AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
(Computer Engineering and Systems)

Efficient Architecture for Controlled Accurate Computation

A Thesis submitted in partial fulfillment of the requirements of
Master of Science in Electrical Engineering
(Computer Engineering and Systems)

by

DiaaEldin Mohamed Abdalla Mohamed Osman
Bachelor of Science in Electrical Engineering
(Computer Engineering and Systems)
Faculty of Engineering, Ain Shams University, 2012

Supervised By

Prof. Ayman Mohammad Bahaa-Eldin Sadek
Dr. Mohamed Ali Sobh
Dr. Ahmed Moustafa Zaki

Cairo, 2018

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer Engineering and Systems

Efficient Architecture for Controlled Accurate Computation

by

DiaaEldin Mohamed Abdalla Mohamed osman
Bachelor of Science in Electrical Engineering
(Computer Engineering and Systems)
Faculty of Engineering, Ain Shams University, 2012

Examiners' Committee

Name and affiliation	Signature
Prof. Hassan Taher Durra Computer Engineering and Systems Faculty of Engineering, Cairo University.	.....
Prof. Hussien Ismail Shahin Computer Engineering and Systems Faculty of Engineering, University.	.....
Prof. Ayman Mohamed Bahaa-Eldin Computer Engineering and Systems Faculty of Engineering, Ain Shams University.	.....
Dr. Mohamed Ali Sobh Computer Engineering and Systems Faculty of Engineering, Ain Shams University.	.....

Date: ... May 2018

Statement

This thesis is submitted as a partial fulfillment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain shams University. The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Name: DiaaEldin Mohamed Abdalla Mohamed Osman

Signature

.....

Date: 29 Decemeber 2018

Researcher Data

Name: DiaaEldin Mohamed Abdalla Mohamed Osman
Date of Birth: 03/03/1990
Place of Birth: Cairo, Egypt
Last academic degree: Bachelor of Science
Field of specialization:
Computer Engineering and Systems

University issued the degree : Ain Shams University
Date of issued degree : 2012
Current job : Software Development Engineer by Vector Informatik GmbH

Thesis Summary

The work of this thesis is to provide an efficient architecture for the accurate arithmetic operations and these operations should also have controlled accuracy. Also the architecture should not require special hardware support. It is proposed in this thesis is to use the streaming vectors that can be found in modern CPUs to accelerate error free transformation algorithms (Multi-Number System) to have an efficient architecture with reasonable performance and accuracy controlled that can be used by various applications that need different level of accuracy with high performance from execution time perspective.

Summary

The thesis is divided into five chapters as listed below:

Chapter 1 Introduction: provides an introduction to the problem, the motivation for this work, objectives and the approach used to achieve the outcome of this thesis.

Chapter 2 Literature Review: provides a literature review of the real number with their different representations and the problems of representing those real numbers on the digital systems, also review the previous work done by other researchers to have error free transformation algorithms for floating-point calculations to have accurate operations.

Chapter 3 Acceleration of accurate floating point operations using Single Instruction Multiple Data: provides detailed explanation of the work done to accelerate the MN-System using both SSE and AVX, and complexity analysis of the accelerated operations.

Chapter 4 Experimental results: provides some case studies where the performance of the accelerated version of MN-System was tested in the applications that require high accuracy and speedup was measured against the original MN version.

Chapter 5 Conclusion and Future Work: contains the conclusion of the whole work documented in this thesis and also Future work suggestions that can enhance the performance.

Key words: Error free transformation, Hilbert Matrix, Floating-point numbers, Ill-conditioned matrices, Vandermonde Matrix, polynomial regression.

Abstract

Faculty of Engineering – Ain Shams University
Computer Engineering and Systems Department

Thesis title: **"Efficient Architecture for Controlled Accurate Computation"**
Submitted by: **DiaaEldin Mohamed Addalla Mohamed Osman**
Degree: **Masters of Science**

Abstract

Many applications suffer from the representation of the real numbers due to the propagation and the accumulation of errors. The real numbers are represented in fixed length format that supports a large dynamic range, but on the other hand it leads to truncation of bits in case of a number that is represented by a long sequence of bits. Researchers proposed several solutions for these errors, one of those proposals is the Multi-Number (MN) system. MN system represents the real number as a vector of floating-point numbers with controlled accuracy by tuning the length of the vector to accumulate non overlapping real number sequences. MN system disadvantage is that the MN computations are iterative and time consuming making it unsuitable for real time applications. In this work, the Single Instruction Multiple Data (SIMD) paradigm found in modern CPUs is exploited to accelerate the MN Computations. The basic arithmetic operation algorithms had been modified to utilize the SIMD architecture and support both single and double precision operations. The new architecture preserves the same accuracy of the original one when implemented for both single and double precision. Also in this work the normal Gaussian Jordan Elimination algorithm was proposed and used to get the inverse of the Hilbert Matrix, as an example of ill-conditioned

matrices, instead of using iterative and time consuming methods. The accuracy of the operations was proved by getting the inverse of the Hilbert Matrix and verify that the multiplication of the inverse and the original matrix producing the unity matrix. Hilbert Matrix inverse calculation time was accelerated and achieved a speedup 3x, compared to the original NM operations. In addition to the previous, the accelerated MN system version was used to solve the polynomial regression problem.

Acknowledgment

All praise is due to Allah, Most Merciful, the Lord of the World, who taught man what he knew not. I would like to thank ALLAH for bestowing upon me the chance, strength, and ability to complete this work. I wish to express my gratitude to my supervisors, Prof. Dr. Ayman Bahaa-Eldin, Dr. Mohamed Ali Sobh, Dr. Ahmed Zaki for their exceptional guidance, encouragement, insightful thoughts, and useful discussions. It was my great pleasure to work with my supervisors. I am in no way capable of appropriately thanking my mother and my wife for their unconditional love and unlimited support. And I ask ALLAH to have mercy on my father, grandfathers, and grandmothers. Thanks for all.

DiaaEldin Mohamed Abdalla Mohamed
Computer Engineering and Systems
Faculty of Engineering
Ain Shams University
Cairo, Egypt
December 2018

Contents