Cluster mesh: a topology for three-dimensional network-on-chip

Junhui Wang[1,3], Huaxi Gu[1,3a], and Yintang Yang[2]

[1] *State Key Laboratory of ISN, Xidian University, Xi'an, China*

[2] *Institute of Microelectronics, Xidian University, Xi'an, China*

[3] *Science and Technology on Information Transmission and Dissemination in Communication Networks Lab, the 54th Institute of CETC, China*

a) *hxgu@xidian.edu.cn*

Abstract: Three-dimensional network-on-chip (3D NoC) is a promising method to overcome the bottlenecks in 3D integrated circuit (IC). Although 3D NoC can provide more efficient inter-layer communication with through-silicon vias (TSVs), the low yield and high overhead become the main challenges. To obtain a balance point between cost and performance, the cluster mesh we proposed applies a new vertical interconnects squeezing scheme which decreases the amount of TSV by sharing vertical links through vertical routers. The simulation results show that the proposed topology can improve the yield of chip, reduce the overhead and provide acceptable performance.

Keywords: network-on-chip, three-dimensional topology, interconnection network, through-silicon via

Classification: Integrated circuits

References

[1] S. Tosun, "FIT: Fast Irregular Topology generation algorithm for application specific NoCs," *IEICE Electron. Express*, vol. 7, no. 15, pp. 1132–1138, Aug. 2010.

[2] B. S. Feero and P. P. Pande, "Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation," *IEEE Trans. Comput.*, vol. 58, no. 1, pp. 32–45, Jan. 2009.

[3] V. F. Pavlidis and E. G. Friedman, "3-D Topologies for Networks-on-Chip," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 10, pp. 1081–1090, Oct. 2007.

[4] L. Cheng, L. Zhang, Y. Han, and X. Li, "Vertical interconnects squeezing in symmetric 3D mesh Network-on-Chip," *Proc. 16th Conf. Asia and South Pacific Design Automation*, Yokohama, Japan, pp. 357–362, Jan. 2011.

[5] M. Motoyoshi, "Through-Silicon Via (TSV)," *Proc. IEEE*, vol. 97, no. 1, pp. 43–48, Feb. 2009.

[6] J. H. Lau, "TSV Manufacturing Yield and Hidden Costs for 3D IC Integration," *Proc. 60th Conf. Electronic Components and Technology*, Las Vegas, USA, pp. 1031–1042, June 2010.

[7] J. Howard, S. Dighe, et al., "A 48-Core IA-32 Processor in 45 nm CMOS Using On-Die Message-Passing and DVFS for Performance and Power

Scaling," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 173–183, Jan. 2011.

[8] A. Kahng, B. Li, L. Peh, and K. Samadi, "ORION 2.0: A Power-Area Simulator for Interconnection Networks," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 1, pp. 191–196, Jan. 2012.

[9] W. Jang, O. He, J. Yang, and D. Z. Pan, "Chemical-mechanical polishing aware application-specific 3D NoC design," *Proc. Conf. International Computer-Aided Design*, San Jose, USA, pp. 207–212, Nov. 2011.

1 Introduction

Three-dimensional integrated circuit (3D IC) is emerged as a promising solution to increase both performance and functionality of system-on-chip (SoC). Meanwhile, network-on-chip (NoC) is proposed to overcome the communication bottlenecks in SoC. 3D NoC, which combines the benefits of 3D IC and NoC, is supposed to be the revolutionary methodology for on-chip design.

The performance of 3D NoC is sensitive to the topology [1], which refers to the interconnection of links and routers. 3D mesh [2] is the most popular topology in the field of 3D NoC due to its simplicity and efficient layout. However, the 7×7 router significantly increases the overhead of area and power. To overcome the disadvantages of 3D mesh and make full use of the negligible inter-layer distance, stacked mesh [2] which applies bus instead of point-to-point link in the vertical direction was proposed. Meanwhile, the 7×7 routers are replaced by 6×6 routers. However, since concurrent communication is prohibited in bus, the performance decreases when contention appears. Most of the existing 3D NoC architectures [2, 3, 4] are on the basics of the mentioned two topologies.

2 Motivations

With abundant through-silicon vias (TSVs) [5], 3D mesh and stacked mesh enable high bandwidth and low power inter-layer communication and show excellent network performance [2]. However, owing to the immaturity of TSV technology, the manufacturing cost increases exponentially with the increase of TSV number [6]. In additional, the negative impact of TSV pads cannot be ignored since they occupy significant part of area and increase the difficulty in wiring. Since the yield of TSV is low, the cost and performance are opposite to each other. Therefore, it is necessary to find an efficient way to reduce the amount of TSV and provide acceptable performance in the same time.

Vertical interconnects serialization and squeezing are two schemes to reduce the amount of TSV. Since serial/parallel and parallel/serial converters increase the complexity and area of router, serialization is not an ideal strategy. Therefore, squeezing is the better choice for 3D topology with low density of TSV. Although the squeezing scheme proposed in [4] decreases the amount of TSV, the packet has to reserve all of the TSV sharing logics between current and destination layers before it can be sent. In additional,

the router has to add a TSV arbiter module due to the existence of TSV sharing logic. To simplify the design of router and reduce the manufacturing cost of 3D NoC, this paper proposes a new topology which utilizes the restricted TSVs in a more efficient way.

3 Cluster mesh

Choosing topology is the most important step in the design of 3D NoC since both the manufacturing cost and network performance depend heavily on the amount of TSV in the topology. The more TSVs a topology applies, the better performance it can provide. However, the severe economic penalty brings by the low yield of TSV should be taken seriously. To obtain a balance point between cost and performance, the cluster mesh we proposed applies a new vertical interconnects squeezing scheme which decreases the amount of TSV by sharing vertical links through an extra router. Meanwhile, the added router is accelerated to guarantee the performance.

3.1 Topology

Unlike 3D mesh which provides a vertical link which consists of a bundle of TSVs for each router, cluster mesh shares the vertical links to reduce the amount of TSV. As shown in Fig. 1.(a), the horizontal routers in the same layer are connected to each other through horizontal links in the term of 2D mesh. Then every four horizontal routers and a vertical router compose a cluster and only the vertical router which locates in the middle of the cluster is connected to the vertical link. The inter-layer connections are realized through the vertical routers and the shared vertical links.
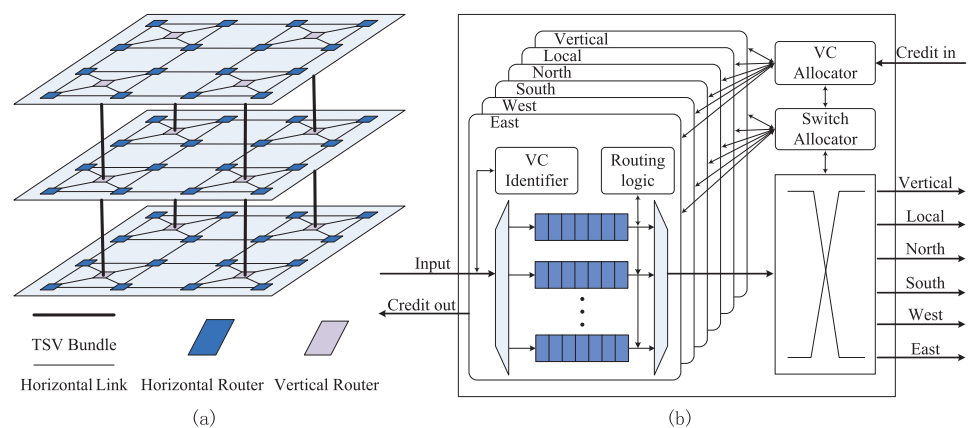

Fig. 1. 4×4×3 cluster mesh (a) and horizontal router (b).

Compared with 3D mesh, cluster mesh eliminates 75% amount of TSV by sharing vertical links through the vertical routers. The most direct benefits are the increase in yield and the decrease in area. To exploit the beneficial attribute of TSV and guarantee the performance, the vertical routers are accelerated to work in a higher frequency through dynamic voltage frequency scaling technology [7].

3.2　Router architecture

The routers applied in cluster mesh are divided into horizontal routers and vertical routers. Horizontal routers are utilized to realize intra-layer communication while vertical routers for inter-layer communication.

As shown in Fig. 1.(b), horizontal router is a typical 6×6 router which consists of six input ports, a routing logic module, a virtual channel (VC) allocator, a switch allocator and a 6×6 crossbar. According to the routing algorithm, routing logic module returns an output port for the head flit in the input port. VC allocator arranges an available VC in the corresponding downstream node for the head flit which has already been given an output port. Switch allocator reserves an idle path for the flit to go through the crossbar and leave the router.

The architecture of vertical router is the same as that of the horizontal one. The only difference between them is where the routers are connected to. For the horizontal router, its neighboring nodes include intellectual property (IP) core, horizontal routers and vertical router in the same layer. For the vertical router, its neighboring nodes only include horizontal routers in the same layer and vertical routers in other layers.

3.3　Inter-layer communication

Since only vertical routers are connected to vertical links, inter-layer communication can just realize by means of the vertical router in the cluster. Therefore, corresponding modifications should be made for the routing algorithm in inter-layer communication. For the dimension order routing algorithm we applied, the packet is sent in X-dimension, Y-dimension and Z-dimension successively. The packet changes the direction to the next dimension only when it arrives at the router which has the same coordinate with the destination node in current dimension. The modification is that when the packet arrives at the horizontal router which has the same X-coordinate and Y-coordinate with the destination node, the packet should be sent to the vertical router which the horizontal router connected to. Then, through the vertical router, the packet realizes inter-layer communication.

4　Simulation results

To evaluate the cost and performance of the cluster mesh, a cycle-accurate 3D NoC simulator is developed with OPNET. In this experiment, each IP core generates fixed-length packets based on exponential distribution. Dimension order routing algorithm is chosen to forward the packet. Energy and area parameters are obtained from the power and area simulator Orion 2.0 [8]. To be fair, the cost and performance of 2D mesh, 3D mesh, stacked mesh and cluster mesh in the scale of 48 IP cores are compared in the simulation.

4.1　Costs comparison

The costs of 3D chip in this paper include area, power and yield. As shown in Fig. 2.(a), the area consists of TSV pad, link and router. The area of TSV

pad is obtained from the equation: $A_{pad} = p^2 \times N_{TSV}$ [9], where A_{pad} means the area of TSV pad, p which obtained from TSV height variation [9] means the pitch of TSV and N_{TSV} means the amount of TSV in a TSV bundle. Meanwhile, the area of link and router are obtained from Orion 2.0. The result shows that the area of cluster mesh decreased by 10.89% compared with 3D mesh. The main reasons are that the overhead of TSV pads has a 75% reduction and the area of 6×6 router is only 74% of 7×7 router. The accelerated cluster mesh (A-cluster mesh) occupies the same area as cluster mesh does since that the only change of accelerated vertical router is the supply voltage and the architecture of router keeps intact.
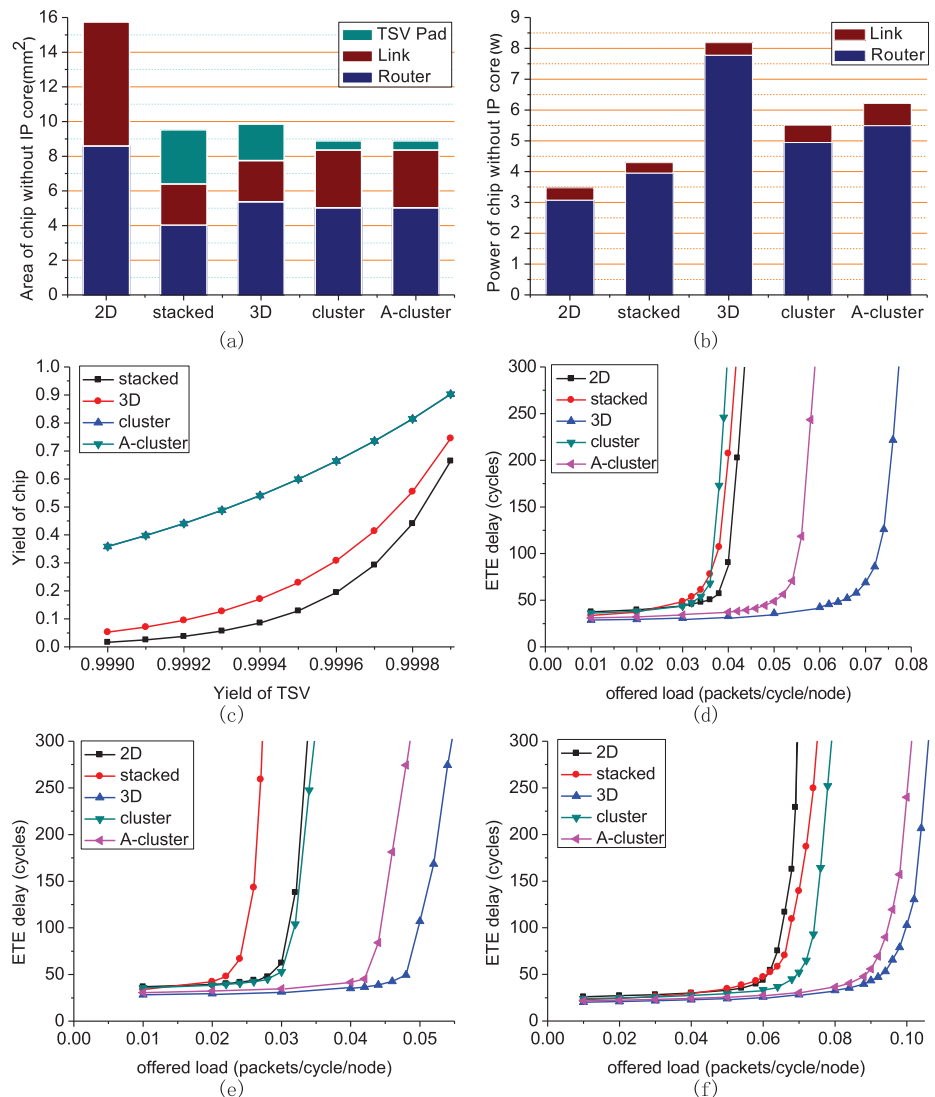

Fig. 2. Area (a), power (b) and yield (c). ETE delay under (d), hotspot (e) and local (f) traffic pattern.

The power which obtained from Orion 2.0 are illustrated in Fig. 2.(b). Since the power consumed by 7×7 router is approximately two times more than that of 6×6 router, the power of cluster mesh has reduced by 32.7%. A-cluster mesh only has a 24.1% reduction since that its vertical routers are

working in 2 GHz which is four times of cluster mesh.

Fig. 2.(c) shows the relationship between the yield of TSV and the yield of 3D chip. Since the TSV height variations of point-to-point link and bus are equal, the yield of a TSV bundle is only determined by the amount of TSV it includes. Therefore, the yield of chip can be obtained from the equation: $Y_C = Y_b^{N_b \times (N_l - 1)}$, where Y_C means the yield of chip, $Y_b = Y_{TSV}^{N_{TSV}}$ means the yield of TSV bundle, Y_{TSV} means the yield of TSV, N_b means the number of TSV bundle in each layer and N_l means the number of layer. Since the yield of TSV is the only considered factor, the curve of cluster mesh and A-cluster mesh are equal. It is clear that the improvement in chip yield brought by reducing the amount of TSV is notable. Unless the yield of TSV is higher than 99.99%, the amount of TSV will be a huge impact factor of 3D NoC.

4.2 Performance comparison

Fig. 2.(d)-(e) illustrate the end-to-end (ETE) delay under uniform, hotspot and local traffic pattern, respectively. For the uniform traffic, cluster mesh performs worse since the evenness of uniform traffic increases the competition probability in the shared vertical links. For the hotspot and local traffic which are more similar to realistic traces, the saturation points of cluster mesh are about 65.4% and 75% of 3D mesh. Meanwhile, after accelerating the vertical routers, the saturation points of A-cluster mesh are improved to 88.9% and 96.2% of 3D mesh. The accelerated routers make sure that the performance is acceptable.

5 Conclusion

Cluster mesh which squeezes the amount of TSV by adding an additional vertical router is proposed in this paper. Unlike introducing new logic module or redesigning router, cluster mesh adopts five typical 6×6 routers which are combined into a cluster. The four horizontal routers are utilized to realize intra-layer communication while the vertical router plays the role of inter-layer correspondent with shared vertical link. To compensate the decrease of inter-layer bandwidth, the vertical routers are accelerated to work in a higher frequency. The simulation results illustrate that cluster mesh can reduce the cost efficiently and provide acceptable performance.

Acknowledgments