Hybrid NoC with Traffic Monitoring and Adaptive Routing for Future 3D Integrated Chips

Ville Rantala[1], Teijo Lehtonen[1,2], Pasi Liljeberg[1], Juha Plosila[1,3]
[1]University of Turku, Department of Information Technology, Joukahaisenkatu 3-5 B, FIN-20520 Turku, Finland
[2]Turku Centre for Computer Science (TUCS)
[3]Academy of Finland, Research Council for Natural Sciences and Engineering
Email: ville.rantala@utu.fi, teijo.lehtonen@utu.fi, pasi.liljeberg@utu.fi, juha.plosila@utu.fi

***Abstract*—Future high-performance 3D integrated chips require a reliable and efficient communication structure for internal communication. It needs to be able to move data between devices on the chip efficiently without limiting the performance of the system. An on-chip communication structure for near-future 3D integrated chips is proposed in this paper. It is based on the Network-on-Chip paradigm and it utilizes an adaptive routing algorithm supported by a distributed network monitoring service which spreads the traffic information over the network. The traffic information diffuses over the network so that locations of the congestion spots can be predicted. The presented hybrid mesh-ring NoC structure improves the performance of the network by balancing the load of the communication resources efficiently.**

I. Introduction

While the complexity of integrated chips increases, the challenge to organize an efficient communication between parts of the system becomes more and more essential. A new emerging technology for future high-performance integrated chips is 3D integration where multiple chip layers containing active devices are vertically stacked. In these systems there is a need for efficient communication between devices on a same layer as well as between devices on different layers. [1] Continuous increase in the number and performance of the processing units causes a situation where on-chip communication becomes a bottleneck of the system and limits its overall performance.

Obtaining high performance of a large 3D integrated system requires careful design of the communication platform as well as efficient utilization of available resources. The balanced resource utilization requires methods to monitor the system and adapt the operation according to the state of the system. A 3D integrated system can include a large number of processing devices which might consume a lot power. Because of the complex stacked structure of a 3D integrated chip, one have to pay attention to the heat management. The stacked structure of a 3D integrated system makes the thermal management especially challenging. Therefore, the energy consumption of the system has to be as low as possible. The heat production can also centralize around the most used processing elements. These hot spots can also hamper the operation of the system. [2] A method to balance heat production and avoid hot spots is to use the computational and communication resources of the system as balanced and as efficiently as possible without for instance increasing the clock frequency of the system. This is an essential subject on the hybrid mesh-ring NoC.

The paper is organized as follows. In Section II the structure of the implemented hybrid mesh-ring Network-on-Chip is presented. Routing and monitoring on the implemented system are discussed in Sections III and IV. The simulation results are presented in Section V and finally the conclusions are drawn in Section VI.

II. Structure of Hybrid Mesh-Ring NoC

The hybrid mesh-ring NoC is a communication structure for near-future 3D integrated chips. It consists of local mesh networks and a global ring network which connects the local networks, or clusters, together (see Fig. 1). The hybrid mesh-ring NoC contains four clusters each of which includes 16 routers and resources. The structure is designed for 3D integrated systems so that the clusters can be located on the different layers of the chip, and the global ring network operates as the inter-layer network. Local mesh networks and the global ring network are scalable in a way that the number of clusters as well as the number of resources on a cluster can be increased. It has to be noted that a significant insertion of new clusters can cause unbearable increase to delays between different clusters, because the inter-cluster traffic flows through other clusters in the ring network.

The hybrid network includes a monitoring system which is used to balance the network load over the communication resources. The components of the monitoring system observe traffic in the network and deliver traffic information to the routers. The modular structure of the monitoring system enables possibility to replace the monitor components if necessary. The current implementation receives router load information from the local router and traffic level information from the monitoring devices attached to the neighboring routers. The implemented system includes two types of monitoring components, one for monitoring on the mesh networks and other for the ring network.

The global ring network has two channels so that the packets can be transferred in both directions. The number of clusters as well as the size of them can be easily increased, which enables scaling of the communication structure for different size 3D systems. The routers in the two opposite corners of each cluster operate as routers of the mesh network as well

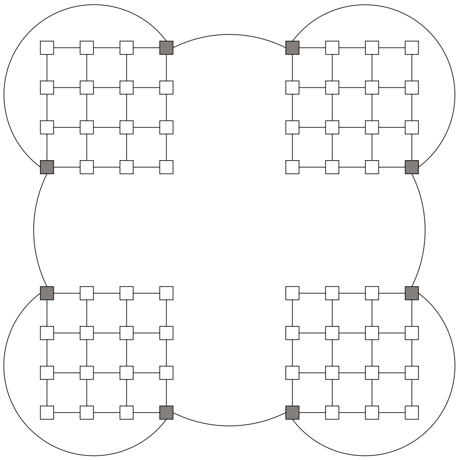

Fig. 1. Hybrid mesh-ring topology including four cluster and 16 routers on each cluster. Global routers are colored gray.

as routers of the ring network (see Fig. 1). The links between these global routers in a certain cluster can also be used to route packets inside the cluster.

The structure of a router is presented in Fig. 2. The router consists of a routing module, two buffer elements (I/O in Fig. 2) on each channel and a monitoring module. Besides these the global routers, which are connected to both local mesh and global ring networks, include a monitoring module for the ring network (gray part in Fig. 2). The 5-channel router is designed to operate in a mesh network, which is the topology of the clusters on the implemented system. The routing module does arbitration, routing decisions and packet directing. On each channel, which are called as channels north, south, east, west and local, there are two buffer elements, which receive packets and send them forward. The routing module is modular in a way that the arbitration method and routing algorithm can be replaced if necessary. Three different routing modules were implemented for the simulation purposes. The primary router design includes an adaptive routing algorithm which uses the traffic level information, produced by the monitoring modules, to balance network load and to avoid congestion. The routing module in this implementation does not contain any local buffer memory, so it moves the packets directly from the input buffers to the output. The two other router implementations are used as reference systems. Both of the reference systems use deterministic routing algorithm without utilization of the monitoring information. One of these deterministic routers moves packets directly from input buffers to output while the other stores packets on its local buffer memory during the routing phase. The routing is discussed more detailed in Section III.

On near-future 3D integrated chips the amount of inter-layer links or die-to-die vias is constrained [3]. On the hybrid mesh-ring NoC the limited number of these vertical vias is taken into account in design of the global links. The structure of the hybrid NoC allows distribution of the global links on different sides of the chip instead of locating them close to each other. Distribution of the global links could be a necessary practice due to the physical structure of 3D integrated chips.

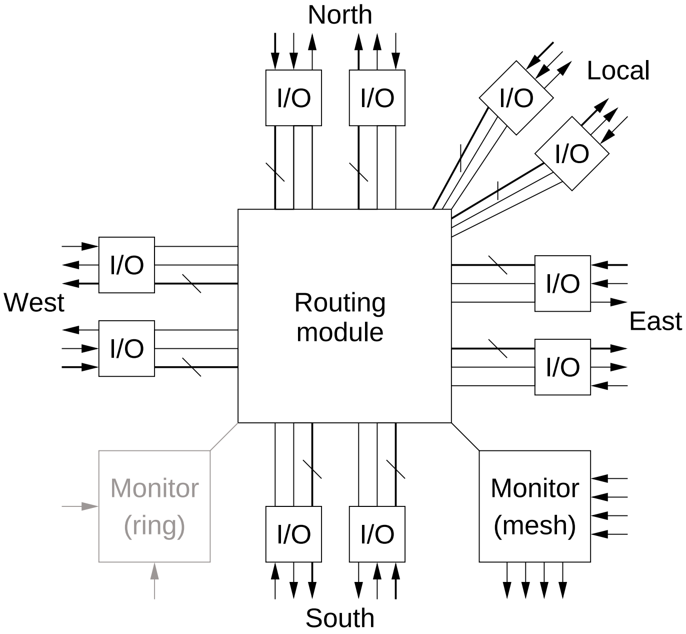

Fig. 2. Structure of the router.

III. Routing

The hybrid mesh-ring NoC uses a dynamically adaptive XY routing on the mesh networks. The algorithm routes packets based on the X and Y coordinates of the mesh. At each router the algorithm decides if the packet is going to be routed towards the X or Y direction based on the information of congestion in different directions. [4]. The adaptive XY algorithm uses the traffic level information, provided by the monitoring modules, for choosing the least congested productive routing direction. Productive routes are routes which lead towards the destination [5].

The traditional deterministic XY routing algorithm prevents cyclic dependencies between different routing dimensions or the X and Y dimensions. The deadlocks can still occur within a dimension [5]. A routing path of the dynamically adaptive XY routing algorithm can be divided to shorter parts, each of corresponding to a path of the deterministic XY routing. The path is divided so that a new deterministic path starts every time when the route turns from the Y direction to the X direction. Based on the properties of the deterministic XY routing algorithm, each of these short paths is deadlock free between the dimensions. Therefore, it is assumed that cyclic dependencies between routing dimensions do not occur along the whole routing path. The adaptive routing algorithm utilizes a deadlock avoidance (DA) method which prevents the occurrence of deadlocks within the routing dimensions. Typically deadlocks can occur when two packets are routed simultaneously to two opposite directions over a link. Adjacent routers are sending packets simultaneously to each other while neither of them receives the packet before the previous packet has been sent. The deadlock avoidance is realized so that a router does not receive a packet until it can be surely routed forward. A packet is not routed to a direction where the output buffer of the router is reserved. A main property on this deadlock prevention is that the packets are not stored to the actual router. A packet lies in the input buffer until the routing decision has been made and the packet is moved straight to the correct output buffer.

The primary adaptive router design utilizes all the features discussed above. Its performance is later compared to performance of two reference systems. The reference routers operate deterministically without information about the current network state. The packets are routed first in the X direction to the accurate column and then in Y direction to the destination [4]. The reference systems differ so that one utilizes the deadlock avoidance method by moving packets straight from the input buffers to the outputs while the other stores the packets on its local buffer during the routing decision, and after that sends them forward. This feature prevents the utilization of the deadlock avoidance method.

The communication between the routers is realized using a synchronous handshake protocol. On the global ring network the packets are routed along the shortest path or the path with the least traffic level. The traffic level of the ring network is indicated using an identifier which indicates if there exists or has recently existed traffic on the corresponding channel. The inter-layer links provide a possibility to move data between the layers, but a continuous utilization of these links would overload the global network and slow down the operation. The deadlocks on the global network are avoided using the same deadlock avoidance method as in the mesh networks. In addition, blocking of the global ring network is avoided assigning higher priorities to the packets which are already traveling in the ring network. This prevents blocking of the global network because it cannot become overloaded. On the deterministic reference routers the packets are routed in the global ring network through the shortest path.

IV. Monitoring

The monitoring on an intelligent NoC can include monitoring of traffic levels on the network as well as monitoring of workloads of the computational devices. The communication medium monitoring aims to balance the network load over the whole network and to avoid congestion. The monitoring of the workloads prevents overloading parts of the system while other parts are idle. The load balancing is essential because that way the energy consumption can be distributed over the whole chip which also spreads the heat production to a larger area.

The implemented monitoring service on the hybrid mesh-ring NoC monitors traffic levels in the mesh and ring networks and provides this information to the routers. The traffic level of a mesh network router is indicated with four-level traffic index which describes the current traffic load of the current router and its neighboring routers. Table I shows how the traffic level indexes are determined. In the table, *local traffic* shows how many output ports of the router are currently loaded. Respectively, *neighbor traffic* describes the number of neighboring routers which are currently handling packets. The neighboring routers with traffic level value less or equal to 1 are considered to be idle while others are used to define the traffic level. The traffic level index is determined using the Table I in a way that the column, which leads to bigger traffic level index, is used. The determined traffic level index illustrates the state of the router and is reported to the neighboring routers.

TABLE I
Traffic level determination.

Local traffic	Neighbor traffic	**Traffic level**
$\geqslant 3$		3
2	$\geqslant 3$	2
1	2	1
0	$\leqslant 1$	0

A monitor observes the state of the local router as well as the traffic indexes of the neighboring routers. Because the traffic level index of a neighboring router is based on its own state as well as the traffic levels of its neighbors, the traffic information spreads indirectly over the whole mesh network. This kind of monitoring is referred as distributed network monitoring. It improves the possibility to distribute traffic to the whole network. The widely spread traffic information steers packets away from the congestion spots and leaves the resources around the congestion spot to be used for resolving the congestion. Used with a dynamically adaptive XY routing algorithm, the neighbor monitoring enables efficient usage of the network resources [6].

In Fig. 3 and Fig. 4 the traffic information distribution using the distributed monitoring is compared to the distribution using a local monitoring method. A traditional local monitoring method determines the traffic levels based only on the state of the local router. Figures illustrate mesh networks with 16 routers. The height of a bar shows the traffic level on a router. In Fig. 3 there is one sender and one receiver in the network and in Fig. 4 two of both. Figures show how the traffic level information spreads.

The global ring network has less optional routes than the local mesh networks, but the traffic load can still be

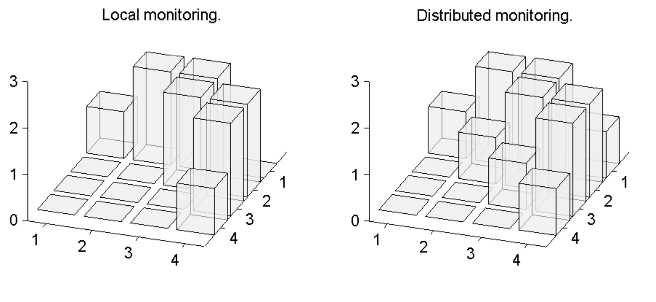

Fig. 3. Spreading of the traffic level information. One sender and one receiver.

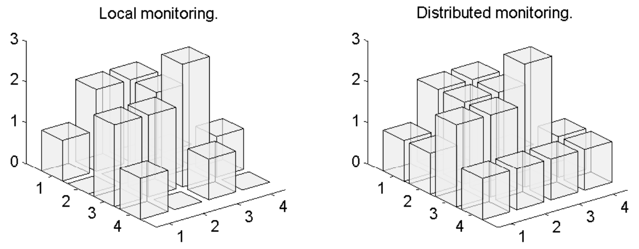

Fig. 4. Spreading of the traffic level information. Two senders and two receivers.

balanced between the channels of the ring network. A ring network monitor monitors the traffic load on the channels and distributes traffic to both of them while the priority is to use the shortest routes.

V. Results

The communication structure was implemented with VHDL and synthesized using 90 nm CMOS technology. The simulations concentrate on the throughput and latency analysis of the hybrid NoC system. The throughput and delay simulations are carried out using random traffic patterns. The throughput values of the adaptive system and two deterministic reference systems are presented in Fig. 5. The throughput and the offered traffic are defined as fractions of the maximum capacity of the system. The capacity is determined as bisection of the four mesh networks [5]. The maximum throughput of the deterministic XY routing is almost independent of the deadlock avoidance method. The impact of the deadlock avoidance can be seen after the moment when the maximum throughput has been achieved. After that point deadlocks start to form on the system without deadlock avoidance and the throughput finally collapses. On the other deterministic system the deadlocks do not exist and the throughput saturates after reaching its maximum value. As can be also noticed the adaptive algorithm increases the performance by 70 % while the system still operates stable.

Average propagation delay of the packets is presented in Fig. 6. The delay is defined as number of clock cycles and proportioned to the offered traffic. As can be noticed the deadlocks, existing in the other reference system, significantly slow down the operation of the system. The efficient network resource utilization of the adaptive routing algorithm can be seen as lowest average delays. The average delay of the adaptive system is about half of the delay of the stable deterministic reference system.

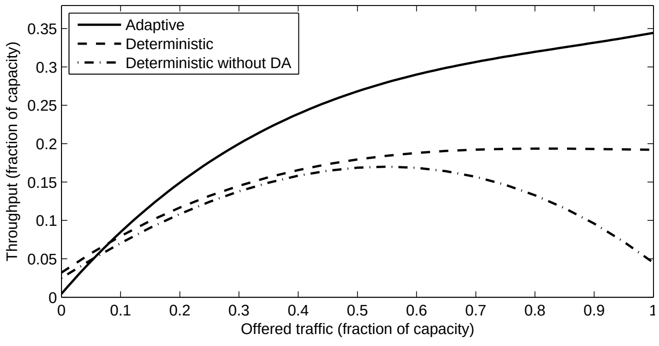

Fig. 5. Comparison of throughput with different routing algorithms.

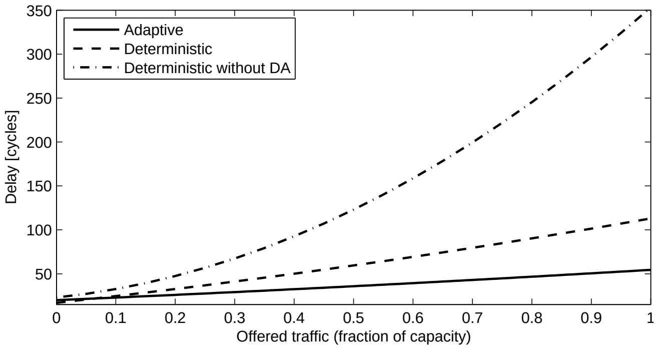

Fig. 6. Comparison of average delays with different routing algorithms.

VI. Conclusions

A hybrid mesh-ring Network-on-Chip was implemented with VHDL and analysed. The system utilizes adaptive XY routing and distributed network monitoring services. The research proves that the use of adaptive routing with presented monitoring technique improves the performance of the communication structure significantly without a need to increase the operating frequency of the system. The achieved performance increase is considerable with respect to the area overhead caused by the monitoring and adaptive features. The adaptive routing and the monitoring services increase the throughput by 70 % and halve the average delays while the size of the implementations increases by approximately 9 %.

VII. Acknowledgement

The authors would like to thank the Academy of Finland for financial support.

References

[1] Y. Xie, G. H. Loh, B. Black, and K. Bernstein, *Design Space Exploration for 3D Architectures*, ACM Journal on Emerging Technologies in Computing Systems, 2(2):65–103, April 2006.

[2] A. Vassighi, M. Sachdev, *Thermal and Power Management of Integrated Circuits*, Springer, 2006.

[3] S. Sapatnekar and K. Nowka, *Guest editors introduction: New Dimensions in 3D Integration*, IEEE Design & Test of Computers, 22(6):496–497, 2005.

[4] V. Rantala, T. Lehtonen and J. Plosila, *Network on Chip Routing Algorithms*, TUCS Technical Report, No 779, August 2006.

[5] W. J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*, Morgan Kaufmann Publishers, 2004.

[6] M. Dehyadgari, M. Nickray, A. Afzali-kusha, and Z. Navabi, *Evaluation of Pseudo Adaptive XY Routing Using an Object Oriented Model for NoC*, In The 17th International Conference on Microelectronics, December 2005.