SMITHA: Scalable Modular Interconnect for Three Dimensional High Performance Applications - A New 3D Topology for NoC Based Systems

Sanju V[1], Niranjan Chiplunkar[2] and **Venkata Krishna[3]**

Submitted in July, 2013; Accepted in October, 2014

Abstract - Today we are in an era of green computing wherein the devices are high performing, occupies smaller chip die area, and consumes low power. These systems are designed and implemented using multi core architectures. Network on chip is establishing itself as interconnect for this high performance multi core systems. Currently the systems are realized using two dimensional topologies like mesh, torus etc. Research outcome in fabrication technology is reducing the feature size of silicon processes which enables more logic to be implanted on silicon. This was well complemented with improvement in packaging technology which led to vertical stacking of logic to form of three dimensional structures. This paper introduces a new three dimensional topology SMITHA (Scalable Modular Interconnect for Three dimensional High performance Applications). The paper discusses the two dimensional base topology along with routing algorithms and performance parameters and its extension to three dimension. Performance parameters for both cases are also discussed.

Index Terms – Network On Chip, Design, SMITHA

1.0 INTRODUCTION

One of the biggest inventions of the last century is semiconductor devices. These devices allowed design and implementation of systems in every domain ranging from consumer electronics to industrial or defense applications. In the initial days these systems were implemented using common bus architecture. The different processing modules shared a traditional bus to communicate and pass information between them. As the complexity of the systems increased, common bus posed a serious problem in terms of performance which led to usage of parallel or multi bus structures. These prevailed in design for decades which always saw performance bottleneck. To overcome these performance issues, Network On Chip (NoC), a new paradigm in design was introduced [1,2]. The idea was to implant the techniques of data communication network on chip. In this the modules are interconnected using topologies like mesh, torus and packets were exchanged for purpose of communication.

[1]*Dept. of Computer Science & Engineering, Nitte Meenakhsi Instituteo of Technology, Bangalore, INDIA*
[2]*NMAMIT, Nitte P.O, Karkala, INDIA*
[3]*School of Computing Science & Engineering, VIT University, Vellore, INDIA*
E-mail: [1]sanjuv21@gmail.com

Today we are in an era of mobile computing running multiple applications. These devices being hand held running in batteries also had additional constraints in term of power consumption also. The present popular topologies namely mesh and torus when scaled poses a problem in terms of performance which in term affects the power consumption of the systems. This paper discusses a new topology for network on chip based systems SMITHA (Scalable Modular Interconnect for Three dimensional High performance Applications). The paper also discusses the performance / area parameters of the same and compares with popular topologies namely mesh and torus. It is observed that the proposed topology performs better with lesser area requirement.

2.0 PROPOSED ARCHITECTURE

The discussion about the proposed architecture is done as two sections. The section below discusses the two dimension topology along with the performance parameters followed by its extension into three dimension.

2.1 Two Dimensional Base Topology

The proposed architecture is obtained by deleting the base node and by interconnecting the neighboring nodes along the level of a complete binary tree. This depicted in the figure below (Fig 1).

The topology is identified by the number of layers, numbered from 1 to K where K is the number of layers in the configuration. A node n in layer K is linked to its neighboring nodes 2n and 2n +1 in layer K + 1. The topology becomes bigger with the number of layer with an increment of nodes in power of two starting with two nodes in layer one to 2^K nodes in layer K. The number of nodes in a configuration of K layer

$$N_{s(1,K)} = \sum_{i=1}^{K} 2^i$$

2.2 Addressing

The nodes are being addressed depending on the layer which they are and position within it. The nodes start its address from zero to 2K and one to k for layers.

2.3 Routing Algorithm

This section brings out an optimal routing algorithm for routing packets in the same level. All the packets are routed through the shortest path between any source destination pairs.

Step 0: Check destination address. Initialize current_src as current node address and current_dest as destination address.

Case 1: current_src and current_dest in same layer.

1. Compute the minimum hop count between current_src and current_dest. If it is greater than 3, set current_src and current_dest as their parent nodes respectively in the adjacent layer below.
2. Repeat Step 1 till minimum hop count is greater than 3.
3. Move to the next node from current_src towards current_dest. Set the current_src as the next node.
4. Repeat Step 3 until current_src is equal to current_dest.
5. Consider a complete binary tree with top right node of the current_src as root. If destination node lies in the tree, mark the right node as set current_src else mark the sibling of the left node as current_src. Move to the current_src.
6. Repeat Step 5 until current_src is equal to destination address.

Case 2: When destination node is in a layer above that of the source node

1. Move one step at a time from the current_dest to the layer below until layer of the current_src is reached. Set current_dest as the node obtained in the layer of the source.
2. Repeat the steps as in Case 1.
3. Repeat Step 5 and 6 of Case 1.

Case 3: when destination node is in a layer below that of the source node

1. Move one step at a time from current_ src to the layer below until the layer of current_dest is reached. Set current_src as the node obtained.
2. Repeat the procedure as in case 1.

The algorithm presented above can be explained as follows. Case 1 represents when the source and destination are in the same level. If they differ by three or less positions then the packet is just transferred to them directly else it is routed downward to the layers below and then it moves to the destination.

Case 2 in the algorithm explains when the destination node is above that of the source node. In this case, the routing logic finds the parent node of the destination in the layer of the source. Now both the source and destination is the same layer and will use Case 1 to route the packet to the intermediate node and then the packet moves upward destination.

Case 3 expresses when the destination node is below than the source node. In this case, the packet moves to the root of the source node in the layer of the destination. Now Case 1 hold goods as the source and destination is in same layer.

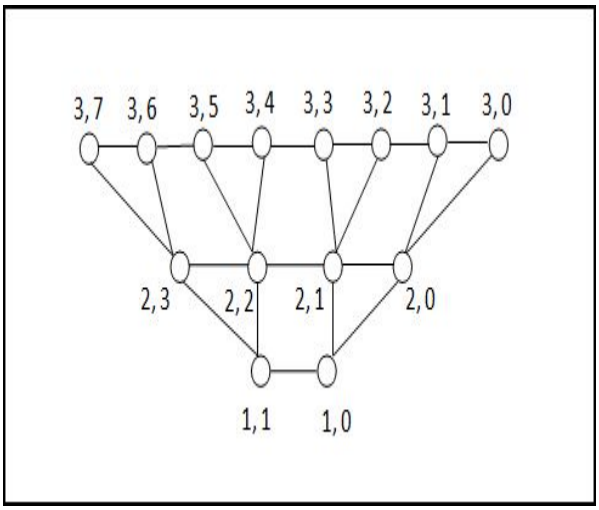

Figure 1: Proposed Topology With Three Layers In Two Dimension

2.4 Performance Parameters

This section deals with the various performance parameters of the proposed topology in two dimension. The performance and area parameter are discussed below.

2.4.1 Maximum & Average Hop: This parameter brings out the performance of the system. Consider a configuration of K layers, we tabulate the number of hop required through the shortest route between every source and destination pair. The maximum value thus obtained in the set is called as maximum hop for a configuration of K layers. The mean value of the set is called as average hop. Maximum hop brings out the time taken for a packet to reach any source destination pair for a given configuration. Higher the value, more will the time taken to reach the destination thereby reduces the performance of the system. Average hop also reflects the performance of the system by considering a cumulative effect of all source destination pair. The maximum hop for a configuration of n nodes is expressed as

$$HC_{(Max)}(1,n) = 2\log_2(n+2) - 3 \approx 2\log_2(n+2).$$

The table (Table 1) below depicts average hop for a single level. From the equation and table quantified above, it should be noted that both the parameters of the proposed topology grows gradually as the topology scales.

2.4.2 Number of Wire Segments & Wirelength: This parameter brings out the area needed by the circuit and the power consumption of the system to an extent. The parameter number of wire segments brings out to the number of interconnection wire segments used to produce the topology. Similarly considering unit length of wire between every pair of nodes, we calculate the wirelength requirement of the system. These parameters also indicate the level of difficulty for the CAD tool to generate the design and complete its process. These parameters for a configuration of n nodes in two dimension is expressed as

$$W_s(1,n) = WL_s(1,n) = 2n - \log_2(n+2) - 1 \approx 2n$$

3.0 Three Dimension Topology

The proposed three dimension topology is made by placing the above discussed base substrate one over the other and by interconnecting the adjacent levels as follows.

- The interconnections between odd level and an even are done by interconnecting the even layers through the right and odd layers though the left. For example interconnection between level one and two.
- Similarly the connections between even level and an odd are done by connecting the odd layers though the right and even layers through the left. For example interconnection between level two and three.

This way makes the proposed topology more scalable by placing one layer over the other. The interconnection discussed above is depicted in the figure (Fig 2) below for a configuration of three levels and each level having three layers.

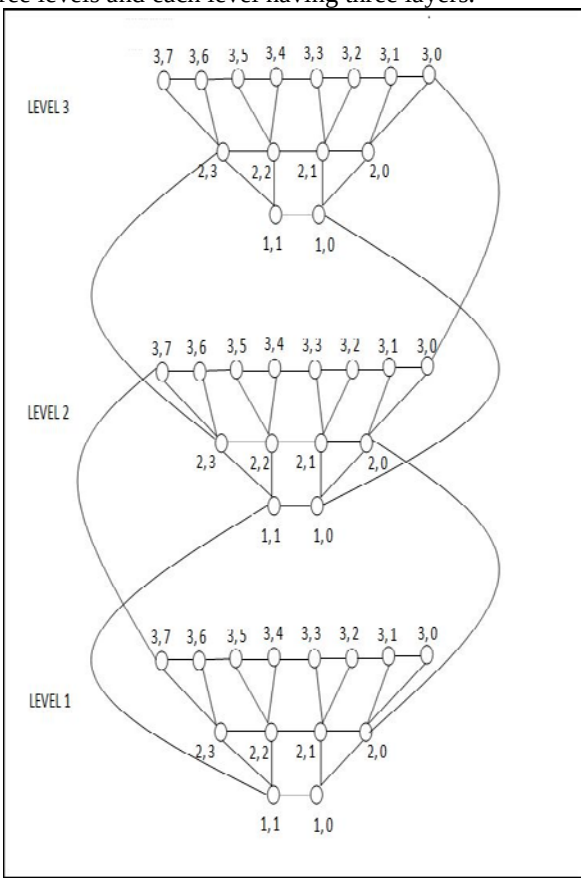

Figure 2: Figure Depicting The Proposed Topology In Three Dimension For A Configuration Of Three Levels And Each Level Having Three Layers

3.1 Addressing

The nodes are placed and are addressed relative to the level, layer within that level and node position along the layer to which the node belongs. For example a node who has an address (1, 2, 0) represents a node in level one layer two and position zero. This is depicted in the figure (Fig 2)

3.2 Routing Algorithm

This section presents a routing algorithm for the proposed topology in three dimension. Whenever the source and destination nodes are in the same level, the routing algorithm presented in case of two dimension applies. The algorithm presented below routes a packet which is in two distinct levels. For clarity in explanation and understandability, we divide the topology vertically into two as shown in the figure (Fig 3)

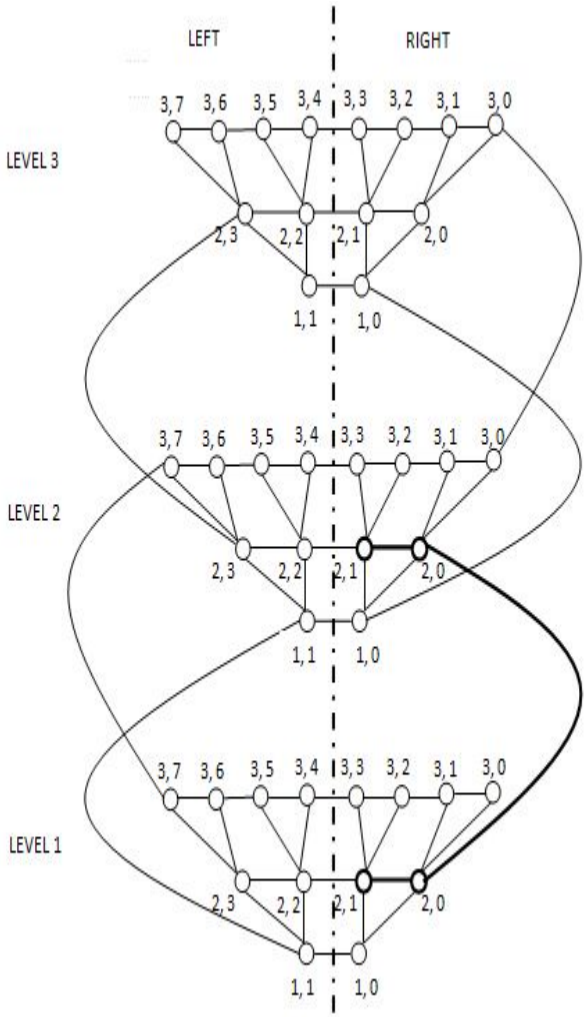

Figure 3: Figure Illustrating the Routing Algorithm

Step 0: Check destination address.
Step 1: Check node number of destination and current node address
Step 2: Check layer of the destination and current node address

Case 1: Current node and destination on the left
Step 3: Move left when current node and destination is in the same layer

Step 4: Move top left when destination is greater than current node
Step 5: Move down when destination is less than current node

Case 2: Current node on the left and destination on the right
Step 3: Move right when current node and destination is in the same layer or when the current node is greater than the destination
Step 4: Move down when destination is less than current node

Case 3: Current node and destination on the right
Step 3: Move right when current node and destination is in the same layer
Step 4: Move top right when destination is greater than current node
Step 5: Move down when destination is less than current node

Case 4: Current node on the right and destination on the left
Step 3: Move left when current node and destination is in the same layer or when the current node is greater than the destination
Step 4: Move down when destination is less than current node

The current node checks the destination address in the packet. The next node of transit is decided by the routing algorithm depending on the position of destination and current node. The routing algorithm routes the packet to the nearest node towards the periphery in the same level of the current node and then to the level of the destination node. Now the packet is in same level as destination node, will follows the two dimensional routing algorithm to reach the destination node. For example, consider a packet currently in node (1,2,1) to be transferred to (2,2,1). Since both the nodes are on the right topology, the packet is routed towards the periphery through the right to (1,2,0), then to the level of the destination to (2,2,0) and then to the destination (2,2,1). Similarly the proposed routing algorithm routes packet to all source destination pairs.

3.3 Performance Parameters

This session discusses the different parameters in case of the proposed three dimension topology. The parameters discussed are same as those of two dimension.

3.3.1 Maximum & Average Hop: As discussed in the case of two dimensional, the performance of the systems will be given by this parameter. As per the routing algorithm presented above, the packet moves towards the node at the periphery which is at most $HC_{s(Max)}(1,n)$. Then it should travel vertically up through (L – 1) nodes to reach the level of the destination node. Now it takes another $HC_{s(Max)}(1,n)$ to reach the destination node. Summing the values above, the parameter can be bounded as

$$HC_{s(Max)}(L,n) = HC_{s(Max)}(1,n) + (L-1) + HC_{s(Max)}(1,n)$$
$$= 2 * HC_{s(Max)}(1,n) + (L-1)$$
$$= 2 *\{ 2\log_2(n+2) - 3\} + (L-1)$$

where L is the number of levels and $HC_{s(Max)}(1,n)$ is the maximum hop in case of two dimension topology with n nodes.

Similarly maximum hop for all source and destination was tabulated and simple arithmetic mean for recorded. The effect was the same as observed in the case of two dimension.

3.3.2 Number of Wire Segments & Wire length: In case of the three dimensional structure, the wires used to interconnect within layers and between the different levels contribute to the wire length and number of wire segments. Considering a unit length of wire for interconnecting the node within the layer and between the layers, the parameters can be expressed as

$$W_s(L,n) = WL_s(L,n) = L(2n - \log_2(n+2) - 1) + (L-1)(\log_2(n+2) - 1)$$

Where n is the number of nodes in each level and L is the number of levels. The first term quantifies the parameters for interconnecting nodes in single level and the second term quantifies the same for interconnecting two successive levels. It should be noted that as the number of levels increases, the parameters for interconnecting the successive levels increases only in logarithmic order. This is very advantageous in case of the proposed three dimension topology.

4.0 COMPARISON OF PERFORMANCE PARAMETERS WITH EXISTING TOPOLOGIES

This session compares the above discussed performance parameters with two dimensional and three dimensional mesh and torus, the popular existing topologies in this concept.

4.1 Comparison with Two Dimension Topologies:

This session compares the performance parameters in case of two dimension with popular existing topologies – mesh and torus

The tables (Table 2,3) below tabulate the parameter maximum hop and average hop for different number of nodes per layer for levels one, two and three

From the tables 2 and 3, it is quite evident that the maximum hop and average hop parameters are good for the proposed topology when it is compared to mesh and torus. Also it should be noted that the performance parameters for the proposed topology when scaled does not grow drastically when compared with mesh and torus This implies that the packets reach the destination in less time and the output generation is faster increasing the throughput of the system in the proposed topology and a system implemented using the proposed topology in any dimension performs better. This has an implication of power consumption also.

The following tables (Table 4,5) quantify the parameters related to area of the systems. The tables below tabulate number of wire segments and wire length for the same.
From the above tables, it is quite clear that the area requirement for the proposed is almost same in lower configuration / levels but as the system scales to higher levels the wire requirement is comparatively very high when compared to the proposed topology. This effect is not only on the area requirement but also on the power dissipation of the system.Summing the above two results it is evident and clear that the proposed architecture performs better with lesser area / power requirement. Apart from the analysis above, the proposed topology was subjected to real time scenarios under different buffer and load / traffic conditions to test the strength of the topology which also gave positive results.

5.0 CONCLUSION

This paper introduces a new three dimensional topology SMITHA (Scalable Modular Interconnect for Three dimensional High performance Applications). The discussion starts with the two dimension variant and extends it to three dimension. The proposed architecture is structured and scalable. The paper quantifies the different performance / area parameters namely maximum hop, average hop. wirelength, number of wire segments in both two / three. dimension. This is compared with those of the current popular topologies namely mesh and torus. The growth of these parameters is found to be slow in case of the proposed topology implying that a system implemented performs better with smaller die area / power consumption.

Patent Information
The architecture discussed in this paper is applied for patent with the following details
Type : India
Number : 1598/CHE/2014
Date : 26/03/2014

6.0 REFERENCES

[1]. William J. Dally , Brian Towles, Route Packets Not Wires : on chip interconnection network, DAC 2001 June 2001.
[2]. Tobias Bjerregaard And Shankar Mahadevan, A Survey of Research and Practices of Network_on_Chip, ACM Computing Survey March 2006.
[3]. Ville Rantala, Teijo Lehtonen, Juha Plosila, Network On Chip Routing Algorithms, TUCS Technical Report No 779, August 2006.
[4]. Ahmed Hemani, Axel Jantsch, Shashi Kumar, Adam Postula, Johnny berg, Mikael Millberg,Dan Lindqvist, Network on a Chip: An architecture for billion transistor era, DAC 2001.
[5]. Radu Marculescu, IEEE, Natalie Enright Jerger, Yatin Hoskote, Umit Y. Ogras and Li Shiuan Peh, Outstanding Research Problems in NoC Design: System, Micro architecture, and Circuit Perspectives, IEEE Transactions on computer aided design of integrated circuits and systems, vol. 28, no. 1, January 2009.
[6]. J. Nurmi: Network on Chip: A New Paradigm for System on Chip Design. Proceedings 2005 International Symposium on System on Chip, 15 17 November 2005.
[7]. L. Benini, G. De Micheli. Networks on chips: A new SoC paradigm. IEEE Computer.35(1), 2002
[8]. S. Kumar, A. Jantsch, J. Soininen, M. Forsell, M.Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, A Network on Chip Architecture and Design Methodology, Proceedings International Symposium VLSI (ISVLSI), pp. 117-124, 2002.
[9]. Sanju V, Niranjan Chiplunkar, Mohd Khalid, P Venkata Krishna, Performance Analysis Through Simulation For Three Dimension Network Of Chip Based System, Anusandhana – Journal Of Science, Engineering And Management, Vol No 01, Issue No 02, Dec 2012, Pg. 18 – 29.

Table 1: Depicting Average Hop For A Single Level

Number of Layers	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of Nodes	2	6	14	30	62	126	252	510	1022	2046
Average Hop	.5	1.2	2.17	3.29	4.65	6.21	7.9	9.75	11.64	13.5

Table 2: Table Quantifying The Parameter Maximum Hop

No Of Nodes	Topology Number of Levels								
	Mesh 1	Torus 1	SMITHA 1	Mesh 2	Torus 2	SMITHA 2	Mesh 3	Torus 3	SMITHA 3
4	2	2	1	3	3	3	4	4	5
8	4	3	3	5	4	4	6	5	6
16	6	4	5	7	5	6	8	6	8
32	10	6	7	11	7	8	12	8	10
64	14	8	9	15	9	10	16	10	12
128	22	12	11	23	13	12	24	14	14
256	30	16	13	31	17	14	32	18	16
512	46	24	15	47	25	16	48	26	18
1024	62	32	17	63	33	18	64	34	20
2048	94	48	19	95	49	20	96	49	22

Table 3: Table Quantifying The Parameter Average Hop

No Of Nodes	Topology Number of Levels								
	Mesh 1	Torus 1	SMITHA 1	Mesh 2	Torus 2	SMITHA 2	Mesh 3	Torus 3	SMITHA 3
4	1.00	1.00	0.50	6.00	6.00	1.25	17.00	15.00	1.94
8	1.75	1.50	1.20	9.00	8.00	2.03	23.75	19.50	2.65
16	2.50	2.00	2.17	12.00	10.00	3.06	30.50	24.00	3.71
32	3.87	3.00	3.29	17.50	14.00	4.35	42.88	33.00	5.05
64	5.25	4.00	4.65	23.00	18.00	5.87	55.25	42.00	6.61
128	7.93	6.00	6.21	33.75	26.00	7.55	79.44	60.00	8.33
256	10.63	8.00	7.90	44.50	34.00	9.35	103.63	78.00	10.16
512	15.96	12.00	9.75	65.88	50.00	11.23	151.72	114.00	12.06
1024	21.31	16.00	11.64	87.25	66.00	13.16	199.81	150.00	14.00
2048	31.98	24.00	13.50	129.94	98.00	15.12	295.86	222.00	15.96

Table 4: Table Quantifying the Parameter Number Of Wire Segments

No Of Nodes	Topology Number of Levels								
	Mesh 1	Torus 1	SMITHA 1	Mesh 2	Torus 2	SMITHA 2	Mesh 3	Torus 3	SMITHA 3
4	4	8	1	12	24	3	20	36	5
8	10	16	8	28	48	18	46	72	28
16	24	32	23	64	96	49	104	144	75
32	52	64	54	136	192	112	220	288	170
64	112	128	117	288	384	239	464	576	361

No Of Nodes	**Topology Number of Levels**								
	Mesh 1	**Torus 1**	**SMITHA 1**	**Mesh 2**	**Torus 2**	**SMITHA 2**	**Mesh 3**	**Torus 3**	**SMITHA 3**
128	232	256	244	592	768	494	952	1152	744
256	480	512	499	1216	1536	1005	1952	2304	1511
512	976	1024	1010	2464	3072	2028	3952	4608	3046
1024	1984	2048	2033	4992	6144	4075	8000	9216	6117
2048	4000	4096	4080	10048	12288	8170	16096	18432	12260

Table 5: Table Quantifying the Parameter Wirelength

No Of Nodes	Topology Number of Levels								
	Mesh 1	Torus 1	SMITHA 1	Mesh 2	Torus 2	SMITHA 2	Mesh 3	Torus 3	SMITHA 3
4	4	8	1	12	48	3	20	72	5
8	10	20	8	28	96	18	46	144	28
16	24	48	23	64	192	49	104	288	75
32	52	104	54	136	384	112	220	576	170
64	112	224	117	288	768	239	464	1152	361
128	232	464	244	592	1536	494	952	2304	744
256	480	960	499	1216	3072	1005	1952	4608	1511
512	976	1952	1010	2464	6144	2028	3952	9216	3046
1024	1984	3968	2033	4992	12288	4075	8000	18432	6117
2048	4000	8000	4080	10048	24576	8170	16096	36864	12260