The Effect of Adaptivity on the Performance of the OTIS-Hypercube under Different Traffic Patterns

H. H. Najaf-abadi [1], H. Sarbazi-Azad [2,1]

[1] School of Computer Science, IPM, Tehran, Iran.
[2] Computer Engineering Dept., Sharif Univ. of Technology, Tehran, Iran.
{h_hashemi, azad}@ipm.ir, azad@sharif.edu

Abstract. The OTIS-hypercube is an optoelectronic architecture for interconnecting the processing nodes of a multiprocessor system. In this paper, an empirical performance evaluation of the OTIS-hypercube is conducted for different traffic patterns and routing algorithms. It is shown that, depending on the traffic pattern, minimal path routing may not have the best performance and that adaptivity may be of no improvement. All judgments made are based on observations from the results of extensive simulation experiments of the interconnection network. In addition, logical explanations are suggested for the cause of certain noticeable performance characteristics.

1. Introduction

In order to exploit the speed and power advantages of optical interconnect (in communication distances exceeding a few millimeters [1, 2]), the OTIS architecture for interconnection networks has been suggested by Marsden et al. [3], Hendrick et al. [4] and Zane et al. [5]. Algorithmic properties of specific, cases such as the OTIS-hypercube and OTIS-mesh, have also been developed in the literature [7-13]. However, previous studies have, to our best knowledge, only considered topological and algorithmic issues in OTIS computers, and no study has evaluated the performance of these systems in sight of parameters such as bandwidth and message latency, in view of realistic implementation assumptions.

The main purpose of this work is to take a step in this direction by initially developing a deadlock-free routing scheme for the OTIS-hypercube, and evaluating the performance of the network under realistic conditions and structural constraints. To this end, extensive simulation experiments have been conducted on the network, with different routing algorithms, traffic patterns, traffic loads, network sizes, message lengths and number of virtual channels.

2. The OTIS-Hypercube and its Router Structure

In the OTIS-hypercube parallel computer, there are 2^{2N} processors organized as 2^N groups of 2^N nodes each. The processors in each group form an N dimensional hypercube that employs electrical interconnect. The inter-group interconnections are realized by optics. In the OTIS interconnect system, processor (i, j), i.e. processor j of group i, is connected via optics to processor (j, i).

A node, in the n-dimensional OTIS-hypercube, or OTIS-H_n for short, consists of a processing element (PE) and a switching element (SE). The PE contains a processor and some local memory. A node is connected, through its SE, to its intra-group neighboring nodes using n input and n output electronic channels. Two electronic channels are used by the PE to inject/eject messages to/from the network. Messages generated by the PE are transferred to the router through the injection channel. At the destination node, messages are transferred to the local PE through the ejection channel. The optical channel is used to connect a node to its transpose node in some other group for inter-group communication. The router contains flit buffers for each incoming channel. A number of flit buffers are associated with each physical input channel. The flit buffers associated with each channel may be organized into several lanes (or virtual channels), and the buffers in each virtual channel can be allocated independently of the buffers in any other virtual channel [6]. The concept of virtual channels has been first introduced in the context of the design of deadlock free routing algorithms, where the physical bandwidth of each channel is multiplexed between a number of messages [6]. However, virtual channels can also reduce network contention. The input and output virtual channels are connected by a crossbar switch that can simultaneously connect multiple input channels to multiple output channels given that there is no contention over the output channels.

3. Message Routing in the OTIS-hypercube

The routing scheme used for inter-group routing and the routing algorithm used for intra-group routing collectively determine the exact routing algorithm in an OTIS-hypercube network. In what follows, we refer to different *routing schemes* in order to identify only the manner in which a message travels between different sub-graphs (groups) of the network to reach its destination. Two basic routing schemes can be suggested for any source-destination pair of nodes in an OTIS-network. In the first scheme, a message is routed in the local sub-graph in which it starts until it reaches the node that has the same node address as the destination node. From that node, the optical channel is taken into another sub-graph. In this sub-graph, the message is routed until it reaches a node that has the same node address as the sub-graph address of the destination node. Once there, the message takes its final optical hop to the destination node. In the second basic scheme, a message is first routed to a node that has a node address equal to the sub-graph address of the destination. Once there, the optical channel takes the message to the sub-graph of the destination node. The message is then routed to the destination node within this sub-graph.

Of the two former routing schemes, the one that takes a shorter path depends on the full address of the source and destination nodes. When considering the OTIS-hypercube, this can be determined easily. If the number of differing bits of the full address of the source and destination nodes is less than that of the source node and the transpose of address of the destination node, the first routing scheme will result in a shorter path. Otherwise, the second scheme will. However, it should be obvious that in the first routing scheme, once the first optical channel has been taken, the remainder of routing can be conducted by the second scheme. Therefore, if in each intermediate node, a message is routed according to the basic scheme that takes a shorter path to the destination of the message (without considering the source node), a minimal-path routing scheme, the third scheme, is obtained.

Routing within a hypercube may be deterministic, partially adaptive or fully adaptive. Any of these routing techniques may be used for intra-group routing in the

OTIS-hypercube network. In order for a routing algorithm to be deadlock-free, cyclic buffer dependencies between messages and the virtual channels they allocate, must not occur. In the hypercube network, dimension order routing, a well-known deterministic routing algorithm, is inherently deadlock-free. Partially adaptive routing algorithms based on the turn model [14], such as p-cube routing, are also deadlock free. For fully adaptive routing to be deadlock free, virtual channel utilization must be restricted in a way, such as that suggested in [15]. But in an OTIS-hypercube, cyclic buffer dependencies between channels may also occur through the optical connections between groups.

To prevent the occurrence of such cyclic buffer dependencies, messages that enter a group through an optical channel must traverse that group through a separate set of virtual channels from those of messages originating in that group. Therefore, we suggest that the virtual channels of each electronic channel be split into two equal sets, i.e. each group be split into two virtual groups, v_1 and v_2. After being injected into the network, a message traverses the source group through v_1. But once an optical channel has been taken and the message has entered another group, that group is traversed through v_2.

When messages traverse only one optical channel in their path (the second routing scheme), no restriction is necessary on the utilization of the virtual channels of optical channels. But when messages traverse two optical channels (the first routing scheme), cyclic dependencies may still occur if all the virtual channels of optical channels are allowed to be utilized by messages taking their first optical hop. Thus, for the first routing scheme (and consequently the minimal scheme), one of the virtual channels of all optical channels must be reserved for messages that are traversing a second optical channel (entering their destination node). All the other virtual channels of optical channels can be allowed to be traversed with no restriction.

Since a message that has traversed its second optical channel has definitely entered its destination node, it can not be part of a cyclic buffer dependency. It is for this reason that reserving one of the virtual channels of each optical channel, specifically for such messages, eliminates the possibility of the occurrence of cyclic buffer dependencies through the optical channels. In this manner, the deadlock-free nature of the specific hypercube routing algorithm, used for inter-group routing, will be is preserved in the OTIS-hypercube.

4. Empirical Performance Evaluation

The traffic patterns considered in our evaluation are
Uniform: destination node can be any network node with an equal probability,
Complement: node $(a_{n-1}\ldots a_1 a_0)$ sends message to node $(\overline{a}_{n-1}\ldots\overline{a}_1\overline{a}_0)$,
Bit-reverse: node $(a_{n-1}\ldots a_1 a_0)$ sends message to node $(a_0 a_1\ldots a_{n-1})$,
Bit-flip: node $(a_{n-1}\ldots a_1 a_0)$ sends message to node $(\overline{a}_0\overline{a}_1\ldots\overline{a}_{n-1})$,
Butterfly: node $(a_{n-1}\ldots a_1 a_0)$ sends message to node $(a_0 a_{n-2}\ldots a_1 a_{n-1})$,
Perfect-shuffle: node $(a_{n-1}\ldots a_1 a_0)$ sends message to node $(a_{n-2}a_{n-3}\ldots a_0 a_{n-1})$.

To evaluate the functionality of the OTIS-hypercube network under different conditions, a discrete-event simulator has been developed that mimics the behavior of the described routing algorithms at the flit level. In each simulation experiment, a minimum of 120,000 messages were delivered and the average message latency calculated. Statistics gathering was inhibited for the first 10,000 messages to avoid distortions due to startup transience. The average message latency is defined as the average amount of time from the generation of a message until the last data flit of that message is consumed at the local PE at the destination node. The network cycle time is defined as the transmission time of a single flit from one router to the next, through an electric channel. The transmission time of a flit, through an optical channel is however a fraction of the network cycle time. Messages are generated at each node according to a Poisson process with a mean inter-arrival rate of λ_g messages per cycle. All messages have a fixed length of *M* flits. The destination node of each message has been determined through a uniform random number generator to simulate a uniform traffic pattern.

Numerous simulation experiments have been performed for different scenarios of the traffic load, traffic pattern and routing algorithm for various message lengths and network configurations. However, message length and network configuration have been observed to be of no effect on the proportional performance of different scenarios. Hence for brevity, we report the results for only a typical setting. This setting consists of a six dimensional OTIS-hypercube with four virtual channels per physical channel. The ratio of optical channel transmission time to electronic channel transmission time is equal to 1/10 and messages have a fixed length of 32 flits.

In the following subsections, we measure the performance of different traffic patterns by means of the *saturation point*. The saturation point is the maximum injection rate at which the average delay is still bounded. It is assumed that when the average message latency is higher than 200000 unit cycles, the network enters saturation region.

4.1. Uniform, Bit-flip, Bit-reverse, and Butterfly Traffic Patterns

In an OTIS-hypercube with uniform traffic, regardless of the routing scheme, the performance of adaptive routing is superior to that of deterministic and p-cube routing, as can be seen in Figure 1. Furthermore, the minimal routing scheme performs better than the first and second routing schemes. But the interesting point is that deterministic routing saturates at a higher generation rate than that of p-cube routing. The OTIS-hypercube inherits this performance characteristic for uniform traffic from the hypercube network (Glass and Ni have reported such a characteristic for the performance of the hypercube network [14]). Due to the fact that, when used individually, the first and second schemes do not always rout messages through an optimal path, one would expect the minimal routing scheme to saturate at much higher generation rates. But for adaptive routing, the difference between the performance of minimal routing and that of the first or second routing scheme is less than what may have been predicted. Thus, considering the extra complexity of implementing the minimal routing scheme, this scheme may not be an efficient option for such a system in which traffic is uniform. But, as will be shown in the following sections, for other traffic patterns, minimal routing may even result in performance poorer than that of the first or second schemes.

As shown in Figure 1 for bit-flip traffic, compared to the minimal routing scheme, the network saturates at a much higher generation rate when the second scheme is used, with bit-flip traffic for all three inter-group routing algorithms. This is while messages travel a longer average distance with the second scheme. An explanation for this is that, with the first routing scheme, all messages generated by bit-flip traffic in a specific group, exit that group through the same optical channel (the optical channel exiting a node whose address is the bit-flip of the group address), creating a bottleneck in the network. It is also apparent from the results that, with the second routing scheme, the generation rate for which the network saturates is greater for P-cube routing than that for deterministic routing. The reason for this is that with bit-flip traffic in an OTIS-hypercube, when the second routing scheme is used, the traffic within each sub-graph is also bit-flip traffic. As shown in [14], a hypercube network with P-cube routing saturates at a higher generation rate than that of deterministic routing for bit-flip traffic. This is while P-cube routing saturates at a lower rate than deterministic routing for uniform traffic.

In the results obtained for bit-reverse traffic, also shown in Figure 1, it is observed that the generation rate for which the second routing scheme saturates is greater than that for minimal routing. However, the difference between p-cube and deterministic routing is less than that for bit-flip traffic. The reason why the performance of the minimal routing scheme is so poor with bit-reverse traffic stems from the inefficiency of the first routing scheme. Similar to bit-flip traffic, the first routing scheme (not shown in this figure) causes all messages that are injected into a group to exit that group through a single optical channel (the optical channel exiting a node whose address is the bit-reverse of the group address). But this is not the case for the second routing scheme where messages use the optical channels to exit their source group evenly. With the second routing scheme used for bit-reverse traffic in an OTIS-hypercube, the traffic within each group is also bit-reverse traffic. This explains the superior performance of p-cube routing over deterministic routing, when the second scheme is used.

With Butterfly traffic, results of which are depicted in Figure 1, the performance of minimal routing is unquestionably better than that of the second routing scheme. Left out from this figure to preserve clarity, are the results of the first routing scheme. These results have, however, shown the performance of the first routing scheme to be very close to that of the minimal routing scheme. But the interesting point is that, with the minimal scheme, there seems to be hardly any difference between the different routing algorithms. This is due to the fact that with butterfly traffic, the Hamming distance between the source and destination nodes of any message is equal to 2. It is thus, unsurprising that the degree of adaptivity with which those two hops are traversed is almost of no effect on the performance of the network. Another point is that, since the Hamming distance between the source and destination nodes is so small, the first routing scheme will almost always be selected by the minimal routing scheme. This explains why, for butterfly traffic, the performance of minimal routing is so close to that of the first routing scheme and why these two schemes are superior to the second scheme.

4.2. Complement and Perfect-shuffle Traffic Patterns

With complement traffic, hardly any difference can be observed between the performance of minimal routing and that of the second routing scheme. This can be observed in the results of Figure 2-a. But the first routing scheme saturates at a much higher generation rate than the other two schemes. The first routing scheme results in the path from source to destination of a message to be equal to the diameter of the network. Therefore, the second routing scheme will never rout messages through a longer path than that of the first scheme. Thus with minimal routing, the second scheme will always be selected. This explains why the minimal scheme and second scheme perform equally. With a complement traffic pattern, all messages injected into a specific group are destined to the same destination group (the address of which is complement to that of the source group address). Thus, with the second routing scheme, all messages are routed to the same node in the source group, i.e. they all exit the source group through the same optical channel. As a result, excessive traffic load is imposed on some optical channels while others are left absolutely unused. Even the traffic load on the electronic channels becomes unequally distributed. But this is not the case for the first routing scheme, by which complement traffic is distributed evenly over the optical channels. This explains why, as depicted in Figure 2-b, the first routing scheme saturates at a much higher generation rate than that of the second scheme, even though messages traverse a longer average distance with the first scheme. In an OTIS-hypercube with complement traffic, there is also complement traffic within each group when the first routing scheme is used.

When the second routing scheme is used for perfect-shuffle traffic, all messages injected into a sub-graph exit that sub-graph, the traffic pattern within each group becomes somewhat similar to the perfect-shuffle pattern. The only difference corresponds to the LSB of the destination address. Therefore, considering that results presented in [14] show that with perfect-shuffle traffic in the hypercube, p-cube routing saturates at a lower generation rate than that of deterministic routing, it is acceptable that through one of two optical channels and the other optical channels exiting that sub-graph are left unused. This, as in the case of complement traffic, results in the uneven distribution of traffic on optical channels, and consequently the second routing scheme suffers from early saturation. But unlike complement traffic, the minimal routing of perfect-shuffle traffic does not always utilize the second scheme. Nevertheless, the poor performance of the second routing scheme does affect the performance of the minimal routing scheme. As a result, minimal routing saturates at a generation rate only slightly higher than that of the second scheme, as revealed in Figure 2-a.

In contrast to complement traffic, even the first routing scheme does not distribute perfect-shuffle traffic equally over the optical channels. Since the MSB (most significant bit) of the group address is rotated into the LSB (least significant bit) of the node address, the first routing scheme causes all messages of the same source

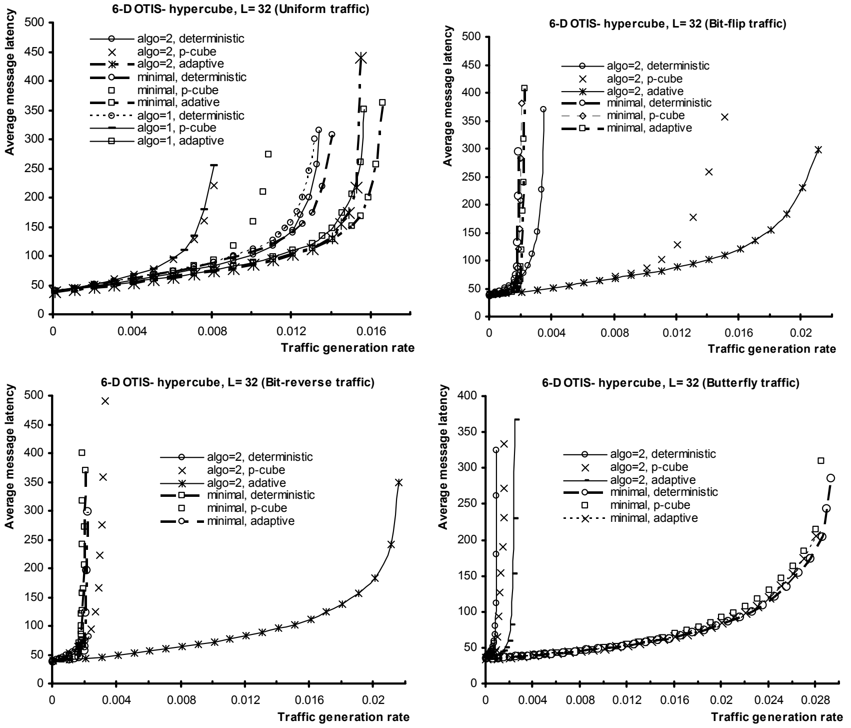
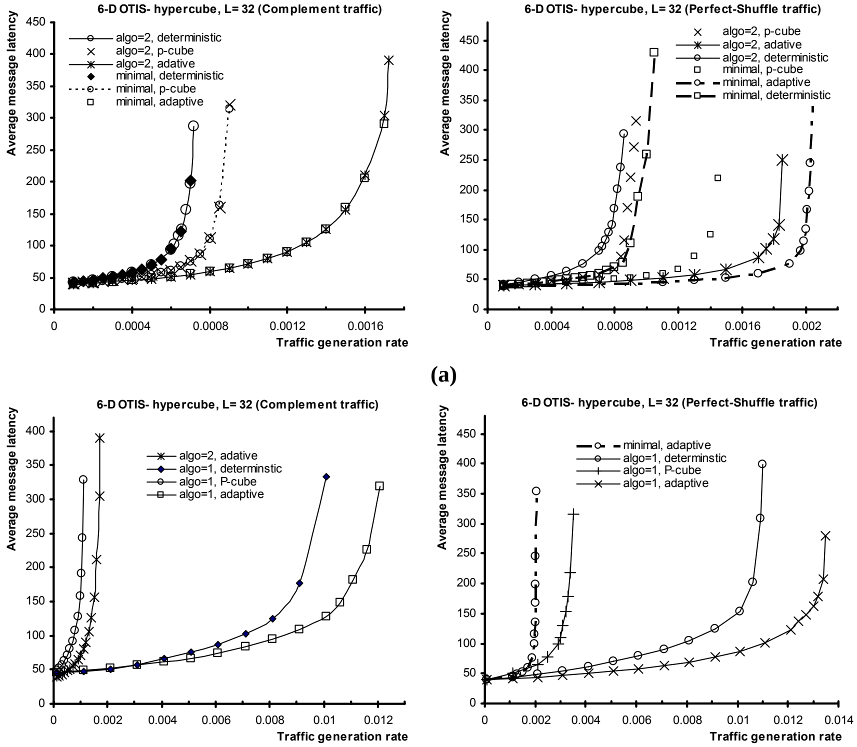


Fig. 1: Average message latency of uniform, bit flip, bit reverse, and butterfly traffic patterns



(a)

(b)

Fig. 2: Average message latency of complement and perfect-shuffle traffic patterns for (a) low and (b) high generation rates.

group to exit that group through the optical channels of nodes with either even or odd addresses, depending on the MSB of the group address. Nonetheless, the first scheme does maintain superior performance over the second scheme. This can be observed in the results of Figure 2-b. The results obtained for adaptive intra-group routing based on the second routing scheme, shown in Figure 2-a, have been included in Figure 2-b once again to facilitate the comparison of the performance of the different routing schemes in this traffic pattern.

5. Conclusions and Future Work

A simulation-based evaluation of the performance of the OTIS-hypercube network has been conducted for three different inter-group routing schemes that we have defined (*first*, *second* and *minimal* routing schemes), three different intra-group routing algorithms (deterministic, fully adaptive and partially adaptive routing) and five different traffic patterns (uniform, complement, bit-reverse, bit-flip, butterfly, perfect-shuffle). We have shown that the method of routing messages between different groups of the network (the inter-group routing scheme) and the intra-group routing algorithm are of considerable influence on the performance of the OTIS-hypercube. However, we observe that (with the exception of uniform traffic) the inter-group routing scheme is generally of greater effect on performance than intra-group routing. Traffic patterns have also been found to be deeply influential on performance. It is found that with bit-flip, and bit-reverse traffic, the network saturates at higher generation rates when the second inter-group routing scheme is used, whereas poor performance is attained with the first routing scheme. The converse holds for butterfly, complement and perfect-shuffle traffic. This is while minimal routing is of superior performance only with uniform traffic.

Consideration of these characteristics can serve as a guideline to the optimal mapping of tasks to nodes by the operating system of such multiprocessor systems. Our next objective is to derive a mathematical performance model of wormhole routing in the OTIS-hypercube, and to validate its prediction accuracy using simulation experiments.

References

1. M. Feldman, S. Esener, C. Guest and S. Lee, “Comparison between electrical and free space optical interconnects based on power and speed considerations”, *applied optics*, 27(9): 1742-1751, May 1988.
2. F. Kiamilev, P. Marchand, A. Krishnamoorthy, S. Esener, and S. Lee, “Performance comparison between optoelectronic and VLSI multistage interconnection networks”, *journal of lightwave technology*, 9(12): 1674-1692, Dec. 1991.
3. G. C. Marsden, P. J. Marchand, P. Harvey, and S. C. Esener, “Optical transpose interconnect system architectures”, *Optical Letters*, 18(13): 1083-1085, July 1993.
4. W. Hendrick, O. Kibar, P. Marchand, C. Fan, D. V. Blerkom, F. McCormick, I. Cokgor, M. Hansen, and Esener, “Modeling and optimization of the optical transpose interconnection system”, *Optoelectronic technology Center*, Sept. 1995.
5. F. Zane, P. Marchand, R. Paturi, and S. Esener, “Scalable network architectures using the optical transpose interconnection system (OTIS)”, In proceedings of the second International Conference on Massively Parallel Processing using Optical Interconnections (MPPOI’96), pages 114-121, San Antonio, Texas, 1996.
6. W.J. Dally, “Virtual channel flow control”, IEEE TPDS, 3 (2) (1992), 194-205.
7. S. Sahni, C.-F. Wang, “BPC permutations on the OTIS-hypercube optoelectronic computer”, Informatica, 22: 263-269, 1998.
8. S. Sahni and C.-F. Wang, “BPC permutations on the OTIS-mesh optoelectronic computer”, In proceedings of the fourth international conference on massively parallel processing using optical interconnections (MPPOI’97), pages 130-135, 1997.
9. C.-F. Wang and S. Sahni, “Matrix multiplication on the OTIS-mesh optoelectronic computer”, In Proceedings of the sixth international conference on Massively Parallel Processing using Optical Interconnections (MPPOI’99), pages 131-138, 1999.
10. C. -F. Wang and S. Sahni, “Image processing on the OTIS-mesh optoelectronic computer”, IEEE TPDS, 11(2): 97-107, 2000.
11. C. -F. Wang and S. Sahni “Basic operations on the OTIS-mesh optoelectronic computer”, IEEE TPDS, 9(12): 1226-1236, 1998.
12. S. Rajasekeran and S. Sahni “Randomized routing, selection and sorting on the OTIS-mesh”, IEEE TPDS, 9(9): 833-840, 1998.
13. A. Osterloh, “Sorting on the OTIS-mesh”, In Proceedings of the 14th International Parallel and Distributed Processing Symposium (IPDPS’2000), pp. 269-274, 2000.
14. C. J. Glass and L. M. NI, “The Turn model for adaptive routing”, *J. ACM*, vol. 5, pp. 874–902, 1994.
15. J. Duato, T. Pinkston, “A general theory for deadlock-free adaptive routing using a mixed set of resources”, IEEE Transaction on Parallel and Distributed Systems, Vol. 12, 2001, pp. 1219-1235.