

The algorithm based on network routing on chip

Chenyang Feng

College of Communication Engineering, Xidian University, Xi 'an, Shaanxi, 710000, China

20012100075@stu.xidian.edu.cn

Abstract. Network-on-Chip (NoC) is a novel communication architecture specifically designed for multi-core System-on-Chip (SoC) integration. Unlike traditional shared bus communication architectures, NoCs offer a promising solution to address issues related to latency, communication performance bottlenecks, and design efficiency. The effectiveness of NoC architecture depends on various factors such as the routing algorithm, switching methods, and network topology. This study focuses on analyzing and summarizing the routing algorithms employed in NoCs. It explores different routing approaches, including independent routing, adaptive routing, and combined independent and adaptive routing. Furthermore, evaluations are conducted to assess the performance of these algorithms concerning congestion awareness, fault tolerance, deadlock prevention, and lock-free operation. By investigating and evaluating various routing strategies, this research aims to enhance our understanding of the intricacies involved in designing efficient NoC architectures. Ultimately, the findings of this study contribute to the advancement of communication protocols in multi-core SoC systems, paving the way for improved performance and scalability in future chip designs.

Keywords: network on chip, routing algorithm, independent routing, algorithm analysis.

1. Introduction

With the development of semiconductor technology, system-on-chip (SoC) technology is widely used in various fields. A series of core tasks of the system on chip need to be realized through certain topologies and interconnection. In a small-scale system, bus interconnection is usually used. Bus contention is however simple to happen when linked nodes grow in quantity. As a result, only two nodes can communicate with each other at a time, resulting in a small system communication bandwidth. The crossbar switch matrix is also a common interconnection mechanism, which realizes the direct interconnection of every two nodes of the network on the chip. It achieves low delay and gaucho vomiting rate, but it is easy to produce the waste of communication channels [1].

The main research contents of NoC include topological structure, switching mechanism, routing algorithm, quality of service, congestion control, and router structure, among which the routing algorithm of network on the chip is one of the key technologies. The routing algorithm finds the path of the packet and is responsible for sending the packet correctly to the destination node. Due to space and power constraints on chips where NoC is implemented, the following criteria must be satisfied while developing the NoC routing algorithm. [2].

- (1) The selected routing algorithm must be able to correctly send packets to the destination node, that is, the algorithm must have no deadlocks or live locks.
- (2) The calculation of routing information must be simple. Only through simple computing, the corresponding hardware facilities can be faster and smaller.
- (3) Due to insufficient network resources, the shortest path should be chosen as far as possible, or the detour routing algorithm should limit the number of detours.
- (4) The routing algorithm should treat all NoC nodes equally.

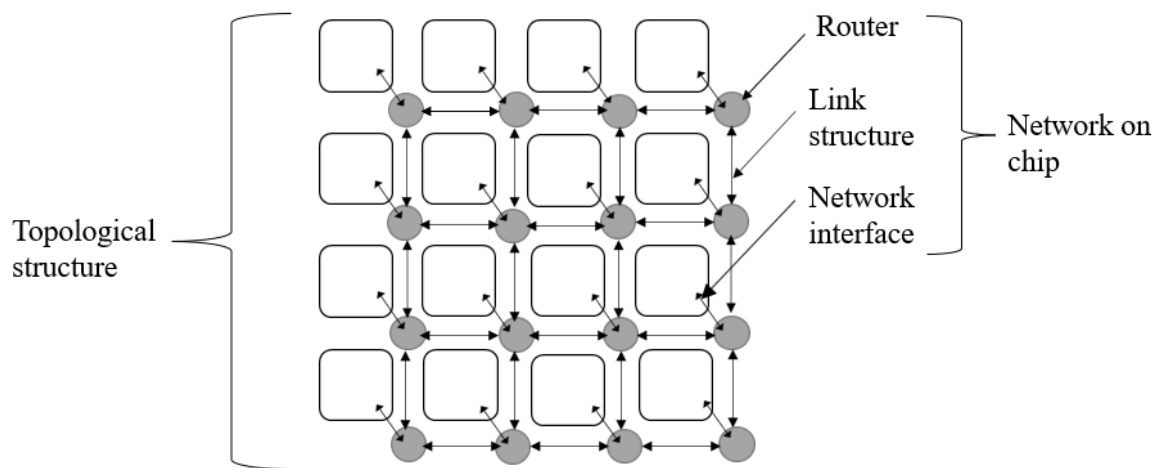


Figure 1. Concept of network-on-chip.

Unicast routing algorithms can be classified as independent, adaptive, or a combination of independent and adaptive routing based on various routing decision-making techniques. Independent routing denotes a fixed routing choice that is unaffected by network conditions. Adaptive routing enables the routing transmission to automatically adjust route selection based on different network status, thus effectively avoiding hotspots, congestion, or faults. This paper introduces several typical routing algorithms classified by route decision mode.

2. Classification of network routing algorithms on chip

Routing algorithms can be categorized initially based on the number of destinations or the location of where routing choices are made. Prior to the message being injected into the network, the source node's central controller can theoretically decide on routing. This is called source routing. Packets can also be sent over the network in a distributed manner known as distributed routing. A hybrid approach is also possible, called polyphase routing.

Different methods can be used to implement routing algorithms. The finite-state machine-based lookup of the routing table or the hardware- or software-based execution of the routing algorithm are the two methods that have been suggested as being the most effective so far. Routing methods can be either deterministic or adaptive in both situations. While adaptive routing algorithms can avoid congested or fault regions in the network based on network traffic and channel condition information, deterministic routing algorithms always offer the same path between a given set of source and destination nodes.

Depending on the selection of the shortest route, the routing algorithms can be divided into the shortest route and the bypass route. The shortest routing algorithm selects only the channel closest to the destination, while bypass routing provides the channel farther away from the destination.

2.1. Irrelevant routing algorithm

Independent routing refers to a class of routing algorithms whose routing decisions are independent of the network state. But its choices are not necessarily certain. For example, the routing table may contain multiple output channels for a destination address. You can select them randomly, iteratively, or in other ways. At present, the commonly used independent routing algorithms are: deterministic routing, random routing algorithm, and so on.

2.1.1. Deterministic routing algorithm. The examples below show how to format different combinations of numbers/callouts. The deterministic routing algorithm designates a fixed path between node A and node B that is independent of network conditions. [3]. When using deterministic routing algorithms, packets are transmitted across the network following a preset fixed path, such as the shortest path algorithm.

XY routing is a prevalent method of dimensional routing used in two-dimensional on-chip network architecture. This algorithm is responsible for establishing the optimal path for data packets between the source node and the destination node in the network. In general, routers that provide brief responses use the XY dimension routing algorithm that employs a deterministic routing method. If the source and destination addresses are provided, a definite routing path can be established.

For multi-dimensional on-chip network topologies, E-cube routing is common. E-cube routing is similar to XY routing in that both routes are routed in one direction (dimension) and then in the other direction (dimension). Specifically, as mentioned earlier, each node in an N-dimensional cube is represented by an n bit binary number. Each node has n output channels, among which the i channel corresponds to the i dimension.

One benefit of using routing is that it involves a straightforward algorithm and produces minimal delay when deployed in a network with low congestion. However, because it cannot respond to dynamic network status changes, its performance deteriorates rapidly when network congestion increases.

2.1.2. Random routing algorithm. In this method, the node that receives the packet selects a random outlet node for the packet among all its neighboring nodes. In order to avoid the problem that the destination node cannot be reached or the arrival time is too long caused by random routing, the maximum routing time is usually specified in the implementation. If no packet is received by the destination node after this time, the packet is discarded. Although the method is simple and reliable, the actual route is not the best route, which increases the unnecessary burden, and the packet transmission delay is unpredictable. Common random routing algorithms include the following three types.

(1) Flood routing algorithm

A packet is used to send the message to the neighbor, and then it is forwarded to another neighbor. The receiver of the broadcast packet does not forward it to the sender, and each node forwards the same broadcast at most once. After the destination node receives the first replication packet, other replication packets are discarded. Based on this idea, flood routing includes direct flooding and probabilistic flooding.

The process of probabilistic flooding involves sending replication packets to all links that have a possibility of connecting to the desired node based on probability. Prior to this, direct flood path preprocessing is used to estimate the general direction in which the packet should be sent towards the destination node, thereby increasing the likelihood of successfully reaching the node by sending the replication packet to the most likely path. Compared with the probabilistic flooding algorithm, the direct flooding algorithm has higher fault tolerance and less network resource occupancy.

(2) Valiant routing algorithm

Valiant Routing algorithm [4]. Because irrelevant routes will generate unbalanced traffic in the network in practical applications [5], packet transmission paths should be further constrained to alleviate network congestion under high loads.

In the Valiant algorithm, routing is divided into two steps: the intermediate destination is the destination of the first route, and the source node follows an unnecessary path to reach the final

destination node. This algorithm converts the original routing problem into two routing problems, which better balances the network load and reduces the congestion probability. The inclusion of random relay nodes can alter the transmission path, which means it may not always be the shortest possible route linking the source and target nodes. As a result, this may cause longer transmission delays.

(3) IVAL routing algorithm

This algorithm is an improvement on Valiant random routing algorithm. The initial stage of the IVAL algorithm involves selecting an intermediate node at random for the packet to be routed to, while the subsequent stage entails routing the packet from the intermediate node to its final destination. The two dimensions are in reverse order and both adopt the dimension order XY routing algorithm. Compared with other routing methods, the IVAL routing algorithm can ensure the throughput in the worst case and improve the local optimization of the network.

2.2. Adaptive routing algorithm

The so-called adaptive routing means that the routing path of the packet is not only dependent on the starting and ending addresses but also takes into account the state of the network. That is, packets with the same start/end address may have different routing paths in different network states. The advantage of adaptive routing is that it adopts the path of adaptive routing to avoid network congestion and obtain a higher network bandwidth saturation value. However, its routing logic is more complex, the cost is larger under the condition of low network security, and there are deadlock problems. Classic adaptive routing includes the turning model, parity turning model, and so on.

2.2.1. Routing algorithm of turning model. Turning model IR is a classical deadlock-free routing algorithm in Mesh networks. The basic idea is to analyze the possible turns of information groups in the network and the rings that these turns may form, and break the ring dependence of the channel by prohibiting a certain number of turns, so as to break the conditions for the formation of deadlock[6]. For a two-dimensional Mesh network, combined with the experimental data, the author shows that the three turning model-based algorithms, West-First, North-Last and Negative-First, perform well in practical application.

2.2.2. Parity turning model algorithm Parity turning model is an improved algorithm of turning model, which is mainly applied in 2D Mesh topology network. Instead of preventing deadlocks by restricting turns in certain directions, as the turning model does, the position where the turn occurs is constrained [7]. The parity steering model constrains the position of the turn. There are concepts of ES, EN, WS, WN, SE, SW, NE and NW steering. The parity turning model places varying limitations on the turning of routing for different nodes, depending on the network position of the node and whether its row number and column number are even or odd.

Some turns forbidden by the parity turn model : (1) Nodes located in odd columns are where NW and SW turns take place. (2) Even column nodes are where turns towards EN and ES directions occur. (3)180° steering (also known as U-turn)

2.2.3. XY-YX routing algorithm. The XY-YX routing algorithm represents an enhancement of the XY routing algorithm. The fundamental concept is that data may not follow a linear path from X to Y, but instead, be determined by the current node and the destination node's location [8]. If the Y-coordinate of the node being currently accessed is higher than the Y-coordinate of the destination node, indicating that the current node is situated towards the north of the destination node, then the transmission of the packet is performed by employing XY routing. If the Y-coordinate value of the current node is less than that of the destination node, which indicates that the current node is situated to the south of the destination node, then the packet's routing is carried out via the YX method, resulting in reduced transmission latency.

2.2.4. Skip fault-tolerant routing algorithm. Skip step fault-tolerant routing algorithm [9] is an improvement of the turning model, which is suitable for Mesh network structure and has the advantage of simple implementation. It can take into account the characteristics of network throughput and delay.

In this algorithm, each node needs to know the link state of its adjacent points, which requires each node to maintain a link state table, and real-time update to ensure that each routing can be carried out correctly. When a negative route encounters a fault or a congested link, it performs a skip step in the vertical running direction and uses the escape virtual channel to bypass the fault area or hotspot area. In order to avoid deadlock caused by a loop turn, the destination node should initially move towards the west if it is located to the left of the source node. Because the routing function restricts the direction of the turn and places an escape virtual channel in the direction of the restricted turn, no deadlock loop is formed, thus avoiding deadlock.

2.2.5. DyXY routing algorithm. In the DyXY routing algorithm [10], nodes select the next hop of the packet route by monitoring the congestion of their neighboring nodes. In this algorithm, the length of the real-time queue of each port on the router is measured and serves as an indication of the pressure. This pressure value is then utilized to represent the congestion state of neighboring nodes. According to the algorithm, the shortest path is used for data transmission. If multiple shortest paths exist, the path with the lowest congestion pressure is selected.

2.2.6. NoC-LS routing algorithm. Ali et al. proposed the The NoC-LS Routing algorithm implements a combination of the Distance Vector Routing and Link State Routing methodologies. [11] Calculates the transmission overhead between nodes and saves the results in a table. The system will refer to this table, use Dijkstra algorithm to calculate the shortest path between two points, and carry out route forwarding. When a network failure or human change occurs, the system reevaluates the transmission between nodes and updates the transmission cost table to ensure that the cost of routed transmission is minimal.

2.2.7. SRN Routing Algorithm. SRN routing algorithm [12] does not need to generate a large number of redundant groupings, and the system enables the network to organize and configure itself without requiring frequent management of routing, while also being able to tolerate faults. This algorithm has better fault tolerance performance than the flooding algorithm.

The implementation of SRN algorithm includes the following two parts: Path discovery: Upon receiving a packet to be transmitted, the source node verifies the destination node details included within the packet and assesses whether the route towards the destination node is presently stored in the local cache. Path maintenance: At the end of the path discovery process, the program executes the packet forwarding program. Each time a packet reaches a node, that node sends a confirmation request to the next node to jump to.

2.2.8. Negative preference routing algorithm based on torus network. Virtual channels are introduced on the basis of Torus network's negative preference routing algorithm [14], and to implement a deadlock-free adaptive routing algorithm, a 2D Torus network requires a minimum of three virtual channels for each physical link. Consequently, the physical network is partitioned into two virtual networks [13]. The adaptive routing algorithm makes use of virtual channels to achieve its purpose. However, there is a significant imbalance in the distribution of the load among the three virtual channels, with virtual channel Number 2 being underutilized resulting in significant wastage. The new virtual channel usage strategy and routing algorithm can be used to evenly distribute traffic to the three virtual channels, which improves the communication efficiency of the 2DTorus network on the whole.

2.2.9. Biological algorithm. Biological algorithm refers to a series of optimization methods that can be applied to other scientific fields based on the observation and analysis of biological behavior in nature. Common biological algorithms include genetic algorithm [16] (from the perspective of biological evolution), ant colony algorithm [15] (from the perspective of ant group foraging), etc.

The ant colony algorithm is a technique based on the behavior of ants when searching for food. It uses probability to determine the optimal route in a graph. The ant colony dynamic routing algorithm applies to Mesh structure, which can balance the traffic and reduce the generation of hot spots in the network. A routing table is employed to direct each node towards other nodes within the network, with options to choose from including East, West, and South directions. The probability of four links in the northbound direction, which selects the shortest path with the least traffic for the packet route. By introducing a biological algorithm, the fault tolerance of the on-chip network routing algorithm can be greatly improved, and the delay can be greatly reduced.

2.3. Routing algorithm combining independent routing and adaptive routing

After analyzing the above routing algorithms, the following three algorithms are created by combining the advantages of the independent routing algorithm and adaptive routing algorithm.

2.3.1. DyAD routing algorithm. A routing technique called DyAD [17] combines parity turn routing with dimensional routing. When the network is under low load, dimension routing is adopted, and when the network is under high load, adaptive routing with parity turning is adopted. The parity turning routing algorithm may choose a non-shortest transmission path for the packet. When more paths and nodes are added, the overall surface area also increases. Therefore, choosing a non-shortest path will significantly increase the transmission delay, which cannot guarantee the transmission quality well. The improved algorithm of DyAD routing algorithm [18] adds a waiting data field at the end of each packet, which will conduct blocking timing when the packet is blocked. The XY dimension routing algorithm is used when the network load is not high. In case of high network traffic, the first attempt is made to detect in the Y-axis direction. If the link in the Y-dimension direction is blocked, the router will immediately initialize the waiting data field of the packet. In case the routing detection fails in both horizontal and vertical directions, the tolerance value in the waiting data field is used for evaluation. This strategy still ensures that packets are transported in the shortest path, reducing unnecessary transmission and thus reducing latency.

2.3.2. SD routing algorithm. Combining deterministic and adaptive routing algorithms creates the SD routing algorithm [19]. Since the nodes in the central area adopt adaptive routing algorithm, it can well solve the hot spot and network congestion in the central node of Mesh and Torus network. As long as the adaptive routing algorithm is adopted in the central area, it can greatly save network hardware resources and reduce the complexity of router implementation.

2.3.3. Pseudo-adaptive XY routing algorithm. This kind of routing is more effective than deterministic XY routing [20]. This algorithm includes two working modes: independent routing mode and adaptive routing mode. When network congestion is reduced, XY routing (independent routing mode) is adopted for packet transmission. When the congestion degree is high, the packet selects the link with relatively low congestion degree around the node and passes through. One of the main purposes of pseudo-adaptive XY routing is to distribute network traffic reasonably and reduce the occurrence of hot spots. When two groups from different directions arrive at the node at the same time, the algorithm uses a fixed priority allocation strategy. For the packets that arrive at a node at the same time, the packets coming from north have the highest priority, followed by those heading south and east, and those heading west have the lowest priority.

3. Routing algorithm of 3D network on chip

The smaller and smaller feature size and higher and higher integration make the original bus structure and the point-to-point interconnection can not meet the needs of communication, and the on-chip communication of the system on chip becomes the bottleneck. Therefore, the network on chip (NoC) has been developed rapidly. Therefore, Three Dimensional Net-work-on-Chip (3D NoC) came into being. The main contents of NoC research include topological routing algorithm, quality of service,

switching mechanism, congestion control and energy consumption, among which 3D routing algorithm is one of the key technologies. The routing algorithm is in charge of precisely transmitting the packet to the target node and determining the path of packet transmission. Therefore, it is very necessary to design an efficient routing algorithm to minimize the cost of routing[21], especially for 3D NOCs with limited physical resources, it is more important to design an efficient routing algorithm[22]. The designed routing algorithm can still work normally in case of abnormal or unforeseeable events. In addition, it also requires low consumption, stability, simplicity, and flexibility, and can avoid deadlock live lock.

4. Conclusion

Network on chip is a kind of interconnection architecture used to connect the core on chip. Compared with other interconnection architectures, network on chip has the characteristics of strong scalability, high communication efficiency and low energy consumption. Therefore, network on chip has a broad prospect of future development. The design space of the network on chip includes network topology, routing node design and routing algorithm design. First of all, this article briefly introduces the network on chip-related content. The study then categorizes the network-on-chip routing techniques. Network routing algorithms on chips may be categorized into three groups based on their degree of adaptability: independent routing, adaptive routing, and the combination of independent routing and adaptive routing.

The design of the routing algorithm of the Mesh structure is relatively simple, because there are many loops in the surface Torus structure and surround channels. Therefore, it is impossible to simply adopt the method similar to the one in Mesh structure, and some improvements are needed. Currently, virtual channel or virtual network is commonly used in research. However, the implementation of virtual channel leads to a significant increase in resource consumption rate, making quality of service optimization the main focus of on-chip network routing algorithm research.

References.

- [1] Lai, Y., & Jing, M. 'e. (2018). Network on chip source routing algorithm based on A" algorithm optimization. *Fudan Journal (Natural Science)*, 57(5), 605-610.
- [2] Millberg, M., Nilsson, E., Thid, R., et al. (2004). The Nostrum backbone - a communication protocol stack for networks on chip. In *Proceedings of the VLSI Design Conference, Mumbai, India* (pp. 693-696). January 2004.
- [3] Yue, P., Chen, J., Liu, J., et al. (2009). Dual-channel Router Applied to On-Chip Network. *Chengdu: Journal of University of Electronic Science and Technology of China*, 38(2), 309-316.
- [4] Rantala, V., Lehtonen, T., & Plosila, J. (2006). Network on chip routing algorithms. *Turku Centre for Computer Science*.
- [5] Xie, P., Gu, H., & Jia, L. (2009). Research on Routing algorithms for on-chip networks. *Computer Engineering and Design*, 13.
- [6] Glass, C. J., & Ni, L. M. (1994). The turn model for adaptive routing. *Journal of the ACM (JACM)*, 41(5), 874-902.
- [7] Chiu, G. (2000). The odd-even turn model for adaptive routing. *IEEE Transactions on parallel and distributed systems*, pp. 729-738.
- [8] Ouyang, Y., Dong, S., & Liang, H. (2009). Design and simulation of NoC routing Algorithm based on 2D Mesh. (Doctoral dissertation).
- [9] Wangyuan, L., & Changshan, W. (2010). Mesh step-fault-tolerant Routing Algorithm. *Computer and Modernization*, 3, 39-42.
- [10] Li, M., Zeng, Q. A., & Jone, W. (2006). DyXY: a proximity congestion aware deadlock-free dynamic routing method for network on chip. In *43rd ACM/IEEE Design Automation Conference, San Francisco, CA, USA: ACM Press* (pp. 849-852).
- [11] Ali, M., Welzl, M., & Hellebrand, S. (2005). A Dynamic Routing Mechanism for Network on Chip. In *23rd NORCHIP Conference* (pp. 70-73). 21-22 November 2005.

- [12] Kim, Y. B. (2007). Fault tolerant source routing for network-on-chip. In 22nd IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems (pp. 12-20).
- [13] Cypher, R., & Gravano, L. (1994). Requirements for Deadlock-Free, Adaptive Packet Routing. *SIAM Journal on Computing*, 23(6), 1262-1274.
- [14] Gu, H., Liu, Z., Wang, K., et al. (2006). Distributed Adaptive Routing Algorithm in Toms Networks. *Journal of Xidian University*, 33(3), 352-358.
- [15] Dorigo, M., Maniezzo, V., & Colnari, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1), 29-41.
- [16] Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press.
- [17] Hu, J., & Marculescu, R. (2004). DyAD: smart routing for networks-on-chip. In *Proceedings of the 41st annual Design Automation Conference*.
- [18] Huan, N., & Changshan, W. (2010). Research on on-chip Network Routing Algorithm for OoS Assurance. *Computer and Modernization*, 26(5), 117-118.
- [19] Zhu, X., Gu, H., & Wang, C. (2009). SD: An Algorithm for on-chip Network Routing Based on mesh Toppling. *Microcomputer Information*, 25(8-2), 63, 93-94.
- [20] Dehyadgari, M., Nickray, M., Afzali-kusha, A., & Na-vabi, Z. (2008). Evaluation of Pseudo Adaptive XY Routing Using an Object-Oriented Model for NOC. In *The 17th International Conference, Microelectronics* (pp. 13-15).
- [21] Xiaoyun Zhang. "Power: Multi-Capability Adaptive Routing for Network-on-Chips", 2023.
- [22] 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), 2023.