ISSN 2186-7437

NII Shonan Meeting Report

No. 2013-8

Many-cores and On-chip Interconnects

Tomohiro Yoneda
José Flich
Jiang Xu
Michihiro Koibuchi

September 23–25, 2013

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Many-cores and On-chip Interconnects

Organizers:
Tomohiro Yoneda (National Institute of Informatics)
José Flich (Universitat Politècnica de València)
Jiang Xu (The Hong Kong University of Science and Technology)
Michihiro Koibuchi (National Institute of Informatics)

September 23–25, 2013

Over the past few decades, a considerable number of studies have been conducted on the improvement and implementation of VLSIs. It has been recognized that the performance improvement of a single processor core is limited due to clock skew, power consumption, heat dissipation, leakage current, instruction level parallelism, and complexity. As a result, the rise of chip multiprocessor (CMP) and multi-processor system-on-a-chip (MPSoC) has rapidly been gaining pace, and they have become accepted as an integral part of the modern processing architecture. In these multiple core architectures, it has been recognized that a simple bus architecture does not scale with the system size as the bandwidth is shared by all the cores attached to it. Thus, the concern with on-chip networks has been growing as a feasible solution to many-core systems. Recently, it is also reported that fully asynchronous on-chip networks for NoCs have many advantages over the corresponding synchronous designs. On the other hand, as semiconductor process technology scales and on-chip networks become large, routers and links that compose on-chip networks should have tolerance against several kinds of faults. For example, even if one link or router goes down, the remaining part of the network should continue to work. Routers and links should adapt to performance degradation caused by effects like PMOS transistor negative bias temperature instability (NBTI), hot carrier degradation (HCI), VDD drop, temperature increase, and so on. Also, transient faults caused by soft-errors or noises should be tolerated. Furthermore, as more and more complicated applications are run on NoCs, the demand for on-chip networks with low-latency/high-throughput is increasing.

In this meeting, we discussed the future direction of many-core and its on-chip interconnect technologies related to the above issues for supporting strong and weak scaling of parallel applications. Our technical interests include on-chip communication technologies, architectures, methods and applications, and asynchronous design for achieving low-power, high reliability, low-latency and high-throughput computing toward high-performance systems that include not only for High Performance Computing (HPC) but also embedded systems.

Participants

Yuichiro Ajima, Fujitsu Limited
Hideharu Amano, Keio University
Peter A. Beerel, University of Southern California
Davide Bertozzi, University of Ferrara
Sébastien LE BEUX, Institut des Nanotechnologies de Lyon
Krishnendu Chakrabarty, Duke University
José Flich, Universitat Politècnica de València
Ikki Fujiwara, National Institute of Informatics
Ran Ginosar, Technion
Paul V. Gratz, Texas A&M University
John Kim, Korea Advanced Institute of Science and Technology
Kenji Kise, Tokyo Institute of Technology
Michihiro Koibuchi, National Institute of Informatics
Tao Li, University of Florida
Yun Liang, Peking University
Olav Lysne, University of Oslo
Hiroki Matsutani, Keio University
Yuichi Nakamura, NEC Corp.
Jiang Xu, The Hong Kong University of Science and Technology
Lin Yang, Institute of Semiconductors, Chinese Academy of Sciences
Tomohiro Yoneda, National Institute of Informatics
Tsutomu Yoshinaga, University of Electro-Communications

Program

Monday (Sep. 23) Chip Design and Networks

8:45 - 9:25 Welcome, Introducing ourselves (Chair: Tomohiro Yoneda)

9:25 - 10:45 Many-core Design (Chair: Tomohiro Yoneda)

Hideharu Amano
Ran Ginosar

11:00 - 12:20 Off-chip Network Design and Layout (Chair: John Kim)

Michihiro Koibuchi
Yuichi Nakamura

1:30 - 3:30 Walking

3:40 - 5:00 On- and Off-chip Network Design (Chair: Hideharu Amano)

Ikki Fujiwara
John Kim

5:10 - 6:30 Off-chip Networks (Chair: José Flich)

Yuuichirou Ajima
Olav Lysne

Tuesday (Sep. 24) NoC

8:45 - 10:45 NoC Design (Chair: Tao Li)

Hiroki Matsutani
Davide Bertozzi
Krishnendu Chakrabarty

11:00 - 12:20 Photonic Interconnect (1) (Chair: Davide Bertozzi)

Sébastien Le Beux
Lin Yang

1:30 - 2:50 Photonic Interconnect (2) (Chair: Sébastien Le Beux)

Jiang Xu
Tsutomu Yoshinaga

3:00 - 5:00 Async Design & NoC (Chair: Ran Ginosar)

Peter Beerel
Tomohiro Yoneda
Paul V. Gratz

5:15 - 6:30 Panel: Interconnects for Manycores

Moderator: Jiang Xu

Panelist:

Yuuichirou Ajima
Ran Ginosar
José Flich
Davide Bertozzi
Paul Gratz
Michihiro Koibuchi

Wednesday (Sep. 25) Many-core Architecture

8:45 - 10:05 Many-core Architecture (Chair: Olav Lysne)

Tao Li
Eric Liang

10:15 - 11:35 Many-core Architecture and NoC (Chair: Eric Liang)

Kenji Kise
José Flich

11:45 - 12:00 Wrap-up (Chair: Jóse Flich)

Overview of Talks

Highly-scalable and light-weight design of the Tofu interconnect

Yuichiro Ajima, Fujitsu Limited

The Tofu interconnect is an interconnection network designed for the K computer and its commercial version Fujitsu PRIMEHPC FX10. Tofu interconnects tens of thousands of nodes. The network topology of the Tofu interconnect is a highly-scalable six-dimensional mesh/torus. Some dimensions are configured as rings and contribute to the availability and the serviceability of the system. The packet delivery system and endpoint system are holistically designed to make the communication protocol of the Tofu interconnect light-weight. The protocol depends on guaranteed and in-order packet delivery.

Building Block Networks with Wireless Inductive Coupling Though-Chip Interface

Hideharu Amano, Keio University

Inductive Coupling Though-Chip Interface (TCI) connects stacked chips by coils only with existing IC interconnections. Over-Gb/s data transfer rate can be achieved with less than 10mW power dissipation. Parallel data bits can be multiplexed into one single coil and burst-transferred. With TCI, a high speed network can be formed just by stacking multiple chips in various forms. A heterogeneous multi-core system called Cube-1 consisting of an embedded CPU and multiple accelerators is now available. By using TCI, a building block network, which has intermediate properties between Network-on-Chips and wireless ad-hoc networks, is formed by combining chips in various structures.

Why Asynchronous Interconnect?

Peter A. Beerel, University of Southern California

I will focus on the advantages of asynchronous interconnect in SoC and NoCs. I will first present an argument of feasibility and design efficiency showing the most recent 1.2B Transistor Switch Chip by Intel that is 90% fully asynchronous. It was designed using an automated asynchronous synthesis and place-and-route flow for logic blocks and a high-performance interconnect using full-custom clock domain crossing logic, asynchronous links and cross-bars. I will then review the advantages of asynchronous design over synchronous counterparts in terms of low latency, high-performance, and robustness to process variations, aging, and temperature.

Evolutionary and Revolutionary Technologies for Low Power On-Chip Communication

Davide Bertozzi, University of Ferrara

The advent of networks-on-chip is far from stabilizing the domain of on-chip communication architectures for multi- and many-core systems. For the high-performance computing domain, NoCs are a non-negligible source of power dissipation. For the embedded computing domain, the NoC design point stems from a trade-off between maximum resource utilization and communication performance, ultimately ending up in a system-level energy optimization issue. Low-power on-chip communication can be achieved via evolutionary design techniques (e.g., by removing the clock and implementing clockless switching), or by means of disruptive technologies such as on-chip optical links. This talk will address the latest research findings on these technologies, taking the viewpoint of their crossbenchmarking against (aggressive) reference NoC implementations. The ultimate source of debate will be where, when and how such technologies will become viable for actual design in industry, and about how to accelerate this process.

Highly regular and reconfigurable ONoC

Sébastien LE BEUX, Institut des Nanotechnologies de Lyon

Optical on-chip interconnects enable significantly increased bandwidth and decreased latency in MPSoC. However, the interfaces between electronic and photonic signals imply strong constraints on the layout of the 3D architecture and may impact the system scalability. The scalability also relies on the flexibility level of the network. This presentation deals with a regular layout for an ONOC used to interconnect processing elements located on different electrical layers. The flexibility issue of the ONoC is addressed by considering the use of reconfigurable blocks on the interfaces.

Test-Delivery Optimization in Manycore SOCs

Krishnendu Chakrabarty, Duke University

A network-on-chip (NOC) enables the integration of the hundreds and even thousands of cores in a many core system-on-chip (SOC). Efficient testing and design-for-testability techniques must be developed for such!Hmonster !Ichips. I will describe test-data delivery optimization algorithms for manycore SOCs with hundreds of cores, where a network-on-chip (NOC) is used as the interconnection fabric. I will first present an optimization algorithm based on a subset-sum formulation to solve the test-delivery problem in NOCs with arbitrary topology that use dedicated routing. Next I will propose an algorithm for the important class of NOCs with grid topology and XY routing. The proposed algorithm is the first to co-optimize the number of access points, access-point locations, pin distribution to access points, and assignment of cores to access points for optimal test resource utilization of such NOCs. Test-time minimization is modeled as an NOC partitioning problem and solved with dynamic programming in polynomial time. Both the proposed methods yield high-quality results and are scalable to large SOCs with many cores. Test scheduling under power constraints is also incorporated in the optimization framework.

On-Chip Networks, what's really new?

José Flich, Universitat Politècnica de València

The on-chip network paradigm appeared in the last decade as a solution for the connection of components inside a chip. While this is a covered need, the questioning is whether this paradigm is just a new domain field of already known solutions and techniques (from the off-chip communication interconnects hugely researched and practiced in the past) or challenges with new milestones to be achieved (asking for new methods and tools yet to discover). While it is commonly agreed constraints and requirements are different from the off-chip interconnect world, the current proposed solutions are highly similar and suggest an evolutionary approach. In this talk I will focus on what's new and what's now and will provide some basic examples of key differentiating research only applicable to on-chip networks.

Does light speed affect topologies?

Ikki Fujiwara, National Institute of Informatics

A massively parallel application run on a future supercomputer is expected to require very low end-to-end latencies. Most of off-chip interconnection topologies does not consider cable delay (i.e. near light speed), because switch delay (some hundred nanoseconds) dominates the end-to-end latency. So, what if a ultra-low-delay (some ten nanoseconds) switch becomes available in the near future? Do traditional off-chip topologies work well in those situations? We would like to discuss about the topology design for the future supercomputing systems, as well as for the future on-chip networks.

Mathematical modeling of many-cores

Ran Ginosar, Technion

Many-cores come in many flavors: mesh-noc tiled arrays (e.g. Tilera), hierarchical multi-cores (e.g. Rigel), hierarchical multi-threading (e.g. Nvidia), SIMD and associative processors. Comparing them for performance, power, area and ease of programming is a fuzzy art at best, typically requiring the construction of complete applications, optimizing them separately for each architecture and executing or simulating them. The results are not always convincing and we often end up just where we started. We attempt to extend mathematical analysis of architecture to this field. A model accounts for performance and power of cores as a function of area and other parameters. On-chip memories are also modeled. Basic axioms such as Amdahl's law and Pollack's rule are employed to formulate the model. However, adapting the model to a variety of many-core architectures remains a challenge.

Fault Prevention in Future Network-on-Chip Interconnected Chip-Multiprocessors

Paul V. Gratz, Texas A&M University

Today multi-core chip-microprocessors (CMPs) contain tens of interconnected cores or tiles. As the number of CMP cores increases, on-chip interconnection networks (NoCs) are necessary to for efficient inter-tile, on-chip communication. Unfortunately, deep sub-micro CMOS is marred by increasing susceptibility to wear-out. Prolonged operational stress gives rise to accelerated wear-out and failure due to one of several failure mechanisms. In many-core CMPs, while an individual core's wear-out may not necessarily be catastrophic for the system, a single fault in the NoC could render the entire chip useless. Recently proposed fault-tolerant routing algorithms respond to wear-out by rerouting traffic around faulty components and topological NoC regions so as to safely deliver inter-core communication, however, these approaches are reactive to faults. This talk explores the viability of proactive techniques to maintain NoC components through data manipulation and wear-leveling. In particular, the talk examines the failure mechanisms expected to dominate in future process technology, and how they are influenced by network traffic generated by real application workloads. The talk then explores how these failure mechanisms, together with inherent process variation, map on a router microarchitecture under typical application workloads. Finally, we develop a wear-resistant router microarchitecture whereby the wear induced by usage is lessened through a management of the per-component, bit-transitions, balancing transition sensitive failure mechanisms against level sensitive ones.

Alternative Interconnection Networks

John Kim, Korea Advanced Institute of Science and Technology

Historically, significant research has been done in large-scale off-chip networks and recently, there has been significant work done on on-chip networks. Each interconnection networks have different constraints and as a result, the resulting architectures can be different while still sharing some similarities. In this talk, we will explore how we can apply interconnection networks to other domains. In particular, we will discuss how interconnection networks can be leveraged in other ways - including the design of an internal switch microarchitecture and be leveraged for use as a memory network.

Challenges for Dependable Many-Core Processors

Kenji Kise, Tokyo Institute of Technology

I will discuss the dependability issues, in particular soft errors and timing errors, for many-core processors. One of our proposals is a NoC-based DMR mechanism named SmartCore to detect transient errors on a many-core processor. It is unique because the packet level comparison for error detection is done by the new designed NoC router.

Is it an innovative technology for Datacenter and HPC networks?

Michihiro Koibuchi, National Institute of Informatics

I would like to discuss our 40Gbps off-chip !Hwireless !Ilink technology. As supercomputers become large, interconnection networks face two problems to be resolved: (1) reduce the aggregate cable length, e.g. over two thousands meters in a machine room, and (2) design a network topology optimized to various communication patterns of parallel applications. To mitigate two problems, we briefly discuss about the possibility of free space optical communication at free space between the ceiling and top of cabinets.

Architecting Technology Enabled Network On Chip: Challenges and Opportunity

Tao Li, University of Florida

Network on chip (NoC) has become an imperative communication fabric in the era of multi-/many- core architecture design. Nevertheless, the impact of semiconductor technology scaling, emerging technology integration, and throughput oriented core architecture have made reliable and power efficient NoC design increasingly challenging. In this talk, I will first discuss the implication of nano-scale semiconductor fabrication and silicon photonic integration on NoC design and introduce cross-layer optimizations to improve NoC dependability and run-time efficiency. I will then address NoC design issues in throughput GPGPU processors and highlight some promising design paradigms.

GPU Acceleration and Performance Optimization

Yun Liang, Peking University

Graphics processing units (GPUs) are increasingly important for general-purpose parallel processing performance. GPU hardware is composed of many streaming multiprocessors, each of which employs the single-instruction multiple-data (SIMD) execution style. This massively parallel architecture allows GPUs to execute tens of thousands of threads in parallel. Thus, GPU architectures efficiently execute heavily data-parallel applications. However, the performance of GPU applications critically depends on the compiler optimization. If it is not done right, it will seriously hurt the performance. In this talk, I will first present a case study of accelerating 3D sound localization using GPUs. Then, I will present the modeling and optimization techniques we have developed including control flow divergence modeling, register and thread structure optimization, and cache passing optimization.

Towards a self-adaptive network architecture for clouds

Olav Lysne, University of Oslo

Recent studies show that there is a significant performance gap between commercial clouds and conventional clusters equipped with equivalent processors, and that cloud computing is not mature enough to support, for instance, high performance computing. This is partly due to the lack of elastic and efficient provisioning of network resources in the cloud. We therefore argue that the following research topics are important:

1. Methods for self-adaptive and predictable provisioning of network resources,
2. Methods for high-granularity service differentiation across a set of resources,
3. Methods for self-detection of failures.
4. Methods for fault-tolerance and robust computing
5. Elastic virtualization methodologies for efficient up and down scaling.

Integrating Hardware Network Stack and Database Processing Engines for Big Data

Hiroki Matsutani, Keio University

We are facing two competing trends in ICT: Big data and green datacenters. Since data reuse and repurposing are now expected to make innovations, IT equipments will be rapidly augmented for Big data, while energy-saving is essential for datacenters from a preventing global warming point of view. To fill in the gap between these trends, we are studying FPGA-based database processing engines that support various structured storages or polyglot persistence. Since main bottleneck of conventional software-based memcached is related to TCP/IP stack, we are now considering the integration of hardware-based network stack and these database processing engines in FPGA-based platforms.

The layout evaluation and hierarchical layout method of MPSoC

Yuichi Nakamura, NEC Corp.

This talk presents how to layout many-core processors SoCs. Currently, the layout design time is quite significant for large scale LSIs, because of the complexity involved with the verification of timing and signal integrity constraints. The size and complexity of many-core SoC, limit the possibility to perform hierarchical layout designs. However, there are various methods for the hierarchical layout designs. In general, a strict hierarchical design method can provide ease of reconfigurability, but it results in worse area and timing with respect to a flat layout method, which, on the other hand, does not provide reconfigurability. To solve these problems and questions about the hierarchical layout design, the trials and the evaluations are applied for Network On Chip (NoC), which is the typical implementation of many-core processors SoC. A NoC which connects IP cores by network interfaces can be easily reconfigured during place and route and it has a strong regularity. In this talk, the several layout evaluation results of NoC and the discussions are presented. For example, it is confirmed that the strict hierarchical design method even makes the poor layout results for NoC. Furthermore, a reconfigurable layout method for Networks-on-Chip (NoCs) based on partial re-layout is introduced. I also welcome discussions on layout issues for many-core processor SoC.

Opportunities and Challenges in Inter/Intra-Chip Optical Networks

Jiang Xu, The Hong Kong University of Science and Technology

The performance and energy efficiency of a multi-core system is determined by not only its processor cores but also how efficiently they collaborate with each other. As new applications continuously require more communication bandwidth, metallic interconnects gradually become the bottlenecks of multi-core systems due to their high power consumption, limited bandwidth, and signal integrity issues. Optical interconnects are promising candidates to bring low power, high bandwidth, and low latency to address inter-chip as well as intra-chip communication challenges. Silicon-based photonic devices, such as optical waveguides and microresonators, have been demonstrated in CMOS-compatible fabrication processes and can be used to build inter/intra-chip optical networks. This talk will discuss the opportunities and challenges of this emerging technology based on our recent findings.

Optical modulators and routers for Photonic Networks-on-Chip

Lin Yang, Institute of Semiconductors, Chinese Academy of Sciences

The performance of chip multiprocessor (CMP) is determined not only by the number of the processor cores integrated on a die, but also by how efficiently they collaborate with each other. With more processor cores being integrated on a die, interconnects in CMP gradually move from traditional bus interconnects to more sophisticated networks-on-chip (NoC). With CMP continuously requiring more communication bandwidths, metallic-based electrical NoC gradually becomes the bottleneck for improving the performance of CMP due to its high power consumption, limited bandwidth and long latency. Recent studies have demonstrated that photonic NoC is a potential solution to overcome the limitations of its electrical counterpart. Many architectures for photonic NoC have been widely studied, such as Mesh, Fat Tree and Clos. While, recent studies are gradually focused on Mesh NoC due to its symmetric architecture, good scalability, simple routing algorithm and easy implementation. A series of optical devices are required to construct the optical interconnect between two processor cores, such as lasers, modulators, multiplexers, waveguides, de-multiplexers and detectors. Such a point-to-point interconnect is the simplest communication mode. However, for network application environment, the processor core is required to communicate with other processor cores. This function is usually completed by optical router, which is located at each node of photonic NoC and connects the local processor core with other remote nodes. In this paper, we will review the status of optical modulators and routers for photonic networks-on-chip and introduce our efforts on these topics. In the first section, we will introduce the 40 Gb/s carrier-depletion Mach-Zehnder silicon optical modulator with very a large optical bandwidth. In the second section, we will introduce a universal method for constructing the N-port non-blocking optical router for photonic NoC.

Executing safety-critical embedded applications on many-core systems

Tomohiro Yoneda, National Institute of Informatics

We are working for an approach to executing safety-critical embedded applications dependably using redundancy available in many-core systems. In this approach, each task of applications is loaded in several processor cores, and usually two cores execute the same task simultaneously using the same inputs. The results of the task are sent to an IO-core, and compared there. If a mismatch is found, the task is executed again, but using three cores, to find the correct results and a faulty core. If a faulty core is successfully detected, it is excluded from the system, and tasks are continuously executed on a reconfigured system. This idea for dependable task execution is simple, but its implementation is not so straightforward. For example, how can we maintain and update the state variables of tasks for the temporary TMR execution, and how can the IO-core be implemented dependably? How about maintaining real-time properties? We have suggestions for some of them, but there are still open issues. This talk will discuss these implementation issues for dependable task execution on many-core systems.

A Fully Optical Ring Network-on-Chip with Static and Dynamic Wavelength Allocation

Tsutomu Yoshinaga, University of Electro-Communications

Silicon photonics Network-on-Chips (NoCs) have ! ! emerged as an attractive ! ! solution to alleviate the high power consumption of traditional electronic ! ! interconnects. In this work, we propose a fully optical ring NoC that ! ! combines static and dynamic wavelength allocation communication mechanisms. A different wavelength-channel is statically ! ! allocated to each destination ! ! node for light weight communication. Contention of simultaneous ! ! communication requests from multiple source nodes to the destination is solved by a token based arbitration for the particular wavelength-channel. For heavy load communication, a multiwavelength-channel is available by requesting it in execution time from source node to a special node that ! ! manages dynamic allocation of the shared multiwavelength-channel among all nodes. We combine these static and dynamic communication mechanisms in a same network that introduces selection techniques based on message size and congestion information. We discuss performance of the proposed photonic NoC based on preliminary simulation results.