Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 3, March 2014, pg.462 – 469

RESEARCH ARTICLE

Design of Router Micro Architecture Based on Runtime Adaptive Selection Strategies for On-Chip Communication Interconnection Network

[1]Ms. Aryadevi S, [2]Mr. T Shanmuganathan

[1]PG Scholar, Department of Electronics and Communication Engineering,
Hindustan University, Chennai, Tamil Nadu, India

[2]Assistant Professor, Department of Electronics and Communication Engineering,
Hindustan University, Chennai, Tamil Nadu, India

[1] arya.punartham@gmail.com; [2] thangatamizh@gmail.com

Abstract—To meet the growing computation-intensive applications and the needs of low-power, high-performance systems, and the number of computing resources in single-chip has enormously increased, because current VLSI technology can support such an extensive integration of transistors. This paper presents adaptive routing selection strategies suitable for network-on-chip (NoC). The main prototype presented in this paper use west first routing algorithm to make routing decision at runtime during application execution time. Messages in the NoC are switched with a wormhole cut-through switching method, where different messages can be interleaved at flit-level in the same communication link without using virtual channels. Hence, the head-of-line blocking problem can be solved effectively and efficiently.

Keywords— Network on Chip; Router; Adaptive Routing; VLSI; West First Routing

1. INTRODUCTION

For many-core processor systems, Network-on-Chip (NoC) is a feasible communication infrastructure because of the scalable bandwidth capacity of the NoC. There are many research challenges in the field of many-core processor systems starting from abstract application layer until physical network layer. In the network layer, optimum network and router architecture design in terms of cost (logic area, power, etc.) as well as its performance issues (network bandwidth capacity, router latency, etc.) are the challenging topics. In addition, many current System-on-Chips (SOCs) use a system bus to connect several functional units. SOC technologies are the packaging of all the necessary electronic circuits and parts for a "system" on a single integrated circuit, generally known as a microchip.

However, these SOC system buses can support only limited number of functional units, and thus will face scaling problems in heterogeneous MPSOCs or large scale CMPs. In order to solve these long global wire delay and scalability issues,

many studies suggested the use of a packet based communication network which is known as Network-on-Chip (NoC). There are three main building blocks of the NoC. The first and most important one are the links that physically connect the nodes and actually implement the communication. The second block is the router, which implements the communication protocol. The last building block is the network adapter (NA) or network interface (NI).This block make a logic connection between the IP cores and the network, since each IP may have a distinct interface protocol with respect to the network.

The, routing major impact of the routing algorithm would be on area and the network performance. The routing algorithm, generally can be made in deterministic (static) or adaptive manner. Network designers are motivated to design adaptive routing algorithms because of two main objectives, i.e., to avoid entering hotspot links such that communication performance can be increased, and to avoid entering faulty network components. The adaptive routing algorithms have the main issue of deadlock configuration problem due to cyclic dependency. Turn models can be principally used to design a deadlock-free adaptive routing algorithm. Most of routing implementations made at design time use routing tables to route messages (packets). The contents of the routing tables are programmed at design time, and then adaptive routing paths are assigned in every routing table in the network nodes by using some technique.

Most of routing implementations made at design time using routing tables to route messages (packets). The contents of the routing tables are programmed at design time, and then adaptive routing paths are assigned in every routing table in the network nodes by using some technique. An offline routing method called "Application Specific Routing Algorithm" is used to increase the degree of routing adaptivity for hotspot avoidance. The "Segment-based Routing" (SR) which is also an offline routing method, in which the network is segmented into some subnets and restrictions are applied to avoid deadlock configurations. Another method is the dynamic routing protocol is used for balancing distribution of traffic in NoCs.

In lookup-table-based routing algorithms, the size of the tables will increase as the network size increases, since all entries must be added in the tables. So, different techniques were used to reduce the size of the routing tables. A region-based routing algorithm aimed at reducing the size of routing tables for NoCs by grouping destination network into network regions. The proposed methodology can be implicitly viewed also as a technique to design a router using West First Algorithm in 2D mesh . Our methodology can be classified into runtime distributed routing approach, where the routing is made locally in every NoC router at runtime during application execution time.

2. ROUTING ALGORITHMS AND SELECTION STRATEGIES

The routing algorithm determines the path that each packet follows between a source-destination pair. Routing is deterministic if only one path is provided, or adaptive if several paths are available and dynamically selected at switches. Generally, an adaptive routing algorithm is preferable since it has the potential of achieving higher performance (low latency, high throughput, and fault tolerance). Routing strategies can also be classified as source or distributed. In source routing, the source node stores the entire path in the packet header. Since the header itself must be transmitted through the network, it reduces the effective network bandwidth (especially with short packets). In distributed routing, the packet header contains the address of the destination. Each switch computes the next link that will be used while the packet travels across the network. Distributed routing is preferred for NoCs as the packet header is reduced and more flexibility and adaptivity can be achieved. Distributed routing can be implemented using two distinct methods.

The first method is called algorithmic routing and is suitable for on-chip and off-chip networks with regular topologies. In this method, a finite-state machine (FSM) is used for computing the routing option as a function of the current and destination node addresses along with possibly other information like status of output ports, network traffic, packet priority, etc. The implementation is very efficient in both area and speed, but it is topology and routing algorithm dependent. The second method uses a table which stores the output ports that should be used for each destination end node. The main advantage of this approach is that the same generic design can be reused for implementing any network topology and routing algorithm. It is also much easier to incorporate fault tolerance with this approach as compared to the algorithmic approach.

3. EXISTING METHOD

In the existing method, we are using a wormhole cut-through switching method, where messages are interleaved at the flit level rather than at packet level. Hence, single message can be associated as single packet, which consists of single header flits, payload data flits and single tail flit header and tail flit, and it is not divided into packets. A single flit has 39-bit width, 32 bits for dataword plus 9 extra bits, i.e., 3-bit field to define the type of flits and 4-bit field to determine the local identity label or ID-tag k of a message. Fig. 1 shows a 2D mesh-planar topology, where the NoC is divided into two subnets, i.e., X^+ (increment) subnetwork depicted in solid lines and X^- (decrement) subnetwork depicted in dashed lines. If a target node offset of a packet is $x_{offset} = x_{target} - x_{source} \geq 0$, then the packet will be routed through the X^+ subnetwork, while if its target node offset is $x_{offset} \leq 0$, then it will be routed through the X^- subnetwork. Once a packet is routed to a subnetwork, it will not move to another subnet. By using such routing rule, the minimal planar adaptive routing algorithm will be free from a cyclic dependency.

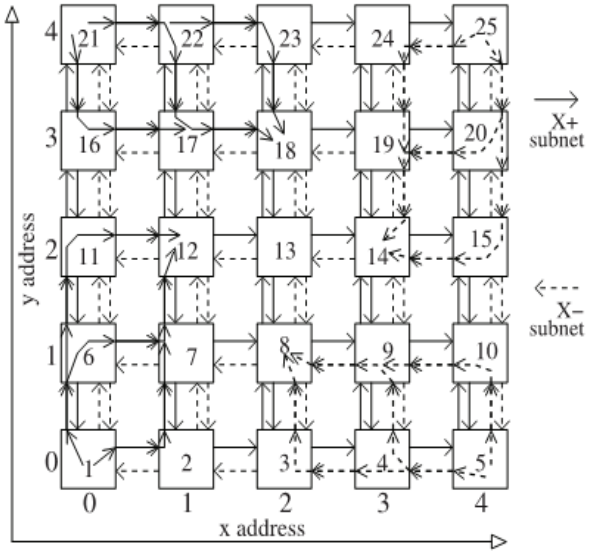

Figure 1: Possible Adaptive Routing

3.1 ADAPTIVE ROUTING SELECTION FUNCTIONS

Five router implementations based on information that are considered to make routing decision and based on the view point of our NoC microarchitecture will be presented. The three considered information are described in the following:

- Identity (ID) slot occupancy (the number of free ID slots). This information can be called also as Contention Information of an output port, i.e., the number of messages that have competed so far to access the output port. Since our router can interleave different wormhole messages at flit-level in the same link without using VCs, then the number of reserved ID slots will represent the number of the wormhole messages that have been mixed in the outgoing link.
- BW space occupancy (the number of free BW space). This information can be called also as BW-Reservation Information of an output port, i.e., the number of BW spaces that have been reserved by messages to access the output port.
- Buffer space occupancy (the number of data queue in a FIFO buffer). This information can be called also as CI of an output port, i.e., the queue length in the FIFO buffer at the input port of the next neighbour switch connected directly to the output port.

In Fig. 2.a, congestion aware routing is shown, where messages are routed to an output direction having less utilized buffer spaces. The BW-ID prototype that uses two information signals to make routing decisions. The first prioritized signal is the number of the reserved bandwidth spaces, and the second one is the number of used ID slots (ID slot occupancy). This adaptive routing strategy can be called as a Contention and BWA Adaptive Routing Selection Strategy, is shown in Fig.2.b. Messages are routed to an output direction having less reserved bandwidth spaces. If the numbers of the reserved BW spaces between two output ports are equal, then the second prioritized signal is used, i.e., the number of reserved ID slots. When the numbers of the reserved BW spaces between the alternative output ports are equal, the messages are then routed to an output direction having less reserved ID slots.

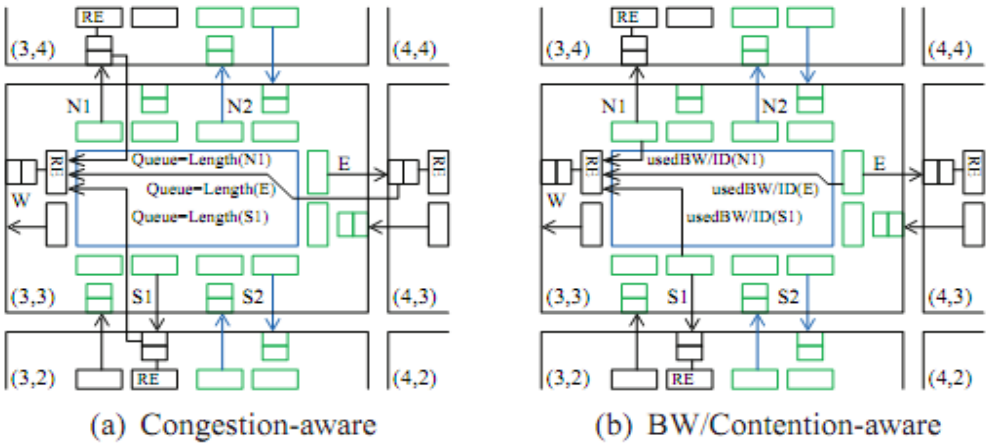

Figure 2: Alternative information that can be used to make adaptive output routing selection.

3.2 ROUTING ARCHITECTUTRE

The microarchitectures of the NoC router that uses the CBWA adaptive routing and the CCA adaptive routing selection strategies are presented in Fig. 5. For simplicity, only the router components in East input port and in West output port are

depicted. The router is designed based on a 2D mesh-planar topology, where each router has seven IO ports, i.e., East, North1, North2, West, South1, South2 and Local ports. Crossbar interconnect is customized to optimize the logic area of the router based on the allowed turns in the 2D planar adaptive routing algorithm. The rest router internal IO connections representing the prohibited turns are removed from the architecture.

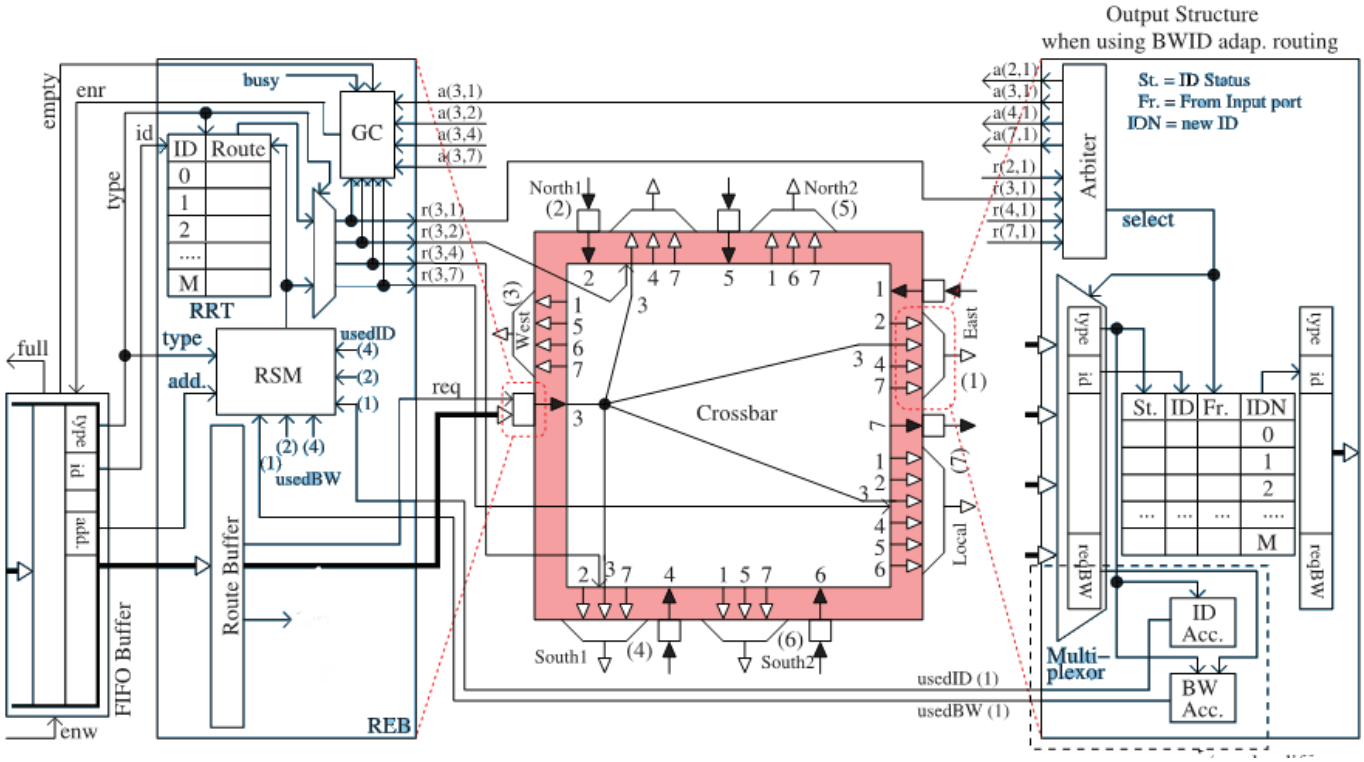

Figure 3: Router Architecture

Set of components at each input and output ports n is the FIFO buffer, the Routing Engine with Data Buffering (REB), the Multiplexor with ID Management unit (MIM) and the Arbiter unit. Based on the crossbar interconnects shown in Fig. 5, the each port name is assigned to a port number as follows: East (1), North (2), West (3), South (4), North2 (5), South2 (6), and Local (7).Set of subcomponents in the REB module at output port n are the Routing State Machine (RSM), the Route Buffer, the RRT and Grant Controller (GC). In the REB unit, the combination of the RSM, in which the planar adaptive routing algorithm is implemented, and the RRT is implemented to support runtime adaptive routing mechanism. The GC unit is used to control the read operation of the FIFO Buffer, and the Route Buffer is used to store data that will be routed to an output port.

4. PROPOSED METHOD

In the proposed method, deadlock free West first routing algorithm is applied to a 2D mesh topology and a wormhole switching method is used along with data link layer flow control. The proposed router is shown in figure. Router is having five ports east, west, north, south and local port. Each port is having its input and output channel, and each input and output channel is having its control and decoding logic, which supports five parallel connections at the same time simultaneously. The input

channel consists of three parts i.e. FIFO, FSM, and West First routing logic. The FIFO here is used as input buffer to store the data temporarily. It is of 8 bits size and is of depth 16 bits. The first 8 bits of FIFO is header containing the coordinates of destination port, thus the size of packet varies from 8 to 120 bits. FSM is used to control the read and write operation of FIFO, according to status. If FIFO is empty, FSM generates the signal to perform the write operation and if FIFO is full, FSM generates signal to perform the read operation.

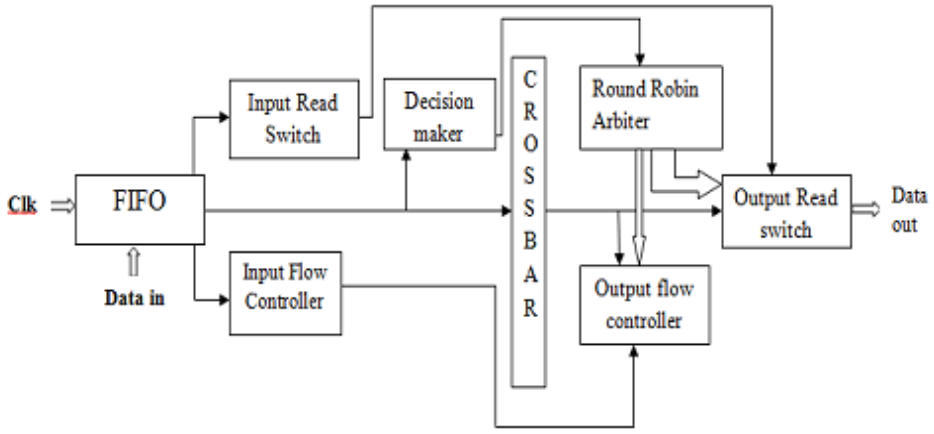

Figure 4: Router for Proposed Method

West First Routing logic is the adaptive minimal logic which analyses the header of packet to find out its destination address. West First logic compares the coordinates values stored in header with the locally stored coordinates values and send the data to its destination port. Similarly the output channel consists of three parts i.e. FIFO, FSM and arbiter. The FIFO here is used as output buffer to store the data temporarily. It is of 8 bits size and is of depth 16 bits the 8 bits are the header, thus the packet size varies from 8 bits to 120 bits. FSM is used to control the read and write operation of FIFO according to its status. If FIFO is empty, FSM generates signal to perform write operation and if FIFO is full FSM generates signal to perform read operation. Arbiter is used in output channel to overcome the problem of multiple input requests coming at single output port. Arbiter is based on rotating priority scheme in which each port get reduced its priority once it has been served.

The mesh based router consists of multi ports such as east, west, north, south and local port. It also has a central cross point matrix. Inside each port there are two channels input and output. Data packet is sent from one port moves in to the input channel of router by which it is forwarded to the output channel of the other port. Each input channel and output channel has its own decoding logic which increases the performance of the router. Buffers are used at all ports to store the data for a short time span. The store and forward method is used here for data transmission. Control logic is present to make decisions to grant access to a port request. In this way communication is established between input and output ports. The transfer of data from source to destination is called packet switching mechanism where the flit size is 8 bits. In router structural modelling is used in which input channel of all 5 ports, output channel of all 5 ports and cross point matrix is used as component to form complete router. The input signals here are five data input signals; five requests signals, and five acknowledgement signals. The output signals are five data output signals, five acknowledgement signals and five request signals. The port mapping of each component is done to connect all.

5. RESULT AND OBSERVATION

Two router architectures are implemented in Verilog and then synthesized in a Xilinx 13.1. Router design is simulated in ISIM 13.1. Synthesis process converts user's hardware description into structural logic description. Figure 7 represents the synthesis of input channel using west first algorithm, Figure 8 represents the synthesis of output channel.

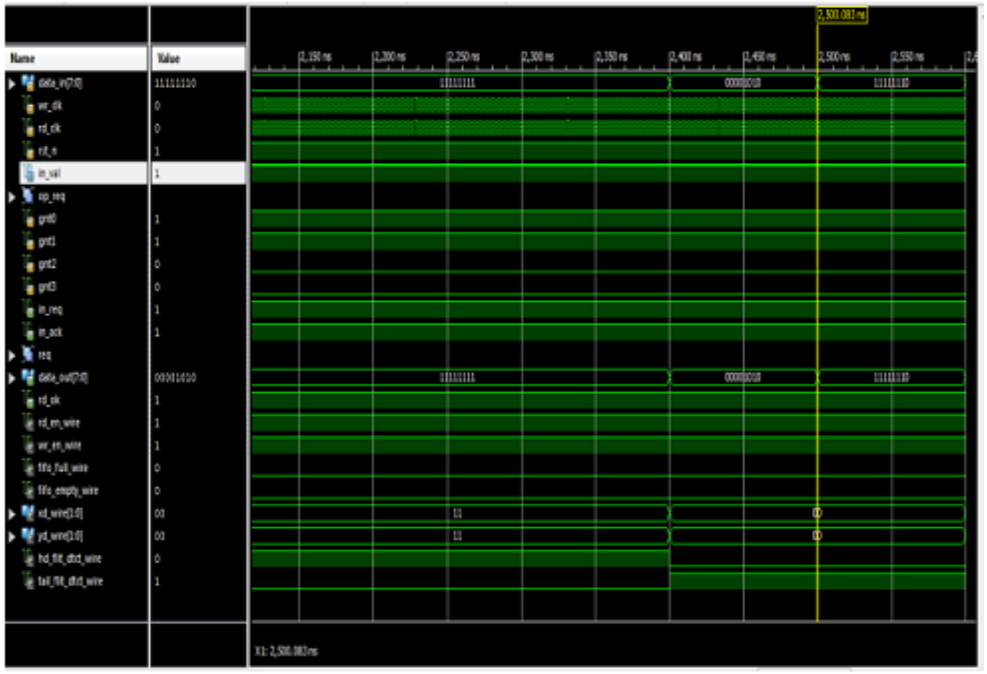

Figure 6: Simulation waveform of input channel

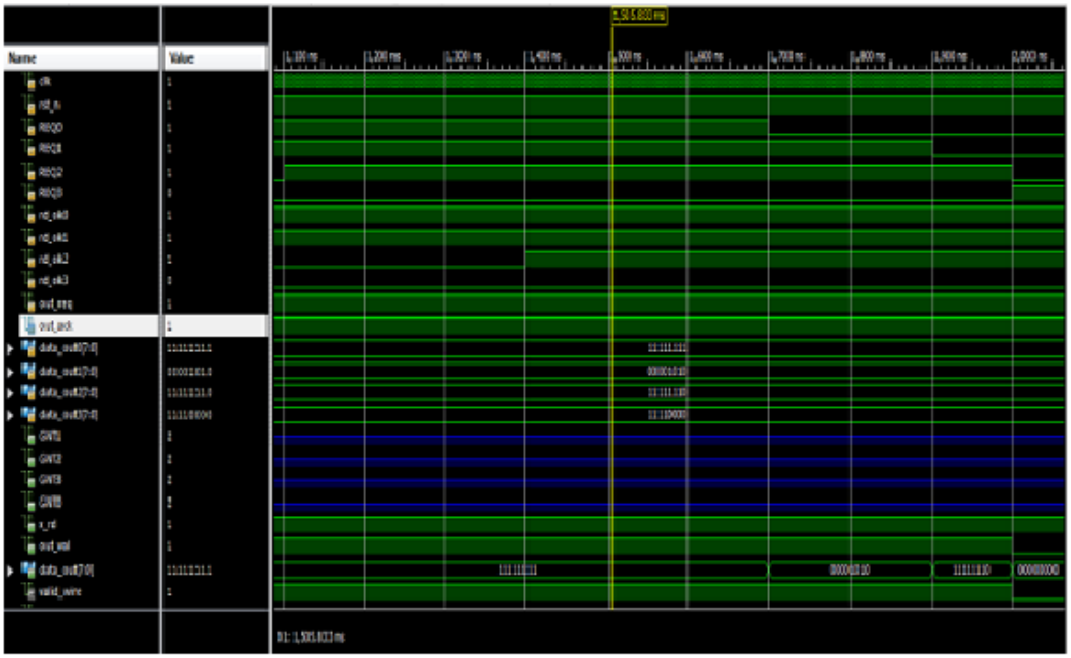

Figure 5: Simulation waveform of output channel

6. CONCLUSION

In this paper, router architecture using west first algorithm is implemented using HDL called Verilog at RTL level. Architecture is synthesized using Xilinx and simulated using ISIM 13.3 for evaluating latency and delay of two routers. The simulation results show that there is reduced latency when using west first algorithm. Network on chip presents a most adapted technology to perform communication in complex SoCs. In this work, we present a west first routing algorithm related router which offers a low latency 1 clock cycles and high speed 216 MHz communication for on chip modules. This architecture offers a variety of SoC communication services owing to its flexibility and adaptability.

REFERENCES

[1] J. Duato, "A New Theory of Deadlock-Free Adaptive Routing in Wormhole Networks," IEEE Trans. Parallel and Distributed Systems, vol. 4, no. 12, pp. 1320-1331, Dec. 1993.

[2] T.T. Ye, L.Benini, and G. De Micheli, "Packetization and Routing Analysis of On-Chip Multiprocessor Networks," J.System Architecture, vol. 50, nos. 2/3, pp. 81-104, 2004.

[3] M.K.F. Schafer, T.Hollstein, H.Zimmer, and M. Glesner, "Deadlock-Free Routing and Component Placement for Irregular Mesh-based Network-on-Chip," Proc. IEEE/ACM Int'l Conf.Computer-Aided Design (ICCAD '05), pp. 238-245, 2005.

[4] M.Li, Q.A.Zeng, and W.-B.Jone, "DyXY: A Proximity Congestion-Aware Deadlock-Free Dynamic Routing Method for Network on Chip," Proc. 43rd Ann. Design Automation Conf. (DAC '06), pp. 849-852, 2006.

[5] G. Ascia, V. Catania, M. Palesi, and D. Patti, "*Implementation and Analysis of a New Selection Strategy for Adaptive Routing in NoC*", IEEE Trans. Computers, vol. 57, no. 6, pp. 809-820, June 2008.

[6] M. Palesi, R. Holsmark, S. Kumar, and V. Catania, "Application Specific Routing Algorithms for Networks on Chip," IEEE Trans. Parallel and Distributed Systems, vol. 20, no. 3, pp. 316-330, Mar. 2009.

[7] F.A. Samman, T.Hollstein, and M.Glesner, "*Adaptive and Deadlock-Free Tree-Based Multicast Routing for Networks-on-Chip*," IEEE Trans. Very Large Scale Integration Systems, vol. 18,no. 7, pp. 1067-1080, July 2010.

[8] F.A. Samman, T.Hollstein and M.Glesner, "Runtime Contention and Bandwidth -Aware Adaptive Routing Selection Strategies for Network –on- Chip," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 7, July 2013

