International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 3, March 2015

Network-on-Chip Architecture Based on Cluster Method

Raj Kumar .S

Assistant Professor, Department of Electronics and Communication Engineering, Dhanalakshmi Srinivasan College of Engineering, Perambalur District, Tamilnadu, India

Abstract: ***Network-on-Chip architectures are emerging for the highly scalable, reliable, and modular on-chip communication infrastructure platform. On-chip networks utilize 2-D mesh topology. As the number of the cores present on-chip is increasing rapidly, the diameter of the network-on-chip is also increasing rapidly, which leads to large delay and energy consumption. The NoC architecture uses layered protocols and packet-switched networks which consist of on-chip routers, links, and network interfaces on a predefined topology. This paper proposed a cluster based topology with long-range links insertion algorithm. Finally, we evaluate the performance of the topology proposed in this paper through simulations.***

Keywords: cluster, on-chip, bus, topology

1. Introduction

Transistors embedded on a chip, the SoC of bus structure is poor at scalability, flexibility, reusability, and programmability. As a result the Network-on-Chip has been proposed and has gradually replaced the System-on-Chip of bus structure. Topology shows the connectivity and distribution of nodes which is a very important aspect to consider when we design an on-chip network. An appropriate topology can help to improve the performance of the on-chip network. Especially for a specific application, topology plays an important role in optimizing the performance of the network. Since different applications may have some different kind of communication requirements, general-purpose topology will be less efficient than application-specific designs. Therefore, in this paper, we will present a methodology of designing a topology for a specific application.

A. The Topology of 2-Dmesh

As 2-Dmesh has lower design complexity, most existing on-chip networks utilize 2-D mesh topology. As shown in Fig. 1(a), 2-D mesh has a regular and simple layout. However, the 2-Dmesh is not qualified to scale for their large diameter and energy inefficient. Even if the shortest path routing algorithms are utilized, the large network diameter still leads to extra router hops, and the router energy is much higher than the link energy. In addition, traversing many hops between two remotely communicating nodes may also lead to higher message blocking probability. To solve these problems, two kinds of solutions have been proposed, the long-range links insertion algorithms and cluster based mapping algorithms.

B. Long-Rang Link Insertion

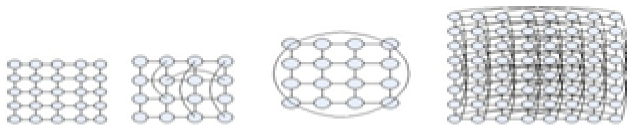
To reduce the diameter of a network, some papers proposed long-range links insertion algorithms [2]. Long-range links allow the shortcuts between two remotely communicating nodes [1]. There are two kinds of inserting methodology. One is called physical express topology [2] with long-range links inserted per regularly, as shown in Figure 1(d). Indeed, this topology can reduce the network diameter, and save the latency and power. However, extra router ports, large crossbars and extra physical channels are required. We may also insert fewer long links like Tmesh [6], as shown in Fig. 1 (c).This topology can optimize the performance of the network in a certain extent, but the long links are fully used while other links are rarely used, which may lead to link inefficiency and load imbalance. For specific applications, some long - range insertion algorithms have been proposed. According to these algorithms, only a few long-range links are inserted to optimize the performance of the network, but different with topologies like Tmesh. In these algorithms, the communication volumes of some couples of nodes are considered to decide which nodes to insert the long-range links. In most specific application networks [1], long-range links are inserted between the nodes which have large traffic volumes, Fig. 1(b) is an example in [1]. Long-range links can help to reduce the hops between remotely communicating nodes. Indeed, the performance between the nodes which have been inserted long links has been optimized, but for the whole network, there still has a large space for the performance improvement. Moreover, for the long-range links insertion, the better logical connectivity comes at the expense of a penalty in the structured wiring. The fewer the links the fewer the penalties, that means both the number and the length of long-range links are limited. Furthermore, application specific long-range links insertion algorithms are complex[1], [5]. According to the insertion algorithm proposed by [1], we have to consider all the possibilities when we just insert one long-range link.

After comparing the performance of all the inserting possibilities, a most efficient one is chosen. We have to repeat such complex computation each time we insert a long-range link to the networks. Moreover, since the long-range links insertion algorithms just take some couples of nodes into consideration, the whole network has not been mostly optimized, and it cannot guarantee the low energy consumption and traffic load balance.

ISSN (Online): 2347-3878
Volume 3 Issue 3, March 2015

C. Clusters Based Mapping Algorithms

To achieve low energy consumption and traffic load balance, some mapping algorithms [11] have been proposed. Especially in some mapping algorithms the nodes are divided into clusters [7][12]. When dividing the nodes, we have to consider two factors, the traffic volumes and physical shortest distance. As in [11], the paper proposed a fast topology partition based mapping algorithm for NoC. The purpose of the mapping algorithm is low energy consumption and traffic load balance. Through mapping algorithms, the performance of the network is optimized. But the diameter of the network is still large, and the performance of the remotely communicating nodes still has not been optimized. Furthermore, mapping algorithms usually cost a large amount of computation.



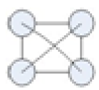
(a) (b)

(c) (d)

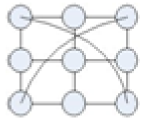
Figure 1: Topologies

D. Cluster Based Long-Range Links Insertion

Only considering the mapping or the long-range links inserting cannot mostly optimize the performance of the networks. Both of the ideas have their own advantages and limitations. So in this paper, we combine the two ideas, cluster based mapping and long-range links insertion, and we call this methodology "cluster based long-range links insertion algorithm". This method does not need to take too complex computation. First, we divide the nodes into clusters and then we insert long -range links between two remotely communicating clusters. According to the characteristics of the Small-World networks, cluster and long -range links make the topology to be a Small-World network. Firstly, we divide all the nodes into 4 global clusters according to the traffic volumes. If there are still a large amount of nodes in a cluster, we divide the nodes in the cluster into 4 small clusters further. That means the division is kept continuing until the bottom cluster is one of the unit clusters shown in Fig.2. Compared with other algorithms, this dividing algorithm needs fewer computations. After mapping the nodes onto a 2Dmesh, we insert long-range links between two remotely communicating clusters, such as, diagonal clusters. This algorithm can solve problems about large network diameter, node congestion, large energy consumption and latency.



a. even unit



b. odd unit

Figure 2: Unit Clusters

This paper is organized as follows. In section I, we will introduce some related works. In section II, we propose a new topology designing algorithm and elaborate the algorithm. We will introduce the routing algorithm of the topology proposed in section III. In section IV, we evaluate the performance of the topology.

2. Related Work

The methodology proposed in this paper comes from the characteristics of Small-world network. Shortcuts of the networks help to reduce the diameter of the networks and as a result, the delay of packets and the energy consumption are both reduced. In the next passage, we will talk about the Small-World as related work. The concept of Small-World comes from an experiment conducted by a social psychologist Stanley Milgram. The experiment shows that two people can be connected with each other through six people in average. This characteristic is named as "six degrees of separation", which is popularly known as "the Small-World" phenomenon. Recent work has shown that the physical connectivity of the Internet exhibits Small-World behavior [9]. In a word, under the two conditions, Small-World will be formed: 1) clusters existence; 2) permission of the shortcuts. For a Small-World network, we check the network performance with two parameters- cluster coefficient and path length [8]. The cluster coefficient is used to indicate the magnitude of group within network which is a parameter of local nature. It is a probability parameter which shows the possibility of two nodes being local nature. It is a probability parameter which shows the possibility of two nodes being neighbors of one node and also neighbors with each other. The Small-World networks usually have high cluster coefficient. Here we regard C as cluster coefficient. In a fully connected network, C=1. Assumed that the number of neighbors of node V is K, and the number of the links between K nodes is l_k, so the cluster coefficient of node V is:

$$C_V = \frac{l_V}{k_V(k_V-1)/2}, \quad C = \frac{1}{N} \cdot \sum_{V=1}^{N} C_V \qquad (1)$$

The average length of shortest paths is a parameter of global nature. Usually Small-World networks have small path length because of shortcuts. I_{ij} is the length of the shortest path between the node i and node j, so for a network, the average path length is:

$$L = [1 / (N(N-1))] \sum_{i \neq j} l_{ij} \qquad (2)$$

ISSN (Online): 2347-3878
Volume 3 Issue 3, March 2015

According to the properties of Small -world network, even only with the local information, we can find the shortest path. The advantages of Small-World should be considered to optimize the performance of network while designing a Small-World based topology. According to the nature of the Small-World networks, we propose a new method to design a NoC topology for a specific application. As the two basic principles of Small- World are clusters and shortcuts, we take both the cluster and long-rang links into consideration. We divide the nodes into clusters, and insert long-range links between remotely communicating clusters. We name this method as “Cluster Based Long-range Links Insertion Algorithm”.

3. Cluster Based Long-Range Links Insertion Algorithm

A. Background

The basic topology is m = n 2Dmesh, as shown in Fig. 1(a). The nodes communicate with each other through the network. The process of designing a topology is shown in Fig.3. Firstly we get the communication graph. We compare the traffic volumes between the nodes, making the nodes whose traffic volumes distinctly larger a cluster. Then we insert the long links between two distant clusters.

Figure 3: Flow chart of insertion algorithm

A cluster can be composed of several cluster units. Usually we define 4 nodes or 9 nodes as a unit cluster, just as Fig. 2 shows. We define the 2*2 unit cluster as even unit cluster and 3*3 unit cluster as odd unit cluster. The diagonal links are inserted to optimize the performance within a cluster. A network is composed of several clusters. For a large scale of networks, we can regard a small cluster as a node. In other words, a network can be made up of several clusters which are also composed of several low level clusters. The relationship between clusters in different levels is the iterative relationship.

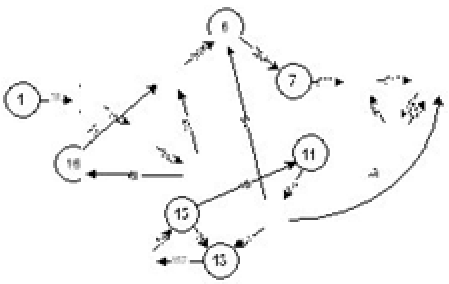


Figure 4: VOPD

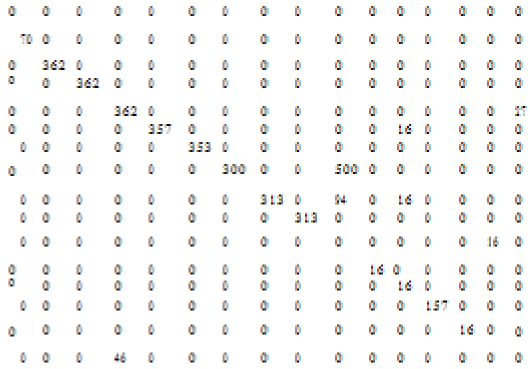
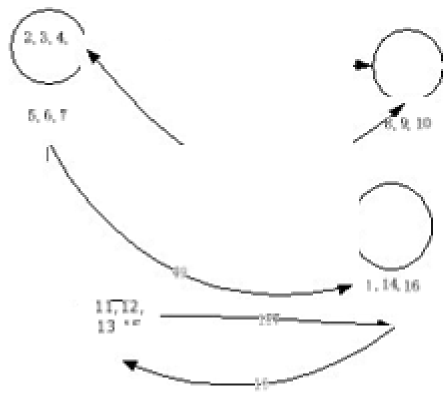


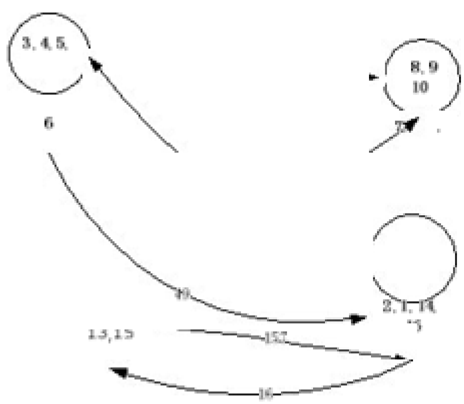
Figure 5: Traffic matrix

Table 1: The Nodes Division

>350	2→3→4→5→6→7,10→8
300-350	7→8→9→10
100-200	13→14
0-100	1→2,4→16→5,12→6,10→9,12→9, 11→12→13→15,14→15→11



(a) The Primary Partition



(b) The Final Result
Figure 6: The Partition of Nodes

B. The Partition of Clusters

For a specific application, we divide the nodes into 4 global clusters according to the traffic volumes and physical distance. For example, Fig. 4 is a communicating core graph of VOPD. To divide the nodes into four global clusters, first we get the traffic matrix as shown in Fig.5, then we draw a communication distribution according to the matrix, as shown in TABLE I. We make the first and second group as two clusters. Then we continue to divide the remaining nodes into clusters. For the VOPD application, the primary result is shown in Fig. 6 (a). Finally we adjust the results to make the partition as simple and regular as possible. The final result is shown in Fig.6 (b). To avoid congestion, we should also take the communication relationship into consideration. The mapping result is shown in Fig. 7. For an application, as long as we divide the nodes according to the traffic volumes, the result is reliable. Fig. 8 is a probable mapping result of 25 nodes; two odd clusters have a common node. Since appropriate partition and mapping can help to avoid link congestion and reduce latency and power consumption, the partition of clusters and mapping is very important

C. Long-Range Links Insertion Algorithm

After mapping, we insert the long-range links between the remotely communicating clusters. To minimize the number of the long-range links inserted into the topology, firstly, we have to confirm which nodes to insert the long-range links. When choosing the nodes to insert the long-range links, we should take the acceptable length of long links into consideration. At the same time, if possible, the long-range links should be inserted into different nodes. Furthermore, we choose the nodes which have large traffic volumes with the remote clusters. Usually there is a need to insert long-range links between the diagonal clusters. Fig. 9 shows several results of long-range links insertion. The diameter of the network and the hops of the packets traveling from one cluster to the remote cluster have been reduced. As a result, the latency and power consumption also have been greatly reduced.

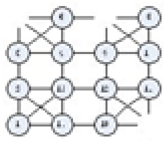


Figure7: Mapping result of 4 x 4 2D mesh

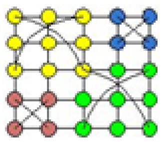
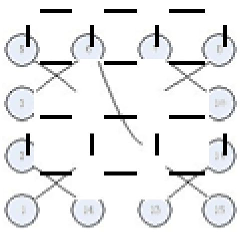
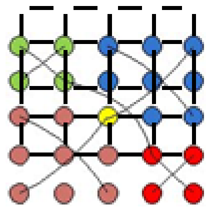


Figure 8: Mapping result of the 5x5 2D mesh



(a) 4 x 4 2D mesh



(b) 5 x 5 2D mesh

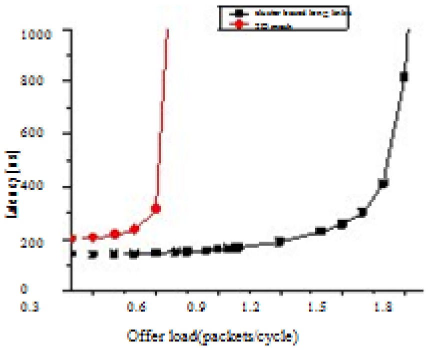
Figure 9: Topologies with Long-Range Links

4. Routing Algorithm with Long-Range Links

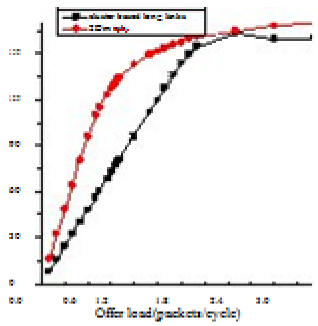
The topologies with long-range links usually adopt deadlock free routing algorithm. We can also enumerate all the possible paths [3]. In this paper, to avoid deadlock [8], we utilize the deadlock detection and recovery mechanism. The routing strategy proposed in this paper is based on the XY dimension -order routing algorithm. First we define the 4 clusters as 4 quadrants and name them with numbers (1 to 4). Within a cluster, the lower level clusters are also named as numbers according to the interior quadrants. As a result, the bottom quadrants of the unit clusters (level n) are named by n-dimension numbers. So the address of the nodes can be expressed as q_1, q_2,, q_n, x, y, x, y is the coordinate in bottom cluster which represents the address of the node in bottom cluster, q_1, q_2,..., q_n is the address of the bottom cluster, n is the depth of the cluster. When the packet arrives, we get the local and destination address. The comparison begins from the global quadrant number, that means, first we compare the two q_1. If local q_1 and destination q_1 are same which means the local node and destination node are in the same global cluster, we continue to compare q_2, and this work continues until local q_i and destination q_i are different. If q_1, q_2,, q_n are all same, local node and destination node hop between nodes while in this topology, to see from level i which means q_i is different, packets hop between clusters in level i. We regard the clusters in level i as nodes. Through default routing algorithm (the XY dimension-order routing algorithm extended with long- rang links), we get the path in level i. To arrive the expected nodes along the path in level i, we have to consider from level i − 1. In lower levels, we the current destination is q_i we have found the path in level i, and it is a progress of iteration. When q_1, q_2,, q_n are all the same, we utilize the default routing algorithm to lead the packet to the destination. The last, we have to apply the dead lock detection and recovery mechanism to avoid dead lock.

5. Performance Analysis and Simulation

We evaluate the topology by simulating in the OPNET simulation platform. We observe the delay, throughput and energy consumption of the network. The energy consumption of the network is defined as formula (3), E_s is the energy consumption of routers, E_l is the energy consumption of links. As shown in formula (4), $E_{bit}^{i,j}$ is the energy consumption of transmitting one bit from router r_i to router r_j, E_t is the total energy consumption of the network, $C_{i,j}$ is the communication volume from router r_i to router r_j, N_r is the number of routers along the path. We compare the topology proposed in this paper with the traditional 2Dmesh with XY dimension-order routing algorithm. We simulate the application of VOPD as shown in Fig. 4 on the topology shown in Fig. 9(a). The simulation results in Fig. 10(a) show the performance of the network was optimized. The delay has been reduced while the throughput has not been changed much. For the reduction of network diameter, the hops that link. The energy consumption is also reduced.



(a) Delay



(b) Throughput

Figure 10: The Simulation Results

6. Conclusion

In this paper, we proposed new Cartesian networks based topology. This topology combines mapping and long-range links insertion, which can greatly optimize the performance of the network. So there is a high speed network transmission and efficiency is higher compared to other algorithm. Power consumption, delay and storage capacity is low, which leads to minimum chip size.

References

[1] U. Y. Ogras and R. Marculescu, “It's a small world after all: NoC performance optimization via long-range link insertion,” Very Large Scale Integration (VLSI) Systems, IEEE Transactions, vol. 14, pp. 693-706, 2006.

[2] C. H. O. Chen, et al., “Physical vs. Virtual Express Topologies with Low-Swing Links for Future Many-Core NoCs, ” in proceedings of 2010 Fourth ACM/IEEE International Symposium on Networks-on-Chip (NOCS), pp. 173-180, 2010.

[3] P.P. Yue, J. Liu, S. Anjum, et al., “Enumeration-based path allocation algorithm in NoC mapping, ” Journal of University of Electronic Science and Technology of China, vol. 37, no.1, pp. 54-57, 2008.

[4] F. Ge and N. Wu, “Low energy topology generation approach for application-specific network on chip, ” System Engineering and Electronics, vol.32, no.8, pp. 1754-1759, 2010.

[5] U. Y. Ogras and R. Marculescu, “Application-specific network-on-chip architecture customization via long-range link insertion, ” in proceed-ings of IEEE/ACM International Conference in Computer-Aided Design, pp. 246-253, 2005.

[6] Y.Q. Sheng and W.Z. Kai, “An Improved Mesh Topology and Its Routing Algorithm for NoC,” in proceedings of International Conference on Computational Intelligence and Software Engineering (CiSE), pp. 1-4, 2010.

[7] L. Zhonghai, X. Lei, and A. Jantsch, “Cluster-based Simulated Annealing for Mapping Cores onto 2D Mesh Networks on Chip, ” in proceedings of 11th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems (DDECS), pp. 1-6, 2008.

[8] B. Yang and S.-h. Xu, “The Analysis of Knowledge Transfer Network Characteristic Based on Small-world Network Model, ” in proceedings of Second International Conference on Future Networks, pp. 428-432, 2010.

[9] S. Jin and A. Bestavros, “Small-world characteristics of the Internet and multicast scaling,” in proceedings of the 11th IEEE/ACM International Symposium on Modeling.