Design and Evaluation of Binary-Tree Based Scalable 2D and 3D Network-on-Chip Architecture

Mohammad Rashid Ansari\textsuperscript{a}, Abdul Quaiyum Ansari\textsuperscript{a} and Mohammad Ayoub Khan\textsuperscript{b}

\textsuperscript{a}Department of Electrical Engineering, Jamia Millia Islamia (JMI) University, New Delhi, India; \textsuperscript{b}Department of Computer Science, SET, Sharda University, Greater Noida, India

ABSTRACT

Network-on-Chip (NoC) has been developed as a most prevailing innovation in the paradigm of communication-centric technology. It solves the limitations of bus-based systems, with the incorporation of 3D IC technology, and it reduces packaging density and improves performance of Multiprocessor System-on-Chip. There is need of suitable NoC topology for these applications and desired performances. This paper proposes a scalable binary tree-based topology for 2D and 3D NoCs. The average degree of the proposed network is reduced around 40% of the torus whereas the diameter also reduced significantly, as compared to other topologies.

1. Introduction

Advancements in the semiconductor technology have increased the integration of many heterogeneous components on single chip with Multiprocessor System-on-Chip (MPSoC). On the single chip, many features are demanded such as high performance and high communication-centric technology. For this demand, Network-on-Chip (NoC) is the suitable option for the communications on single chip to remove the problems faced by bus-based communications. NoC on single chip is based on miniaturization computer network...
concepts, which reduces the cost of communication and time-to-market.

3D NoC has the advantage of 3D integrated chip (IC) and combines it with NoC. This gives an option to connect many heterogeneous chips vertically with the small vertical links. For vertical connection on different chip layers, Through-Silicon-Vias (TSVs) are mostly used, as it consumes less power and have low delay and high bandwidth.

The physical structure of NoC is decided by its topology. Regular and fixed topologies like mesh, torus, ring, etc. are commonly used in NoC. Mesh is mostly used because of its simple and regular structure, but it has many limitations, for example, for a larger network, the links’ requirement is more, which requires extra area and power of the chip. By reducing the links, there is an advantage of both the area and power in NoC. Several topologies have already been proposed topologies such as torus and mesh [1] are mostly used. Torus topology overcomes the large diameter of the mesh topology. The 3D torus topology has been shown in Figure 1, it has long wrap-around links and complex that can be seen from the figure. Quadrant-based routing designed for 3D torus in [2], it used few TSVs in place of all vertical connection.

It suffers from long wrap-around links causing more delay, which was reduced by using folded torus topology. Hierarchical binary tree [3] have small bisection width through better management. Fat tree overcome small bisection bandwidth of binary tree but the problem of high-degree node comes [4]. Mesh of tree is the hybrid of mesh and tree architecture [5]. Binary search tree-based ring topology was proposed and that had drastic reduction in degree and meter of network [5]. The proposed topology is the combination of binary search tree and ring topology which drastically removes the number of links used in [6]; thus, the proposed topology has the advantage of both area and power saving on chip along with reduction in degree.

The remaining part of this paper is organized as follows: Section 2 presents the related works, Section 3 presents the proposed topology and its characteristics, in Section 4, discussion, comparison, and analysis are presented, and in final section, conclusions are presented.

2. Related Works

Most of the NoC uses the mesh-based topology for implementation as it has regular and simple structure but the drawback in mesh topology is that it requires more resources, like, more links required for larger size network which results in needless links and energy overhead [7].

A cross-by-pass-mesh (CBP-Mesh) topology has been proposed which reduces the diameter and average hop counts [8]. It has high scalability and is based on mesh topology architecture. The extra bypass links provide small route to destination from the source, and this results in improvement in the performance of the NoC, but at little complexity and cost of power and energy requirement.

With the emergence of 3D IC technology, 3D NoC is a compelling option for chip interconnection. As 3D Recursive Network Topology (3D RNT) was given in [9], it has partial vertical links by using TSVs and also proposed the routing algorithm for the topology. But the issue of heat dissipation is still an issue with it.

In [10], an optimization technique, the ant colony optimization (ACO) was applied to 3D networks having torus, mesh, and hypercube topologies for routing protocols. The 3D Ant Colony Routing (3D-ACR) and the another optimization technique used in [11] is based on Hopfield Neural Network (HNN) for 2D mesh topology, and they have some shortcomings like the scalability of the algorithm and chip area overhead problem for implementation.

In [12], exhausted survey of 3D NoCs is given. The transistor scalability issues can be solved by using the 3D NoC which is well suited for heterogeneous multiprocessor SoC.

From the various current papers and articles, we can say that there is still more focus on designing a 2D and 3D NoCs’ architecture which can give better scalability and performance improvement and to overcome the issues of area overhead, energy, or power consumption and to reduce the number of links.

The proposed topology is the combination of binary tree and ring topology for 2D and 3D NoCs’ architecture, that drastically removes the number of links. The less links means reduction in chip resources which saves area and energy consumption. The average degree of the proposed network is reduced around 40% of the torus which reduces the routing cost and complexity. The diameter also reduced significantly, as compared to other topologies which give shorter routing paths and the advantage for the latency of the routing, and the scaling of the network improves as the diameter of the topology directly linked with it.

3. Proposed Topology

A binary tree is a well-established structure in which each node has two child nodes named as left child node
and right child node originated from a core node. Binary search tree-based ring topology was proposed in [13]. The basic module of it consists of tree nodes that can communicate directly with each other as shown in Figure 2. Although this topology drastically reduces in the diameter and degree of the node, it has higher number of links. As the number of nodes increases, more levels of BST-ring topology are needed, which gives rise to redundant links, and that adds to unnecessary area and power consumption of the NoC architecture. In this work, a combination of BST and BST-ring topology is proposed which gives the significant improvement in the node degree with the less number of communication links. The proposed topology is shown in Figure 2.

For the proposed topology, following parameters are derived:

- \( N_t \): Total number of nodes
- \( \bar{d}_a \): Degree (average)
- \( D_N \): Network diameter

Number of terminal nodes \( (T_N) \) at level \( (L) \):

For level \( L = 1 \), \( T_N = 3 \)
- \( L = 2 \), \( T_N = 3 \times 2 \);
- \( L = 3 \), \( T_N = 3 \times 2 \times 2 \);
- \( L = 4 \), \( T_N = 3 \times 2 \times 2 \times 2 \);
-...
- \( L = n; T_N = 3 \times 2^{n-1} \);

Hence, the number of terminal nodes at level \( L = n \) is given by the below-given expression:

\[
T_N = 3 \times 2^{n-1}, \text{if } n > 0
\]

Total number of nodes \( N_t \) at level \( L \) can be calculated as:

For level \( L = 1 \), \( N_t = 3 \times 1 = 3 \)

\( L = 2 \), \( N_t = 3 \times 2 + 3 \times 1 \)
\( L = 3 \), \( N_t = 3 \times 2 \times 2 + (3 \times 2 + 3 \times 1) \)
\( L = 4 \), \( N_t = 3 \times 2 \times 2 \times 2 + (3 \times 2 \times 2 \times 3 \times 2 + 3 \times 1) \)
\( L = n, N_t = 3 \times 2^{n-1} + 3 \times 2^{n-2} + 3 \times 2^{n-3} + \ldots + 3 \times 2^0 \)

Hence, the total number of nodes at level \( L = n \) is given by:

\[
N_t = \sum_{n=1}^{n} 3 \times 2^{n-1}
\]  

Diameter of network \( (D_N) \) for level \( L = n \):

For level \( L = 1 \), \( D_N = 1 \)
- \( L = 2 \), \( D_N = 1 + 2 = 3 \)
- \( L = 3 \), \( D_N = 1 + 2 + 2 = 5 \)
- \( L = 4 \), \( D_N = 1 + 3 + 3 = 7 \)
-...
- \( L = n, D_N = 1 + (n-1) + (n-1) \);

Hence, the diameter of the proposed topology is given below:

\[
D_N = 1 + 2(n - 1)
\]

Average degree \( (\bar{d}_a) \) of network can be derived as follows:

Depending on the level of topology, the internal nodes and terminal nodes have different degrees, like if level is odd, the terminal nodes have degree 2, and when level \( L \) is even, the degree of terminal node is 1.

For level \( L = 1 \), Nodes total degree: \( d_T = 3 \times 2^0 = 6 \)

Figure 2. Basic module of the proposed topology for level 1, 2, and 3.

Figure 3. Possible 2D layout of the proposed topology.
The degree of the proposed topologies is compared with torus topology and ring-based tree topology as shown in Figure 5. It has found to be better than them, and value of degree of the proposed topology is around 40% less than the torus topology. The possible distribution of the cores and nodes are shown in Figure 3 and Figure 6. It shows the layout of the proposed topology.

We can have a generalized expression as follows for level \( L = n \) as:

\[
d_t = m \times (3 \times 2^{n-1}) + \sum_{k, \text{even}}^{N} 4 \times (3 \times 2^k) + \sum_{k, \text{odd}}^{N-1} 3 \times (3 \times 2^k)
\]

where \( k \geq 0, m = 1 \) and \( N = n - 1 \), when \( L = n \) is even and \( m = 2 \) and \( N = n - 1 \), when \( L = n \) is odd.

From the above Equations (2) and (4), we can find the average degree \( d_a \) of the network as:

\[
d_a = \frac{\text{Total degrees}}{\text{Total Nodes}} = \frac{d_t}{N_t}
\]

### 4. Discussion and Analysis

The performance and scaling property of the network-on-chip can be analyzed by theoretical and mathematical modeling of the network. Analysis of proposed network parameters is given in Table 1. In scaling property, the performance is highly dependent on the diameter of the network. Diameter \( (D_{\text{N}}) \) measures the number of maximum nodes traveled from source node to destination node. Graphical analysis shown in Figure 4 represents comparison of the diameter \( (D) \) of proposed topology and exiting topologies such as mesh and WK-recursive [14]. From the Figure 4, it can be seen that the proposed topology has the lowest diameter among other topologies. At level 5, number of nodes are 93, the mesh has diameter 17.28 whereas proposed topology has diameter 9.

| Level | \( N_t \) | \( T_{\text{N}} \) | Average degree \( (d_a) \) | Diameter \( (D_{\text{N}}) \) |
|-------|----------|----------------|----------------|----------------|
| 1     | 3        | 3              | 2              | 1              |
| 2     | 9        | 6              | 2              | 3              |
| 3     | 21       | 12             | 2.5            | 5              |
| 4     | 45       | 24             | 2.26           | 7              |
| 5     | 93       | 48             | 2.64           | 9              |
| 6     | 189      | 96             | 2.31           | 11             |
| 7     | 381      | 192            | 2.66           | 13             |
| 8     | 765      | 384            | 1.99           | 15             |
| 9     | 1533     | 768            | 2.66           | 17             |
| 10    | 3069     | 1536           | 2.33           | 19             |

The degree of the proposed topologies is compared with torus topology and ring-based tree topology as shown in Figure 5. It has found to be better than them, and value of degree of the proposed topology is around 40% less than the torus topology. The possible distribution of the cores and nodes are shown in Figure 3 and Figure 6. It shows the layout of the proposed topology.
5. Conclusion

Topology is the basic building block for designing the NoC; it decides the roadmap for traversal of the packets. Performance, complexity, and scalability mainly depend on the topology of the network. Various metrics of the proposed topology are explored and compared with the other existing most common topologies. It is found that diameter of the proposed topology is considerably small compared to other topologies such as mesh and WK-recursive. The degree of proposed topology is significantly lower than the other existing topology. This results in lower number of links required that not only save the area overhead but also reduce the complexity and cost of router.

With the extension of the topology for 3D NoC, the number of TSVs required is reduced which is important while designing 3D ICs. As extension to present work in future, it is planned to design routing algorithm that can deal with the faults and congestion and to calculate the power consumption.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Mohammad Rashid Ansari  
http://orcid.org/0000-0003-0092-8389

References

[1] Dally WJ, Towles B. Principles and Practices of Interconnection Networks. San Francisco (CA): Elsevier; 2004.
[2] Ansari AQ, Ansari MR, Khan MA. Modified quadrant-based routing algorithm for 3D Torus Network-on-Chip architecture. Perspect Sci. 2016;8:718–721.
[3] Jeang Y, Jou J, Huang W. A binary tree based methodology for designing an application specific network-on-chip (ASNOC). IEICE Trans Fundam Spec Sect VLSI Des CAD Algorithms. 2005;E88-A(12):3531–3538.
[4] Guerrier P, Greiner A, Pierre U, et al. A generic architecture for on-chip packet switched interconnections. Proceedings of the Conference on Design, Automation and Test in Europe; Paris, France; 2000. p. 250–256.
[5] Balkan AO, Qu G, Vishkin U. A mesh-of-trees interconnection network for single-chip parallel processing. In Proceedings International Conference Application-specific Systems, Architectures and Processors; Steamboat Springs, CO, USA; 2006. p. 73–80.
[6] Khan MA, Ansari AQ. An efficient tree-based topology for network-on-chip. In World Congress on Information and Communication Technology (WICT 2011); Mumbai, India; 2011. p. 1316–1321.
[7] Balkan AO, Qu G, Vishkin U. Mesh-of-trees and alternative interconnection networks for single-chip parallelism. IEEE Trans Very Large Scale Integr Syst. 2009;17(10):1419–1432.
[8] Gulzari UA, Anjum S, Agha S, Khan S, Torres FS. Efficient and scalable cross-by-pass-mesh topology for networks-on-chip. IET Comput Digit Tech. 2017;11(4):140–148.
[9] Viswanathan N, Paramasivam K, Somasundaram K. An optimised 3D topology for on-chip communications. Int J Parallel Emergent Distrib Syst. 2017;29(4):346–362. Taylor & Francis.
[10] Junior LS, Nedjah N, Mourelle LDM. Efficient routing in network-on-chip for 3D topologies. Int J Electron. 2015;102(10):1695–1712.
[11] Esmaelpoor J, Ghafouri A. Performance evaluation of a routing algorithm based on hopfield neural network for network-on-chip. Int J Electron. 2015;102(12):1963–1977.
[12] Taylor P, Kumar MP, Murali S, Veezhinathan K. Network-on-chips on 3-D ICs: past, present, and future. IETE Tech Rev. 2014;29(4):318–335. Taylor & Francis.
[13] Khan MA, Ansari AQ. Efficient topologies for 3-D network-on-chip multicore technology: architecture, reconfiguration and modeling. In Qadri Muhammad Yasir, Sangwine Stephen J, editors. Multicore technology architecture, reconfiguration, and modelling. London: CRC Press; 2013 Jan. 2015, p. 335–353.
[14] Suboh SA, Narayana VK, Bakhhouya M, et al. A scalability study of interconnect architectures for system-on-chip. Proceedings 2012 International Conference High Performance Computing and Simulation, HPCS; Madrid, Spain; 2012. p. 300–306.