Pipeline-Based Interlayer Bus Structure for 3D Networks-on-Chip

Masoud Daneshtalab, Masoumeh Ebrahimi, Pasi Liljeberg, Juha Plosila, Hannu Tenhunen
Department of Information Technology, University of Turku, Turku, Finland
{masebr, masdan, paslil, juhplo, hanten}@utu.fi

Abstract- The structure of direct vertical interconnections, called Through Silicon Vias (TSVs), is an important issue in the realm of 3D ICs. The bus-based and network-based structures are the two dominant architectures for implementing TSVs as interlayer connection in 3D ICs. Both implementations have some disadvantages. The former suffers from poor scalability and deteriorates the performance at high injection rates, and the latter consumes more area and power dissipation. In this paper, we propose a novel pipeline bus structure for TSVs to improve the performance of the prior bus-based architecture. The presented structure can utilize bi-synchronous FIFO for synchronization between stacked layers if each layer is fabricated by different technologies. Experimental results with synthetic test cases demonstrate that the proposed architecture gives significant improvements in average network latency. Also, the hardware area and power consumption of the presented bus structure are 9% and 11% less than the typical bus structure of TSVs, respectively.

I. INTRODUCTION

The Network-on-Chip (NoC) architecture paradigm, based on a modular packet-switched mechanism, can address many of the on-chip communication design issues such as performance limitations of long interconnects, and integration of high numbers of intellectual property (IP) cores in a chip [1][2][3]. However, two-dimensional (2D) chip fabrication technology is facing lots of challenges in the deep submicron regime even by utilizing NoC architectures [4][5], e.g. designing the clock-tree network for a large chip, limited floor-planning choices, increasing the wire delay and power consumption, integrating divers components that are digital, analog, MEMS and RF, and etc. The Three Dimensional (3D) integration has emerged as a potent solution to address these problems and the design complexity of MPSoC in 2D Integration Circuits (IC). 3D ICs reduce the interconnect delay problem by stacking vertically active silicon layers as well as offering a number of advantages over the traditional 2D chip [5][6][7][8]: (1) shorter global interconnects; (2) higher performance; (3) lower interconnect power consumption due to wire-length reduction; (4) higher packing density and smaller footprint; and (5) support for the implementation of mixed-technology chips. In this paper we focused on wafer stacking technology. In wafer-to-wafer bonding technology, one of the popular choices for 3D integrations, 3D chips are vertically stacked with short and fat Through Silicon Vias (TSVs). The distance between wafers can range from 5µm to 50µm [8][9], which is much shorter than the wire length between cores on a tier, and the pitches of a TSV can range from 1µm to 10µm square [8][9]. That is, the wire delay, power consumption and chip form factor are significantly reduced [10][11][13]. If 3D ICs are adopted as the basic fabrication standard, the performance of NoCs will be significantly improved. Therefore, 3D NoCs have been introduced as a new communication infrastructure of multi-core architectures [4][5][6][7][8].

The power-performance and efficiency of 3D NoCs essentially depend on the underlying topology formation. On-chip network topology is a crucial factor of 3D chips in terms of performance, cost, and energy consumption [5]. Various network topologies have been studied for 3D NoCs [5][6][7][10][12][14]. 3D Symmetric NoC and 3D NoC-Bus Hybrid (stacked mesh) structures are popularly used in 3D systems, because their grid-based regular structure is intuitively considered to be matched to the 2D VLSI layout for each stack layer [5][6][7][8][10]. By adding two additional physical ports to each baseline-router (one for up and one for down) in the popular 2D mesh-based system the 3D symmetric NoC structure is formed (Fig. 1(a)) [5][8]. This structure suffers from: adding two additional ports requires larger crossbar incurring significant area and power overhead. That is, the more input ports employed the more blocking probability occurred inside the router. This increases the latency of moving flits through the upward and downward ports. Thereby, the 3D NoC-Bus Hybrid structure using a single-hop communication (bus) among layers is presented (Fig. 1(b)) [5][8]. Routers in this structure have at most 6 ports, one to the IP-core, one to the bus, and four for cardinal directions. Due to the fact that the 3D IC is emerging as a promising solution to continue the growth of the number of transistors on a chip, in this research work we explore the topology formation in 3D ICs.

Our contribution of this work is to propose a new 3D topology based on stacked mesh architecture. In this topology, we introduce a high-performance pipeline bus structure to overcome the drawbacks of the previously presented bus structures such as dTDMA bus [15] and segmented bus [16] to communicate among multiple layers. The benefits of the proposed bus structure are as follows:

First, this novel bus architecture, improve the performance, by reducing the delay and complexity of previously proposed bus' arbitration module which is the foremost impediment in bus communications.

Second, if in some cases, each layer built with different vendors and with completely different processes, each layer should have its own clock tree with associated clock buffers [17]. Furthermore, because there is no clear solution of

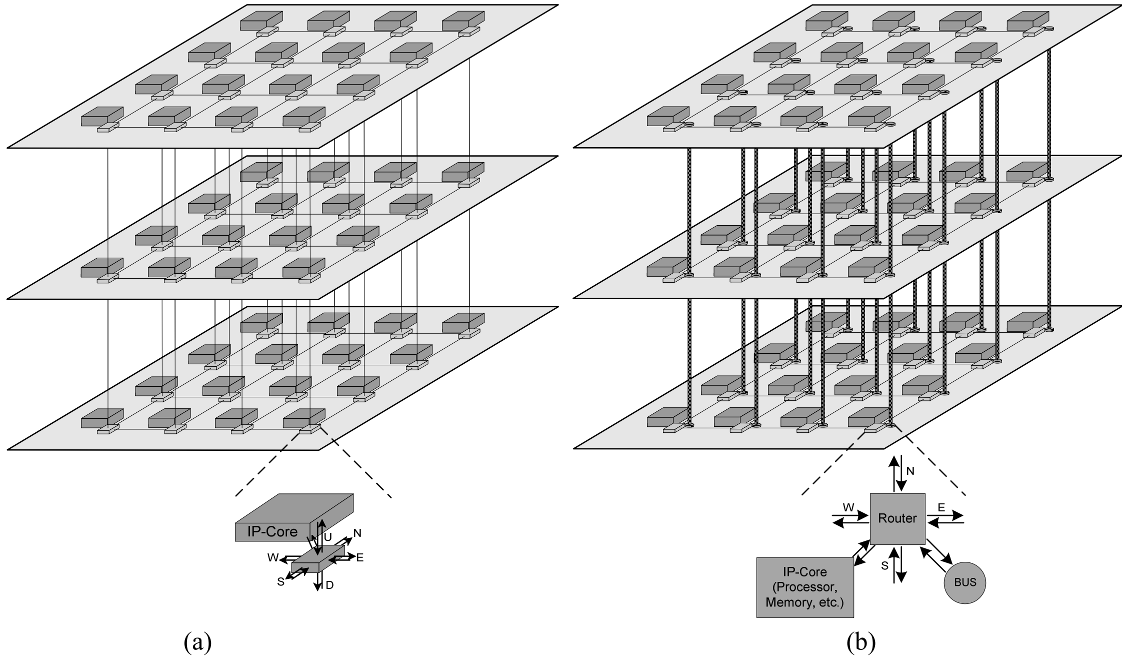

Fig. 1. Mesh-based NoC architectures: (a) 3D symmetric NoC (b) 3D NoC-Bus Hybrid structures.

modular and skew-free clock distribution in 3D ICs, a clock synchronization mechanism between active layers through vertical connections (interlayer vias) is required [17]. In this work, we introduced a programmable synchronous and asynchronous pipeline-bus structure to cope with the multi clock domains communication. This structure can be employed as an interlayer connection to handle the communication among layers with different clock frequencies in 3D chips.

II. RELATED WORKS

Design techniques and methodologies for 3D architectures have been investigated to efficiently exploit the benefits of 3D technologies. Several NoC topologies for 3D systems have been exhaustively investigated in [5][6][7][8][12][18]. The authors in [5] demonstrate that besides reducing the footprint in a fabricated design, 3D systems provide a better performance compared to traditional 2D systems. They have also demonstrated that both mesh and tree topologies for 3D systems achieve better performance compared to traditional 2D systems. However, the mesh topology shows significant performance gains in terms of throughput, average latency, and energy dissipation with a small area overhead [5]. In [18] different 3D mesh-based architectures have been compared in the zero-load latency, but the performance of the network with different traffic patterns and loads is also necessary to be evaluated.

To form an optimistic 3D mesh-based system, several 3D structures have been presented. Baseline-routers in 2D mesh-based systems have 5 ports, i.e. 4 ports to adjacent routers and one for the resource node. The straightforward extension for 3D mesh-based systems (*3D symmetric NoC*) is to utilize routers with two additional interlayer links by adding two physical ports to baseline-routers (one for up and one for down) [5][8][10][12][19]. As mentioned earlier, the 3D structure using such routers, not only increase the area and power overhead of the routers but also degrades the performance. The electrical behavior of the relatively short and wide TSV, i.e. the low resistance, and supporting much higher signaling speeds led the authors of [8] to propose the 3D hybrid structure. This 3D structure exploits the Dynamic Time Division Multiple Access (dTDMA) bus [15] with a centralized arbiter for the vertical communication link. Thus, moving from one layer to any of the other layers takes only one hop. However, contention issues in the bus limit the attainable performance gains [5]. That is, such structures inherently suffer from the limitation of buses since only one transmission is allowed each time over a vertical bus.

In [7], the DimDe router for 3D architectures has been proposed. The presented router uses a full 3D crossbar and a simple bus structure spanning all layers of the chip and fusing them into a single router entity. This router can minimize vertical traversal to one hop between any layers, but requires huge number of vertical connections and significantly complicates the control and arbiter of the router.

A multilayered 3D router architecture, named MIRA, is introduced for 3D systems by D. Park et al [6]. The router components are classified as separable components (buffers, crossbar, and inter-router links) and non-separable components (arbiter and routing modules).

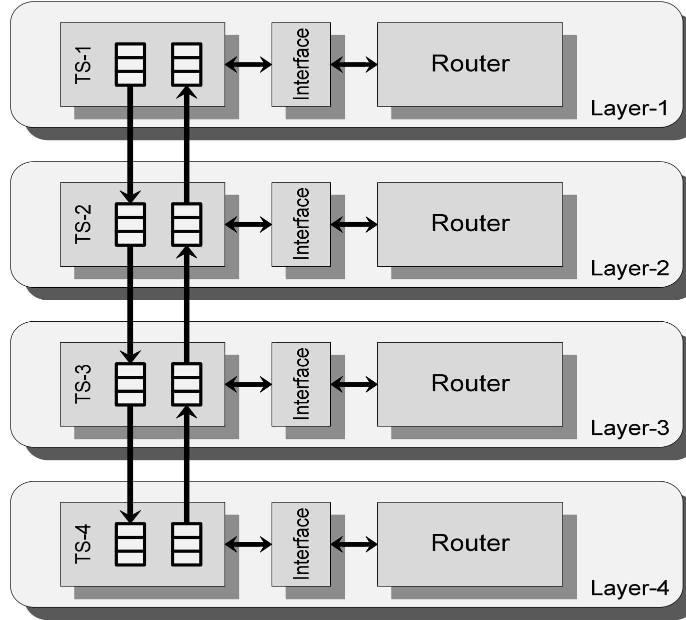

Fig. 2. Proposed Bus Architecture.

The separable components are laid out across multiple layers to save chip area and reduce power by dynamically shutting down some inactive layers. However, such routers are too aggressive in the current technology [20].

Due to the above concerns, in this paper we have focused on the 3D symmetric structure (7-port switch design) and the 3D hybrid structure (bus-based vertical interconnect). As described in [12], the 3D hybrid structure was shown to perform the worst out of the other structures in terms of scalability under local traffic, as it is physically limited by its raw bandwidth due to a smaller links per node ratio and contention issues as the number of layers increases. Although shown to be weak in [12][20], the bus may be appropriated for hot spot traffic injection where many packets may need to be sent through several layers to a hot spot frequently. This may be akin to a processor on one layer, and a memory stack directly above it. In sum, in 3D architectures, the 3D hybrid structure performance degrades as the number of layers and number of processing nodes increase [12], thereby the 3D symmetric structure is more feasible, mature, and more efficient than the Hybrid structure as network size increases [21]. In this work, we present an efficient bus structure containing the benefits of both Hybrid and Symmetric structures with lower hardware overhead and power dissipation.

III. PROPOSED TSV ARCHITECTURE

Traditionally, a bus is described as a shared link which can be owned by one attached subsystem at the time, i.e., when one module is transmitting via the bus, the others can only be receiving. Parallelism can be added to the structure by partitioning the bus into segments with bridges and allowing these segments to operate concurrently [5]. However, on one side, the overall system performance in such designs is still limited by the lack of parallel bus transactions, and on the other hand, because of using many control wires for the central arbitration in such segmented buses, it is not a suitable approach to use them as vertical bus (TSVs) in 3D ICs. Our solution for these bottlenecks for vertical buses is to consider the system bus with a bidirectional pipeline which is capable of transferring data concurrently from one or more sources to several destinations. The proposed architecture, named Novel Interlayer Structure (NIS), is illustrated in Fig. 2. The system is partitioned into a set of modules each of which forms its own timing domain. In fact, each module is used to connect the corresponding layer to the pipeline bus. As the system is based on GALS design paradigm, the layers can internally operate at different clock frequencies. The interface between the layers can be either self-timed, i.e., based on asynchronous handshake signaling, or synchronous in which case the interface forms a clock domain of its own. The layers are independent of each other, in case there are some interlayer transactions, the layers exchange data synchronously or asynchronously through the pipelined system bus, a segmented communication link which allows simultaneous transfer in both directions. The layers can concurrently access the bus without waiting for any grant signals, because of the pipelined structure of the proposed bus architecture. The interface module role is to act as a dedicated adapter between the internal, possibly synchronous timing domain of the layer and the bus domain. Indeed, it acts as a synchronizer between the chip layer and the pipeline bus. To form the pipelined bus, the physical wires that

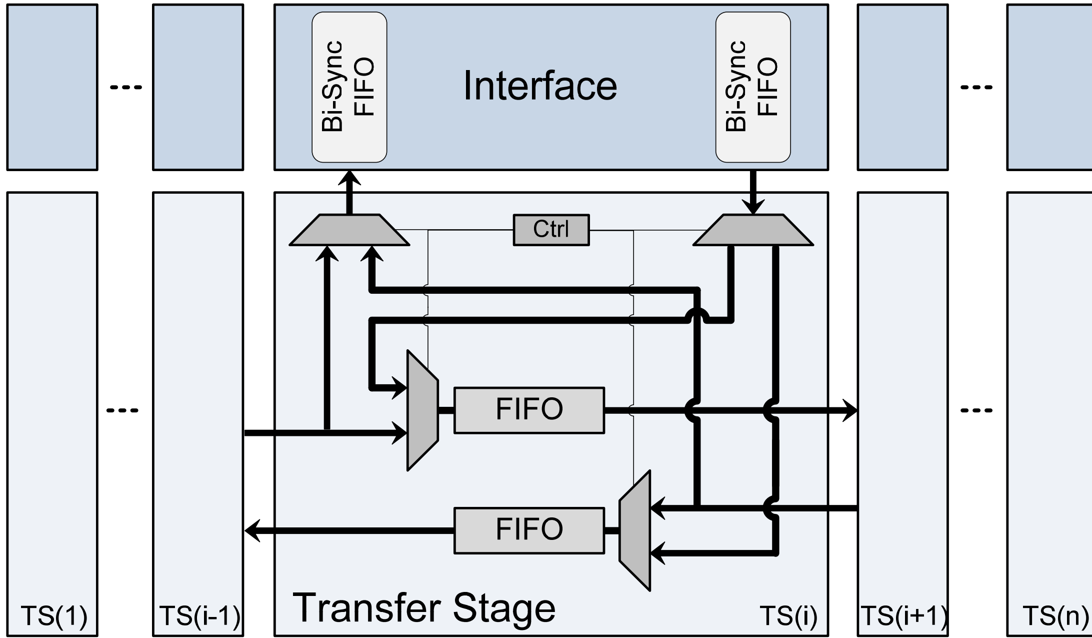

Fig. 3. The micro-architecture of the transfer stage and interface.

implement the bus are divided into a set of segments separated from each other by Transfer Stages (TS), one attached to each layer (Fig. 2). Each transfer stage contains internal FIFO queues for pipelining the data flow, and a bus segment between adjacent stages consists of two separate unidirectional point-to-point interconnects which transfer data synchronously (or asynchronously) between the stages in opposite directions. These two links of a segment can operate in parallel, and due to pipelining, all segments of the system bus can transfer data simultaneously. Each layer has a unique address for inter-layer communication. Furthermore, each IP-core in a layer has its own address which makes addressing of a specific module in a given layer possible. Hence, a datagram propagating along the bus has a header containing both the layer address and the IP-core address. The former is analyzed at each transfer stage, and the latter is decoded by routers in each layer.

A. Transfer Stage micro-architecture

A transfer stage has three different functions:

1. It forwards incoming data from the preceding stage to the next stage through a buffer, in both directions.
2. If incoming data from an adjacent transfer stage is intended to be processed by the layer, the transfer stage transfers data to the interface module of the layer through a FIFO queue.
3. When the layer decides to send data to another layer, the transfer stage operates as an output buffer. This means that it takes care of first receiving data from the interface module of the attached layer through a FIFO queue and then sending this data to one of the two adjacent transfer stages, depending on the direction in which, the target layer is located.

The micro-architecture of the transfer stage is illustrated in Fig. 3. A transfer stage includes two identical pipelines which transfer data to the opposite directions. Each pipeline contains three registers to pipe data between segments.

Apart from the pipelines the interface contains FIFO queues used as input and output buffers of the host port. Their capacity has to be chosen according to the speed of the bus interface and the estimated data rate of the attached router.

When data arrives to a pipeline from the neighboring transfer stage, it will be either forwarded to the next stage or transferred to the host router via the interface, determined by the destination address. An arbitration in the controller module has to be performed to prevent the two parallel operating pipelines from writing simultaneously to the FIFO in the interface.

Data sent by a host router is divided into two pipeline FIFOs according to the destination address and will be sent to one of the adjacent transfer stages. It is also possible to use different priority schemes. Incoming data from a router can be prioritized so that it will be sent to the next stage prior to the forwarded data from the previous stage, or the other way around. If no priority scheme is used and there is a continue dataflow from both directions, the output data will be selected by alternating equally between both sources. In this paper, the round robin mechanism [22] is exploited. In addition, because the electrical behavior of short and wide TSVs provides much higher signaling

speeds, the credit-based flow control [2] has been implemented for the transmission protocol on a segment between transfer stages.

B. Synchronizing FIFO

Bi-synchronous (Bi-Sync) FIFOs are widely used in multi-clock system to synchronize signals from different clock/frequency domains. Each domain is synchronous to its own clock signal but can be asynchronous with others in either clock frequency or phase [23]. The challenges of designing Bi-Sync FIFOs include the enhancement of reliability and reducing latency and power/area cost. We identify the Bi-Sync FIFO structure presented in [24] as a suitable synchronizer to be used in the interfaces.

The structure of the Bi-Sync FIFO is described in [24]. The FIFO implementation uses two pointers, one defining the next writing position and another defining the next reading position. The FIFO state is either full or empty when both pointers refer to the same address. Thus, it is necessary to compare the pointers. Although this procedure is trivial in synchronous circuits, it implies some complexity in the Bi-Sync FIFO, because the pointers are generated by different clocks. The usual solution to solve this problem is to transfer and synchronize the writing pointer (reading pointer) with the receiver clock domain (the sender clock domain) which generates the empty signal (the full signal). Exchanging the pointers via a handshake protocol implies additional latency. Therefore, two synchronizers are utilized for exchanging the pointers [24]. The addresses are translated to Gray code which guarantees that consecutive addresses are at a Hamming distance of 1. In this way, the metastability problem is confined to a single bit and synchronizers can be employed without handshake. Utilizing the Bi-Sync FIFO in the interfaces, allows each layer to work with its own clock source.

IV. EXPRIMENTAL RESULTS

In this section, we compare the proposed interlayer structure with the hybrid bus-based (dTDMA) and symmetric 7-port structures by measuring the average network latency under different traffic patterns. Hence, a 3D NoC simulator is implemented with VHDL to assess the efficiency of the proposed architecture. The simulator models all major components of the NoC such as network interface, routers, and wires.

A. System Configuration

In this work, we use a 36-node (3×3×4) 3D mesh on-chip network configuration for the entire architecture. In this configuration, illustrated in Fig. 4, out of 36 nodes, 12 nodes are assumed to be processors and other 24 nodes are memories. The processors are 32b AXI and the memories are DDR2-512MB (tRP-tRCD-tCL=2-2-2, 32b, 4 banks) [25][26]. In addition, the on-chip network considered for experiment is formed by a typical state-of-the-art router structure including input buffers, a VC (Virtual Channel) allocator, a routing unit, a switch allocator and a crossbar. Each router has 5 input/output ports, and each input port of the router has 2 VCs. Packets of different message types (request and response) are assigned to corresponding VCs to avoid message dependency deadlock [27]. The arbitration scheme of the switch allocator in the typical router structure is round-robin.

The array size, routing algorithm, link width, number of VCs, buffer depth of each VC, and traffic type are the other parameters which must be specified for the simulator. The routers adopt the XY [2] routing and utilize wormhole switching. For all routers, the data width (flit) was set to 32 bits, and the buffer depth of each VC is 5 flits. The presented configuration uses 1 flit for messages related to read requests and write responses, and the size of read request messages typically depends on the network size and memory capacity of the configured system. The message size of the read responses and write requests is variable and depends on the request/response length produced by a master/slave core. As the performance metric, we use latency defined as the number of cycles between the initiation of a request operation issued by a master (processor) and the time when the response is completely delivered to the master from a slave (memory). The request rate is defined as the ratio of the successful read/write request injections into the network interface over the total number of injection attempts. All the cores and routers are assumed to operate at 1 GHz. For fair comparison, we keep the bisection bandwidth constant in all configurations. All memories (slave cores) can be accessed simultaneously by each master core with continuously generating memory requests.

B. Performance Evaluation

To evaluate the performance of the proposed interlayer structure, the uniform and non-uniform synthetic traffic patterns have been considered separately for the specified configuration. These workloads provide insight into the strengths and weakness of the different interlayer structures in the interconnection networks, and we expect applications stand between these two synthetic traffic patterns. The random traffic represents the most generic case, where each processor sends in-order read/write requests to memories with the uniform probability. Hence, the memories and request type (read or write) are selected randomly. Eight burst sizes, among 1 to 8, are stochastically chosen regarding the data length of the request. In the non-uniform mode, the traffic consists of 70% local requests, where the destination memory is one hop away from the master core, and the rest 30% traffic is uniformly distributed to the non-local memories. Fig. 5 and Fig. 6 show the simulation results under uniform and non-uniform traffic models, respectively. The proposed interlayer structure, NIS, has been compared with the Hybrid (bus-based) and Symmetric (7-port router) structures.

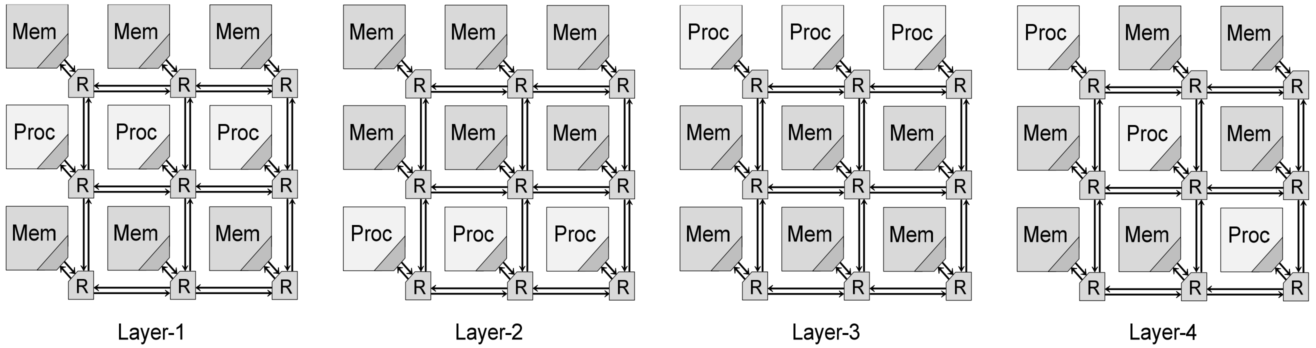

Fig. 4. 3×3×4 stacked mesh layout.

As demonstrated in both figures, compared with Hybrid and Symmetric structures, the presented architecture reduces the average latency when the request rate increases under uniform and non-uniform traffic models. One of the foremost reasons of such an improvement is that because NIS has a small local arbiter in the transfer stage, the arbitration delay is reduced significantly. The average latency of the network has been computed near the saturation point (0.4) under the uniform traffic profile. As a result of using NIS structure, compared with the Hybrid structure, the average network latency is reduced by 24%.

Table 1. Hardware implementation details.

Router	Area (μm^2)	Power (mW)
2D Symmetric structure	96328	52.42
3D Symmetric structure	148715	74.68
Hybrid structure (+ central bus arbitration)	122455	67.88
NIS structure	110895	60.35

C. Hardware Implementation

The NIS, Hybrid, and Symmetric structures were synthesized by Synopsys Design Compiler using the UMC 0.09µm technology. For NIS we have considered that each router contains a transfer stage as well as an interface, described in Fig. 2 and Fig. 3. The layout areas and power consumptions of routers in the three different structures (NIS, Hybrid, and Symmetric) have been summarized in Table 1. Routers in the Hybrid and NIS structures are composed of 6 ports, but routers in Symmetric structure are constructed with 7 ports. As the results from Table 1, the hardware area and power consumption of NIS's router are 9% and 11% less than that of the Hybrid's router. In addition, the hardware implementation details in Table 1 are based on the typical FIFO without considering the synchronization issue. If supposed to consider the synchronization between layers, the FIFOs in the interfaces are needed to be replaced by Bi-Sync FIFO for Hybrid and NIS structures whereas in the Symmetric structure, the down-port and up-port FIFOs of routers should be replaced. In this case, the overhead of the Symmetric structure increased significantly because the number of Bi-Sync FIFOs replaced by the typical FIFO in this structure is two times more than the other two structures.

V. CONCLUSION

3D stacked architectures provide significant benefits in performance, footprint and yield. It has also been demonstrated that combining 3D ICs and on-chip networks can be a promising option for designing large multiprocessor architectures. One critical issue in 3D design is that the vertical interconnections are very fast and the existent architectures of interlayer connection exacerbate the performance. In this paper, we presented a novel pipeline bus structure to improve the performance of the prior bus-based architecture for interlayer connection. A 3D simulator was used to evaluate the efficiency of the proposed architecture. Under both uniform and non-uniform traffic models, in high traffic load, the proposed structure had lower average communication delay in comparison with the other architectures. The power dissipation and hardware area of the presented structure is also lower than the traditional structures.

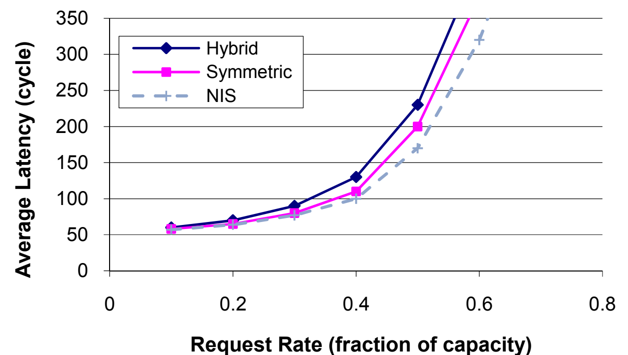

Fig. 5. Performance evaluation under uniform traffic model.

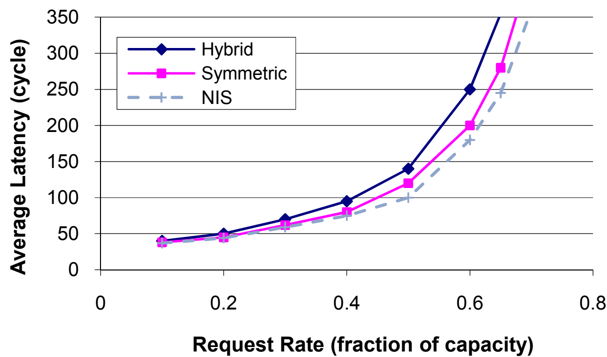

Fig. 6. Performance evaluation under non-uniform traffic model.

ACKNOWLEDGMENT

The authors wish to acknowledge the academy of Finland and Nokia Foundation for the financial support during the course of this research.

REFERENCES

[1] W. J. Dally and B. Towles, “Route packets, not wires: On-chip interconnection networks,” in Proceedings of the 38th Design Automation Conference, June 2001, pp. 684–689.

[2] A. Jantsch and H. Tenhunen, Networks on Chip. New York: Kluwer, 2003.

[3] M. Daneshtalab, M. Ebrahimi, P. Liljeberg, J. Plosila, and H. Tenhunen, "Input-Output Selection Based Router for Networks-on-Chip" in Proceedings of 9th IEEE International Symposium on VLSI (ISVLSI), pp. 92-97, May 2010, Greece.

[4] K. Banerjee, S. J. Souri, P. Kapur, K. C. Saraswat, “3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration”, Proc. of the IEEE, 89(5):602-633, May 2001.

[5] B. S. Feero, P. P. Pande, “Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation,” IEEE Transactions on Computers, vol. 58, no. 1, pp. 32-45, Jan. 2009.

[6] D. Park, S. Eachempati, R. Das, A. K. Mishra, Y. Xie, N. Vijaykrishnan, and C. R. Das, “MIRA: A Multi-Layered On-Chip Interconnect Router Architecture”, ISCA 2008, pp. 251-261, Pennsylvania State, USA.

[7] J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, V. Narayanan, and C. R. Das, “A novel dimensionally-decomposed router for on chip communication in 3D architectures,” in Proc. of the ISCA, pp. 138-149, Boston, USA, 2007.

[8] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir, “Design and Management of 3D Chip Multiprocessors Using Network-in-Memory,” In 33rd International Symposium on Computer Architecture (ISCA), pp. 130–141, 2006.

[9] W. R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M. Sule, M. Steer, and P. D. Franzon, “Demystifying 3D ICs: The Pros and Cons of Going Vertical,” IEEE Design and Test of Computers, 22(6):498–510, Nov. 2005.

[10] I. Loi, F. Angiolini, L. Benini, “Supporting vertical links for 3D networks on chip: toward an automated design and analysis flow,” in Proc. Nanonets, 2007.

[11] K. Snoeckx, E. Beyne, and B. Swinnen, “Copper-nail TSV technology for 3D-stacked IC integration,” Solid State Technology, 50(5):53, 2007.

[12] A. Y. Weldezion, M. Grange, D. Pamuwa, Z. Lu, A. Jantsch, R. Weerasekera, and H. Tenhunen. Scalability of the Network-on-Chip communication architecture for 3-D meshes. In International Symposium on Networks-on-Chip (NoCS), pp. 114-123,2009.

[13] I. Savidis, S. M. Alam, A. Jain, S. Pozder, R. E. Jones, R. Chatterjee, “Electrical modeling and characterization of through-silicon vias (TSVs) for 3-D integrated circuits,” Microelectronics Journal, Vol. 41(1), pp. 9-16, 2010.

[14] S. Pasricha, “Exploring Serial Vertical Interconnects for 3D ICs,” in Proc. IEEE/ACM DAC, pp. 581-586, 2009.

[15] T. Richardson, C. Nicopoulos, D. Park, V. Narayanan, Y. Xie, C. Das, and V. Degalahal, “A hybrid SoC interconnect with dynamic TDMA-based transaction-less buses and on-chip networks”, In Proc. VLSID, pp. 8-15, 2006.

[16] R. Lu, A. Cao, and C. Koh, “SAMBA-Bus: A High Performance Bus Architecture for System- on-Chips”, IEEE Transactions on VLSI Systems, Vol. 15, Issue 1, pp. 69-79, Jan 2007.

[17] I. Loi, F. Angiolini, L. Benini, “Developing mesochronous synchronizers to enable 3D NoCs”, Proc. of DATE, Germany, pp. 1414-1419, 2008.

[18] V.F. Pavlidis and E.G. Friedman, “3-D Topologies for Networks-on-Chip,” IEEE Transactions on Very Large Scale Integration Systems, 15(10):1081, 2007.

[19] S. Murali, C. Seiculescu, L. Benini, and G. D. Micheli, “Synthesis of networks on chips for 3D systems on chips,” In Proceedings of the 14th Asia and South Pacific Design Automation Conference (ASPDAC’09), pp. 242-247, Jan. 2009.

[20] Y. Qian, Z. Lu and W. Dou, “From 2D to 3D NoCs: A Case Study on Worst-Case Communication Performance,” In. Proc. of the International Conference on Computer-Aided Design (ICCAD), pp. 555-562, 2009.

[21] A. Y. Weldezion, Z. Lu, R. Weerasekera, and H. Tenhunen, “3-D Memory Organization and Performance Analysis for Multi-processor Network-On-Chip Architecture,” in Proc. of IEEE International 3D System Integration Conference, San Francisco, USA, 2009.

[22] A. Demers, S. Keshav and S. Shenkar, “Analysis and Simulation of a Fair Queuing Algorithms,” Proceedings of SIGCOMM’89, pp. 3–12, August 1989.

[23] T. Ono, M. Greenstreet, “A modular synchronizing FIFO for NoCs,” in proc. of the 3rd ACM/IEEE International Symposium on Networks-on-Chip, USA, pp. 224-233, 2009.

[24] C. E. Cumings, “Simulation and Synthesis Techniques for Asynchronous FIFO Design”. In: SNUG, 2002.

[25] Micron Technology, Inc. Micron 512Mb: x4, x8, x16 DDR2 SDRAM Datasheet, 2006.

[26] M. Daneshtalab, M. Ebrahimi, P. Liljeberg, J. Plosila, and H. Tenhunen, "A Low-Latency and Memory-Efficient On-Chip Network," in Proceedings of 4th IEEE/ACM International Symposium on Network-on-Chip (NOCS), pp. 99-106, May 2010, France.

[27] S. Murali, and et al. “Designing message-dependent deadlock free networks on chips for application-specific systems on chips,” In Proc. VLSI-SoC, pages 158-163, 2006.