Performance Assessment of All-Reduce Communication Operation in OTIS-Mesh Optoelectronic Architecture

Basel A. Mahafzah, Sami I. Serhan, and Ruby Y. Tahboub

***Abstract*—** As a barrier synchronization communication operation, all-reduce communication operation is used in many parallel and distributed algorithms. In this paper, the all-reduce communication operation is developed using Extended Dominating Node (EDN) approach on OTIS-Mesh (Optical Transpose Interconnection System Mesh) optoelectronic architecture. Also, the performance assessment of all-reduce communication operation is presented mathematically and by simulation in terms of number of communication steps, latency, and latency improvement; among three optoelectronic architectures: the single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh. The obtained mathematical and simulation results show that the all-reduce communication operation on all-port EDN-OTIS-Mesh significantly outperforms the single-port and all-port OTIS-Mesh.

***Keywords*—** All Reduce Operation, Barrier Synchronization, Extended Dominating Node, Interconnection Network, Mesh.

I. Introduction

DESIGNING and implementing efficient communication operations in optoelectronic architectures plays great role in the performance of parallel and distributed algorithms. In optoelectronic architecture, the communication links between processing elements are electronic and optical, where for short distances electronic links are used and for long distances optical links are used, since optical links offer faster data transmission speeds [1], [2]. An attractive optoelectronic architecture that has gained a considerable attention in the recent years is the OTIS (Optical Transpose Interconnection System) architecture [2]. In OTIS optoelectronic architecture, processors are organized into groups, where in each group; processors are connected electronically along short distances forming a basis network, such as a mesh, hypercube, etc., whereas the longer interconnection among groups is achieved optically. OTIS-Mesh, under study, is an example of OTIS, in which each of the constituting groups is a mesh network. Thus, more details can be found in [3], [4] about mesh and hypercube basis networks.

In parallel and distributed algorithms including load balancing, Traveling Salesman Problem (TSP), and matrix multiplication [5]–[7], the all-reduce communication operation can be used as a barrier synchronization involving two or more processors that perform the same operation. Thus, efficient implementation of all-reduce communication operation forms crucial design and yet performance challenges. One graph theoretic approach to all-reduce is called the Extended Dominating Set (EDS) [8], [9]. This approach defines a set of nodes (processors), referred to as Extended Dominating Nodes (EDNs), which are capable of delivering a message to all other processors in a group within a single communication step.

As a continuous work of [10], [11], in this paper, an all-reduce communication operation for OTIS-Mesh optoelectronic architecture is presented and evaluated mathematically and by simulation in terms of number of communication steps, latency, and latency improvement. However, the all-reduce communication operation is based on the EDN approach [8]–[12], where this communication operation has been modified and embedded on OTIS-Mesh, which is referred as all-port EDN-OTIS-Mesh. Moreover, the all-reduce operation on all-port EDN-OTIS-Mesh has been compared with the all-reduce operation on single-port OTIS-Mesh and on all-port OTIS-Mesh. However, the difference between all-port OTIS-Mesh and all-port EDN-OTIS-Mesh is that the later uses all-reduce operation based on the EDN approach.

This paper is organized as follows: Section 2 provides background and related work on all-reduce communication operation, EDN approach, and OTIS optoelectronic architectures. The all-reduce communication operation on all-port EDN-OTIS-Mesh is presented in Section 3. In Section 4, the mathematics assessment of the all-reduce communication operation is presented. The simulation results are discussed in Section 5. Finally, the conclusions and suggested future work are presented in Section 6.

B. A. Mahafzah is with the Department of Computer Science, The University of Jordan, Amman 11942, Jordan (corresponding author phone: +962-6-535-5000; fax: +962-6-515-5522; e-mail: b.mahafzah@ju.edu.jo).

S. I. Serhan is with the Department of Computer Science, The University of Jordan, Amman 11942, Jordan (e-mail: samiserh@ju.edu.jo).

R. Y. Tahboub is with the Department of Computer Science, Purdue University, West Lafayette, IN, USA. This work has been done at the Department of Computer Science, The University of Jordan, Amman 11942, Jordan (e-mail: ruby.tahboub@gmail.com).

II. Background and Related Work

This section presents background and related work on all-reduce communication operation, EDN approach, and OTIS optoelectronic architectures. The all-reduce operation is a synchronization point in parallel and distributed algorithms, where all participating processors must reach a certain point before any processor precedes execution. It consists of two phases: *reduction* and *distribution*; in the *reduction* phase, one processor acts as a barrier processor. When a processor reaches the synchronization point it sends a message to the barrier processor. In the *distribution* phase, after all processors have reached the synchronization point, the barrier processor sends a message to all participating processors so they can continue execution.

All-reduce communication operation plays a great role in developing parallel and distributed algorithms with their ability to rapidly distribute or collect large amounts of data [13], [14]. For example, Matsuda *et al.* [14] modified some collective operations, such as broadcast and reduction algorithms, in MPI (Message Passing Interface) tool to effectively utilize fast wide-area inter-cluster networks and to control the number of nodes, which can transfer data concurrently through wide-area networks to avoid congestion. The all-reduce communication operation can be used on different OTIS optoelectronic architectures; such as OTIS-Mesh, OTIS-Hypercube, etc.

The dominating set approach [12] is applied in graph theory. However, the dominating node approach defines a set of dominating nodes that have direct links to all other processors in a group. Such a feature grants a processor the capability of delivering a message to all other processors in a single step. For example, Fig. 1 shows the dominating nodes 1, 7, 8, and 14 (gray squares) can deliver a message to all other nodes in a single step. The extended dominating nodes preserve the single step message delivery by also delivering to nonadjacent processors. Furthermore, the definition of EDNs can be recursively applied to form multiple levels of EDN processors. The EDN approach was applied in the design of collective communication operations, such as reduction, broadcasting, and global combine operations in all-port wormhole-routed two-dimensional mesh [8], [9].

OTIS was introduced by Marsden, Marchand, Harvey, and Esener [2] as an optoelectronic architecture that combines the advantages of electronic and optical interconnection links. Processors in OTIS are divided into groups, where intra-group (short-distance) communication is realized by electronic links and longer inter-group communication is achieved through optical links. Generally, OTIS optoelectronic architecture is divided into P groups, each of which consists of P processors. A processor within an OTIS group is modeled as a tuple (G, N), where G denotes the group's number and N is the processor's number within the group. Processor (G, N) is connected directly to its transpose processor (N, G) via optical interconnection. Intra-group processors, on the other hand, are connected electronically forming a common interconnection network's topology, such as mesh, hypercube, etc. For example, OTIS-Mesh, under study, consists of P groups with P processors in each group organized as a two-dimensional $\sqrt{P} \times \sqrt{P}$ mesh.

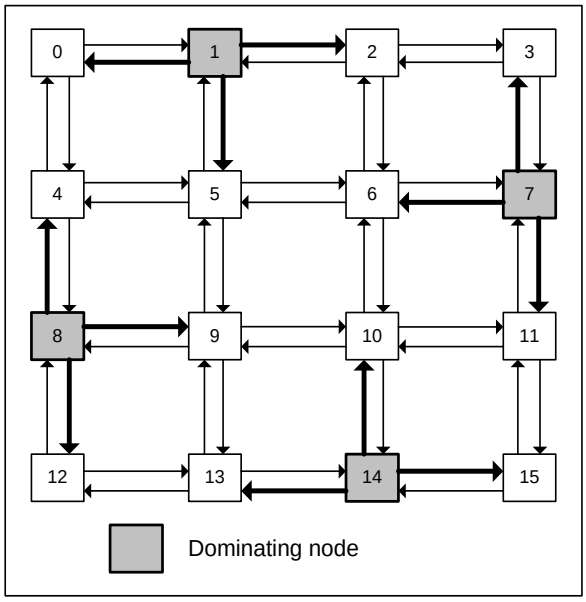

Fig. 1 Dominating nodes in a 4×4 two-dimensional mesh [10].

OTIS optoelectronic architecture has several attractive features. For example, it was verified that OTIS bandwidth is maximized and the power consumption is minimized when the number of groups is equal to the number of processors within each group [15]. Also, several research efforts have achieved significant performance optimization in OTIS through broadcasting, routing, and load-balancing algorithms [16]–[19]. Moreover, Wei and Xiao [20] developed basic communication operations such as: broadcast, prefix sum and data sum on Swapped Network; an optoelectronic interconnection network that resembles OTIS with as number of processors as in OTIS. McKinley, Tsai, and Robinson [21] presented broadcast and global combine algorithms in single- and all-port wormhole-routed parallel computers. Also, the broadcast and global combine operations using EDN approach were applied and evaluated in terms of number of communication steps, latency, and latency improvement on OTIS-Mesh interconnection networks in [10]. Moreover, the all-reduce communication operation is presented and evaluated only analytically in terms of minimum and maximum number of communication steps in [11].

III. All-Reduce Operation in EDN-OTIS-Mesh

This section presents an efficient all-reduce communication operation on all-port OTIS-Mesh using EDN approach referred to EDN-OTIS-Mesh. However, the all-reduce in the EDN-OTIS-Mesh only can be applied on the all-port model, since the all-port model allows multiple messages to be sent to neighbor processors in parallel in a single step, whereas the single-port model allows sending single message at a time.

In particular, OTIS-Mesh is organized as an interconnection of *P* groups, each of *P* processors interconnected as a mesh, where each processor is presented as a two-element tuple (*G*, *N*); where *G* is the group number and *N* is the processor number within *G*. For each processor (*G*, *N*), the corresponding transpose processor is denoted by (*N*, *G*). In OTIS-Mesh, each processor is connected to its corresponding transpose via an optical link. Moreover, on each group of OTIS-Mesh, the EDN approach is applied forming an EDN-OTIS-Mesh.

A 16×16 EDN-OTIS-Mesh is shown in Fig. 2; it consists of 16 groups (*G0*, *G1*… *G15*), each of which contains 16 processors (0, 1, …, 15) interconnected in the form of mesh. As shown by the figure, each processor is interconnected with its transpose processor via an optical link; for example, processor (0, 2) (processor 2 in group 0) shares an optical link with processor (2, 0) (processor 0 in group 2). The shaded processors in each group present the EDN processors, where a set of processors can deliver a message to all other processors in a group in a single communication step. For example, within group *G0*, processors 1, 7, 8 and 14 are called level-1 EDN processors, where they can deliver a message to level-0 processors in a single communication step, while level-0 processors form the rest of the processors.

All-reduce operation is a barrier synchronization point in parallel and distributed programs, where all participating processors must reach before any processor proceeds execution. It consists of two phases: *reduction* and *distribution*.

In the *reduction phase*, the root processor acts as a barrier. When each processor in the network reaches the synchronization point, it sends a message to the barrier processor. However, the following steps present the *reduction phase*:

Step 1: Level-0 processors send a synchronization message to level-1 EDN processors and block execution. Next, level-1 EDN processors send the collected synchronization messages to the next EDN level and block execution until the processors in the highest EDN level collect the entire synchronization messages within the group. Fig. 3 illustrates the *reduction phase* in 64×64 EDN-OTIS-Mesh, where G_0 is the control group. Within each group (G_0 to G_{63}) level-0 processors (white squares) send their synchronization messages to level-1 EDN processors (dotted squares). Next, level-1 EDN processors send the collected messages to level-2 EDN processors (gray squares).

Step 2: The synchronization messages in the highest EDN level are then sent to the processor that has an optical link with the control group. For example, in Fig. 3, level-2 EDN processors sent the synchronization messages to processors in the control group via the optical link.

Step 3: The previous two steps are performed again in the control group. Processors that represent the highest EDN processors level send the collected synchronization messages to the root processor. Fig. 4 illustrates the *reduction phase* in the control group. That is after the control group receives the messages via the optical links, level-0 processors transmit the messages to level-1 EDN processors. Next, level-1 EDN processors transmit the collected messages to level-2 EDN processors. Finally, level-2 EDN processors transmit the messages to the root processor (barrier).

Step 4: The root processor counts the received messages and checks whether all the participating processors has sent a synchronization message.

In the *distribution phase* and after all processors has reached the synchronization point and sent a message to the barrier processor, the root processor (barrier) sends a message to all participating processors so they can continue execution. However, the following steps present the *distribution phase:*

Step 1: The root processor sends permission messages to the EDN processors in the highest level so they can continue their execution. Next, the EDN processors in the highest level send the messages to the next lower level EDN processors until level-0 processors receive the permission messages. Fig. 5 illustrates the *distribution phase* in the control group. The root processor sends level-2 EDN processors a permission message so they can continue execution. Next, level-2 EDN processors forward the permission message to level-1 EDN so they can continue execution. After that, level-1 EDN processors send the permission message to level-0 processors.

Step 2: Each processor has an optical link with another group sends the permission message to that group and the previous steps are performed again. Fig. 6 illustrates the *distribution phase* in 64×64 EDN-OTIS-Mesh. Within each group of (*G1* to *G63*), level-2 EDN processors send the permission message to level-1 processors. Next, level-1 EDN processors send the permission message to level-0 processors.

IV. Mathematics Assessment

This section presents mathematics assessment of all-reduce communication operation in single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh optoelectronic architectures in terms of minimum and maximum number of communication steps. Also, the definitions of latency and latency improvement are presented.

The number of communication steps to perform all-reduce communication operation is the sum of optical and electronic communication steps needed to complete the operation. The required number of optical communication steps between the groups of OTIS-Mesh is two; one step is required in reduction phase and another step is required in the distribution phase. For this reason, we consider the number of communication steps metric is only the sum of the electronic communication steps needed to perform the all-reduce operation.

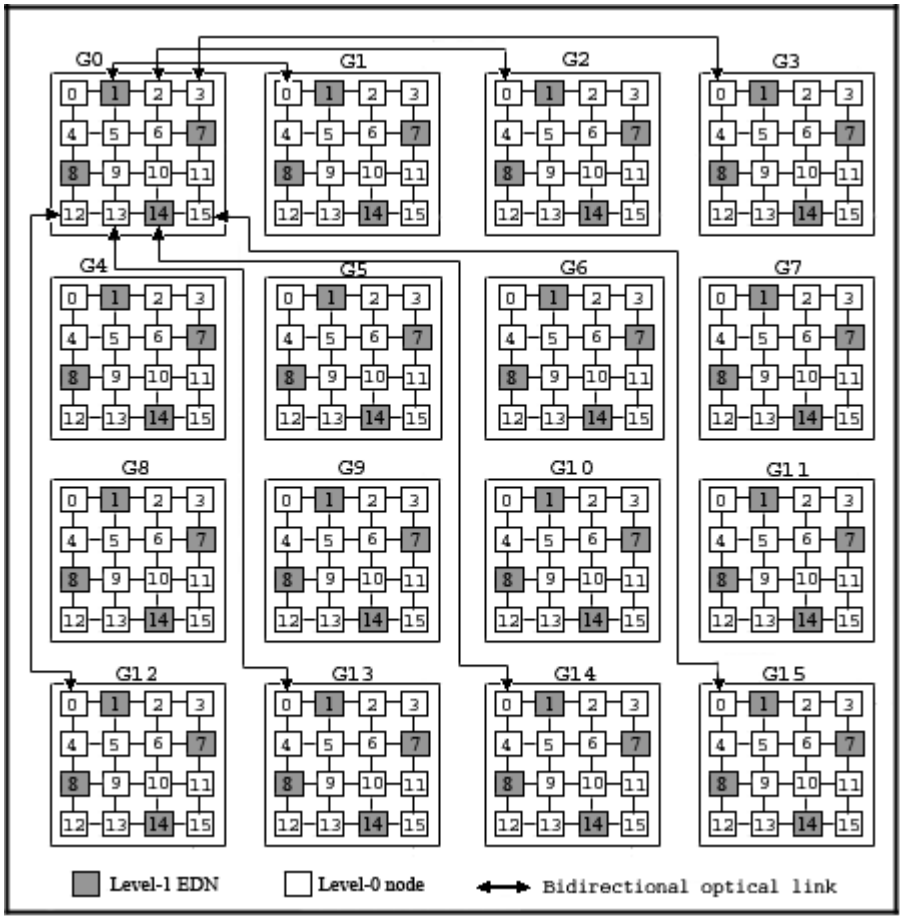

Fig. 2 A 16×16 EDN-OTIS-Mesh [10].

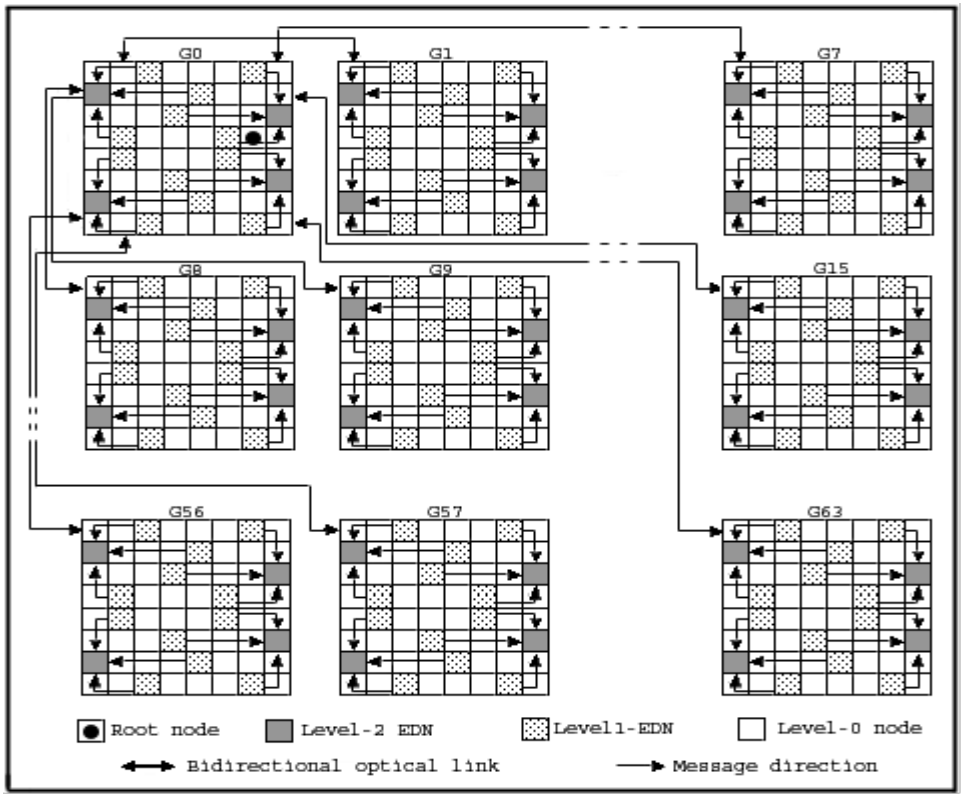

Fig. 3 All-reduce operation in 64×64 EDN-OTIS-Mesh in the reduction phase.

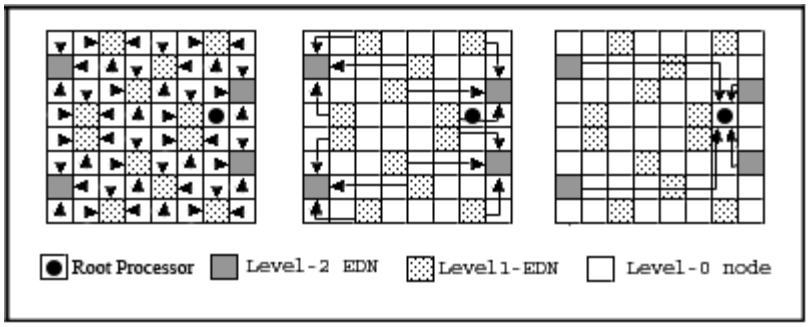

Fig. 4 All-reduce operation in 64×64 EDN-OTIS-Mesh control group in the reduction phase.

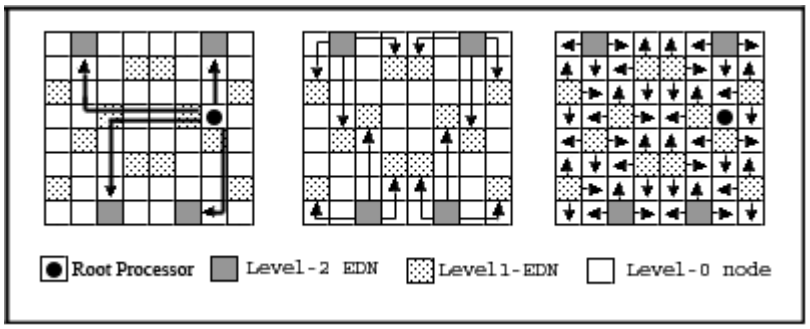

Fig. 5 All-reduce operation in 64×64 EDN-OTIS-Mesh control group in the distribution phase.

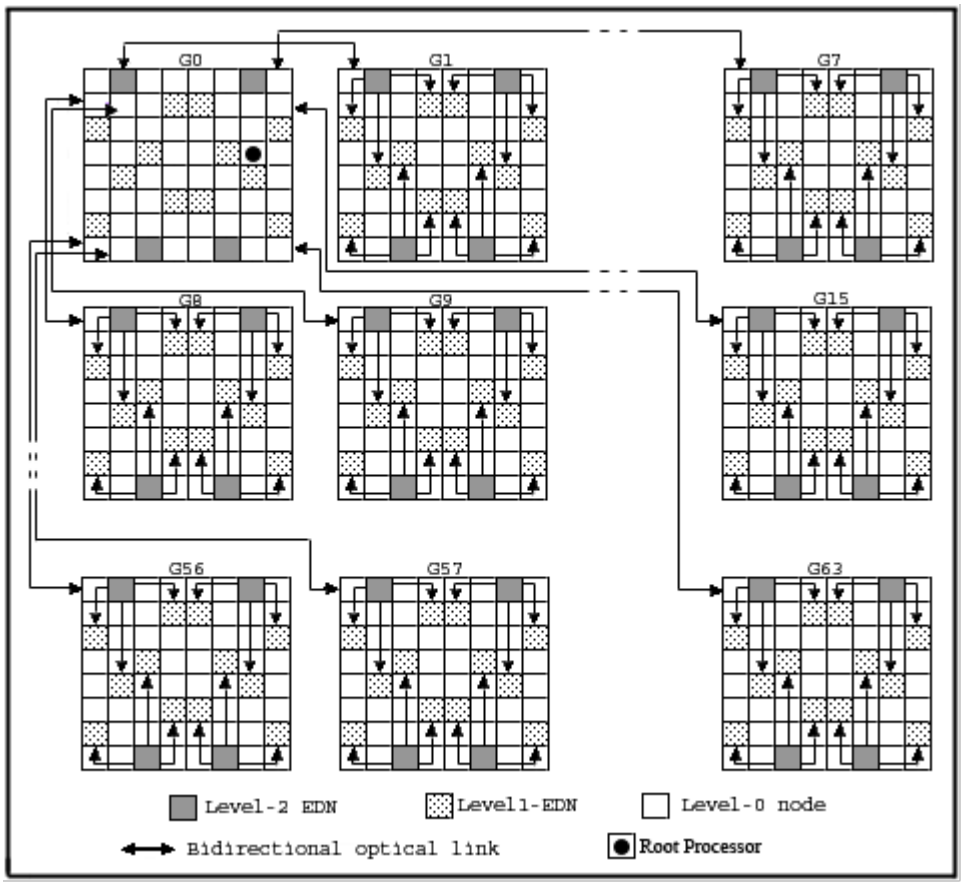

Fig. 6 All-reduce operation in 64×64 EDN-OTIS-Mesh in the distribution phase.

In all-port OTIS-Mesh and all-port EDN-OTIS-Mesh; the number of communication steps is affected by the root processor's location. So accordingly, the minimum (best-case) and maximum (worst-case) number of communication steps is calculated. The minimum number of communication steps is obtained when the root processor is located in the middle of the control group, which enables the root processor to perform simultaneous send/receive from all of its input/output channels. Whereas, the maximum number of steps is obtained when the root processor is located at the end-most of the control group, which allows the root processor to perform a limited number of simultaneous send/receive from all of its input/output channels.

In order to mathematically evaluate the single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh in terms of the minimum and maximum number of communication steps, the following assumptions hold: the *routing algorithm* used is deterministic, and the dimension order routing traverses the *X* dimension first then the *Y* dimension in OTIS-Mesh, where wormhole-routed is used. Moreover, the *number of processors* in an OTIS-Mesh group is $P = 4^k$, where $k = 2, 3\ldots n$ and 4 is the number of EDN processors in the smallest OTIS-Mesh group. However, k starts from 2 in order to have at least a group with $P = 16$ and one EDN level of 4 processors [10]. The *height of EDN tree* (i.e., the number of EDN levels) within each EDN-OTIS-Mesh group is equal to $H = (\log_4 P) - 1$, where P is the number of processors in the OTIS-Mesh group and 4 is the number of EDN processors in the smallest OTIS-Mesh group [10].

Next, (1) shows the number of communication steps to perform all-reduce operation in single-port OTIS-Mesh. Also, (2)–(5) show the minimum and maximum number of the communication steps required to perform all-reduce operation in both all-port OTIS-Mesh and all-port EDN-OTIS-Mesh.

In particular, (1) presents the number of electronic communication steps needed to perform all-reduce operation on single-port OTIS-Mesh.

$$4\times(P-1) \tag{1}$$

The number of communication steps required to perform the *reduction phase* is equal to $2\times(P-1)$ and the number of communication steps required to perform the *distribution phase* is equal to $2\times(P-1)$. Therefore, collectively the number of communication steps is equal to $4\times(P-1)$.

Equation (2) presents the minimum number of electronic communication steps needed to perform all-reduce operation on all-port OTIS-Mesh.

$$4\times((\sqrt{P}/2)\times\sqrt{P}) \tag{2}$$

The minimum number of communication steps required to perform the *reduction phase* is equal to $2\times((\sqrt{P}/2)\times\sqrt{P})$ and the minimum number of communication steps required to perform the *distribution phase* is $2\times((\sqrt{P}/2)\times\sqrt{P})$. Therefore, collectively the minimum number of communication steps is equal to $4\times((\sqrt{P}/2)\times\sqrt{P})$.

Equation (3) presents the maximum number of electronic communication steps needed to perform all-reduce on all-port OTIS-Mesh.

$$4\times((\sqrt{P}-1)\times\sqrt{P}) \tag{3}$$

The maximum number of communication steps required to perform the *reduction phase* is equal to $2\times((\sqrt{P}-1)\times\sqrt{P})$ and the maximum number of communication steps required to perform the *distribution phase* is $2\times((\sqrt{P}-1)\times\sqrt{P})$. Therefore, collectively the maximum number of communication steps is equal to $4\times((\sqrt{P}-1)\times\sqrt{P})$.

Equation (4) presents the minimum number of electronic communication steps needed to perform all-reduce operation on all-port EDN-OTIS-Mesh.

$$4\times(H+2) \tag{4}$$

The minimum number of communication steps required to perform the *reduction phase* is equal to $2\times(H+2)$ and the minimum number of communication steps required to perform the *distribution phase* is $2\times(H+2)$. Therefore, collectively the minimum number of communication steps is equal to $4\times(H+2)$.

Equation (5) presents the maximum number of electronic communication steps needed to perform all-reduce on all-port EDN-OTIS-Mesh.

$$4\times(H+3) \tag{5}$$

The maximum number of communication steps required to perform the *reduction phase* is equal to $2\times(H+3)$ and the maximum number of communication steps required to perform the *distribution phase* is $2\times(H+3)$. Therefore, collectively the maximum number of communication steps is equal to $4\times(H+3)$.

The latency of all-reduce communication operation is the time elapses from the beginning of communication until the moment the last processor finishes communication, and is given by the latency of the last processor finishes communication (i.e., the processor that has the maximum latency value). However, the latency is defined in more details in [10].

Latency improvement is the improvement gained from using the all-port EDN-OTIS-Mesh over single-port OTIS-Mesh or all-port OTIS-Mesh. In all-reduce, the latency improvement is the ratio of the latency of single-port OTIS-Mesh or all-port OTIS-Mesh over all-port EDN-OTIS-Mesh [10].

V. Simulation Results and Discussion

This section presents a detailed discussion on the obtained simulation results. The performance of the all-reduce communication operation is evaluated on the following three optoelectronic architectures: single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh, where the network switching method used for these optoelectronic architectures is wormhole switching [4], [10], [11], under the following performance metrics: number of communication steps, latency, and latency improvement.

The performance assessment of the simulation runs were conducted on a Dual-Core Intel Processor (CPU 1.5 GHz), with 14 pipeline stages and a multithreaded architecture, 2 MB L2 cache per CPU, and 2 GB RAM. The simulation was developed using C++ programming language, within the Microsoft Visual Studio 6.0 programming environment. The simulation runs were performed under Windows Vista operating system.

A. Number of Communication Steps

In this section, the minimum and maximum number of communication steps to perform all-reduce communication operation on single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh, are evaluated and compared.

Figs. 7 and 8 show that the minimum and maximum number of communication steps to perform all-reduce operation in all-port EDN-OTIS-Mesh are significantly less than in single-port OTIS Mesh and all-port OTIS-Mesh, because the EDN processors in each group increase the number of parallel send/receive operations. For example, as a best-case scenario, the all-reduce in all-port EDN-OTIS-Mesh performs about 170 times less number of communication steps than in single-port OTIS-Mesh and 85 times than in all-port OTIS-Mesh for 1024×1024 network size, respectively, as shown Fig. 7. Also, in Fig. 7 the minimum number of communication steps to perform all-reduce operation in network sizes 16×16, 64×64 and 256×256 single-port OTIS-Mesh are about as twice more as to perform it in all-port OTIS-Mesh, because the single-port OTIS-Mesh can perform one send/receive operations at a time. Moreover, in Fig. 8 the maximum number of communication steps to perform all-reduce operation in single-port OTIS-Mesh for various network sizes are slightly higher than to perform it in all-port OTIS-Mesh, because in this case the all-port OTIS-Mesh performs a limited number of parallel send/receive operations. For example, as a worst-case scenario, the all-reduce in single-port OTIS-Mesh performs about 3% more number of communication steps than in all-port OTIS-Mesh for 1024×1024 network size, as shown Fig. 8. Also, as the size of the network increases the number of communication steps to perform all-reduce in single-port and all-port OTIS-Mesh increases by at least four times from previous size, whereas in all-port EDN-OTIS-Mesh the number of communication steps increase by only four steps from previous size, as shown in Figs. 7 and 8.

B. Latency

The latency is the elapsed time from the beginning of communication until the last processor finishes communication. The latency of all-reduce communication operation is evaluated on single-port OTIS-Mesh, all-port OTIS-Mesh, and all-port EDN-OTIS-Mesh optoelectronic architectures, as both; best-case and worst-case scenarios.

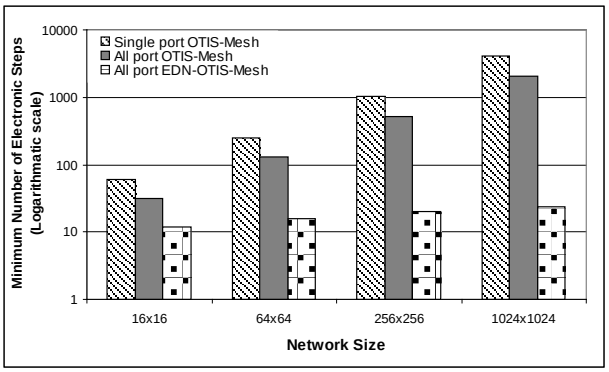

Fig. 7 Minimum number of communication steps to perform all-reduce operation.

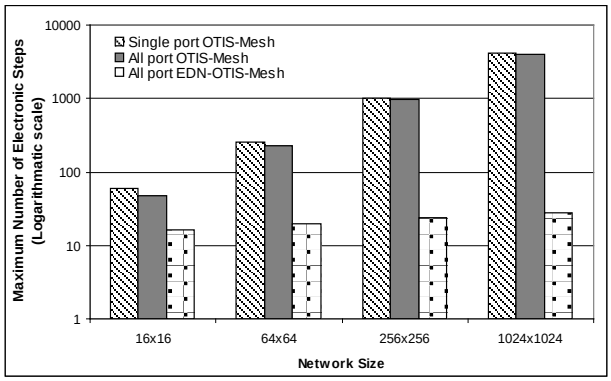

Fig. 8 Maximum number of communication steps to perform all-reduce operation.

Fig. 9 shows that the minimum latency (best-case) to perform all-reduce operation in single-port OTIS-Mesh is higher than to perform it in all-port OTIS-Mesh and all-port EDN-OTIS-Mesh; because the single-port OTIS-Mesh performs one send/receive operation at a time. Also, Fig. 10 shows that the maximum (worst-case) latency to perform all-reduce operation in single-port OTIS-Mesh and all-port OTIS-Mesh is higher than to perform it in all-port EDN-OTIS-Mesh, because the single-port OTIS-Mesh performs one send/receive operation at a time and the all-port OTIS-Mesh performs limited number send/receive operations. On the other hand, the EDN processors in the EDN-OTIS-Mesh influences the numbers of parallel send/receive operation in the reduction and distribution phases of the all-reduce operation.

In general, both Figs. 9 and 10 show that the minimum and maximum latency to perform all-reduce operation in all-port OTIS-Mesh are higher than to perform it in all-port EDN-OTIS-Mesh, because the EDN processors in the EDN-OTIS-Mesh increases the parallel send/receive operations. For example, for network size 1024×1024, the maximum latency to perform all-reduce operation in all-port OTIS-Mesh is about 80.35 milliseconds, whereas to perform it in all-port EDN-OTIS-Mesh is about 1.69 milliseconds, as shown in Fig. 10. Moreover, from Figs. 9 and 10 the latency increases as the network size increases. That is explained by the fact that large size networks such as 1024×1024 are comprised of large number of processors compared to small size networks such as 16×16. As a result, the number of send/receive operations increases and consequently the latency to perform the all-reduce operation increases as well. For instance, the minimum latency to perform all-reduce operation in 16×16 single-port OTIS-Mesh is equal to 1.13 milliseconds whereas it is 82.94 milliseconds to perform it in 1024×1024 single-port OTIS-Mesh, as shown in Fig. 9. However, this increase in latency as the network size increases is slightly small in all-port EDN-OTIS-Mesh, due to the fact that EDN processors increases the parallel send/receive operations. For instance, the minimum latency to perform all-reduce operation in 16×16 all-port EDN-OTIS-Mesh is equal to 0.34 milliseconds whereas it is 1.2 milliseconds to perform it in 1024×1024 all-port EDN-OTIS-Mesh, as shown in Fig. 9.

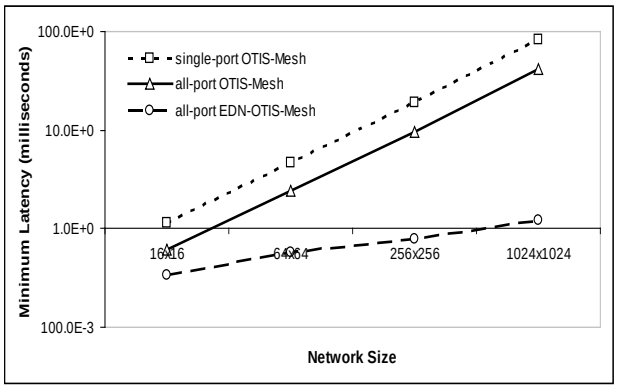

Fig. 9 Minimum latency to perform all-reduce operation.

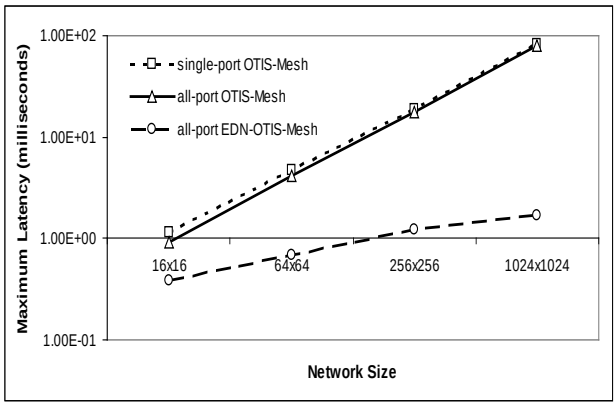

Fig. 10 Maximum latency to perform all-reduce operation.

C. Latency Improvement

The latency improvement metric (i.e., reduction in latency) is defined as the improvement achieved by the all-reduce communication operation using all-port EDN-OTIS-Mesh over single-port OTIS-Mesh or over all-port OTIS-Mesh. In particular, the latency improvement is the ratio of the latency of single-port OTIS-Mesh or all-port OTIS-Mesh over all-port EDN-OTIS-Mesh.

Fig. 11 shows that the minimum latency improvement (best-case) of the all-reduce operation on all-port EDN-OTIS-Mesh outperforms single-port OTIS-Mesh by 3.3, 8.1, 24 and 69 times for network sizes 16×16, 64×64, 256×256, and 1024×1024 respectively, and all-port OTIS-Mesh by 1.8, 4.1, 12.1 and 34.5 times for the same previous network sizes, because the all-port EDN-OTIS-Mesh performs more parallel send/receive operations. Also, Fig. 12 shows that the maximum latency improvement (worst-case) of all-reduce operation on all-port EDN-OTIS-Mesh outperformed single-port OTIS-Mesh by 2.9, 6.8, 15.3, and 49.1 times for network sizes 16×16, 64×64, 256×256 and 1024×1024 respectively, and all-port OTIS-Mesh by 2.3, 6, 9.7, and 47.6 times on the same previous network sizes, because the all-port EDN-OTIS-Mesh performs more parallel send/receive operation.

In general, Figs. 11 and 12 show that the latency improvement of performing all-reduce operation on the EDN-OTIS-Mesh over performing it on single-port OTIS-Mesh and all-port OTIS-Mesh increases when the network size increases. This is because the ratio of latency on single-port OTIS-Mesh over the latency on all-port EDN-OTIS-Mesh and the ratio of all-port OTIS-Mesh over all-port EDN-OTIS-Mesh increases as the network size increases. This is explained by the fact that the EDN-OTIS-Mesh performs more parallel send/receive operations; therefore, the latency decreases and consequently the latency improvement ratio increases.

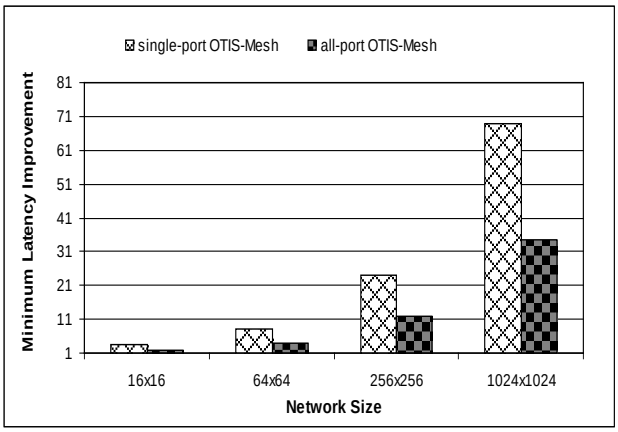

Fig. 11 Minimum latency improvement of all-reduce operation on all-port EDN-OTIS-Mesh.

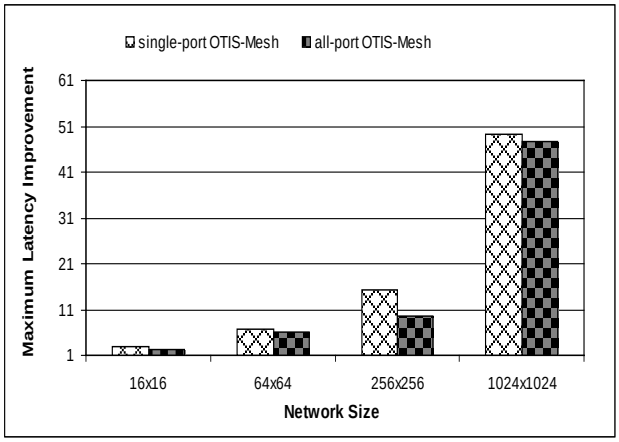

Fig. 12 Maximum latency improvement of all-reduce operation on all-port EDN-OTIS-Mesh.

VI. Conclusion and Future Work

In this paper, the performance assessment is presented mathematically and by simulation for all-reduce communication operation on single-port and all-port OTIS-Mesh in addition to all-port EDN-OTIS-Mesh optoelectronic architectures, under the following performance metrics: number of communication steps, latency, and latency improvement.

Mathematics assessment showed that the all-reduce communication operation in all-port EDN-OTIS-Mesh performed significantly better, in terms of minimum and maximum number of communication steps than both single-port and all-port OTIS-Mesh.

Simulation results showed that the all-reduce communication operation in all-port EDN-OTIS-Mesh performed significantly better, in terms of number of communication steps, latency, and latency improvement than both single-port and all-port OTIS-Mesh. The reason behind this is that the EDNs in all-port EDN-OTIS-Mesh reduces the latency as the number of EDNs increases the parallel send or receive operations, so the time to perform the operation is improved. For example, as a worst-case scenario, the all-reduce in all-port EDN-OTIS-Mesh performs about 146 and 142 times less number of communication steps than in single-port and all-port OTIS-Mesh, respectively, for 1024×1024 network size. Another example, for same network size (1024×1024), the maximum latency to perform all-reduce operation in all-port EDN-OTIS-Mesh is about 1.69 milliseconds, whereas in single-port and all-port OTIS-Mesh is about 82.94 milliseconds and 80.35 milliseconds, respectively. Therefore, for network size 1024×1024, the maximum latency improvement of all-port EDN-OTIS-Mesh over single-port and all-port OTIS-Mesh is about 49.1 and 47.6, respectively.

As future works, this work can be extended to include various collective communications operations, such as scatter, reduction, barrier, etc. on other optoelectronic architectures, such as Extended OTIS-Cube, OTIS-Hypercube, OTIS Hyper Hexa-Cell, and Optical Chained-Cubic Tree [16], [17], [22], [23]. Also, these communication operations can be evaluated and validated by applying them on different parallel and distributed algorithms; such as matrix multiplications, load balancing, TSP, and sorting.

References

[1] T. de Sousa and M. Fernandes, "Multilayer perceptron equalizer for optical communication systems," *WSEAS Transactions on Computers*, vol. 13, pp. 462–469, 2014.

[2] G. Marsden, P. Marchand, P. Harvey, and S. Esener, "Optical transpose interconnection system architectures," *Optics Letters*, vol. 18, pp. 1083–1085, 1993.

[3] J.-C. Lin, "Fault-tolerant mapping of a mesh network in a flexible hypercube," *WSEAS Transactions on Computers*, vol. 8, pp. 1587–1596, 2009.

[4] A. Grama, A. Gupta, G. Karypis, and V. Kumar, *Introduction to Parallel Computing*. Second Ed., USA: Addison Wesley, 2003, ch. 2.

[5] M. Tripathy and C. Tripathy, "A distributed shared memory cluster architecture with dynamic load balancing," *WSEAS Transactions on Computers*, vol. 11, pp. 121–130, 2012.

[6] I. Aziz, N. Haron, M. Mehat, L. Jung, A. Mustapa, and E. Akhir, "Solving traveling salesman problem on cluster compute nodes," *WSEAS Transactions on Computers*, vol. 8, pp. 1020–1029, 2009.

[7] M. Amiripour and H. Abachi, "Impact of new sub-classes of hypercube topology on execution time of matrix multiplication," *International Journal of Mathematics and Computers in Simulation*, vol. 2, pp. 118–124, 2008.

[8] Y. Tsai and P. McKinley, "An extended dominating nodes to collective communication in wormhole-routed 2D meshes," in *Proceedings of the IEEE Scalable High Performance Computing Conference*, 1994, pp. 199–206.

[9] Y. Tsai and P. McKinley, "An extended dominating node approach to broadcast and global combine in multiport wormhole-routed mesh networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 8, pp. 41–58, 1997.

[10] B. Mahafzah, R. Tahboub, and O. Tahboub, "Performance evaluation of broadcast and global combine operations in all-port wormhole-routed OTIS-mesh interconnection networks," *Cluster Computing*, vol. 13, pp. 87–110, 2010.

[11] B. Mahafzah, S. Serhan, and R. Tahboub, "All-reduce communication operation in OTIS-mesh interconnection network," in *Proceedings of the 14th International Conference on Software Engineering, Parallel and Distributed Systems (SEPADS '15)*, Dubai, United Arab Emirates, 2015, pp. 63–70.

[12] Y. Tsai and P. McKinley, "A dominating set model for broadcast in all-port wormhole-routed 2D mesh networks," in *Proceedings of the Eighth ACM International Conference on Supercomputing*, 1994, pp. 126–135.

[13] J. Pjesivac-Grbovic, T. Angskun, G. Bosilca, G. Fagg, E. Gabriel, and J. Dongarra, "Performance analysis of MPI collective operations," *Cluster Computing*, vol. 10, pp. 127–143, 2007.

[14] M. Matsuda, T. Kudoh, Y. Kodama, R. Takano, and Y. Ishikawa, "The design and implementation of MPI collective operations for clusters in long-and-fast networks," *Cluster Computing*, vol. 11, pp. 45–55, 2008.

[15] A. Krishnamoorthy, P. Marchand, F. Kiamilev, and S. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks," *Applied Optics*, vol. 31, pp. 5480–5507, 1992.

[16] J. Al-Sadi, "Broadcasting and routing algorithms for the extended OTIS-cube network," *International Journal of Communications*, vol. 5, pp. 95–102, 2011.

[17] H. Najaf-Abadi and H. Sarbazi-Azad, "An empirical comparison of OTIS-mesh and OTIS-hypercube multicomputer systems under deterministic routing," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05)*, vol. 15, 2005.

[18] B. Mahafzah and B. Jaradat, "The load balancing problem in OTIS-hypercube interconnection networks," *Journal of Supercomputing*, vol. 46, pp. 276–297, 2008.

[19] C. Zhao, W. Xiao, and B. Parhami, "Load-balancing on swapped or OTIS networks," *Journal of Parallel and Distributed Computing*, vol. 69, pp. 389–399, 2009.

[20] W. Wei and W. Xiao, "Algorithms of basic communication operation on the biswapped network," in *Proceedings of the 8th International Conference on Computational Science (ICCS 2008)*, Part I, Lecture Notes in Computer Science, vol. 5101, Springer-Verlag, Berlin, 2008, pp. 347–354.

[21] P. McKinley, Y. Tsai, and D. Robinson, "Collective communication in wormhole-routed massively parallel computers," *Computer*, vol. 28, pp. 39–50, 1995.

[22] B. Mahafzah, A. Sleit, N. Hamad, E. Ahmad, and T. Abu-Kabeer, "The OTIS hyper hexa-cell optoelectronic architecture," *Computing*, vol. 94, pp. 411–432, 2012.

[23] B. Mahafzah, M. Alshraideh, T. Abu-Kabeer, E. Ahmad, and N. Hamad, "The optical chained-cubic tree interconnection network: Topological structure and properties," *Computers & Electrical Engineering*, vol. 38, pp. 330–345, 2012.