Simulation and Analysis of Network on Chip Architectures: Ring, Spidergon and 2D Mesh

Luciano Bononi, Nicola Concer
Dipartimento di Scienze dell'Informazione, Università degli Studi di Bologna,
Mura Anteo Zamboni 7, 40126, Bologna, Italy
{bononi, concer}@cs.unibo.it

Abstract

NoC architectures can be adopted to support general communications among multiple IPs over multi-processor Systems on Chip (SoCs). In this work we illustrate the modeling and simulation-based analysis of some recent architectures for Network on Chip (NoC). Specifically, the Ring, Spidergon and 2D Mesh NoC topologies have been compared, both under uniform load and under more realistic load assumptions in the SoC domain. The main performance indexes considered are NoC throughput and latency, as a function of variable data-injection rates, source and destination distributions, variable number of nodes. Results show that the Spidergon topology is a good trade-off between performance, scalability of the most efficient architectures inherited from the parallel computing systems design, constraints about simple management, small energy and area requirements for SoCs.

1. Introduction

A new generation of communication infrastructures called Networks on Chip (NoCs) [1-15] have been recently considered as a novel alternative to existing On Chip Communication Architectures (OCCAs) based on shared communication medium like on-chip buses (as an example, ARM AMBA [16], Wishbone [17], STBus [20], Core Connect [18], Sonics Backplane [19]. The design of Network on Chip solutions can be considered in between the classical networking solutions (good for scalability and flexibility to adapt to general communication patterns) and the more specific communication and switching architectures for high-performance parallel computing (good for performances but less scalable and less flexible under the dynamic configuration viewpoint). A natural step has been the attempt to inherit the consolidated solutions in these two domains into the SoC world. This led to proposals for packet-switched micro-network backbones based on appropriate protocol stacks. Most recent NoC architectures have been implemented on top of ring, 2D mesh or custom topologies: UPMC/LIP6's DSPIN [2], Nostrum (KTH) [15], Æthereal (Philips Research Lab) [11], Raw network (MIT) [12], Eclipse (VTT) [13], Xpipes (University of Bologna) [14]. One of the most recent architectures and topologies proposed is the novel Spidergon (STMicroelectronics) NoC architecture [9, 10], which will be sketched in section II.

Many research issues are still open and the complex design space for NoCs requires a deep exploration, by involving at least three identified fields: synthesis of communication infrastructure, selection of communication paradigms and application-mapping optimization [7]. This led to the definition of many, correlated, open research problems, still requiring global solutions, techniques and tools to assist designers at many levels [7, 8]. The list includes: topology synthesis, channel and buffer sizing, floorplanning, routing and switching techniques, flow control techniques, data scheduling, buffer and queues management, IP mapping over the NoC architecture, performance evaluation and resource planning, end-to-end services, Quality of Service, packet and message format, deadlock avoidance [7, 8].

Recent works have investigated and compared many NoC architectures under general assumptions [6].

The main contributions of this paper are the following: *i)* the modeling and simulation-based analysis of low degree topologies (Ring, 2D Mesh and Spidergon) focusing the on chip domain requirements and overcoming the classical parallel computing and networking results; to the best of our knowledge, this is the first work considering irregular mesh topologies, and this analysis is motivated since regular meshes cannot be always assumed as realistic topologies, *ii)* the first deep analysis of the novel Spidergon NoC is presented, resulting in a good compromise among the well-known topologies, and *iii)* we propose the OMNeT++ framework [21] for a fast and high level simulation environment for NoC topologies' exploration.

This work was supported by MIUR and University of Bologna and STMicroelectronics project funds "modeling and analysis of network protocols for Spidergon-based networks on chip".

The paper structure is the following: in section II we sketch the Ring, Mesh and Spidergon topologies considered in our study, by illustrating some results of the analysis of their characteristics; in section III we illustrate the modeling and simulation scenarios, and the performance results; in section IV we draw conclusions and future work.

2. The considered NoC architectures

NoCs must have regular, scalable and simple network topology, characterized by the space locality of modules connected by short links, the high correlation of the link traffic, the severe energy and latency constraints, and the need of low cost solutions. In packet-based NoC communication each packet is split into data units called flits. The buffer queues for channels are defined as multiples of the flit data unit. The packet forwarding among nodes is performed with a flit-by-flit (adaptive, source, arithmetic or table-driven) routing and local signal-based flow control. The most generally adopted switching scheme is the *wormhole* scheme. In wormhole, the head flit of a packet is actively routed towards the destination by following the forwarding indications on routers, while subsequent flits are passively switched by pre-configured switching functions on the output queue of the channel belonging to the path opened by the head flit. When the channel buffer space is available on the input queue of the channel towards the next switch in the path the next flit of a packet is forwarded on. Flit-based wormhole is an interesting solution compared to virtual cut-through and packet-based circuit switching because its pipelined nature facilitates flow control and end-to-end performances, with low packet-overheads and low buffering space. Wormhole realizes a tradeoff between circuit-switching performances and packet-switching flexibility and resources utilization. Due to the distributed and partial capture of buffers and channel resources and the possible circular waiting, deadlock and livelock conditions are possible. The management of deadlock solutions and the efficiency of link utilization often introduced the virtual channels (VCs) management. VCs are implemented by multiple output queues for each physical link, and respective buffers. The IPs are connected to a NoC switch by a Network Interface (NI) incorporating the connection management and the data fragmentation functions.

Some of the most common NoC architectures belong to the classes of the Ring (figure 1.b) and m*n 2D Mesh (figure 1.c). The Spidergon architecture with N (even) nodes is similar to a ring enriched by across links between opposite nodes (see figure 2.a). For a tagged node, a clockwise, counterclockwise and across links are present. Some of the most interesting characteristics of the Spidergon scheme are: i) network with regular topology, ii) vertex symmetry (same topology appears from any node), iii) edge-transitivity, iv) constant node degree (equal to 3) translating in simple router HW and efficiency. High node degree reduces the average path length but increases complexity. By assuming channels as unidirectional pairs of links, the number of network links in a N nodes' network is 2N for Ring, 3N for Spidergon and 2(m-1)n+2(n-1)m for a (m*n=N) 2D Mesh.

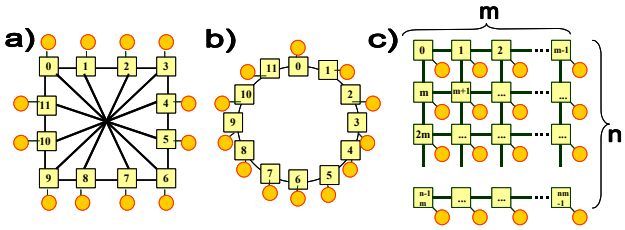

Figure 1: NoC topologies: IPs connected to numbered nodes on a) Spidergon, b) Ring, c) (m*n) 2D Mesh

A significant worst case index, named the network diameter *ND* is defined as the maximum shortest path length between any pair of nodes in the topology. The average network distance *E[D]* is defined as the average path length of all different paths in the network. By assuming a NoC of N nodes, in a Ring topology, *ND=floor(N/2)* and *E[D]=N/4*, in (m*n) 2D Mesh *ND=(m+n-2)* and *E[D]=(m+n)/3*, in Spidergon, *ND=ceiling(N/4)* and $E[D]=(2x^2+4x+1)/N$ (if $N=4x$) and $E[D]=(2x^2+2x-1)/N$ if ($N=4x+2$).

Under the worst case analysis assumptions, the network diameter of real 2D mesh topologies with N nodes shows quite unpredictable fluctuations between the ideal ($\sqrt{N}*\sqrt{N}$) mesh values and the Ring diameter values, as shown in figure 2. The analysis shows that the Spidergon NoC has lower ND than regular 2D meshes at least up to 40-45 nodes (and after, depending on the value of N, see figure 2). In figure 3, we show the analysis results for the average network distance E[D] for ring, ideal and real 2D meshes, and Spidergon. It results that Spidergon outperforms Ring, and works on the middle of the value range of the real mesh implementations. Ideal mesh behavior is obtained by real meshes only under specific N values (that is when $N=m*n$ and $m \approx n$). These results are quite indicative of the difference that may exist between theory results in ideal cases and real scenarios, for mesh topologies. Results in figures 2 and 3 show that Spidergon is expected to have competitive and linear behavior, on the average and worst case scenarios, due to node symmetry and regular topology with respect to real ring and mesh topologies.

In the following we will investigate the NoC support for communication under some optimal routing strategies (that is, resulting in the lowest path length) for the proposed topologies. Specifically, the Spidergon NoC

will adopt the *Across-first routing scheme*: first, if the target node for a packet is at distance $D > N/4$ on the external ring (that is, in the opposite half of the Spidergon external ring) then the across link is traversed first, to reach the opposite node. Second, clockwise or counterclockwise direction is taken and maintained, depending on the target's position. In Ring-based NoC the routing strategy is straightforward: clockwise or counterclockwise direction is taken from the source to the target node, depending on the shortest path direction. In 2D Mesh NoC, *Dimension order* routing is adopted: flits from the source node migrate along the X (horizontal link) nodes up to the column of the target, then along the Y (vertical link) nodes up to the target node.

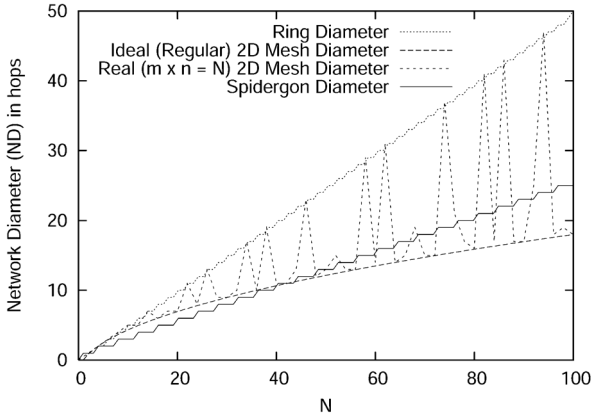

Figure 2: Network Diameter ND vs. number of nodes N in Ring, ideal and real 2D Mesh and Spidergon NoCs.

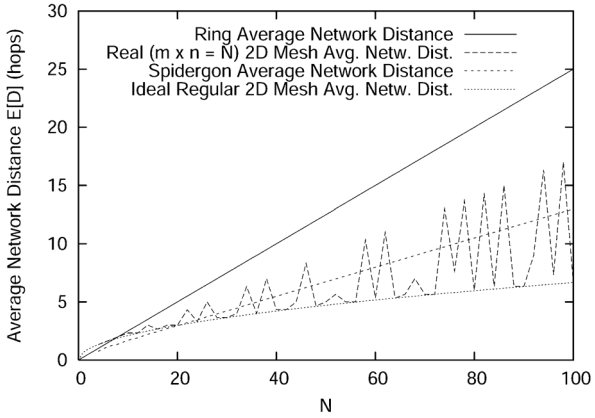

Figure 3: Average Network Distance vs. number of nodes in Ring, ideal and real 2D Mesh and Spidergon NoCs.

3. Performance evaluation

The modeling and simulation of the NoC architectures have been performed with the OMNeT++ simulation framework [21]. OMNeT++ is a public source, generic and flexible simulation environment with strong GUI support that allows a fast and high-level simulation environment for NoC exploration topologies.

The node model for the Spidergon NoC is shown in figure 4. Each node has an external network interface to connect the IP to the NoC. The external IP can act as a packet source and/or as a packet destination (sink) depending on the simulated scenario. Packet sources adopt a Poisson interarrival distribution of constant size packets (6 flits in our simulations), with variable parameter Lambda. The first (head) flit of a packet is sent to the routing mechanism of the node, and then transferred on the output queue of the target channel (if room). Once the head flit has been processed by the routing element of a node, a switching mechanism is defined to forward all immediately following packet-flits to the buffers of outgoing links of the target path to the destination node. Application packets are consumed from the IP memory in a FIFO order.

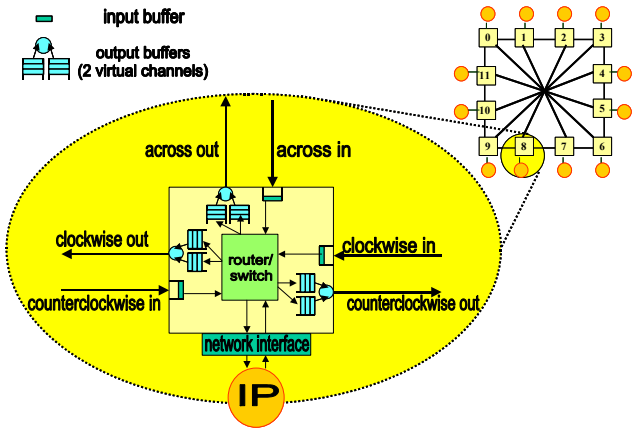

Figure 4: a node model for Spidergon NoC

The scheme in figure 4 refers to Spidergon nodes. On the other hand, Ring and Mesh nodes considered in this analysis have been defined with the same node architecture, excepted the number of links, the cumulative buffers sizes, and the routing policies. Specifically, Ring nodes have clockwise and counterclockwise links only, and mesh nodes may have from 2 up to 4 links, by including N, S, W and E direction links. Incoming links have a one-flit buffer, while outgoing links have a pair of output buffers (used both for virtual channel management and deadlock avoidance) in Ring and Spidergon topologies, and one single buffer in Mesh topologies. All output buffers may contain up to three-flits.

Experiments have been performed by modifying the overall buffer capacity of nodes and buffer symmetry depending on the expected link usage. Results indicated that small buffer tuning have some marginal impact on the peak performances. These results have not been presented in this work due to space limitations. Due to

space limitations, in the following we will illustrate and comment the results obtained in three basic scenarios: the *single and double hot-spot target* scenario, and the *homogeneous sources and destinations* scenario.

3.1 Simulation Results

The first set of data shown is related to the validation of the simulation and analytical model. Figure 5 shows

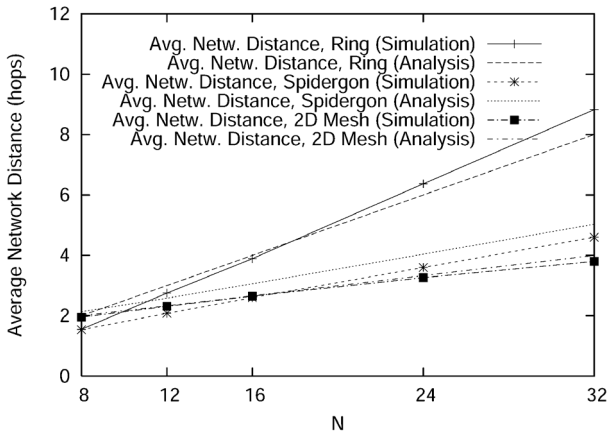

Figure 5: analytical and simulation-based average network distances (hops)

the analytically estimated average distance E[D] and the simulation-based value obtained. Despite some differences in the data, due to stochastic variability, the figure confirms that Ring has the worst average performances, while Spidergon and 2D Mesh topologies

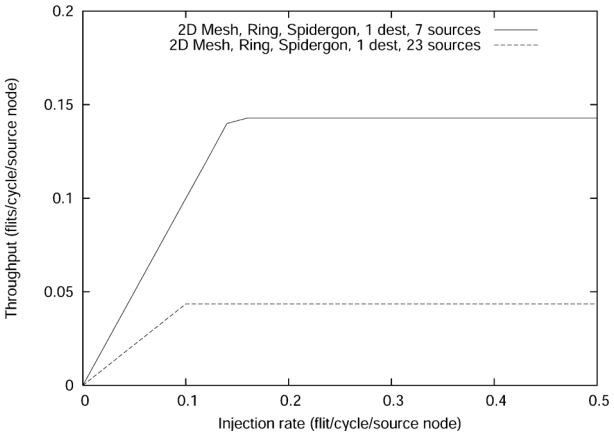

Figure 6: NoC throughput, one hot-spot destination node

work close to each other in the range from 8 to 32 nodes.

3.1.1 Single hot-spot target scenario

Figure 6 shows the throughput index of the NoC architecture as a function of the injection rate parameter of the source nodes when *hot-spot target* is present in the system (that is, one single destination node for all packets). Destination nodes have been taken in different points on the Mesh topology (in symmetric Ring and Spidergon this would not have difference). The result from figure 6 is that the throughput index presents no differences with respect to the implemented topology when one single target destination is adopted for all communications. The only difference is given by varying the number of source nodes. When all the sources homogeneously increase the injection rate, this translates to linear absorption from the (single) destination node, up to the destination node saturation is obtained. This means that the most significant system bottleneck under hot-spot traffic destination scenarios is the destination node, and not the NoC architecture and the channel buffering resources. This result is quite different from the interpretation that can be obtained by assuming a uniform load distribution among many sources and many destinations. This does not mean that the NoC architecture is irrelevant, because the NoC architecture behaves better when parallel local communication is present. On the other hand, in today's common SoCs scenarios, when the system memory is external, the behavior obtained with different NoC topologies would converge to the behavior shown in figure 6.

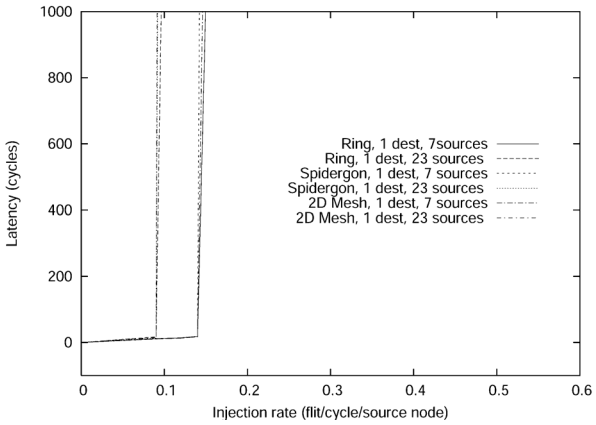

Figure 7: NoC latency, one hot-spot destination node

In other words, the scalable and symmetric architecture of Spidergon would give the same advantages of more complex solutions, like 2D Mesh, under the hot-spot communication viewpoint. In addition, Spidergon can outperform ring or a complex bus hierarchy when multiprocessors are presents (these data have been obtained and were not included in this paper due to space limitations).

Moreover, Spidergon introduces a degree of scalability and flexibility that would not be found in current bus architectures. For this reason, Spidergon appears as the good trade-off solution for obtaining the same

performances of more complex architectures, under common scenarios in current SoCs.

Figure 7 shows the average latency obtained by Spidergon, 2D Mesh and Ring topologies under one single hot-spot destination node, as a function of the number of nodes N and the injection rate parameter of multiple source nodes. Data show that the latency sharply increases when the target node saturation is obtained, with little differences due to the NoC topology adopted. By assuming an homogeneous injection rate, the latency increases early when the number of source nodes increases, as expected.

3.1.2 Double hot-spot scenario

Simulations have been performed by considering a *pair of hot-spot target* scenarios, and by allocating the targets in different positions inside the NoC topologies.

For 2D Mesh, scenario A is with 2 targets on the opposite corners (nodes 1 and N), scenario B is with one target in the corner (node 1) and the second one in the middle (node 5 with 2*4=8 mesh and node 14 with 4*6=24 mesh), and scenario 3 is with both targets in the middle (nodes 5 and 6 with 2*4=8 mesh, and nodes 14 and 15 with 4*6=24 mesh). In Ring and Spidergon, scenario A is with two targets in opposition (North-South position) on the ring, and scenario B is with two targets in North and West positions on the ring. The results (see figures 8 and 9) basically confirm the system behavior and conclusions discussed for one hot-spot target.

3.1.3 Homogeneous sources/destinations scenario

Figure 10 shows the throughput results with respect to the NoC topology and the number of nodes, under homogeneous scenarios with uniform distribution of sources and destinations. Specifically, all the nodes behave like sources and can be addressed as destination for packets, with uniform probability distribution. When all node sources increase the injection rate, this translates to linear absorbtion from all the destination nodes, up to the set of destination nodes and/or the network become saturated. This performance index illustrates that Spidergon and 2D Mesh topologies outperform Ring, and scale better when the number of nodes is low. Under this scenario, 2D Mesh shows a better throughput than Spidergon only with many nodes and when the local injection rate of all source nodes is greater than 0.3 flits/cycle. On the other hand this scenario is hardly obtained in real systems, and this does not constitutes a good motivation to prefere the adoption of 2D Mesh in favour of the Spidergon topology. As expected, the bottleneck emerging in this scenario is basically given by the communication infrastructure. This is confirmed also by the worst performances obtained by the Ring topology.

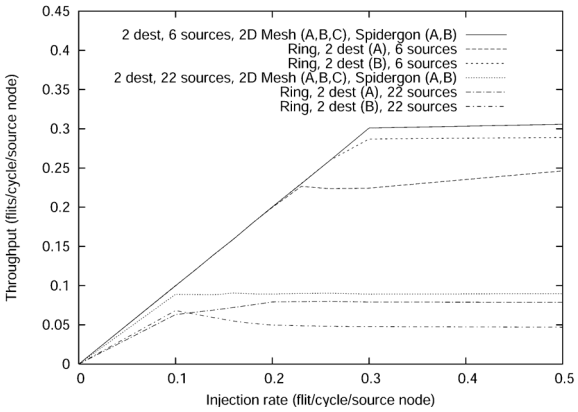

Figure 8: NoC throughput, two hot-spot destination nodes

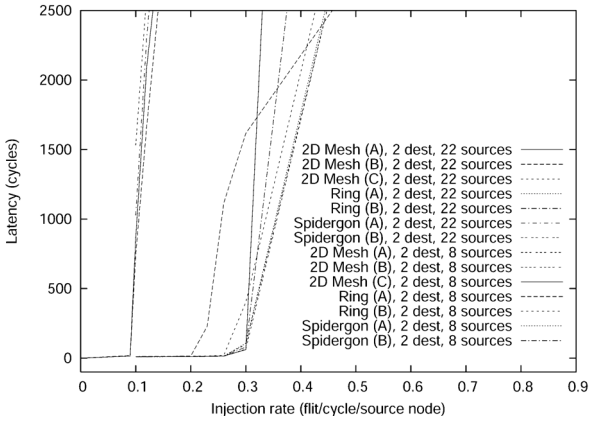

Figure 9: NoC latency, two hot-spot destination nodes

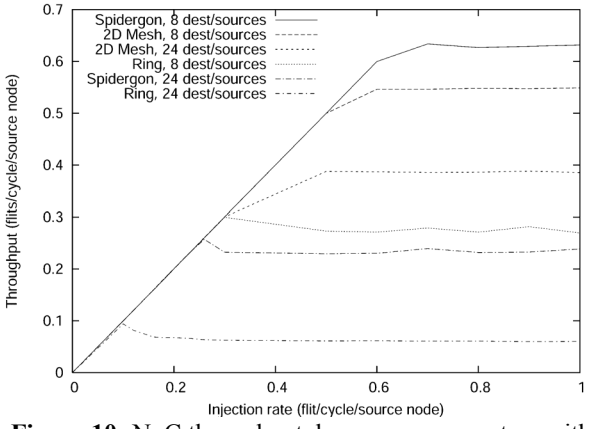

Figure 10: NoC throughput, homogeneous system with all nodes working as packet sources and destinations

Figure 11 illustrates the average latency obtained by Spidergon, 2D Mesh and Ring topologies under homogeneous source and destination distribution scenarios. All the nodes behave like sources and can be

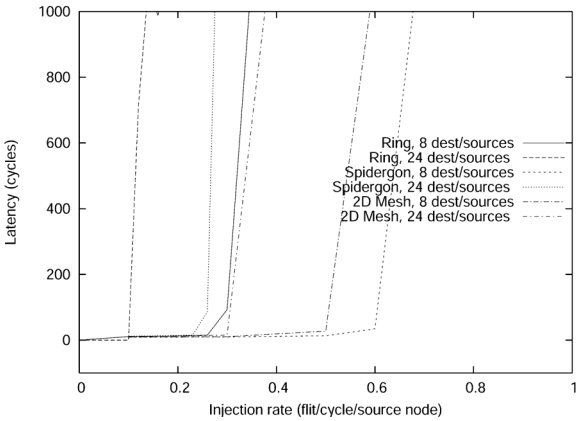

Figure 11: NoC latency, homogeneous system with all nodes working as packet sources and destinations

addressed as destination for packets with uniform probability distribution. Latency is shown as a function of the number of nodes N, and the injection rate parameter of multiple source nodes.

Data show that the latency sharply increases when the network saturation is obtained, with some differences due to the different saturation properties of the NoC topology adopted. By assuming an homogeneous injection rate, Ring topology saturates first, and the latency generally increases early when the number of system nodes increases, accordingly with the throughput results obtained for the same scenarios.

4. Conclusions

In this work we illustrated the modeling and simulation-based analysis of low degree topologies (Ring, 2D Mesh and Spidergon) focusing the on-chip domain requirements and overcoming the classical parallel computing and networking results. To the best of our knowledge, this is the first work considering irregular mesh topologies, whose analysis is motivated since regular meshes cannot be always assumed as realistic topologies in SoCs. On the other hand, the first deep analysis of the novel Spidergon topology is presented, resulting in a good compromise among the well-known topologies. This has been demonstrated by analyzing Spidergon characteristics with respect to real 2D Mesh and Ring topologies, and by considering common hot-spot communication scenarios that characterize current SoC architectures. This analysis has shown that Spidergon can be considered a good solution under the system design, the ease of implementation and management viewpoint, with performance and scalability results that are in line with other more complex solutions under most common assumptions and scenarios.

Future work will include the extension of the analysis and simulation with more NoC nodes, specific traffic patterns originated by common applications, and analysis of routing protocols and additional NoC topologies.

Acknowledgments

The authors wish to thank Marcello Coppola and Riccardo Locatelli (STMicroelectronics) for their contributions and constructive discussions about the topics and analysis presented in this paper.

References

[1] L. Benini and G. De Micheli, “Networks on Chips: A New SoC Paradigm,” Computer, vol. 35, no. 1, pp. 70-78, Jan. 2002.
[2] DSPIN, www.lip6.fr/Direction/2005-05-13-DSPIN.pdf
[3] W.J. Dally and B. Towles, “Route Packets, Not Wires: On-Chip Interconnection Networks,” Proc. Design Automation Conf. (DAC), pp. 683-689, 2001.
[4] J.Duato, S.Yalamanchili, and L.Ni, Interconnection Networks, An Engineering Approach, M. Kaufmann, 2002.
[5] A. Jantsch and H. Tenhunen, Networks on Chip, eds. Kluwer Academic, 2003.
[6] P.P. Pande, C. Grecu, M. Jones, A. Ivanov and R. Saleh, “Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures”, IEEE Trans. on Computers, Vol. 54, No. 8, pp. 1025-1040, Aug. 2005.
[7] U.Y. Ogras, J. Hu, R. Marculescu, “Key Research Problems in NoC Design: A Holistic Perspective”, in Proc. of the Int’l Conf. on HW-SW Codesign and System Synthesis, Sep. 2005.
[8] M. Coppola, S. Curaba, M. D. Grammatikakis, G. Maruccia, F. Papariello, “OCCN: A Network-On-Chip Modeling and Simulation Framework”, in proc. DATE 2004.
[9] M. Coppola, R. Locatelli, G. Maruccia, L. Pieralisi, M.D. Grammatikakis, “Spidergon: A NoC Modeling Paradigm”, book chapter in Model Driven Engineering for Distributed Real-time Embedded Systems, ISBN: 1905209320, Aug. 2005
[10] M. Coppola et al. “Spidergon: a novel on chip communication network”, proc. Int’l Symposium on System on Chip 2004, Tampere, Finland, Nov. 2004
[11] E. Rijpkema et al., “A router architecture for networks on silicon”, Progress 2001, 2nd Workshop on Embedded Systems.
[12] M. B. Taylor et al., “The Raw microprocessor: a computational fabric for software circuits and general-purpose programs”, IEEE Micro, pp. 25-35, March 2002.
[13] Forsell, M. “A scalable high-performance computing solution for networks on chips”, IEEE Micro, 22 (5), 2002.
[14] M. Dall’Osso et al., “Xpipes: a latency insensitive parameterized network-on-chip architecture for multi-processor SoCs”, proc. ICCD’03
[15] M. Millberg et al., “The Nostrum backbone - a communication protocol stack for networks on chip”, VLSI Design Conference, 2004
[16] AMBA Bus Specification, http://www.arm.com, 1999
[17]Wishbone Service Center, www.silicore.net/wishbone.htm
[18] CoreConnect, www.ibm.com/chips/products/coreconnect.
[19] Sonics Backplane Micro-Network, www.sonicsinc.com
[20]A.Scandurra, G. Falconeri, B. Jego, “STBus communication system: concepts and definitions” int.l report, STM, 2002
[21] A. Varga, "OMNeT++", in the column "Software Tools for Networking", IEEE Network Interactive. July 2002, Vol.16 No.4, www.omnetpp.org