Routing Algorithms in Networks-on-Chip

Maurizio Palesi • Masoud Daneshtalab
Editors

Routing Algorithms in Networks-on-Chip

Editors

Maurizio Palesi
Facoltà di Ingegneria
Kore University of Enna
Cittadella Universitaria
Enna, Italy

Masoud Daneshtalab
Department of IT
University of Turku
Turku, Finland

ISBN 978-1-4614-8273-4

ISBN 978-1-4614-8274-1 (eBook)

DOI 10.1007/978-1-4614-8274-1

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013950182

© Springer Science+Business Media New York 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

In the year 2000 when the idea of Networks on Chip (NoC) was proposed, many people looked at it sarcastically as “bizarre,” “too complicated,” and “unacceptably inefficient.” In a few years they were proved wrong. But even the greatest optimists did not predict that this new paradigm, for designing inter-core connections for multi-core systems using packet-switched communication, will get unanimous endorsement from academicians as well as semiconductor industries in just a matter of 10 years. But it has happened with the announcement of NoC-based 48-core chip by Intel. Now annually there are many international workshops and special track sessions held in important international conferences specifically dealing with issues related to NoC architecture. Within the NoC architecture area, new ideas in routing algorithm design continue to dominate the research publications. During the earlier years, routing algorithm proposals attempted communication performance improvement by maximizing routing adaptivity (while avoiding any possibility of deadlock) and by reducing congestion in general or application-specific contexts. Recently, researchers have expanded the scope of routing algorithm design by including fault tolerance and lowering power consumption as added objectives along with high performance. Motivated by advances in new technologies, proposals of routing algorithms for 3D architectures and mixed electro-optical or pure optical NoCs have also started appearing in literature.

As is obvious from research publications in various conferences and workshops, NoC is becoming an important topic of research and postgraduate teaching in universities all over the globe. Routing algorithm design is a challenging topic for researchers since it provides the possibility of graph theoretic analysis of any proposed new solution. Thanks to the availability of free NoC simulators, this area provides the possibility of concrete and speedy experimental evaluation of new ideas in NoC routing. Availability of ASIC design tools and FPGA prototyping tools also allows evaluation of cost and power consumption implications of the new ideas. Research in the area of routing algorithms is still flourishing and by no means has reached the saturation point. This book, entitled *Routing Algorithms in Networks-on-Chip*, is a collection of papers describing representative solutions to important aspects and issues related to routing algorithms. This collection does not claim

to include best solutions for any aspect of routing algorithms, nor does it claim a complete coverage of topics related to this important area of NoC architecture. But the book does provide a good source of reference to postgraduate students and researchers getting started in this exciting area.

For many years, both Maurizio Palesi and Masoud Daneshtalab have been very active in research related to various aspects of NoC architecture design in general, and design of routing algorithms in particular. They have made significant and distinctive contributions in the area of routing algorithms. Their contributions, through the organization of NoC-related workshops and special sessions in international conferences, as well as through special issues for various international journals, are well known and highly appreciated by the NoC community. The contacts and knowledge gained by them through these experiences have placed them in a unique position to put together this excellent collection of papers in a book form.

The book is organized in six logical parts such that each part contains papers related to a common theme. For example, Part I contains papers proposing ideas to improve routing performance in NoC platforms. Similarly, Part II collects ideas related to multicast routing in NoC platforms. There are parts, each with multiple chapters, dealing with fault tolerance routing in NoC, power/energy-aware routing, and routing for 3D and optical NoC. The single chapter in the last part describes an industrial case study of routing algorithm in a tera-scale architecture. This organization makes the book directly useable as a reference or as a textbook in a special topic graduate course.

I recommend this book to all those who are new to the area of NoC architecture and NoC routing and want to understand the basic concepts and learn about important research issues and problems in this area. The book will also be useful as a reference source to established research groups as well as industry involved in NoC research. I feel this book will make an important contribution in promoting education and research in NoC architecture.

Jönköping University,
Jönköping, Sweden

Shashi Kumar

Preface

Modern Systems-on-Chip (SoCs) today contain hundreds of Intellectual Properties (IPs)/cores, including programmable processors, coprocessors, accelerators, application-specific IPs, peripherals, memories, reconfigurable logic, and even analog blocks. We have now entered the so-called many-core era. The International Technology Roadmap for Semiconductors foresees that the number of Processing Elements (PEs) that will be integrated into an SoC will be in the order of thousands by 2020. As the number of communicating elements increases, there is a need for an efficient, scalable, and reliable communication infrastructure. As technology geometries shrink to the deep submicron regime, however, the communication delay and power consumption of global interconnections become the major bottleneck. The Network-on-Chip (NoC) design paradigm, based on a modular packet-switched mechanism, can address many of the on-chip communication issues, such as performance limitations of long interconnects and integration of a large number of PEs on a chip.

The overall performance of a network depends on several network properties such as topology, routing algorithm, flow control, and switching technique. This book is focused on routing algorithms. The routing algorithm has a strong impact on several nonfunctional requirements of an NoC-based system. Performance, reliability, energy consumption, power dissipation, thermal aspects, and fault tolerance represent just a short list of the major common metrics affected by the routing algorithm.

The scientific literature related to NoC architectures and design methodologies is mostly dominated by works that address issues concerning the routing algorithms. Unfortunately, although the topic is so important and so widely discussed in the NoC community, there is a lack of structured resources (i.e., books, articles, etc.) aimed at organizing the great deal of literature and information on routing algorithms used in the NoC-based systems. The goal of this book is to provide a unified platform for students (master's, Ph.D.), researchers (from both academy and industry), and practitioners for building the basis of routing algorithms for NoC-based systems as well as providing in-depth discussions on advanced solutions applied (and to be applied) in the current and the next generation of NoC-based many-core SoCs. After

a brief introduction on the basic concepts of on-chip networks, routing algorithms for NoC architectures will be presented and discussed at different abstraction levels – starting from the algorithmic level to their actual circuitual implementation. The impact on current and future key design objectives, namely, power dissipation, energy consumption, thermal aspects, reliability, and performance, will be analyzed and discussed.

This book is organized based on the key problems which affect the current and (may also affect) the next generation of many-core SoCs by putting emphasis on the role played by the routing algorithm. The book is organized into six parts, each of them aimed at presenting a selection of routing algorithms specifically designed for addressing the key issues which characterize the many-core era. The six parts will cover the following macro-topics: Performance Improvement, Multicast Communication, Fault Tolerance and Reliability, Power/Energy and Thermal Issues, Emerging Technologies, and Industrial Case Study.

The book can be used as a reference for university (post)graduate courses and by Ph.D. students, as well as for advanced courses on SoCs. The audience of the book is not limited to various NoC-related research groups; it also might be attractive for faculty and students from other fields when high-performance computing, supercomputers, and many-core systems become the key elements. In fact, routing algorithms play an important role for these elements. As there is much literature related to this subject, we attempt to collect and adjust the state-of-the-art research results, which have been recently published, into the chapters.

Chapter 1 provides the basic concepts of on-chip networks, including network topologies, switching techniques, and routing algorithms. Such topics represent the conceptual bases exploited by the strategies, the mechanisms, and the methodologies discussed in the rest of the book.

The first part of the book consists of a set of chapters related to performance improvement. Chapter 2 presents a system-level framework for designing minimal deterministic routing algorithms for NoCs that are customized for a set of applications. Chapter 3 studies deadlock detection and recovery strategy in NoCs as opposed to deadlock avoidance. Chapter 4 discusses the abacus turn model which allows to keep the network deadlock-free by dynamically applying forbidden turns. Chapter 5 investigates routing algorithms based on learning approaches for balancing the traffic over the network.

The second part of the book focuses on multicast communication. Chapter 6 presents a new method and concept for implementing efficient and deadlock-free tree-based multicast routing algorithms. Chapter 7 addresses how to implement unicast and multicast routing efficiently in 2D and 3D mesh networks.

Fault tolerance and reliability are discussed in the third part of the book. Chapter 8 presents a fault-tolerant routing algorithm that keeps the negative effect of faulty components on the NoC power and performance as low as possible. Chapter 9 provides extensive knowledge on how to develop a fault-tolerant routing algorithm based on the characteristics of a system.

The fourth part of the book addresses power, energy, and thermal issues. Chapter 10 presents the use of bufferless deflection routing for removing input buffers which

are responsible for a significant fraction of the total power budget. Chapter 11 proposes a routing algorithm to reduce the hotspot temperature for application-specific NoCs.

Emerging technologies are explored in the fifth part of the book. Chapter 12 introduces some design concepts of traffic- and thermal-aware routing algorithms in 3D NoC architectures, which aim at minimizing the performance impact caused by the run-time thermal managements. A new architecture for nanophotonic NoCs which consists of optical data and control planes is proposed in Chap. 13.

Finally, in the last part of the book (Chap. 14), an industrial case study illustrates a comprehensive approach in architecting (and micro-architecting) a scalable and flexible on-die interconnect and associated routing algorithms that can be applied to a wide range of applications in an industry setting.

Enna, Italy
Turku, Finland

Maurizio Palesi
Masoud Daneshtalab

Acknowledgments

“Life is like riding a bicycle. To keep your balance you must keep moving.”

“Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.”

– Albert Einstein

A collective work of this nature is not possible without a considerable amount of mutual support. For that, we would like to express our gratitude to all the contributing authors for their hard work and cooperation. We would also like to thank the Springer staff involved in the publication of this book.

Maurizio Palesi would like to pay special thanks to his wife, Cristina Galvagno, for constant support and encouragement and for taking care of Alberto (their son) especially as the deadlines approached.

Masoud Daneshtalab would like to express his deepest gratitude to his wonderful wife, Masoumeh Ebrahimi, for her excellent guidance and patience. Without her love, encouragement, and patience, he would not be able to finish this work. She is always the first one he would go to whenever he needs support, and the first one to share his happiness on his success. Finally, Masoud would like to thank his mother, Mahnaz, for her constant love, support, and prayers.

“Let yourself be drawn by the stronger pull of that which you truly love”

“Two there are who are never satisfied – the lover of the world and the lover of knowledge”

– Rumi

Contents

1	Basic Concepts on On-Chip Networks	1
	Masoud Danashtalab and Maurizio Palesi	

Part I Performance Improvement

2	A Heuristic Framework for Designing and Exploring Deterministic Routing Algorithm for NoCs	21
	Abbas Eslami Kiasari, Axel Jantsch, and Zhonghai Lu	
3	Run-Time Deadlock Detection	41
	Ra’ed Al-Dujaily, Terrence Mak, Fei Xia, Alex Yakovlev, and Maurizio Palesi	
4	The Abacus Turn Model	69
	Binzhang Fu, Yinhe Han, Huawei Li, and Xiaowei Li	
5	Learning-Based Routing Algorithms for On-Chip Networks	105
	Masoumeh Ebrahimi and Masoud Daneshtalab	

Part II Multicast Communication

6	Efficient and Deadlock-Free Tree-Based Multicast Routing Methods for Networks-on-Chip (NoC)	129
	Faizal Arya Samman and Thomas Hollstein	
7	Path-Based Multicast Routing for 2D and 3D Mesh Networks	161
	Masoumeh Ebrahimi, Masoud Daneshtalab, Pasi Liljeberg, Juha Plosila, and Hannu Tenhunen	

Part III Fault Tolerance and Reliability

- 8 Fault-Tolerant Routing Algorithms in Networks On-Chip** 193
 Reyhaneh Jabbarvand Behrouz, Mehdi Modarressi,
 and Hamid Sarbazi-Azad
- 9 Reliable and Adaptive Routing Algorithms for 2D and 3D
 Networks-on-Chip** 211
 Masoumeh Ebrahimi

Part IV Power/Energy and Thermal Issues

- 10 Bufferless and Minimally-Buffered Deflection Routing** 241
 Chris Fallin, Greg Nazario, Xiangyao Yu, Kevin Chang,
 Rachata Ausavarungnirun, and Onur Mutlu
- 11 A Thermal Aware Routing Algorithm
 for Application-Specific Network-on-Chip** 277
 Zhiliang Qian and Chi-Ying Tsui

Part V Emerging Technologies

- 12 Traffic- and Thermal-Aware Routing Algorithms for 3D
 Network-on-Chip (3D NoC) Systems** 307
 Kun-Chih Chen, Chih-Hao Chao, Shu-Yen Lin,
 and An-Yeu (Andy) Wu
- 13 Scalable Architecture for All-Optical Wavelength-Routed
 Networks-on-Chip** 339
 Somayyeh Koochi and Shaahin Hessabi

Part VI Industrial Case Study

- 14 On Chip Network Routing for Tera-Scale Architectures** 379
 Aniruddha S. Vaidya, Mani Azimi, and Akhilesh Kumar