Performance of Multi–Hop Communications Using Logical Topologies on Optical Torus Networks *†

Xin Yuan
Dept. of Computer Science
Florida State University
Tallahassee, FL 32306
xyuan@cs.fsu.edu

Rami Melhem
Dept. of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
melhem@cs.pitt.edu

Rajiv Gupta
Dept. of Computer Science
University of Arizona
Tucson, AZ 85721
gupta@cs.arizona.edu

Abstract

We consider multi–hop communication in optical networks with time–division multiplexing. By using time–division multiplexing, multiple communication channels are supported on each link. As a result, more sophisticated logical topologies can be realized on top of a simpler physical network to improve the communication performance. These logical topologies reduce the number of intermediate hops that a packet travels at the cost of a larger multiplexing degree. On the one hand, the large multiplexing degree increases the packet communication time between hops. On the other hand, reducing the number of intermediate hops reduces the time spent at intermediate nodes. We study the trade–off between the multiplexing degree and the number of intermediate hops needed to realize the logical topologies on top of physical torus networks. Specifically, we examine four logical topologies ranging from the most complex logical all–to–all connections to the simplest logical torus topology. We develop an analytical model that models the maximum throughput and the average packet delay, verify the analytical model through simulations, and study the performance and the impact of system parameters on the performance for these logical topologies.

Keywords: Optical Network, Multi–hop Communication, Time–division Multiplexing, Performance Evaluation, Single–hop Communication

*A preliminary version of this paper appears in the proceedings of the Seventh International Conference on Computer and Communication Networks (IC3N'98), 1998.

†This work was supported in part by NSF award CCR-9157371, CCR-9704350 and MIP-9633729.

1 Introduction

With the increasing computation power of parallel computers, interprocessor communication has become an important factor that limits the performance of supercomputing systems. Optical interconnection networks are promising networks for future supercomputers due to their large bandwidths. Multiplexing techniques are used in optical networks to fully utilize the large bandwidths. Many research efforts have focused on two multiplexing techniques, *time–division multiplexing* (TDM) [1, 15, 21] and *wavelength–division multiplexing* (WDM) [4, 6, 9, 20]. In TDM, optical links are multiplexed by having different virtual channels communicating in different *time slots*, while in WDM, optical links are multiplexed by having different virtual channels using different *wavelengths*. By using TDM, WDM or TWDM (a combination of TDM and WDM), each link can support multiple channels with each channel operating at a speed close to the electronic processing speed.

Two types of communication mechanisms are used in multiplexed optical networks, *single–hop* communication [13] and *multi–hop* communication [14]. In single–hop communication, each data message travels from the source to its destination completely in the optical domain, that is, without optical to electronic (O/E) and electronic to optical (E/O) conversions at intermediate nodes. While single–hop communication eliminates the E/O and O/E conversions at intermediate nodes, it requires significant amount of dynamic coordination among the nodes in the network. In multi–hop communication, intermediate nodes are responsible for routing packets such that a packet sent from a sender will eventually reach its destination, possibly after being routed through a number of intermediate nodes. Clearly, multi–hop networks require E/O and O/E conversions at intermediate nodes. Since the electronic processing speed is relatively slow in comparison to the optical data transmission speed, it is important to reduce the number of hops that a packet visits to obtain high communication performance in multi–hop networks. In a multiplexed optical network, reducing the number of intermediate hops can be achieved by routing packets through a more efficient logical topology, as opposed to routing packets through the physical topology.

While shared media optical networks such as buses and stars were proposed as interconnects for parallel computers [4], point–to–point networks, such as meshes, tori and hypercubes, can offer larger aggregate throughput and better scalability than shared media networks by exploiting space diversity and traffic locality and by utilizing the *channel–routing* capability in optical switches. In this paper, we study multi–hop communication on top of physical optical TDM torus networks using different logical topologies in the multi–processor environment. We chose the torus topology as our underlying physical topology because it has nice characteristics, such as a fixed number of ports in each node, good scalability, and because it is currently used by many commercial supercomputers. Different logical topologies have different logical connectivity and require different multiplexing degree. On the one hand, each packet travels less number of intermediate hops in logical topologies with higher connectivity, which reduces the time that the packet spends at intermediate hops. On the other hand, logical topologies with higher connectivity require larger multiplexing degree, which results in smaller bandwidth in each channel and increases the packet communication time between hops. Thus, the trade–off between the multiplexing degree and the number of intermediate hops needed must be carefully studied to design efficient logical topologies for optical TDM networks.

Four logical topologies are considered in this paper. The first logical topology is the logical all–to–all connections. This logical topology represents an extreme case where the number of intermediate hops is traded in favor of the multiplexing degree. This topology will be called the logical *all–to–all* topology. The second topology is the logical *torus*, which has the same topology as the physical network. This represents another extreme case where the multiplexing degree is traded in favor of the number of intermediate hops. The other two logical topologies are in between these two extremes. The third topology is a 1-hop system formed by having all–to–all connections along each dimension of the physical torus. Thus, a packet passes at most 1 intermediate hop to reach its destination. We will call this topology the *allXY* topology. The fourth topology is the logical *hypercube* topology. We will discuss these topologies in detail in Section 4.

We develop an analytical model that models the maximum throughput and the average packet delay for the four topologies. We verify the analytical model with simulations and study the impact of system parameters, such as the packet routing time, on the performance of these topologies. Through the study of the four representative logical topologies, we obtain the general performance trends in multi–hop communication using logical topologies on optical TDM networks.

The rest of the paper is organized as follows. Section 2 presents the related work. In Section 3, we describe system assumptions. In Section 4, we describe how the logical topologies are realized on top of physical torus networks. In Section 5, we present the analytical model and verify the model with simulations. Section 6 studies the performance of the logical topologies. Section 7 concludes the paper.

2 Related work

The study of logical topology design for optical networks has mainly focused on WDM networks [3, 10, 11, 12, 19]. A survey can be found in [5]. Most of the work investigates the logical topology problem in the Wide–Area–Network (WAN) environment [3, 10, 12, 19]. Other work considers the problem on broadcast-based optical networks [11]. This paper considers the logical topology problem for channel–routed TDM networks. The logical topology problem in optical TDM networks is similar to that in WDM networks in that the solutions to both problems make the best use of the virtual channels in the system. Thus, many techniques can apply to both problems. However, the difference between the logical topology problem in TDM and WDM networks lies in the way that virtual channels are created using the two multiplexing techniques. In WDM, a new channel is created by adding a new wavelength into the system. Hence, larger number of virtual channels means higher bandwidth in the links, better communication performance and higher system cost. The logical topology design problem in WDM networks focuses on achieving the best performance for a given cost (the number of channels). In TDM, new channels are created by time–multiplexing a communication link. Since the bandwidth on a given link is fixed, TDM creates more channels with each channel having less bandwidth. Hence, in a TDM system, larger number of virtual channels does not always mean better communication performance since the overall bandwidth on a link is fixed. The logical topology design involves carefully choosing the multiplexing degree to realize a logical topology that achieves the best performance. Thus, the optimization objective for designing logical topologies for WDM networks and TDM networks is different. Although the WDM technique is currently very popular, we choose to study optical TDM networks because in the future, each wavelength will be able to support more bandwidth than an electronic device can process and TDM or TWDM will be a natural method to utilize the large bandwidth in optical links.

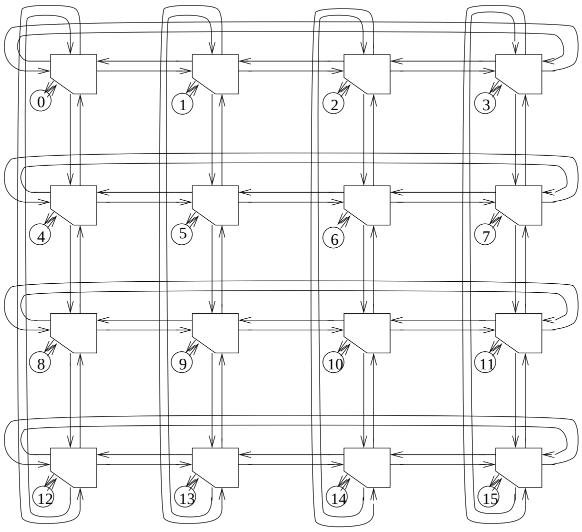


Figure 1: A torus network

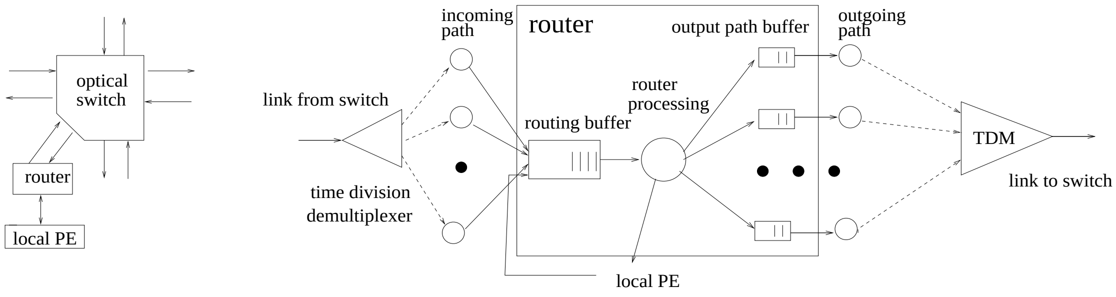
Much work has been done in the areas of routing and channel assignment [18] and connection scheduling [7, 16, 23]. Some of the techniques [7, 23] are adopted in this paper. However, our work focuses on comparing the performance of different logical topologies while routing and channel assignment finds the most efficient way to realize a traffic pattern (logical topology). Notice that although our work studies optical TDM networks in the multi-processor environment, most of the related work are in the areas of WDM networks in the WAN environment. This is due to the similarity between TDM and WDM networks. Although, many optimization techniques can apply to both TDM and WDM, the two multiplexing paradigms are sufficiently different (as discussed in the previous paragraph) to necessitate separate studies.

3 System assumptions

As shown in Figure 1, a torus network consists of switches with a fixed number of input and output ports. All but one input port and one output port are used to interconnect with other switches, while one input port and one output port are used to connect to a local processing element (PE). Each link in the network is multiplexed to support multiple virtual channels. The number of virtual channels supported by each link is called the *multiplexing degree*. We assume that there are no channel (time-slot) converters in the network, thus the same channels on all links along a path must be used to establish a connection. We will use a *path* to refers to a lightpath that may span a number of optical switches and links without O/E and E/O conversions. Each connection in a logical topology is a path. A packet may cross a number of paths to reach its destination. We assume that in each time slot, a packet can be transferred over a path. For example, if a 1Gbps channel is used with a 53–byte packet (or cell) as defined in the ATM standard, then the slot duration is $0.424\mu s$. All other delays in the system are normalized with respect to this slot duration. Note that we consider a multiprocessor environment, where nodes are close to each other and the signal propagation delay is negligible (on the order of one nanosecond per foot).

The nodal switching architecture consists of an optical component and an electronic component, as shown in Figure 2 (a). The optical component is an all–optical switch (e.g. a Ti:LiNbO$_3$ switch [8]), which can switch optical signals from input ports to output ports without E/O and O/E conversions, and which can locally terminate lightpaths by directing them to the node's electronic component. The electronic component is a store–and–forward packet router overlaid on top of the optical virtual topology. The router can be implemented either in system software or in hardware. We assume that each router contains a *routing buffer* that buffers all incoming packets. For each packet, the router determines whether to deliver the packet to the local PE or to the next path toward the packet destination. A separate *output path buffer* is used for each outgoing path that buffers the packets to be sent on that path and thus accommodates the speed mismatch between the electronic router and the optical path. Figure 2 (b) depicts the structure of a router. Note that the output paths are multiplexed in time over the physical link that connects the local PE to its corresponding switch.

In the rest of the paper, we will use *routing delay* to denote the time a packet spends in routing buffers and the time for routers to make routing decisions for the packet (packet routing time). We will use *transmission delay* to denote the time a packet spends on path buffers and the time it takes for the packet to be transferred on paths.



(a) a nodal switching architecture (b) the router

Figure 2: Nodal switching architecture

4 Logical topologies

Given a physical optical torus network with TDM, logical topologies can be realized on top of the physical network by exploiting both the channel–routing capability and the electronic switching capability. The performance of the logical topologies depends on many factors including the traffic pattern, the number of multiplexing degree (d) needed to realize the topology, the number of intermediate hops (h) for a packet to reach the destination and the electronic switching speed (γ). This section summarizes the property of the logical topologies and describes how the logical topologies can be realized on top of a physical $N \times N$ torus network.

The logical all–to–all topology establishes direct connections between all pairs of nodes and thus, totally eliminates the intermediate hops, resulting in $h = 0$. The algorithm in [7] schedules the all–to–all connections in phases such that within each phase, all links in the network are used and that each node can send or receive at most once within each phase. For an $N \times N$ torus, where $N \geq 8$ and $N = 2^r$, a total of $\frac{N^3}{2}$ links are needed to realize a broadcast from one node to all other nodes. A total of $\frac{N^3}{2} \times N$ links are needed to realize the all–to–all communication. Since each phase has $4N$ links, $\frac{N^3}{8}$ phases are needed to realize the all–to–all communication. Since each phase can be realized by one channel, using the algorithm in [7], a multiplexing degree of $d = \frac{N^3}{8}$ can be used to realize the logical all–to–all topology.

A logical torus topology can be realized using a multiplexing degree of $d = 4$. Notice that although the logical paths are the same as the physical links, the logical torus topology cannot be realized using a multiplexing degree of 1 due to the contention for the links connecting local PEs to switches. Specifically, in one time slot each router can only access one channel, and since each router has four outgoing logical paths, one to each neighbor, a multiplexing degree of 4 is needed to realize this topology. For a logical $N \times N$ topology, the average number of intermediate hops is $h = \frac{N}{2} - 1$.

For $N = 2^r$, the algorithm in [23] can be used to realize a logical hypercube topology on an $N \times N$ torus using a multiplexing degree of $\lfloor \frac{N}{3} + \frac{N}{4} \rfloor + 2$, if r is odd, and $\lfloor \frac{N}{3} + \frac{N}{4} \rfloor + 1$, if r is even. For a logical N^2 node hypercube, the average number of intermediate hops is $h = \frac{lg(N^2)}{2} - 1 = lg(N) - 1$.

Finally, let us consider the logical *allXY* topology. By using the 1-dimension communication patterns for all–to–all connections on rings from [7] and mixing the 1-dimensional communication patterns to form 2-dimensional patterns for the allXY topology, it can be shown that the logical topology can be realized using a multiplexing degree of $2N - 2$ if $N \leq 8$, and $\frac{N^2}{8}$ if $N > 8$. In this topology, no intermediate hop is needed to route packets between two nodes in the same column or the same row. When the source and the destination are not in the same column or the same row, one intermediate hop is needed. Hence, the average number of intermediate hops is given by:

$$\frac{2N-2}{N^2-1} \times 0 + \frac{(N^2-1)-(2N-2)}{N^2-1} \times 1 = \frac{N^2-2N+1}{N^2-1}.$$

Logical topology	Number of intermediate hops (h)	multiplexing degree (d)	total number of connections (P)
all–to–all	0	$\frac{N^3}{8}$†	$N^2(N^2-1)$
allXY	$\frac{N^2-2N+1}{N^2-1}$	$\frac{N^2}{8}$ ‡	$N^2(2N-2)$
hypercube	$lg(N)-1$	$\lfloor \frac{N}{3} + \frac{N}{4} \rfloor + 1$ ⋆	$2N^2 lg(N)$
torus	$\frac{N}{2}-1$	4	$N^2 \times 4$

† Here, we assume that $N \geq 8$ and $N = 2^r$.
‡ Here,we assume that $N \geq 8$. If $N < 8$, the value is $2N - 2$.
⋆ Here, we assume that $N = 2^r$ and r is even. If r is odd, the value is $\lfloor \frac{N}{3} + \frac{N}{4} \rfloor + 2$.

Table 1: Summary of the logical topologies

Table 1 summarizes the average number of intermediate hops (h), the multiplexing degree (d) and the total number of logical connections (P) for the four topologies. As can be seen from the table, as the number of (logical) connections decreases in the logical topologies, the number of intermediate hops increases and the multiplexing degree needed to realize the topologies decreases. A study of the performance of these four topologies will give us a good indication about the trade–offs between the multiplexing degree and the number of intermediate hops when using multi–hop communication.

5 An analytical model and its verification

In this section, we develop an analytical model that models the maximum throughput and the average packet delay for the logical topologies and verify the model through simulations. We model the routers and the paths in a network as a network of queues. As shown in Figure 2, each router has a routing queue that buffers the packets to be processed. The router places packets either into one of the output path queues that buffer packets waiting to be transmitted, or into the local processor. The exact model for such network is very difficult to obtain. We approximate the analysis by making the following assumptions: 1) the queues are independent of each other, and 2) each queue has a Poisson arrival and constant service time. These assumptions enable us to derive expressions for the maximum throughput and the average packet delay of the four logical topologies by dealing with M/D/1 queues independently. Our simulation results confirm that these approximations are reasonable. We use the following notation in the model:

- N. Size of each dimension of the torus. The network has a total of N^2 nodes.
- d, h and P are defined in the previous section. A *frame* is defined to consist of d time slots. Within a frame, one time slot is allocated to each path. As discussed earlier, the average number of paths that a packet traverses is equal to $h+1$. The average number of routers that a packet traverses is $h+2$.
- λ. Average packet generation rate at each node per time slot. This implies that the average generation rate of packets to the entire network is $N^2\lambda$. We assume that the arrival process is Poisson and is independently and identically distributed on all nodes. Furthermore, we assume that all packets are equally likely to be destined to any one of the nodes. At each router, the newly generated packets and the packets arriving from other nodes are maintained in a routing buffer of infinite size.
- λ_s. Average packet arrival rate at a router per time slot, including both newly generated packets and packets received from other nodes. This composite arrival rate, λ_s, can be derived as follows. In any time slot the total number of newly generated packets that arrive at all the routing buffers is λN^2. On average, each of these packets traverses $h+2$ routers within the network. Therefore, under steady state condition, there will be $\lambda N^2(h+2)$ packets at all routers in the network in each time slot. Under the assumption that each packet is equally likely to be at each router, the final arrival rate is given by $\lambda_s = \lambda(h+2)$.
- λ_p. Average packet arrival rate at a path buffer per time slot. This arrival rate, λ_p, can be derived as follows. Under steady state condition, in any time slot, the total number of packets at all routers in the network is $\lambda N^2(h+2)$. Of all these packets, λN^2 packet will exit the network and $\lambda N^2(h+2) - \lambda N^2 = \lambda N^2(h+1)$ packets will be transmitted through paths in the network. Since there are P paths in the system, the average arrival rate is given by $\lambda_p = \frac{\lambda N^2(h+1)}{P}$.
- γ. The routing time per packet at a router. Since packets are of the same length, the routing time is a constant value. The average packet departure rate from the routing buffer, denoted by μ_s, is $\mu_s = \frac{1}{\gamma}$.

- μ_p. The average packet departure rate from each path buffer per time slot. Since in our model, each path will be served once in every frame, $\mu_p = \frac{1}{d}$. The average service time for each path buffer is $S_p = \frac{1}{\mu_p} = d$.

Maximum throughput

With the above notation, we can now study the performance of the logical topologies. We will first study the maximum throughput and then the average packet delay. Two bottlenecks can potentially limit the maximum throughput.

- A router may not have enough bandwidth to process all arriving packets. The maximum packet generation rate allowed by the router bandwidth, λ_s^{max}, is achieved when $\lambda_s = \mu_s$. That is $(h+2)\lambda = \frac{1}{\gamma}$, or $\lambda = \frac{1}{\gamma(h+2)}$. Thus,

$$\lambda_s^{max} = \frac{1}{\gamma(h+2)} \tag{1}$$

- A path may not have enough bandwidth to process all arriving packets. The maximum packet generation rate allowed by the path bandwidth, λ_p^{max}, is achieved when $\lambda_p = \mu_p$. That is $\frac{(h+1)\lambda N^2}{P} = \frac{1}{d}$, or $\lambda = \frac{P}{(h+1)N^2 d}$. Thus,

$$\lambda_p^{max} = \frac{P}{(h+1)N^2 d} \tag{2}$$

The maximum packet generation rate allowed by the network is the minimum of λ_s^{max} and λ_p^{max}, that is, $\lambda^{max} = min(\lambda_s^{max}, \lambda_p^{max})$. The maximum throughput of the network is thus, $N^2\lambda^{max}$. For a given topology, $\lambda^{max} = \lambda_s^{max}$ indicates that the router speed is the bottleneck, while $\lambda^{max} = \lambda_p^{max}$ indicates that the path speed is the bottleneck.

Average packet delay

As mentioned in Section 2, we divide the packet delay into (1) the *routing delay*, which includes the time a packet spends in routing buffers and the time for routers to process the packets, and (2) the *transmission delay*, which includes the time a packet spends in path buffers and the actual packet transmission time on the paths.

Let us first consider the routing delay at each router. It takes γ timeslots for a router to process a packet when the packet reaches the front of the routing buffer. As for the packet waiting time in the routing buffer, since we model the routing buffer as an $M/D/1$ queue, the average queuing delay depends on the arrival rate λ_s and is given by:

$$Q = \frac{\lambda_s(\gamma)^2}{2(1 - \frac{\lambda_s}{\mu_s})} \tag{3}$$

where λ_s is the average packet arrival rate, γ is the expected service time, μ_s is the average packet departure rate. Given that $\mu_s = \frac{1}{\gamma}$, the total time that a packet spends in each router is given by:

$$routing\ delay = \gamma + \frac{\lambda_s(\gamma)^2}{2(1 - \lambda_s\gamma)} \tag{4}$$

Now, let us consider the two components of the transmission delay on each path. The first component is the delay required by a packet to synchronize with the appropriate outgoing slot in the frame on which the node transmits and the actual packet transmission time. The average value of this delay is $\frac{1+2+...+d}{d} = \frac{d+1}{2}$. The second component is the $M/D/1$ queuing delay that a packet experiences at the buffer before it reaches the head of the buffer. This follows the same formula as in the routing delay case, and is given by,

$$\frac{\lambda_p S_p^2}{2(1-\frac{\lambda_p}{\mu_p})} = \frac{\lambda_p d^2}{2(1-\lambda_p d)} \tag{5}$$

Combining the two components, we obtain the total delay a packet encounters on a path,

$$transmission\ delay = \frac{d+1}{2} + \frac{\lambda_p d^2}{2(1-\lambda_p d)} \tag{6}$$

As discussed earlier, each packet passes $h+2$ hops and $h+1$ paths on average. Thus, given that on average, a packet spends *routing delay* in each router and *transmission delay* on each path, the average packet delay can be expressed as follows:

$$delay = (h+2) \times routing\ delay + (h+1) \times transmission\ delay \tag{7}$$

Using formula (4) and (6), we obtain the following average delay that a packet encounters from the source to the destination.

$$delay = (h+2) \times (\gamma + \frac{\lambda_s(\gamma)^2}{2(1-\lambda_s\gamma)}) + (h+1) \times (\frac{d+1}{2} + \frac{\lambda_p d^2}{2(1-\lambda_p d)}) \tag{8}$$

Model verification

To verify our analytical model and to further study the performance of these logical topologies, we developed a network simulator that simulates all four logical topologies on top of physical torus networks. The simulation results are obtained with 98% confidence level and 1% confidence interval. The simulator takes the following parameters.

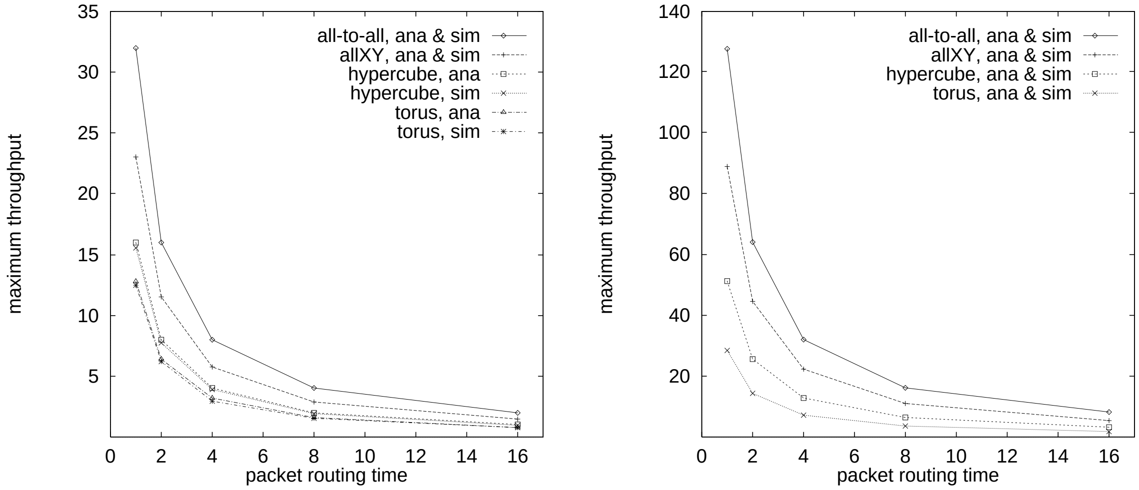
- *System size*, $N \times N$. This specifies the size of the network. For a given logical topology, the system size also determines the multiplexing degree in the system.

- *Packet generation rate*, λ. This is the average rate at which fresh packets are generated in each node. The inter–arrival of packets follows a Poisson distribution. When a packet is generated at a node, the destination is generated randomly among all other nodes in the system with a uniform distribution.

- *Packet routing time*, γ.

Figure 3 shows the maximum throughput obtained from the analytical model and from the simulation. We examine physical 8×8 and 16×16 torus networks with different packet routing times. As can be seen from the figure, the analytical results and the simulation results almost have a perfect match for all cases.

Figure 4 and Figure 5 show the average packet delay obtained from the analytical model and from the simulation. Here, the packet routing time, γ, is equal to 1 time slot. For 8×8

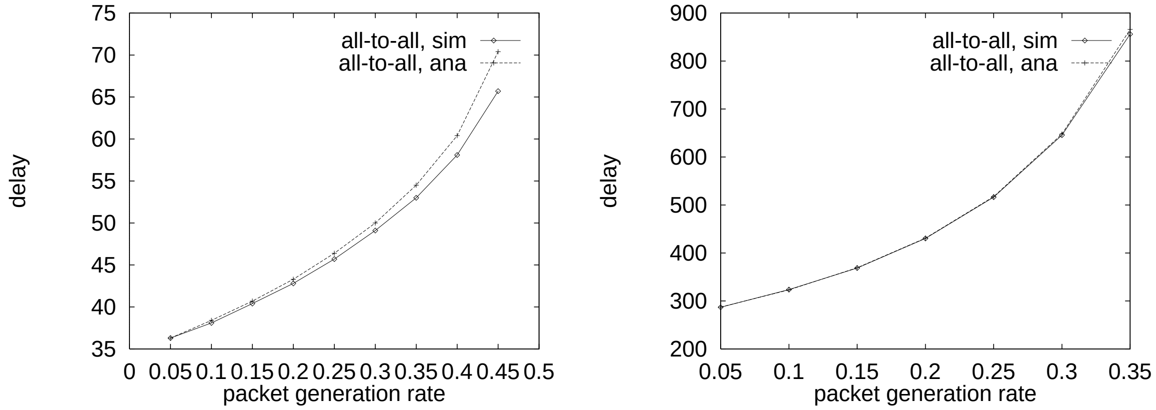


(a) physical 8×8 torus

(b) physical 16×16 torus

Figure 3: Maximum throughput from the analytical model and from the simulator

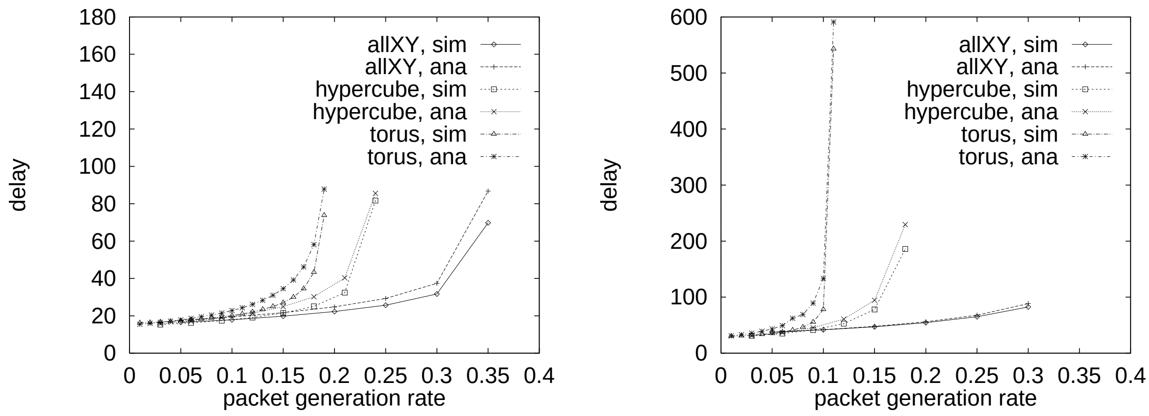
torus, the analytical results matches the simulation results fairly well for all topologies except when the network load is close to saturation. The difference between the results from the analytical model and those from simulations is around 10%. For the 16×16 physical topology, the analytical results match the simulations results for the all–to–all, allXY and hypercube topologies. For the torus topology, the difference is about 20% due to the approximation. We also conducted studies for other values of γ. The analytical results and the simulation results on those studies match slightly better than the ones shown in Figures 4 and 5. Thus, overall the analytical model gives a good indication of the actual performance.



(a) physical 8×8 torus

(b) physical 16×16 torus

Figure 4: The average packet delay for the logical all–to–all topology ($\gamma = 1$)



(a) physical 8×8 torus

(b) physical 16×16 torus

Figure 5: The average packet delay for the logical allXY, hypercube and torus topologies ($\gamma = 1$)

6 Performance of the logical topologies

In the previous section, we developed an analytical model for the logical topologies and verified the analytical model with simulations. In this section, we study the performance of the logical topologies. Since the simulation results and the analytical results match reasonably well, we will only use the analytical model in this section to study the performance.

Figure 6 shows the impact of the packet routing time on the maximum throughput. The underlying topology is a 32×32 torus. For all logical topologies, increasing the speed of routers increases the maximum throughput up to a certain limit. For example, for the all–to–all topology, the router speed of 1 packet per 4 time slots is sufficient to overcome the router performance bottleneck. When the routing speed is faster than the threshold value, the maximum throughput is bound by the link speed. Table 2 shows the bandwidth limits of routers and links for $N = 32$.

Figure 6 also shows that the all–to–all topology achieves higher maximum throughput than the allXY topology, which in turn achieves higher maximum throughput than the hypercube topology. The logical torus has the worst maximum throughput. This observation holds for all packet routing speeds. When the network is saturated, the logical topologies with more connectivity utilizes the bandwidth on links more efficiently than the logical topologies with less connectivity.

Figure 7 shows the impact of the network size on the maximum throughput. The results in this figure are based upon a packet routing time of one time slot. We also studied different packet routing times and found similar trends. In terms of the maximum throughput, the all–to–all topology scales the best, followed by the allXY topology, followed by the hypercube topology. The logical torus topology scales worst among all these topologies. Figures 6 and 7 show that by using time–division multiplexing to establish complex logical topology, we can exploit the large aggregate bandwidth in the network and deliver higher throughput when the network is under high load.

The average packet delay is another performance metric to be considered. For a network

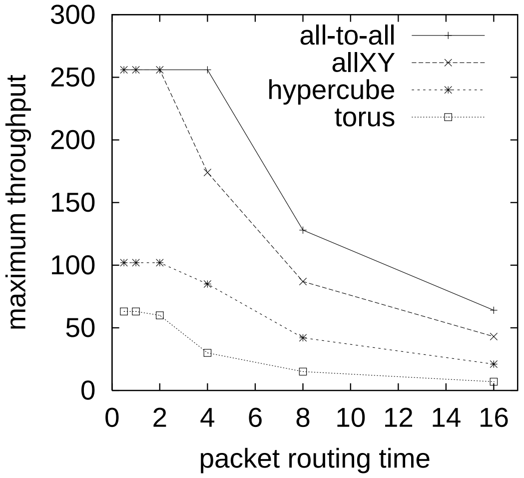


Figure 6: Impact of the packet routing time on the maximum throughput ($N = 32$)

topology	bottleneck	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
all–to–all	λ_s^{max}	2.0	1.0	0.5	0.25
	λ_p^{max}	0.25	0.25	0.25	0.25
	λ^{max}	0.25	0.25	0.25	0.25
allXY	λ_s^{max}	1.36	0.68	0.34	0.17
	λ_p^{max}	0.25	0.25	0.25	0.25
	λ^{max}	0.25	0.25	0.25	0.17
hypercube	λ_s^{max}	0.67	0.33	0.17	0.09
	λ_p^{max}	0.1	0.1	0.1	0.1
	λ^{max}	0.1	0.1	0.1	0.09
torus	λ_s^{max}	0.24	0.12	0.06	0.03
	λ_p^{max}	0.06	0.06	0.06	0.06
	λ^{max}	0.06	0.06	0.06	0.03

Table 2: Maximum throughput for the logical topologies on 32×32 torus

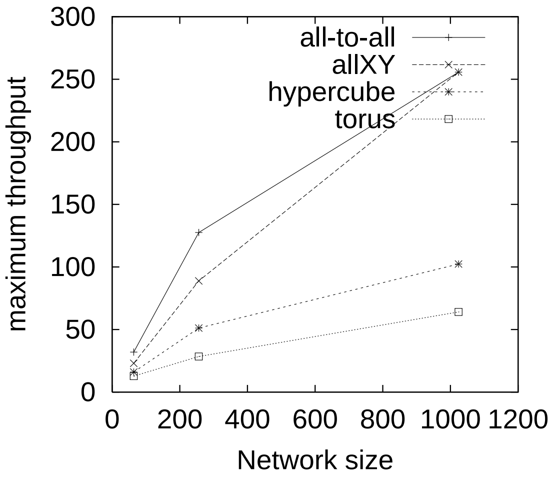


Figure 7: Impact of the network size on the maximum throughput ($\gamma = 1$)

to be efficient, it must be able to deliver packets with small delays. It is well known that time–division multiplexing results in large average packet delay due to the sharing of the links. However, as we have discussed earlier, while using time–division multiplexing techniques to establish logical topologies increases the per hop transmission time, it reduces the average number of hops that a packet travels. Thus, the overall performance depends on many factors and needs further study.

Figure 8 shows the average packet delay with respect to the packet generation rate. The results are based upon a physical 16×16 torus network and $\gamma = 1$. As we can see from the figure, the all–to–all topology incurs very large delay compared to other logical topologies, this is because of the large multiplexing degree needed to realize the logical all–to–all topology. Other topologies have similar delays when the packet generation rate is small. However, the allXY topology has a larger saturation point than the hypercube and torus topologies, and thus has a small delay even when the network load is reasonably high (e.g. $\lambda = 0.25$). These results also hold for larger packet routing times.

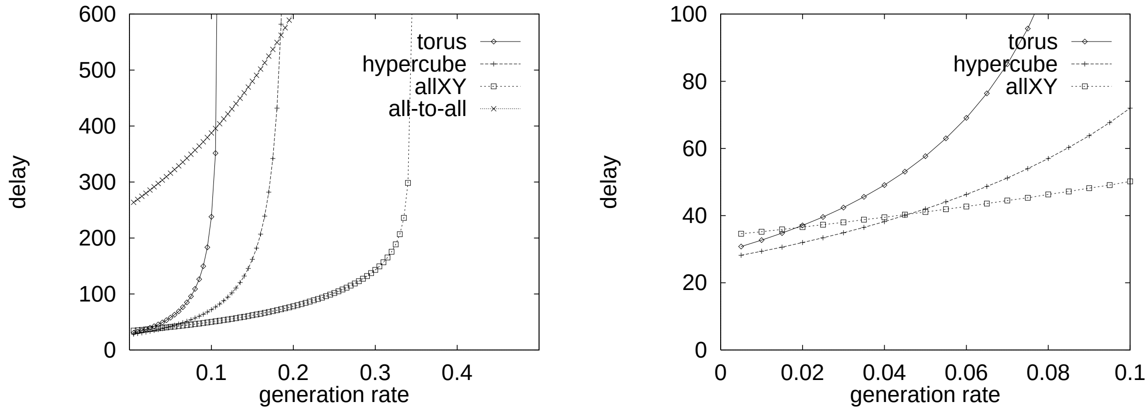


Figure 8: The average packet delay as a function of the packet generation rate ($\gamma = 1.0, N = 16$)

Figure 9 shows the impact of the packet routing time on the average packet delay. The results are based upon a physical 16×16 torus network and a packet generation rate of 0.005. The packet routing speed affects the average packet delay for all topologies. When the packet routing time is very small ($\gamma = 0.25$), the torus topology has the smallest delay. When the packet routing time increases, the delay in torus increases drastically, while the delays in the all–to–all and allXY topologies increase slightly. In the all–to–all and allXY topologies a packet travels through fewer number of routers than it does in the torus topology. Hence the contention at routers does not affect the delay in the all–to–all and allXY topologies as much as it does in the torus and hypercube topologies. This study implies that to achieve low packet delay for the logical torus topology, fast routers are crucial, while a fast router is not as important in the all–to–all and allXY topologies.

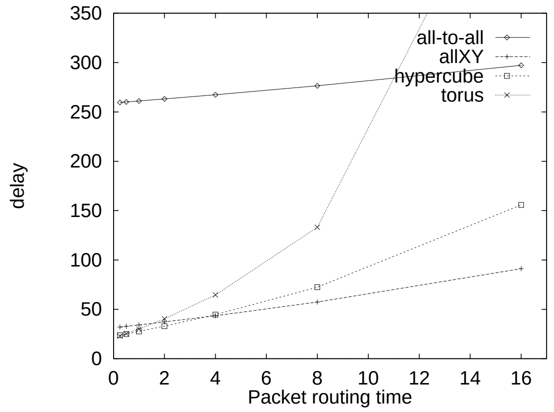
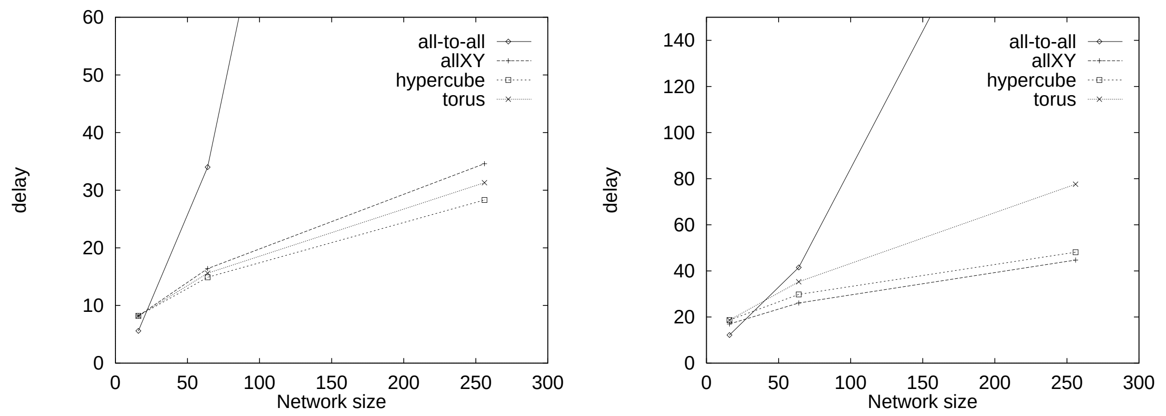


Figure 9: Impact of the packet routing time on the average packet delay ($\lambda = 0.005, N = 16$)

Figure 10 shows the impact of the network size on the packet delay. The results are based upon a packet routing time of 1 time slot and a packet generation rate of 0.01. As shown in the figure, the all–to–all topology has very large delay when the network size is large while the other three logical topologies have similar average packet delays. When the packet routing time is small ($\gamma = 1.0$), the hypercube topology scales slightly better than the torus and the allXY topologies as shown in Figure 10 (a). When γ is large ($\gamma = 4.0$), the hypercube topology and the allXY topology are better than the other two topologies as shown in Figure 10 (b).

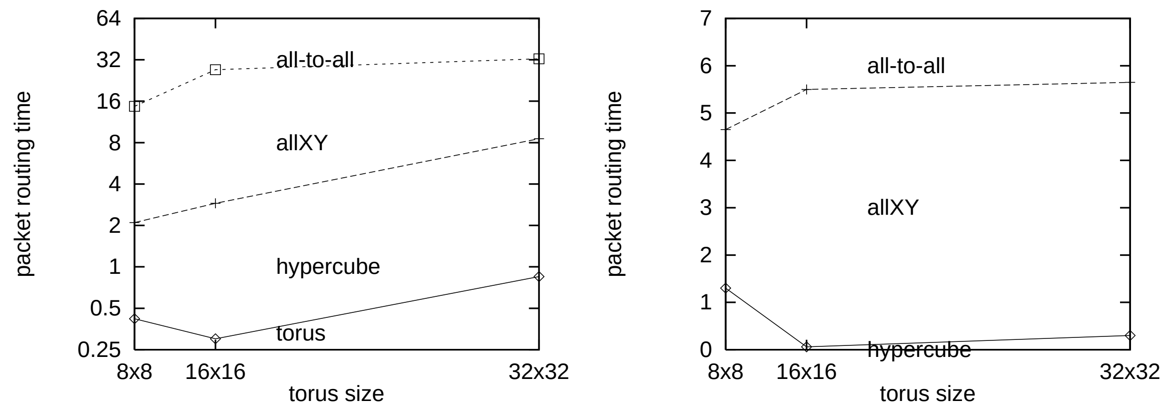
From the above discussions, three parameters, N, γ and λ affect the average packet delay for all the logical topologies. Next, we will identify the regions in the (N, γ, λ) parameter space, where a logical topology has the lowest packet delay. Figure 11 shows the best topologies in the parameter space (N, γ) for a given packet generate rate. A topology is the best in that it offers the smallest average packet delay under the given set of parameters. As can be seen from the Figure 11 (a), with a small packet generation rate, all four logical topologies occupy part of the (N, γ) parameter space, which indicates that under certain conditions, each of the four topologies out–performs the other three topologies. In the case of a larger packet generation rate as shown in Figure 11 (b), the logical torus topology is pushed out of the best topology picture. These results show that logical topologies with low connectivity are quite sensitive to the network load while logical topologies with high connectivity are not.

Figure 12 shows the best logical topologies on the (γ, λ) parameter space for a given network size. Here, the underlying network is a 16×16 torus. Networks of different size exhibit similar characteristics. The majority of the (γ, λ) parameter space is occupied by the logical hypercube



(a) $\gamma = 1.0$ (b) $\gamma = 4.0$

Figure 10: Impact of the network size on the average packet delay ($\lambda = 0.01$)



(a) $\lambda = 0.01$ (b) $\lambda = 0.06$

Figure 11: The best logical topologies for a given packet generation rate

and allXY topologies. The logical torus topology is good only when λ is small and γ is small. The logical all–to–all topology out–performs other topologies only when the network is almost saturated, that is, large λ or large γ. This indicates that in general, the logical hypercube and allXY topologies are better topologies than the logical torus and all–to–all topologies in terms of the average packet delay.

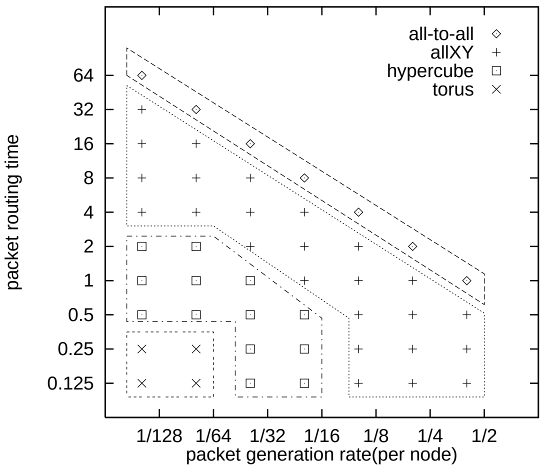


Figure 12: The best logical topologies for a 16×16 torus

Figure 13 compares the performance of the logical hypercube and allXY topologies. Given a fixed γ, there is a packet generation rate, λ, above which the allXY topology out–performs the logical hypercube topology. When γ increases, the line in the figure moves down. In other words, the hypercube topology is more sensitive to the packet routing time γ.

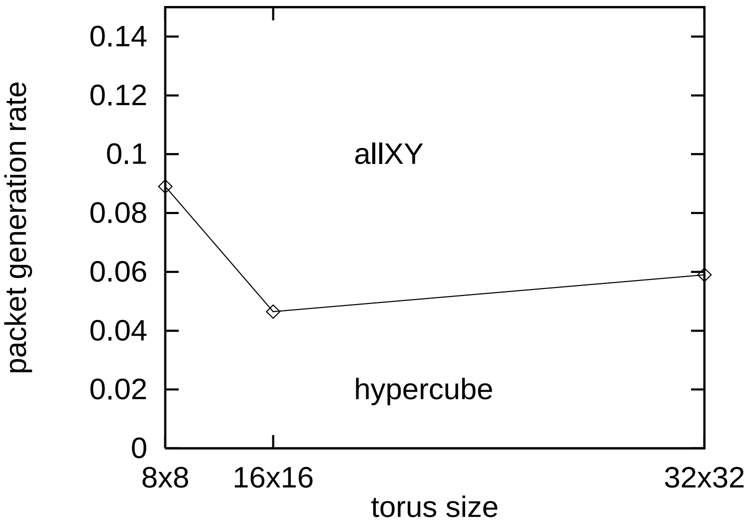


Figure 13: The best logical topologies for a given packet routing time ($\gamma = 1.0$)

7 Conclusion

In this paper, we have studied the logical topologies for routing messages on top of physical torus networks. We developed an analytical model for the maximum throughput and the average packet latency for multi–hop communication, verified the model with simulations, studied the performance of these topologies and identified the cases where each logical topology out–performs the other topologies.

In general, the performance of the logical topologies with less connectivity, such as the torus and hypercube topologies, are more sensitive to the network load and the router speed while the logical topologies with dense connectivity, such as the all–to–all and allXY topologies, are more sensitive to the network size. Logical topologies with dense connectivity achieve higher maximum throughput than the topologies with less connectivity. In addition, they also scale better. In terms of the maximum throughput, the topologies can be ordered as follows:

$$\textit{all–to–all} > \textit{allXY} > \textit{hypercube} > \textit{torus}.$$

In terms of the average packet delay, the logical torus topology achieves best results only when the router is fast and the network is under light load, while the logical all–to–all topology is the best only when the network is almost saturated. In all other cases, the logical hypercube and allXY topologies out–perform the logical torus and all–to–all topologies. Comparing the logical allXY to the logical hypercube, the allXY topology is better when the network is under high load. These results hold for all network sizes.

References

[1] R.A. Barry, V.W.S. Chan, K.L. Hall, E.S. Kintzer, J.D. Moores, K.A. Rauschenbach, E.A. Swanson, L.E. Adams, C.R. Doerr, S.G. Finn, H.A. Haus, E.P. Ippen, W.S. Wong, and M. Haner "All–Optical Network Consortium – Ultrafast TDM Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 5, June 1996.

[2] I. Chlamtac, A. Ganz and G. Karmi. "Lightpath Communications: An Approach to High Bandwidth Optical WAN's" *IEEE Trans. on Communications*, Vol. 40, No. 7, pages 1171–1182, July 1992.

[3] I. Chlamtac, A. Ganz and G. Karmi. "Lightnets: Topologies for high-speed optical networks." *Journal of Lightwave Technology*, 11(5/6):951-961, May/June 1993.

[4] P. Dowd, K. Bogineni and K. Ali, "Hierarchical Scalable Photonic Architectures for High-Performance Processor Interconnection", *IEEE Trans. on Computers*, vol. 42, no. 9, pp. 1105-1120, 1993.

[5] R. Dutta and G. N. Rouskas, "A Survey of Virtual Topology Design Algorithms for Wavelength Routed Optical Networks." *Optical Network Magazine*, 1(1):73-89, January 2000.

[6] P.E. Graan, "Optical Networking Update." *IEEE Journal on Selected Areas of Communications*, 14(5):764-779, June 1996.

[7] S. Hinrichs, C. Kosak, D.R. O'Hallaron, T. Stricker and R. Take. "An Architecture for Optimal All–to–All Personalized Communication." In *6th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 310-319, June 1994.

[8] H. Scott Hinton, "Photonic Switching Using Directional Couplers", *IEEE Communication Magazine*, Vol 25, no 5, pp 16-26, 1987.

[9] E. Karasan and E. Ayanoglu "Performance of WDM Transport Networks" *IEEE Journal on Selected Areas in Communications*, 16(7):1081-1096, September 1998.

[10] R. M. Krishnaswamy and K.N. Sivarajan "Design of Logical Topologies: A Linear Formulation for Wavelength Routed Optical Networks with No Wavelength Changers." *Proc. of IEEE INFOCOM*, pages 919-927, 1998.

[11] J.F.P. Labourdette "Traffic Optimization and Reconfiguration Management of Multiwavelength Multihop Broadcast Lightwave Networks." *Computer Networks and ISDN Systems*, 30:981-998, May 1998.

[12] M.A. Marsan, A. Bianco, E. Leonardi and F. Neri "Topologies for Wavelength-Routing All-Optical Networks", *IEEE/ACM Trans. Networking*, 1(5):534-546, Oct 1993.

[13] B. Mukherjee, "WDM-Based Local Lightwave Networks Part I: Single–Hop Systems." *IEEE Network*, Vol. 6, No. 3, pages 12–27, May 1992.

[14] B. Mukherjee, "WDM-Based Local Lightwave Networks Part II: Multi–Hop Systems." *IEEE Network*, Vol. 6, No. 4, pages 20–33, July 1992.

[15] C. Qiao and R. Melhem, "Reconfiguration with Time Division Multiplexed MIN's for Multiprocessor Communications." *IEEE Trans. on Parallel and Distributed Systems*, Vol. 5, No. 4, pages 337–352, April 1994.

[16] C. Qiao and Y.Mei, "Wavelength Reservation Under Distributed Control." to *Proc. of IEEE/LEOS summer topical meeting: Broadband Optical Networks*, paper TuB5, August 1996.

[17] C. Qiao and R. Melhem, "Reducing Communication Latency with Path Multiplexing in Optically Interconnected Multiprocessor Systems", *Proc. of the International Symposium on High Performance Computer Architecture*, pages 34–43, Jan. 1995.

[18] R. Ramaswami and K. N. Sivarajan, "Routing and Wavelength Assignment in All–Optical Networks," in *Proc. IEEE INFOCOM'94*, 1994, pages 970–979.

[19] R. Ramaswami and K.N. Sivarajan, "Design of Logical Topologies for Wavelength-routed Optical Networks." *IEEE Journal on Selected Areas of Communications*, 14(5):840-851, June 1996.

[20] S. Subramanian, M. Azizoglu and A. Somani, "Connectivity and Sparse Wavelength Conversion in Wavelength-Routing Networks." *Proc. of INFOCOM'96*, pages 148–155, 1996.

[21] B. Y. Yu, R. Runser, P. Toliver, K.-L. Deng, D. Zhou, T. Chang, S.W. Sea, K. Kang, I. Glesk, and P.r. Prucnal "TDM 100Gbits/s Packet Switching in an Optical ShuffleNet." *Hot Interconnects Symposium*, August 21-23, 1997, Stanford, CA.

[22] X. Yuan, R. Melhem and R. Gupta "Distributed Path Reservation Algorithms for Multiplexed All-optical Interconnection networks" *the Third International Symposium on High Performance Computer Architecture(HPCA 3)*, pages 38–47, San Antonio, Texas, Feb.1-5, 1997

[23] X. Yuan and R. Melhem "Optimal Routing and Channel Assignments for Hypercube Communication on optical Mesh–like Processor Arrays." *Fifth International Conference on Massively Parallel Processing Using Optical Interconnections*(MPPOI'98), Las Vegas, June 1998