

Mapping of Processing Elements of Hardware-based Production Systems on Networks on Chip

by

Mostafa Wasiuddin Numan

B. Sc. (Engineering) in Computer Science and Engineering
Shahjalal University of Science and Technology, Bangladesh, 2006

M. Sc. (By Research) in Electrical, Electronic and Systems Engineering
National University of Malaysia (UKM), Malaysia, 2010

Thesis submitted for the degree of

Doctor of Philosophy

in

Electrical and Electronic Engineering
Faculty of Engineering, Computer and Mathematical Sciences
The University of Adelaide, Australia

2017

Supervisors:

Professor Michael Liebelt

Dr Braden J. Phillips

© 2017

Mostafa Wasiuddin Numan

All Rights Reserved



THE UNIVERSITY
of ADELAIDE

*To my dearest parents,
to my wonderful wife, Fathima, and our little princess, Zoharin,
to my sister, Mili, and brother, Shahan,
without whom none of my success would be possible.*

Contents

| | |
|---|-------------|
| Contents | v |
| Declaration | ix |
| Acknowledgments | xi |
| Conventions | xiii |
| Abstract | xv |
| List of Figures | xvii |
| List of Tables | xxi |
| Chapter 1. Introduction | 1 |
| 1.1 Background | 2 |
| 1.2 Research motivations | 4 |
| 1.3 Thesis objectives | 5 |
| 1.4 Statement of original contributions | 6 |
| 1.4.1 On-chip communication of cognitive workload | 6 |
| 1.4.2 Optimised mapping techniques | 7 |
| 1.5 Overview of the thesis | 7 |
| Chapter 2. Background and Related Work | 11 |
| 2.1 Introduction | 12 |
| 2.2 Artificial general intelligence | 13 |
| 2.3 Cognitive architectures | 14 |
| 2.3.1 Characteristics of cognitive architectures | 16 |
| 2.3.2 Related work on cognitive architectures | 17 |
| 2.3.3 Soar | 19 |

| | | |
|--|--|-----------|
| 2.3.4 | ACT-R | 22 |
| 2.4 | Production systems | 24 |
| 2.4.1 | Basic structure of production systems | 24 |
| 2.4.2 | Related work on production systems | 26 |
| 2.5 | Multi-processor systems on chip | 28 |
| 2.6 | Networks on chip | 30 |
| 2.6.1 | NoC communication model | 31 |
| 2.6.2 | Related work on NoC architectures | 33 |
| 2.6.3 | Related work on NoCs mapping | 34 |
| 2.7 | Chapter summary | 35 |
| Chapter 3. Hardware-based Production System | | 37 |
| 3.1 | Introduction | 38 |
| 3.2 | Production systems | 39 |
| 3.2.1 | Rete matching algorithm | 40 |
| 3.2.2 | TREAT | 41 |
| 3.3 | Street Processor: Hardware-based production system | 42 |
| 3.4 | Street language | 43 |
| 3.4.1 | Working memory | 44 |
| 3.4.2 | Production rules | 44 |
| 3.4.3 | Synchronisation | 48 |
| 3.5 | Street Processor architecture | 50 |
| 3.5.1 | Micro-architecture of a producter | 50 |
| 3.5.2 | Big Productor | 52 |
| 3.6 | Dependency graph | 53 |
| 3.7 | Sleep period | 54 |
| 3.8 | Conclusion | 55 |
| Chapter 4. Interconnect Platform of SoCs with Homogeneous PEs | | 57 |
| 4.1 | Introduction | 58 |
| 4.2 | Limitations of conventional on-chip communication | 58 |

| | | |
|---|---|-----------|
| 4.3 | NoC architecture | 59 |
| 4.3.1 | Physical components | 60 |
| 4.3.2 | Topology | 61 |
| 4.3.3 | Routing schemes | 62 |
| 4.3.4 | Routing implementation | 64 |
| 4.3.5 | Flow control techniques | 65 |
| 4.3.6 | Virtual channels | 67 |
| 4.4 | Router micro-architecture | 67 |
| 4.5 | NoC of the Street Processor | 71 |
| 4.6 | Conclusion | 71 |
| Chapter 5. Mapping of Homogeneous PEs to NoC Routers | | 73 |
| 5.1 | Introduction | 74 |
| 5.2 | Mapping considerations | 75 |
| 5.3 | Mapping constraint and goal | 76 |
| 5.3.1 | Energy cost function | 76 |
| 5.3.2 | Problem definition | 78 |
| 5.4 | Simulated Annealing | 79 |
| 5.4.1 | Simulated Annealing algorithm | 80 |
| 5.4.2 | Annealing schedule | 80 |
| 5.4.3 | Acceptance test | 80 |
| 5.4.4 | Swapping function | 82 |
| 5.4.5 | Stop criterion | 82 |
| 5.5 | Branch and Bound | 82 |
| 5.5.1 | Branch and Bound algorithm | 83 |
| 5.5.2 | Lower bound cost calculation | 86 |
| 5.5.3 | Upper bound cost calculation | 86 |
| 5.6 | Test case: the self-configurable Street agent | 87 |
| 5.6.1 | Entity relationship | 88 |
| 5.6.2 | Production rules | 89 |
| 5.7 | Experiments | 90 |
| 5.8 | Analysis of Branch and Bound mapping | 94 |
| 5.9 | Conclusion | 96 |

| | |
|--|------------|
| Chapter 6. Priority-based Simulated Annealing of Grouped PEs | 97 |
| 6.1 Introduction | 98 |
| 6.2 Graph partitioning | 99 |
| 6.2.1 Multi-level graph partitioning | 100 |
| 6.2.2 Multi-level partitioning of traffic graph | 103 |
| 6.3 Priority-based Simulated Annealing | 104 |
| 6.3.1 Initial mapping | 104 |
| 6.3.2 Annealing | 107 |
| 6.4 Test case: the Subsumption Cockroach agent | 108 |
| 6.4.1 Subsumption architecture | 109 |
| 6.4.2 Development of the Cockroach agent | 110 |
| 6.5 Experiments | 111 |
| 6.6 Conclusion | 115 |
| | |
| Chapter 7. Conclusions and Future Work | 119 |
| 7.1 Summary | 120 |
| 7.2 Conclusions | 122 |
| 7.3 Future work | 123 |
| | |
| Appendix A. List of Street Rules of the Self-configurable Agent | 127 |
| | |
| Appendix B. List of Street Rules of the Cockroach Agent | 131 |
| | |
| Bibliography | 141 |
| | |
| List of Acronyms | 155 |
| | |
| Publications | 157 |
| | |
| Biography | 159 |

Declaration

I certify that this work contains no material, which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan, photocopying, and dissemination through the library digital thesis collection subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search, the Australian Digital Thesis Program (ADTP), and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

September 26, 2017

Date

Acknowledgments

First and foremost, I take this opportunity to convey sincere gratitude to my supervisors, **Professor Michael Liebelt** and **Dr Braden J. Phillips** for their guidance and support throughout my candidature. Their remarkable expertise and unwavering optimism have always been helpful in propelling my research forward. I owe many thanks to my principal supervisor, Professor Liebelt for his generous knowledge sharing, insightful discussions and inspiring words in times of needs. His critical and thoughtful comments were always constructive and fruitful to improve the quality of my research. I am thankful to Dr Phillips for being a constant source of research ideas and constructive suggestions. I would like to gratefully acknowledge his enthusiastic supervision and encouraging attitude.

I recognise that this research would not have been possible without the financial support of Australian Government via a generous International Postgraduate Research Scholarship (IPRS) and Australian Postgraduate Award (APA).

There are many people who helped me throughout my PhD candidature. I would express appreciation to my colleagues and friends in the Centre for High Performance Integrated Technologies and Systems (CHiPTec), Jesse Frost, Francis Li, Peng Wang, Thanh Thi Thanh Bui, Simon Keen and Muhammad Usman Khan for their valuable comments time to time. My sincere gratitude goes to Jesse for his passion in discussing around different research issues and critical suggestions.

I am thankful to the office and support staff of the school for their kindness and assistance. I would also like to thank the head of the school, Associate Prof. Cheng-Chew Lim, for his support during my PhD research. In addition, I sincerely thank my friends and colleagues in the school, Syed Imranul Islam, Mehdi Kasaei, Dr Amir Ebrahimi, Dr Sarah Immanuel, Robert Moric, Madhulika Tripathi, Yansong (Garrison) Gao, Nicholas P. Lawrence, Dr Sam Darvishi, Shengjian (Jammy) Chen, Dr Ali Karami, Dr Zahra Shaterian, Mariam Ebrahimpour, Dr Mostafa Rahimi and Nazmul Huda for making such a friendly research environment. It has been a privilege to get acquainted with so many talented individuals.

Acknowledgments

I appreciate the selfless support of my friends, Dr Md. Ayub, Kingshuk Nandy, Dr Manabendra Saha, Dr Raihan Rumman, Shuvra Shaha, Asafuddoulah Suzon, Javed Husain, Hassan Jahangir and Dr Zaheed Hasssan for making life so easy since my very early days in Adelaide. I also thank my good old friends, Majba Uddin, Saiful Islam, Rumon Rahman, Zaidul Alam, Razeeb Saleheen, Amit Roy and their families for always staying around me and my family. You all make me feel at home while away from home.

My endless gratitude goes to my parents, who always bestowed me with infinite support, wishes and continuous love. They always wanted me to be a 'doctor' (physician). I somehow failed to pursue their dreams, but they never complained, rather encouraged me for whatever I did. I hope, they can now get a contentment seeing 'Dr' in front of my name. Abbu and Ammu, you are the best blessings of Almighty Allah to me. I also thank Apamoni, Dulabhai and Najnin for their tremendous support and inspiration throughout my journey. As always, my heartfelt gratitude goes to my younger brother, Shahan, for taking my responsibilities on his shoulder since very young age. I also wish to express my warm and sincere thanks to my mother-in-law for her kindness and earnest wishes.

Last but not least, my wholehearted appreciation are due to my beloved wife, Fathima. The way she took care of me and our little princess, Zoharin, besides her own study and work, was really commendable. My study would not have been completed so smoothly without her patience and sacrifices. She stood by me in ups and downs, not only during my PhD study, but also in all stages of my life, and always endowed me with her endless love and support. My dear, you are very special.

Mostafa Wasiuddin Numan

September 2017

Adelaide

Conventions

The following conventions have been adopted in this thesis:

Typesetting

This document was compiled using L^AT_EX2_ε. Texmaker and TeXstudio were used as text editor interfaces to L^AT_EX2_ε. Inkscape and Matlab were used to produce schematic diagrams and other drawings.

Referencing

The Harvard style has been adopted for referencing.

System of units

The units comply with the international system of units recommended in an Australian Standard: AS ISO 1000–1998 (Standards Australia Committee ME/71, Quantities, Units and Conversions 1998).

Spelling

Australian English spelling conventions have been used, as defined in the Macquarie English Dictionary (A. Delbridge (Ed.), Macquarie Library, North Ryde, NSW, Australia, 2001).

Abstract

This thesis investigates network on chip (NoC) architecture, most particularly, NoC mapping algorithms for homogeneous processing elements of a system on chip (SoC) designed for AI and cognitive computing.

Production systems are used in cognitive architectures and knowledge-based systems to produce appropriate reasoning behaviours by matching the symbolic information of the environment with the production rules stored in their knowledge bases. General purpose computers are not specifically manufactured for the purpose of continuous matching involved in production systems, and often fail to deliver the performance and speed required in real-time applications. A reconfigurable and parallel computer architecture, named the Street Processor, has been developed by the research group of which the author is a member, to address the performance gap. The processor has its own instruction set, called the Street language, to define the production rules. The production rules are implemented on simple and identical PEs of the Street Processor that conduct the matching operations in parallel and asynchronously. Special steps can be taken to make these operations synchronous if required. Two artificial agents demonstrate the capability of the Street Processor, and are also used as test cases to measure the performance of NoC mapping techniques.

The Street Processor is expected to contain thousands of fine-grained homogeneous PEs to build a complex cognitive agent. To make the continuous and simultaneous communication among the PEs more efficient, a regular and generic NoC architecture is considered in this work. The network architecture allows multiple PEs to be associated with a single NoC router to optimise its resources. The mapping of PEs to NoC routers, which is an NP-hard optimisation problem, is addressed in this work using two alternative approaches. The Branch and Bound (BB) and Simulated Annealing (SA) techniques are analysed for use as a preferred mapping technique. Although the BB technique provides a mapping solution faster than SA, the latter is considered more promising for large systems, e.g. the Street Processor, since BB achieves the computation time advantage at the cost of a high memory requirement.

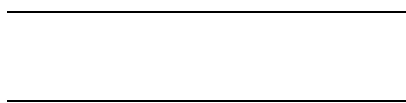
To reduce the computation time of the SA method by shrinking the search space, the dependency graph, which captures the communication volume among PEs over a period of time, is partitioned into smaller groups of PEs (GPEs). By assigning each GPE to a router, this approach also reduces the number of required routers and the inter-router traffic of the network. A Priority-based Simulated Annealing (PSA) technique is proposed, which takes advantages of the relative placements of the routers and inter-dependencies of the GPEs to determine a heuristic initial mapping to start annealing. The experiments show that this approach significantly improves the computation time for finding a solution without sacrificing mapping quality. Considering the inherent memory utilisation advantage over the BB technique, and the computation time improvement over the original SA technique, the proposed approach is suggested to be the most suitable for NoC mapping for the Street Processor and similar homogeneous SoCs.

List of Figures

| | | |
|-------|--|----|
| 1.1 | Transistor counts in 1965 – 2015 | 3 |
| 1.2 | Thesis outline and contributions | 8 |
| <hr/> | | |
| 2.1 | Block diagram of the Soar architecture | 20 |
| 2.2 | Soar processing cycle | 21 |
| 2.3 | Block diagram of the ACT-R architecture | 22 |
| 2.4 | Processing cycle of a production system | 25 |
| 2.5 | Data flow through different layers in NoC and OSI models | 32 |
| <hr/> | | |
| 3.1 | An example of Rete discrimination network | 41 |
| 3.2 | A Street production rule | 44 |
| 3.3 | Example of a inequality check | 45 |
| 3.4 | Example of a negative condition element | 46 |
| 3.5 | Example of a negative action | 47 |
| 3.6 | Example of multiple instantiation of a production rule | 47 |
| 3.7 | A rule to demonstrate bundled actions | 48 |
| 3.8 | A Street production rule susceptible to race conditions | 49 |
| 3.9 | Hardware and software stack on a conventional computer and the Street Processor | 51 |
| 3.10 | Block diagram of a productor | 51 |
| 3.11 | An example of a Big Productor | 53 |
| 3.12 | An example of dependencies between two production rules | 54 |
| <hr/> | | |
| 4.1 | A NoC structure and its major components | 60 |
| 4.2 | Regular NoC topologies | 61 |
| 4.3 | A deadlock situation | 63 |

| | | |
|-------|--|-----|
| 4.4 | Permitted turns in dimension order routing | 64 |
| 4.5 | Data decomposition in wormhole routing | 66 |
| 4.6 | Example of virtual channel flow control | 68 |
| 4.7 | Router micro-architecture with n interfaces, each having v virtual channels | 69 |
| 4.8 | Pipeline stages of a NoC router | 70 |
| <hr/> | | |
| 5.1 | Examples of two mappings M_1 and M_2 | 77 |
| 5.2 | Simulated Annealing algorithm for NoC mapping | 81 |
| 5.3 | An example of Branch and Bound search tree of solution space | 83 |
| 5.4 | Branch and Bound algorithm for NoC mapping | 85 |
| 5.5 | Relationship between the major entities | 89 |
| 5.6 | Traffic graph of the agent solving Branch and Bound mapping | 91 |
| 5.7 | Energy cost in each iteration of Simulated Annealing | 92 |
| 5.8 | Router-wise traffic after Simulated Annealing mapping | 92 |
| 5.9 | Minimum UBC and maximum LBC limiting the best energy cost | 93 |
| 5.10 | Router-wise traffic after Branch and Bound mapping | 94 |
| 5.11 | Energy cost comparison between SA and BB mapping | 95 |
| 5.12 | Computation time comparison between SA and BB mapping | 95 |
| <hr/> | | |
| 6.1 | Example of graph partitioning | 100 |
| 6.2 | The phases of multi-level graph partitioning | 101 |
| 6.3 | The collapse of an edge | 102 |
| 6.4 | Priority-based Simulated Annealing algorithm for GPE mapping | 105 |
| 6.5 | Routers indicating their priorities in 5×5 and 6×6 network structures . | 107 |
| 6.6 | Pseudocode of localised swap | 108 |
| 6.7 | Traffic graph of the Cockroach agent | 112 |
| 6.8 | Computation time comparison between mappings of two Street agents . | 113 |
| 6.9 | Partitioned traffic graph of the Cockroach agent | 114 |
| 6.10 | Energy cost comparison between PSA and SA in each iteration for the Cockroach agent | 115 |

| | | |
|------|--|-----|
| 6.11 | Energy cost comparison between PSA and SA | 116 |
| 6.12 | Computation time comparison between PSA and SA | 116 |



List of Tables

| | | |
|-----|--|-----|
| 2.1 | Approaches to building AGI agents | 15 |
| 2.2 | Classifications of cognitive architectures | 17 |
| 4.1 | Node table based routing of a 3×3 mesh using XY routing | 65 |
| 5.1 | Important Street rules to realise the self-configurable agent | 90 |
| A.1 | List of Street production rules of the self-configurable agent | 128 |
| B.1 | List of Street production rules of the Cockroach agent | 132 |