Chapter 2
Survey of Literature

2.1 Direct Networks

Bus based systems are not ideal when scalability is of prime concern. On the other hand, interconnection networks offer better throughput for interconnection of computers in a multi-processor environment. Jose Duato et al. (31) classified interconnection networks as: shared-medium networks, direct networks, indirect networks and hybrid networks. In shared medium networks the transmission medium is shared by all communicating devices. Alternatively, point to point communication mode involves communication from each device to a subset of other communicating devices. Communications to all non-neighboring stations in point-to-point mode pass through several intermediate stations. These networks are called direct networks. In indirect networks interconnection is through one or more switches. The switches are connected between them using point to point links. Hybrid approach follows combination of all the three modes described above. The thesis work concentrates on direct networks.

In a direct Network architecture each node has a direct point-to-point connection to some other node called neighboring node. Popular direct networks are mesh, k-ary-cube, and tori, and ring networks. Each node is a separate processor by itself, supported by local memory and other supporting devices. Most common device at each node is a router, that handles all communications into and out of the node. A dedicated router is often employed to speed up the routing process. Hence Direct Networks are also called router based Networks. A direct network is characterized by
three factors: topology, routing and switching. The fully connected topology would connect every node to every other node. No message would even have to pass through an intermediate node before reaching destination. This fully connected topology requires a router to have \( n \) links at each node for a network of \( n \) nodes. The cost of such a router is prohibitively high. Also the number of physical connections is limited by hardware constraints such as available pins and available wiring area. As a consequence, many other topologies have been proposed. In these topologies messages may have to traverse some intermediate nodes before reaching the destination and such topologies are called direct network topologies. Most commonly used direct network topologies are k-ary-\( n \) cube, mesh, tori and hypercube. Examples of parallel computers with interconnection using direct networks are

- Cray TSE: Bidirectional 3-D torus, 14-bit data links.
- Cray TSD: Bidirectional 3-D torus up to 1,024 nodes, 24-bit data links.
- Intel Cavalline: Bidirectional 3-D topology. 16-bit links.
- SGI SPIDER: Router with 20-bit links supporting regular and irregular topologies.
- MIT M-Machine: 3-D mesh.
- MIT reliable router: 2-D mesh, 23-bit links.
- Chaos router: 2-D torus topology, 8-bit links.
- Intel ipsc-8 Hypercube: Binary hypercube bit serial channel.

The time required to move data between nodes is critical to system performance. In (64) and (67), the network latencies due to block switching is presented. A performance measure commonly used to evaluate direct network systems is communication latency defined as the sum of start-up latency, network latency and blocking time (61). Start-up latency is the time required for a message framing/unframing, memory/buffer copying, validation and so on at both source and destination nodes. The start-up latency mainly depends on the design of system software within the nodes and the interface between node and routers. The network latency is the elapsed time after the head of a message has entered the network at the source until the tail emerges from the network at the destination. If the network is contention free, given a source and destination node, the start-up and network latencies are
static values. These delays are mainly due to conflict over the use of shared resources, for example, a busy channel or a full buffer. Blocking time reflects the dynamic behavior of the network due to the passing of multiple messages whose traffic is unevenly distributed.

The unit of information exchange between the nodes is a message. For efficient and fair use of network resources, a message is often divided into fixed size packets prior to transmission. A packet carries the destination address and sequence number along with the data. The routing algorithms determine the path taken by a packet to reach its destination and indicates at each node indicates the next channel to be used. That channel may be selected among a set of possible choices. If all the potential channels are busy the packet is said to be blocked and the packet cannot advance further. Hence efficient routing is critical to the network performance. Routing algorithms that use minimal path(or shortest path) between source and destination node pairs for message delivery are called minimal routing algorithms. A routing algorithm is non-adaptive, if it uses fixed paths between source destination node pairs. A routing algorithm is called adaptive if it provides multiple paths to use between source destination node pairs. The minimal adaptive routing algorithms do not impose any restriction on the choice of shortest path to be used in routing the messages.

When a packet header reaches an intermediate node, a switching mechanism determines how and when the router switch is set i.e., the input channel is connected to the output channel selected by the routing algorithm. The switching mechanism determines allocation of network resources. In circuit switching all channels required by a message are reserved before starting transmission. In packet switching a packet is transmitted through a reserved channel but the next channel is not reserved until the packet releases the current channel. Obviously some buffer space is required to store the packet until the next channel is reserved. That buffer should be allocated before starting transmission. So buffer allocation is closely related to switching mechanism. The need to buffer a complete packet within the router makes it difficult to construct small, compact and fast routers. In wormhole switching, message packets are pipelined through the network. A message is broken into flits and is pipelined at
the flit level. Thus at any instant of time a blocked packet occupies buffer in several routers. By this, the buffer requirements within a given router are substantially reduced. In direct network topologies, this method of wormhole switching has become quite popular in which the network latency is also independent of path length. In the next section, ring networks in distributed systems are discussed.

2.2 Ring networks in Distributed Systems – Loop Topologies

When reliability is the prime concern, distributed systems offer attractive options. In these systems processing, control, and data bases are distributed. In a distributed system, topology directly controls the network delays. In (29), (93) topologies and their relationship with delays have been dealt in detail. Inter node distances directly contribute to message delays and the complexity of routing algorithm depends largely on topologies. In (35), (90) authors have proposed several topologies. There are several commercial multi processor systems, designed with ring topology as interconnection network (43), (49). For example ILLIAC IV uses 64 processors interconnected to achieve massively parallel processing. CDC Cyberplus also uses these interconnection networks in their two level slotted token ring for massively parallel processing. In 1992 Kendall Research center used these technologies in their 2 level ring for KSR1 model, which supports 8 to 1088 processors and supports every known form of parallelism through KSR Mach based operating system. Interconnection networks like torus, mesh etc. have been well studied for performance analysis (12). B Care (15) provides mathematical foundations on Graphs and Networks.

Ring networks have larger diameter and hence suffer longer network communication delays. Further ring networks are unreliable as a failure of single node would disrupt the entire ring. Loop networks with multiple hops offer smaller diameters, path lengths and better fault-tolerance. These networks, also known as multi connected distributed loop (MCDL) networks in the literature, have extensive uses in LAN, parallel processing, and multi processing environments. In this chapter the study the related work in the areas of shortest path, diameter, delay performance, fault tolerance, congestion, and communication latencies has been carried out.
In a ring network of N processors, processor \( i \) is connected to processor \( (i+1) \mod N \) is called a direct ring and the edges are directed. Directed rings suffer from larger communication latencies and larger diameter. By increasing the number of connections amongst the processors we can reduce the diameter. For example, \( i \) can be connected to processor \( (i-1) \mod N \) too. The ring network then is called doubly connected loop network. Maximum diameter of a double loop network is \( N/2 \). Therefore systematic addition of links to Ring can enhance performance. One such a way is to connect every node in doubly connected loop to \( (i+h) \mod N \) and \( (i-h) \mod N \), where \( h \) is called hop length or skip distance. These networks are called \textit{Multi connected Distributed Loop (MCDL) Networks}. MCDL Network for \( n=20 \) and hop size = 6 has been shown in Figure 2.1. These type of networks are widely used in design implementation of distributed systems and also in local area networks (63), (45), (72), and (79).

Fig 2.1 : MCDL Network with \( N = 20 \) and hop size = 6
2.3 Terminology of MCDL Network

Several studies have been carried out on mesh, tori and hyper cube and k-ary n - cube direct network topologies for multi computer and multi processor organizations (26), (30), and (51). MCDL is a fixed degree (four) network with hop and loop connections (+h, -h, +1, and -1) at every node. Packets travel to destination via one or more intermediate node depending on the shortest path algorithm. A node has four bidirectional channels. The routing is Source routing in which source decides the entire path before sending the packet. Each packet carries this information and selects next node depending on hop or loop links. The routing can be distributed also in that a packet can find shortest path when it reaches intermediate node. The routing algorithms decide which output channel is selected for a packet arriving on a input channel. This is referred to as output selection policy. Similarly a particular output channel can be requested by several of the packets arriving on many input channels. This is termed as input selection policy. An output channel that is requested by several packets is allocated by first come first service basis.

A ring topology of N processors, labeled 0 to (N-1) is formed by connecting processor ‘i’ to processor (i+1) mod N and is called a directed ring as the channels are unidirectional. Instead, if the channels are bi-directional, the network is called an undirected ring network and every node ‘i’ is connected to nodes, (i+1) mod n and (i-1) mod n. In a ring network, if at least one of the links breaks down the transmission of packets becomes difficult, since there is no alternate path, which can be used to bypass the fault. Hence, this network does not offer much reliability. Moreover the length of the paths that the packets have to travel is quite large. For example, in a directed ring network having n processors, the maximum length of the path for a packet is (n-1). In an undirected ring network this reduces to floor(N/2) (where floor(x) implies the largest integer less than x), which is still quite large and causes unacceptably large delays for large network traffics.
One way to improve the reliability of the network is to add a second ring. But even then, the delay is not reduced. Systematic addition of links to an undirected ring network can increase the reliability and speed considerably. One way to do this is to add two links \( +h, -h \) to every node (here \( h \) is the hop size) such that a node 'i' is connected to \( (i+h) \mod n \) and \( (i-h) \mod n \). These networks are called Multi connected Double-loop networks (MCDL). Therefore every node is adjacent to four other nodes \((i+1), (i-1), (i+h), (i-h) \mod N\) and four different links leading to it. A node will not be isolated unless all the four links are faulty and this brings greater reliability to the system.

In a MCDL network, there can be four types of links/channels termed \([+h], [+1], [-h], [-1]\). A traversal along the \([+h]\) link or the \([-h]\) link is called a long hop and a traversal along \([+1], [-1]\) is called a short hop or loop traversal. In routing a packet from one processor to another, a number of links need to be traversed. All these links together form the path and the number of \([+h], [-h], [+1], [-1]\) links in a path are denoted by \(w, x, y, z\) respectively. Depending on the link combinations that make up the path the paths are termed, \((+h, +1), (+h, -1), (-h, -1), (-h, +1)\) paths. A path in \((+h, +1)\) combination using \(w\) \([+h]\) links and \(y\) \([+1]\) links is denoted as \(w[xh] + y[+1]\) and the length of the path is equal to \(w + y\). The paths and path lengths in the remaining three path patterns can be defined similarly. It is important to note that the order in which the links are traversed is immaterial. MCDL network is denoted with \(N\) nodes is denoted as \(G(n; h, 1)\). A sub-graph of the above graph using \(+h, +1\) links only is denoted as \(G(n; +h, +1)\). Similarly, the graphs \(G(n; +h, -1)\), \(G(n; -h, -1)\) and \(G(-n; -h, +1)\) are used to denote sub-graphs formed by \((+h, -1), (-h, -1)\) and \((-h, +1)\) link patterns respectively.

It is interesting to note that in a MCDL network, finding a path from node 'i' to 'm' is similar to finding a path from 0 to \((m-i) \mod N\). The proof has been given in N Chalamalai et al. (23). Moreover the four paths in the MCDL network are isomorphic. If a path exists between a pair of nodes in one link pattern, an equivalent path can be found in the other link patterns. All the three paths \((+h, -1), (-h, -1)\) and \((-h, +1)\)
+1) can be converted to (+h, +1) link combination. N Chalamaiah et al. enumerated following important Lemmas of MCDL Network, viz

Lemma 1: If \( w[h] + x[h] + y[+1] + z[-1] \) is a path from \( I \) to \( m \), then it is a path from \( 0 \) to \( p \) also where \( p = m - (mod n) \).

Lemma 2: If \( w[h] + z[-1] \) is a path from \( 0 \) to \( p \) in (+h, -1) link combination, then \( w[n-h] + z[+1] \) is a path from \( 0 \) to \( (n-p) \) in (+h, -1) link combination.

Lemma 3: If \( x[-h] + y[+1] \) is a path from \( 0 \) to \( p \) in (-h, +1) link combination, then \( x[+(n-h)] + y[+1] \) is a path from \( 0 \) to \( (p) \) in (-h, +1) link combination.

Lemma 4: If \( x[-h] + z[-1] \) is a path from \( 0 \) to \( p \) in (-h, -1) link combination, then \( x[+h] + z[+1] \) is also a path from \( 0 \) to \( (n-p) \) in (+h, +1) link combination. Table 2.1 shows the transformations required to convert (+h, -1), (-h, +1), and (-h, -1) combinations to (+h, +1) link.

Table 2.1: Transformations for converting (+h, -1), (-h, +1), and (-h, -1) to (+h, +1) links

<table>
<thead>
<tr>
<th>Link</th>
<th>Source-destination</th>
<th>Equivalent (+h, +1)</th>
<th>Equivalent src-dest</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+h, -1)</td>
<td>0 to p</td>
<td>(+n-h, +1)</td>
<td>0 to (n-p)</td>
</tr>
<tr>
<td>(-h, +1)</td>
<td>0 to p</td>
<td>(+n-h, +1)</td>
<td>0 to p</td>
</tr>
<tr>
<td>(-h, -1)</td>
<td>0 to p</td>
<td>(+h, -1)</td>
<td>0 to (n-p)</td>
</tr>
</tbody>
</table>

In (16), (17), (18), (19), (20), and (59) authors N Chalamaiah et al. have dealt in detail about MAC access protocol issues in MCDL Networks. In (87) Vasco et. discuss analysis and simulation of Multiple-Ring Token ring architecture. MT Liu (53) proposed a fault tolerant system using [+1] and [-1] links and termed it as Distributed Double Loop Computer network (D DLCN). Similarly, Raghavendra et
al (72) proposed the forward loop backward hop (FLBH) network using [+1] and [-h] links and forward loop parity hop (FLPH) network using [+1] and [+h] or [-h]. The hop connections in FLPH are formed by connecting all odd nodes to \((i+h) \mod N\) and all even nodes to \((i-h) \mod N\). The FLPH with additional [-i] links, odd \(h\) and even \(N\) is called a chordal ring by Arden and Le (1).

\(G(N; +1, +h)\) is used to denote a loop network of \(n\) nodes with link connections [+1] and [+h] only. Similarly \(G(N; 1, h)\) denotes the networks having all the four link types. Their diameters are represented by \(\text{dia} (N; +1, +h)\) and \(\text{dia}(N, h)\) respectively. Thus \(G(n; +1, +h)\) is the directed version and \(G(n; 1, h)\) is the undirected version. Networks \(G(N; +1, -h)\), \(G(N; -1, +h)\) and \(G(N; -1, -h)\) with the other three link combinations are also possible. Hence this work for the most part considers \(G(N; +1, +h)\) type of subnet works only for performance analysis.

The shortest path, in the \(G(N; +1, +h)\) from node \(l\) to node \(m\) is represented by \(sp_{++}(l, m)\). Similar notation is used in networks with the other three link combinations. The overall shortest path from \(l\) to \(m\) in \(G(N; 1, h)\) is represented by \(sp(l, m)\). Hence

\[
sp(l, m) = \text{Min} \{ sp_{++}(l, m), sp_{-+}(l, m), sp_{+-}(l, m), sp_{-+}(l, m) \}
\] (2.1)

Average path length of the network, \(ava' (n; +1, +h)\) and \(ava(N, h)\) for the directed and undirected versions respectively are defined as (71):

\[
ava((N; +1, +h)) = \frac{1}{N} \sum_{l=0}^{N-1} \left( \frac{1}{N-1} \sum_{m=0}^{N-1} sp_{++}(l, m) \right) \quad l \neq m
\] (2.2)

\[
ava((N; 1, h)) = \frac{1}{N} \sum_{l=0}^{N-1} \left( \frac{1}{N-1} \sum_{m=0}^{N-1} sp(l, m) \right) \quad l \neq m
\] (2.3)

\(\forall (l, m) \in \{0, 1, \ldots, N-1\}\)
As the network is symmetric we can define

\[
N - 1 \quad \text{ava}'(N;+1,+h) = \left[ \frac{1}{N-1} \sum_{p=0} \text{sp}_{++,0,p} \right]
\] (2.4)

\[
N - 1 \quad \text{ava}'(N;h) = \left[ \frac{1}{N-1} \sum_{p=0} \text{sp}_{0,p} \right]
\] (2.5)

The diameters of the directed and undirected network can now be defined as:

\[
\text{dia}'(N;+1,+h) = \text{Max} [\text{sp}_{++,l,m}]
\] (2.6)

\[
\text{dia}(N,h) = \text{Max} [\text{sp}(l,m)]
\] (2.7)

\[\forall l,m \in \{0,1,...,(N-1)\} \text{and } l \neq m. \text{ Again in view of symmetry of the topology,}
\]

\[
\text{dia}'(N;+1,+h) = \text{Max} [\text{sp}_{++,0,p}]
\] (2.8)

\[
\text{dia}(N,h) = \text{Max} [\text{sp}(0,p)]
\] (2.9)

\[\forall p \in \{1,2,...,(n-1)\}.
\]

The diameter of a loop network is said to be optimal diameter if it is the smallest diameter among networks with all the possible hop sizes. The optimal diameters are represented as \(\text{dia}'(N)\) and \(\text{dia}(N)\) for directed and undirected loops respectively. Hence they are,

\[
\text{dia}'(N;+1,+h) = \text{Min}_{h} [\text{dia}'(n;+1,+h)]
\] (2.10)

\[
\text{dia}(N,h) = \text{Min}(h [\text{dia}(N,h)])
\] (2.11)

Loops which achieve this diameter are called optimal diameter loops. Similarly optimal average distance loops are defined as loops which achieve the minimum average distance defined below:

\[
\text{ava}(N) = \text{Min}(h [\text{ava}(N;+1,+h)])
\] (2.12)
\[ \text{ave}(N) = \text{Min}(h[\text{ave}(N,h)]) \quad (2.13) \]

The average path length and diameter for the other three directed versions of MCDL is similarly defined. It is observed that loops which are optimal in diameter are not necessarily optimal in their average distance and will be corroborated in section 4.4.

In (50) computation of shortest path between a pair of node in \( O(5) \) time where \( d \) is the diameter of MCDL Network has been reported. This algorithm presupposes the knowledge of diameter apriori. In (29), (90), (8), the optimal forms of \( N \) for which the diameter is minimum overall possible hop sizes were reported for only specific values of \( N \).

### 2.4 Wormhole Routing

High performance interconnection networks can be achieved by using wormhole routing (24), cut thru switching (91), adaptive routing (3). In these systems communications occur by sending packets independently through network. Both these methods suffer from network saturation. The main advantage of wormhole switching are low memory requirements in routers, pipe lined data movement, and low communication latency. Let \( L \) and \( F \) be the length of packet and flit(in bits). A packet is divided into a sequence of fixed sized units of data called ‘flits’. Let \( W \) be channel band width in bits/sec and \( L \) and \( F \) be length of packet and flit respectively. If the packet has to travel a distance of \( D \) links, the latency is given by \( (L/W + (F/W)^*D) \). If the size of the flit is such that \( F << L \) then the latency is independent of \( D \), the distance between source and destination.

Alternatives to wormhole switching are virtual cut-through (91) and store and forward (39) and they require more storage space at each node. For example, in a virtual-cut-through switching technique, a packet can be forwarded as soon as its destination is known, even before the entire packet has arrived. In the store-and-forward messaging technique each node has to receive and store
the complete packet before it is forwarded to the next node. This requires large buffers at each node. In (37) S Yalamanchili et al. discuss a pipe lined circuit switching, a fault tolerant variant of wormhole routing. The advantage of using the wormhole technique is that it reduces the latency (the delay for forwarding a packet) and storage requirements at each node. However, each node has to keep track of the messages currently flowing through it, which adds greater complexity to the technique. Moreover, this method does not guarantee freedom from deadlocks. A deadlock can occur if more than one packet competes for the same channel, and some parts of packet, flits, get blocked in some channel. If the two messages mutually block and none of them can forward to its destination, a 'routing deadlock' is said to have occurred. Congestion in the networks is another predominant problem in wormhole routing. It generally results when any node is generating traffic much more than what the network can handle. Moreover, the network should be made tolerant to faults. Achieving fault tolerance is much more difficult owing to the constraints inherent in the wormhole routing technique. One way out from these problems is to simulate multiple virtual channels on each physical channel and to enforce a pre-defined order in the allocation of the virtual channel to the messages. Generally, the allocation of virtual channels is done on the basis of the direction in which the packet should be forwarded.

The time required to move the data in between nodes is a critical parameter for assessing network performance. A performance metric used frequently is communication latency which is defined as the sum of start-up latency, network latency and blocking time (92). Start up time is time required to get the message ready to leave i.e. framing/unframing, memory and buffer copying, and validation at both source and destination nodes. Start up latency depends on the software and interface between the node and the router. Network latency is the elapsed time between header message at source to tail message reaching the destination. If a network is congestion and fault free, then these two latencies are static values. Blocking time refers to all possible delays encountered due to conflicts over the use of shared resources.
2.5 Shortest Path Studies

In any network configuration, the paths through which the packets are routed from the source to the destination are of utmost importance. Irrespective of how the systems are configured, if the packets were not routed through the shortest paths, the performance of the system would be less than expected. Therefore in most networks, the shortest path information is calculated and stored in some form or the other. A number of algorithms have been proposed in order to find the shortest paths for routing the packets in Multi Connected Double-Loop Networks.

A number of algorithms have been proposed in order to find the shortest paths for routing the packets in MCDL networks. Several researchers also considered directed double loop networks (11), (71), and (79) which have directions assigned to the four links of each node. In a directed double loop network, for each pair of links of same hop (i.e., a pair of links consisting of ±1-links only or consisting of ±h-links only), one of the two links is an inward arc and the other link is an outward arc, uniformly throughout the network. Raghavendra et al (71), (72), and (73) proposed a forward loop and backward hop (FLBH) network using +1- and -h- links and further discussed on reliable loop topologies and their performance studies. In (1) Arden and Lee proposed a chordal ring in which hop connections are formed by connecting node i to (i+h) mod N or (i-h) mod N, depending on whether i is even or odd. In (97) Chen Bao-Xing et al presented a method to compute the diameter of undirected double loop networks with certain restrictions, i.e. when \( q \leq r \) and \( a > (b+1)q+1 \), where \( q, r, a \) and \( b \) satisfy \( N = qr + r, 0 \leq r < h, \) and \( h-r = br + a, 0 \leq a < r, \) and the algorithm runs in time \( O(\log r) \), where \( 0 < r < h \) and \( h \) is hop length.

In (50) computation of shortest paths in MCDL topology has been reported in \( O(\delta) \) time, where \( \delta \) is the diameter. But this algorithm requires knowledge of diameter apriori. In (23) Chalamalaih et. al have developed algorithms for finding shortest paths using number of theoretic properties and linear congruencies. In Bapeswara Rao et al. (40) authors have developed algorithm for finding shortest path
by filling and array of the order $O(i^3)$. The main feature of these algorithms is that shortest path is computed in time independent of the diameter, unlike in (50).

### 2.6 Average Path Length Studies

A way of measuring the performance of interconnection networks is by the maximum distance (or diameter) overall pairs of nodes. A large diameter would contribute to large transmission delays in the network. Another related quantity is the average distance between all pairs of nodes in the network. A way to increase the connectivity and decrease the diameter (and average distance) in a ring network is by adding additional links to the loop network. These links must be added in a homogeneous way and must be as few as possible so that messages can be routed in a systematic way and the switching mechanism at every node can be implemented easily. In MCDL this is accomplished by hop connections. In an MCDL, for a given number of nodes $N$, one wants to find a smallest diameter loop as it reduces the overall transfer delay of packets. But most of the time the exact value is unknown. In the following, a survey for some bounds on this is presented.

Arden and Lee in (1) initially proposed chordal rings, a variation of MCDL. Several studies on MCDL topology have been carried out. For example, routing has been studied by Arden and Lee (1), Raghavendra and Silvestre (72), Mukhopadhyaya and Sinha (50). Fault tolerance has been studied by Masuyama and Ichimori (57), Peha and Tobagi (70), Hu and Hwang (44), Hwang and Wright (48). The diameter of these networks were studied by Boesch and Wang (11), Hsu and Xu (29), Zerovnik and Pisanski (89).

The problem of diameter optimality was studied by Bermond (6), (8), Tzvieli (90). Some of the recent studies in parallel processing applications for MCDL is reported in (45), (63) in which reliability and fault tolerance were considered. An exhaustive survey on the MCDL networks is found in (72), (6).
For a directed ring network, the diameter is as big as $(n-1)$. For a DDLCN of $G(n;1)$ diameter is \( \left\lfloor \frac{n}{2} \right\rfloor \) and average distance is \( \frac{n}{4} \). The daisy chain which is $G(n; +1, -2)$ has diameter \( \left\lfloor \frac{n}{3} \right\rfloor +1 \) and average distance of \( \frac{n}{6} \). Moreover Wong and Coppersmith (94) have shown that for $G(t^2; +1,+t)$ has diameter exactly $2t-2$ (roughly $\sqrt{n}$, for large $n$) and average distance of $t-1$ (again roughly $\sqrt{n}$, for large $n$). It was also shown that

\[
\left\lfloor \sqrt{3n} \right\rfloor -2 \leq d'(N) \leq \left\lfloor \sqrt{3n} \right\rfloor -1
\]

\[
\frac{5}{9} \left\lfloor \sqrt{3n} \right\rfloor -1 \leq a'(N) \leq \frac{5}{9} \left\lfloor \sqrt{3n} \right\rfloor -1
\]

where

\[
d'(N) = \text{Min}(h_1, h_2 [\text{dia}'(n;+h_1, +h_2)])
\]

\[
a'(N) = \text{Min}(h_1, h_2 [\text{ava}'(n;+h_1, +h_2)])
\]

Loop networks which achieve the lower bound on diameter i.e., $\text{lb}(N) = \left\lfloor \sqrt{3n} \right\rfloor -2$ as given above are called tight directed loops. Not all $G(n; +1, +h)$ can achieve this lower bound. Fiol et at. (94) have shown by an exhaustive search that for most values of $N$, there exist networks such that $\text{dia}'(N) = d'(N)$ and the smallest $N$ for which the equality fails is for $N = 450$. In this case, $\text{dia}'(450) = 36 = \text{dia}(450; +1, +59)$ and $d'(450) = 35 = \text{dia}(450; +2, +185)$. There are infinite number of such loops that are not as tight as given by Esque and Aguilo [EAF93]. They gave through computer search that for $1 \leq n \leq 30000$ there exist several values of $n$ for which $\text{dia}'(N)-\text{lb}(N) = 3$. But for $1 \leq n \leq 75000$ there exist only three values for $n$ that show $\text{dia}'(N)-\text{lb}(N) = 4$. They are $n = 53749$ ($\text{dia}'(53749) = \text{dia}'(53749; +1, +985) = 404$), $n = 64729$ ($\text{dia}'(64729) = \text{dia}'(64729; +1, +394) = 443$) and $n = 69283$ ($\text{dia}'(69283) = \text{dia}'(69283; +1, +1764) = 458$). On the other hand it was not known until recently whether $\text{lb}(N)$ could be achieved for infinite values of $n$. Erdos and Hsu (32), and Esque et al (96) reported a number of families of $n$ in this direction. Thus the problem of optimizing $G(n; +1, +h)$ is not completely solved.
It was also shown by Fiol et al. (94) that \( \text{diam}'(N) \) does not increase monotonically. For example \( \text{diam}'(20) = 7 = \text{diam}(20; +1, +4) \) while \( \text{diam}(21) = 6 = \text{diam}(21; +1, +9) \). Similarly, \( \text{ava}'(N) \) and \( \text{ava}'(N; +1, +h) \) also are not regular. As an example consider \( \text{ava}'(20) = \text{ava}'(21; +1, +4) = 3.5 \), but \( \text{ava}'(21) = \text{ava}'(21; +1, +9) = 3.429 \). It is also interesting to note that some loops that are optimal in diameter are not necessarily optimal for average distance. As an example \( \text{ava}'(59) = \text{ava}'(59; +1, +8) = 6:5768 \). But \( \text{diam}'(59; +1, +8) = 13 > \text{diam}'(59) = \text{diam}'(59; +1, +25) = 12 \).

Similar to the bound given for the directed version, the undirected version \( G(n; 1, h) \) of MCDL is also studied. A more general form of this would be \( G(n; h1, h2) \), where every node \( i \) is connected to \((ih1) \) and \((ih2) \mod n\). They are connected if \( \text{GCD}(n; h1, h2) = 1 \). DJ Du (29) showed that

\[
\text{d}(N) \geq \left\lfloor \frac{\sqrt{2n-1}-1}{2} \right\rfloor
\]

Where

\[
\text{d}(N) = \text{Min}_{h1,h2} \left\{ \text{diam}(n; h1, h2) \right\}
\]

is the lower bound for the undirected version. The graphs \( G(N, h1, h2) \) which achieve this \( d(N) \) are called tight undirected loops. Tavieli (90) has shown that \( d(N) \) can be achieved if \( h1 = d(N) \) and \( h2 = d(N) + 1 \). The problem of finding optimal diameter \( \text{diam}(N) \) for \( G(N; 1, h) \) is not completely solved. But it is known that \( \text{diam}(2t^2+2t+1, 2t+1) = \text{diam}(2t^2+2t+1) = t = d(2t^2+2t+1) \) from Du (29). Tavieli (90) conjectured that \( \text{diam}(N) \) is at most \( d(N) + 1 \) and verified it up to \( N = 8000000 \). Du and Hsu (29) also classified \( N \) which are tight i.e., that achieve \( d(N) \). Bermond and Tzvieli (6), (8), (90) determined families of \( N \) for which \( \text{diam}(N) = d(N) \). Tzvieli (90) also defined suboptimal loops, graphs which achieve \( \text{diam}(N) = d(N) + 1 \) and obtained complete lists of \( N \) that are also suboptimal. Using this census Hsu (29) gave that
A class of loop networks proposed by Arden and Lee (1) called chordal rings have the diameter and average distance of $O(\sqrt{n})$. They also gave a distributed algorithm for routing and determination of alternate route under node or link failures.

### 2.7 Fault-Tolerance Studies:

In a network containing $N$ processors, it is possible that some of the processors are faulty. In some cases though the processors may be functional one or more links of the processor may be faulty. It is generally assumed that the mean time to manually repair the faults is quite large, sometime extending to a few days. Yet, the remaining processors are connected and functional and should be allowed to exchange packets. To provide for this, we resort to fault tolerant routing algorithms, which can route packets in spite of the presence of faults.

The fault information in the network may be made global—each fault-free node knows about all the faults in the system—or local—each fault-free node knows the status of its neighbors only. The global knowledge model requires some form of routing tables, which are necessary for fault-free routing. Distributing fault information to each and every node of the network is very expensive. Moreover, dynamic update of these routing tables may be required to be performed at frequent intervals, which further increases the network traffic. Also routing algorithms that depend on global fault information should have alternate schemes to transmit fault status messages during the transition period—the interval between the occurrence of a fault and the time at which this fault is known globally. Therefore, algorithms that use local fault information are generally used.

Fault tolerant routing algorithms can be further classified into non-adaptive routing algorithms and fully-adaptive routing algorithms. In (81) L Schiewbert et al. describe a fully adaptive routing algorithm...
for meshes. In a non-adaptive algorithm, once the path of the message is fixed it cannot be changed. Fully adaptive routing algorithms can be further classified into strongly adaptive and weakly adaptive algorithms. In a strongly adaptive algorithm, a blocked message can wait, indefinitely, for any of the legal links at an intermediate node; a legal link is a link, which takes the message closer to the destination. In a weakly adaptive algorithm, a blocked message can wait indefinitely only on a strict subset of the set of the legal links. Most fully-adaptive algorithms are strongly adaptive algorithms and weakly adaptive algorithms are rarely used. Fault-tolerant algorithms generally make use of the concept of virtual channels to route packets in the presence of faults. Each physical channel is logically divided into channels, which are called virtual channels, and the messages are assigned to a virtual channel depending on the direction in which they have to be forwarded. Algorithms differ on the number of virtual channels that are used. Use of four virtual channels is common, though, algorithms which use lesser number of virtual channels- two or three-have been proposed for networks with less stringent fault requirements.

The routing of messages is called a static routing if a predetermined fixed path between a given node pair in a network is used for routing. In a directed ring network, messages travel in a fixed path and hence it is a static routing. Moreover failure of a link or node brings down the entire network. Therefore, a directed ring is not reliable unless all the links and nodes are reliable. In multi-connected distributed loop networks discussed in Section-2.2, several alternate paths exist between a pair of nodes. Hence, they can sustain node or link failures by routing via an alternate path. The routing of messages under such conditions is called adaptive routing. Such adaptive routing is possible in distributed loops in view of availability of multiple source-destination paths. In this work adaptive routing under faults (single and up to two nodes or links failures) has been studied.

Fault tolerant routings were investigated in MCDL interconnection networks by several researchers in respect of node and edge faults (22), (50), (57), (68) (82). In (45) Lendert M. Huisman et al. reported reliability and fault tolerance in MCDL networks. Some of the studies are based on the
reliability measurements under single and multiple faults F K Hwang et al. (48), Rennels (88). This section's approach is based on BP Sinha et al. (50), and Ting Yi Sung et al. (82). In (50) the existence of an alternative routing under node and edge is considered. In (83) some topological properties of multi connected loops are reported for single node and edge fault tolerance.

Several results exist for fault tolerant routing in hyper cubes (52),(80),(24) that depend on rich interconnection structure of hyper cube and hence inherently unsuitable for low dimension network like MCDL network. In (60) Reddy et al. use global knowledge of faults which is difficult to maintain in a massively parallel processor and routing tables. In (37) Yalamanchili et al. use a pipe lined circuit switching mechanism with back tracking for fault tolerant routing. In (38) Glass et al. present the negative first algorithm that tolerates up to n-1 faults in a n dimensional mesh with out any extra virtual channels. In (26) Dally et al. presents fault tolerant algorithms based on the concept of dimension reversal which occurs whenever a message takes a hop, in a dimension lower compared to that of the previous hop. A message can be routed adaptively if the number of dimensional reversals it has taken is less than the number of highest virtual channel class, if a message finds a free channel in other outgoing channel of the current host in a finite amount time. In (26) Dally et al. present a partially adaptive algorithm to handle block faults in Meshes by using three virtual channels for fault tolerant routing. The limitation of their method is that they can not handle faults on the boundaries of the meshes with out computational overheads(4). Bopanna et al. in their work (4) have presented fully adaptive and non adaptive algorithms and handles faults on the network boundaries using four virtual channels. Case study for handling simpler faults using only two virtual channels has been provided. In (33) Felparin et al. discuss design of wormhole routing using as few virtual channels as possible. Further multiple virtual channels could be multiplexed on single physical channel using additional flit buffers and multiplexers to improve performance.
2.8 Congestion Control studies

Network performance degrades rapidly beyond network saturation. This saturation occurs when multiple packets contend for same resource like same link between two nodes. These packets occupy buffer space and thus delay other packets. This process of saturation extends and makes several packets to wait, even though they are destined for different station, but are using only current congested link. This type of saturation can be overcome by a process called source throttling, that prevents source from injecting a packet when congestion is detected. This is difficult to achieve because nodes will have no information about global state of congestion and have to depend only on local state of information. Algorithms that adapt to the network and traffic conditions are called adaptive routing algorithms. When a message arriving at a switch finds its preferred output port busy, it doesn't wait. Rather, it is detected from its shortest path and immediately transmitted on another available link. Thus, it takes a detour, which increases its path length and transit delay. Adaptive routers are classified as fully or partially adaptive depending on whether they allow all possible paths between source and destination or only a subset of them. Adaptive algorithms can be broken down into two broad categories as minimal or non-minimal.

Minimal Adaptive Routers:

Since most networks provide multiple shortest paths between most source destination pairs, it is intuitive to design a router that allows flexibility in choosing among these paths. Routers, which allow some choice among minimal-length paths based on local or temporal conditions, are known as minimal adaptive routers.

Non-minimal Adaptive Routers:

While optimal and minimal adaptive routers require that a packet always be routed on a minimal length path, non-minimal adaptive routers relax this constraint. If all minimal paths for a packet are heavily congested, but longer paths are uncontested, it may be possible for a packet to travel out of its
way on a longer path but arrive more quickly than otherwise. This is the basic principle of non-minimal adaptive routers. Therefore, resolving congestion using minimal adaptive routing is not that preferable as it gives rise to the problem of deadlock. The main problem with minimal adaptive routing is that, in general, the number of paths available decreases as the distance to the destination decreases. Hence, use of non-minimal adaptive routing has been explored. With non-minimal adaptive routing, whenever a packet finds its path heavily congested, it looks for other paths though longer ones that are uncontested. Then, it travels out of its way on a longer path but arrive more quickly than otherwise. Moreover, if hardware faults result in a link becoming permanently unavailable, there will be some pair of nodes, which can no longer communicate if minimal routing is used. By allowing non-minimal paths, these nodes can be avoided and communication is still possible.

Congestion can occur even in a moderate traffic conditions in wormhole routing. Dead locks also lead to congestion in the network. In (5) Boppana et al. presented a methodology for designing a dead lock free wormhole Routing Algorithms. Global knowledge based self tuned source throttling mechanism for multi processor interconnection network offers attractive scheme for achieving congestion control. These networks automatically determine when network near saturation occurs by monitoring all the buffers state thus facilitating earlier detection rather than wait for network back pressure to create locally observable indicators for congestions. Self tuning mechanism automatically monitors global network throughput and determines appropriate threshold value. Adaptive routing dynamically chooses from the multiple potential routes available based on current local network state. This can alleviate problems of deterministic routing algorithms under heavy loads. In (62) JY Njai et al. present a frame work for adaptive routing in Multi Computer networks. However a fully adaptive algorithms can cause congestion. Thus to keep network below saturation and avoid resulting performance degradation, it is necessary to implement congestion control mechanism that adapts itself to communication loads and topologies. Lopez et al.(54), (55) use the number of busy output virtual channels in a node to estimate congestion. Baydal et al.(14) proposed an approach that
counts a sub set (free and useful) of virtual channel buffers to decide to throttle or not. In (82) YJ Njai and CI Seitz, presented a frame work for adaptive routing in Multi Computer networks.

Smai and Thorelli (86) and Mithuna Thottethodi et al (56), describe a form of global congestion control. In (86), a node that detects congestions signal all the other nodes in the network to limit packet injection. In (56) authors have used global knowledge of full network buffers to detect potential congestion and limit packet injection. The technique proposed by Kim et al.(42) allows sender to kill any packet that has experienced more delay than a threshold. Both (86) and (42) do not discuss self tuning which becomes critical when different communication patterns are encountered. Scott and Sohi (85) uses explicit feedback to report to node when tree saturation is imminent in multi stage interconnection network. In (65) authors LS Peh et al., have proposed a flit reservation flow control as an alternate flow control scheme to improve network utilization at which saturation occurs. It uses control flits to schedule bandwidth and buffers ahead of arrival of data flits. In (27) virtual channels as a means to control flow has been discussed. In (34) FW Farmer et al. have proposed a distributed switching mechanism to achieve control inc case of busty traffic.

2.9 Communication Latency Studies

Usually, evaluation of the performance of interconnection networks is done using diameter, bisection width, number of edges per node, constant edge length, etc., but the parallel algorithm is not explicitly taken into consideration while evaluating the networks. Parallel algorithms are usually mapped on to processor interconnections such as mesh, hypercube and tor to get the benefits of parallelism in the algorithms. Each mapping has its own advantages. Effectiveness of implementing a parallel algorithm on MCDL network is important for a network planner to decide optimal target organization amongst options available like hyper cube, mesh, and MCDL network etc. Interconnection networks can be evaluated based on latency for implementing a parallel algorithm.
Parallel organization in information processing emphasizes the concurrent processing of data in a single problem. An interconnection network is a topological representation of a processor organization. The distribution of data structures among the processors usually confirms to processors and its data manipulations. A data graph DG is a representation of data manipulation patterns in which vertices represent a data subsets, and edge represents a computation involving data from two data subsets. A processor organization is represented by a target graph TG in which the nodes represent processors with local memory and edges represent the communication paths between pairs of processors. The data graph has to be embedded into the target graph with minimum dilation (58). An embedding of data graph DG=(V, E) into a target graph TG=(V', E') is a mapping function f, between data graph and target graph such that for each vertex u in DG, f(u) is a unique vertex in TG. The dilation in embedding is defined as \( \text{dil}(f) = \max \{ \text{dist}_{TG} \{ f(u), f(v) \} \mid (u, v) \in E \}, \) u \neq v and also \( f(u) \neq f(v) \). An embedding of data graph in to target graph with dilation of 1 is possible if the former is a sub graph of the later. Though a data graph can be embedded into target graph of many processor organizations, there exists variation in their performances due to the dilation factor. By Some might exhibit a lower latency, some higher. This means that there is a need to evaluate processor organization based on dilation and latency.

Distribution of data plays an important role in performance of a processor organization for implementing a parallel algorithm. Data, which resides in the local memory of a processor, can be accessed faster than non-local data or remotely located data. A data graph represents the data manipulation patterns of a parallel algorithm. This has to be suitably mapped into the corresponding target graph of the processor organization graph to achieve minimal latency. Performance may suffer if the data graph is not sub graph of the target graph, because the latency increases as the dilation increases. Passing a message from one processor to nonadjacent processor requires longer time than it takes to pass a message between adjacent processors. In (25), the embedding of hypercube network on 2D Mesh has been discussed, by considering issues like communication pipe lining and optimal scheduler for lowering latency. Inter processor communications, arising due to reallocation of workloads and data, and under limited buffer conditions has been discussed by Al Pinar et.al in (66).