Editorial

Networks on chips

Networking has been proven in the computer system arena to be an extremely effective means of managing parallel communication flows in distributed systems. By distilling the most applicable concepts from this domain and by applying them in a way that suits the constraints of semiconductor design, Networks-on-chip (NoCs) have been proposed as the communication backbone for large-scale integrated systems.

NoCs are already used for Multi-Processor Systems-on-Chip (MPSoC) in the embedded systems domain, where multiple programmable processors are accompanied by large numbers of hardware accelerators. NoC-based SoCs can achieve higher performance at lower cost, in combination with higher programmability. Designers of high-performance microprocessors plan to take full advantage of this disruptive interconnect technology, as early research-oriented prototypes of many-core microprocessors like the Intel Polaris chip prove. For both domains, latency minimisation and/or tolerance remain a big challenge.

Currently, the superiority of NoCs with respect to state-of-the-art interconnect fabrics, mostly in terms of operating speed and scalability, is well understood, NoCs being competitive already in a 130 nm technology node. Even the implications of bringing NoCs to a 65 nm technology node have been investigated. Moreover, guiding principles for the design of basic network building blocks (switches and network interfaces) are consolidated. These achievements can be considered a major milestone in network-on-chip research.

At this time, it is becoming apparent that removing roadblocks for an effective NoC utilisation in future industrial products implies not only a knowledge of basic architecture design techniques and physical design principles, but also a deeper understanding of cross-layer design trade-offs. NoC design should be integrated into a cooperative design platform filling the gap between the design layers (from application to physical design) and enforcing cross-layer design and optimisation.

This Special Issue serves the purpose of collecting timely and selected research contributions on this new frontier of NoC design. The specific focus will be on the architecture layer, while trying to capture how the awareness of the upper and lower design layers affects NoC architecture design.

In this direction, the network architecture might be configured and its parameters tuned based on the knowledge of application requirements. For instance, if we are designing a NoC specifically for a set of applications, then it is desirable to determine minimal sizes for the network buffers, as they are major contributor to NoC power and silicon area, while meeting application real time requirements. In 'Enabling Application-Level Performance Guarantees in Network-Based Systems on Chip by Applying Dataflow Analysis', authors A. Hansson et al. propose to capture the behaviour of the NoC and the application using a dataflow model. This allows them to verify the temporal behaviour and to compute buffer sizes using existing dataflow analysis techniques.

Especially in the embedded computing domain, where an accurate characterisation of the communication traffic is usually feasible, it should be possible to design highly efficient routing algorithms by using information regarding the communication behaviour of the application. In 'Bandwidth Aware Routing Algorithms for Networks-on-Chip Platforms', authors M. Palesi, S. Kumar and V. Catania propose an embodiment of this approach. They use off-line analysis to estimate expected load on various links in the network. The result of this analysis is used along with the available routing adaptivity in each router to distribute less traffic to links and paths which are expected to be congested.

Finally, many MPSoC applications require multicast communication services from the network and not just unicast ones, such as replication, barrier synchronisation, cache coherency and clock synchronisation. Although multicast communication can be implemented by multiple unicast messages, this is clearly an inefficient solution. In '*Low-Distance Path-Based Multicast Routing Algorithm for Networks-on-Chips*' from M. Daneshtalab *et al.*, authors refine an inspiration from the multi-computer domain and come up with an adaptive multicast routing for wormhole switched 2D meshes. The algorithm makes use of network partitioning, optimised destination ordering, and the odd-even turn model adaptive routing technique for both multicast and unicast messages. Additionally, the algorithm invokes non-congested paths in routing the messages to prevent creating highly congested areas.

Application-specific networks and multicast routing can be even co-designed. In '*Joint Multicast Routing and Network Design Optimisation for Networks-on-Chip*', authors S. Yan and B. Lin consider the problem of synthesising custom network-on-chip architectures that support both unicast and multicast traffic flows. They formulate the joint multicast routing and network design problem using a rip-up and reroute procedure, where each multicast step is formulated as a minimum directed spanning tree problem. This formulation efficiently captures the best routing solutions for multicast flows during the topology synthesis procedure.

While the above works show how the on-chip network can be optimised based on the requirements of the software layers, the unrelenting pace of technology scaling to the nanoscale regime also causes a widening gap between abstract architecture models and the intricacies of physical design. As a workaround, silicon-aware decision making should be enforced at each layer of the design hierarchy. In practice, NoC designers should be aware that the manufacturing process is a growing source of uncertainty, resulting in the spread of circuit parameters and/or in the malfunctioning of some logic blocks. The ability to deal with such uncertainty is a distinctive challenge for the architecture designers of on-chip networks. In '*Efficient Implementation of Distributed Routing Algorithms for NoCs*', authors S. Rodrigo *et al.* move from the assumption that although regular 2D mesh topologies are designed, manufacturing faults may break their regularity, thus calling for topology agnostic routing algorithms. Although forwarding tables are usually employed at the switches to implement such algorithms, their scalability in terms of access latency, area and power is questioned. This motivates their proposal for a logic based distributed routing, able to mimic the performance of routing algorithms implemented with routing tables, both in regular and irregular topologies.

Even the use of cross-layer design and optimisation techniques does not avoid the need for efficient and effective test strategies for the newly introduced NoC-based systems to reduce their manufacturing costs and meet today's stringent time-to-market requirements. While on-chip networks pave the way for new testing strategies of NoC-based systems (they could be reused as a test access mechanism to the integrated cores, thus saving dedicated busses to transfer test data), they raise the new challenge of testing the network structures themselves (routers and communication channels). In '*DfT-Based External Test and Diagnosis of Mesh-Like NoCs*', authors J. Raik and R. J. Ubar propose a new concept of test and diagnosis in regular mesh-like network-on-chip architectures. The goal of their work is to propose a method for targeting manufacturing faults in the NoC switches based on external test configurations to be applied from the network boundaries. Authors also consider diagnosis of faulty links in the NoC routing framework.

Asynchronous design offers an attractive solution to address the problems faced by NoC designers, especially timing constraints and power consumption. Nevertheless, post-fabrication testing is a big challenge to bring the asynchronous NoC to the market due to a lack of testing methodologies and support. In '*Design-for-Test Approach of an Asynchronous Network-on-Chip Architecture and its Associated Test Pattern Generation and Application*', authors T. Xuan-Tu *et al.* present the design and implementation of a design-for-test architecture, which improves the testability of an asynchronous NoC architecture. The testing strategy for the full NoC is also illustrated, leveraging a simple method for test pattern generation which leads to high fault coverage for single stuck-at fault models.

While the above works build up the first section of this Special Issue and push the envelope on NoC architecture design techniques, the second section takes a wider perspective. It shows how the use of NoC interconnect technology can even lead to novel solutions for the global system architecture.

Non-uniform cache architectures (NUCA) have been proposed as a novel design paradigm for large last-level on-chip caches in order to reduce the effects of wire delays, which significantly limit the performance scaling of today's high clock frequency microprocessors. This is achieved by the adoption of a storage structure partitioned into sub-banks and by the adoption of a fast interconnection network to connect the banks and the cache controller. In '*Impact of On-Chip Network Parameters on NUCA Cache Performance*', authors A. Bardine *et al.* prove that the high sensitivity of the NUCA-based system to the cut-through latency of network routers limit the effectiveness of the NUCA solution. Moreover, varying the buffer capacity has almost negligible effects on the overall performance. From these considerations, authors have identified an alternative

NUCA organisation based on the clustering of banks. Assuming 4 banks per node, the strong constraints on router latency can be relaxed and multi-cycle routers can be more easily employed in high-performance systems.

Stream processing applications such as video and image processing usually comprise series of tasks organized in the form of a pipeline with data dependencies mainly between adjacent tasks in the pipeline. For efficient utilisation of multi-core processors for various stream applications, it is important to design on-chip interconnections capable of managing concurrent data transactions initiated by the multiple processing cores. In '*Memory-Centric NoC for Power Efficient Execution of Task-Level Pipeline on a Multi-Core Processor*', authors K. Donghyun *et al.* propose a memory-centric network-on-chip to facilitate mapping various types of task-level pipelines on multi-core processor architectures. In practice, the memory-centric NoC manages producer-consumer data transactions between the pipelined tasks, thus relieving the burden of the software layers and resulting in power-efficient stream processing.

In '*Processing While Routing: a NoC-based Parallel System*', authors S. Fernandes *et al.* propose a novel and exotic system architecture which also addresses the problem of NoC area and power overhead. In practice, the on-chip network is considered not only as the system interconnect but also as the processing datapath, thus affecting the model of computation. Beyond moving packets closer to their destination, routers also perform the most common logic, arithmetic and synchronisation operations usually found in applications. These latter are described in a packet format, including instructions which are executed as the packets flow across the network.

The final paper of this Special Issue addresses a concern which is orthogonal to all the research efforts presented beforehand. In '*Application Modeling and Hardware Description for Network-on-Chip Benchmarking*', authors E. Salminen *et al.* lay the groundwork for a standardised NoC benchmark set. This work meets the need of NoC developers to assess the gain brought by novel features and of system integrators to select the most suitable NoC and its configuration for the platform at hand. Currently, test cases are generally proprietary, small and poorly documented, thus preventing a common reference for comparative evaluation of NoC solutions.

We sincerely hope you will enjoy this Special Issue and that it will inspire further research in this very important area of electronic system design. We would like to thank all authors who submitted papers to this Special Issue. Special thanks go to the referees for their time and diligence during the review process and for coming up with high-quality reviews. In conclusion, we would like to thank Bashir Al-Hashimi, Editor-in-Chief of *IET Computers and Digital Techniques*, for offering us the opportunity to bring about this Special Issue.

DAVIDE BERTOZZI
ENDIF, University of Ferrara
Ferrara 44100, Italy
E-mail: dbertozzi@ing.unife.it

KEES GOOSSENS
Delft University of Technology, Delft
The Netherlands.
NXP Semiconductors, Eindhoven,
The Netherlands.
E-mail: kees.goossens@nxp.com