Postprint: Regular Tetrahedron Fission Topology for Three-Dimensional Network-on-Chip

Authors: Zheng Yazhen, Dakun Zhang

Date: 2018-05-20T00:00:00+00:00

Abstract

This study investigates a novel regular tetrahedron fission topology for three-dimensional Network-on-Chip (3D NoC), presenting its generation process along with encoding and routing designs. Performance simulation experiments are conducted on the regular tetrahedron fission topology by extending the gpNoCsim NoC simulator to three dimensions. The simulation results demonstrate that under uniform traffic pattern, both the average latency and average hop count of the regular tetrahedron fission topology are lower than those of the Mesh topology; at an injection rate of 0.02, the average latency is 16.8% lower and the average hop count is 5.5% less compared to the Mesh topology. Under localized traffic pattern, when the injection rate exceeds 0.008, both metrics exhibit significant improvement over the Mesh topology; at an injection rate of 0.014, the average latency is reduced by 18.7% and the average hop count is reduced by 9.6% compared to the Mesh topology. These findings indicate that the regular tetrahedron fission topology is viable for 3D NoC topology design.

Full Text

Preamble

Title: Research on Tetrahedron Fission Topology in Three-Dimensional Network-on-Chip

Authors: Zheng Yazhen, Zhang Dakun
Affiliation: School of Computer Science & Software Engineering, Tianjin Polytechnic University, Tianjin 300387, China

Abstract: This paper investigates a novel three-dimensional Network-on-Chip (3D NoC) topology called tetrahedron fission topology, presenting its generation process along with corresponding encoding and routing designs. Through a three-dimensional extension of the gpNoCsim simulator, performance simulation experiments were conducted on the tetrahedron fission topology. Simulation results demonstrate that under uniform traffic patterns, the tetrahedron fission topology achieves lower average latency and fewer average hops compared to Mesh topology. At an injection rate of 0.02, the average latency decreases by 16.8% and the average hops reduce by 5.5% compared to Mesh. Under localized traffic patterns, when the injection rate exceeds 0.008, both average latency and average hops show significant improvement over Mesh. At an injection rate of 0.014, the average latency decreases by 18.7% and the average hops reduce by 9.6% compared to Mesh. These results indicate that the tetrahedron fission topology is viable for 3D NoC topology design.

Keywords: three-dimensional network on chip; tetrahedron fission; topology; average latency; average hops

0 Introduction

With the development of large-scale integrated circuit technology, System-on-Chip (SoC) and two-dimensional Network-on-Chip (2D NoC) have emerged successively. As 2D NoC scales continue to increase, they have reached bottlenecks in area, power consumption, layout and routing, and packaging density [1], leading to the emergence of three-dimensional Network-on-Chip (3D NoC), which has become a research hotspot [2–5]. In 3D NoC research, topology structure is one of the key issues [6]. 3D NoC topologies include regular and irregular categories. While numerous studies have analyzed performance based on classical regular topologies such as 3D Mesh [7] and 3D Torus [8], research on novel irregular 3D NoC topologies remains limited. Compared to regular topologies, irregular topologies can be customized for specific application requirements and demonstrate promising prospects in various domains, with several advances already achieved. Examples include the honeycomb 3D NoC topology proposed by Yin et al. from the University of Turku [9], the three-dimensional hypercube NoC topology proposed by He Xu et al. from Hunan University [10], and research by Liu Youyao et al. from Xidian University on NoC topology structures and communication methods, which summarized typical 3D NoC architectures and proposed three novel architectures including a 3D hypercube dual-ring architecture, a Petersen graph architecture with 3D Torus connections, and a 3D rectangular twisted torus mesh architecture [11].

This paper focuses on a novel irregular 3D NoC topology, proposing the tetrahedron fission topology (originating from National Natural Science Foundation project: 61272006). This topology offers advantages including constant node degree (always 3), good symmetry, and simple routing. Moreover, it transforms network expansion from an "additive" approach to a "fission" approach, enabling personalized 3D NoC topology design. This research is exploratory and innovative.

1 Tetrahedron Fission Topology and Routing Design

The tetrahedron fission topology is generated by successively completing the fission and interconnection of each node in [Figure 1: see original paper] according to the fission process described above. This structure contains 12 network nodes, each with a degree of 3, and forms 4 fission clusters labeled A, B, C, and D. A ball-and-stick model of the tetrahedron fission topology is shown in [Figure 3: see original paper].

1.1 Tetrahedron Fission Topology Generation

A regular tetrahedron is one of the Platonic solids [12]. The tetrahedron fission topology is derived from the "fission" of a regular tetrahedron, with an annotated tetrahedron model shown in Figure 1. The first-level fission process for a tetrahedron node proceeds as follows: select any node No0 from Figure 1, move along its three connected edges for a certain distance to generate three new nodes No00, No01, and No02, then connect these three nodes sequentially (the three interconnected nodes are called a fission cluster) and delete No0, thus completing the fission of No0. This fission process for node No0 is illustrated in [Figure 2: see original paper], where circles represent network nodes. A typical network node contains a router and several IP cores, which may be processors, memory units, or other functional devices, with IP cores connecting to the router through local ports.

1.2 Node Encoding and Routing

1) Node Encoding

A reasonable node encoding scheme can simplify routing protocol complexity, improve network performance, and reduce latency. Since the tetrahedron fission topology is derived from a regular tetrahedron, the fission clusters A, B, C, and D must be encoded for convenient addressing. The structure contains 12 network nodes requiring 4 bits for encoding. Each node' s code is generated by shifting its cluster code left by two bits and adding the node' s internal cluster number, as detailed in .

2) Router Structure Design

In the tetrahedron fission topology, each network node' s router has three ports connecting to adjacent network nodes and several local ports. In this design, each node hosts 4 IP cores, thus requiring 4 local ports, with one end connected to the router and the other to local IP cores. Each port includes input and output channels, with each channel able to receive or transmit data packets from connected routers or IP cores. Each physical channel is divided into multiple virtual channels, typically 2, 4, or 8; this design selects 4 to effectively avoid deadlock. The router includes routing logic, a switch, and virtual channel arbitration units, sending received data packets to the correct output channel through routing computation, with physical channels shared among virtual channels using a round-robin approach.

3) Routing Algorithm Design

The tetrahedron fission topology contains 12 network nodes and 4 fission clusters. The routing algorithm employs a hierarchical approach due to the structure' s hierarchical characteristics, where *dest* represents the destination node, *curr* represents the current node, *a* determines whether the packet has reached the destination' s fission cluster, and *result* determines whether the packet has reached the destination node. The routing process is as follows:

a) The data packet contains the destination node address *dest*. When a node *curr* in the tetrahedron fission topology receives a packet, it calculates $result = curr \quad dest$, where represents XOR operation, then proceeds to step b).

b) Evaluate the *result* value. If *result* is not 0, proceed to step d); otherwise, proceed to step f).

c) Calculate $currA = curr/4$, $destA = dest/4$, $currB = curr \ mod \ 4$, $destB = dest \ mod \ 4$, $a = currA \quad destA$, and evaluate a. If a is not 0, proceed to step d); otherwise, proceed to step e).

d) Based on the values of *currA*, *destA*, and *destB*, forward the packet to an adjacent fission cluster or to the node nearest to the destination cluster, update the current node to *curr*, and return to step a).

e) The destination and current nodes are in the same fission cluster. Then evaluate *currB* and *destB*, forward the packet to the corresponding node, update the current node to *curr*, and return to step a).

f) The packet has reached the destination node; forward it to the corresponding IP core.

2 Simulation Experiments and Performance Analysis

2.1 Simulation Experiment Design

1) Simulation Platform and Experimental Scheme

This experiment employs the gpNoCsim [13] NoC simulator developed by Hossain et al. from the University of Chester, extending it to three dimensions to implement encoding for the tetrahedron fission topology. The tetrahedron fission topology is compared with 2D Mesh under equivalent network scale through simulation experiments, evaluating performance metrics including throughput, average latency, and average hop count to verify the superiority of the tetrahedron fission topology. The tetrahedron fission topology contains 12 network nodes, so the 2D Mesh network is configured as 4×3, also with 12 network nodes. Each network node' s router connects to 4 IP cores through local ports, so both architectures host 48 IP cores. In the experiments, 2D Mesh employs XY routing algorithm, while the tetrahedron fission topology uses the routing algorithm designed in Section 1.2. Simulations are conducted under both uniform and localized traffic patterns. Under uniform traffic, network traffic is evenly distributed with equal probability of each node receiving packets. Under localized traffic, 70% of traffic is confined within 4 IP cores of a cluster, while 30% is randomly sent to other nodes outside the cluster.

2) Parameter Configuration
All required parameters for the gpNoCsim simulator are configured as follows: number of virtual channels per physical channel is 4, each virtual channel can store 4 flits, message length is 200 bytes, flit length is 150 bits, each simulation runs for 20,000 clock cycles with a warm-up period of the first 10% cycles. To ensure precision, the simulator runs 20 times under identical parameter configurations, with arithmetic averages used as output results.

2.2 Throughput Comparison Analysis

1) Uniform Traffic Pattern Throughput Analysis
Simulation experiments under uniform traffic pattern yielded throughput data for both topologies. Analysis shows that when injection rate is below 0.016, the network is uncongested and both topologies exhibit essentially identical throughput. As packet injection rate continues to increase, the network becomes progressively busier. For the tetrahedron fission topology, saturation occurs at injection rate 0.018 with saturation throughput of 0.17. For 2D Mesh, when the tetrahedron fission topology saturates, the 2D Mesh throughput is 0.18. The tetrahedron fission topology’ s saturation throughput is 5.6% lower than 2D Mesh. The relationship between throughput and injection rate for both topologies under uniform traffic is shown in [Figure 4: see original paper].

2) Localized Traffic Pattern Throughput Analysis
Under localized traffic pattern, analysis of experimental data shows that when injection rate is below 0.004, both topologies exhibit essentially identical throughput. As injection rate increases, network traffic grows and both topologies successively reach saturation. For the tetrahedron fission topology, saturation occurs at injection rate 0.008 with throughput of 0.39. For 2D Mesh, saturation also occurs at injection rate 0.008 with throughput of 0.41. In saturation, the tetrahedron fission topology’ s throughput is 4.9% lower than 2D Mesh. However, after saturation, as injection rate increases, 2D Mesh throughput declines significantly compared to the tetrahedron fission topology. At injection rate 0.01, the tetrahedron fission topology throughput is 6.1% higher than 2D Mesh; at injection rate 0.012, it is 16.7% higher. The relationship between throughput and injection rate under localized traffic is shown in [Figure 5: see original paper].

2.3 Average Latency Comparison Analysis

1) Uniform Traffic Pattern Average Latency Analysis
Experimental data demonstrates that under uniform traffic pattern and vari-

ous injection rates, the tetrahedron fission topology consistently achieves lower average latency than 2D Mesh, though both increase with injection rate. At injection rate 0.016, the tetrahedron fission topology' s average latency is 7.2% lower than 2D Mesh; at 0.018, it is 11.0% lower; at 0.02, it is 16.8% lower. The relationship between average latency and injection rate under uniform traffic is shown in [Figure 6: see original paper].

2) Localized Traffic Pattern Average Latency Analysis
Under localized traffic pattern and low injection rates, both topologies exhibit low average latency. As injection rate increases, average latency grows for both. When injection rate is below 0.008, the tetrahedron fission topology's average latency increases slightly more than 2D Mesh. When injection rate exceeds 0.008, 2D Mesh average latency increases significantly more than the tetrahedron fission topology, because 2D Mesh' s outer nodes have smaller degree and are more prone to congestion at higher injection rates, resulting in noticeably increased packet waiting times. At higher injection rates, average latency under localized traffic exceeds that under uniform traffic for both topologies, as network congestion-induced delays far outweigh time saved by reduced hop counts. At injection rate 0.01, the tetrahedron fission topology' s average latency is 19.1% lower than 2D Mesh; at 0.014, it is 18.7% lower. The relationship between average latency and injection rate under localized traffic is shown in [Figure 7: see original paper].

2.4 Average Hop Count Comparison Analysis

1) Uniform Traffic Pattern Average Hop Count Analysis
Under uniform traffic pattern and various injection rates, experimental data for total packets received and average hop count for both topologies is summarized in . The results show that under this traffic pattern, the tetrahedron fission topology' s average hop count is approximately 5.5% lower than 2D Mesh.

2) Localized Traffic Pattern Average Hop Count Analysis
Under localized traffic pattern and various injection rates, experimental data for total packets received and average hop count for both topologies is summarized in . The results show that under this traffic pattern, the tetrahedron fission topology' s average hop count is approximately 9.6% lower than 2D Mesh. Both topologies exhibit lower average hop counts under localized traffic than under uniform traffic.

---

3 Conclusion

Currently, 3D NoC technology represents a crucial solution to communication bottlenecks arising from increasing SoC integration density, with topology structure being an important research branch. The tetrahedron fission structure is an irregular topology. This paper conducted performance comparison analysis between this topology and 2D Mesh under equivalent network scale and identical experimental conditions, focusing on three metrics: throughput, average latency, and average hop count. Experimental results indicate that under uniform traffic pattern and various injection rates, the tetrahedron fission topology achieves lower average latency and fewer average hops compared to 2D Mesh. Under localized traffic pattern and injection rates above 0.008, the tetrahedron fission topology demonstrates lower average latency than 2D Mesh, and exhibits fewer average hops across all injection rates. In terms of throughput, both topologies show similar performance. These experiments demonstrate that the proposed tetrahedron fission topology is feasible for 3D NoC topology design. Current work has focused on simulator implementation and performance analysis; future work will emphasize physical structure design and implementation.

References

[1] Zhang Dakun, Song Guozhi, Lin Huazhou, et al. Double improved genetic algorithm and low power task mapping in 3D Networks-on-chip [J]. Journal of Computer Research and Development, 2016, 53 (4): 921-931.

[2] Dang K N, Meyer M, Okuyama Y, et al. A low-overhead softhard fault-tolerant architecture, design and management scheme for reliable high-performance many-core 3D-NoC systems [J]. Journal of Supercomputing, 2017: 1-25.

[3] Lee D, Das S, Pande P P. Performance-thermal trade-offs for a VFI-enabled 3D NoC architecture [C]// Proc of International Symposium on Quality Electronic Design. 2017: 271-276.

[4] Rezaei S H S, Modarressi M, Daneshtalab M, et al. A Three Dimensional Networks-on-Chip architecture with dynamic buffer sharing [C]// Proc of Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2016: 771-776.

[5] Eghbal A, Yaghini P M, Bagherzadeh N, et al. Analytical fault tolerance assessment and metrics for TSV-Based 3D Network-on-Chip [J]. IEEE Trans on Computers, 2015, 64 (12): 3591-3604.

[6] 张大坤, 宋国治, 王莲莲, 等. 三维片上网络拓扑结构研究综述 [J]. 计算机科学与探索, 2015, 9 (2): 129-164.

[7] Feero B S, Pande P P. Networks-on-chip in a three-dimensional environment: a performance evaluation [J]. IEEE Trans on Computers, 2009, 58 (1): 32-45.

[8] Jiao Jiajia, Fu Yuzhuo, Liu Ting, et al. Performance analysis and optimization for homogenous multi-core system based on 3D torus network on chip [C]// Proc of Newcas Conference. 2010: 313-316.

[9] Yin A W, Xu T C, Liljeberg P, et al. Explorations of honeycomb topologies for network-on-chip [C]// Proc of the 6th IFIP International Conference on Network and Parallel Computing. 2009: 73-79.

[10] He Xu, You Zhiqiang, Kuang Jishun. Performance analysis of three-dimensional hypercube network topology [OL]. China Science and Technology Papers Online, 2011.

[11] 刘有耀, 韩俊刚. 片上网络拓扑结构与通信方法研究 [D]. 西安: 西安电子科技大学, 2009.

[12] 张大坤, 王光兴. 基于群论的柏拉图立体着色方案三维模型构造 [J]. 软件学报, 2004, 15 (2): 292-299.

[13] Hossain H, Ahmed M, Al-Nayeem A, et al. Gpnocsim: a general purpose simulator for network-on-chip [C]// Proc of International Conference on Information and Communication Technology. 2007: 254-257.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.