

An Improved Algorithm with Load Balancing Method for Fault-Tolerant in Mesh Based NoCs

Hamed Sadatmehrizi

Dept. of Electrical, IT & Computer Sciences
Qazvin Branch, Islamic Azad University
Qazvin, Iran
h.mehrizi@qiau.ac.ir

Esmail Zeinali

Dept. of Electrical, IT & Computer Sciences
Qazvin Branch, Islamic Azad University
Qazvin, Iran
zeinali@qiau.ac.ir

Abstract— this paper presents an adaptive routing algorithm for 2D mesh network-on-chips (NoCs). The algorithm is based on DyXY routing algorithm. Our proposed routing algorithm is a very low cost fault-tolerant routing method to tolerate at least one faulty link in mesh-based on chip networks. It is a distributed, adaptive, congestion-aware and deadlock-free routing algorithm where only two virtual channels are used for adaptiveness and fault-tolerance. This is obtained by using two congestion flags between each two nodes which demonstrate the existence of congestion in a row or column. The same flags may be used to alarm a link failure in a row or column. The network performance, fault-tolerance capability and hardware overhead are computed through appropriate simulations. The experimental results show that the overall reliability and throughput of a NoCs is significantly elevated against at least one link failure with only a small latency overhead.

Index Terms— Congestion, fault tolerance, network-on-chip, reconfiguration, routing algorithm.

I. INTRODUCTION

Decreasing feature sizes in deep sub-micron technologies and increasing the complexity of evolving systems have led complex systems such as system-on-chip (SoC) and chip multiprocessor (CMP) to be more sensitive and susceptible to environmental effects. Thus, Network-on-Chip as a viable communication infrastructure in these complex systems may incur some device failures in both manufacturing phase and operational mode. Therefore, fault-tolerance methods are required in different levels to maintain the system operational and healthy against component defects or failures. For this purpose, at the network level a routing algorithm that tolerates both router and link failures can be very beneficial for NoCs.

In this paper, we construct our method based on faulty link tolerance. The proposed method needs only two virtual channels (VC) to be deadlock free and to achieve both fault-tolerance and adaptivity when selecting the paths. Considering the fact that a fully

adaptive routing algorithm in a mesh network requires two virtual channels, our method does not require any extra virtual channel to achieve fault-tolerance; but instead, it utilizes the existing two available virtual channels. This is an important aspect since requiring more VCs leads to more cost especially for NoCs. There are many VC-based fault-tolerant routing algorithms primarily designed for parallel computers domain that require three or four VCs [1–5]. However, implementing these fault-tolerant methods in the NoC domain is not suitable due to power restrictions in addition to on-chip area overheads.

There are many fault-tolerant wormhole routing algorithms to tolerate faulty components in two dimensional (2D) mesh networks. To avoid deadlock conditions some of these algorithms use turn model [6–8] and the others use virtual channels [1,2,9,10]. The turn-based routing methods have lower hardware costs but also lower performance because of having less adaptivity. In [7] a powerful routing algorithm is proposed that can tolerate many numbers of faulty links. Another method is presented in [8] that can be accordingly reconfigured to tolerate all one-faulty-link situations in the networks. But, these algorithms cannot route packets in such a way the traffic is balanced. In [11] a routing algorithm is presented for NoCs that dynamically detects the faulty links while routing the packets and also selects the routing paths based on the received congestion information from the adjacent nodes, but it is basically a one-faulty-link tolerant algorithm.

The LFX and LFC methods introduced in this paper are based on a basic one-faulty-link tolerant routing algorithm, which can be considered as a fault-tolerant version of DyXY method proposed in [8]. LFC does not need to exchange any information between the healthy routers to discover where the working routers are located relative to a faulty link. Those algorithms can route the packets around a faulty link. But, in the LFC method, we use the contour concept and configuration registers to store the fault information obtained in fault detection process. The stored information implicitly includes the exchange information needed for routing algorithms. In other words, LFC uses fault information in different

way, better, simpler and more cost-effective in the NoC domain. In addition, we do not apply any restriction on the location of faults.

The rest of the paper is organized as follows. In Section 2 DyXY routing methods and their fundamentals are described. Then, the LFXXY routing method is explained in Section 3. In section 4 LFeXY routing algorithm is introduced. The experimental results are presented in Section 4. Finally, conclusions are drawn in Section 5.

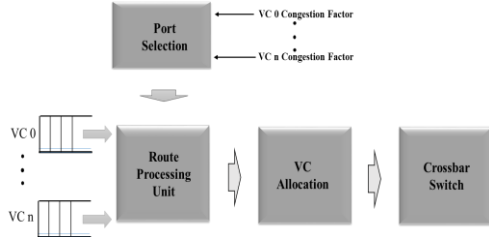


Figure 1. Architecture of a DyXY router.

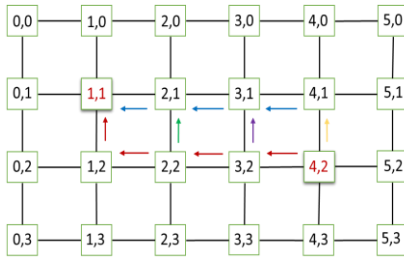


Figure 2. A NoC with mesh structure

II. DYXY ALGORITHM AND ITS RESTRICTIONS

In the DyXY routers, relevant to the network situation, the route processing unit may pick different paths at different times for the same source and destination pair. For this to happen, a port selection unit is appended to the router that based on the current network situations, selects the best candidate for every adjacent ports (i.e., North vs. East, North vs. West, South vs. East, and South vs. West) and provide route processing unit with this information (Fig. 1). One of the factors for choosing an output port is the number of free buffers at the corresponding input port in the next hop [14]. This technique has been used in the DyXY routing algorithm where the free buffer count at a downstream node is used for congestion approximation. To exchange the count, which can be considered as a congestion factor, some wires are added between adjoining routers.

Due to the fact that in each node, packets can be routed in both X and Y directions without restriction, this routing algorithm needs a mechanism to guarantee deadlock avoidance. In networks having virtual channels (general case), usually the following method is used to guarantee deadlock avoidance. Virtual channels in Y dimensions are divided into two parts. The network is partitioned into two sub-

networks called X^+ sub-network and X^- sub-network each having half of the channels in the Y dimension. If the destination node is to the right of the source, the packet will be routed through the X^+ sub-network. If the destination node is to the left of the source, the packet will be routed through the X^- sub-network. Otherwise that packet can be routed using either sub-network [15]. For networks without virtual channels, other deadlock avoidance strategies should be used (e.g., Odd-Even [15]).

Now, we discuss the weakness of the DyXY algorithm in routing packets from switches whose X position (Y position) is one unit apart from that of the destination. Let us consider the simple example shown in Fig. 2 where (4, 2) and (1, 1) are the source and destination switches. In the DyXY routing, switch (4, 2) compares the current length of the east queue of switch (3, 2) and the south queue of the switch (4, 1) and the packet is sent to the direction with a less occupied queue. Note that if switch (4, 1) is selected, the entire path has been determined and hence switches (3, 1) and (2, 1) are the next hops without any choice. If they are congested, the DyXY algorithm has selected wrong path based on local information. However, if switch (3, 2) is selected, there are other choices for the next hops, and hence, the congestion may be avoided. The wrong path was selected since the decision was based on the local information of switch (4, 1). Note that this undesired effect may occur when routing from a switch whose X (or Y) position are just one unit apart from that of the destination and the congestion is happening further away from the current hop.

III. LFXXY ALGORITHM TO TOLERATE A FAULTY LINK

Preliminaries

The used high-level fault model is based on the assumption that when any type of permanent faults occurs in any direction of a bidirectional link, the entire link will be considered as faulty. Moreover, an appropriate detection mechanism like built in-self-test (BIST) mechanism is used to detect the faulty links and to update the configuration registers in a reconfiguration process. In addition, it is assumed that the routers know how to resume their operation after finishing the reconfiguration process.

Def. 1: In a 2D mesh each link has two direct neighboring nodes west and east (W and E) or north and south (N and S) and at most four indirect neighboring nodes (NW, NE, SW, and SE) or (WN, EN, WS, and ES) based on its horizontal or vertical orientation.

Def. 2: Each link has a contour encompassing all direct and indirect neighboring nodes and their interconnecting links.

Def. 3: Each node has at most four direct neighboring links in its north, south, east or west directions.

In Fig. 3 “ \times ” represents a faulty link. In this Figure there are 6 different contours for links (broken or healthy) corresponding to 6 types of positions. Each mesh border

has one contour (S1, S3, S4, S6) and other positions have the contour S2 or S5. Following Dally's condition the Channel Dependency Graphs (CDG) can be used to prove that the contours S1, S3, S4 and S6 are deadlock free. But, the contours S2 and S5 are not deadlock free because each has two cycles in its CDG. If we use an additional VC, the contours S2 and S5 will become deadlock free. The additional VC can be one of two VCs used in an adaptive routing algorithm and is dynamically selected. This is due to the fact that a six-node rectangular with the faulty middle link (the sub-network inside S2 or S5) can be treated the same as a four-node rectangular in a non-faulty sub-network by an adaptive routing algorithm. These two virtual channels can be used only to increase the performance for the contours S1, S3, S4 and S6 since there is only one path between the nodes inside these contours.

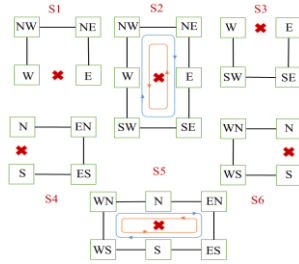


Figure 3. All situation for horizontal and vertical faulty links

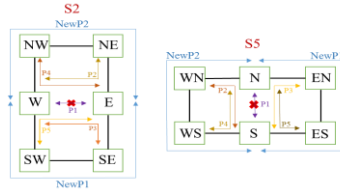


Figure 4. The old paths (P1-P5) broken by a faulty link and the new paths (NewP1, NewP2) for a horizontal and a vertical faulty link.

Routing With Reconfiguration

As stated before, our basic routing method to tolerate faulty links is based on a fully adaptive and minimal routing algorithm in which if there are two possible directions to move towards a destination, the direction with smaller traffic load will be selected. To achieve fault-tolerance in addition to adaptiveness, we use a reconfiguration process and define a new routing algorithm for all routers inside a contour surrounding a faulty link. In this reconfiguration process, the configuration registers of faulty component's direct and indirect neighbors are updated to be used in the routing process. But, this process does not comprise any message passing between adjacent routers. Other routers use the paths selected by the adaptive routing algorithm when no faulty link exists in the network.

In Fig. 4 the old broken paths and the new paths are shown for the contours S2 and S5 based on the reconfigurable routing. In addition, it shows the location of each node with respect to the faulty link which is maintained in each node's configuration register. In this figure, P1-P5 are the old broken bidirectional paths and NewP1 and NewP2 are the new paths which can be used instead of P1. In this way, if the current and destination

nodes belong to the same row or column like W and E nodes in Fig. 4, the new routing algorithm selects a new path direction from two new possible output channels (for W to E or vice versa it selects N or S direction which leads to one of the paths NewP1 or NewP2) perpendicular to the old path. The new direction which leads to a new shortest path between the current and destination nodes is determined in such a way that it incurs less traffic load using a congestion-aware method. For some node pairs, the faulty link breaks only one of two possible paths (P2-P5) and thus, the remaining path is used. For example, for W to SE the path from W to SW to SE is used. This is also the case for the nodes inside the contours S1, S3, S4 and S6. But, for node pairs in diagonal such as NW to SE in Fig. 4, there are two possible directions to move towards the destination. Hence, the direction selection process is not affected by the faulty link and it is the same as when there is no faulty link. It should be noted that this algorithm is an oblivious routing one and the path direction selection is independently performed in each router. In addition, for the destination nodes beyond these contours a similar approach is used.

According to the main contours S2 and S5 for a faulty link in a 2D mesh, a given router can be in 13 different positions: six positions for the neighboring of horizontal faulty link regarding six nodes inside S2, six positions for the neighboring of vertical faulty link regarding six nodes inside S5 and one position for a router which is not a neighbor of a faulty link and thus is configured as Fault Free. Therefore, only a four-bit configuration register in each router is enough to store its position, which will be used in the routing function. In addition, after detecting a faulty link by a detection mechanism, configuration registers of maximum six routers should be updated.

LFX procedure is based on using the new or alternative paths described before instead of faulty paths. It is an adaptive and distributed routing, and can tolerate all one-faulty link situations. In this algorithm, the perpendicular directions are the directions orthogonal to the straight path between source and destination. The congestion factor used in this algorithm is the number of occupied cells in all input buffers of an adjacent router [8]. To distribute the traffic while selecting between two directions, one of the outputs shown in Fig. 5 is used. Each output in this figure is one bit and indicates the preferred direction. For example, if NE is zero (one) the north direction has smaller (bigger) congestion factor relative to the east direction. The outputs NS and EW are used to select between two opposite directions, and this only occurs when a straight path between the current and destination nodes on the same row or column is faulty. However, the other outputs may be used in faulty or non-faulty conditions.

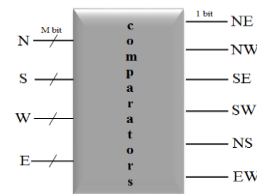


Figure 5. Path selection using congestion factors from adjacent routers.

Deadlock Freeness

Deadlock may occur when a cyclic dependency is formed in a group of packets i.e., each packet of the group requires a channel which is occupied by another packet of the group. The LFX algorithm prevents deadlock by using only one extra virtual channel. When the network is not faulty or a faulty link does not affect shortest paths directions of a specific source and destination nodes based on their locations, LFX acts similar to the DyXY method and needs two virtual channels for its adaptivity since two virtual channels are sufficient for an adaptive minimal routing algorithm in a mesh network to be deadlock free. When a faulty link affects the shortest paths, two cases may occur. If the source and destination nodes are not on the same row or column, then, the remaining direction or output channel towards the destination should be selected without any concern about its congestion factor. This selection does not incur more than two virtual channels since this type of selection still follows a shortest path which an adaptive minimal routing may offer, too. If the source and destination nodes are on the same row or column, then, one of two perpendicular directions should be selected based on congestion factors. In this case, because of creating the contour S2 or S5, it needs two virtual channels for its adaptivity. Therefore, the LFX method is a deadlock free routing algorithm.

IV. LFCXY ALGORITHM FOR IMPROVING DYXY

The goal of the LFCXY routing algorithm is to avoid the problem of the DyXY algorithm. This is obtained by using a flag which demonstrates congestion along the path of a row or column. This flag spreads in a row or column and demonstrates to the adjacent rows (or columns) that this row (or column) is near saturation and should be avoided. Since congestion flag should spread along a row (or column), each switch transparently spreads its prior switch congestion flag. Also, each router monitors its input buffer. If the occupied percentage of the buffer is larger than a threshold value, the corresponding switch will activate its congestion flag. The congestion information for the West, East, North, South input port is given to the switches which are to the left, right, up, and down-side of this switch, respectively. To effectively track the congestion condition, two flags are added to each row (or column): one informs left side switches of congestion in a right side switch and the other one informs right side switches of congestion in a left side switch in the row. For this purpose, between every two adjacent switches, two congestion wires, one in each direction, are added. The congestion wires are added to the wires used to transmit the free buffer count. The congestion flag transmitted by any switch is obtained by ORing its flag and that of the previous switch in the direction that the flag is transmitted. Therefore, if the flag is active, it shows that either the current or at least one of the previous switches is congested.

In LFCXY routing, if a switch is one hop apart from the destination in a row (or column), the congestion flag in the destination column (or row) is used in routing decision. If the congestion flag is one, the algorithm select the other path to the destination. For the example shown in Fig. 2, the congestion wire in the destination row is passed

to switch (4, 2) by switch (1, 2). Therefore, the packet will be routed to switch (4, 1).

In LFCXY, every router first looks at the destination address of the received packet. If the router address is not one hop apart from that of the destination in either the X or Y direction, the LFCXY algorithm ignores the congestion wire and routes the packet the same way as the DyXY algorithm does. However, if the destination address is just one hop apart from the router in either the X or Y direction, in addition to the current queue length, one of the congestion wires is also used for routing. We decided to use this congestion value as most significant bit of the stress value. In these cases, we refer to the port which is one hop apart from the destination in either X or Y directions as critical port while the other port is called non-critical port. For the non-critical port, we ignore congestion flag value and a zero is used for the congestion value to select this port against the other port.

It should be noted that congestion wires may report false positive information because they do not provide information regarding the location of the congestion node in a row (or column). We compared shared congestion wires with the general case in which the position of the congestion node is also spread in a row or column. The results show only slight differences between these two mechanisms. We proposed shared congestion wires because they are simple and have the desired characteristics for reducing latency.

For the implementation of the LFCXY routing algorithm, extra hardware should be added to the DyXY router. This extra hardware is divided into two parts. An extra unit is needed in each router to drive the congestion wires in four directions. In addition, each router should have extra logic to use the congestion wires in the routing decisions. In each direction, the output congestion wire is set if the input congestion wire due to the congestion in the previous routers in that direction is set or the occupied part of the input buffer for routing in that direction is larger than the threshold value. For implementing this logic, a comparator and an OR gate are used. The comparator compares the queue length with the predefined threshold value. In the case of exceeding the threshold value, the output of the comparator becomes one. The output congestion wire is the OR of the comparator output and the input congestion wire as shown in Fig. 6. The high level description of LFCXY routing algorithm is illustrated in fig. 7.

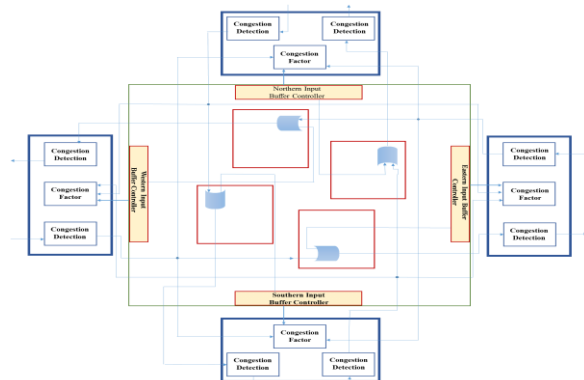


Figure 6. LFCXY switch implementation

V. EXPERIMENTAL RESULTS

Simulation Setup

To evaluate fault-tolerance capabilities, appropriate simulations were conducted using a cycle-accurate SystemC simulator [16]. We modified this simulator to support fault injection capability and related computations. In all NoCs with different sizes, the input buffer size of routers and the packet length are set to four flits. Each simulation was initially run for 1000 cycles to allow transient effects to stabilize and afterwards, it was executed for 2000 cycles in each iteration where each one included one or more random faults. The number of iterations is between 10000 and 20000 dependent on the network size. To evaluate performance, a 5×5 VHDL-based NoC architecture is used with the packet length of 16 flits and the input buffer size of four flits. In this architecture, router delay equals four clock cycles.

Performance Evaluation

To demonstrate the network performance of the LFcXY method, we compare the simulation results for the average packet latencies and saturation points in a 5×5 NoC under the uniform traffic pattern for three cases: A normal network with no fault and two examples of faulty networks, one with three and another with five faulty links as depicted in Fig. 8a. As shown in Fig. 8b under low traffic loads, the impact on the average latency is negligible. In addition, the saturation points of the normal, three-fault and five fault NoCs are equal to 0.31, 0.25 and 0.24 of maximum traffic load, respectively. The differences between the latencies in the normal and faulty NoCs are inevitable since the mandatory turns around faulty links naturally increase average latencies. In addition, networks throughput in this situation is demonstrated in Fig. 9.

Reliability evaluation

To compute the reliability enhancements of the LFCXY and LFCXY routing algorithms, NoCs with different sizes are imposed to many random faults under the uniform traffic pattern. It is sufficient to only simulate under this traffic pattern which uses all NoC links. In some synthetic or real traffic patterns some NoC links are not used. Thus, simulating under these traffics achieves reliabilities equal or higher than the ones obtained under the uniform traffic since unused length and packet injection rates. Both LFCXY and LFCXY methods tolerate all single-faulty-link situations. In addition, LFCXY algorithm avoid the problem of the DyXY algorithm. Although adaptive routing methods such as DyXY are more reliable compared to deterministic ones in transmitting the packets to their destinations, they cannot guarantee the correct delivery of all packets in faulty links do not decrease network reliability. In addition, the obtained results are independent of packet fault situations. Even if a retransmission method is used, it is probable that the redundant packets pass again via the previous faulty path due to congestion. Thus, we consider that a network structure is reliable or fault-tolerant if it correctly delivers all packets to their destinations in spite of existence of faulty components. Therefore, Fig. 10 shows the

percentage of reliable faulty network structures amongst all faulty networks including one faulty link. In other words, it shows the probabilities for NoCs to be still reliable after incurring one faulty link in manufacturing or operational phases.

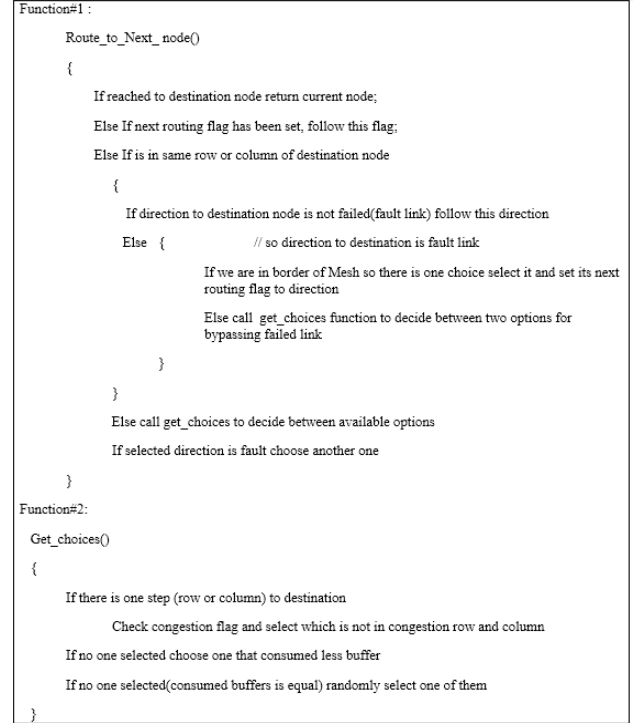


Figure 7. High-level description of the LFcXY routing algorithm.

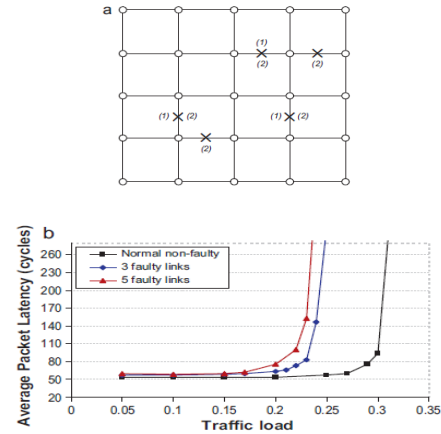


Figure 8. (a) Two faulty link scenarios: three faults (1) and five faults (2). (b) Average packet latencies in the normal and faulty 5×5 NoCs

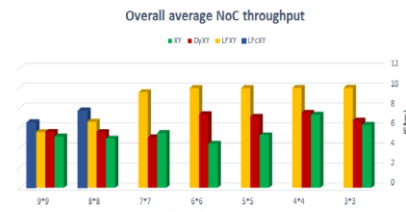


Figure 9. Throughput against faulty link

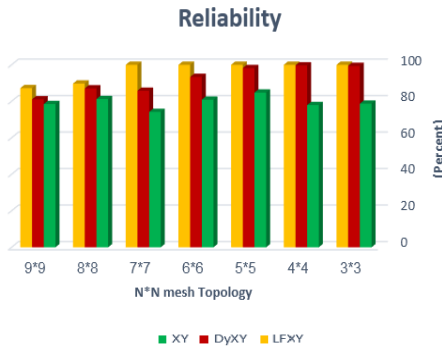


Figure 10. Reliability against faulty link.

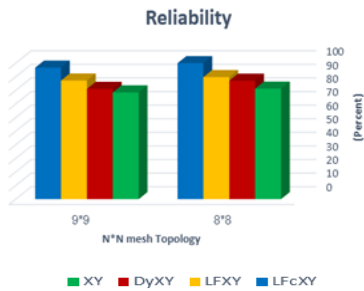


Figure 11. Reliability against faulty link.

Fig. 11 shows the percentage of reliable network structures for 8×8 and 9×9 NoCs using the LFCXY routing algorithm when one bidirectional link is broken. As this method tolerates the permanent faults occurred in manufacturing phase in addition to operational phase, the yield parameter which is the percentage of correctly working chips amongst all produced chips, directly relates to these reliabilities and therefore it is greatly enhanced in faulty situations. For example, based on Fig. 10 for N equal to 8 when one faulty links exist in a network, the yield increases from 84.9% to 100% when we use LFCXY instead of LFXXY. It should be noted that an adaptive method such as DyXY does not have any yield for this faulty situation.

CONCLUSION

In this paper, a general, reconfigurable and fault-tolerant routing method is presented, suitable for applications targeting NoC designs in unreliable environments. This method uses simple and efficient approaches to tolerate one link; however, it is designed such that it favorably supports the mesh networks with many more faults. In addition, this method can incorporate different congestion-aware approaches to adaptively distribute congested traffic, especially caused by faulty components. We presented the details of two variants of this method. Both of them without any extra virtual channel compared to adaptive routing algorithms. Hence, they need only two virtual channels. Based on evaluations, this method causes small hardware overhead while it considerably increases the reliability and yield of mesh-based NoCs. Thus, this method enables the deployment of NoC architectures where correct operation and network connectivity should be maintained possibly at a cost of

graceful performance degradation as network components fail. Moreover, this method can be extended to support 2D torus topology in addition to unidirectional faulty links.

REFERENCES

- [1] Chalasani S, Boppana RV. Fault-tolerant wormhole routing algorithms for mesh networks. *IEEE Trans Computers* 1995; 44(7):848–64.
- [2] Sui PH, Wang SD. An improved algorithm for fault-tolerant wormhole routing in meshes. *IEEE Trans Computers* 1997; 46(9):1040–2.
- [3] Park S, Youn JH, Bose B. Fault-tolerant wormhole routing algorithms in meshes in the presence of concave faults. In: *Proceedings of International Parallel and Distributed Processing Symposium (IPDPS)*. 2000. p. 633–8.
- [4] Ubar R, Raik J. Testing strategies for Network on Chip. In: Jantsch A, Tenhunen H, editors. *Networks on Chip*. Kluwer Academic Publisher; 2003. p. 131–52.
- [5] Li M, Zeng QA, Jone WB, DyXY. A proximity congestion-aware deadlock-free dynamic routing method for Network on Chip. In: *Proceedings of Design Automation Conference (DAC)*. 2006. p. 849–52.
- [6] Gratz P, Grot B, Keckler SW. Regional congestion awareness for load balance in Networks-on-Chip. In: *Proceedings of International Symposium on High-Performance Computer Architecture*. 2008. p. 203–14.
- [7] Lotfi-Kamran P, Daneshmand M, Lucas C, Navabi Z. BARP-A dynamic routing protocol for balanced distribution of traffic in NoCs. In: *Proceedings of Design, Automation and Test in Europe (DATE)*. 2008. p. 1408–13.
- [8] Kim J, Park D, Theodorides T, Vijaykrishnan N, Das CR. A low latency router supporting adaptivity for on-chip interconnects. In: *Proceedings of Design Automation Conference (DAC)*. 2005. p. 559–64.
- [9] Glass CJ, Ni LM. Fault-tolerant wormhole routing in meshes. In: *Proceedings of Annual International Symposium on Fault-Tolerant Computing (FTCS)*. 1993. p. 240–9.
- [10] Zhang Z, Greiner A, Taktak S. A reconfigurable routing algorithm for a fault tolerant 2D-mesh Network-on-Chip. In: *Proceedings of Design Automation Conference (DAC)*. 2008. p. 441–6.
- [11] Fick D, DeOrio A, Chen G, Bertacco V, Sylvester D, Blaauw D. A highly resilient routing algorithm for fault-tolerant NoCs. In: *Proceedings of Design, Automation and Test in Europe (DATE)*. 2009. p. 21–6.
- [12] Zhou J, Lau FCM. Adaptive fault-tolerant wormhole routing in 2D meshes. In: *Proceedings of IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 2001.
- [13] Nunez-Yanez JL, Edwards D, Coppola AM. Adaptive routing strategies for fault tolerant on-chip networks in dynamically reconfigurable systems. *IET Compute Digit Tech* 2008; 2(3):184–98.
- [14] TRAN, A. T. 2012, “On-Chip Network Designs for Many-Core Computational Platforms”, Ph.D. Thesis, Office of Graduate Studies of the University of California. California, USA.
- [15] A. M. Rahmani-Sane., Dec. 2012, “Exploration and Design of Power-Efficient Networked Many-Core Systems”, Ph.D. Thesis, University of Turku Department of Information Technology, Turku, Finland.
- [16] NIRGAM. <http://www.nirgam.ecs.soton.ac.uk>, Ver. 1.2; 2010.