TECHNIQUE REPORT FOR AN ADAPTIVE 2D MESH NETWORK-ON-CHIP (NOC) ARCHITECTURE ASIC BASED DESIGN

Pham Dang Lam*, Nguyen Van Viet, Hoang Trang

Ho Chi Minh University of Technology, 268 Ly Thuong Kiet, Ho Chi Minh City, Vietnam

*Email: lamd.pham@hcmut.edu.vn

ABSTRACT

Not only is 2D-mesh Network-On-Chip (NoC) supposed to be an effective scheme to apply in multi-core systems and take the place of traditional bus system but it is also proposed to enhance the bandwidth mainly because of both simply architecture and extended ability feasibly. However, in order to apply the NoC, an efficient routing algorithm is always required to tackle the bottleneck. As regards integrated circuit fabrication, the low yield is also the root cause of leading to decrease productivity. Particularly, if any router node in NoC meets trouble causing of fault connection in fabrication process, the system throughput is affected critically. In this work, an effective 2D-mesh NoC architecture is proposed not only to halt bottleneck but also to be able to solve one fault-tolerant issue automatically. In order to perform proposed NoC obviously, a 4x4 configuration of 2D-Mess NoC generated by own tool using Perl language is verified and synthesized on 90 nm technology following ASIC base design. The final results on 4x4-configuration confirm that proposed NoC has experienced an average clock number at 7.1 for one flit to go through one router, bandwidth of proposed NoC at 0.56 GBs as well as maximum 500 MHz of high frequency.

Keywords: Network-On-Chip (NoC), adaptive routing, first in first out (FIFO), multi-core system, fault-tolerant routing algorithm, bottleneck.

1. INTRODUCTION

With development in semiconductor industry, more and more multi-core systems have been developed to adapt various applications. Presently, there is a fact that the traditional bus is not suitable to adapt strict requirements of multi-core systems, while the NoC models have been studied and achieved many successes. However, approaching NoC models have experienced many obstacles such as channel sharing, bottleneck, or fault tolerance [1]. One of popular solution to tackle these issues is adaptive routing algorithms [2]. To be more specific, it is separated in to two principal approaches known as fault tolerance routing for enhancing the yield of fabrication and adaptive routing for innovating the bottleneck issue. Firstly, the low yield index in fabrication is always the critical issue to apply the real products. Similarly, the products approaching NoC models not only depend on this index strictly mainly because of high density but their performances are also affected seriously. In order to improve the productivity, some fault tolerance

routing adaptive methods such as Probabilistic Flooding, Direct Flooding Algorithm, and Walk Random algorithms have witnessed successes; however, these algorithms have still lacked of many requirements such as resource limitation, complicated algorithm, reused ability and deadlock issue [3]. Another tendency concentrates to apply routing algorithm for promoting the speech, bottleneck issues. One of typical NoC schemes as Figure 1 in which router groups are controlled by central unit is described by Hamed S. Kia and Cristinel Ababei [4]. In this structure, the corresponding experiments confirm the double of throughput improvement to compare with the traditional OXY routing.

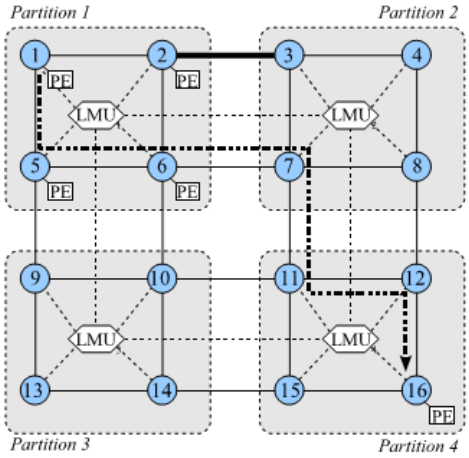

Figure 1. Central control unit NoC scheme.

Another dynamic routing concept based on the 2D Mesh topology as Figure 2 is illustrated by Terrence Mak at al. to confirm the average delay on different traffic by 22.3% [5]. In this architecture, one dynamic routing unit is designed to ensure the routing successfully.

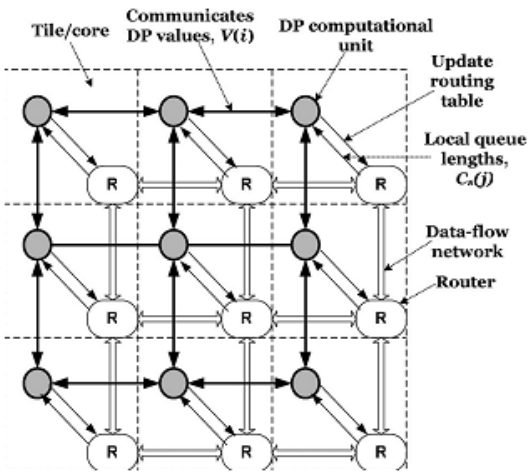

Figure 2. 2D-Mesh applying dynamic routing algorithm.

Overall, the upper solutions generally use private units to control the routing process. With the support from these units, the network performance is enhanced significantly; however, it becomes the considerable issue in enormous networks mainly because of reused property, resource limitation as well as how to control the complicate network. From such circumstances, a proposed NoC is presented not only to solve the bottleneck but also to halt one fault-tolerant issue.

The rest of this paper is organized follows. Section 2 gives the overview of related concepts of NoC to promote the chosen techniques. Next, Section 3 illustrates effective router architecture as well as techniques to solve the bottleneck and fault-tolerant issue. Section 4 shows the experiment results. Finally, Section 6 presents the conclusion and future works.

2. INTRODUCTION

2.1. Topology and connection

In order to develop a complete NoC, NoC topology is always decided at the first time probably because most of other techniques applied to NoC architecture are strongly affected by the utilized topology. The 2D-Mesh NoC as Figure 3 has been introduced as basic topology to expand to diversified schemes feasibly [6]. As matter of fact, most of researches usually approach the 2D-Mesh topology before achieving significant innovations on their own architectures. Consequently, this research also chooses 2D mesh topology to approach hardware architecture.

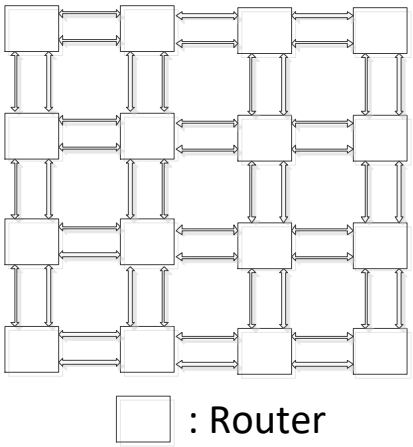

Figure 3. 4x4 configuration of 2D-Mesh NoC.

A basic router interface of 2D-Mesh NoC including in five connections namely West, East, South, North, and IP is described in Figure 4 obviously. At the West, East, South, or North direction of router, it is connected to the neighbor router, while an IP, CPU core, or any hardware structure is connected to IP direction.

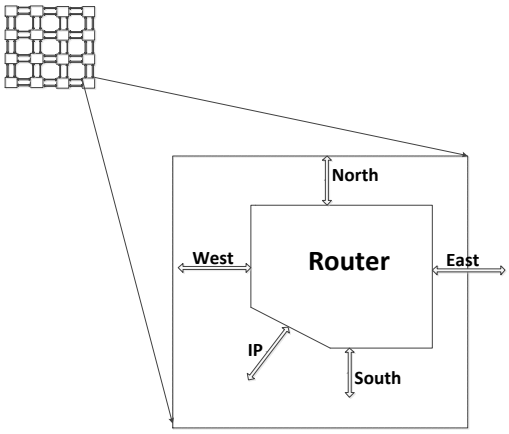

Figure 4. Connections of single router.

As regards one connection to any neighbor router, a single router in 2D-Mesh topology has two directions consisting of input and output as Figure 5. In order to transfer data between routers, a flow control normally affected by switched technology is always necessary. The detail descriptions for connection between routers as well as flow control methods are described below in detail.

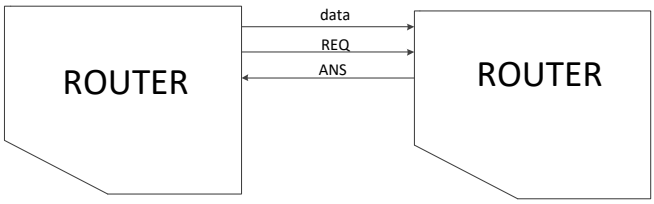

Figure 5. Connections between two routers.

2.2. Switched technology and flow control

At the present, some switch technologies denoted as Figure 6 have approached widely [7].

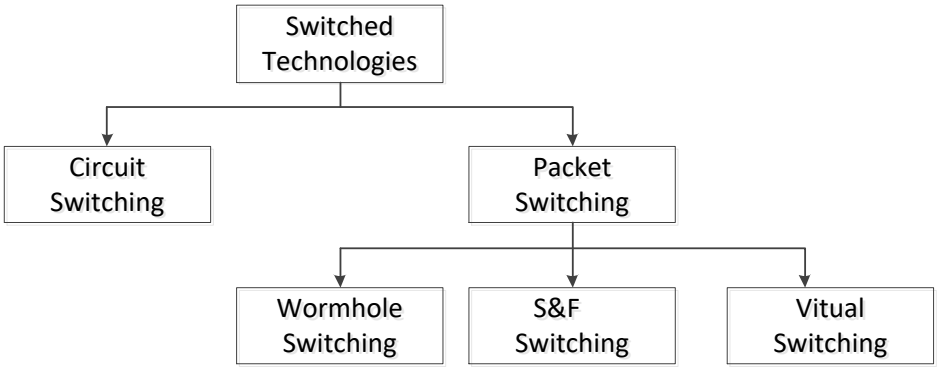

Figure 6. Switched technologies.

According to Figure 6, there are two main groups of switched technologies comprising of circuit switching and packet switching. While circuit switching is typical technique presenting a ready roadmap for data before transferring process, packet switching technology illustrates the connection setup on pair of two routers partly. As the result, the circuit switching technique has experienced the significant transferred process mainly because of ready roadmap identifying the exact routers to serve the complete process; however, it is really serious issue in high traffic situation in which serviced routers are always busy to complete long processes [8]. It is fact that packet switching technologies including in Wormhold, S&F, or Virtual have applied more popularly because serviced routers will be free more quickly after transferring only flit. Moreover, other techniques as virtual channel or adaptive routing algorithm are also applied to support packet switching to obtain the best performance. As the result, a REQ/ANS flow control based on package switching technology is proposed in this work.

2.3. Virtual channel and adaptive routing

In NoC development, virtual channel technique or adaptive routing algorithms have become effective solutions to tackle the bottleneck issue. Firstly, analyzing virtual channels points out a tradeoff between used resource and ability of halt the bottleneck issue. Particularly, the virtual channel number decides the increasing of internal buffers in router opposed to increase ability of sharing buffer to store data from many directions to solve the bottleneck problem. Figure 7 describes a virtual channel structure between two routers basically. As Figure 7, data from different buffers will share only one data bus. As the result, increasing virtual channel number also requires a complicated controller to arbitrate the data flows.

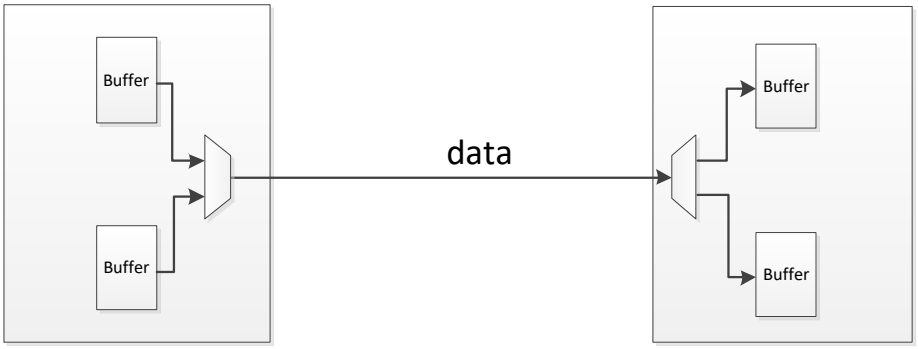

Figure 7. Two virtual channels.

As regards adaptive routing techniques, mainly these promoted algorithms are applied to halt the bottleneck issues in high traffic density. There are two principle tendencies to approach adaptive algorithms to NoC structure. The first applies an additional architecture to be able to control a router group [4, 5]; however, these algorithms have become unsuitable to enormous NoC requiring a hundred of integrated IPs. Another method makes effort to develop the complicate router in which adaptive ability is controlled by single router automatically. Figure 8 describes a hierarchy of routing algorithm following different criteria. In this paper, instead of using an adaptive routing, a combination among virtual channel, XY deterministic routing [9] and mask technique is proposed to solve bottleneck issue. Moreover, in order to enhance the yield after

fabrication, this proposed architecture has also integrated adaptive routing to be able to solve one fault tolerance.

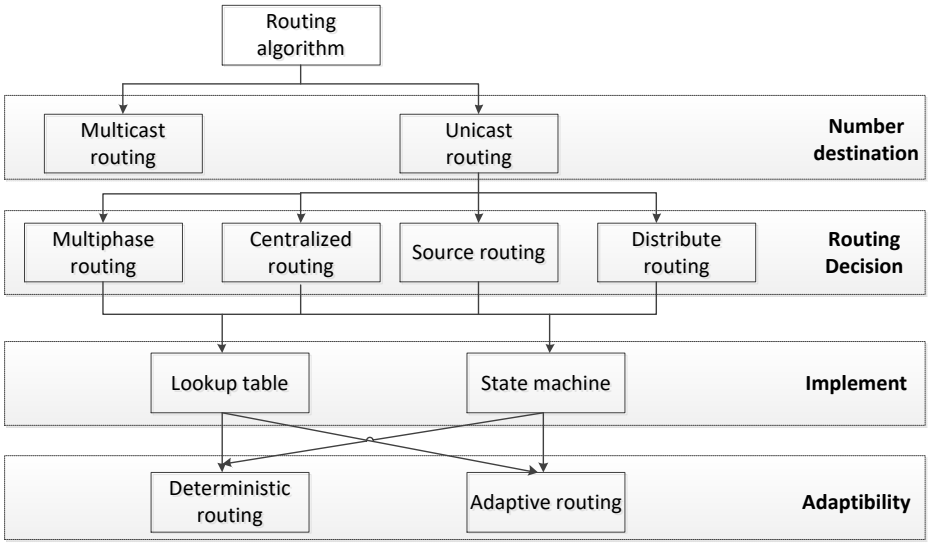

Figure 8. Routing algorithms [7].

3. ROUTER ARCHITECTURE

3.1. Router connection and REQ/ANS

According to presentation in the Section II, the proposed NoC also approaches the 2D-Mesh topology. In our work, a 4x4 configuration of 2D-Mesh NoC is generated to estimate the performance. After confirming the topology and corresponding configuration, the architecture of single router in NoC is illustrated. Firstly, the router connection and REQ/ANS protocol are introduced in Figure 9 below.

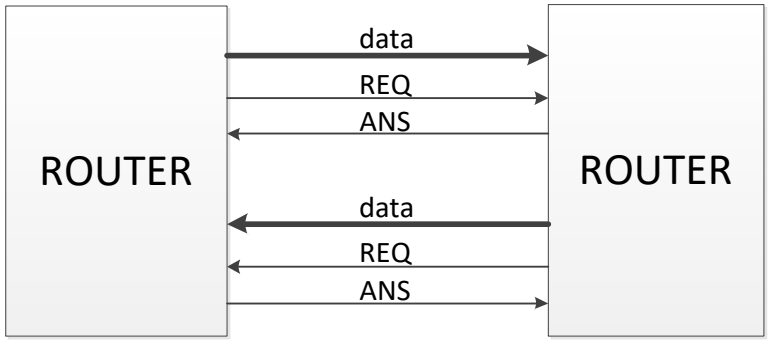

Figure 9. Proposed router interface.

Regarding to REQ/ANS protocol, the clock diagram as Figure 10 identifies how to establish a connection as well as delay of 2 clock cycles per each connection.

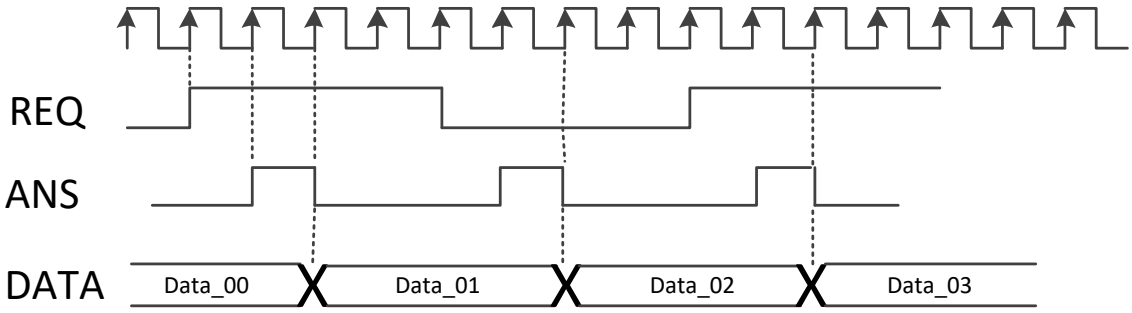

Figure 10. REQ/ANS protocol.

Based on the clock diagram, whenever data is ready in output buffer, REQ signal is active. One connection is start by changing logic of REQ signal. At the neighbor router nearly, the ANS pulse replies to notice the accepted connection. After receiving the active ANS signal, a flit is transfer between routers. As the result, every connection is corresponded to one flit. If all flits are finished transferring, the REQ signal is stop the inversion. Furthermore, this protocol is not only applied to setup connection between routers but also utilized as the main protocol in the router architecture to transfer data among input and output buffers.

3.2. Virtual channel and mask technique

As presented in previous section, the virtual channel number decides the buffer resource in a single router. Moreover, it also affects the interface of single router significantly. To be more specific, it is fact that a single router need more control signals for added virtual channels. Figure 11 shows the interface of a single router integrated two virtual channels.

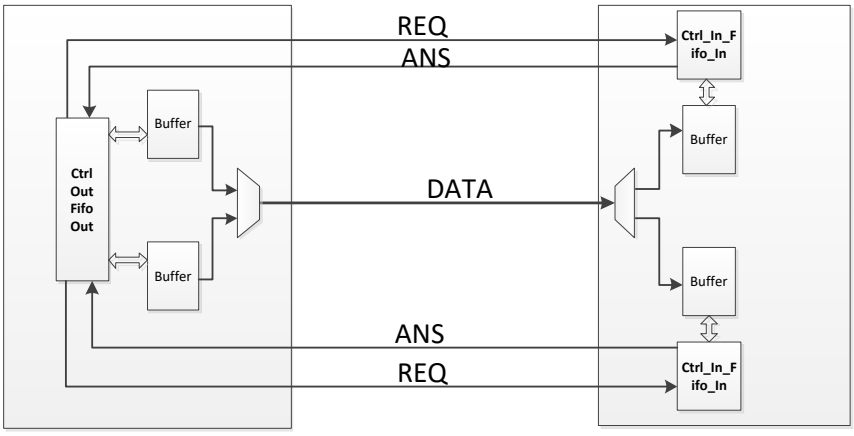

Figure 11. Interface of single router integrated two virtual channels.

Configurating two virtual channels requires two output buffers and two control signal groups comprising of REQ and ANS essentially. Similarly, it means that the neighbor routers also need two input buffers. In this proposed structure, data belonging to exact virtual channel will be stored in corresponding buffer. It can be explained in more detail that data on the first output buffer is only stored on the first input buffer.

Furthermore, increasing virtual channel number is similar to increase the number internal input control structure. As illustrated in Figure 11, two *Ctrl_FiFo_In* structures for two input buffers but only one *Ctrl_FiFo_Out* structure for output buffers are implemented. At the same time, if data is ready on both two output buffers, the data bus is shared by two channels respectively.

Basing on the virtual channel configuration, the bottleneck issue on output buffers is solved feasibly; however, in order to solve the problem in which many data flows from different input buffers share only one output buffer, the mask technique is proposed.

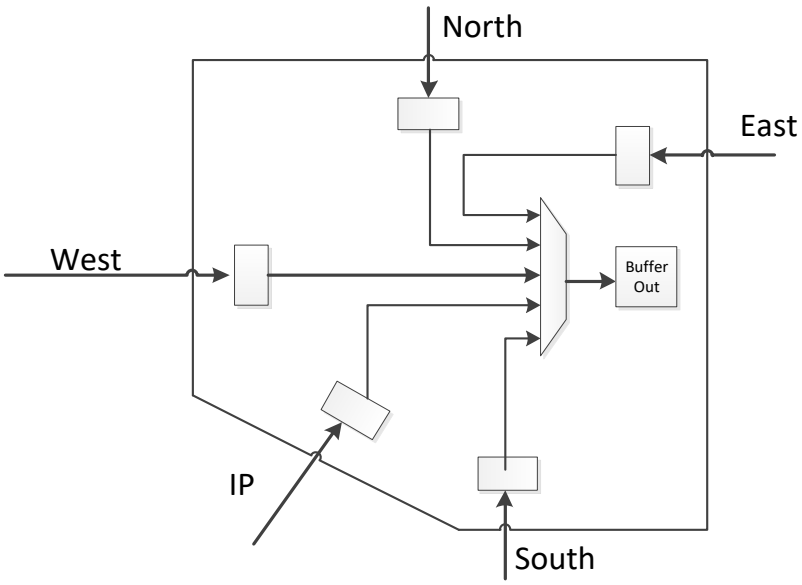

Figure 12. Data flows to one output buffer.

As shown in Figure 12, the highest traffic situation of proposed router is that there are five requests, maximum number of requests, to one direction. To be able to solve this issue, a mask register is implemented to catch the requirements. Following the fixed priority, the request with highest priority is serviced firstly. Then, the next requests are solved respectively. It is noticeable that the mask register is only updated new value until the router finishs solving the submitted previous requests. A request is corresponding to a data package consisting of many flits. In proposed architecture, perfectly serving a request means that a complete package is finished transferring between buffers in router or different routers. Consequently, the number of flit per data package affects to speech of mask service.

One of other factors particularly affects to the bottleneck problem is the depth of buffer and how to separate the data package to flit. While data structure related to proposed adaptive property is described in next phase, advantages of choosing buffer depth is analyzed. It is fact that not only does the large buffer negatively affect to the speech of mask service but it is also serious issue in using resource. However, a large buffer can store more flits and enhance the throughput of complete NoC. From such circumstances, the depth of buffer at eight in proposed router is implemented to balance the resource limitation and throughput strictly.

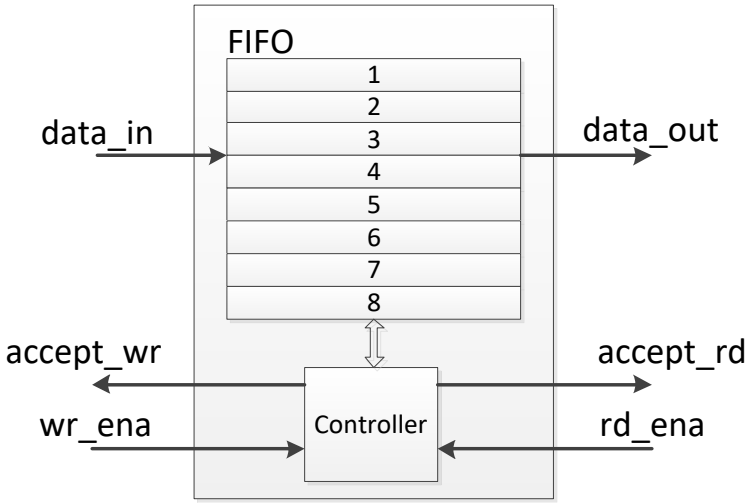

Figure 13. The buffer structure.

To be more specific, the buffer presented in Figure 13 includes two structures known as FIFO and controller. Controller structure receives *wr_ena* and *rd_ena* to control write and read processes respectively. The situation of FIFO is confirmed by *accept_rd* and *accept_wr* signals. The input and output buffers have similar structure.

3.3. Data structure and adaptive property

Before illustrating the adaptive property, the data structure of flit is shown firstly. Particularly, the flit shown in Figure 14 consists of 16 bits separated to two 8-bit groups called as 8-bit control and 8-bit data. 8-bit control is continuously divided to 4 sub-groups including 2 bits per group. The meaning of bit groups is classified in Table 1.

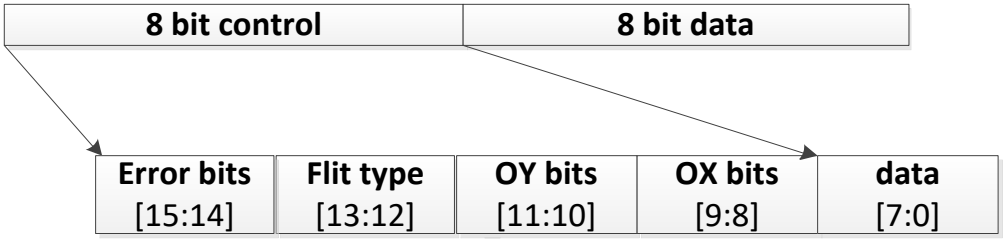

Figure 14. The data structure.

Table 1. Description of bits in flit.

Error bits [15:14]	Describe the error situation .00: No Error, OX is higher priority than OY .01: Not used .10: Avoid the error router, OX is higher priority than OY .11: Avoid the error router, OY is higher priority than OX
Flit type [13:12]	Describe the type of flit. One data package is divided to 8 flit comprising of three flit type .00: Not used .01: Tail flit .10: Body flit .11: Header flit
OY bits [11:10]	OY position in 4x4 configuration
OX bits [9:8]	OX position in 4x4 configuration
Data [7:0]	Data

In this research, the XY routing, one of deterministic routing algorithms, is approached. According to XY routing algorithm, the flit is transferred on OX axes firstly and continuously followed OY direction. In this paper, a proposed adaptive routing following XY routing algorithm is applied to halt one fault tolerance on any router in NoC after fabrication. In order to perform the adaptive routing algorithm, the position of error router needs to be recognized integrally. It can be detected feasibly by forwarding data on OX and OY respectively as Figure 15.

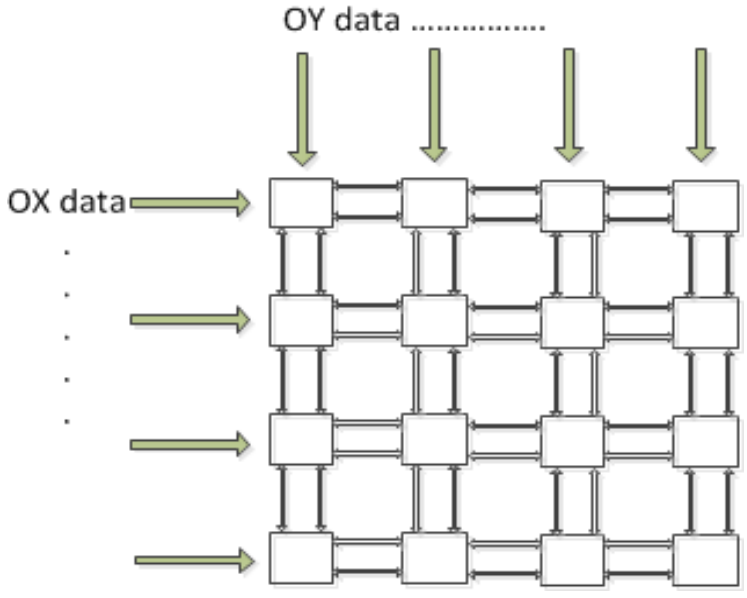

Figure 15. Error detection method.

After identifying the error router position, 4 error bits are setup and help all single routers know that the neighbor router is error or not. As regards XY routing algorithm, OX direction is higher priority than OY direction; however, this rule can be inversed to void the error neighbor router.

To be more detail, Figure 16 illustrates all cases to avoid the error router. Changing direction is not always occurred whenever detecting to error neighbor router. In some cases, OX direction is still kept higher priority. In order to understand the priority at every router, it is based on the first two bits of data [15:14] as description of Table 1.

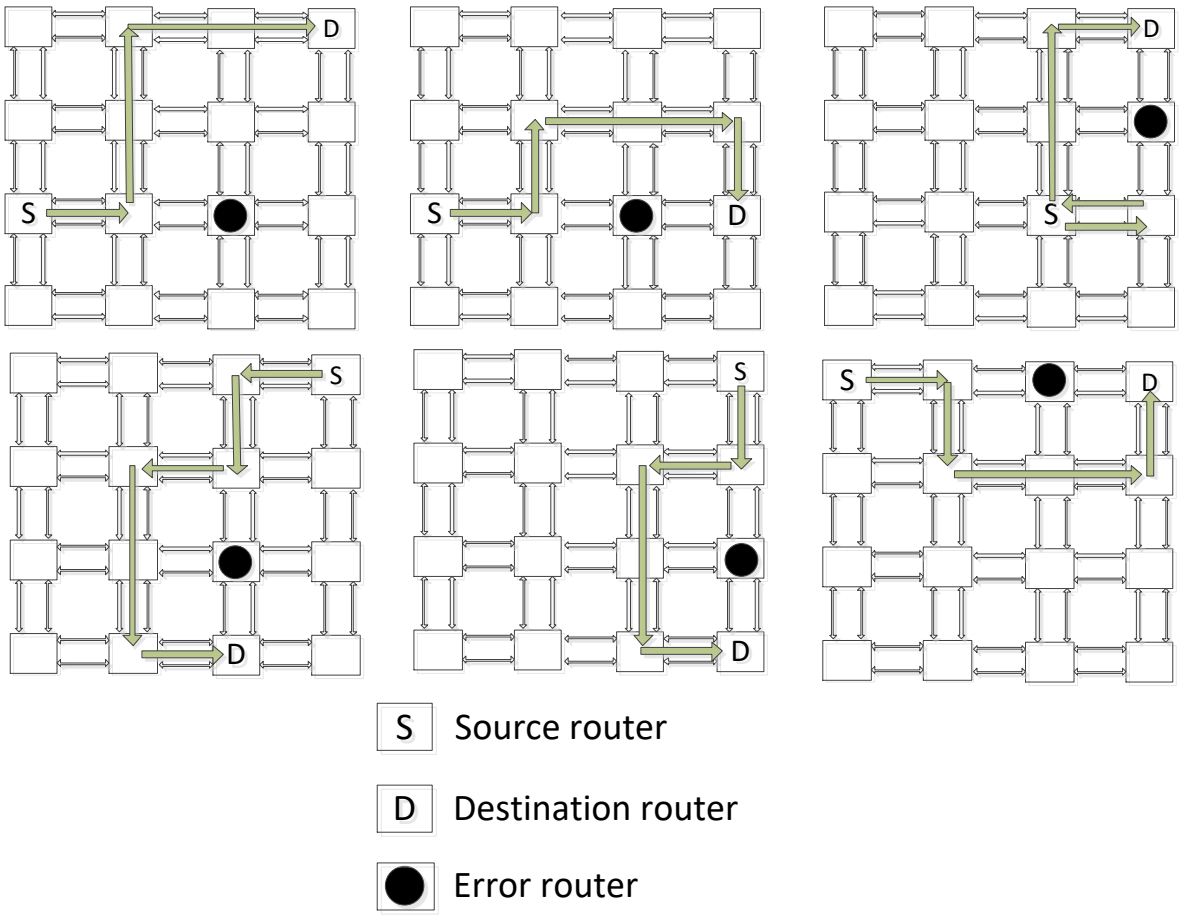

Figure 16. The adaptive routing after fabrication.

From such circumstances, the proposed router architecture implemented and evaluated in this work can be collected to Table 2 below.

Table 2. Router architecture property.

NoC configuration	4x4 2D-Mesh
Virtual channel number	2
Buffer depth	8
Data structure (18 bit)	. 8 bit for control . 8 bit for data
Connection Protocol	REQ/ANS
Internal routing technique	Dynamic XY
Ability of fixing bug after fabrication	One bug

According to proposed router architecture, the control bits have to be concerned whenever increasing the NoC configuration; however, the throughput is enhanced feasibly by expanding the data bits from 8 bits to 16 bit or 32 bits. The combination between virtual channel and mask techniques can solve the bottleneck issue. Moreover, the low depth of buffer also promotes the speech of solving mask cycle. In detail, Figure 17 describes internal connections of router generally.

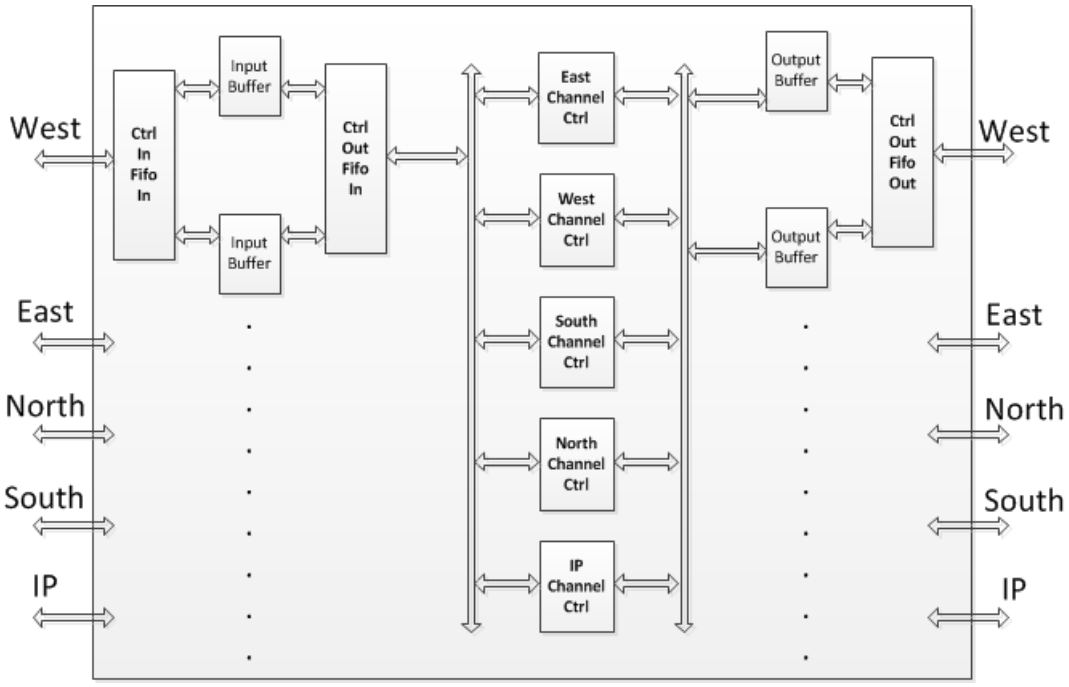

Figure 17. Internal structure of proposed router.

4. EXPERIMENTS

In order to verify the 4x4 configuration as well as others, an automatic tool is developed by Perl language to generate any configuration to estimate the exact performance. In this work, a 4x4 configuration of routers is generated and experimented. Because this research only develops the router architecture, data package generators are required. To tackle this problem, adaptive input and output structure models taking the role of Network Interface functions are developed to generate flits to router as Figure 18.

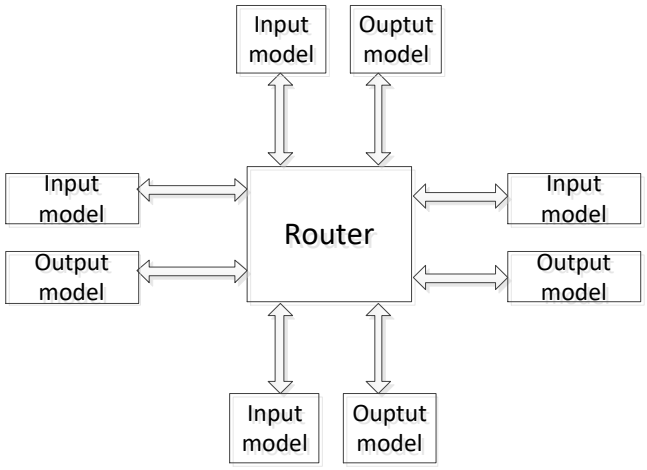

Figure 18. Router test environment.

Table 3. 4x4 2D-MESH NoC configuration verification.

NoC configuration	4x4
Package Number	1000
Flit Number	8
Bit number in Flit	16
Data bit in Flit	8
Maximum Frequency	500Mhz
The first flit	After 171 cycles
The final flit	After 7171 cycles
Time Average/ Packet	142 ns
Bandwidth	0.56GBs

Likewise, both 4x4 NOC configuration and input/output models are also generated automatically. The verifications of sharing data flow on virtual channel or mask behavior are

configured by the data in input/output models. 1000 data packages including 8000 flits are generated by input models. At the output models, the first flit and final flit are obtained to calculate the delay time average as well as bandwidth. Next, both single router and completed 4x4 NoC configuration are synthesized on 90nm technology following ASIC base design to estimate used resource, maximum frequency as well as power consumption. To be more specific, Table 4 describe the final results of physical steps.

Table 4. Performances of single router and 4x4 NoC configuration.

	4x4 NoC configuration	Single Router
Maximum Frequency (MHz)	500	500
Total Power (mW)	569.2902	53.6717
Area (NAND gate number)	506386	42335

5. CONCLUSION

In this paper, one feasible 2D-Mesh NoC architecture is presented. Basing on small buffers and short package, the bottleneck issue is solved by mask and virtual channel techniques obviously. Moreover, the adaptive routing to tackle one fault-tolerant issue after fabrication is also implemented to enhance the productivity. In proposed architecture, the bandwidth is not affected by fault-tolerant routing algorithm. Actually, the bandwidth in error and normal situations are similar. Besides, the proposed structure also confirms that deadlock issue is not occurred if there is only one connection fault because all techniques are implemented and integrated in router. Consequently, network expanding and reused ability is feasible obviously.

However, this architecture has experienced disadvantages as expanding control bits in flit for enormous NoC configurations. Furthermore, the internal complicate control is also an issue in improving the throughput. As the result, improving the control structures and developing Network Interface to match general interface standards are our future works.

REFERENCES

1. Umit Y. Ogras, Jingcao Hu, Radu Marculescu - Key Research Problems in NoC Design: A Holistic Perspective, IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, 2005, pp. 69-78.
2. Masoud Daneshtalab - Exploring Adaptive Implementation of On-Chip Networks, ISBN 978-951-29-4787-4, 143p, Department of Information Technology, University of Turku, Finland 2011.
3. Shahram Jamali, Fateme Sarkohakiand, Reza Fotohi - Providing a fault-tolerant algorithm for on-chip interconnection networks, International Research Journal of Applied and Basic Sciences, Vol.4,2003, pp1700-1705.
4. Kia, H.S, Ababei, .C - A new Fault-tolerant and Congestion-aware Adaptive Routing Algorithm for Regular Networks-on-Chip, Congress on Evolutionary Computation (CEC), 2011, pp. 2465-2472.
5. Mak T., Cheung P.Y.K., Kai-Pui Lam, Luk W. - Adaptive Routing in Network-on-Chips Using a Dynamic-Programming Network , IEEE transactions on Industrial Electronics, Vol.58, 2011, pp. 3701-3716.
6. Grot B., Hestness, J. ; Keckler, S.W. ; Mutlu, O. - Express Cube Topologies for On-chip Intercornnects, High Performance Computer Architecture (HPCA), 2009, pp. 163-174.
7. Ankur Agarwa, Boca Raton, Cyril Iskander, Ravi Shankar - Survey of Network on Chip (NoC) Architecture & Contributions, Engineering, Computing and Architecture ISSN, 2009.
8. Shaoteng Liu, Jantsch A., Zhonghai Lu - Analysis and Evaluation of Circuit Switched NoC and Packet Switched NoC, Digital System Design (DSD), 2013, pp. 21-28.
9. Shubhangi D Chawade, Mahendra A Gaikwad, Rajendra M Patrikar - Review of XY Routing Algorithm for Network-on-Chip Architecture, International Journal of Computer Applications 21 (2012) 20-23.