Analysis of Routing Algorithms Generation for Irregular NoC Topologies

Ronaldo Milfont, Paulo Cortez, Alan Pinheiro,
Joao Ferreira, Jarbas Silveira, Rafael Mota
LESC - DETI - UFC
Federal University of Ceara
Fortaleza, Brazil
milfont@lesc.ufc.br

Cesar Marcon
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil
cesar.marcon@pucrs.br

***Abstract*—Network-on-Chip (NoC) is the most promising communication architecture for a scalable and high parallel System-on-Chip (SoC). Recent manufacturing technologies are very close to physical limitations increasing the number of faults during fabrication and operation; as a result, NoCs designed to be regular often will become irregular. Thus, the analysis of irregular topologies derived from regular ones is an important task in the NoC design. This work uses a tool that evaluates restriction paths of irregular NoCs to ensure deadlock freedom and generates analytical metrics for assessing the quality of routing at design time. We used a fault model for generating NoC topologies with faulty links, which allows exploiting manufacturing processes scenarios with 65 nm and 22 nm CMOS technology, from the links delay variability and spatial correlation strength. We also used a segmentation algorithm to avoid deadlocks. The experimental results show that the percentage of faults significantly increases with the spatial correlation strength, while the average distance between topological faults decreases. Furthermore, the greater the fault percentage gets, the longer the routing average distance becomes.**

Keywords: Irregular NoC, routing algorithm, fault tolerance

I. Introduction

System-on-Chip (SoC) integrates into a single chip all or near all the parts found on the printed circuit board of the past. They are typically composed of several processor cores together with application-specific circuitry. In a SoC, the design of an efficient interconnection structure is an important task since that can affect not only the overall computing performance but also the operation frequency, area, power dissipation, design time, test time and fault tolerance issues.

Due to a very high nonrecurring engineering cost of manufacturing a SoC, researchers and industry have been looking for reusable, scalable and structured solutions to the interconnection systems, leading to the proposal of packet switched networks for on-chip communication as Network-on- Chip (NoC) [1]. The NoC is composed of several routers and links that follow some connection pattern or topology where each node of the SoC is connected to the NoC through a network adapter. The packets of a node are forwarded through intermediate routers until reaching their destination.

Concerning topology, most NoCs implement regular topologies that can be efficiently laid out on a chip surface. Although, in many high-end complex applications, some specific SoCs are heterogeneous in nature because of module shape and the need to physically separate module internals of the NoC infrastructure , with each core having different sizes,functionalities and communication requirements. This led to a no longer regular interconnection topology.

Besides that, just like all components of a modern system, the interconnections are also prone to process variability and susceptibility to wear-out and aging. The aggressive scaling into the nanoscale regime has made the reliability level decrease, with some studies showing that up to 10% of future on-chip components could be defective. Therefore, the necessity for designing robust NoCs is more pronounced [1].

In general, faults are classified as either transient or permanent. Transient faults are introduced by power supply noise, ground bounce, energetic particles, and interconnection noise such as crosstalk and electromagnetic interference [2]. These faults, which are temporary and unpredictable, are often difficult to be detected and corrected. On the contrary, permanent faults are introduced by physical damages, such as manufacturing defects and device wear out [3]. This paper takes into account permanent faults. Therefore, and without lack of generality, we have chosen the variability model proposed by Hargreaves et al. [4] to generate the fault scenarios. This model considers the effect of the variability in router-to-router links, router to router, using two parameters: the variability of delay (σ) and variability in spatial correlation (λ).

Several routing algorithms have been studied in areas of high performance and fault-tolerant SoCs. In this work, we chose the Segment-based Routing (SBR) [3] which is responsible for deadlock prevention.

- The analysis of the routing algorithms generation employing fault scenarios based on fault distributions that reflect real manufacture and operation faults;
- The NoIA graphical tool that uses SBR to engender routing tables for simulation in NoC fault scenarios, which were acquired using the variability model proposed in [4];
- The analytical metrics obtained by the tool for routing quality evaluation for irregular NoCs [5].

978-1-5386-0415-1/17/$31.00 ©2017 IEEE

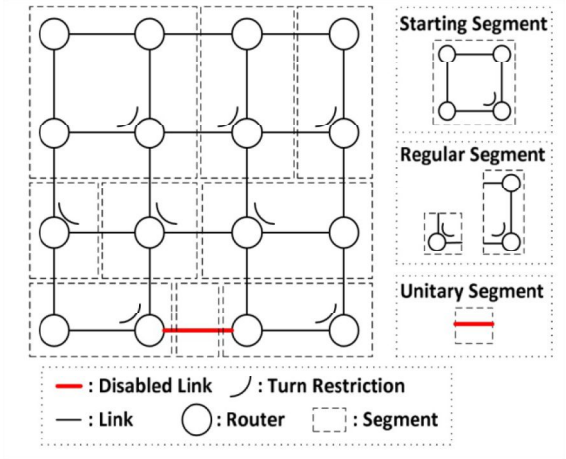

Fig. 1. Segments and turn restrictions computed by the SBR algorithm in a 44 2D Mesh NoC (based on [3]).

II. Related Work

A defective router or link could ruin the 2D mesh communication structure, which would lead to an irregular topology (i.e., topologies derived from regular networks with induced faults or faulty links, also known as agnostic topologies).Therefore, deterministic and static routing algorithms tailored to a regular NoC must provide some degree of fault-tolerance while ensuring deadlock-freedom.

One approach to guarantee deadlock-freedom for irregular topologies is based on virtual channels (e.g., [6]- [11]), which requires a large area for multiplexing schemes and implies significant energy consumption by the buffers present in routers. Another approach is based on turn prohibition (e.g., [3] [12] [15]), which avoids deadlocks by eliminating a subset of network path turns.

Our work employs this latter approach, using the SBR mechanism [3]. SBR is composed of two phases: (i) segment computation and (ii) routing restrictions placement. At segment computation phase, SBR partitions the NoC into segments comprising routers and links. As shown in Figure 1, each segment is characterized by a turn restriction that avoids routing deadlocks. SBR classifies segments into three types: (i) starting, which starts and ends at the same router forming a loop; (ii) regular, which starts at a link, contains at least one router and ends on another link; and (iii) unitary, which includes a single link that does not allow traffic. SBR tries to create small segments to reduce the occurrence of unitary segments. At the routing restrictions placement phase, SBR breaks all possible cycles by placing one bidirectional restriction at each segment. Globally, SBR guarantees deadlock freedom and connectivity among all components in the network; i.e., the turn restrictions still allow communication among all source-destination pairs.

III. Methodology

This section provides a description of all test scenarios and evaluation metrics. It also shows how topologies based on the fault model are generated and analyzed. The reference topology is a mesh type for all the experiments. We used the variability model of deep-submicron CMOS manufacturing process described by Hargreaves et al. [4] to generate synthetic failure scenarios. This variability model is composed by two parameters: variability in delay (σ) and variability in spatial correlation (λ), which is the inverse of the strength of the spatial correlation. According to ITRS [16], the 65 nm and 22 nm CMOS technologies are modeled using σ = 0.05 and σ = 0.18, respectively. Moreover, to induce typical correlation of manufacture processes [4], the experiments were produced with λ = 0.4 as the high variability and λ = 1.2 as the low variability for both CMOS technologies. Table 1 illustrates the experiments encompassing four sizes of mesh NoCs: 55, 66, 77, 88, which combined with the delay and spatial correlation produce 16 fault scenarios. Aiming to explore the randomness of the variability model, we generated each of these scenarios three times, resulting in 48 irregular NoC mesh topologies. Finally, an in-house tool expanded these 48 topologies into 12,224 scenarios by combining all of the possibilities of faulty links.

TABLE I
EXPANSION OF FAULT SCENARIOS FOR 48 IRREGULAR MESH NOCS

NoC Size	Distribution σ	Distribution λ	Amount of faulty channels (3 samples)	Quantity of expanded scenarios (3 samples)
5x5	0.05	1.2	4,4,4	15,15,15
	0.18	1.2	4,4,4	15,15,15
	0.05	0.4	6,6,6	63,63,63
	0.18	0.4	6,6,6	63,63,63
6x6	0.05	1.2	5,5,5	31,31,31
	0.18	1.2	5,5,5	31,31,31
	0.05	0.4	7,7,7	127,127,127
	0.18	0.4	7,7,7	127,127,127
7x7	0.05	1.2	5,5,5	31,31,31
	0.18	1.2	5,5,5	31,31,31
	0.05	0.4	9,9,9	511,511,511
	0.18	0.4	8,8,8	255,255,255
8x8	0.05	1.2	4,4,4	15,15,15
	0.18	1.2	5,5,5	31,31,31
	0.05	0.4	11,11,10	2047,2047,1023
	0.18	0.4	10,10,10	1023,1023,1023
	Total		48 NoC topologies	12224 scenarios

Figure 2 shows the experimental setup including the generation of routing tables and analytical metrics.

In *Step 1* we obtain the expanded combination described before. The following steps are implemented in Java programming language: *Step 2* initiates the SBR processes that divides the topology into subnets and, then, into segments. *Step 3* performs the restriction placement, blocking some routes to avoid deadlocks [4]. Next, *Step 4* is responsible for routing tables generation. Finally, during *Step 5*, NoIA provides the analytical metrics and the routing tables for each topology produced in *Step 1*. The routing tables are written in VHDL and join in a VHDL NoC project, whom is simulated on Modelsim. This work employs the following analytical metrics (in bold):

1) Effectivity of Segmentation and Routing Methods encompasses the **Average Quantity of Unitary Segments (AQUS)**, which is calculated dividing the number of uni-

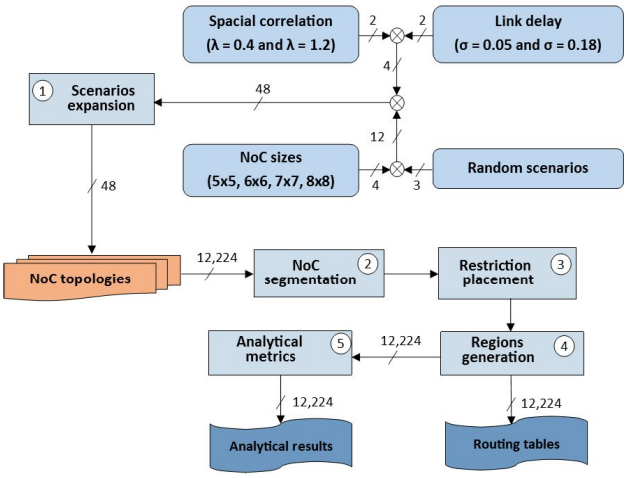

Fig. 2. Experimental setup for analytical metrics evaluation

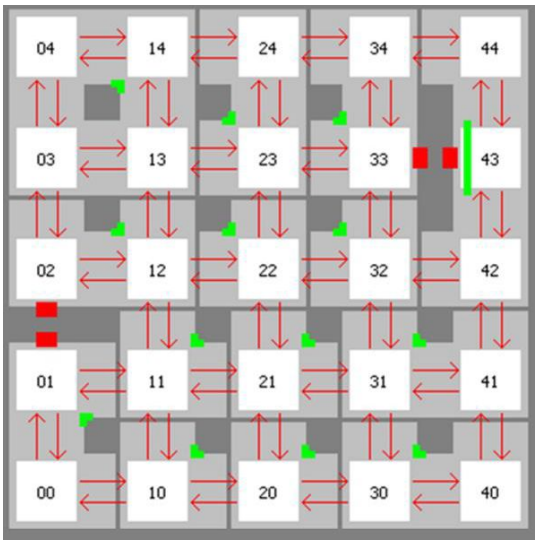

Fig. 3. Graphical interface of NoIA showing segments, restrictions and faulty links of synthetic 5x5 mesh NoC.

tary segments by the number of links. It represents the efficacy of the segmentation step since unitary segments frequently increase the routing paths delaying the NoC communications.

2) **Average Communication Latency (ACL)** is computed through a composition of the (a) Average Routing Distance (ARD), (b) Network Link Weight (NLW), and (c) Standard Deviation Link Weight (STDLW). ACL is measured as an average quantity of hops. A hop is a communication path between two neighbor routers. Equation 1 describes ARD, which is the sum of all hops of all communications divided by the number of paths, representing the average of all path lengths.
Where P is the total of NoC paths, and is the number of hops in the path p. Equation 2 shows LW, which is the average number of paths per link; and, STDLW is the standard deviation of LW [6].
3) Fault Distribution is characterized by the (a) **Average Quantity of Faults (AQF)**, which is the sum of all faulty links divided by the quantity NoC links and (b) **Average topological Distance between Faults (ADF)**. We consider as a topological distance between two faulty links the quantity of hops existing in a minimum path between these links. As AQUS, AQF and ADF are also used to evaluate the efficiency of the segmentation method according to the fault distribution, which depends on the process variability and the NoC size.

Figure 3 exemplifies an irregular 55 mesh NoC illustrated in the graphical interface of NoIA.

IV. Experimental Results and Discussion

The first set of preliminary results illustrates the AQF per delay variability and spatial correlation variability of all scenarios detailed in Table 1 (results are presented in percentage). Figure 4 displays that the variability of delay does not affect the fault rate; however, the fault percentage greatly increases when the strength of the spatial correlation increase (i.e., λ decreases). This happens because when the value of spatial correlation is stronger, faults become more frequent, engendering more severe failure scenarios.

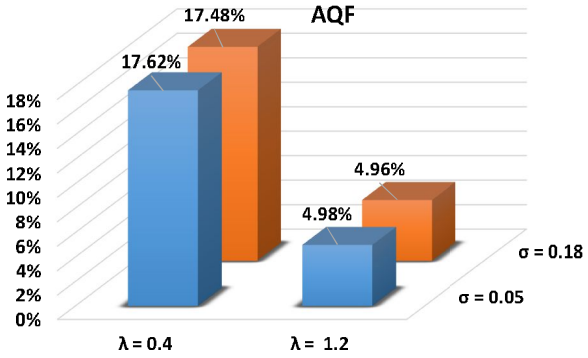

Fig. 4. Average Quantity of Faults (AQF) of all fault scenarios, regarding variability and spatial correlation distributions, in percentage.

The first experiment is confirmed by the second set of experimental results shown in Figure 5, where it is possible to see that ADF is lower when the strength of the spatial correlation is bigger; in other words, it shows that faults are closer to each one. Hence, we conclude that recent technologies produce three times more faults than the old ones.

ACL enables to evaluate the effect of faulty links on the communication delay. Figure 6 illustrates that increasing the correlation strength increases ACL, which is justified by the increase in the percentage of faulty links, which tends in average to make paths longer.

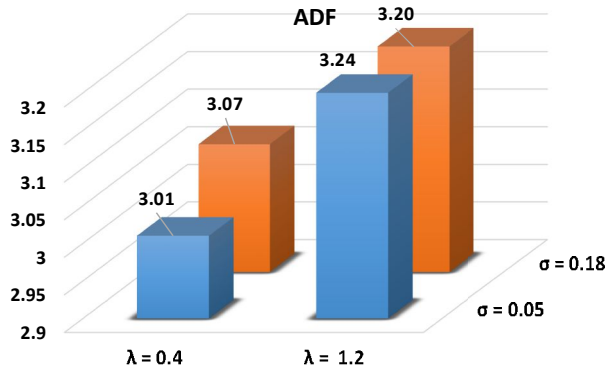

Fig. 5. Average topological Distance between Faults (ADF) according to variability and spatial correlation distributions, in hops.

The severity of failure distribution also complicates the search for deadlock-free paths, requiring more time for routing calculation and implying larger routing tables. This circumstance is confirmed in the fourth set of experimental results shown in the Figure 7.

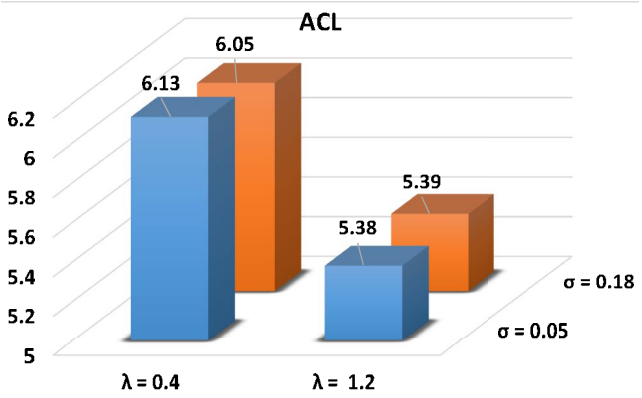

Fig. 6. Average Communication Latency (ACL) in hops the packets passes through, regarding variability and spatial correlation distributions.

A non-zero number of unitary segments, arising from the SBR process means that the segmentation was not optimal; since no communication is allowed to pass through an unitary link, which increases the average routing path. Thus, the distribution with bigger correlation strength presents a higher number of unitary segments.

The following graphics take into account the grouping of values obtained in the experimental results, according to the variation of the NoC sizes. Figure 8 illustrates that AQF decreases as the NoC size increase, because the quantity of NoC links rises quadratically, whereas the number of faulty links grows linearly. However, this reduction is not observed in the faults grouping. As one can see in Figure 8(b), ADF grows as the NoC size increases.

The same fact depicted in Figure 9 can be found in Figure 10, where we can see that ACL increases as the size of NoC

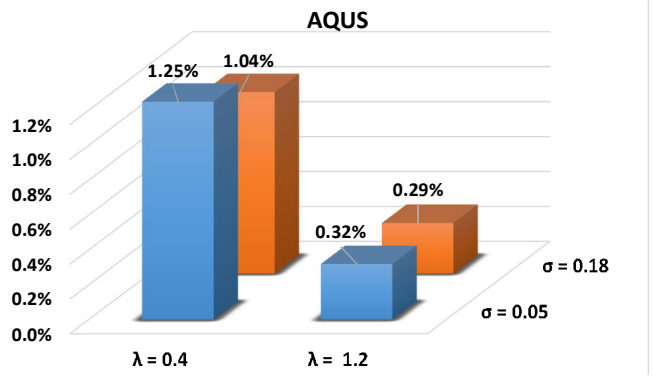

Fig. 7. Average Quantity of Unitary Segments (AQUS) of all fault scenarios, regarding variability and spatial correlation distributions, in percentage.

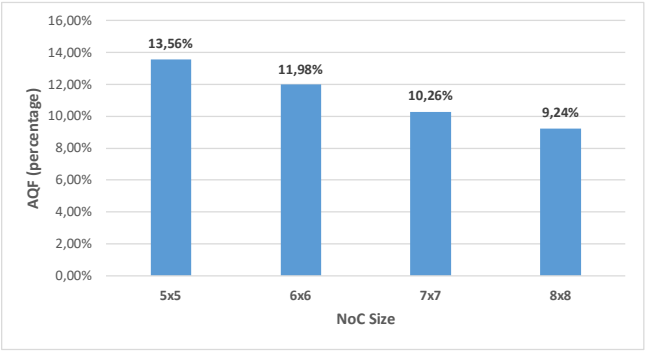

Fig. 8. AQF of all fault scenarios according to the NoC Size, in percentage.

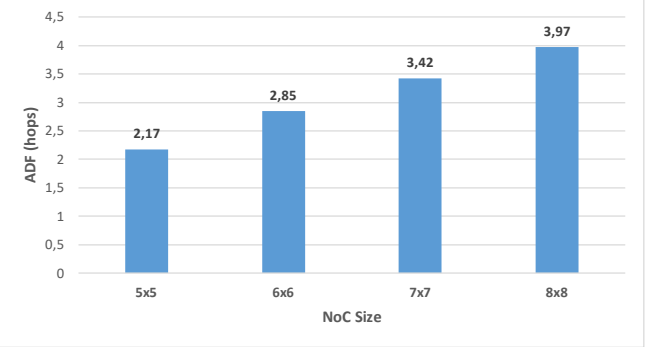

Fig. 9. ADF of all fault scenarios according to the NoC Size, in hops.

increases, this happens due to the increase of the distance between source-destination pairs.

V. Conclusion

In this paper, we have proposed the use of SBR in order to explore its flexibility in irregular topologies, allied to the variability model proposed by Hargreaves et al. [4] to generate fault scenarios. All scenarios were evaluated according to the (i) effectivity of segmentation and routing methods, represented by the Average Quantity of Unitary Segments (AQUS), (ii) Average Communication Latency (ACL) which

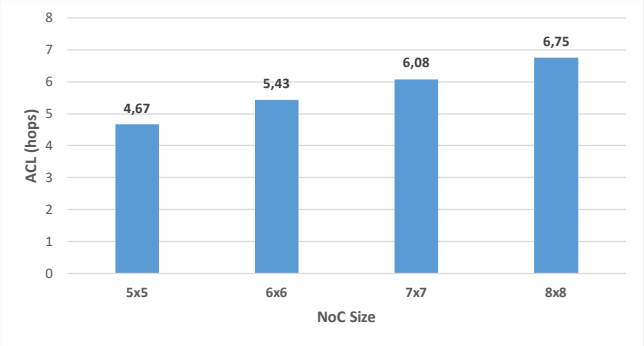

Fig. 10. ACL of all fault scenarios regarding NoC size, in hops.

depends on the Average Routing Distance (ARD), the Network Link Weight (NLW) and the Standard Deviation of LW, and (iii) fault distribution, which encompasses the metrics Average Quantity of Faults (AQF) and Average topological Distance between Faults (ADF). From our set of analysis, we could correlate the metrics finding several remarkable conclusions: the percentage of faults significantly increases with the strength of the spatial correlation, and so do the Average Routing Distance (ARD). Additionally, we can conclude that the greater the percentage of faults is the higher ACL becomes. Thus, the severity of the faults distribution can make the segmentation process more complex, since a higher spatial correlation has a larger percentage of faults, thereby increasing the number of unitary segments.

REFERENCES

[1] L. Benini and G. De Micheli, "Networks on chips: a new SoC paradigm," in Computer, vol. 35, no. 1, pp. 70-78, Jan 2002.

[2] T. Bengtsson, S. Kumar, R. Ubar, A. Jutman. Off-line Testing of Crosstalk Induced Glitch Faults in NoC Interconnects. Norchip Conference, pp. 221-225, 2006.

[3] A. Mejia, J. Flich, J. Duato, S.-A. Reinemo, T. Skeie, Segment-Based Routing: An Efficient Fault-Tolerant Routing Algorithm for Meshes and Tori. International Parallel and Distributed Processing Symposium (IPDPS), pp. 1-10, 2006.

[4] B. Hargreaves, H. Hult, S. Reda. Within-Die Process Variations: How Accurately can they be Statistically Modeled? Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 524-530, 2008.

[5] J. Flich, T. Skeie, A. Mejia, O. Lysne, P. Lopez, A. Robles, J. Duato, M. Koibuchi, T. Rokicki, J. Sancho, A Survey and Evaluation of Topology-Agnostic Deterministic Routing Algorithms. IEEE Transactions on Parallel and Distributed Systems, v. 23, n. 3, pp. 405-425, Mar. 2012.

[6] W. Dally, H. Aoki. Deadlock-Free Adaptive Routing in Multicomputer Networks Using Virtual Channel. IEEE Transactions on Parallel and Distributed Systems (PDS), v. 4, n. 4, pp. 466-475, Apr. 1993.

[7] F. Chaix, D. Avresky, N.-E. Zergainoh, M. Nicolaidis. Fault-Tolerant Deadlock-Free Adaptive Routing for Any Set of Link and Node Failures in Multi-cores Systems. IEEE International Symposium on Network Computing and Applications (NCA), pp. 52-59, 2010.

[8] M. Valinataj, S. Mohammadi, J. Plosila, P. Liljeberg. A Fault-Tolerant and Congestion-Aware Routing Algorithm for Networks-on-Chip. IEEE International Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS), pp. 139-144, 2010.

[9] C. Jackson, S. Hollis. A Deadlock-Free Routing Algorithm for Dynamically Reconfigurable Networks-on-Chip. Microprocessor and Microsystems, v. 35, n. 2, pp. 139151, Mar. 2011.

[10] S. Pasricha, Y. Zou. NS-FTR: A Fault Tolerant Routing Scheme for Networks on Chip with Permanent and Runtime Intermittent Faults. Asia and South Pacific Design Automation Conference (ASPDAC), pp. 443-448, 2011.

[11] M. Ebrahimi, M. Daneshtalab, J. Plosila, H. Tenhunen. MAFA: Adaptive Fault-Tolerant Routing Algorithm for Networks-on-Chip. Euromicro Conference on Digital System Design (DSD), pp. 201-207, 2012.

[12] C. Glass, L. Ni. Fault-Tolerant Wormhole Routing in Meshes without Virtual Channels. IEEE Transactions on Parallel and Distributed Systems (PDS), v. 7, n. 6, pp. 620-636, Jun. 1996.

[13] K.-H. Chen, G.-M. Chiu. Fault-Tolerant Routing Algorithm for Meshes without Using Virtual Channels. Journal of Information Science and Engineering (JISE), v. 14, n. 4, pp. 765-783, Jan. 1998.

[14] K. Aisopos, A. DeOrio, L.-S. Peh, V. Bertacco. ARIADNE: Agnostic Reconfiguration in a Disconnected Network Environment. International Conference on Parallel Architectures and Compilation Techniques (PACT), pp. 298309, 2011.

[15] A. DeOrio, D. Fick, V. Bertacco, D. Sylvester, D. Blaauw, J. Hu, G. Chen. A Reliable Routing Architecture and Algorithm for NoCs. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), v. 31, n. 5, pp. 726-739, May 2012.

[16] International Technology Roadmap for Semiconductors. ITRS 2007. //www.itrs.net/reports.html, 2016.