A Fault Tolerance NoC Topology and Adaptive Routing Algorithm

Pengfei Yang, Quan Wang*, Wei Li, Zhibin Yu, Hongwei Ye
School of Computer, Xidian University
Xi'an, China
*Corresponding Author: qwang@xidian.edu.cn

***Abstract*—The congestion in any regions or the faults in routing nodes, communication links and processing elements(PEs) will affect the NoC performance seriously in the multiprocessor system-on-chip (MPSoC) based on network on chip (NoC). In this paper, the NoC topology is designed which the fault tolerance function is enabled. Its dual-port network interface (NI) makes the key PE have two links with network at least, which can ensure high reliability of the system. The adaptive fault tolerant routing algorithm is also designed that could sense congestion on the network, and the routing nodes can well sense the congestion regions or error nodes on the network to work around the issues effectively. Experiment results demonstrate that the design proposed in this paper can ensure normal network communication and good system performance even if the congestion in routing nodes, communication links, processing elements, or any regions occurs.**

***Keywords*—NoC; fault tolerance; adaptive routing algorithm**

I. INTRODUCTION

As CMOS technology scales down into the atomic domain, the deviation and vulnerability to aging of large-scale integrated circuit chips become more obvious, and the system fault rate increases greatly [1][2]. The faults, transient or permanent [3][4], will affect the system performance and challenge the system reliability. In view of this, the large-scale digital integrated circuit must have capable of self-repair(SR) both at circuit level and system level to improve the rate of finished chips, lower costs, and construct a stable, reliable, and performance-predictable system.

The system reliability is affected by the system design solution, quality of components, and operating environment. To improve the system performance, a good design solution and high quality components are necessary. However, this will increase costs and only decrease the fault rate, it is impossible to eliminate faults from the system. Therefore, a proper fault tolerance mechanism is a shield against the system faults, which is an important part of the reliability design. The research of Intel shows that in the future, resources used to design the fault tolerance mechanism will occupy 5% to 10% of the whole chip resources [5].

In the chip-scale interconnection aspect of SoC which is the major design and integration method of embedded system, NoC replaces the bus on chip has become the main internal interconnection method of MPSoC system. Multiple hop data packet transmission is used to shorten the communication links among multiple processors and reduce soft faults, such as crosstalk and electromagnetic interference. At the same time, the application environment of NoC is fixed, so human-made destruction on links and routers is reduced, which improves the chip reliability. However, there are some issues of small inside size and large integration scale while NoC appears. It is hardly inevitable for the process of manufacturing and application to the invalidation or faults of some inside components. These issues can affect that the normal communication cannot achieve among multiple processors so that the rate of accepted chips decreased. So it is necessary to make a research on NoC reliability and fault tolerance.

This paper is supported by National Natural Science Foundation of China under Grant 61572385, Science & Technology Program of Shaanxi Province, China (Grant NO.2015KTCXSF-01), NSFC under Grant 61303034

II. RELATED WORK

NoC reliability fault tolerance means that taking effective measures to make network activities can still meet performance requirements when its components become invalid or data transmission is affected by external environment. As shown in Fig.1, the communication architecture of NoC can be classified into application layer, transmission layer, network layer, and physical layer, similar to the hierarchical communication architecture of Open System Interconnect(OSI). As one of the major service quality issues, NoC reliability can be studied from different layers.

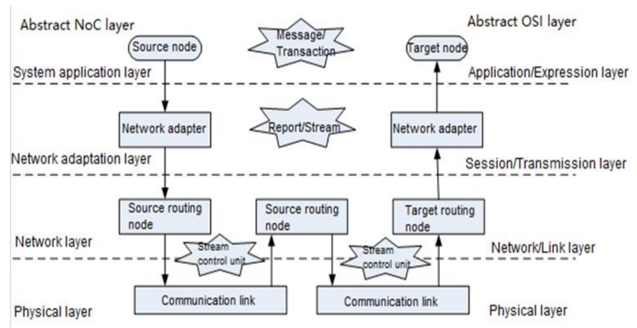


Fig.1. Communication architecture model of NoC

The physical layer defines the routing method and signaling method of NoC. NoC reliability and fault tolerance on the physical layer is selecting or designing proper signals which can meet communication requirements and designing rational line structure based on characteristics of electric apparatus to reduce interference.

The network layer is used to send data packages from source node to target node. For this layer the fault tolerance can be designed from two aspects: one is the architecture level(topology), the other is routing algorithm [6].

The topology should ensure that at least two paths exist between any source node and target node. In this way, data packages can be transmitted without interruption when a link or routing node fail. At the same time, the topology can [7] optimized the design of links and routing nodes to make better fault tolerance. In addition, the topology [8] generates the best

978-1-5090-3727-8/16 $31.00 © 2016 IEEE
DOI 10.1109/ICESS.2016.20

fault tolerance topology algorithm from the redundancy topology as required. This design is actually a kind of hardware redundancy design, which will add extra topology components. Algorithm [9] helps to find a way to enhance NI reliability and flexibility.

Traditional network fault tolerance methods of the routing algorithm are extensively used for reference, but the limitation of those algorithm restricts its application on NoC, for example: broadcast routing algorithm has large power consumption, and the error retransmission routing algorithm causes a long time delay and a low throughput rate.

If the processing element connected to NoC fail, the task assigned to this PE on the application layer cannot be timely and correctly executed. That will affect the execution of the entire task queue and then influence the system performance. Therefore, it is necessary to allocate the task that should be executed by the failed processing element to other processing elements. The topology reset concept is proposed [10]: the redundant PEs configured in the system do not execute tasks in normal status, if some PEs in the system fail, the redundant PEs will replace the failed elements to execute tasks using remapping from physical topology to logical topology. The topology reset ensures that structured topology can be used by the application layer when the physical topology changes. However, parameters, such as communication link length and transmission delay of the PEs at the same logically location, have changed after the topology reset. If these parameters are allocated in the same manner, the allocation results will have very large fluctuations in practice. Therefore, relevant parameters should be updated when using the topology reset.

The issue of reliability and fault tolerance is systematic subject. It is necessary to take the fault tolerance into account during each design process to build a system with high reliability. Based on the preceding studies, the fault tolerance design is performed from topology and routing algorithm in this paper to solve NoC issues and guarantee the system reliability.

III. Fault Tolerance Topology Design

A Honeycomb-like NoC is designed in preceding research. As shown in Fig.2, according to different requirements for network communication performance, the topology consists of two different units. To meet the large communication demand of intensive communicating units, an additional switching node is placed in the hexagon's center which only connects to the hexagon's six vertexes to provide more data transmission paths. The topology can easily extend in 2D or 3D space (Fig. 3), comparing with common topology like Mesh, Torus and so on, the new topology have advantages on system resources occupancy and performance. But the reliability fault tolerance is not considered at the early stage. It would have three main failure parts on the network: transmission link failures, routing node failures, and PE failures. Transmission link failures will cause frustration of communication; PE failures will cause that tasks assigned on it cannot be executed timely. If a routing node connected to a PE fail, not only its links connected in the three directions fail, but also the connected PE cannot transmit data. As a result, the communication and processing performance of a whole region of the system will be greatly reduced or become invalid. The system performance will be seriously affected. To solve this problem, dual-port network nodes are added to the original topology units in this article.

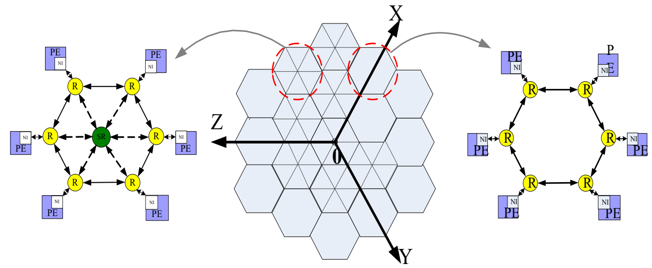


Fig.2. Basic units of cellular-alike topology

The NI of traditional NoC has a structure with single input and single output that connects with PE and routing node respectively. To enhance the system fault tolerance and improve reliability, the dual-port NI is created by adding multiplexers and arbitration units. Fig.4 shows its internal structure.

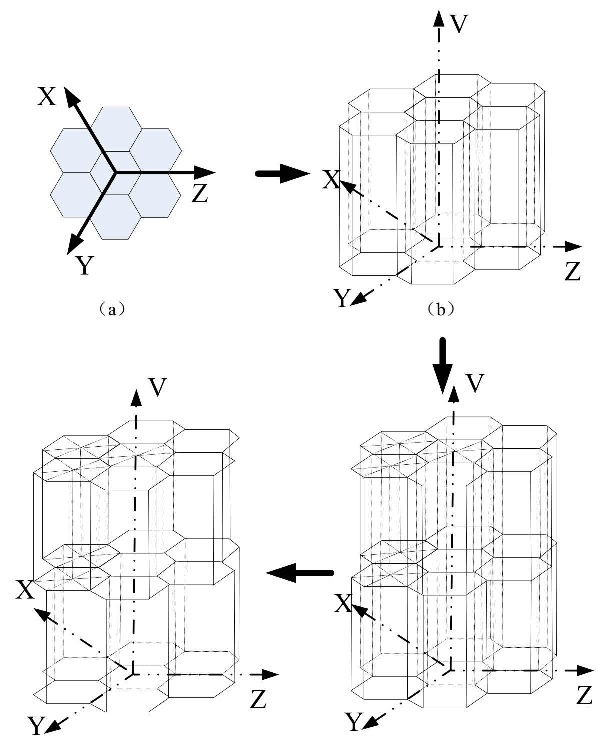


Fig.3. 3D extension of cellular-alike topology

Encoders and decoders are used to convert the data format between PEs and NoC. The send/receive buffer is used to store the data generated during the send and receive process. The send and receive control logic units are used to control the data sending or receiving. The receive arbitration and multiplexer on the receive end and the send arbitration and multiplexer on the send end are used to select paths to receive and send data respectively.

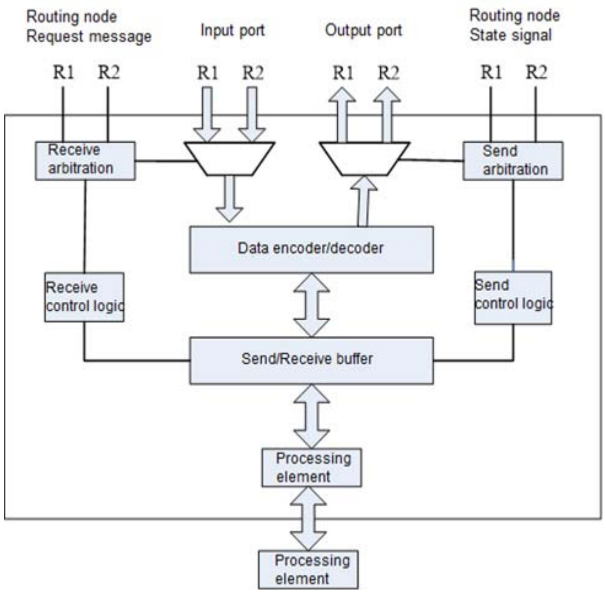


Fig.4. Structure of a dual-port NI

The dual-port NI changes the structure of the original topology, as shown in Fig.5. In the original topology, each PE is connected with a unique routing node. The data sent and received by the PE has to pass the routing node, which has poor fault tolerance. In the new topology, a switching routing (SR) node is added in the middle of basic units. Each PE has two paths for receiving and sending data packages at least.

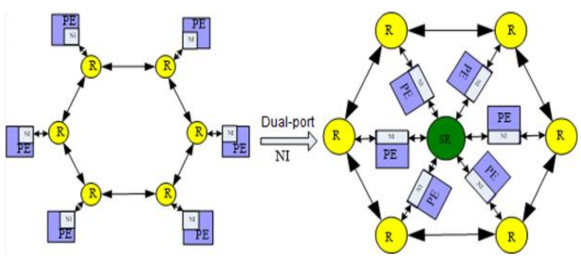
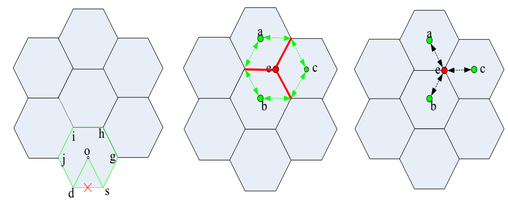


Fig.5. Basic units of network topology

In this NoC topology with the dual-port NI, two hops will be added if the data is transmitted to the network through NI near to SR comparing whit the edge routing node. So, in the NI arbitration, preferentially choose the port at the edge of hexagon to send data packages. The dual-port NI provides more paths for data package transmission, and it improved the fault tolerance capability and the system reliability. If the communication channels between routing nodes are disturbed, some transient failures may occur, such as bit error and bit loss. Too much data that exceeds the capacity of link resources is transmitted, high congestion may also emerge. In Fig.6(a), at least two paths exist between any two nodes. If one path fail, data can be sent through the other path without causing system failures. For example, if s-d fail, data can be transmitted through s-g-h-i-j-d to d. If a switching node o exists in the center, data also can be transmitted through s-o-d to d. Data transmission is not blocked.

Traditionally, if one routing node failed, not only its connected links become invalid, but also the processing element connected with the node cannot be invoked because of losing Link between NIs. However, the network designed in this paper is different from the traditional network. In Fig.6(b), if routing node e failed, the three standby nodes a, b, and c can form a closed loop around the fault region, thereby ensuring the normal operation of the system. At the same time, in Fig.6(c), the PE connected with routing node e uses the dual-port NI. Even if routing node e failed, this PE can still send and receive data through node b, wide area failures will not occur.



(a) (b) (c)

Fig.6. Failure topology model

The PE is independent from the routing node, but its failure can also not be ignored. A proper PE must be selected to replace the failed PE to execute tasks assigned on it, which involves the task remapping. The task remapping is performed on the application layer of NoC, when a failed PE is detected by the system, it will be excluded at the task scheduling stage. The system will re-update parameters of communication processors according to available resources and then conduct task scheduling and mapping algorithm.

For 3D Honeycomb-like NoC, except for the preceding failures, transmission failures may occur in vertical layer links. Using fault tolerance algorithm can work around these issues and select other available vertical layer links to transmit data.

IV. FAULT TOLERANT ROUTING DESGIN

First, a parameter called "Buffer occupancy rate Buffer_n(n=X, Y, Z)" is defined. During data transmission, the transmission direction of X, Y, Z is deterministic according to the location of the current node (Xc, Yc, Zc) and the target node (Xd, Yd, Zd), Denote the VC buffer occupancy rate of the input port of the next routing node in the X direction by Buffer_x. Denote the VC buffer occupancy rate of the input port of the next routing node in the Y direction by Buffer_y. Denote the VC buffer occupancy rate of the input port of the next routing node in the Z direction by Buffer_z. According to the value of Buffer_n (n=X, Y, Z), divide the network congestion into three levels: No congestion (Buffer_n is empty), low congestion (Buffer_n ≤ 50 %), and high congestion (Buffer_n>50%). If the link in one direction failed, the value of Buffer_n in this direction is set to 1. If one routing node failed, the value of Buffer_n in all directions of the node is set to 1.

For 2D topology, before each data hop, determine whether offsets exist in the three directions. Select the path with minimum Buffer_n to transmit data. If the offsets exist only in one direction, the congestion comparison is not required, directly transmit data packages in this direction to the target node. In Fig.7, S1 is the source node and D1 is the target node. Assume that the coordinate of S1 is (x1, y1, z1) and D1 is (x2, y2, z2), and the offset in the X direction is X_offset=[x2-x1], in the Y direction is Y_offset=[y2-y1], and in the Z direction is Z_offset = [z2-z1]. The shortest path between S1 and D1 is formed by side |X_offse| parallel to the X axis, |Y_offse| parallel to the Y axis, and |Z_offse| parallel to the Z axis. See the paths denoted by blue arrows. However, data may not be transmitted in accordance with the blue paths in actual situations. According to the link congestion level, flexible path selection and adjustment are used in data transmission. The paths denoted by green or orange arrows are also possible to be selected to transmit data.

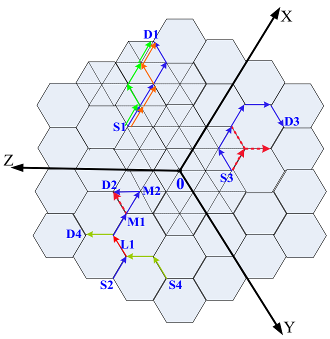


Fig.7. Example of NoC routing algorithm

When a number of data packages are transmitted on the network at the same time, it is inevitable that two or several data packages request the same data links. For example, in Fig.7, L1 is simultaneously requested by data from S2 to D2 and data from S4 to D4, the contention of network resources is generated. If this contention issue is not solved, data lovelocks as well as transmission delay will occur, plenty of received data cannot be processed because of the absence of a key data package, the system performance will also be affected, in addition, large-scale data caching will occupy limited storage resources of the system and restrict the transmission efficiency. An effective routing algorithm must be capable of sending the data packages with long transmission delay to the target node as soon as possible.

When designing the data package format on the network adapter layer, we use the register Hop to count the data package hopping times during the entire transmission process. When the data package is sent from the source node to the network, the value of Hop is set to 1. Then, the value of Hop automatically adds 1 for each store and forward on the intermediate node until the data package is received by the target node. When conflicts occur between two or several data packages because of resource contention, the routing node assigns priorities based on the hop times of each data package, the data package with more hop times has higher priority. If two or several data packages have the same Hop value, assign priorities by determining the congestion on the next forward link of the contention link, the package with less congestion has higher priority. In this way, the congestion caused by resource contention can be avoided.

For the contention of L1, in Fig.7, the data package from S2 takes one hop when arriving at L1, so its Hop value is 1, the data package from S4 has taken two hops when arriving at L1, so its Hop value is 2, the priority of data package from S4 is higher than that from S1, L1 should be assigned to the data package from S4.

If a link on the network fails, the two values of Buffer_n in the two directions of this link are set to 1, if a routing node on the network fails, the values of Buffer_n in all the directions of this node will be set to 1. During the transmission of a data package, if the Buffer_n of the next available link is greater than 50% or equal to 1, another link with Buffer_n less than 1 will be requested instead of this link. For example, if link M1D2 in Fig.7 failed, buffer_Y in the Y direction of node M1 is set to 1, a detour is required for transmitting data packages from M1 to D2, which selects a link which lower Buffer_n to implement one step furth er, link M1M2 in the X direction is the only choice.

For 3D topology, if the source node and the target node are in the same layer, the routing algorithm is the same as 2D topology. If the two nodes are in different layers, firstly, find a corresponding node with the same location as the target node in the source node layer, as shown in Fig.8(a), S is the source node and D is the target node, the routing algorithm finds D1 corresponding to D in the layer which S located. Then transmit data packages using preceding algorithm, when meeting the first vertical link, use this link to transmit data packages along the direction of V value reduced. That means, during the data package transmission, the priority of cross-layer links is higher than the links on the same layers. When the data packages arrive at the layer of the target node, reuse the routing algorithm of 2D topology.

Similar to the structure of the same layer, NoC in vertical cross-layer links will encounter failures or congestion links. In Fig.8(b), when meeting the first vertical link which is in high congestion or fault state, continue to transmit data packages on the same layer until meeting the next vertical link which is not in congestion to implement cross-layer transmission.

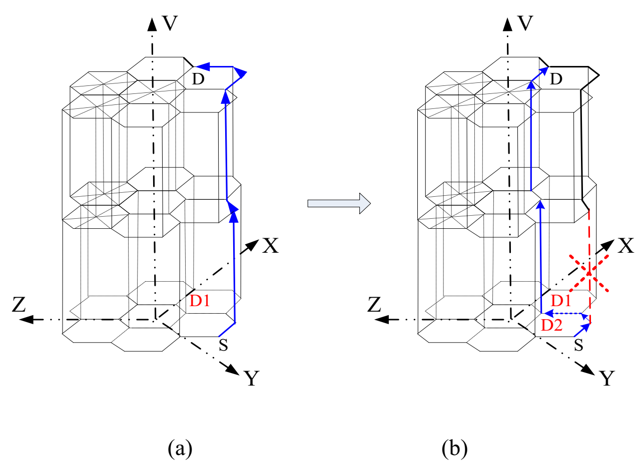


Fig.8. 3D fault Tolerance Routing Algorithm

The process of routing algorithm is:

1) Receive a data package and parse its target node.

2) If the addresses of the target node are same as the current node of router, send the data package to the PE connected with the router and the algorithm is over, if the addresses are different, go to step 3.

3) Determine the transmission trend and offsets in direction of V, if the Buffer_v is less than 50%, send the data one step further and go to step 2. If the Buffer_v is more than 50%, go to step 4.

4) Determine the transmission trend and offsets in four direction of X, Y, Z. Compare the Buffer_n of available links, select a path with minimum Buffer_n as the transmission link to implement one step further towards the target node, go back to step 2. If only one path is available and its.

5) Buffer_n is greater than 50% or equal to 1, go to step 5.

6) Do not consider the offset of the three directions of X, Y, Z and select a path with minimum Buffer_n in the available directions, send the data package one step further and go back to step 3.

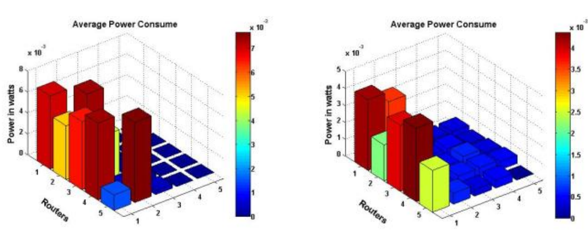
V. Experiments And Comparison

For the NoC system, the power consumption and packet delay during the transmission are two important parameters showing the performance. We use NIRGAM emulator as the experimental environment, and make a comparison between the 5x5 Mesh topology and Honeycomb structured with 24 nodes. Firstly, if not considering the congestion or fault area in the network, the main parameters are shown in TABLE 1 and the comparison results are shown in Fig.9 and Fig.10.

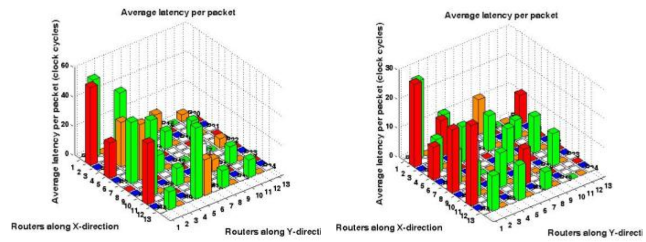
TABLE I. Main Parameters Setting

Parameters type	Parameters setting	
	Mesh	***Honeycomb-like***
PE number	25	24
Routing algorithm	dimension-ordered	dimension-ordered
NUM BUFS	16	
FLITSIZE	4 bytes	
SIM NUM	2000	

The maximum power consumption of the system node is reduced from 0.007 watts to 0.005 watts, besides, no router is much busier than others, and four high-power nodes spread out into six lower nodes. It is obvious that the honeycomb-like topology are significantly better than the Mesh structure in terms of both average latency and power consumption which takes less system resources.



(a(Dimension ordered routing (b) Proposed method
Fig.9. Average power consumption



(a(Dimension ordered routing (b) Proposed method
Fig.10. Average latency per packet

Then, conduct fault tolerance experiments to verify the fault tolerance feature of adaptive fault tolerant algorithm. In Fig.11, build a 3D honeycomb-like topology on the NIRGAM platform, set two faults in the system: one is node 7 which fails on the paths S1 to D1; the other link error occurs on the path S2 to D2. Set node 1 and 20 as constant bit rate, and their fixed target node are 9 and 18, implement dimension-ordered routing algorithm and advanced adaptive fault tolerant(AFT) routing algorithm respectively, the results are shown in Fig.12 and Fig.13.

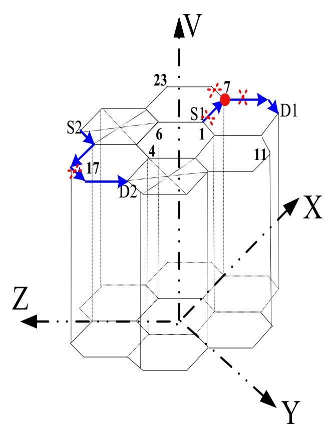
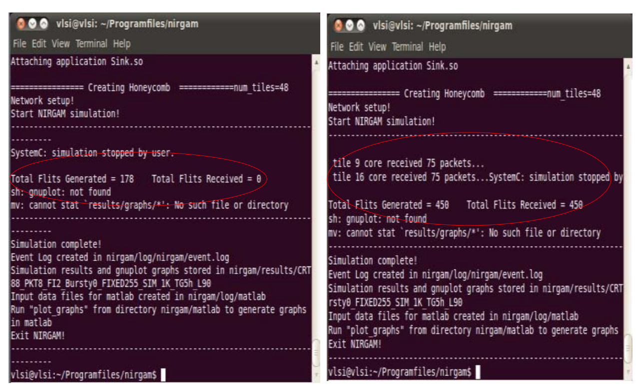
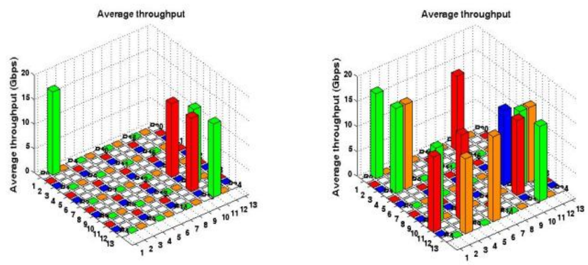


Fig.11. 3D topology with two failure points



(a) dimension-ordered routing (b)AFT routing
Fig.12. Comparison of data reception

For the dimension-ordered routing strategy, the target nodes cannot receive any data packages because of the failed node and link. However, for the advanced adaptive routing algorithm, data packages can bypass the failed node or link successfully and arrive at the target node. Such change can also be observed from the network throughput figure(Fig.13). Because of the failed points, the dimension-ordered routing algorithm can only calculate the throughput of data packages of the first half of the link. However, the throughput of the entire transmission link can be obtained for the advanced adaptive routing algorithm.



(a) dimension-ordered routing (b)AFT routing

Fig.13. Average throughput statistical chart

VI. Conclusion

In conclusion, the topology and routing algorithm we designed can ensure good system performance even if the congestion in routing nodes, communication links, processing elements, or any regions occurs. It has great application advantages.

References

[1] Fick D, DeOrio A, Hu J, et al. Vicis: a reliable network for unreliable silicon[C]//Proceedings of the 46th Annual Design Automation Conference. ACM, 2009: 812-817.

[2] Vitkovskiy A, Soteriou V, Nicopolous C. A fine-grained link-level fault-tolerant mechanism for networks-on-chip. In: 28th IEEE international conference on computer design, 2010. 447–54.

[3] Nicolaidis, Michael, ed. Soft errors in modern electronic systems. Vol. 41. Springer Science & Business Media, 2010.

[4] Yu Q, Ampadu P. Dual-layer adaptive error control for network-on-chip links. IEEE Trans Very Large Scale Integr VLSI Syst 2012;20(7):1304–17.

[5] Borkar S, Dubey P, Kahn K, et al. Platform 2015: Intel processor and platform evolution for the next decade[J]. Technology, 2005: 1.

[6] Liu J, Harkin J, Li Y, et al. Low cost fault-tolerant routing algorithm for Networks-on-Chip[J]. Microprocessors and Microsystems, 2015, 39(6): 358-372.

[7] Poluri P, Louri A. A Soft Error Tolerant Network-on-Chip Router Pipeline for Multi-Core Systems[J]. IEEE Computer Architecture Letters, 2015, 14(2): 107-110.

[8] Tosun S, Ajabshir V B, Mercanoglu O, et al. Fault-tolerant topology generation method for application-specific network-on-chips[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2015, 34(9): 1495-1508.

[9] Fiorin L, Sami M. Fault-Tolerant Network Interfaces for Networks-on-Chip[J]. Dependable and Secure Computing, IEEE Transactions on, 2014, 11(1): 16-29.

[10] Liu L, Ren Y, Deng C, et al. A novel approach using a minimum cost maximum flow algorithm for fault-tolerant topology reconfiguration in NoC architectures[C]. Design Automation Conference (ASP-DAC), 2015 20th Asia and South Pacific. IEEE, 2015: 48-53.