A Spanning Multichannel Linked Hypercube: A Gradually Scalable Optical Interconnection Network for Massively Parallel Computing

Ahmed Louri, *Senior Member, IEEE*, Brent Weech, *Student Member, IEEE*, and Costas Neocleous, *Student Member, IEEE*

Abstract—A new, scalable interconnection topology called the Spanning Multichannel Linked Hypercube (SMLH) is proposed. This proposed network is very suitable to massively parallel systems and is highly amenable to optical implementation. The SMLH uses the hypercube topology as a basic building block and connects such building blocks using two-dimensional multichannel links (similar to spanning buses). In doing so, the SMLH combines positive features of both the hypercube (small diameter, high connectivity, symmetry, simple routing, and fault tolerance) and the spanning bus hypercube (SBH) (constant node degree, scalability, and ease of physical implementation), while at the same time circumventing their disadvantages. The SMLH topology supports many communication patterns found in different classes of computation, such as bus-based, mesh-based, and tree-based problems, as well as hypercube-based problems. A very attractive feature of the SMLH network is its ability to support a large number of processors with the possibility of maintaining a constant degree and a constant diameter. Other positive features include symmetry, incremental scalability, and fault tolerance. It is shown that the SMLH network provides better average message distance, average traffic density, and queuing delay than many similar networks, including the binary hypercube, the SBH, etc. Additionally, the SMLH has comparable performance to other high-performance hypercubic networks, including the Generalized Hypercube and the Hypermesh. An optical implementation methodology is proposed for SMLH. The implementation methodology combines both the advantages of free space optics with those of wavelength division multiplexing techniques. A detailed analysis of the feasibility of the proposed network is also presented.

Index Terms—Interconnection networks, scalability, massively parallel processing, optical interconnects, wavelength division multiplexing, product networks.

1 Introduction

Progress in VLSI technology, combined with the escalating demands for more processing power and speed, have recently produced a technological environment in which Massively Parallel Processors (MPPs), with hundreds or even thousands of processing elements (PEs), are becoming commonplace (examples include Intel Paragon, Cray T3D and T3E, IBM SP-1 and 2, MasPar MP-1 and 2, Stanford Dash, etc.). The interconnection network, not the PEs or their speed, is proving to be the decisive and determining factor in terms of cost and performance [1], [2], [3], [4]. Ideally, a network topology intended for massively parallel computation should have the following characteristics:

1) small or bounded degree (a small degree implies low design complexity and low cost, a bounded or fixed degree implies scalability and fixed PE complexity),
2) ability to incrementally add PEs to an existing network with minimal or no changes at all to the existing configuration,
3) a large number of PE-disjoint paths between any two pairs of PEs for increased reliability and fault tolerance,
4) the message routing should be simple to implement and flexible to route around faulty PEs in the network,
5) the diameter should be small to minimize communication delays,
6) must have a large communication bandwidth, and
7) must possess good embedding capabilities (ability to efficiently emulate a large number of commonly required interconnection patterns).

We should note that the diameter of a network remains an important criterion in characterizing the network delay (latency) even with the wide use of wormhole routing [5], [6]. Wormhole routing does not reduce the time required for intermediate PEs to process small packets (flits). Larger diameter networks must therefore keep a larger number of PEs busy for a single packet transfer, which leads to a reduction in throughput for the entire system [7].

To this end, several topologies have been proposed to fit different styles of computation. Examples include crossbars, multiple buses, multistage interconnection networks, and hypercubes, to name a few. Unfortunately, most of the presently known network topologies do not satisfy all the above characteristics. Among these, the hypercube has received considerable attention, due mainly to its good topological characteristics (small diameter, regularity, high connectivity, simple control and routing, symmetry, and fault tolerance) and its ability to efficiently permit

• *The authors are with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721. E-mail: {louri, weech}@ece.arizona.edu.*

Manuscript received 17 June 1996; revised 16 June 1997. For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number 100226.

1045-9219/98/$10.00 © 1998 IEEE

the embedding of numerous topologies, such as rings, trees, meshes, and shuffle-exchange, among others [8]. However, a drawback of the hypercube is its lack of scalability, which limits its use in building large size systems out of smaller size systems. The lack of scalability of the hypercube stems from the fact that the node degree is not bounded and varies as $\log_2 N$, where N is the total number of nodes. This property makes the hypercube cost prohibitive for large N. Most hypercube-based interconnection networks proposed in the literature [9], [10], [11], [12], [13], [14], [15] suffer from similar size scalability problems.

Recently, some networks have been introduced that are a product of hypercube topology with fixed degree networks such as the mesh, the tree, and the deBruijn [4], [14], [16] in the quest of preserving the properties of the hypercube while improving its scalability characteristics. In a previous study, the authors introduced the Optical Multi-Mesh Hypercube (OMMH) [17], [18], [19]. The OMMH is a network that combines the positive features of the hypercube (small diameter, regularity, high connectivity, simple control and routing, symmetry, and fault tolerance) with those of a mesh (constant node degree and size scalability). The OMMH can be viewed as a two-level system: a local connection level representing a set of hypercube modules and a global connection level representing the mesh network connecting the hypercube modules. The OMMH network has been physically demonstrated using a combination of free-space and fiber optics technologies, and has shown good performance characteristics [20] for a reasonable size network. However, for very large networks (greater than 1,000 PEs), the OMMH experiences a logarithmic increase in diameter and requires a large amount of fiber, which makes the implementation complicated and expensive.

In this paper, we propose a novel network that improves the topological characteristics, as well as the implementation and performance aspects, of the OMMH network. The new network topology proposed is called *Spanning Multichannel Linked Hypercube* (SMLH), and possesses a constant degree and a constant diameter while preserving many properties of the hypercube. The SMLH, similar to the OMMH, employs the hypercube topology at the local connection level. The global connection level connecting the hypercube modules is called a *Spanning Multichannel Network* (SMN) and resembles a Spanning Bus Hypercube (SBH) network [21], where each bus is replaced with a multichannel link. The SBH is a D-dimensional lattice of width w in each dimension. Each node is connected to D buses, one in each of the orthogonal dimensions; w nodes share a bus in each dimension which permits only one data transfer at a time. The spanning bus hypercube offers small node degree, small diameter, and scalability. It can be scaled up by expanding the size of the spanning buses [21]. However, expanding the size of the buses leads to an $O(w)$ increase in traffic density [21] which in turn leads to bus congestion problems [22]. In the SMN, each multichannel link spanning a dimensional axis represents w distinct logical channels allowing up to w simultaneous data transfers. The advantages of the SMLH network are that it utilizes both the hypercube local interconnection level and the SMN multiple logical channels to decrease traffic density, which alleviates the bus congestion problems encountered in pure SBH networks. This feature allows larger systems to be built, since it allows the SMLH multichannel links to support a larger number of processors than the SBH network. Another attribute of the SMLH is incremental scalablilty with a high degree of connectivity and a low diameter. Additionally, we propose an optical implementation of such a network. Optical interconnects offer many desirable features, such as very large communication bandwidth, reduced crosstalk, immunity to electromagnetic interference, and low power requirements [3], [4], [23], [24], [25], [26], [27], [28], [29], [30].

TABLE 1
PROPERTIES OF PRODUCT NETWORKS

Network Property	Relationship
Resultant Graph	$G_1 \times G_2$
Size (N)	$N_1 \times N_2$
Degree (k)	$k_1 + k_2$
Diameter (d)	$d_1 + d_2$
Number of Links (L)	$N_1 L_2 + N_2 L_1$

2 Structure of the Spanning Multichannel Linked Hypercube Network

In this section, we formally define the structure of the SMLH network and discuss its properties.

2.1 Topology of the SMLH Interconnection Network

The topology of the SMLH can be described as an undirected graph, $G_{SMLH} = (V, E)$, where V represents a set of nodes and E represents a set of edges. The SMLH can also be viewed as a product hybrid graph because it combines a D-dimensional spanning multichannel graph and a binary hypercube graph in such a way that if $G = G1 \times G2$, where $G1$ represents the spanning multichannel graph and $G2$ the Binary Hypercube (BHC) graph, then the Cartesian product of their vertices is $V1 \times V2 = [(u2, u1) \mid u2 \in V2 \text{ and } u1 \in V1]$ [14]. This resultant graph relationship and other performance properties for product networks are shown in Table 1 [31]. The SMN resembles the SBH topology introduced by Wittie [21], with a major improvement in performance and complexity. Each multichannel link with w PEs represents w distinct logical channels which allow up to w simultaneous data transfers. In contrast, in the SBH, each bus can only allow one data transfer at a time.

The SMLH can be characterized by a three-tuple (w, n, D), where w, n, and D are positive integers. The first parameter, w, defines the number of nodes attached to each multichannel link. The second parameter, n, is the degree of the point-to-point n-cube (hypercube). The third parameter, D, identifies the number of dimensions spanned by a node in the SMN. To this end, the SMLH can be viewed as a product hybrid network combining two subnetworks: a D-dimensional SMN and a hypercube.

For an SMLH(w, n, D), the number of nodes, $|V|$, is equal to $w^D 2^n$. A node address in the SMLH is denoted by a $(w + 1)$-tuple $(a_1, a_2, \cdots, a_w, a_{w+1})$ using a mixed radix system, where, for $i = 1$ to $i = w$, $0 \le a_i < w$, and $0 \le a_{w+1} < 2^n$.

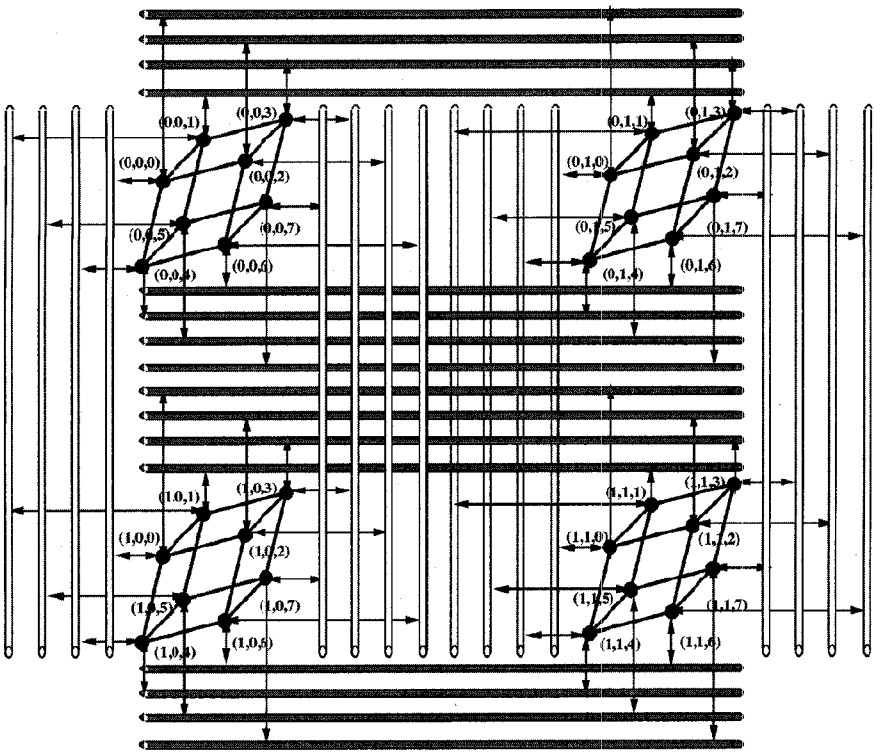


Fig. 1. An example of the spanning multichannel linked hypercube network: An SMHL(w = 2,n = 3) (32 nodes) interconnection is shown. Thick lines represent SMN connections, while bold, thin lines represent point-to-point hypercube connections.

Given the set of nodes (V), the set of edges (E) is constructed as follows: For two nodes (a_1, a_2, $\cdots$, a_w, a_{w+1}) and (b_1, b_2, $\cdots$, b_w, b_{w+1}) where, for $i = 1$ to $i = w$, $0 \le a_i < w$, for $j = 1$ to $j = w$, $0 \le b_j < w$, $0 \le a_n < 2^n$, and $0 \le b_{w+1} < 2^n$:

1) The two nodes span the same SMN if
   a) $a_{w+1} = b_{w+1}$, and
   b) if, for $i = 1$ to $i = w$, there are only two components, a_j and b_j, that are different, while all the other components are identical.
2) There is a link (called a hypercube link) between two nodes if and only if, for $i = 1$ to $i = w$,
   a) $a_i = b_i$, and
   b) a_{w+1} and b_{w+1} differ by one bit position in their binary representation (Hamming distance of one).

Fig. 1 shows an SMLH(w = 2, n = 3, D = 2) interconnection where the bold, thin lines represent point-to-point hypercube links and thick lines represent SMNs. Small, dark circles represent nodes of the SMLH network which are, in this paper, abstractions of processing elements, memory modules, or switches. Note that, because D = 2, each node spans two SMNs, one along each dimension. Furthermore, there are three bidirectional point-to-point links attached to a node which corresponds to the hypercube links. A careful observation of Fig. 1 shows that the node addresses satisfy the connection rules outlined earlier.

In this paper, we only consider SMLH networks with D = 2 and leave higher dimensioned SMLH network analysis for future work. Therefore, in the notation, the third parameter, D, will be dropped. Consequently, an SMLH(w, n, D = 2) network will be referred to as SMLH(w, n). As can be seen in Fig. 1, the SMLH(w = 2, n = 3) consists of $2^2 \times 2^3 = 32$ nodes, and it can be viewed as eight concurrent 2D SMNs. Note that w horizontal multichannel links and w vertical multichannel links are needed to form one $w \times w$, 2D SMN. Fig. 2 shows one such 2D SMN formed by nodes with the same hypercube addresses and belonging to different hypercube modules. Similar considerations take place for the other seven 2D SMNs in Fig. 1. The SMLH(w = 2, n = 3) network can also be viewed as four concurrent three-dimensional hypercubes in which four nodes, having identical hypercube addresses, form a 2×2 SMN. The SMLH(w = 2, n = 3) in Fig. 1 looks like a hypercube-clustered, spanning multichannel network. In general, there are 2^n 2D SMNs and w^2 hypercube modules. Note that when w is equal to one, the SMLH becomes a pure hypercube network, while when n is equal to zero, it becomes a pure spanning bus network if the multichannel link is constrained to operate on a single channel. This implies that both the hypercube and the SBH can be thought of as special cases of the SMLH network.

The choice of two parameters, w and n, completely determines the size of the network, the resources and implementation requirements, and the scaling complexity. The w parameter determines the size of the multichannel link lattice, while the n parameter defines the size of the hypercubes. From a scaling viewpoint, two scaling rules can be applied for an SMLH(w, n) network. The first rule, which we call "fixed-w" rule, keeps the size of the multichannel links constant and increases the size of the network by increasing n. The second rule, which we call "fixed-n" rule, keeps the size of the hypercube constant and increases the size of the network by increasing w. Clearly, the advantage of the SMLH(w, n) network is its flexibility to scale up using either, or a combination of, the two scaling rules.

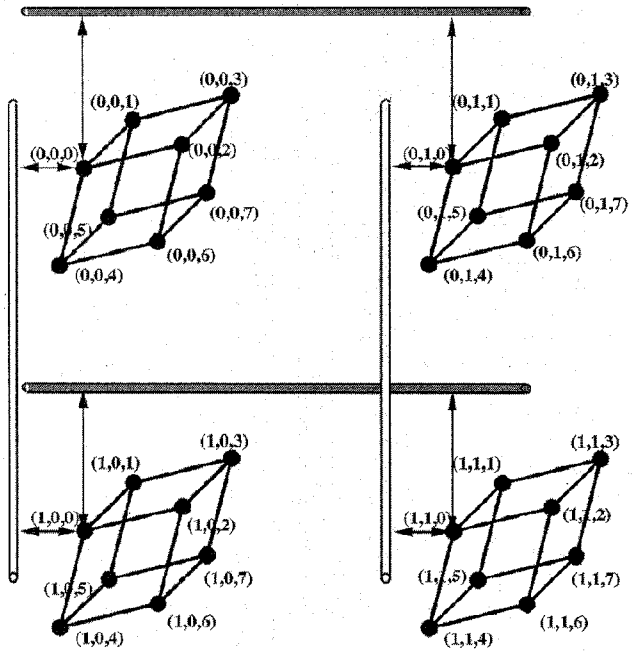


Fig. 2. An example of a 2D SMN within an SMHL($w = 2$, $n = 3$) network. Note that the nodes that construct the 2D SMN belong to different hypercube modules but they possess the same binary hypercube address representation within their corresponding hypercube modules. Eight such 2D SMNs coexist in the SMLH($w = 2$, $n = 3$) interconnection.

For instance, the size of the SMLH can grow without altering the number of links per node by expanding the size of the multichannel links. For example, three-dimensional hypercubes can be added on the perimeter of the 2D SMN of Fig. 1. Fig. 3 illustrates an SMLH($w = 3$, $n = 3$), which is constructed by expanding the SMLH($w = 2$, $n = 3$) network by adding hypercube modules along an outer row and an outer column. The existing configuration of the nodes of the SMLH($w = 2$, $n = 3$) network did not change because each node still spans two multichannel links and still has three bidirectional point-to-point links for the hypercube connections. This option allows the SMLH to be size scalable. Some discussion of the relationship between the two scaling rules and tradeoffs in performance are discussed in Section 2.3.

2.2 Message Routing in the SMLH Interconnection Network

Due to the regularity and symmetry of the SMLH architecture, a distributed routing scheme can be implemented without global information. At the source node, the message is formatted with the source address, the destination address, message length, and a few control bits, such as semaphore bits. The interprocessor message traffic of a node gets redistributed into two categories, i.e., the hypercube communication and the spanning multichannel communication. If the source and the destination of the message are within the same hypercube subnetwork of the SMLH network, the routing procedure is exactly the same as that of the regular hypercube network. Similarly, if the source and the destination of the message are within the same SMN of the SMLH network, the routing procedure is exactly the same as that of a regular bus connected network [18].

If the source and the destination of the message share neither a hypercube nor a 2D SMN, the routing scheme uses the hypercube routing scheme until the message arrives at the same hypercube node as the destination and, then, uses the SMN routing scheme for the message to arrive at the destination. Or, the SMN routing scheme can first be applied to forward the message to the same hypercube where the destination resides and, then, the message can reach the destination using the hypercube routing scheme. We can also mix the hypercube and the SMN routing until the message is forwarded to the same hypercube or to the same SMN where the destination resides and, then, we can forward the message to the destination using the hypercube or the SMN routing scheme, respectively.

2.3 Properties of the SMLH Interconnection Network

2.3.1 Diameter and Link Complexity

The diameter of a network is defined as the maximum distance between any two processors in the network. Thus, the diameter determines the maximum number of hops that an average message may have to take. The diameter of a 2D SMN is two, as it is for a 2D SBH, and the diameter of a binary hypercube with N_{BHC} nodes is $n = \log_2 N_{BHC}$. Therefore, the diameter of SMLH(w, n) is ($n + 2$). For the SMLH(w, n) network with N nodes, $N = w^2\, 2^n$, therefore, $n = \log_2 \frac{N}{w^2}$. Consequently, the diameter of the SMLH(w, n) network can be written as $\log_2 \frac{N}{w^2} + 2$. Using the fixed-w scaling rule, the diameter of the SMLH(w, n) network experiences a logarithmic increase ($O(\log_2 N)$) when the network size increases. However, using the fixed-n scaling rule would make the diameter constant, $n + 2$, for any network size.

Link complexity or node degree is defined as the number of physical links per node. For a regular network, where all nodes have the same number of links, the node degree of the network is that of a node. The node degree of a hypercube with N_{BHC} nodes is $n = \log_2 N_{BHC}$ and that of a 2D SMN is two, since there are only two physical SMNs links connected to each node. A node of an SMLH(w, n) network possesses links for both the hypercube connections and the SMN connections. Consequently, the node degree of the SMLH network is ($n + 2$) or $\left(\log_2 \frac{N}{w^2} + 2\right)$. Again, when using the fixed-w scaling rule, the SMLH network experiences a logarithmic increase in degree ($O(\log_2 N)$); however, when the network is expanded using the fixed-n scaling rule, the degree becomes constant, $n + 2$.

2.3.2 Bisection Width

The bisection width of a network is defined as the minimum number of links that have to be removed to partition the network into two equal halves [32]. The bisection width indicates the volume of communication allowed between any two halves of the network with an equal number of nodes. The bisection width of a n-dimensional hypercube is $2^{n-1} = N_{BHC}/2$ since that many links are connected between two $(n-1)$-dimensional hypercubes to form an n-dimensional hypercube. The equation of bisection width for the SMLH

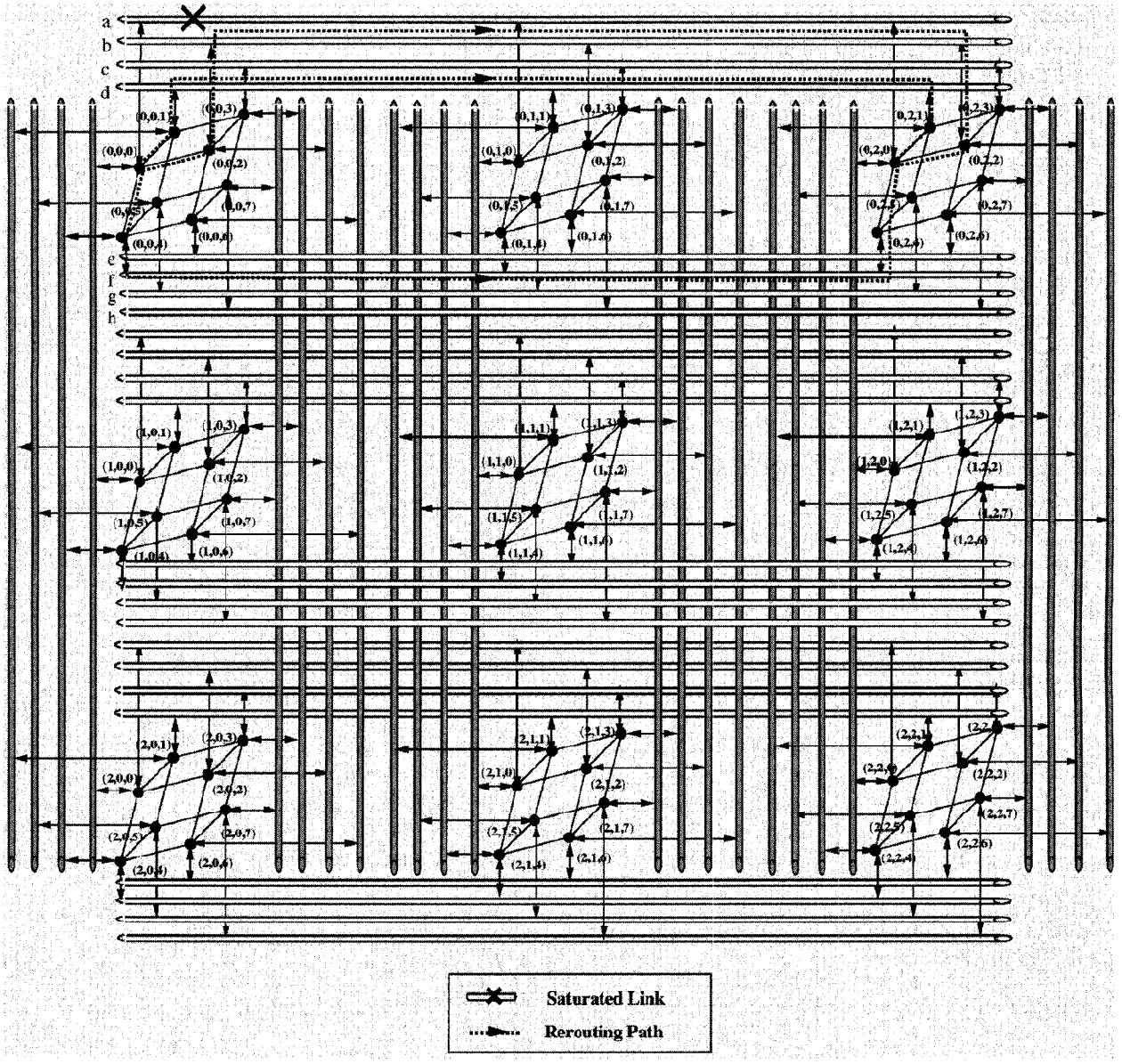


Fig. 3. An SMLH($w = 3$, $n = 3$) (72 nodes) interconnection. This SMLH network can be constructed by adding hypercube modules along a row and a column.

varies depending on whether w is even or odd, since w odd dictates cutting at least one hypercube in half. For w even, there are w such n-dimensional hypercubes connecting 2^n 2D SMNs, the bisection width of an SMLH(w even, n) is equal to $w \times 2^n = N/w$. For w odd, there is one such n-dimensional hypercube that must be cut in half in addition to cutting all multichannel links in one dimension and 2^{n-1} multichannel links in the second dimension. Therefore, the bisection width of an SMLH(w odd, n) is equal to $w \times 2^n + 2^n + 2^{n-1} = N/w + 3N/2w^2$.

2.3.3 Granularity of Size Scaling

The granularity of size scaling is the ability of the system to increase in size with minor or no change to the existing configuration, and with an expected increase in performance proportional to the extent of the increase in size. For a 2D SMN, the granularity of size scaling is only $2w + 1$ since, at a minimum, one multichannel link per dimension could be added to the network in order to increase its size. Therefore, the granularity of the size scaling in a $w \times w$ 2D SMN of $N_{SMN} = w^2$ nodes is $2N_{SMN}^{1/2} + 1$. However, the size of a hypercube can only be increased by doubling the number of nodes; that is, the granularity of size scaling in an n-dimensional hypercube is 2^n. When the fixed-w scaling rule is applied, the granularity of size scaling follows the hypercube size scaling. Therefore, the granularity of size scaling for the SMLH, using the fixed-w rule, is $w^2 \times 2^n = N$. When the fixed-n scaling rule is used, the granularity of size scaling follows that of the SMN. Therefore, the granularity of size scaling following the fixed-n rule is $2^n(2w + 1) = 2(N/w) + 2^n$. Note that the granularity of size scaling using the fixed-w rule is $O(N)$ while, for the fixed-n rule, it is $O(N/w)$. SMLH is competitive with other networks in this smooth scalability property [33].

2.3.4 Average Message Distance

The average message distance in a network is defined as the average number of links that a message should travel between any two nodes. Let N_i represent the number of nodes at a distance i, then the average distance $\bar{l}$ is defined as [12]:

$$\bar{l} = \frac{1}{N-1}\sum_{i=1}^{n} iN_i, \quad (1)$$

where N is the total number of nodes, and n is the degree. Since the SMLH is a product network of the BHC and SMN, the average message distance can be derived from the individual average message distances for the BHC and SMN. For the BHC, the number of nodes at a fixed distance, $N_{i,BHC}$, is given as [12]:

$$N_{i,BHC} = \binom{n}{i}(w-1)^i, \quad (2)$$

since, for i differing digits which are each able to vary in $(w-1)$ ways, the enumeration is $(w-1)^i$. Substituting the equation into (1) and computing the summation yields the common form of the BHC average message distance $\bar{l}_{BHC}$ [12]:

$$\bar{l}_{BHC} = \frac{n}{2}\left(\frac{N_{BHC}}{N_{BHC}-1}\right). \quad (3)$$

For a 2D SMN, the number of nodes at distance 1, $N_{1,SMN}$ is $2(w-1)$ since there are $w-1$ nodes in each direction, not including the source node. Since the diameter of the 2D SMN is two, the number of nodes at distance 2, $N_{2,SMN}$, is equal to all nodes not at distance 1, minus the source node, i.e.,

$$N_{2,SMN} = w^2 - 2(w-1) - 1. \quad (4)$$

Since the SMLH network is not required to use time division multiple access (TDMA), as the SBH is, to access a shared medium, the above equations for $N_{i,SMN}$ will result in an average message distance for a 2D SMN, l_{SMN}, which differs from the traditional average message distance equation for a SBH of [21]:

$$\bar{l}_{SMN} = \frac{D(w-1)}{2}\left(\frac{N_{SMN}}{N_{SMN}-1}\right). \quad (5)$$

Assuming a collision free environment, an average message can use the hypercube routing scheme until the message arrives at the same hypercube position and then use the SMN routing scheme to arrive at the destination, as described in Section 2.2. Or, the SMN routing scheme can first be applied to forward the message to the same hypercube where the destination resides and, then, use the hypercube routing scheme. Therefore, an average message has the potential of encountering any of the $(N_{BHC}-1)$ nodes in a given hypercube and any of the $(N_{SMN}-1)$ SMN nodes, where N_{BHC} is the total number of nodes in a single binary hypercube and N_{SMN} is the number of nodes in the SMN network. Therefore, the average message distance in the SMLH, $\bar{l}_{SMLH}$, can be calculated as:

$$\bar{l}_{SMLH} = \frac{1}{(N_{BHC}-1)+(N_{SMN}-1)}\left(\sum_{i=1}^{n} iN_{i,BHC} + \sum_{i=1}^{2} iN_{i,SMN}\right), \quad (6)$$

which can be rewritten using the above equations as:

$$\bar{l}_{SMLH} = \frac{2w(w-1)+n2^{n-1}}{(w^2-1)+(2^n-1)}. \quad (7)$$

2.3.5 Fault Tolerance

Due to the concurrent presence of SMNs and hypercubes in the SMLH, rerouting of messages in the presence of a single faulty link or a single faulty node can easily be done with little modification of existing fault-free routing algorithms. In the SMLH, any single faulty link or any single faulty node can be side stepped by only two additional hops as long as that particular node is not involved in the communication, namely, the node is neither the source nor the destination for any message. This can be shown as follows: A message in the SMLH is routed using a SMN routing function if both the source and the destination of the message are in the same 2D SMN, or a hypercube routing function if they are in the same hypercube module, or a combination of these two routing functions if they are neither in the same SMN nor in the same hypercube module. Consider the rerouting scheme in the presence of a single faulty link when the SMN routing function is being applied. When we refer to a faulty link, we mean that a PE cannot access the SMN due to a failure. In such a case, the PE would not be able to communicate with other PEs that share the same SMN. The problem can be solved by forwarding the data to the neighboring SMN via one hop of the hypercube link (n such neighboring two-dimensional multichannel links exist in SMLH(w, n)). By using the neighboring SMN, the message arrives at a node which is one hop away from the destination, since the message has been routed in the neighboring SMN to detour the faulty link. Similarly, a single faulty link when the hypercube routing function is being applied can be side stepped by forwarding the message to the neighboring hypercube via a SMN operation, as shown in Fig. 3. In general, for an SMLH(w, n) network, n two-hop rerouting schemes are available to bypass a faulty link which is competitive with other networks [12], [33].

3 Comparisons of SMLH with Popular Networks

In this section, we compare the SMLH network with existing, well-known topologies. These include the Binary Hypercube (BHC) [8], the Generalized Hypercube (GHC) [12], the Torus [34], the Spanning Bus Hypercube (SBH) [21], the Hierarchical Cubic Network (HCN) [9], the Cube-Connected-Cycle (CCC) [35], the Hyper-deBruijn (HdB) [36], the Folded Peterson (FPT) [37], the Hypermesh (HM) [33], and the Optical Multi-Mesh Hypercube (OMMH) [17]. The comparison parameters include diameter, degree, number of links, average message distance, average traffic density, and queuing delay. The topological characteristics of the above networks are indicated in Table 2. The results of the comparison are shown in Figs. 4, 5, 6, and 7.

In the figures, the SMLH(w = 32, n) notation denotes that the network is expanded following the fixed-w (w = 32) rule; that is, the size of the multichannel links is kept constant (32 PEs each) and the size of the hypercube module is changed to have the same network size for comparison purposes. The

TABLE 2
TOPOLOGICAL CHARACTERISTICS OF SEVERAL POPULAR NETWORKS

Network	Size (N)	Degree (k)	Diameter (d)	Number of Links (L)
BHC	2^n	$\log_2 N$	$\log_2 N$	$\frac{N}{2}(\log_2 N)$
Torus(w, n)	w^n	$2\log_w N$	$\frac{w}{2}\log_w N$	$N\log_w N$
GHC(w, n)	w^n	$(w-1)\log_w N$	$\log_w N$	$\frac{N(w-1)}{2}\log_w N$
SBH(w, D)	w^D	$\log_w N$	$\log_w N$	$\frac{N}{w}\log_w N$
HCN	2^{2n}	$(\frac{1}{2})(\log_2 N)+1$	$\log_2 N$	$((\frac{N}{4})(\log_2 N)+\frac{N}{2})$
CCC(c, n)	$c2^n$	3	$\lfloor\frac{(5c-2)}{2}\rfloor$	$\frac{3}{2}N$
HdB(c, n)	$2^{(n+c)}$	$\log_2 N - c + 4$	$\log_2 N$	$\frac{N}{2}(\log_2 N - c + 4)$
FPT	10^n	$3\log_{10} N$	$2\log_{10} N$	$\frac{3N}{2}\log_{10} N$
HM(w, n)	w^n	$\log_w N$	$\log_w N$	$N\log_w N$
OMMH(l, l, n)	$l^2 2^n$	$4+\log_2 \frac{N}{l^2}$	$l+\log_2 \frac{N}{l^2}$	$\frac{N}{2}\left(4+\log_2 \frac{N}{l^2}\right)$
SMLH(w, n)	$w^2 2^n$	$2+\log_2 \frac{N}{w^2}$	$2+\log_2 \frac{N}{w^2}$	$\frac{N}{2}\left(\frac{4}{w}+\log_2 \frac{N}{w^2}\right)$

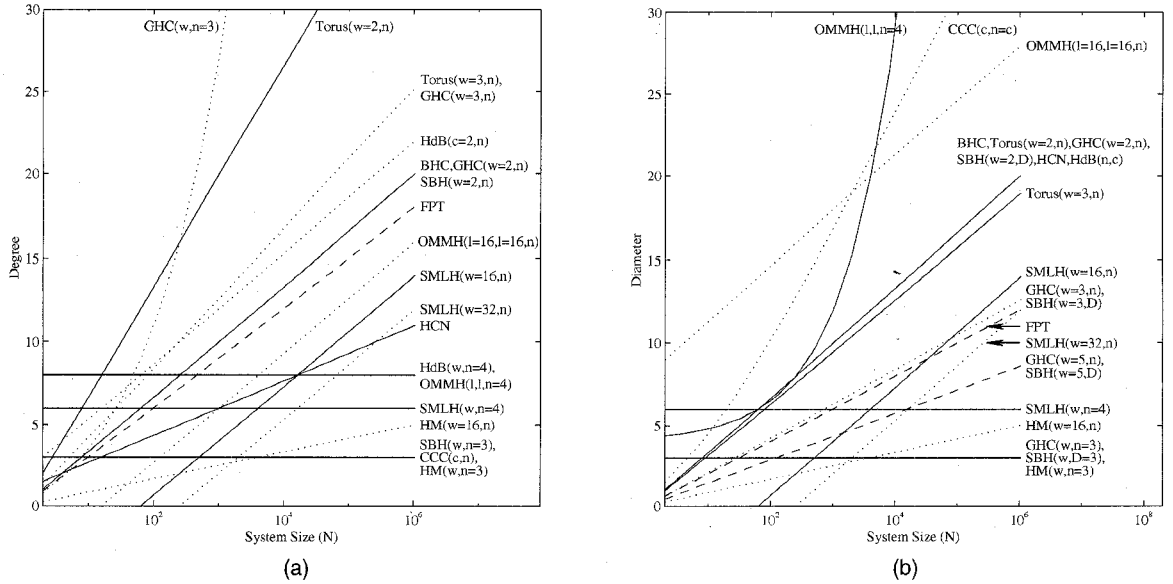
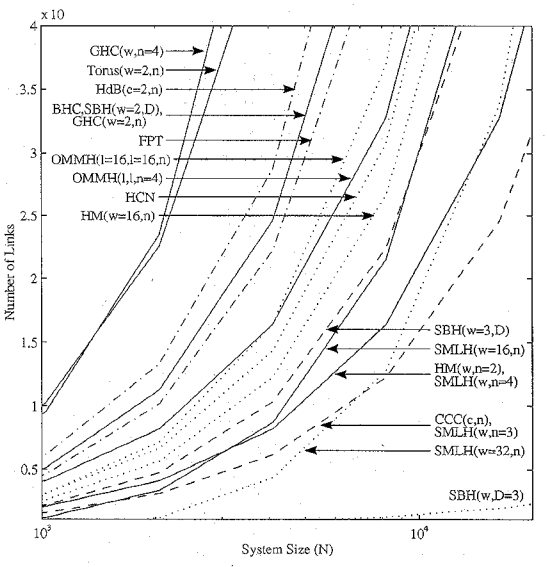


Fig. 4. Network comparisons for (a) degree (b) diameter.

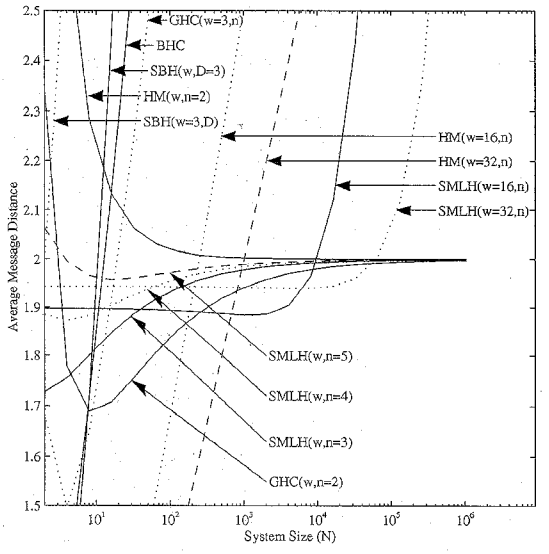
SMLH(w, n = 4) notation denotes that the network is expanded following the fixed-n rule, which means that the size of the hypercube module (4) is kept fixed and the size of the multichannel links are increased. Note that when expanding the SMLH network, some mathematical constraints exist. In Section 2.3, the degree and diameter of the SMLH network were derived; they are both equal to $\log_2 \frac{N}{w^2} + 2$. The first term of the equation is a factor of both the number of nodes in the entire network (N) and the size of the multichannel links (w). The constraint is that $N \geq w^2$ because, otherwise, the $\log_2 \frac{N}{w^2}$ factor of the degree/diameter equation will give a negative number, which would be unacceptable. The notation Torus(w = 3, n) means the number of nodes per ring is fixed while the hypercube dimension varies. The notation SBH(w, D = 3) means that the dimension of the SBH network is kept constant and the size of the buses is changed. The notation GHC(w, n = 4) denotes that the diameter of the network is fixed while the number of nodes per dimension is varied. Similarly, the notation HM(w, n = 2) means that the node dimension is fixed, while the number of nodes per dimension is varied. The notation OMMH(l = 16, l = 16, n) denotes that the size of the mesh network in the OMMH is fixed while the size of the hypercube is varied. Similarly, the OMMH(l, l, n = 4) notation denotes that the size of the hypercube is fixed and the mesh size is varied. The notation used for the other networks is similarly given.

3.1 Degree and Diameter

Figs. 4a and 4b show the graph comparisons in terms of degree and diameter as network size is increased. At the key mark of 10,000 nodes (desirable for MPPs), SBH(w, D = 3), HM(w, n = 3), HM(w = 16, n), SMLH(w = 32, n), SMLH(w, n = 4), and SMLH(w = 16, n) exhibit very good performances in terms of diameter and degree with values 3, 3, 3.3, 5.3,



(a)



(b)

Fig. 5. Network comparisons for (a) number of links and (b) average message distance.

6, and 7, respectively. The CCC(c, n) and HCN reveal good degree values (3 and 7.6), but they also exhibit fairly large diameters (17 and 13). The OMMH(l, l, n = 4) experiences the worst diameter (29) and the GHC(w, n = 3) the worst degree (61). Even though at 10,000 nodes the SMLH(w = 32, n) reveals better characteristics than the SMLH(w, n = 4), the latter is more desirable because it possesses constant degree and diameter, features that allow it to be scalable. The SMLH(w = 32, n), on the other hand, experiences a logarithmic increase in degree and diameter, features that make it difficult to scale up to a larger number of processors. In general, from Fig. 4a, the hybrid networks show a logarithmic increase in their degree which makes them difficult to scale them up in size.

3.2 Number of Links

Fig. 5a shows the graph comparisons in terms of the number of links as the networks scale up in size. The SBH(w, D = 3),

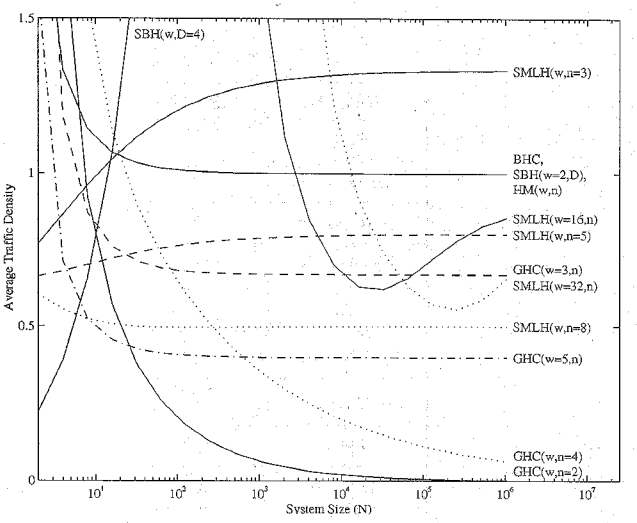


Fig. 6. Network comparisons for average traffic density.

CCC(c, n), SMLH(w, n = 3), SMLH(w = 32, n), HM(w, n = 2), SMLH(w, n = 4), SMLH(w = 16, n), and SBH(w = 3, n) reveal the best performance characteristics in terms of number of links, while the GHC(w, n = 4), and Torus seem to require a larger number of links. We should note that, of the SMLH links, $2N/w$ of them are actually multichannel links.

3.3 Average Message Distance

Fig. 5b illustrates graph comparisons between the BHC, SBH, SMLH, GHC, and HM networks in terms of average message distance. The average message distance of the SMLH is given in (7), the BHC in (3), and the SBH in (5), while equations for the other topologies of interest are [12], [33]:

$$\bar{l}_{GHC} = \frac{n(w-1)}{w}\left(\frac{N_{GHC}}{N_{GHC}-1}\right) \tag{8}$$

$$\bar{l}_{HM} = n\left(\frac{N_{HM}}{N_{HM}-1}\right). \tag{9}$$

At 10,000 nodes the SMLH(w = 16, n), SMLH(w = 32, n), SMLH(w, n = 3), SMLH(w, n = 4), SMLH(w, n = 5), HM(w, n = 2), and GHC(w, n = 2) exhibit the best average message distances with values close to two. The SBH(w, D = 3), SBH(w = 3, D), and BHC experience the worst average message distances with values close to 31, 13, and 6.6, respectively. Fig. 5b indicates the average message distance of the SMLH and GHC, both having fixed size hypercubes, as well as the HM(w, n = 2) approach a constant value for large N. However, the other networks (excluding the SBH with fixed size hypercubes) grow logarithmically with respect to the network size.

3.4 Average Traffic Density

As emphasized in Section 2, the advantages of the SMLH(w, n) network over the SBH network is its ability to use the multichannel links and point-to-point hypercube links to alleviate the spanning bus congestion. A very good measure of that is the average traffic density. The average traffic density is defined as the product of the average distance and the total number of nodes, divided by the total number of communication links [26]. Using the definition

stated above, the average traffic density of the SMLH(w, n) network can also be calculated:

$$\rho = \frac{\bar{l}_{SMLH} N}{\frac{N}{2}\left(\frac{4}{w} + n\right)}. \tag{10}$$

Therefore, the average traffic density can be rewritten as

$$\rho = \frac{2\bar{l}_{SMLH}}{\left(\frac{4}{w} + n\right)}, \tag{11}$$

where $\bar{l}_{SMLH}$ can be obtained from (7). Equation (11) reveals that, when the fixed-n rule is followed to expand the network, the average traffic density of the SMLH(w, n) is essentially independent of w. This feature allows the network to utilize a much large number of nodes along the multichannel links.

Fig. 6 presents the average traffic density comparisons between the BHC, SBH, SMLH, GHC, and HM networks. The HM, BHC, GHC with fixed w, and the SMLH with fixed sized hypercubes, have low traffic density and exhibit no sensitivity to size for larger networks. As the size of the hypercubes within the SMLH is increased, the traffic density approaches zero, as it also does for the GHC with fixed n. On the other hand, the SBH network with fixed dimension (SBH(w, D = 4)) shows an increase in traffic density. Therefore, for larger networks, the SBH network would most likely experience severe bus congestion problems, which would lead to large message delays.

3.5 Queuing Delay Analysis

For an exact analysis of the queuing delay of the SMLH network, we based our model on similar analysis of hypercubes [12] and optical hypercubes, meshes and hypermeshs [33]. The model is a communication net with the ith channel represented as an $M/M/1$ system with Poisson arrivals at a rate λ_i and exponential service time of mean $1/\mu c_i$. $1/\mu$ is the average packet size and c_i is the capacity of the ith channel. Additionally, the following assumptions are made [38]:

- Each node is equally likely to send a message to every other node in a fixed time period.
- The routing algorithm traverses dimensions in a static order.
- The load is evenly distributed, i.e., λ_i is the same for all i.
- The link capacities of the network have been optimally assigned.
- The cost per capacity per link is unity.

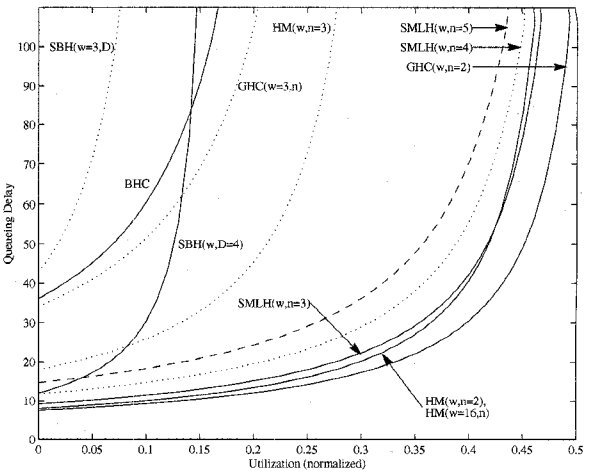
Under theses assumptions, an exact expression for the queuing delay is an M-channel N-node network is given by [38]

$$T = \frac{\bar{l}\left(\sum_{i=1}^{M}\sqrt{\frac{\lambda_i}{\lambda}}\right)^2}{\mu C\left(1 - \bar{l}\gamma\right)}, \tag{12}$$

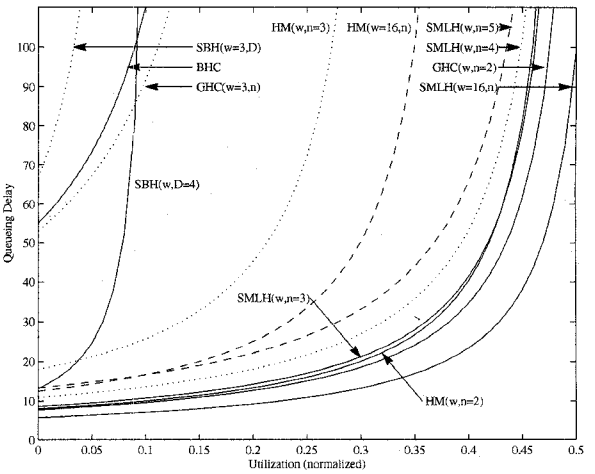
where M = total number of directed links,

$$\lambda = \sum_{i=1}^{M} \lambda_i = M\lambda_i,$$

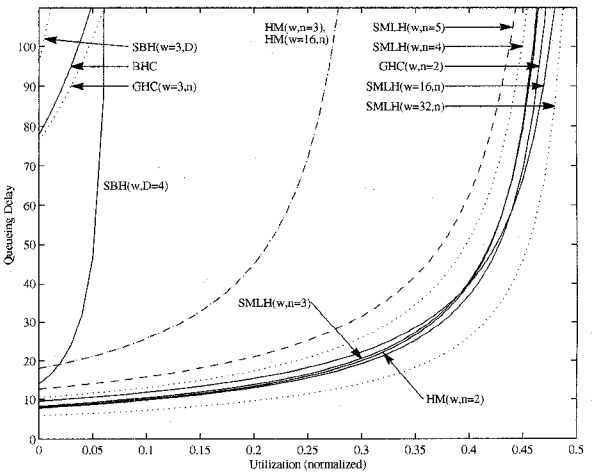
γ = the utilization factor, and $C = \sum_{i=1}^{M} c_i$ = aggregate bandwidth of the topology. Removing constants μ, C, and N, the above delay can be normalized to



(a)



(b)



(c)

Fig. 7. Queuing delay for networks of size (a) N = 256, (b) N = 1,024, and (c) N = 4,096.

$$T = \frac{\bar{l}M}{1 - \bar{l}\gamma}. \tag{13}$$

Figs. 7a, 7b, and 7c graph the average queuing delay in terms of increasing utilization for the BHC, SBH, SMLH, GHC, and HM with 256, 1,024, and 4,096 nodes, respectively. The delay increases exponentially with increasing utilization and

saturates at a particular load, given by $\gamma_{sat} = 1/\bar{l}$. The graphs illustrate the maximum traffic capability of each topology, with the SMLH, GHC with fixed n, and HM with fixed n having higher utilization and lower queuing times than the other networks.

The graphs further prove that the SBH network, despite its better topological characteristics (diameter, degree, etc.), has severe bus saturation problems for large numbers of PEs. The advantage of the SMLH network is that, in addition to its good topological characteristics, it demonstrates insensitivity to traffic density when the network scales up in size. This feature allows the SMLH network to grow up in size while maintaining good traffic capability. Nevertheless, even if saturation problems appear, the SMLH network does not experience the same message delays as the SBH because it can utilize the point-to-point hypercube links to redirect the packets from another path. In the SMLH network, the saturated multichannel link can be side stepped by only two additional hops as long as that particular saturated multichannel link is not involved in the communication. Fig. 3 demonstrates a rerouting scheme assuming a multichannel link saturation problem. Assume that, in Fig. 3, node (0, 0, 0) wants to send a package to node (0, 2, 0) via the horizontal multichannel link *a*. In case of a multichannel link saturation problem, the package will need to follow a different routing path. Three different routing paths are available and are all shown in Fig. 3 with thick dotted lines. The packet can utilize three hypercube links to access multichannel link *b* or multichannel link *d* or multichannel link *f*. By using one of the three multichannel links, the packet will arrive at a PE, which is one hop away from the destination link. Note that in each of the three rerouting paths, two additional hops were necessary to bypass the saturated multichannel link.

4 Optical Implementation of the SMLH Network

Obviously, an electronic implementation of the proposed SMLH network is feasible. One methodology would be to use multiprocessor board technology (e.g., Multichip Module technology) for the hypercube module connections and backplanes for the multichannel link connections. To limit the number of boards required, k hypercube modules can be clustered together on a single multiprocessor board. However, for a large number of PEs and a greater bandwidth and interconnection density, conventional backplanes have major limitations [3], [4], [39]. These include signal skew, wave reflection, impedance mismatch, skin effects, interference, and many others. A possible alternative is the use of optical interconnects. Optical interconnects offer many communication advantages over electronics, including gigahertz transfer rates in an environment free from capacitive loading effects and electromagnetic interference, high interconnection density, low power requirements, and, possibly, a significant reduction in design complexity through the use of multiple access techniques and the third dimension of free-space optics. The effectiveness of optical interconnects has been extensively examined [3], [4], [23], [24], [25], [26], [27], [28], [29], [30], [40]. In the following, we propose an all optical implementation of the SMLH(w, n) network, where the hypercube modules are implemented using free-space, space-invariant optics [17] and the multichannel link modules are implemented using Wavelength Division Multiple Access (WDMA) techniques.

4.1 Optical Implementation of the Hypercube Modules Using Holographic Optical Elements (HOEs)

The free space optical implementation of the hypercube network has been rigorously studied and analyzed [4], [17], [18]. The main objective is to exploit the third dimension and the communication advantages of free-space optics to provide efficient and adequate implementation of the hypercube network. This implementation is for illustration only and may not handle the simultaneous transfer of multiple, different packets, as a pure hypercube can. However, optics is quite capable of handling this function and a subsequent paper will address this in more detail.

The design methodology is based on an observation that PEs in a bipartite interconnection network can be partitioned into two different sets of PEs such that any two PEs in a set do not have a direct link. This is a well-known problem of bipartitioning a graph if the interconnection network is represented as a graph. For a binary n-cube, PEs whose addresses differ by more than one in Hamming distance can be in the same partition, since no link exists between two PEs if their Hamming distance is greater than one. Besides bipartitioning the graph, we arrange the PEs in each partition onto the plane such that interconnection between two planes becomes space-invariant (the connection pattern is identical for every PE in the plane). This self-imposed requirement reduces the design complexity of the optical setup [41]. The two partitions of PEs are called $plane_l$ and $plane_r$. Optical sources and detectors are assumed to be resident on processor-memory boards located on $plane_l$ and $plane_r$. Free-space holographic optical elements (HOEs) are used to implement the connection patterns required between PEs of the two planes [4], [17], [18].

Fig. 8 depicts the conceptual 3D free-space, space-invariant optical implementation of a three-cube network. A three-cube network consisting of eight PEs is bipartitioned into $plane_l$ and $plane_r$. Each plane consists of four PEs which are arranged in a two by two square configuration. The HOE provides the connection patterns required for the three-cube. The hypercube modules have been experimentally implemented in the lab [20]. Issues relating to actual fabrication of the hologram, size, misalignment, power budget, components used, actual BER, and more can be found in [20]. We should note that this is one way of implementing the Hypercube module in free space and several other means exist that are currently under study.

4.2 Implementation of the Spanning Multichannel Network Using WDMA Techniques

In this subsection, the implementation of the spanning multichannel network using WDMA techniques is presented. To exploit the large communication bandwidth of optics, WDMA techniques that enable multiple multiaccess channels to be realized on a single physical channel can be utilized. In a WDMA system, the optical spectrum is divided into many different logical channels, each channel

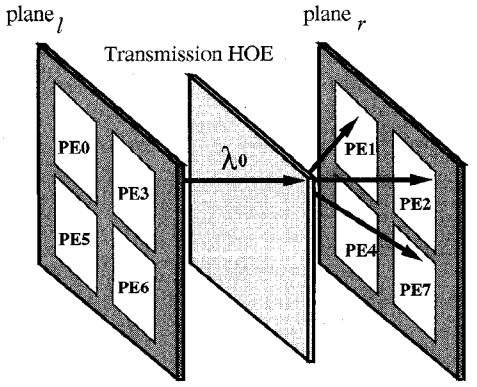


Fig. 8. Space-invariant optical implementation of a three-cube network with a HOE.

corresponding to a unique wavelength. These channels are carried simultaneously on a small number of physical channels, e.g., a fiber. Additionally, each network node is equipped with a small number of transmitters and receivers, some of which are dynamically tunable to different wavelengths.

The SMLH(w, n) network consists of w^2 hypercube modules and 2^n $w \times w$ 2D SMNs. From the discussion above, every hypercube module is bipartitioned into two planes called $plane_l$ and $plane_r$. In the SMLH network, all $plane_l$s are grouped together to form a plane called $Plane_L$ while all $plane_r$s are grouped to form another plane called $Plane_R$. The multichannel links can be implemented by interconnecting the individual $plane_l$ s of $Plane_L$ and $plane_r$ s of $Plane_R$. The hypercube modules are implemented using free space optics to provide the connectivity between $plane_l$s and $plane_r$s. Additionally, 2^{n-1} 2D spanning multichannel networks per plane ($Plane_L$ or $Plane_R$) need to be implemented. Each 2D SMN consists of $2w$ multichannel links, therefore, a total of $2w \times 2^{n-1}$ multichannel links per Plane are required.

A trivial implementation of the SMN is to assign a distinct wavelength for every PE in $Plane_L$ and $Plane_R$ and then perform WDMA techniques to implement the multichannel links. However, such a straightforward method requires a prohibitively large number of different wavelengths and fibers. For example, for an SMLH($w = 4$, $n = 3$) consisting of 128 PEs, a total of 64 wavelengths would be necessary. A wavelength assignment technique [17], [27] can be employed to reduce the number of wavelengths used in the system. Let's take a running example, an SMLH($w = 4$, $n = 3$). Fig. 9 shows how wavelengths are assigned for each PE of $Plane_L$. The following wavelengths are assigned to the first row: $\lambda_1, \lambda_2, \cdots, \lambda_8$. Then, $\lambda_2, \cdots, \lambda_8, \lambda_1$ are assigned as wavelengths in the second row. In general, wavelength assignment in a row is achieved by rotating the wavelength assignment of the previous row by one column. This wavelength assignment results in no two PEs in the same row or column of $Plane_L$ having an identical wavelength. Similar considerations take place for PEs of $Plane_R$. With this wavelength assignment technique, the total number of wavelengths required to implement the SMLH($w = 4$, $n = 3$) network is reduced from 64 to 8. In general, for an SMLH(w, n) the following wavelength assignment for the first row must be performed: $\lambda_1, \lambda_2, \ldots, \lambda_{w \times 2^{\lceil (n-1)/2 \rceil}}$, and then, $\lambda_2, \ldots, \lambda_{w \times 2^{\lceil (n-1)/2 \rceil}}$, are assigned to the PEs of the second row and so on. Thus, an implementation of an SMLH(w, n) with the above wavelength assignment requires no more than $w \times 2^{\lceil (n-1)/2 \rceil}$ wavelengths.

Referring to Fig. 9, the wavelengths assigned to the PEs of the first row are divided into two groups of four wavelengths each. The groups are: (λ_1, λ_3, λ_5, λ_7) and (λ_2, λ_4, λ_6, λ_8). Each of these groups correspond to the implementation of a row-wise multichannel link. Every PE in the group should be capable of tuning in to any of the wavelengths assigned to that group. For example, the node of group 1 with wavelength λ_1 must be able to tune to wavelengths λ_3, λ_5, λ_7 which correspond to wavelengths that were assigned to the other PEs of that group. Rotating the wavelength assignments of the previous rows will form the new wavelength groups that correspond to every row. Similarly, each column of Fig. 9 must be divided into two groups of four wavelengths each. For example, for the second column of Fig. 9, the following groups are formed: (λ_2, λ_4, λ_6, λ_8) and (λ_3, λ_5, λ_7, λ_1). Each of these wavelength groups correspond to the implementation of a column-wise multichannel link. Again, rotation of the column-wise wavelength assignment will result in the formation of the wavelength groups for the other columns.

We now consider the overall optical implementation of an SMLH(w, n). For simplicity and without loss of generality, we consider the implementation of an example network of size SMLH($w = 4$, $n = 3$). Fig. 9 shows an example $Plane_L$ of the SMLH($w = 4$, $n = 3$) network. We assume that each PE has three light sources: one fixed source, S_h, which illuminates the HOE to generate the required hypercube links, and the other two relatively tunable sources, S_r and S_c, coupled into optical fibers to implement the two multichannel links. It should be noted that full tunability is not required, as explained above. In fact, each source should be tunable for a single wavelength group only. This reduced range increases the efficiency, the yield, and the tuning speed of the light sources. Furthermore, each PE is equipped with three receivers; one fixed receiver, R_h, receives light from the free-space optics implementing the hypercube, and the other two receivers, R_r and R_c, receive light from fibers coming from demultiplexers. The key component that provides multichannel link connectivity here is the tunable-transmitter, fixed-receiver scheme. The wavelength assignment shown in Fig. 9 corresponds to the receiver wavelength assignment of every PE. Other PEs can communicate with a particular PE by simply tuning in to the wavelength assigned to that PE. Rapid progress is being made in the development of tunable devices, both in the range over which they are tunable, and their tuning times [42], [43]. Current tuning ranges are in the 4-10 nm and the tuning times vary from nanoseconds to milliseconds [42].

We also assume the availability of the following optical components: A $k \times 1$ passive optical star coupler and a diffraction grating. The $k \times 1$ coupler acts as an optical multiplexer by funneling light from k light beams with k

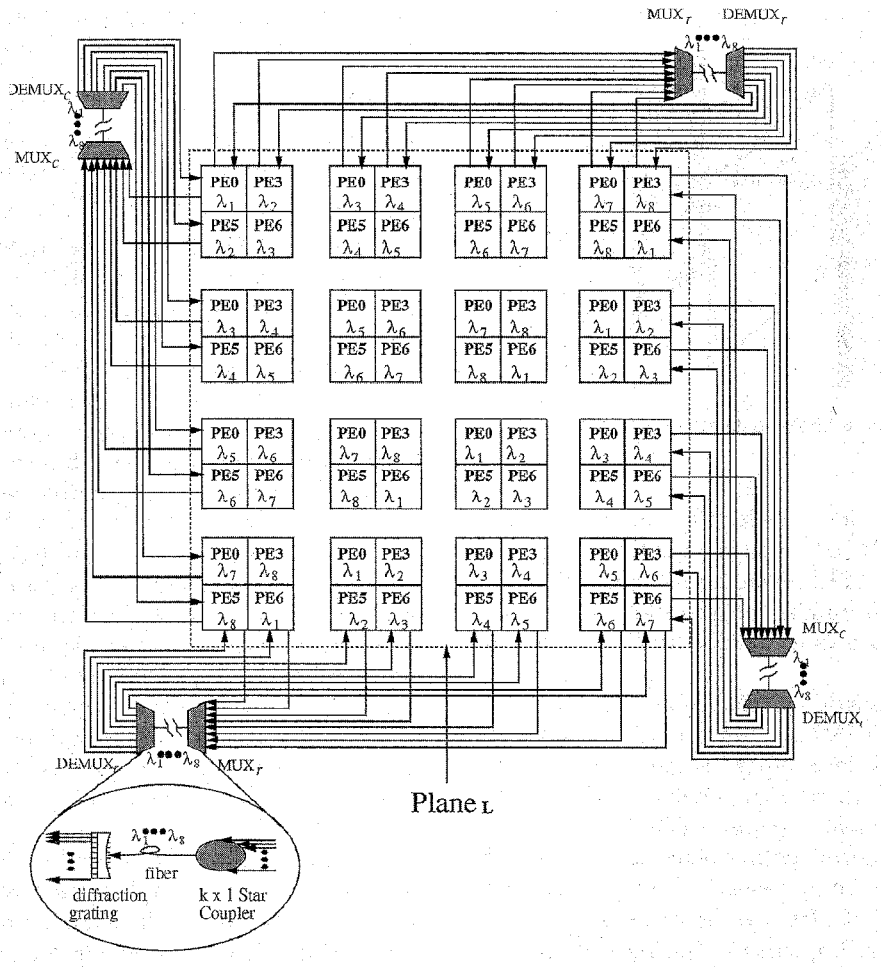


Fig. 9. An optical implementation of $Plane_L$ of an SMLH(w = 4, n = 3) network using WDMA. For an SMLH(w = 4, n = 3) network, 16 MUX-DEMUX pairs are needed per plane. For clarity of the figure, we show only four of them.

different wavelengths onto the same waveguide. The diffraction grating acts as an optical demultiplexer by separating and incoming combined optical signal into many signals, each corresponding to a different wavelength. Many examples that use a conventional lens, GRIN rod lenses [44], and no lens have been reported as optical demultiplexers. Recently, the idea of free-space concave grating demultiplexers has been reported [45], [46]. These demultiplexers use a wavelength range of 16 nm (40 channels) and channel spacing of 0.4 nm. The fiber-to-fiber loss was 6-9 dB and channel crosstalk was better than 25 dB. For clarity, in the rest of the paper, we denote the $k \times 1$ star coupler as a multiplexer (MUX) and the concave grating as a demultiplexer (DEMUX), bearing in mind that these are passive devices.

We should note that the functionality of a MUX-DEMUX pair can also be realized using a single $k \times k$ passive optical star coupler. The purpose of an $k \times k$ star coupler is to couple light from one of its k input ports to all the k output ports uniformly. Star couplers with 128×128 ports and the capability of handling more than one hundred different wavelengths are feasible with currently available technology. An experimental ISDN switch architecture using eight 128×128 multiple star couplers to handle over 10,000 input port lines has been reported [42].

In Fig. 9, MUX$_r$ and DEMUX$_r$ represent the MUX and DEMUX that implement row-wise multichannel links. Similarly, MUX$_c$ and DEMUX$_c$ represent the MUX and DEMUX that implement column-wise multichannel links. A MUX$_r$ multiplexes light signals from S_r sources emanating from PEs belonging to the same row of $Plane_L$, and DEMUX$_r$ demultiplexes a light beam into $w \times 2^{\lceil (n-1)/2 \rceil} = 8$ wavelengths that are destined through other fibers to their corresponding destinations. Similarly, light beams emanating from S_c sources and belonging to PEs along the same column of $Plane_L$ or $Plane_R$ are multiplexed by MUX$_c$ and demultiplexed by DEMUX$_c$. For clarity of Fig. 9, only the implementation of the multichannel links for two rows and two columns is shown. Similar connections exist for the other rows and columns of $Plane_L$ of the SMLH(w = 4, n = 3) network. Each row-wise (or column-wise) multichannel link in Fig. 9 can be implemented using one pair of MUX-DEMUX. Fig. 10 shows a top view of both planes of the SMLH(w = 4, n = 3) network. In the middle of the figure, the HOEs that implement the hypercube modules are shown. Only two MUX-DEMUX pairs are shown. The top MUX-DEMUX pair implements the two

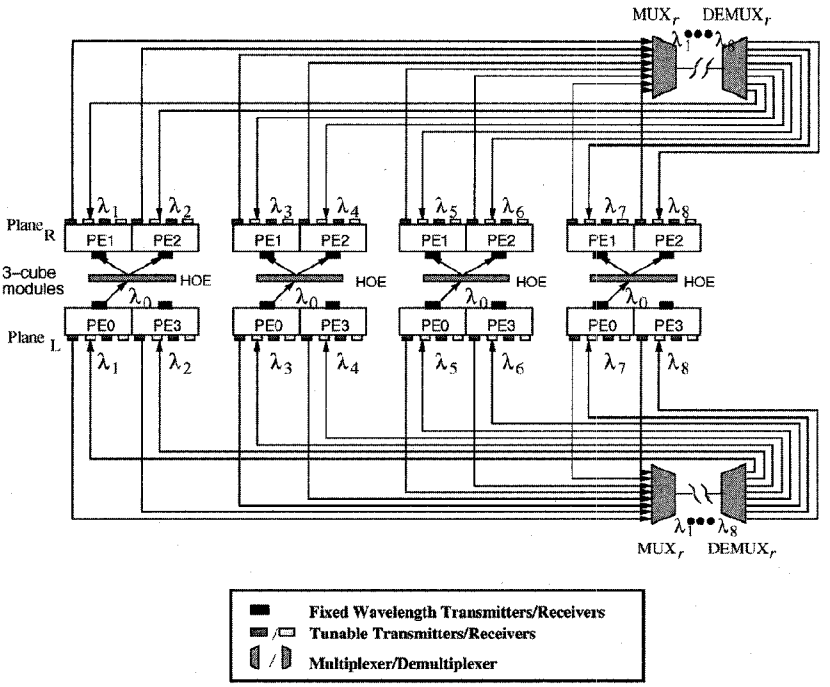


Fig. 10. Top view of an SMLH($w = 4$, $n = 3$) network. The figure shows the implementation of the first row-wise multichannel links of $Plane_L$ and $Plane_R$. Similar connections exist for the other rows and columns of the SMLH($w = 4$, $n = 3$).

row-wise multichannel links of the first row of $Plane_R$ while the bottom MUX-DEMUX pair implements the respective multichannel links of the first row of $Plane_L$.

Limitations on fan-out capabilities of MUXs and DEMUXs dictate the number of multichannel links and the number of PEs attached to a multichannel link using this implementation method. On the other hand, if one uses optical star couplers for implementing the multichannel links, the fan-out will no longer be a limitation, but power losses associated with star couplers become the limiting factor. These issues are analyzed in the next section.

In order to alleviate multichannel link collisions (e.g., different messages destined to the same PE at the same time), the time domain along each subchannel can be utilized. Time division multiple access techniques can be combined with the proposed WDMA scheme. This issue is beyond the scope of this paper and will not be treated any further.

5 Power and Delay Analysis of the Optical Implementations

5.1 Bit-Error Rate Estimation

In this section, we present some system noise calculations to investigate the bit-error rate (*BER*) capabilities and the delay of the proposed optical implementation of the SMLH network. Calculation of *BER* of an optical system requires estimation of the Signal-to-Noise ratio (*SNR*). Estimation of total power losses leading into the receiver sensitivity calculation is required for the *SNR* calculation. In what follows, the optical power loss of the implementation methodology is calculated. Then, the receiver sensitivity is estimated, and consequently the *BER* of the proposed implementation is evaluated.

The number of nodes that an optical system can support is determined by the emitting power of the transmitter, the required receiver sensitivity, and the losses incurred between the transmitter and the receiver. Let L_{MUX} be the total insertion loss of the WDMA multiplexers, and L_{DEMUX} be the total insertion loss of the WDMA demultiplexers. Let L_{sf} be the source-to-fiber coupling loss, and L_{fd} be the fiber-to-detector coupling loss. Assume that the fiber loss is L_f and that all nodes are equidistant, meaning that if we denote $d(i, i + 1)$ as the fiber distance from node i to node $i + 1$, then d is the same for all i's. The total transmission loss using WDMA techniques for the multichannel link implementation is then given by the following equation:

$$L_{total} = L_{MUX} + L_{DEMUX} + dL_f + L_{sf} + L_{fd} - 3. \quad (14)$$

To estimate the total loss of the optical system, values from commercially available components were considered. We assume laser diodes sources with characteristics of +7 dBm. A dBm is defined as 10 times the logarithmic ratio of the power with a reference power quantity of 1 mW. For example, for a value of 5 mW, the ratio in dBm is $10 \times \log\left(\frac{5mW}{1mW}\right) = +7$ dBm. Positive values in dBm indicate that the required power is greater than the reference power (1 mW) and negative values show that the required power is less than 1 mW.

The insertion loss for a commercially available fiber coupler is taken as –1 dB, while fiber to detector losses are –0.46 dB. The fiber loss is taken as 0.3 dB/Km, but since their lengths are in the order of centimeters, the total fiber loss is negligible. The insertion loss of the $k \times 1$ star coupler is taken as –2 dB [26], while, for the concave grating demultiplexer, the insertion loss is –9 dB. In addition, a –3 dB loss is being added for engineering errors. The total power loss is –15.5 dB, which is equivalent to an optical system efficiency of 2.8 percent. If

we consider the detectors to be GaAs Metal-Semiconductor FET Transimpedance [47] with a quantum efficiency of 60 percent at 670 nm, and if the data rate is 2 Gbps and if the required error probability after amplification is less than 10^{-17}, then the receiver sensitivity required is determined to be 12 μW or –19.2 dBm. Given the total power loss estimated above (–15.5 dB), and considering optical sources of 5 mW, the receiver sensitivity should be 140 μW or –8.5 dBm. This is well over the receiver sensitivity requirement for 10^{-17} *BER*. Therefore, the WDMA implementation proposed for the SMN subnetwork is feasible with *BER*s better than 10^{-17}. Depending on the optical components used, the size of the multichannel links is bounded either by the fan-out of the diffraction grating demultiplexers if the MUX-DEMUX scheme is used, or by the power loss of the star couplers if the optical star couplers are used. For the diffraction grating components, a fan-out of up to 40 channels has been reported [45], [46]. Taking this value as an upper limit, a fairly large SMLH network can be realized. For example, an SMLH(w = 20, n = 3) network consisting of N = 3,200 PEs and a total of 160 MUX-DEMUX pairs is feasible. Larger networks are also feasible by implementing every multichannel link in the network with a different MUX-DEMUX pair instead of utilizing a single pair to implement $2^{\lceil (n-1)/2 \rceil}$ multichannel links. This would increase the number of PEs in the network, but it would also increase the hardware complexity. For example, an SMHL(w = 40, n = 3) network consisting of N = 12,800 PEs and a total of 320 MUX-DEMUX pairs is feasible. Therefore, when designing the optical implementation of an SMHL network, trade-offs between the number of PEs desired and the hardware cost should be considered.

If a MUX-DEMUX pair were to be realized by a single optical star coupler, then (14) would have to include the excess (L_e) and splitting(L_{sp}) losses of the star coupler. Let P_{in} be the power entering the coupler from an input channel, and P_{out} be the power obtained from an output channel. Then, $L_{sp} = -10 \log \frac{P_{out}}{P_{in}}$, and the total transmission loss is

$$L_{total} = L_e + L_{sp} + dL_f + L_{sf} + L_{fd} - 3. \tag{15}$$

P_{out} is equal to P_{in}/k, where k is the fan-out of the star coupler. For the SMLH, k is $w \times 2^{\lceil (n-1)/2 \rceil}$ (number of PEs on a row or column of $Plane_L$ or $Plane_R$). Consequently, (15) can be rewritten as

$$L_{total} = L_e - 10log\ k + dL_f + L_{sf} + L_{fd} - 3. \tag{16}$$

The excess losses of the star coupler are taken as –1 dB [48]. Rearranging (16) and using the above losses, the number of PEs supported by the star coupler implementation method, given a desirable *BER*, can be determined. For a *BER* of 10^{-17} and for 5 mW sources, the total loss in the system should be –26.2 dB, yielding a star coupler fan-out of k = 118. This value is well within the capabilities of current star coupler technology. The optical fanout of star couplers reported to date is 128 × 128 [42]. For k = 118, large SMLH networks are feasible. For example, an SMHL(w = 30, n = 5) network supporting about 28,800 PEs could be realized.

5.2 Communication Delay

The data transfer delay in the SMLH network is a function of the transmitter, signal propagation, receiver delays, packet collisions, and channel bandwidth. The signal propagation delay includes the propagation time of the signal through the medium (i.e., fiber or free-space) and the delay when passing through optical components, such as star couplers and free space optics. In this paper, we estimate the communication delay without considering contention and collisions. Estimating the delay including contention and collisions would require the inclusion of access protocols and conflict resolution schemes which is beyond the scope of this paper.

Let τ_{tx} be the delay of the transmitter, τ_{rx} be the delay of the receiver, and τ_{prop} be the delay due to signal propagation. Therefore, the total delay of a single link can be expressed as

$$\tau_d = \tau_{tx} + \tau_{rx} + \tau_{prop} + W/B, \tag{17}$$

where W is the packet size in bits and B is the bandwidth of the communication medium. The value for the τ_{tx} delay includes driver turn on, laser turn on, and tuning delays, while the value for the τ_{rx} delay includes the photodetector circuit delay and skew. The photodetector circuit delay consists of the preamplifier delay, decision circuit delay, as well as the decision circuit skew. The propagation delay of an optical link (without any components) is

$$\tau_{prop} = \frac{L}{c/\eta}, \tag{18}$$

where L is the length of the link, and c/η is the speed of light in the medium. Hence, the total delay of an optical link is

$$\tau_d = \tau_{tx} + \tau_{rx} + \frac{L}{c/\eta} + \tau_m + W/B, \tag{19}$$

where τ_m is the propagation delay through optical components.

Delays for commercially available optical fixed transmitters and receivers are in the range of 1 ns and 1.8 ns, respectively [49], [50], while those of tunable transmitters, are 50 ns [51]. In SMLH, the optical star couplers and the grating demultiplexers are all passive components and, therefore, their time delays are very small and can be ignored in this estimation.

The worst case scenario for a collisionless SMLH network is when a packet travels through both multichannel links and through n hypercube links in order to reach its destination. In this case, the maximum collision free delay is:

$$\tau_d^{\max} = 2\tau_{tx}^{tune} + n\tau_{tx}^{fixed} + (n+2)\tau_{rx} + n\tau_{prop}^{BHC} + 2\tau_{prop}^{SMN} + W^{\max}/B^{\min}, \tag{20}$$

where τ_{tx}^{tune} is the delay of the tunable transmitters in the SMN subnetwork, τ_{tx}^{fixed} is the delay of the fixed transmitters in the hypercube subnetwork, τ_{rx} is the delay of the receivers, τ_{prop}^{BHC} is the propagation delay through the hypercube links, τ_{prop}^{SMN} is the propagation delay through the SMN links, W^{max}

is the maximum packet size in bits, and B^{min} is the minimum bandwidth. Clearly, the communication delay using WDMA techniques is dominated by the packet size more than by any other delays.

6 Conclusions

In this paper, we proposed a novel hybrid network which improves, or is comparable to, hypercube-based topologies in general and, in particular, the spanning bus hypercube (SBH) and the Optical Multi-Mesh Hypercube (OMMH). The key, attractive features of the proposed network include the possibility of a constant diameter and a constant degree network while it is feasible to interconnect thousands of processors at an adequate performance level. Additionally, the network is incrementally scalable and fault-tolerant. These features make SMLH very suitable for massively parallel systems. The topological characteristics of the proposed network were compared with several other well-known networks and it was shown that SMLH compares extremely well with the SBH, the OMMH, the Binary Hypercube, the Torus network, the Hierarchical Cubic Network, the Cube-Connected Cycle, the Hyper-deBruijn, and the Folded Peterson, and comparable performance with the Generalized Hypercube and the Hypermesh. A WDMA technique has been proposed for the optical implementation of the SMLH network. While the structural properties of the proposed network were fully analyzed in this paper, the optical implementation was presented for illustration purposes only. Detailed analysis of the optical feasibility of the proposed network including device and optical setups is under study and is subject of a subsequent paper.

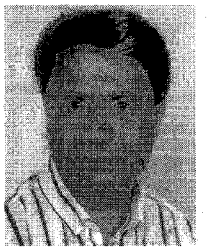
Acknowledgments

The authors would like to thank the anonymous referees for their constructive remarks. This work was supported by the U.S. National Science Foundation under Grant MIP 93-10082.

References

[1] K. Hwang, *Advanced Computer Architecture: Parallelism, Scalability, Programmability*. New York: McGraw-Hill, 1993.
[2] H.J. Siegel, *Interconnection Networks for Large-scale Parallel Processing*. New York: McGraw-Hill, 1990.
[3] H.S. Stone and J. Cocke, "Computer Architecture in the 1990s," *Computer*, vol. 24, no. 9, pp. 30–38, Sept. 1991.
[4] A. Louri and H. Sung, "3D Optical Interconnects for High-Speed Interchip and Interboard Communications," *Computer*, vol. 27, no. 10, pp. 27–37, Oct. 1994.
[5] L.M. Ni and P.K. McKinley, "A Survey of Wormhole Routing Techniques in Direct Networks," *Computer*, vol. 26, no. 4, pp. 62–76, Feb. 1993.
[6] S.A. Felperin, L. Gravano, G.D. Pifarre, and J.L.C. Sanz, "Fully-Adaptive Routing: Packet Switching Performance and Wormhole Algorithms," *Supercomputing '91 Proc.*, pp. 654–663, 1991.
[7] J. Kim and C.R. Das, "Hypercube Communication Delay with Wormhole Routing," *IEEE Trans. Computers*, vol. 43, no. 7, pp. 806–814, July 1994.
[8] Y. Saad and M.H. Schultz, "Topological Properties of Hypercubes," *IEEE Trans. Computers*, vol. 37, no. 7, pp. 867–872, July 1988.
[9] K. Ghose and K.R. Desai, "Hierarchical Cubic Networks," *IEEE Trans. Computers*, vol. 6, no. 4, pp. 427–435, Apr. 1995.
[10] J.M. Kumar and L.M. Patnaik, "Extended Hypercube: A Hierarchical Interconnection Network of Hypercubes," *IEEE Trans. Parallel and Distributed Systems*, vol. 3, no. 1, pp. 45–57, Jan. 1992.
[11] N.-F. Tzeng and S. Wei, "Enhanced Hypercubes," *IEEE Trans. Computers*, vol. 40, no. 3, pp. 284–294, Mar. 1991.
[12] L.N. Bhuyan and D.P. Agrawal, "Generalized Hypercube and Hyperbus Structures Constructing Massively Parallel Computers," *IEEE Trans. Computers*, vol. 33, pp. 323–333, 1984.
[13] Q.M. Malluhi and M.A. Bayoumi, "The Hierarchical Hypercube: A New Interconnection Topology for Massively Parallel Systems," *IEEE Trans. Parallel and Distributed Systems*, vol. 5, no. 1, pp. 17–30, Jan. 1994.
[14] C. Chen, D.P. Agrawal, and J.R. Burke, "dBCube: A New Class of Hierarchical Multiprocessor Interconnection Networks with Area Efficient Layout," *IEEE Trans. Parallel and Distributed Systems*, vol. 4, no. 1, pp. 1,332–1,344, Jan. 1993.
[15] K. Efe, "The Crossed Cube Architecture for Parallel Computation," *IEEE Trans. Parallel and Distributed Systems*, vol. 3, no. 9, pp. 513–524, Sept. 1992.
[16] J.R. Goodman and C.H. Sequin, "Hypertree: A Multiprocessor Interconnection Topology," *IEEE Trans. Computers*, vol. 30, pp. 923–933, 1981.
[17] A. Louri and H. Sung, "A Scalable Optical Hypercube-Based Interconnection Network for Massively Parallel Computing," *Applied Optics*, vol. 33, pp. 7,588–7,598, Nov. 1994.
[18] A. Louri and H. Sung, "An Optical Multi-Mesh Hypercube: A Scalable Optical Interconnection Network for Massively Parallel Computing," *IEEE J. Lightwave Technology*, vol. 12, pp. 704–716, Apr. 1994.
[19] A. Louri and S. Furlonge, "Feasibility Study of a Scalable Optical Interconnection Network for Massively Parallel Processing Systems," *Applied Optics*, vol. 35, pp. 1,296–1,308, Mar. 1996.
[20] A. Louri, S. Furlonge, and C. Neocleous, "Experimental Demonstration of the Optical Multi-Mesh Hypercube: A Scalable Interconnection Network for Multiprocessors and Multicomputers.," *Applied Optics*, vol. 35, no. 35, pp. 6,909–6,920, Dec. 1996.
[21] L.D. Wittie, "Communication Structures for Large Networks of Microcomputers," *IEEE Trans. Computers*, vol. 30, no. 4, pp. 264–273, Apr. 1981.
[22] G. Lerman and L. Rudolph, *Parallel Evolution of Parallel Processors*. New York: Plenum Press, 1993.
[23] M.R. Feldman, C.C. Guest, T.J. Drabik, and S.C. Esner, "Comparison Between Electrical and Free Space Optical Interconnects for Fine Grain Processor Arrays Based on Connection Density Capabilities," *Applied Optics*, vol. 28, pp. 3,820–3,829, 1989.
[24] J.W. Goodman, F.J. Leonberger, S.Y. Kung, and R.A. Athale, "Optical Interconnections for VLSI systems," *Proc. IEEE*, vol. 72, pp. 850–866, July 1984.
[25] D.M. Chiarulli, S.P. Levitan, R.G. Mehlen, M. Bidnurkar, R. Ditmore, G. Gravenstreter, Z. Guo, C. Qiao, M.F. Sakr, and J.P. Teza, "Optoelectronic Buses for High-Performance Computing," *Proc. IEEE*, vol. 82, no. 11, pp. 1,701–1,710, Nov. 1994.
[26] P.W. Dowd, "Wavelength Division Multiple Access Channel Hypercube Processor Interconnection," *IEEE Trans. Computers*, vol. 41, no. 10, pp. 1,223–1,241, Oct. 1992.
[27] Y. Li, A.W. Lohmann, and S.B. Rao, "Free-Space Optical Mesh-connected Bus Networks Using Wavelength-Division Multiple Access," *Applied Optics*, vol. 32, pp. 6,425–6,437, 1993.
[28] P.B. Berra, A. Ghafoor, M. Guizani, S.J. Marcinkowski, and P.A. Mitkas, "Optics and Supercomputing," *Proc. IEEE*, vol. 77, pp. 1,797–1,815, Dec. 1989.
[29] D.S. Miller, "Optics for Low Energy Communication Inside Digital Processors: Quantum Detectors, Sources, and Modulators as Efficient Impedance Converters," *Optics Letters*, vol. 14, pp. 146–148, 1989.
[30] A.A. Sawchuk, C.S. Raghavandra, B.K. Jenkins, and A. Varma, "Optical Crossbar Networks," *Computer*, vol. 20, no. 6, pp. 50–62, June 1987.
[31] D.P. Agrawl, C. Chen, and J.R. Burke, "A Comparison of Hybrid Graph Based Topologies," *IEEE Computer Society Techinal Committee on Computer Architecture Newsletter*, E.F. Gehringer, ed., pp. 5–8, Winter 1994-95.
[32] V. Kumar, A. Grama, A. Gupta, and G. Karypis, *Introduction to Parallel Computing: Design and Analysis of Algorithms*. New York: Benjamin/Cummings Publishing Co., 1994.

[33] T. Szymanski, "Hypermeshes: Optical Interconnection Networks for Parallel Computing," *J. Parallel and Distribted Computing*, vol. 26, pp. 1–23, 1995.
[34] D.A. Reed and R.M. Fujimoto, *Multicomputer Networks: Message-Based Parallel Processing*. Cambridge, Mass.: MIT Press, 1987.
[35] F.P Preparata and J. Vuillemin, "The Cube-Connected Cycles: A Versatile Network Parallel Computation," *Comm. ACM*, vol. 24, pp. 300–309, May 1981.
[36] E. Ganesan and D.K. Pradhan, "The Hyper-deBruijn Networks: Scalable Versatile Architecture," *IEEE Trans. Parallel and Distributed Systems*, vol. 4, no. 9, pp. 962–978, Sept. 1993.
[37] S. Ohring and S.K. Das, "Folded Petersen Cube Networks: New Competitors for the Hypercubes," *IEEE Trans. Parallel and Distributed Systems*, vol. 7, no. 2, pp. 151–168, Feb. 1996.
[38] L. Klienrock, *Queueing Systems: Vol. II, Computer Applications*. New York: Wiley, 1976.
[39] P. Sweazey, "Limits of Performance of Backplane Buses," *Digital Bus Handbook*. New York: McGraw-Hill, 1990.
[40] A.D. McAulay, *Optical Computer Architectures: The Application of Optical Concepts to Next Generation Computers*. New York: Wiley, 1991.
[41] A. Louri and H. Sung, "Efficient Implementation Methodology for Three-Dimensional Space-Invariant Hypercube-Based Free-Space Optical Interconnection Networks," *Applied Optics*, vol. 32, pp. 7,200–7,209, Dec. 1993.
[42] A. Cisneros and C.A. Brackett, "A Large ATM Switch Based on Memory Switches and Optical Star Couplers," *IEEE J. Selected Areas Comm.*, vol. 9, pp. 1,348–1,360, Oct. 1991.
[43] P.F. Moulton, "Tunable Solid State Lasers," *Proc. IEEE*, vol. 80, pp. 348–364, Mar. 1992.
[44] B.D. Metcalf and J.F. Providakes, "High-Capacity Wavelength Demultiplexer with a Large-Diameter GRIN Rod Lens," *Applied Optics*, vol. 21, pp. 794–796, 1982.
[45] F.N Tinofeev, P. Bayvel, J.E. Midwinter, and M.N. Sokolskii, "High-Performance, Free-Space Ruled Concave Grating Demultiplexer," *Electronics Letters*, vol. 31, pp. 2,200–2,201, Dec. 1995.
[46] F.N Tinofeev, P. Bayvel, J.E. Midwinter, M.N. Sokolskii, E.G Churin, and A. Stavdas, "Free-Space Aberration-Corrected Grating Demultiplexer for Application in Densely-Spaced, Subnanometer Wavelength-Routed Optical Networks.," *Electronics Letters*, vol. 31, pp. 1,368–1,370, Aug. 1995.
[47] T.V. Moui, "Receiver Design for High-Speed Optical Fiber Systems," *IEEE J. Lightwave Technology*, pp. 243–267, 1984.
[48] D. Israel, R. Baets, M.J. Goodwin, N. Shaw, M.D. Salik, and C.J. Groves-Kirkby, "Comparison of Different Polymetric Multimode Star Couplers for Backplane Optical Interconnect," *IEEE J. Lightwave Technology*, vol. 13, pp. 1,057–1,064, June 1995.
[49] R.A. Morgan, "Advances in Vertical Cavity Surface Emitting Lasers," *Proc. SPIE*, vol. 2,147, pp. 97–119, 1994.
[50] A. Takai, T. Kato, S. Yamashita, S. Hanatani, Y. Motegi, K. Ito, H. Abe, and H. Kodera, "200–*Mb/s/ch* 100m Optical Subsystem Interconnections Using 8-channel 1.3 μ Laser Diode Arrays and Single-Mode Fiber Arrays," *IEEE J. Lightwave Technology*, vol. 12, pp. 260–269, Feb 1994.
[51] N.K. Shankaranarayanan, U. Koren, B. Glance, and G. Wright, "Two-Section dBr Laser Transmitters with Accurate Channel Spacing and Fast Arbitrary-Sequence Tuning for FDMA Networks," *Technical Digest of Fiber Optical Comm.*, pp. 36–37. Washington, DC: Optical Soc. of America, 1994.



Ahmed Louri received the PhD degree in computer engineering in 1988, the MS degree in computer engineering in 1984, both from the University of Southern California, and the Diplome D'Engenieur (Engineer Degree) in electrical engineering from the University of Science and Technology, Oran, Algeria, in 1982. He is currently an associate professor of electrical and computer engineering at the University of Arizona and director of the Optical Networking and Parallel Processing Laboratory. His research interests include computer architecture, parallel processing, optical computing, and optical interconnects. Dr. Louri has published numerous journal and conference articles on the above topics. In 1991, he received the "Best Article of 1991 Award" from *IEEE Micro*. In 1988, he was the recipient of the U.S. National Science Foundation Research Initial Award. In 1994, he was the recipient of the Advanced Telecommunications Organization of Japan Fellowship, Ministry of Post and Telecommunications, Japan. In 1995, he was the recipient of the Centre Nationale de Recherche Scientifique (CNRS), France fellowship. In 1996, he was the recipient of the Japanese Society for the Promotion of Science fellowship. Prior to joining the University of Arizona, he worked as a researcher with the Computer Research Institute at the University of Southern California, where he conducted extensive research in parallel processing, multiprocessor system design, and optical computing. He has served as a member of the Technical Program Committee of several conferences, including OSA Topical Meeting on Optics in Computing, OSA/IEEE Conference on Massively Parallel Processors Using Optical Interconnects, among others. In 1996, he was the general chair of the Workshop on Optics in High-Performance Computing at Euro-Par '96, Lyon, France.

Dr. Louri is a senior member of the IEEE and a member of the ACM, OSA, and SPIE.

Brent Weech is an MS candidate at the University of Arizona. He received the BS degree in electrical engineering from Arizona State University in 1994. His research interests include parallel and distributed processing, communication networks, and optical interconnects. He is a student member of the IEEE.

Costas Neocleous received the MS degree in electrical and computer engineering in 1996 and the BS degree in electrical engineering in 1994, both from the University of Arizona. He is a student member of the IEEE.