Performance Analysis of Massively Parallel Architectures

Z. A. Khan[1], J. Siddiqui[2] and **A. Samad[3]**

Submitted in October 2012, Accepted in April 2013

Abstract - Cube based networks have received much attention over the past decade since they offer a rich interconnected structure with a number of desirable properties such as low diameter, high bisection width, lesser complexity and Cost. Among them the hypercube architecture is widely used network for parallel computer system due to its low diameter. The major drawback of hypercube based architectures is the difficulty of its VLSI layout. Several variations of hypercube have also been reported which are designed by considering a specific topological property. Nevertheless, no particular topology claims to have better performance with all the desirable topological properties. In this paper the performance analysis of various interconnection networks is presented. The performance is compared by considering cube type architectures as well as linear type architectures on different parameters such as degree, diameter, bisection width, scalability and cost etc. The Analysis indicates that cube based architectures have a rich interconnected structure with high cost and complexity. On the other hand linear type architectures are scalable, simpler and better in terms of cost and complexity. The comparative study suggests the various aspects to the design of new multiprocessor architectures.

Index Terms - Interconnection network, Performance evaluation, Topological properties, Parallel system, Cube Architectures

1. INTRODUCTION

systems interconnection networks play an important role in the overall performance of the system. Deciding the appropriate network is an important issue in the design of parallel and distributed systems. In general, determining the optimal network to implement any parallel application does not have a known theoretical solution. There are different ways to determine efficient topologies that trade-off high level performance issues against various implementation constraints [1]. A Topology is evaluated in terms of a number of performance parameters such as degree, diameter, bisection width and cost. Several researchers have developed various architectures which are considered better in terms of particular parameters.

[1,2]*Dept. of Computer Science, Aligarh Muslim University, Aligarh (India)-202002*
[3]*University Women's Polytechnic, Aligarh Muslim University, Aligarh (India)-202002*
E-Mail: [1]*khanzaki05@gmail.com,*
[2]*jamshed_faiza@rediffmail.com and*
[3]*abdussamadamu@gmail.com*

Some variations focus on the reduction of the diameter [10] [18], some of them focused on the design of simple routing and communication algorithm [4]. Scalability is also an important issue to evaluate the performance of interconnection networks. However, it can't be clearly mentioned that which interconnection network is working better by considering all the parameters. In terms of complexity interconnection networks may be classified into two major categories. The first is cube based architectures which posses a rich interconnection topology. The Binary hypercube or n-cube has been widely used interconnection network in the design of parallel systems [12]. Several variations of hypercube architecture are reported in the literatures some examples are –folded hypercube (FDC), metacube (MC), folded metacube (FMC) and folded dualcube (FDC) etc. [8] [7] [12] [13] [15] [11] . The major drawback in such networks is the increase in the number of communication links for each node and the increase in the total number of nodes in the system which ultimately enhances the complexity of such interconnection networks [19] [20]. Therefore, there is a need to carry out the performance analysis of various interconnection networks by considering their topological properties.

The second class of the network is linearly extensible networks such as linear array, ring, linearly extensible tree and linearly extensible cube etc [10] [16]. The complexity of these networks is lesser as they do not have exponential expansion. Besides the scalability, other parameters to evaluate the performance of such networks are degree, number of nodes, diameter, bisection width and fault tolerance. The main purpose of this paper is to study and analyse the various multiprocessor networks along with their properties to help in the design of a new interconnection architecture. Selection of a better interconnection network may have several applications with lesser complexities and improved power-efficiency. One such modern application is network on chip (NoC) paradigm where different cores are embedded with appropriate connectivity. Some examples may include mesh, torus, star, etc. [1] [9].

In this paper, the study of five cubes based architectures as well as several linear extensible architectures are carried out. Section 2 describes, the various parameters used to make the performance analysis. Various parameters used to compare the performance of cube based architectures and their characteristic is discussed in section 3. Similarly, the comparative analysis of linearly extensible architectures is carried out in section 4. A comparative study of both the type of architectures is made in section 5 and finally concluded the paper in section 6.

2. PERFORMANCE PARAMETERS

The need for architectural performance evaluation exists from design phase to its installation. The various parameters decide the design alternatives and gives a criterion of selection known as cost performance trade off [6] [3]. In general, the performance of various architectures is measured by the following parameters.

A. Degree (d)

It is connectivity among different nodes in a network. The connectivity of the nodes determines the complexity of the network. The greater number of links in the network means greater is the complexity.

B. Diameter (D)

It is defined as the maximum shortest path between the source and destination node. The path length is measured by the number of links traversed. This virtue is important in determining the distance involved in communication and hence the performance of parallel systems.

C. Bisection width (B)

The bisection width of a network is the minimum number of edges whose removal will result in two distinct sub networks. Greater bisection width is better for a network to be fault tolerant.

D. Cost (C)

It is defined as the product of the diameter and the degree of the node for the asymmetric network. (i.e. Cost = D*d). This factor is widely used in performance evaluation.

E. Extensibility

This is the virtue which facilitates large sized system out of small ones with minimum changes in the configuration of the nodes. It is the smallest increment by which the system can be expanded in a useful way.

In the Present work the above parameters are compared for different types of multiprocessor architectures. The values are computed based on a certain mathematical formula designed for specific topology.

3. CUBE BASED ARCHITECTURES

A. Hypercube

The Binary hypercube or n-cube has been one of the most popular interconnection networks having logarithm diameter [12]. Each node in this network is connected through bidirectional asynchronous point-to-point communication link to other nodes. The major drawback of the hypercube is the increase in the number of communication links for each node with the increase in the total number of nodes in the system[17]. The hypercube has a high bisection width $b=2^{n-1}$ and has good capability of fault tolerance.

B. Folded Hypercube

The folded hypercube (FHC) is a standard hypercube with some extra links established between its nodes [2]. A folded hypercube of dimension n is called FHC (n). The FHC (n) is constructed from a standard hypercube by connecting each node to the unique node that is farthest from it. The FHC (n) is a regular network of node connectivity (n+1) and the hypercube of degree 3 is converted to FHC network as show in Figure 1. The diameter of an FHC (n) is (n/2) and bisection width is $2^n/4$.

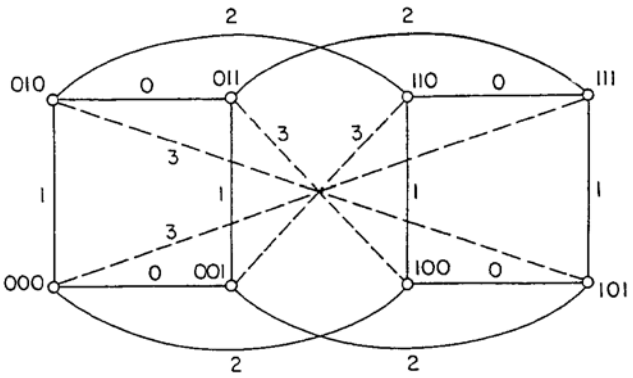

Figure 1: Folded Hypercube FHC (3)

C. Metacube

The metacube (MC) is an interconnection network for a very large parallel computer. In this network, the number of nodes is much larger than hypercube with a small number of links per node [4] [14]. The metacube network shares many desired virtues of the hypercube such as small diameter. The metacube (MC) network includes the dual-cube as a special case. The MC network has two level cube structure a high-level cube (classes) represented by the k- dimension and low- level cube (cluster) represented by m-dimension. An MC (k, m) network can connect 2^{k+m2k} nodes with (k+m) links per node. The degree is $m+k= (n-k)/2^k+k$ and the bisection width of an MC (k, m) is $2^{m2k}/2$.

D. Folded Metacube

The folded metacube is an interconnection topology which inherits some of the useful properties of the metacube and folded hypercube (FHC) [5]. The folded metacube is graph G (V, E) as show in Figure 2. Where V represents a set of vertices and E represents a set of links. The graph is a modified of metacube. The diameter of folded metacube is 2(m+k)-1 and the Bisection width of G is $2^{m2k}/2 + 2^{m2k+k-2}$.

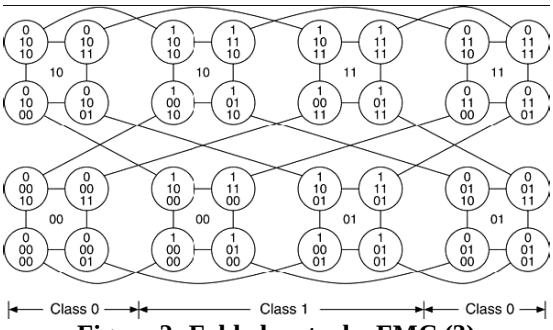

Figure 2: Folded metcube FMC (3)

E. Folded Dualcube

The Folded dualcube (FDC) is a cube based topology which inherits some of the useful properties of the dualcube [8] and the folded hypercube (FHC) [2]. The folded dualcube, which is constructed by connecting each node farthest from it and is show in Figure 3.

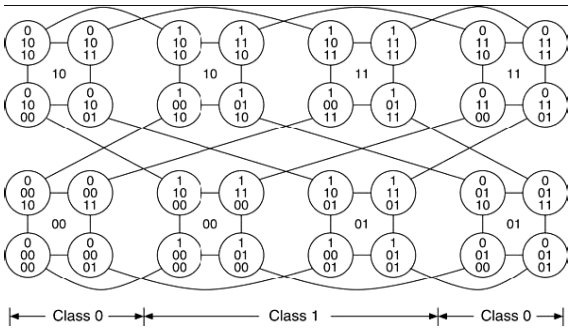

Figure 3: Folded dualcube FDC (3)

The nodes connectivity of folded dualcube is (n+3)/2, the diameter is n-1 and having bisection width is $2^n/2$ [5]. The Various parameters of cube based architectures along with the topological properties are summarized in Table 1.

Type	Nodes	Degree (d)	Diameter (D)	B.W	Cost	Extensibility
HC	2^n	n	n	2^{n-1}	n^2	Exponential
FHC	2^n	n+1	n/2	2^n	n/2*n+1	Exponential
MC	2^n	$(n-k)/2^k+k$	2^{k+1}	$2^{2k}/2$	$(n-k)/2^k+k*2^{k+1}$	Exponential
FMC	2^n	(n+1)	2n-1	$2^{2k}/2+2^{2k}+k-2$	(n+1) *(2n-1)	Exponential
FDC	2^{2n-1}	(n+3)/2	n-1	$2^n/2$	(n+3)/2 * (n-1)	Exponential

Table 1: Various parameters of Cube based Architectures

4. LINEAR EXTENSIBLE ARCHITECTURES

A. Linear Array

It is one dimensional network having the simplest topology with n-nodes having N-1 communication links. The internal nodes have degree 2 and the termination nodes have degree1. The diameter is N-1, which is long for large N and the bisection width is 1. It is asymmetric network.

B. Binary tree

The binary tree is scable architecture with a constant node degree and constant bisection width. In general, an n-level, complexity balanced binary tree should have $N=2^n-1$ nodes. The maximum node degree is 3 and the diameter is 2(n-1).

C. Linearly Extensible Tree

A binary type network topology has been reported [10] shown in Figure 4. The Linearly Extensible Tree (LET) architecture exhibits better connectivity, lesser number of nodes over cube based networks. The LET network has low diameter, hence reduce the average path-length traveled by all message and contains a constant degree per node. The LET network grows linearly in a binary tree like shape. In a binary tree the number of nodes at level n is 2n whereas in LET network the number is (n+1).

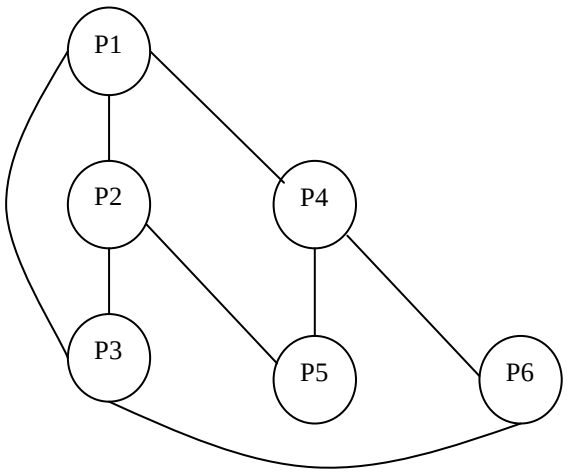

Figure 4. Linearly Extensible Tree (LET) network

D. Linearly Extensible Cube

The Linearly Extensible Cube (LEC) network grows linearly and posses some of the desirable topological properties such as small diameter [10], high connecting constant node degree with high scalability. It has a constant expansion of only two processors at each level of the extension while preserving all the desirable topological properties. The LEC network can maintain a constant node degree regardless of the increase in size (i.e. number of nodes) in a network.

The number of nodes in LEC network is 2*n for n>0 where the number of nodes in the hypercube is 2^n. The diameter of network is $\lfloor N \rfloor$. It has a constant node degree 4. The LEC has a bisection width equal to N**,** as show in Figure 5**.**

Figure 5: Linearly Extensible Cube (LEC) network

E. Ring

A ring is obtained by connecting the two terminal nodes of a linear array with one extra link. A ring network can be uni-or bidirectional and it is symmetric with a constant. It has a

constant node degree of d=2, the diameter is $\lfloor N/2 \rfloor$ for a bidirectional ring and N for unidirectional ring. A ring network has a constant width 2. The different performance parameters of Linearly Extensible Architectures are summarized in Table2.

Type	Nodes	Degree (d)	Diameter (D)	B.W	Cost	Extensibility
LET	$N=\sum_{k=1}^{n} k$	4	$\sqrt{N}$	2log(n-2)	$4\sqrt{N}$	Linear
LEC	N=2*n	4	$\lfloor N \rfloor$	N	$^4\lfloor N \rfloor$	Linear
L.A	N	2	N-1	1	2(N-1)	Linear
Ring	N	2	$\lfloor N/2 \rfloor$	2	$^2\lfloor N/2 \rfloor$	Linear
B.T	$N=2^n-1$	3	2(n-1)	1	6(n-1)	Linear

Table 2: Various parameters of Linearly Extensible Architectures.

5. COMPRATIVE STUDY OF VARIOUS ARCHITECTURE

For multiprocessor network parameters such as diameter, degree, bisection width, cost regularity and symmetry are crucial and determine the performance of the network to compare the performance. We proceed to consider the three important parameters namely, number of processors, diameter and cost. The curves are plotted for each of the parameters for both the class of interconnection networks. Figure 6 shows the trained of increasing number of processors for each level of the extension. It is observed that all the linearly extensible architectures except binary tree have lesser number of processors. Therefore, the complexity of linearly extensible architectures is lesser, when they are expanded on higher level. Having lesser number of processors to implement a parallel algorithm is always economical. On the other hand the cube based architectures have exponential expansions which make the network highly complex. The Figure 6 also shows that among linearly extensible architectures, the LEC network produces better results.

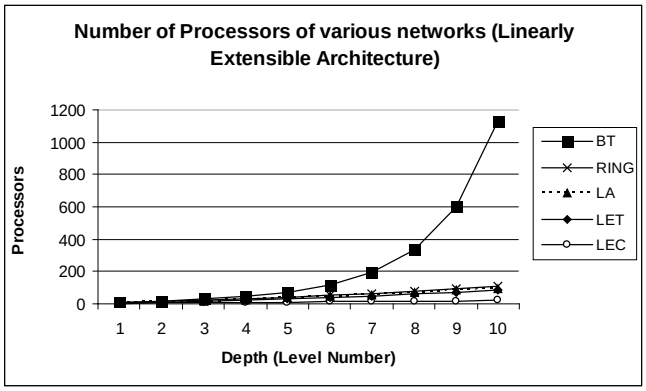

Figure 6: Performance of level extensible architectures

The second parameter when analyzing the performance of both the type of architectures is diameter. To analysis the diameter of various networks the curves are plotted and show in Figure 7 and 8. The study of the results in both the curves shows that the results in both the types of network are comparable. Among

cube based architectures, folded hypercube architectures has lesser diameter as compare to other cube based architectures (Figure 7). When comparing the results of linearly extensible architectures the LEC networks has lesser diameter as compare to other similar architectures.

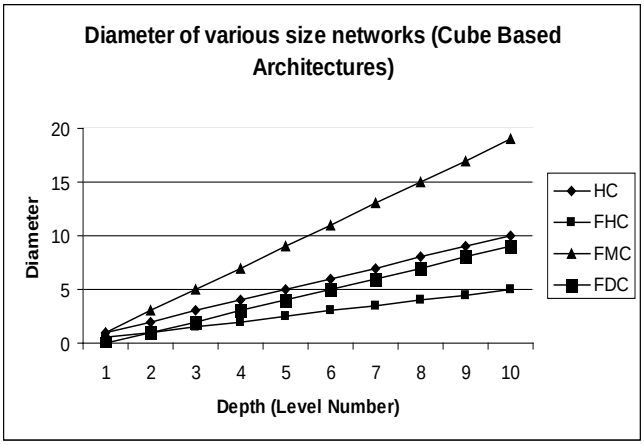

Figure 7: Performance of Cube based architectures

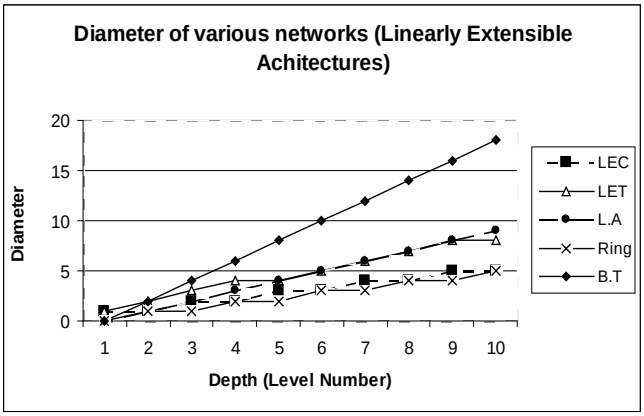

Figure 8: Performance of linearly extensible architectures

The main parameter in terms of evaluating the performance is cost which is defined as the product of the degree and the diameter. Figure 9 and 10 depicts the patterns of the cost analysis of both the class of networks. In cube based network FHC is having lesser cost at greater level as compare to other similar cubical architectures (Figure 9). Similarly, when comparing the cost of linearly extensible architectures, Figure 9 shows that LET is having lesser cost in comparing to other linear types of architectures.

To clearly draw the conclusion, the cost analysis of those architectures is carried out which are giving better results in their respective categories. Therefore, when comparing the cost of FHC and LET, it is observed that LET network has lesser cost at higher level as compare to FHC However, the results are comparable.

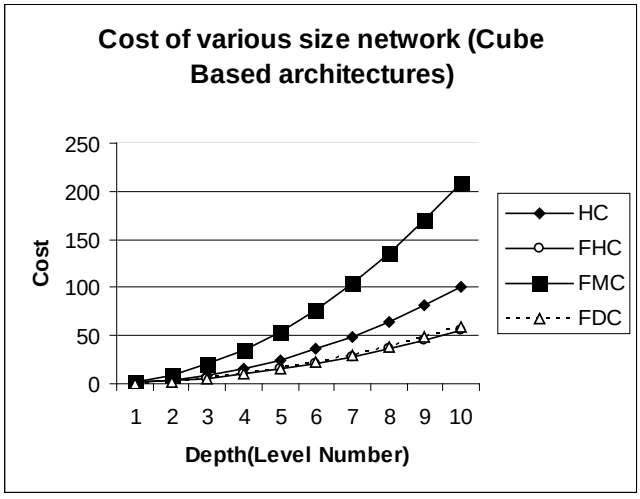

Figure 9: Performance of Cube based architectures

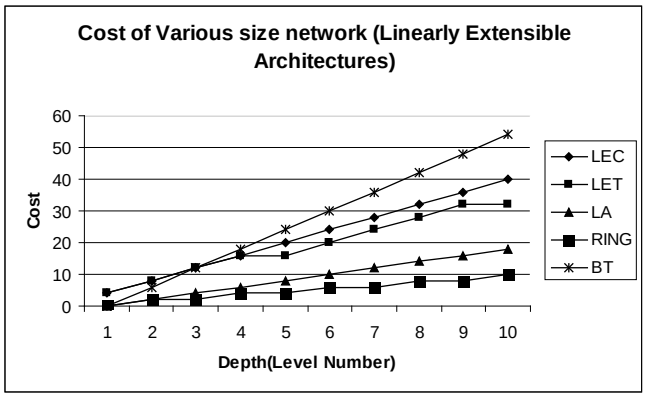

Figure 10: Performance of linearly extensible architectures

The bisection width is also an important parameter for measuring the performance of multiprocessor architectures. The bisection width in cube type architectures is of exponential value. In case of linearly extensible architectures the bisection width is either constant or increases linearly with the increase in number of processors. The linear increment is not desirable, as such, connection at higher level of architectures do not seem to reflect the practical fault tolerance capability of the network.

6. CONCLUSION AND FUTURE SCOPE

In this paper the performance of various multiprocessor architectures are analyzed by considering their topological properties. The comparative study of cube based as well as linearly extensible architectures is made. In cube based networks, it is evaluated that the FHC is giving better performance in terms of diameter and cost. However, all the cube based architectures have exponential expansion which increases the complexity of the system. If we limit the number of processors in FHC it can be considered as best multiprocessor network with high degree of fault tolerance. There is a great scope to modify this network so that it can have approximately all the desirable topological properties with lesser number of processors. As far as linearly extensible architectures are considered they are less complex and easily extensible. However, the common drawback is that they are having low bisection width, which is not a desirable property to make the network fault tolerant.

The important issue in the design of multiprocessor systems is how to cope with the problem of an adequate design of the interconnection network in order to achieve the desired performance at low cost. The choice of the interconnection network may affect several characteristics of the system such as node complexity, scalability and cost etc. The present study is carried out on the basis of several characteristics of various multiprocessor interconnection networks. There have been more work related to design of appropriate multiprocessor network; however no one claims a particular design which entrenched all the desirable properties. The present study gives more scope to design high performance interconnection network that can be used in the design of multiprocessor server.

REFERENCES

[1]. Shafi Patel, Parag Parandkar, Sumant Katiyal and Ankit Agarwal, "Exploring Alternative Topologies for Network-on-Chip Architectures," BIJIT - BVICAM's International Journal of Information Technology - 2011 Vol. 3 No. 2, ISSN 0973 – 5658

[2]. Amway, A. E. and Latifi, S, "Properties and performance of folded hypercubes," IEEE Transactions on Parallel and Distributed Systems, Vol. 2, 1991, pp. 31– 42.

[3]. Harvendra Kumar and A.K. Verma, "Comparative Study of Distributed Computing Paradigms," BIJIT - BVICAM's International Journal of Information Technology – 2009 Vol. 1 No. 2, ISSN 0973 – 5658.

[4]. Yamin Li, Shietung Peng and Wanming Chu, "Metacube: A New Interconnection Network for Large Parallel System," ACSAC02, Australian Computer Science Communications, Vol. 24, No. 4, 2001, pp.29-36.

[5]. N. Adhikari and C.R. Tripathy, "Folded Metacube: An Efficient Large scale Parallel network," 04/2009; DOI:10.1109/IADCC.2009.4809200 in Proceedings of advanced Computing conference, 2009, IACC 2009, IEEE International.

[6]. Neeraj Kumar, "Simulation Study for Performance and Prediction of Parallel Computers," BIJIT - BVICAM's International Journal of Information Technology - 2012, Vol. 4 No. 2, ISSN 0973 – 5658.

[7]. N. Adhikari and C.R. Tripathy, "Folded Dualcube: A New interconnection for Parallel Systems," Proceedings

of 11th Int. Conf. on Information Technology, 2008, 17-18 Dec, pp.-75-78, IEEE, Comp. Society.

[8]. Yamin Li and Shietung Peng, "Dualcube: A New Interconnection Network for High-Performance computer Clusters," International computer symposium, workshop on Computer architecture. December 6-8, 200, ChiaYi, Taiwan.

[9]. Parag Parandkar, Jayesh Kumar Dalal and Sumant Katiyal, "Performance Comparison of XY, OE and DY ad Routing Algorithm by Load Variation Analysis of 2-Dimensional Mesh Topology Based Network-on-Chip," BIJIT - BVICAM's International Journal of Information Technology -2012 Vol. 4 No. 1 ISSN 0973 – 5658

[10]. Samad A, Rafiq M. Q. and Farooq O., "*LEC: An Efficient Scalable Parallel Interconnection Network*," in proceedings of International Conference on Emerging Trends in Computer Science, Communication and Information Technology (CSCIT2010) Nanded, India.

[11]. Vinay Kumar, "Restricted Backtracked Algorithm For Hamiltonian Circuit in Undirected Graph," BIJIT - BVICAM's International Journal of Information Technology – 2010 Vol. 2 No. 2 ISSN 0973 – 5658.

[12]. Y. Saad and M.H. Schultz, "Topological properties of hypercubes," IEEE Trans. Computer. Vol.37, No. 7, 1988, pp.867–872.

[13]. N. Adhikari and C.R. Tripathy, "Folded Crossed cube: A New interconnection for Parallel Systems," International Journal of computer Application (0975-8887) Volume 4, p43. July 2010.

[14]. Yamin Li, Shietung Peng and Wanming Chu, "Efficent Collective Communications in Dual-cube," The journal of super computing, 28, pp.71-90, 2004.

[15]. Y. Q. Zhang, "Folded-crossed hypercube: A CompleteInterconnection Network," Journal of SystemArchitecture, vol. 47, 2002, pp. 917-922, Elsevier Science.

[16]. Nishant Doshi, Tarun sureja, Bhavesh akbari, Hiren Savaliya and Viraj Daxini, "width of Binary Tree,"International journal of computer Applications *0975 – 8887)* Volume*9– No.2, November 2010.*

[17]. S.R. Deshpande and R.M. Jenevein, "Scalability of a binary tree on a hypercube," in proc. Int. Conf. Parallel Processing, 1986, pp. 661 -668.

[18]. L.N. Bhuyan and D.P. Agrawal, "Performance of Multiprocessor Interconnection Network," IEEE Computer,1989 .

[19]. Y.C. Liu, J.F. Fang and C.C. Wu, "On Extensibilities Of Interconnection Networks," IE.(I) Journal-CP,PP.13-16, 2004.

[20]. S. P. Mohanty, B. N. B. Ray, S. N. Patro and A. R. Tripathy, "Topological Properties Of a New Fault Tolerant Interconnection Network for Parallel Computer," icit, pp. 36-40, 2008 International Conference on Information Technology, 2008.