PARAMETER OPTIMIZATION FOR HIGH PERFORMANCE COMPUTING SYSTMS USING GENETIC ALGORITHM

K. I. A. I. KARIYAWASAM

208558H

MSC IN ARTIFICIAL INTELLIGENCE

DEPARTMENT OF COMPUTATIONAL MATHEMATICS

FACULTY OF INFORMATION TECHNOLOGY

University of Moratuwa

Sri Lanka

JULY 2023

PARAMETER OPTIMIZATION FOR HIGH PERFORMANCE COMPUTING SYSTMS USING GENETIC ALGORITHM

K. I. A. I. KARIYAWASAM

208558H

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree

MSC IN ARTIFICIAL INTELLIGENCE

DEPARTMENT OF COMPUTATIONAL MATHEMATICS

FACULTY OF INFORMATION TECHNOLOGY

University of Moratuwa

Sri Lanka

JULY 2023

DECLARATION

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 11th July 2023

The above candidate has carried out research for the Master's thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Thushari Silva

Signature of the Supervisor: Date: 12/07/2023

DEDICATION

I would also like to dedicate this thesis to my wife, Anuradha, who has been my rock and my constant source of motivation. Also, my two children Dhyana and Veditha; their patience, understanding, and steadfast support have been invaluable to me throughout this journey. I am grateful for their unwavering love, and for always believing in me, even when I didn't believe in myself.

I would like to dedicate this thesis to my parents, who have always supported me throughout my academic journey. Their unwavering encouragement and belief in me have been instrumental in helping me achieve my goals. I am grateful for their unwavering love and guidance.

I would also like to dedicate this thesis to my colleagues Romain Bottier and Indika Prasad Kumara at Dell Technologies and London Stock Exchange Group for their support on accessing real life scenarios and their valuable technical inputs to shape my work into success.

Finally, I would like to dedicate this thesis to all the people who have supported and encouraged me along the way, including my friends, family, and mentors. Without their guidance and support, this achievement would not have been possible.

Thank you all for your love and support.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Thushari Silva, my thesis supervisor, for her invaluable guidance, support, and encouragement throughout the research process. Her feedback and advice were instrumental in shaping the direction of this work and bringing it to fruition.

I am grateful to my colleagues in MSc in Artificial Intelligence for their support and encouragement, and for the stimulating discussions that helped me to refine my ideas and approach.

Thank you all for your contributions to this work and for making this journey a memorable one.

K. I. A. I. Kariyawasam

ABSTRACT

Though Von Neumann Architecture (vNa) has had unprecedented success for general-purpose computing for the past seven decades. While being heavily used in High Performance Computing (HPC) systems, vNa is becoming unpopular among industry professionals due to vNa bottleneck and unable to address the ever-increasing demand for performance gain by latest HPC applications. Though there are few methodologies to overcome vNa bottleneck; namely Processor caching, Prefetching, multithreading, Processing in Memory (PIM) and RAMBUS, which we will not discuss in this paper, the latency between ALU and memory still hold back the performance of vNa systems, even in HPC environments. Alternatively, there are few different technologies are emerging as opposed to vNa such as Quantum computing, Parallel computing, Harvard Computing architecture (HCA) which are addressing the vNa bottleneck and taking computing performance in to a whole new level in terms of performance in Floating Points (Flops). On the other hand, another alternative is to use accelerators like Graphic Processing Units (GPU, Field Programmable Gate Arrays (FPGA) and Dell Process Acceleration Tool (DPAT) to overcome the latency between ALU and Memory subsystem.

Even though there are several alternatives available to overcome vNa bottleneck, the performance requirement from almost every HPC application seems never to be achieved by the underlying HPC hardware systems. Hardware system configurations are always performed at a sub-optimal level, and no one has the time or patience to go through thousands of different permutations of hardware parameter finetuning. Moreover, we simply cannot blame any system administrator who is responsible for HPC system configurations, especially processor, memory, and network parameter finetuning to provide best performance benchmarks, as it's a very time consuming and tedious task of many man-days of manually configuring all those parameters to find an optimal convergence point. On the other hand, one cannot be satisfied of a particular combination of parameters is the peak performance point as there always be a doubt whether there can be another set of combination exist for the optimal performance. Thus, it will be a never-ending manual task to find the optimal performance of HPC hardware systems.

To overcome HPC hardware parameter finetuning to find the best performance benchmarks, we proposed use of Genetic Algorithm based approach to find the best combination of Processor, memory, and network parameters to deliver highest hardware performance for any HPC application.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES