists in random pieces which does not contain any hidden information other than text; a normal text-based information is a good example of unstructured information.

Many existing web sites and digital libraries include huge number of scientific papers and books in digital format that contain mathematical expressions and formulae; an essentiality to researchers of different scientific areas like: Mathematics, Physics, Biology, Chemistry, etc. [4]. As Arabic math content has grown extensively on the Internet lately, and with the need by Arab researchers and/or students to search for such content on the Internet, we feel the need to provide a mechanism to search for such content easily. Unfortunately, existing search engines do not provide suitable ways to support the search for such documents containing ME. Figure 1 gives a screenshot of an existing Arabic content from a physics topic on the web [5].

السؤال الخامس: حدد مجال الاقترانات الآتية:

|                                  |                                     |                                     |
|----------------------------------|-------------------------------------|-------------------------------------|
| (أ) $y = \frac{2x-3}{x+1}$       | (ب) $y = \frac{x-2}{x^2-4}$         | (ج) $y = \frac{x-1}{x^2-2x+1}$      |
| (د) $y = \frac{3}{x^2-2}$        | (هـ) $y = \frac{x^2-9}{x^2-10x+25}$ | (و) $y = \frac{x-3}{x^2-2x}$        |
| (ز) $y = \sqrt{\frac{x-5}{x+3}}$ | (ح) $y = \frac{3}{x-2}$             | (ط) $y = \frac{x^2-49}{x^2-14x+49}$ |

Figure 1: Arab math content on the web [5]

As noted by [6], a math expression; whether written in Arabic or English scriptures, could be categorized under structured information and are very formal since they contain special characters and symbols and contain both the structure and the semantic of the information that tell much about the expression itself; for which special attention and processing tools are needed.

With traditional text-based search engines, when searching for functions of symbols like **sin**, **cos**, **log**, and many others, the search will be for the best text matching of symbols only, and not their meaning. With the existence of many expressions with the same meaning presented in different ways, a math search engine must be able to retrieve all related expressions of the same meaning and with different structures than a given query, regardless of the language used; English