International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 5, May2014)

HEADFIRST SLIDING ROUTING FOR 3-D CHIP MULTIPROCESSORS

Sowmyashree S
Student, VTU University
The Oxford College of Engineering, Banglore, india
Sowmyashree6@gmail.com

Laya Tojo
Assistant Professor, VTU University
The Oxford College of Engineering, Banglore, india
layatojo@gmail.com

***Abstract*—Long interconnects are becoming an increasingly important problem from both power and performance .The Headfirst sliding routing scheme to overcome the too much delay problem in simple TDMA-based vertical buses. Each vertical bus gives permission to communication time-slot for different chips at the same time periodically, which means these buses work with different phases. To avoid deadlocks, two VCs are required for all the routers. The main contribution of this paper is reduce the area, power and delay in the NOC chip by using 3D router with additionally gives Headfirst Sliding Routing. Network simulations show that Headfirst sliding routing reduces the area 184mm^{2}and Delay upto reduced 0.120ns . Synthesis results show that the area and critical path delay overheads are modest.**

***Keywords*—Minimum-Hop(MH) routing,RPM routing, Headfirst sliding routing .**

I.INTRODUCTION

Currently, microchips can only pass digital information in a very limited way — from either left to right or front to back, the researchers say. In the future, a 3D microchip would enable additional storage capacity on chips by allowing information to be spread across several layers instead of being compacted into one layer, as is currently the case.

This Long interconnects are becoming an increasingly important problem from both power and performance perspective. This motivates designers to adopt on-chip network-based communication infrastructures and three-dimensional (3D) designs where multiple device layers are stacked together. Considering the current trends towards increasing use of chip multiprocessing, it is timely to consider 3D chip multiprocessor design and memory networking issues.

The three-dimensional integration is a promising VLSI architecture that stack several smaller wafers or dies in order to reduce the wire length and wire delay, and three-dimensional Network-on-Chip (3-D NoC) [1] has been extensively including in terms of its network topology: The topology of an NOC specifies the physical organization of the interconnection network. It defines how nodes, switches and links are connected to each other [2][3][4], router architecture: Its responsible for correctly and efficiently routing packets[5][6][7], and routing strategy: It determines how data flows through the routers in the network and also defines the data transfer and applied switching techniques[8].

There has been considerable discussion in recent years on the benefits of 3-D silicon integration in which multiple device layers are stacked on top of each other with direct vertical interconnects tunneling through them [9] [10] [11] [12] [13]. 3-D integration promises to address many of the key challenges that arise from the semiconductor industry's relentless push into the deep Nano-scale regime. The increasing viability of 3-D technology has opened new opportunities for chip architecture innovations. One direction is in the extension of two-dimensional (2-D) tiled chip-multiprocessor architectures [14] [15] [16] [17] into three dimensions [18] [19]. Many proposed 2-D tiled chip-multiprocessor architectures have relied on a 2-D mesh network topology as the underlying communication fabric. Extending mesh-based tiled chip-multiprocessor architectures into three dimensions represents a natural progression for exploiting 3-D integration. The focus of this paper is on providing efficient routing for such 3-D Various interconnection techniques have been developed to connect multiple chips in a 3-D IC package: wire-bonding, micro-bump [20][21], wireless (e.g., capacitive- and inductive coupling) [22][23][24][25] between stacked dies, and through silicon via (TSV) [22][26] between stacked wafers. These 3-D IC technologies are compared in [6]. Many recent studies on 3-D IC architectures focus on micro-bump and TSV techniques that offer the highest level of interconnect density. On the other hand, as another 3-D integration technique, the inductive coupling can connect more than two examined dies without wire connections.

Toward this purpose, the vertical communication interfaces should be simplified, while arbitrary or customized topologies should be used for intra-chip networks; thus, we focus on mesh networks.static Time Division Multiple Access (TDMA) buses for the inter-chip communication. In this paper, we propose the Headfirst sliding routing scheme to overcome the delay problem and reduced area in simple static TDMA-based vertical buses. The static TDMA-based vertical buses grants a communication time-slot for different chips at the same time periodically, which means they are working with different periodic scheduling. For example, at a certain moment, vertical bus 0 gives a time-slot for chip 0, vertical bus 1 allows chip 2, and vertical bus 2 allows chip 1. At the next phase, vertical bus 0 gives a time-slot for chip 2, vertical bus 1 allows chip 0, and vertical bus 2 allows chip 1. Eachvertical bus behaves just like an elevator in an office building.

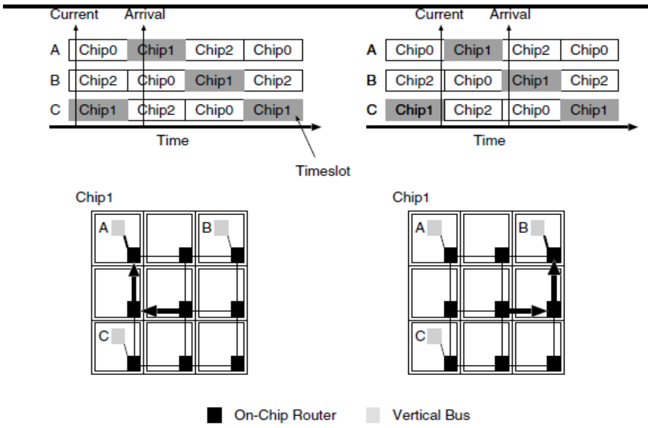
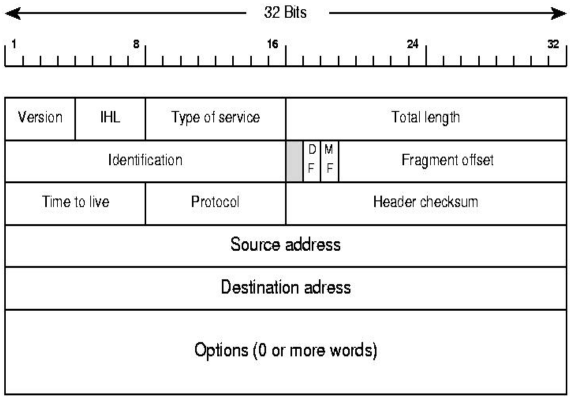

Current
Arrival
A Chip0 Chip1 Chip2 Chip0
B Chip2 Chip0 Chip1 Chip2
C Chip1 Chip2 Chip0 Chip1
Time
Timeslot
Chip1
A
B
C
On-Chip Router
Vertical Bus

Fig 1: Concept of Headfirst sliding routing

In concept of Headfirst sliding routing is after receiving the from 3D router switch then that data should be stored in the one of the communication time slots in the vertical bus, likewise one particular time send the more packets then this vertically connected TDMA buses gives the permission to communication time slots for different chips at same time. If send again more data then TDMA buses occupied in the next phase periodically.

For example in the Fig1 shows dark the chip 1 time slots because here consider chip 1 is the receives the packet from switch. When the data is depending on the current and arrival time then we concentration on our arrival time where appeared means which bus is allow the chip 1 and routs the packet towards the destination through the accepted bus (here accepted bus is A).

II.IPV4 PACKET HEADER



Version	IHL	Type of service	Total length		
Identification			D F	M F	Fragment offset
Time to live	Protocol		Header checksum		
Source address					
Destination adress					
Options (0 or more words)					

III. MINIMUM-HOP(MH) ROUTING

Minimum-hop (MH) routing and Headfirst sliding (HS) routing. MH routes packets using a minimal path between a source and a destination via an elevator. Our observation is that MH routing achieves a high saturated throughput while its zero-load communication waiting time or delay is longer than that of the dynamic TDMA (ideal case) due to the waiting time at switch. But here occur deadlock situation because without connecting virtual channels(VCs).

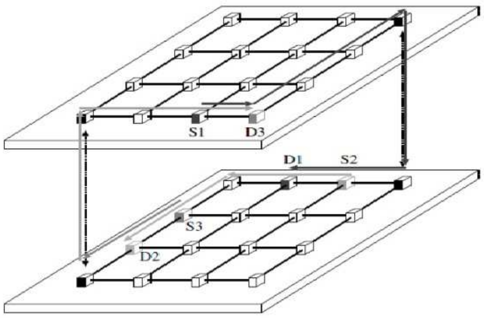


Fig 3:example of Deadlock situation

Minimum Hop Routing Packets are routed based on the following rules:

Case 1: If the source and destination are on the same chip, packets are routed based on arbitrary deadlock free routing on the chip (e.g., XY routing on 2-D mesh topology).

Case 2: If the source and destination are on different chips, packets are first routed to an elevator on the source chip, moved to the destination chip, and routed to the destination. An elevator is selected so that the hop count is minimized. Figure 2 illustrates an example of deadlock situation. Each chip employs 4 × 4 2-D mesh topology, in which XY routing is used for intra-packet transfers. In this case, S1 sends a message to D1, S2 sends a message to D2, and S3 sends a message to D3; thus they cause the cyclic dependency which introduces deadlocks.To reduce this deadlock virtual channels are added to each router.

IV.ROUTER IMPLEMENTATION

In this section, we discuss how 3D with Headfirst sliding router can be efficiently integrated into a typical on-chip router. We first explain how Headfirst sliding router can be made reduce delay and area.

a. RPM Router:

Fig. 5 shows the architecture of a typical 7-port router for 3-D mesh networks. But in the 2-D mesh networks have only with the addition of two extra ports for vertical communication.

At each input port, buffers are organized as separate FIFO queues, one for each VC. Flits entering the router are placed in one of these queues depending on their VC ID. The router is generally pipelined into five stages comprising *route computation, VC allocation, switch allocation, switch traversal and link traversal.*

The route computation stage determines the output port of a packet based on its destination. This is followed by VC allocation where packets acquire a virtual channel at the input of the downstream router. A packet that has acquired a VC arbitrates for the switch output port in the switch arbitration stage. Flits that succeed in switch arbitration traverse the crossbar before finally traversing the output link to reach the downstream router. Head flits proceed through all pipeline stages while the body and tail flits skip the route computation and VC allocation stages and inherit the output port and VC allocated to the head flit. The tail flit releases the reserved VC after departing the upstream router.

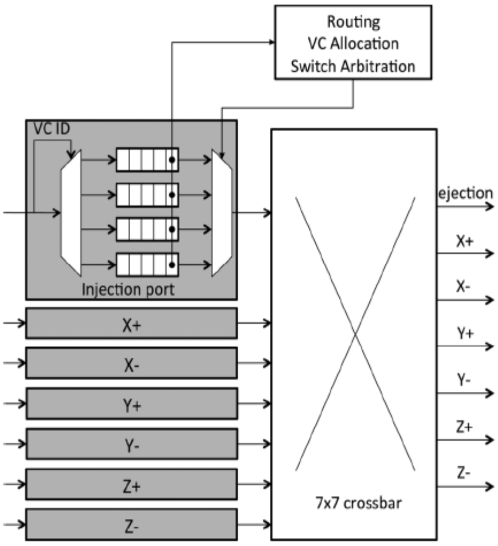
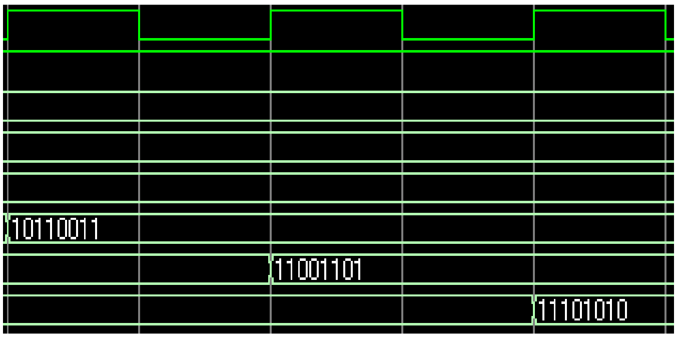


Fig 5: 3D router architecture



Too much delay problem in simple TDMA-based vertical buses. here given three channel data’s but three only consumed more delay and area. To overcome this area and delay problem “Headfirst sliding Routing concept” is introduced.

V. PROPOSED RPM ROUTER ARCHITECTURE

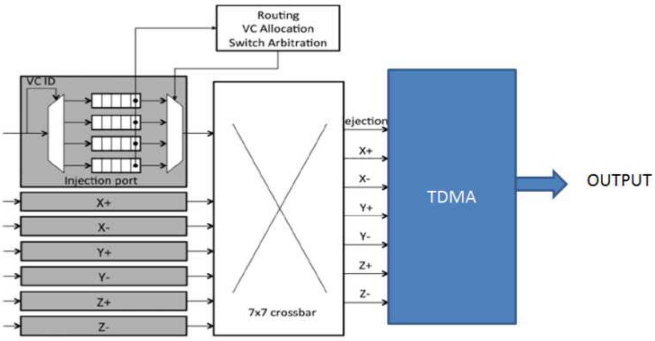


Fig 6:Proposed 3d router architecture

In the proposed router architecture have an advantages i.e reduced the delay ,area and power dissipation.
Headfirst Sliding Routing concept:

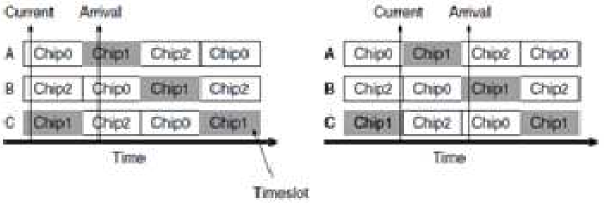
- we propose the Headfirst sliding routing scheme to overcome the simple static TDMA-based vertical buses. The static TDMA-based vertical buses grants a communication time-slot for different chips at the same time periodically, which means they are working with different periodic scheduling.
- Fortunately, a waiting time to obtain the time-slot of vertical bus (elevator) is predictable for each chip, thus a key design of packet routing is to select the best elevator that minimizes the waiting time.

TDMA (Headfirst Sliding Routing concept) is reduced the delay. if we compare with old paper (Randomized Partially-Minimal Routing:
Near-Optimal Oblivious Routingfor 3-D Mesh Networks ,page no.2092) and headfirst sliding routing concept.

POWER	AREA
3D ROUTER : 413.46mW	3D ROUTER: 626114mm$^{2.}$
HEADFIRST SLIDING ROUTING : 50.2mW	HEADFIRST SLIDING ROUTING : 184mm^{2}

Table :comparison between area and power.

VI. Headfirst Sliding Routing concept architecture



VII. MATHEMATICAL DESCRIPTION

Packets are first routed to an elevator on the source chip, moved to the destination chip, and routed to the destination. An elevator is selected so that the expected transfer time is minimized.

T is formulated as follows.

$T = RHsd + Twait$---(1)

where R is a flit transfer time at a router, Hsd is the number of hops from source to destination, and $Twait$ is an expected waiting time at an elevator.

$Twait$ is calculated as follows. First the arrival time of a packet to an elevator $Tarrive$ is calculated as follows, assuming no packet contentions.

$Tarrive = CurrentTime + RHsb$------------------ (2)

where Hsb is the number of hops from source to elevator, which is depending on the routing algorithm. The transfer start and finish times can be estimated based on this $Tarrive$.

Let $Talloc$ is a time-slot allocation time and $Tslot$ is the length of a time-slot. If a packet transfer start time is greater than or equal to $Talloc$ and a packet transfer finish time is less than $Talloc + Tslot$, $Twait$ is zero. Otherwise, $Twait$ is set to the next time-slot allocation time.

VIII. POWER AND AREA EVALUATION

In this section, we evaluate the power and area of a baseline 3-D router. To provide accurate comparisons, we implemented the baseline 3-D router with Headfirst sliding routing down to the gate level, and we used post-layout power and area results for our comparisons. In particular, the Verilog RTL implementation for the baseline 3-D router +Headfirst sliding routing was generated using modelsim SE 6.2c [-------], a fully-synthesizable parameterized router generator that implements an input-buffered pipelined virtual channel router. We consider a baseline 7-port router with 8 VCs per port, 5 flits/VC and a flit width of 16 bytes. For the 3D routers, we extended the Verilog design of the baseline router by incorporating the additional logic needed to implement Headfirst sliding routing. The router RTLs were synthesized with target version is xa3s250e-3tqg144.

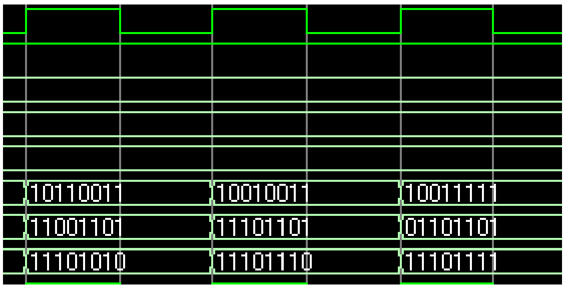
IX. EVALUATION

Port	7 input/output port
Buffer	7 flit
Routing	XYZ routing
Switching	Wormhole 2 virtual channels
Pipeline stage	Router computation, virtual channel allocation, switch allocation, switch traversal and link traversal
Flit size	128 bit

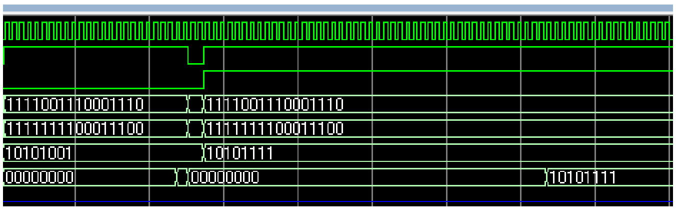
Table 1:Baseline Router

X. RESULTS

To used the virtual channels waiting time or delay would be reduced and also inter chip interconnects occupy the small chip area or reducing chip area.



And BUS not even check the error or packet loss but router would be very much helpful ERROR CHECKING and CORRECTION.



Results shows the 4X4 NOC Router. used two chips , both chips have 16 nodes. from the first chip out of 16 nodes packet used the only six nodes and from the second node out of 16 node used only five nodes. This means on the chips router routes the packet to destination used less nodes.

XI. CONCLUSION

As the demands on high performance and multi-function systems increase, vertically stacked 3DICs have the advantage is short communication distance between chips, leading to high data rate and low power consumption.In this paper, we proposed a new oblivious routing algorithm for 3-D mesh networks called RPM routing. Long interconnects are becoming an increasingly important problem from both power and performance . A wire-less approach that connects chips in vertical dimension has a great potential to customize interconnects or components in 3-D chip multiprocessors (CMPs). The Headfirst sliding routing scheme to overcome the simple static TDMA-based vertical buses.The result of the network performance showed that Headfirst Sliding Router reduces power and area , although Headfirst Sliding Router performs better than the normal 3D router at a high workload.
MH routing does not guarantee deadlock-freedom without virtual channels. To avoid deadlocks, two VCs are required for all the routers.

XII. REFERENCES

[1]. A. Sheibanyrad, F. Petrot, and A. Janstch.3D Integration for NoC-Based SoC Architectures.Springer, 2010.

[2]. H. Matsutani, M. Koibuchi, Y. Yamada, D. F. Hsu, and H. Amano. Fat HTree: A Cost-Efficient Tree-Based On-Chip Network. IEEE Transactionson Parallel and Distributed Systems, 20(8):1126– 1141, Aug. 2009.

[3]. B. Feero and P. P. Pande. Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation. IEEE Transactions on Computers, 58(1):32–45, Jan. 2009.

[4]. V. F. Pavlidis and E. G. Friedman.3-D Topologies for Networks-on-Chip.IEEE Transactions on Very Large Scale Integration Systems,15(10):1081–1090, Oct. 2007.

[5]. J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, N. Vijaykrishnan, M. Yousif, and C. Das. A Novel Dimensionally-Decomposed Router for On-Chip Communication in 3D Architectures. In Proceedings ofthe International Symposium on Computer Architecture (ISCA'07), pages 138 149, 2007

[6]. F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir. Design and Management of 3D Chip Multiprocessors UsingNetwork-in-Memory. In Proceedings of the International Symposium on Computer Architecture (ISCA'06) , pages 130–141, June 2006.

[7]. D. Park, S. Eachempati, R. Das, A. K. Mishra, V. Narayanan, Y. Xie, and C. R. Das. MIRA: A Multi-layered On-Chip Interconnect Router Architecture. In Proceedings of the International Symposium on Computer Architecture (ISCA'08), pages 251–261, 2008.

[8]. R. S. Ramanujam and B. Lin. Randomized Partially-Minimal Routing on Three-Dimensional Mesh Networks. IEEE Computer ArchitectureLetters, 7(2):37–40, July 2008.

[9]. B. Black, D. Nelson, C. Webb, and N. Samra, "3-D processing technology and its impact on IA32 microprocessors," in Proc. Int. Conf.Comput. Des., 2004, pp. 316–318.

[10]. W. R. Davis et al., "Demystifying 3-D ICs: the pros and cons of going vertical," IEEE Des. Test Comput., vol. 22, no. 6, pp. 498–510, Nov./ Dec. 2005.

[11]. M. Kawano et al., "A 3-D packaging technology for 4 gbit stacked DRAM with 3 Gbps data transfer," in IEEE Int. Electron DevicesMeeting, Dec. 2006, pp. 1–4.

[12]. K. W. Lee, T. Nakamura, T. Ono, Y. Yamada, T. Mizukusa, H. Hashimoto, K. T. Park, H. Kurino, and M. Koyanagi, "Three-dimensional shared memory fabricated using wafer stacking technology," in Int. Electron Devices Meeting Tech. Dig., 2000, pp. 165–168.

[13]. L. Xueet al., "Three-dimensional integration: technology, use, and issues for mixed-signal applications," IEEE Trans. Electron Devices, vol. 50, no. 3, pp. 601–609, Mar. 2003.

[14]. A. Agarwal, L. Bao, J. Brown, B. Edwards, M. Mattina, C.-C. Miao, C. Ramey, and D.Wentzlaff, "Tile processor: Embedded multicore for networking and multimedia," presented at the Hot Chips'19, Stanford, CA, 2007.

[15]. P. Gratz, K. Changkyu, R. McDonald, S. W. Keckler, and D. Burger, "Implementation and evaluation of on-chip network architectures," presented at the Int. Conf. Comput. Design, San Jose, CA, Oct. 2006.

[16]. M. B. Taylor, W. Lee, S. Amarasinghe, and A. Agarwal, "Scalar operand networks: On-chip interconnect for ILP in partitioned architectures," presented at the Int. Symp. High Perform. Comput.Arch., Anaheim, CA, 2003.

[17]. S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, P. Iyer, A. Singh, T. Jacob, S. Jain, S. Venkataraman, Y. Hoskote, and N. Borkar, "An 80-tile 1.28 tflops network-on-chip in 65 nm CMOS," presented at the Int. Solid-State Circuits Conf., San Francisco, CA, 2007.

[18]. T. Kgil, A. Saidi, N. Binkert, R. Dreslinski, S. Reinhardt, K. Flautner, and T.Mudge, "Picoserver: Using 3-D stacking technology to enable a compact energy efficient chip multiprocessor," in Proc. 12th Int. Conf. Archit. Support Program. Lang. Oper. Syst. (ASPLOS-XII), 2006, pp. 117–128.

[19]. F. Li, C. Nicopoulos, T. Richardson, Y. Xie, V. Narayanan, and M. Kandemir, "Design and management of 3-D chip multiprocessors using network-in-memory," in Proc. Int. Symp. Comput. Archit., 2006, pp. 130–141.

[20]. B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, D. McCaule, P. Morrow, D. W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. P. Shen, and C. Webb. Die Stacking (3D) Microarchitecture. In Proceedings of the International Symposium onMicroarchitecture (MICRO'06), pages 469–479, Dec. 2006.

[21]. K. Kumagai, C. Yang, S. Goto, T. Ikenaga, Y. Mabuchi, and K. Yoshida. System-in-Silicon Architecture and its application to an H.264/AVC motion estimation fort 1080HDTV. In Proceedings of the InternationalSolid-State Circuits Conference (ISSCC'06), pages 430–431, Feb. 2006.

[22]. 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.

[23]. K. Kanda, D. D. Antono, K. Ishida, H. Kawaguchi, T. Kuroda, and T. Sakurai.1.27-Gbps/pin, 3mW/pin Wireless Superconnect (WSC) Interface Scheme.In Proceedings of the International Solid-State Circuits Conference (ISSCC'03), pages 186–187, Feb. 2003.

[24]. N. Miura, H. Ishikuro, T. Sakurai, and T. Kuroda.A 0.14pJ/b Inductive-Coupling Inter-Chip Data Transceiver with Digitally-Controlled Precise Pulse Shaping.In Proceedings of the International Solid-State CircuitsConference (ISSCC'07), pages 358–359, Feb. 2007.

[25]. N. Miura, D. Mizoguchi, M. Inoue, K. Niitsu, Y. Nakagawa, M. Tago, M. Fukaishi, T. Sakurai, and T. Kuroda.A 1Tb/s 3W Inductive-Coupling Transceiver for Inter-Chip Clock and Data Link.In Proceedings of theInternational Solid-State Circuits Conference (ISSCC'06), pages 424– 425, Feb. 2006.

[26]. J. Burns, L. McIlrath, C. Keast, C. Lewis, A. Loomis, K. Warner, and P. Wyatt. Three-Dimensional Integrated Circuits for Low-Power High-Bandwidth Systems on a Chip.In Proceedings of the InternationalSolid-State Circuits Conference (ISSCC'01), pages 268–269, Feb. 2001.

[27]. W. J. Dally and C. L. Seitz, "Deadlock-free message routing in multiprocessor interconnection networks," IEEE Trans. Comput., vol. 36, no. 5, pp. 547–553, May 1987.

[28]. D. Seo,A.Ali,W.-T. Lim, N. Rafique, and M. Thottethodi, "Near-optimal worst-case throughput routing for two-dimensional mesh networks," presented at the Int. Symp. Comput. Arch., Madison,WI,Jun.2005.

[29]. H. Sullivan, T. R. Bashkow, andD. Klappholz, "A large scale, homogenous, fully distributed parallel machine," in Proc. 4th Annu.Symp.Comput. Archit., 1977, pp. 105–117.

[30]. L. G. Valiant and G. J. Brebner, "Universal schemes for parallel communication," inACM Symp. Theory Comput., 1981.

[31]. T. Nesson and S. L. Johnsson, "ROMM routing on mesh and torus networks," in ACM Symp. Parallel Algorithms Archit., 1995, pp. 275–287.

[32]. D. Seo,A.Ali,W.-T. Lim, N. Rafique, and M. Thottethodi, "Near-optimal worst-case throughput routing for two-dimensional mesh networks," presented at the Int. Symp. Comput. Arch., Madison, WI, Jun. 2005.

[33]. J. Duato, "A new theory of deadlock-free adaptive routing in wormhole networks," IEEE Trans. Parallel Distrib. Syst., vol. 4, no. 12, pp. 1320–1331, Dec. 1993.