UNIT 3 INTERCONNECTION NETWORK

Structure **Page Nos.**

3.0 Introduction 46
3.1 Objectives 47
3.2 Network Properties 47
3.3 Design issues of Interconnection Network 49
3.4 Various Interconnection Networks 50
3.5 Concept of Permutation Network 55
3.6 Performance Metrics 63
3.7 Summary 63
3.8 Solution /Answers 63

3.0 INTRODUCTION

This unit discusses the properties and types of interconnection networks. In multiprocessor systems, there are multiple processing elements, multiple I/O modules, and multiple memory modules. Each processor can access any of the memory modules and any of the I/O units. The connectivity between these is performed by interconnection networks.

Thus, an interconnection network is used for exchanging data between two processors in a multistage network. Memory bottleneck is a basic shortcoming of Von Newman architecture. In case of multiprocessor systems, the performance will be severely affected in case the data exchange between processors is delayed. The multiprocessor system has one global shared memory and each processor has a small local memory. The processors can access data from memory associated with another processor or from shared memory using an interconnection network. Thus, interconnection networks play a central role in determining the overall performance of the multiprocessor systems. The interconnection networks are like customary network systems consisting of nodes and edges. The nodes are switches having few input and few output (say n input and m output) lines. Depending upon the switch connection, the data is forwarded from input lines to output lines. The interconnection network is placed between various devices in the multiprocessor network.

The architecture of a general multiprocessor is shown in *Figure 1*. In the multiprocessor systems, these are multiple processor modules (each processor module consists of a processing element, small sized local memory and cache memory), shared global memory and shared peripheral devices.

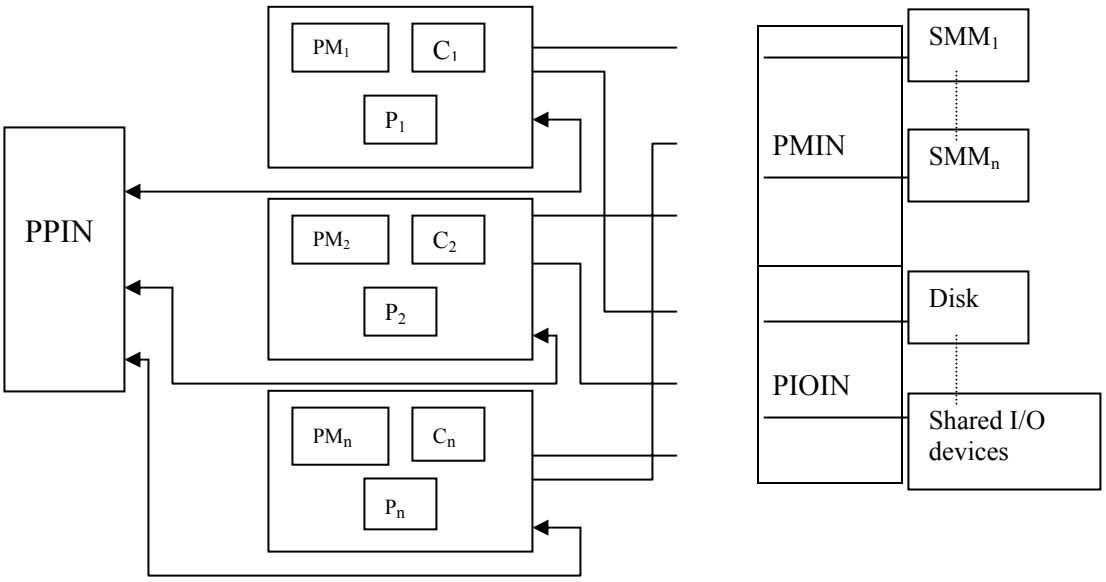


Figure 1: General Multi-Processor

PMIN = Processor to Memory Interconnection Network
PIOIN= Processor to I/O Interconnection Network
PPIN = Processor to Processor Interconnection Network
PM = Processor Module

Module communicates with other modules shared memory and peripheral devices using interconnection networks.

3.1 OBJECTIVES

After studying this unit, students will be able to understand

- discuss the meaning and needs of interconnection network;
- describe the role of interconnection network in a multiprocessor system;
- enumerate the types of interconnection network;
- explain the concept of permutation network;
- discuss the various interconnection networks, and
- describe how matrix multiplication can be carried out on an interconnection network.

3.2 NETWORK PROPERTIES

The following properties are associated with interconnection networks.

1) *Topology:* It indicates how the nodes a network are organised. Various topologies are discussed in Section 3.5.

2) *Network Diameter*: It is the minimum distance between the farthest nodes in a network. The distance is measured in terms of number of distinct hops between any two nodes.

3) *Node degree*: Number of edges connected with a node is called node degree. If the edge carries data from the node, it is called out degree and if this carries data into the node it is called in degree.

4) *Bisection Bandwidth:* Number of edges required to be cut to divide a network into two halves is called bisection bandwidth.

5) *Latency:* It is the delay in transferring the message between two nodes.

6) *Network throughput:* It is an indicative measure of the message carrying capacity of a network. It is defined as the total number of messages the network can transfer per unit time. To estimate the throughput, the capacity of the network and the messages number of actually carried by the network are calculated. Practically the throughput is only a fraction of its capacity.
In interconnection network the traffic flow between nodes may be nonuniform and it may be possible that a certain pair of nodes handles a disproportionately large amount of traffic. These are called "hot spot." The hot spot can behave as a bottleneck and can degrade the performance of the entire network.

7) *Data Routing Functions:* The data routing functions are the functions which when executed establishe the path between the source and the destination. In dynamic interconnection networks there can be various interconnection patterns that can be generated from a single network. This is done by executing various data routing functions. Thus data routing operations are used for routing the data between various processors. The data routing network can be static or dynamic static network

8) *Hardware Cost:* It refers to the cost involved in the implementation of an interconnection network. It includes the cost of switches, arbiter unit, connectors, arbitration unit, and interface logic.

9) *Blocking and Non-Blocking network:* In non-blocking networks the route from any free input node to any free output node can always be provided. Crossbar is an example of non-blocking network. In a blocking network simultaneous route establishment between a pair of nodes may not be possible. There may be situations where blocking can occur. Blocking refers to the situation where one switch is required to establish more than one connection simultaneously and end-to-end path cannot be established even if the input nodes and output nodes are free. The example of this is a blocking multistage network.

10) *Static and Dynamic Interconnection Network:* In a static network the connection between input and output nodes is fixed and cannot be changed. Static interconnection network cannot be reconfigured. The examples of this type of network are linear array, ring, chordal ring, tree, star, fat tree, mesh, tours, systolic arrays, and hypercube. This type of interconnection networks are more suitable for building computers where the communication pattern is more or less fixed, and can be implemented with static connections. In dynamic network the interconnection pattern between inputs and outputs can be changed. The interconnection pattern can be reconfigured according to the program demands. Here, instead of fixed connections, the switches or arbiters are used. Examples of such networks are buses, crossbar switches, and multistage networks. The dynamic networks are normally used in shared memory(SM) multiprocessors.

11) *Dimensionality of Interconnection Network:* Dimensionality indicates the arrangement of nodes or processing elements in an interconnection network. In single dimensional or linear network, nodes are connected in a linear fashion; in two dimensional network the processing elements (PE's) are arranged in a grid and in cube network they are arranged in a three dimensional network.

12) *Broadcast and Multicast :*In the broadcast interconnection network, at one time one node transmits the data and all other nodes receive that data. Broadcast is one to all mapping. It is the implementation achieved by SIMD computer systems. Message passing multi-computers also have broadcast networks. In multicast network many nodes are simultaneously allowed to transmit the data and multiple nodes receive the data.

3.3 DESIGN ISSUES OF INTERCONNECTION NETWORK

The following are the issues, which should be considered while designing an interconnection network.

1) *Dimension and size of network:* It should be decided how many PE's are there in the network and what the dimensionality of the network is i.e. with how many neighburs, each processor is connected.
2) *Symmetry of the network:* It is important to consider whether the network is symmetric or not i.e., whether all processors are connected with same number of processing elements, or the processing elements of corners or edges have different number of adjacent elements.

3) *What is data communication strategy?* Whether all processors are communicating with each other in one time unit synchronously or asynchronously on demand basis.

4) *Message Size:* What is message size? How much data a processor can send in one time unit.

5) *Start up time:* What is the time required to initiate the communication process.

6) *Data transfer time:* How long does it take for a message to reach to another processor. Whether this time is a function of link distance between two processors or it depends upon the number of nodes coming in between.

7) *The interconnection network is static or dynamic:* That means whether the configuration of interconnection network is governed by algorithm or the algorithm allows flexibility in choosing the path.

Check Your Progress 1

1) Define the following terms related with interconnection networks.
   i) Node degrees
   ii) Dynamic connection network
   iii) Network diameter

……
……
……

2) What is the significance of a bisection bandwidth?

……

……

……

3.4 VARIOUS INTERCONNECTION NETWORKS

In this section, we will discuss some simple and popularly used interconnection networks

1) Fully connected: This is the most powerful interconnection topology. In this each node is directly connected to all other nodes. The shortcoming of this network is that it requires too many connections.

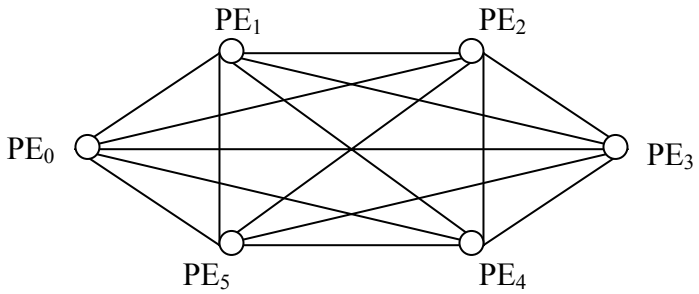


Figure 2: Fully connected interconnection topology

2) Cross Bar: The crossbar network is the simplest interconnection network. It has a two dimensional grid of switches. It is a non-blocking network and provides connectivity between inputs and outputs and it is possible to join any of the inputs to any output.

An N * M crossbar network is shown in the following *Figure 3 (a)* and switch connections are shown in *Figure 3 (b)*.

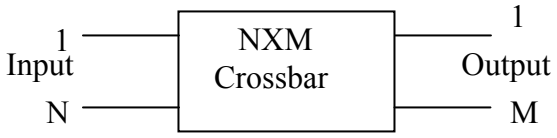


Figure: 3(a)

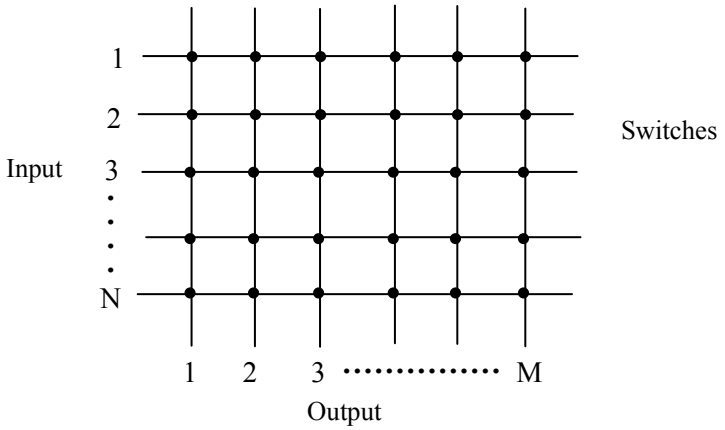


Figure: 3(b)
Figure 3: Crossbar Network

A switch positioned at a cross point of a particular row and particular column. connects that particular row (input) to column (output).

The hardware cost of N*N crossbar switch is proportional to N^2. It creates delay equivalent to one switching operation and the routing control mechanism is easy. The crossbar network requires N^2 switches for N input and N output network.

3) *Linear Array:* This is a most fundamental interconnection pattern. In this processors are connected in a linear one-dimensional array. The first and last processors are connected with one adjacent processor and the middle processing elements are connected with two adjacent processors. It is a one-dimensional interconnection network.

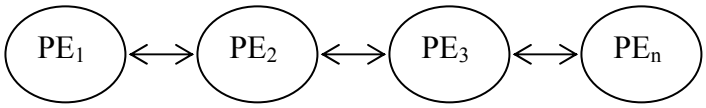


Figure 4: Linear Array

4) *Mesh:* It is a two dimensional network. In this all processing elements are arranged in a two dimensional grid. The processor in rows *i* and column *j* are denoted by PE_i.

The processors on the corner can communicate to two nearest neighbors i.e. PE_{00} can communicate with PE_{01} and PE_{10}. The processor on the boundary can communicate to 3 adjacent processing elements i.e. PE_{01} can communicate with $PE_{00,}$ PE_{02} and PE_{11} and internally placed processors can communicate with 4 adjacent processors i.e. PE_{11} can communicate with $PE_{01,}$ $PE_{10,}$ $PE_{12,}$ and PE_{21}

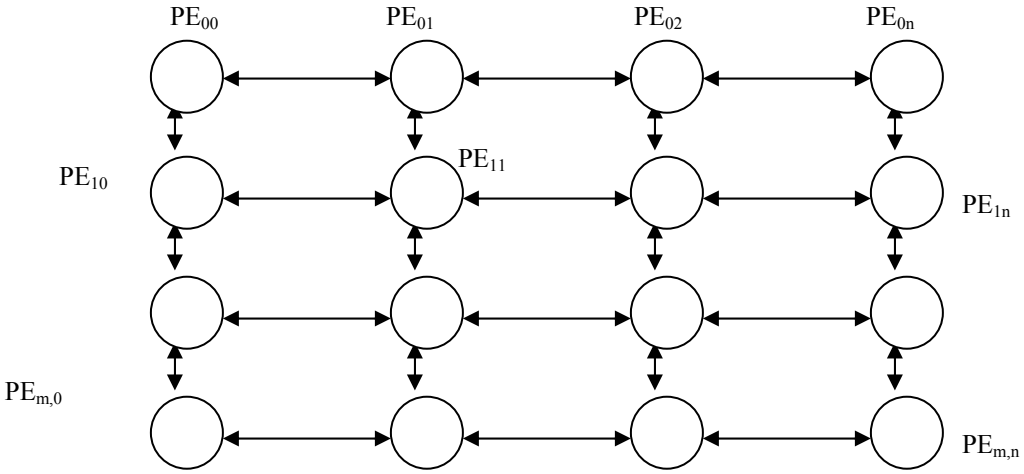


Figure 5: Mesh Network

5) *Ring:* This is a simple linear array where the end nodes are connected. It is equivalent to a mesh with wrap around connections. The data transfer in a ring is normally one direction. Thus, one drawback to this network is that some data transfer may require N/2 links to be traveled (like nodes 2 & 1) where N is the total number of nodes.

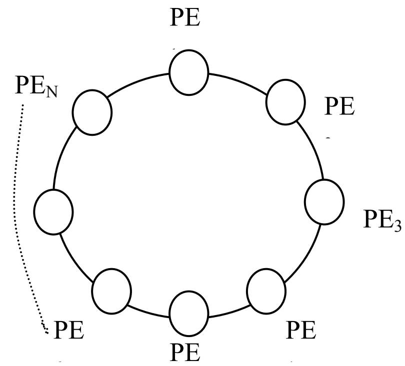


Figure 6: Ring network

6) Torus: The mesh network with wrap around connections is called Tours Network.

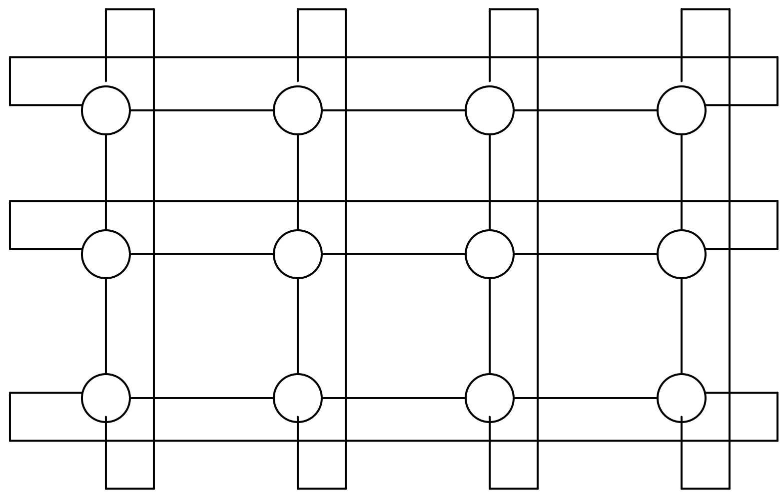


Figure 7: Torus network

7) Tree interconnection network: In the tree interconnection network, processors are arranged in a complete binary tree pattern.

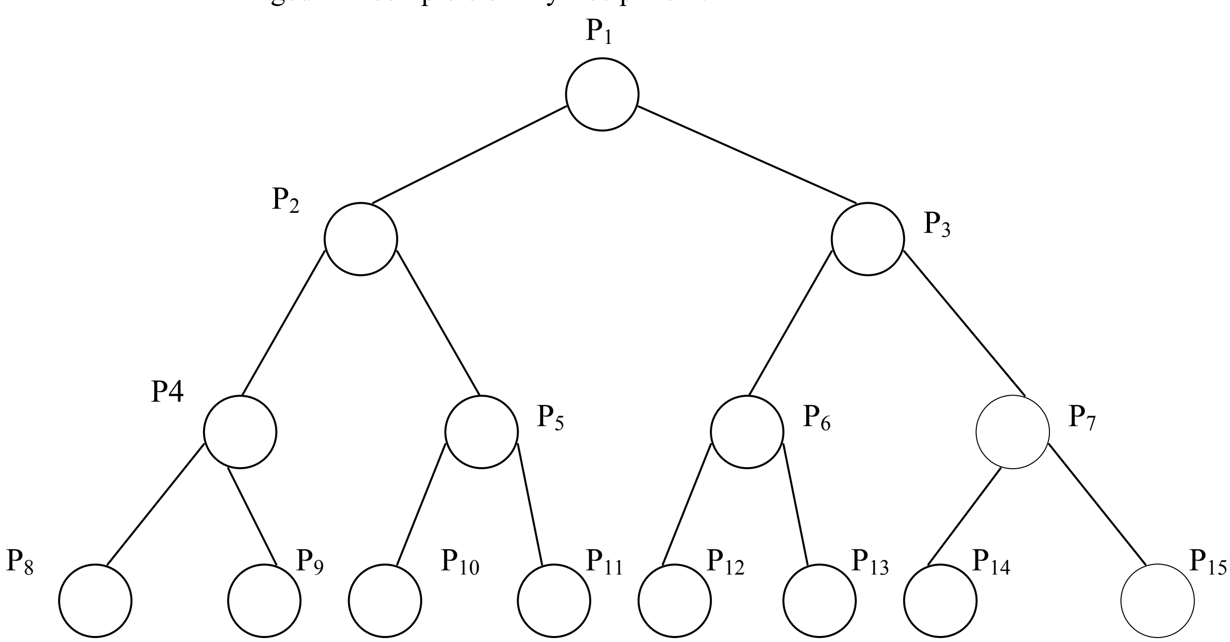


Figure 8:Tree interconnection network

8) Fat tree: It is a modified version of the tree network. In this network the bandwidth of edge (or the connecting wire between nodes) increases towards the root. It is a more realistic simulation of the normal tree where branches get thicker towards root. It is the more popular as compared to tree structure, because practically the more traffic occurs towards the root as compared to leaves, thus if bandwidth remains the same the root will be a bottleneck causing more delay. In a tree this problem is avoided because of higher bandwidth.

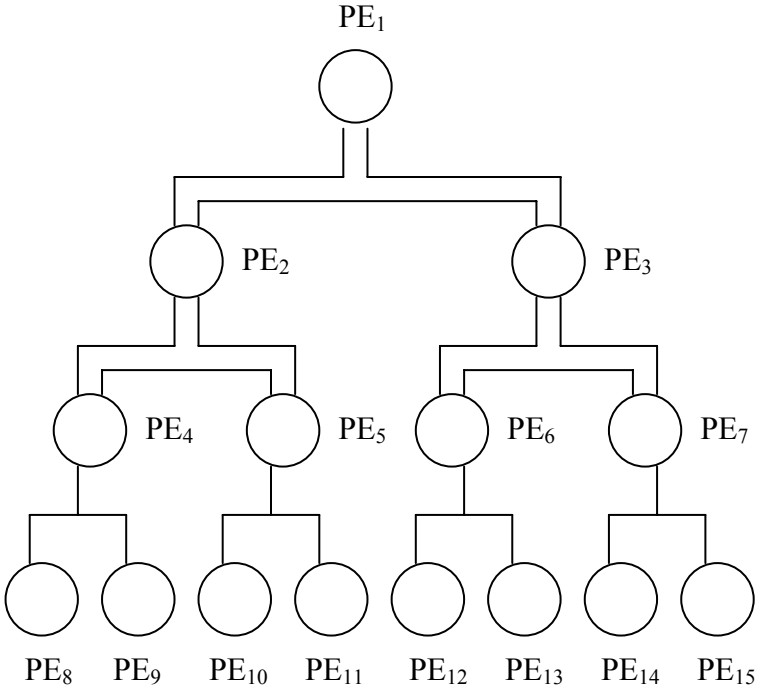


Figure 9: Fat tree

9) Systolic Array: This interconnection network is a type of pipelined array architecture and it is designed for multidimensional flow of data. It is used for implementing fixed algorithms. Systolic array designed for performing matrix multiplication is shown below. All interior nodes have degree 6.

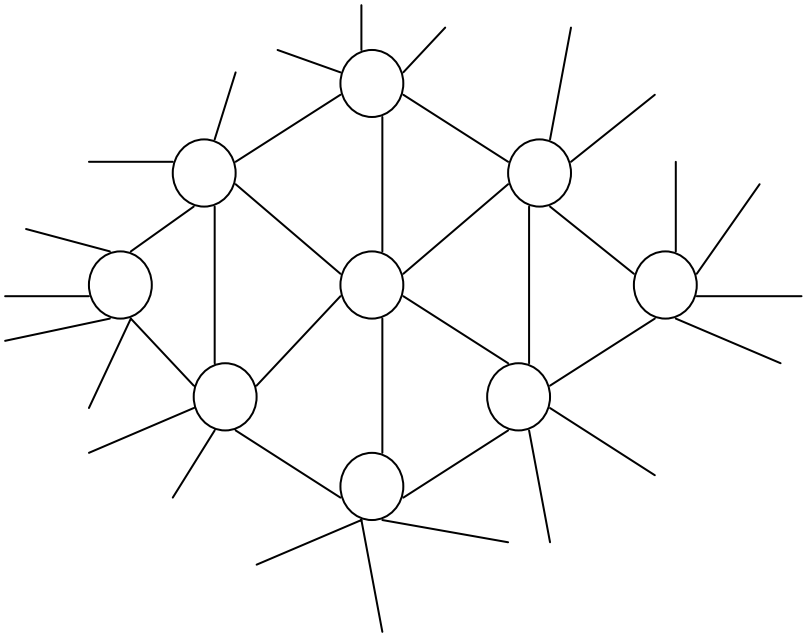


Figure 10: Systolic Array

10) Cube: It is a 3 dimensional interconnection network. In this the PE's are arranged in a cube structure.

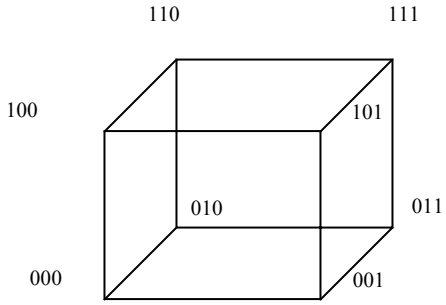


Figure 11: Cube interconnection network

11) Hyper Cube: A Hypercube interconnection network is an extension of cube network. Hypercube interconnection network for n ≥ 3, can be defined recursively as follows:

For n = 3, it cube network in which nodes are assigned number 0, 1, ……,7 in binary. In other words, one of the nodes is assigned a label 000, another one as 001…. and the last node as 111.
Then any node can communicate with any other node if their labels differ in exactly one place, e.g., the node with label 101 may communicate directly with 001, 000 and 111.

For n > 3, a hypercube can be defined recursively as follows:
Take two hypercubes of dimension (n – 1) each having (n –1) bits labels as 00….0, ……11…..1

Next join the two nodes having same labels each (n – 1) -dimension hypercubes and join these nodes. Next prefix '1' the labels of one of the (n – 1) dimensional hypercube and '0' to the labels of the other hypercube. This completes the structure of n-dimensional hypercube. Direct connection is only between that pair of nodes which has a (solid) line connecting the two nodes in the pair.

For n = 4 we draw 4-dimensional hypercube as show in *Figure 12*:

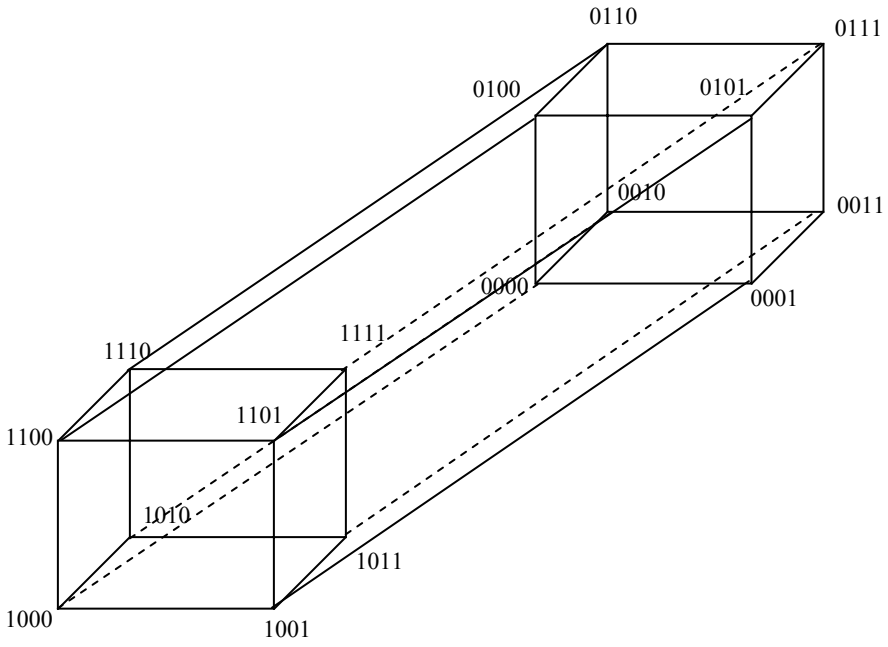


Figure 12: 4-Dimensional hypercube

3.5 CONCEPT OF PERMUTATION NETWORK

In permutation interconnection networks the information exchange requires data transfer from input set of nodes to output set of nodes and possible connections between edges are established by applying various permutations in available links. There are various networks where multiple paths from source to destination are possible. For finding out what the possible routes in such networks are the study of the permutation concept is a must.

Let us look at the basic concepts of permutation with respect to interconnection network. Let us say the network has set of n input nodes and n output nodes.
Permutation P for a network of 5 nodes (i.e., n = 5) is written as follows:

$$P = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 1 & 3 & 2 \end{bmatrix}$$

It means node connections are 1↔5, 2↔4, 3↔1, 4↔3, 5↔2.

The connections are shown in the *Figure 13*.

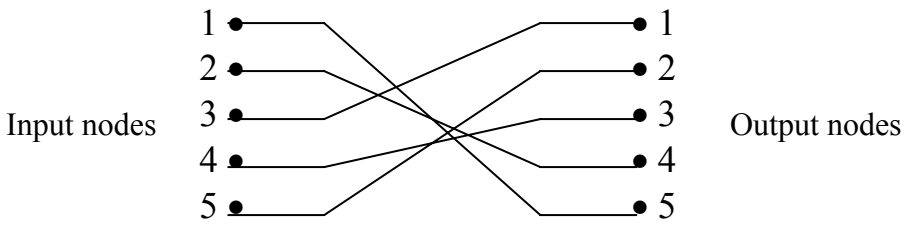


Figure 13: Node-Connections

The other permutation of the same set of nodes may be

$$P = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 1 & 4 \end{bmatrix}$$

Which means connections are: 1↔2, 2↔3, 3↔5, 4↔1, and 5↔4
Similarly, other permutations are also possible. The Set of all permutations of a 3-node network will be

$$P= \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

Connection, $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$ indicates connection from node 1 to node1, node 2 to node 2, and node 3 to node 3, hence it has no meaning, so it is dropped.

In these examples, only one set of links exist between input and output nodes and means it is a single stage network. It may be possible that there exist multiple links between input and output (i.e. multistage network). Permutation of all these in a multistage network are called permutation group and these are represented by a cycle e.g. permutation

P= (1,2,3) (4,5) means the network has two groups of input and output nodes, one group consists of nodes 1,2,3 and another group consists of nodes 4,5 and connections are 1→2, 2→3, 3→1, and 4→5. Here group (1,2,3) has period 3 and (4,5) has period 2, collectively these groups has periodicity 3×2=6.

Interconnection from all the possible input nodes to all the output nodes forms the permutation group.

The permutations can be combined. This is called composition operation. In composition operation two or more permutations are applied in sequence, e.g. if P_1 and P_2 are two permutations defined as follows:

$$P_1 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \quad P_2 \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

The composition of P_1 and P_2 will be

$$P_1 . P_2 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

$$P_1 . P_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

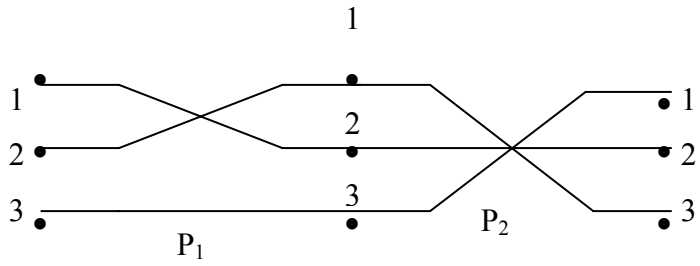


Figure 14 (a)

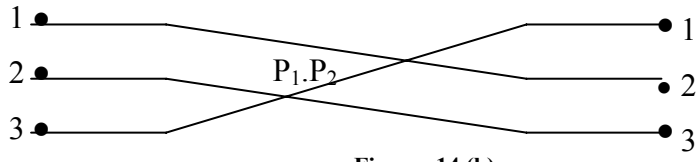


Figure: 14 (b)

Similarly, if $P_3 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{bmatrix}$ and $P_4 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 2 & 4 \end{bmatrix}$

then $P_3 . P_4 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 5 & 4 \end{bmatrix}$

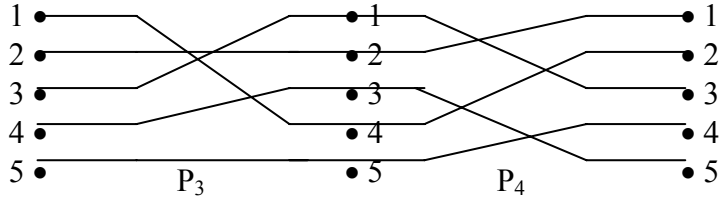


Figure: 15

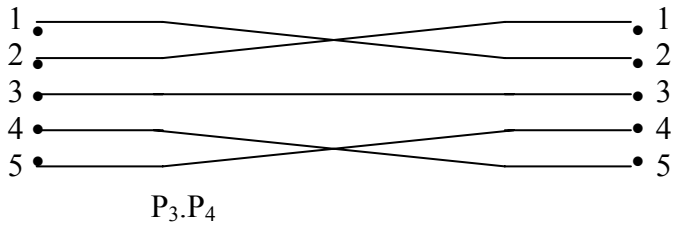


Figure: 16

Composition of those permutations P_1 P_2 are represented in Figures 14 (a) and 14 (b)

There are few permutations of special significance in interconnection network. These permutations are provided by hardware. Now, let us discuss these permutations in detail.

1) **Perfect Shuffle Permutation:** This was suggested by Harold Stone (1971). Consider N objects each represented by n bit number say $X_{n-1,} X_{n-2,} X_0$ (N is chosen such that N = 2n.) The perfect shuffle of these N objects is expressed as

$$X_{n-1,} X_{n-2,} X_0 = X_{n-2,} X_0 \quad X_{n-1.}$$

That, means perfect shuffle is obtained by rotating the address by 1 bit left. e.g. shuffle of 8 objects is shown as

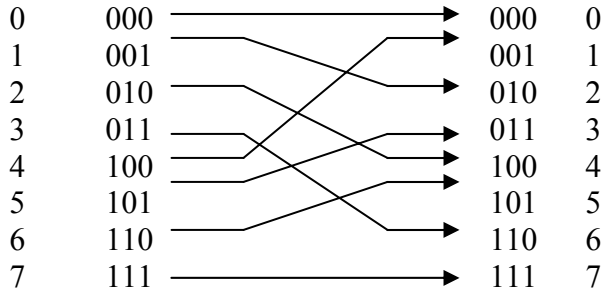


Figure 17: Shuffle of 8 objects

2) **Butterfly permutation:** This permutation is obtained by interchanging the most significant bit in address with least significant bit.

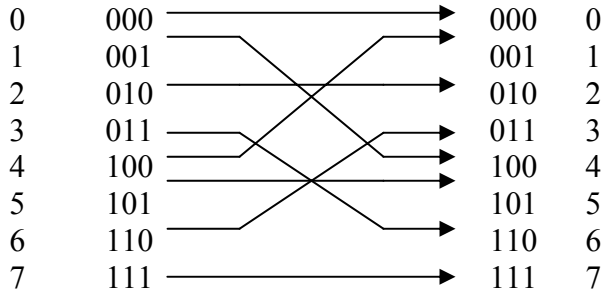


Figure 18: Butterfly permutation

e.g. $X_{n-1,}$ $X_{n-2,}$ and $X_1.X_0 = X_0$ X_{n-2} ……. X_1 X_{n-1}

001↔100, 010↔010
011↔110,

An interconnection network based on this permutation is the butterfly network. A butterfly network is a blocking network and it does not allow an arbitrary connection of

N inputs to N outputs without conflict. The butterfly network is modified in Benz network. The Benz network is a non-blocking network and it is generated by joining two butterfly networks back to back, in such a manner that data flows forward through one and in reverse through the other.

3) **Clos network:** This network was developed by Clos (1953). It is a non-blocking network and provides full connectivity like crossbar network but it requires significantly less number of switches. The organization of Clos network is shown in *Figure 19*:

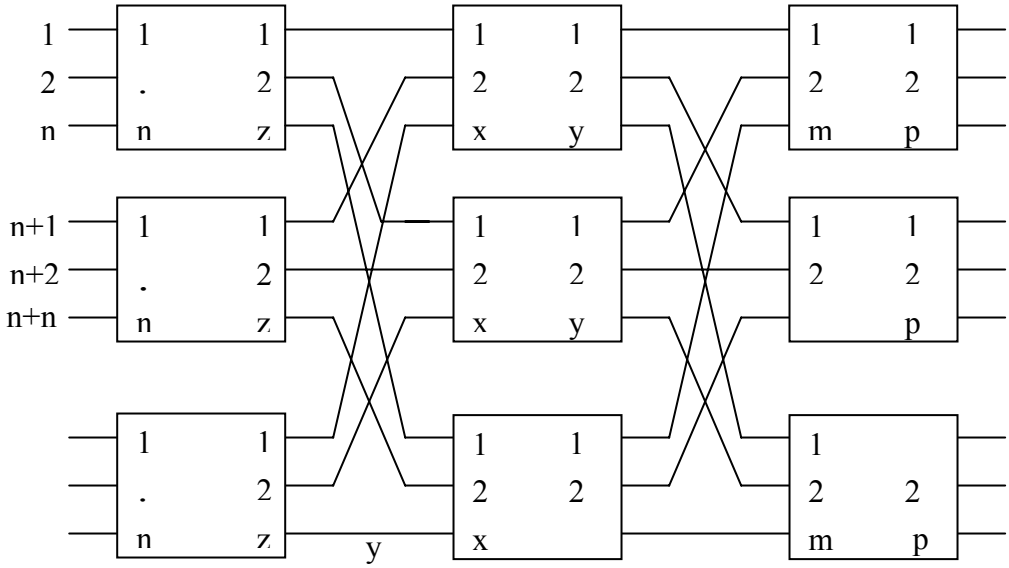


Figure 19: Organisation of Clos network

Consider an I input and O output network
Number N is chosen such that (I= n.x) and (O=p.y).

In Clos network input stage will consist of X switches each having n input lines and z output lines. The last stage will consist of Y switches each having m input lines and p output lines and the middle stage will consist of z crossbar switches, each of size X × Y. To utilize all inputs the value of Z is kept greater than or equal to n and p.

The connection between various stages is made as follows: all outputs of 1st crossbar switch of first stage are joined with 1st input of all switches of middle stage. (i.e., 1st output of first page with 1st middle stage, 2nd output of first stage with 1st input of second switch of middle stage and so on…)

The outputs of second switch of first stage. Stage are joined with 2nd input of various switches of second stage (i.e., 1st output of second switch of 1st stage is joined with 2 input of 1st switch of middle stage and 2nd output of 2nd switch of 1st stage is joined with 2nd input of 2nd switch of middle stage and so on…

Similar connections are made between middle stage and output stage (i.e. outputs of 1st switch of middle stage are connected with 1st input of various switches of third stage.

Permutation matrix of P in the above example the matrix entries will be n

Bens Network: It is a non-blocking network. It is a special type of Clos network where first and last stage consists of 2×2 switches (for n input and m output network it will have n/2 switches of 2 ×2 order and the last stage will have m/2 switch of 2×2 order the middle stage will have two n/2 X m/2 switches. Numbers n and m are assumed to be the power of 2.

Thus, for 16×16 3-stage Bens network first stage and third stage will consist of 8 (2×2) switches and middle stage will consist of 2 switches of size (8×8). The connection of crossbar will be as follows:

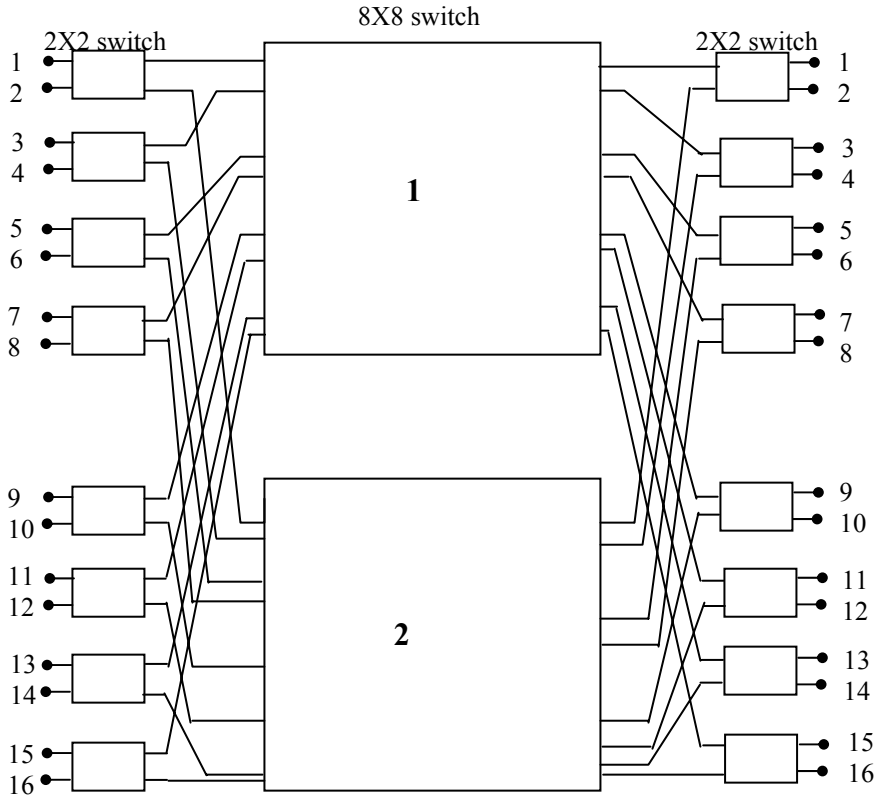


Figure 20: 16X16 3-stage benz network

The switches of various stages provide complete connectivity. Thus by properly configuring the switch any input can be passed to any output.

Consider a Clos network with 3 stages and 3×3 switches in each stage.

Permutation $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 7 & 6 & 2 & 5 & 3 & 9 & 4 & 1 \end{bmatrix}$

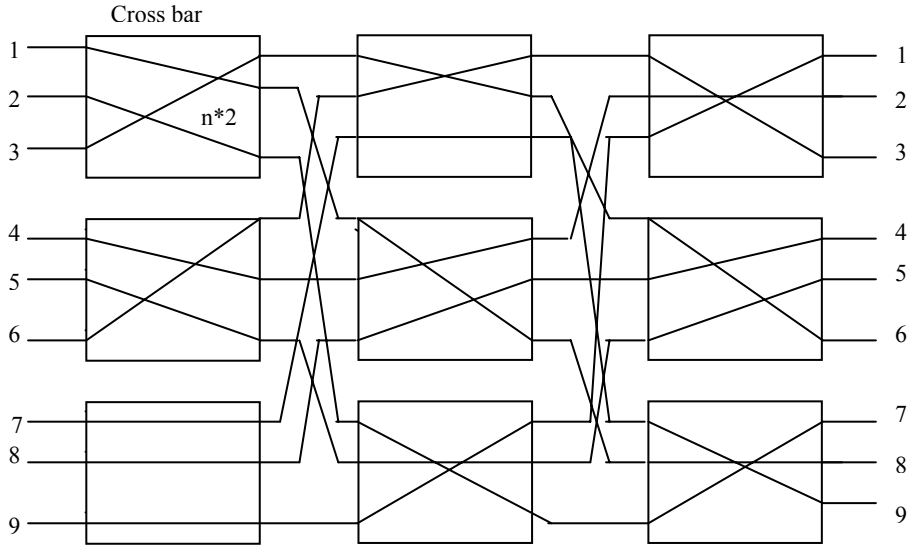


Figure 21: Clos Network

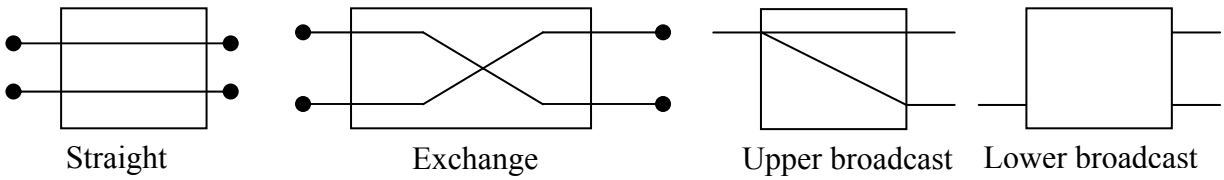
The implementation of this above permutation is shown in *Figure 20*.
This permutation can be represented by the following matrix also,

Input \ Output	1	2	3	4	5	6	7	8	9
1								X	
2							X		
3						X			
4		X							
5					X				
6			X						X
7									
8				X					
9	X								

Figure 22: Permutation representation through Matrix

The upper input of all first stage switches will be connected with respective inputs of 1st middle stage switch, and lower input of all first stage switches will be connected with respective inputs of 2nd switch. Similarly, all outputs of 1st switch of middle stage will be connected as upper input of switches at third stage.
In Benz network to reduce the complexity the middle stage switches can recursively be broken into N/4 × N/4 (then N/8 × M/8), till switch size becomes 2×2.

The connection in a 2×2 switch will either be straight, exchange, lower broadcast or upper broadcast as shown in the Figure.



Switch 2x2

The 8×8 Benz network with all switches replaced by a 2×2 is shown in *Figure 23(a)*:

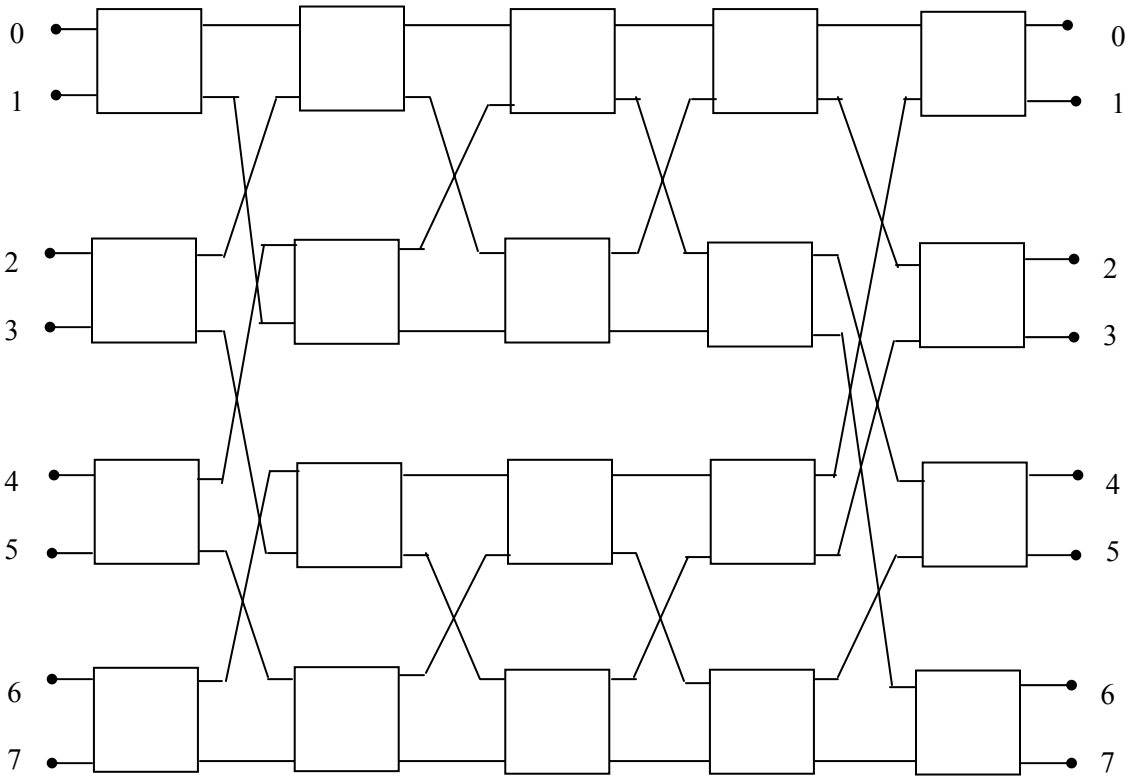


Figure 23 (a): Benz Network

The Bens network connection for permutation

$$P \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 7 & 3 & 4 & 5 & 6 & 2 & 0 \end{bmatrix}$$

will be as follows:-

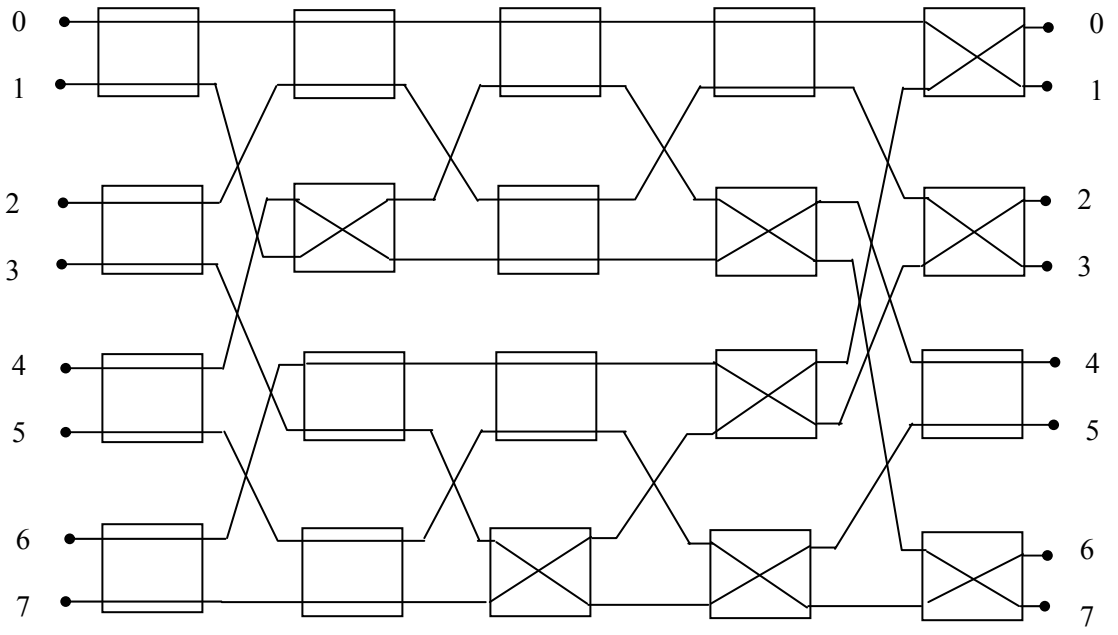


Figure 23 (b): 8X8 BENZ NETWORK OF 4 STAGE

The permutation for P = $\begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 0 & 7 & 6 & 2 & 5 & 3 & 9 & 1 \end{bmatrix}$ will be as follows

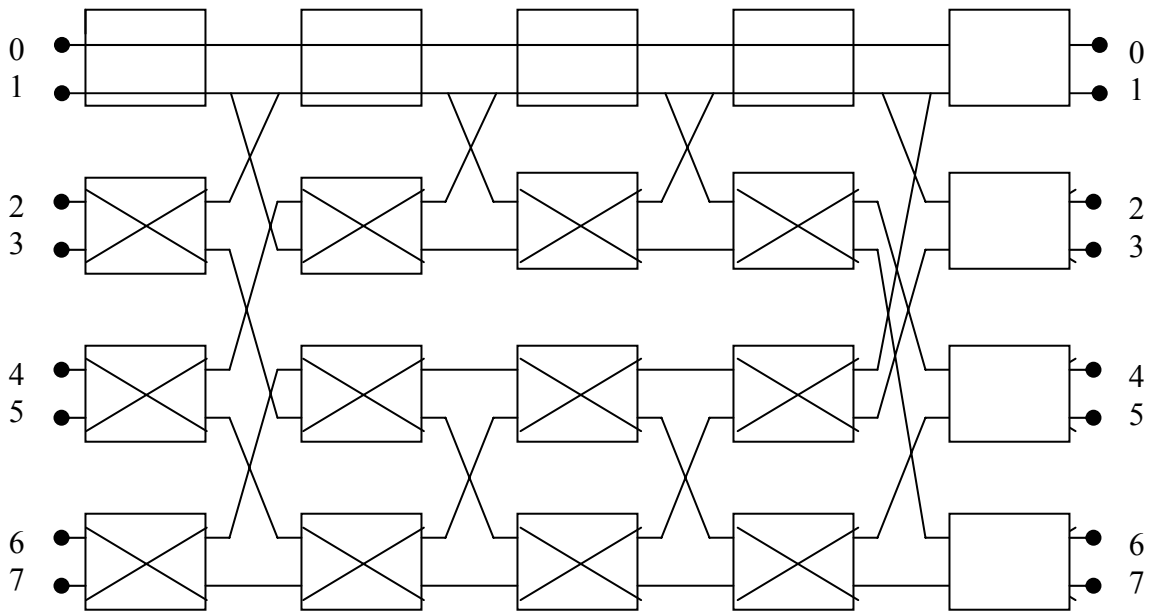


Figure 24: Line number for n

Hardware complexity of Benz Network: - Benz network uses lesser switches and it provides good connectivity. To find hardware complexity of Benz network let us assume that

$$N = 2^n \Rightarrow n = \log_2 N$$

Number of stages in N input Benz network $= 2n-1 = 2 \log_2 N - 1$

Number of 2X2 switches in each stage = N/2.

Total number of cells in network = $(N/2)(2 \log_2 N-1) = N \log_2 N - N/2$

Number of switches in different networks for various inputs is shown in the following table:

Input	No. of cross bars	Switches Clos Net	Benz Net
2	4	9	4
8	64	69	80
64	4096	1536	1408
256	65536	12888	7680

Thus we can analyze from this table that for larger inputs the Benz Network is the best as it has the least number of switches.

Shuffle Exchange Network:- These networks are based on the shuffle and exchange operations discussed earlier.

3.6 PERFORMANCE METRICS

The performance of interconnection networks is measured on the following parameters.

1) **Bandwidth:** It is a measure of maximum transfer rate between two nodes. It is measured in Megabytes per second or Gigabytes per second.

2) **Functionality:** It indicates how interconnection networks supports data routing, interrupt handling, synchronization, request/message combining and coherence.

3) **Latency:** In interconnection networks various nodes may be at different distances depending upon the topology. The network latency refers to the worst-case time delay for a unit message when transferred through the network between farthest nodes.

4) **Scalability:** It refers to the ability of interconnection networks for modular expansion with a scalable performance with increasing machine resources.

5) **Hardware Complexity:** It refers to the cost of hardware logic like wires, connectors, switches, arbiter etc. that are required for implementation of interconnection network.

3.7 SUMMARY

This unit deals with various concepts about interconnection network. The design issues of interconnection network, types of interconnection network, permutation network and performance metrics of the interconnection networks are discussed.

3.8 SOLUTIONS/ANSWERS

Check Your Progress 1

1) i) **Node degrees:** Number of edges connected with a node is called node degree. If the edge carries data from the node, it is called out degree and if this carries data into the node it is called in degree.

   ii) **Dynamic connection network:** In dynamic network the interconnection pattern between inputs and outputs can be changed. The interconnection pattern can be reconfigured according to the program demands. Here, instead of fixed connections, the switches or arbiters are used. Examples of such networks are buses, crossbar switches, and multistage networks. The dynamic networks are normally used in shared memory(SM) multiprocessors.

   iii) **Network diameter:** It is the minimum distance between the farthest nodes in a network. The distance is measured in terms of number of distinct hops between any two nodes.

2) Bisection bandwidth of a network is an indicator of robustness of a network in the sense that if the bisection bandwidth is large then there may be more alternative routes between a pair of nodes, any one of the other alternative routes may be chosen. However, the degree of difficulty of dividing a network into smaller networks, is inversely proportional to the bisection bandwidth.