TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 3, ISSUE 3
ISSN: 2054 - 7420

The Non-Uniform Communication Performance of Adaptive Routing for Hierarchical Interconnection Network for 3D VLSI

[1]Yasuyuki Miura, [2]Shigeyoshi Watanabe, [3]M.M. Hafizur Rahman
[1,2]School of Information Technology, Shonan Institute of Technology, Fujisawa, Japan;
[3]International Islamic University, Malaysia (IIUM), Kuala Lumpur, Malaysia;
miu@info.shonan-it.ac.jp; watanabe@info.shonan-it.ac.jp; hafizur@iium.edu.my

ABSTRACT

The Tori-connected mESH (TESH) Network is a k-ary n-cube networks of multiple basic modules, in which the basic modules are 2D-mesh networks that are hierarchically interconnected for higher level k-ary n-cube networks. Many adaptive routing algorithms for k-ary n-cube networks have already been proposed. Thus, those algorithms can also be applied to TESH network. We have proposed three adaptive routing algorithms - channel-selection, link-selection, and dynamic dimension reversal - for the efficient use of network resources of a TESH network to improve dynamic communication performance.

In this paper, we have evaluated the dynamic performance of a TESH network using different non uniform traffic patterns. In this paper, we have evaluated by local communication traffic pattern in addition to the hotspot, perfect shuffle, and complement traffic patterns. It was shown that the dynamic communication performance was improved when inter-BM communication appeared frequently such as perfect shuffle and local communication traffic patterns.

Keywords: TESH network, adaptive routing algorithm, communication performance

1 Introduction

Interconnection networks are the key elements for building massively parallel computers consisting of hundreds or thousands of processors[1]. Recent progress in very-large-scale integration (VLSI) and network-on-chip (NoC) technology has led to multicomputer systems on three-dimensional LSIs.

Through Silicon Via (TSV) is widely used in 3D-LSI implementation. Like the via of printed circuit board, TSV pierces between silicon chips. By the inter-chip connection through TSV, the complex 3D VLSI implementation can be realized. However, in the LSI structure, large amount of layout area is necessary for inter-chip connection. So, the number of inter-chip connections should be reduced. Using 30-40 nm CMOS gate length, the length of a TSV is several μm. Based on this issue, we have been studied the hierarchical interconnection network which can reduce the number of wires.

A Tori connected mESH (TESH) network [2]-[6] is a hierarchical interconnection network for large-scale on-chip multicomputers. It consists of multiple basic modules (BMs) which are 2D-mesh networks and the BMs are hierarchically interconnected by a 2D-torus (*k*-ary 2-cube) to build higher level networks. Such networks with hierarchical structure assumed to be implementing in 3D-VLSI.

DOI: 10.14738/tnc.33.1293
Publication Date: 02nd July 2015
URL: http://dx.doi.org/10.14738/tnc.33.1293

On the other hand, restricted use of physical links between silicon planes reduced performance of the TESH network. We have proposed a deterministic, dimension-order routing algorithm[7]-[10] for the TESH network and have shown that Level-3 TESH networks have higher performance than a *k*-ary 2-cube. The minimum number of virtual channels[11] per link in dimension order routing has been proven to be two[7].

An adaptive routing algorithm can also be implemented using additional virtual channels. Based on adaptive routing algorithms of *k*-ary *n*-cube[12]-[16], we have proposed three adaptive routing algorithms for TESH[17]-[20]. In [17], three adaptive routing algorithms were proposed, and those algorithms were implemented by HDL (Hardware Description Language)[18]-[20]. By those studies, time delay and hardware cost for implement different adaptive routing algorithm were evaluated. It was shown that the time delay of adaptive routing algorithm were almost same as that of deterministic dimension order routing.

In our previous studies, we have evaluated various traffic patterns for the performance evaluation using dimension order routing. In [20] and [21], we have evaluated the hotspot and two types of non-uniform traffic patterns of *Bit Permutation and Communication* (BPC) traffic patterns. The main objective of this paper is evaluate the dynamic communication performance of a TESH network under different non-uniform traffic patterns using our previously proposed adaptive routing algorithms. In addition to the hotspot and two types of BPC traffic patterns (perfect shuffle and complement), we consider the local traffic patterns to show the suitability of our proposed adaptive routings on a TESH network.

The remainder of the paper is organized as follows. We briefly describe the basic structure of the TESH network and dimension order routing algorithm on it in Section 2 and 3, respectively. Different adaptive routing algorithms and their time delay and hardware implementation is discussed in Section 4. The dynamic communication performance of the TESH network using these adaptive routing algorithms under the various traffic patterns is discussed in Section 5. Finally, Section 6 concludes this study.

2 Structure of the TESH network

The Tori-connected mESH (TESH) Network is a hierarchical interconnection network consisting of Basic Modules (BM) that are hierarchically interconnected to form a higher level network. The BM of the TESH network is a 2D-mesh network of size $2^m \times 2^m$. In this paper, unless specified otherwise, BM refers to a Level-1 network. Successively higher level networks are built by recursively interconnecting immediately lower level subnetworks in a 2D-torus network. A higher-level network is built using immediate lower level networks as subnet modules as a 2D-torus network of size $2^m \times 2^m$[2]. Here *m* is a positive integer and in this paper, we have considered *m*=2 i.e., 4×4 2D-mesh as BMs and 4×4 2D-torus (*k*-ary 2-cube) as higher level networks as shown in Figure 1.

A $2^m \times 2^m$ BM has 2^{m+2} free ports at the contours for higher level interconnection. For each higher level interconnection, a BM uses $4 \times 2^q = 2^{q+2}$ of its free links, 2×2^q free links for vertical interconnections and 2×2^q free links for horizontal interconnections. Here, $q \in \{0,1,\cdots,m\}$ is the inter-level connectivity. *q*=0 leads to minimal inter-level connectivity, while *q*=*m* leads to maximum inter-level connectivity. Considering the size of the basic module *m*, level of hierarchy *n*, and inter-level connectivity *q*, we can define the TESH network as TESH(*m*,*L*,*q*) networks. Since we have considered *m* = 2, a Level-2 network, can be formed by interconnecting $2^{2\times2}$=16 BMs. Similarly, a Level-3 network can be formed by interconnecting 1 Level-2 subnetworks, and so on. Each BM is

connected to its logically adjacent BMs. To avoid clutter, the wraparound links of the BMs are not shown in Figure 1. In the rest of this paper we consider m=2, therefore, we focus on a class of TESH(2,L,q) networks.

The highest level network which can be built from $2^m \times 2^m$ BM is $L_{\max} = 2^{m-q} + 1$. With *m*=2 and *q*=0, $L_{\max} = 2^{2-0} + 1 = 5$. The total number of nodes in a TESH network is $N = 2^{2mL}$. Using maximum level of hierarchy, $L_{\max} = 2^{m-q} + 1$, the maximum number of nodes which can be interconnected by a TESH(*m*,*L*,*q*) is $N = 2^{2m(2^{m-q}+1)}$. With *m*=2 a Level-2 TESH network consists of 256 nodes. Similarly a Level-3 networks consists of 4096 nodes.

Processing elements (PEs) or node in a TESH(*m*,*L*,*q*) network are addressed using base-2^m numbers as follows.

$$\begin{aligned} n &= n_{2L-1}\, n_{2L-2} \cdots n_3 n_2 n_1 n_0 \\ &= (n_{2L-1}\, n_{2L-2}) \cdots (n_3 n_2)(n_1 n_0) \end{aligned} \tag{1}$$

Here, $(n_{2i-1}\, n_{2i-2})$ is the location of a subnetwork at level *i*-1. For example, in a Level-3 TESH with *m*=2, each PE is addressed by a base-4 number $n = n_5 n_4 n_3 n_2 n_1 n_0$, where n_5 and n_4 address of a PE in the Level-3 network, n_3 and n_2 address of a PE in the Level-2 network, and n_1 and n_0 address of a PE in the BM. The numerical values shown in Figure 1 are at BM address $n_3 n_2$ of a Level-2 TESH network.

The assignment of free ports for inter-level connections for the higher level networks has been done quite carefully so as to minimize the higher level traffic through the BM. The address of a node n^1 encompasses in BM_1 is represented as $n^1 = n^1_{2L-1} n^1_{2L-2} \cdots n^1_3 n^1_2 n^1_1 n^1_0$. The address of a node n^2 encompasses in BM_2 is represented as $n^2 = n^2_{2L-1} n^2_{2L-2} \cdots n^2_3 n^2_2 n^2_1 n^2_0$. The node n^1 in BM_1 and n^2 in BM_2 are connected by a link if the following condition is satisfied.

$$\exists_i \left\{ n_i^1 = (n_i^2 \pm 1) \bmod 2^m \cap \ \forall_j (j \neq i \ \rightarrow \ n_i^1 = n_i^2) \right\} \tag{2}$$

where $i, j \geq 2$.

It is shown in Figure 1 that for a Level-2 TESH network, BM(0,0) connects with either BM(0,1) or BM(0,3) in the *x*-direction, i.e., ($n_3 = 0$ and $n_2 = 1$ or $n_2 = 3$), and with either BM(1,0) or BM(3,0) in the *y*-direction, i.e., ($n_2 = 0$ and $n_3 = 1$ or $n_3 = 3$).

This hierarchical interconnection of the TESH network has the following possible implementation.

- A BM is laid out in one VLSI chip and the BMs are connected by TSV. In this way Level-2 TESH network can be realized.
- A Level-2 TESH network is laid out in one chip and then they are connected by TSV. In this way Level-3 TESH network can be realized. In a Level-2 TESH network on a chip, the yield of the chip and fault tolerance is improved by carrying out hierarchical reconfiguration algorithm. The improvement of yielding in an array network using hierarchical reconfiguration algorithm is depicted in [22].

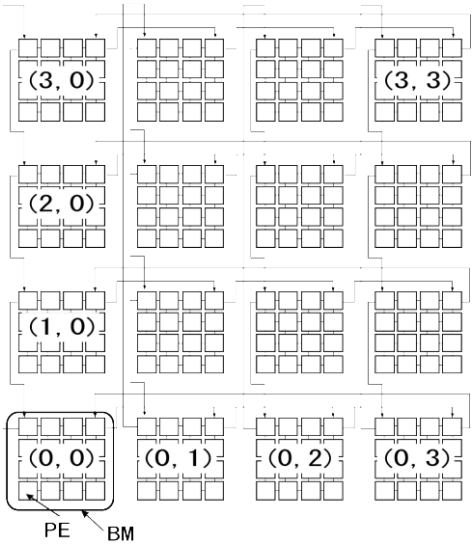

Figure 1: Hierarchical interconnection of Level-2 TESH network.

3 Dimension-Order Routing

A routing algorithm determines the path a packet takes as it travels through the network from its source to destination. A deterministic, dimension-order routing for a TESH network[7] transfers packets from higher-levels to lower-levels. That is, it is first done at the highest level network; then, after the packet reaches its highest level sub-destination, routing continues within the subnetwork to the next lower level sub-destination. This process is repeated until the packet arrives at its final destination. Packets passing through inter-BM links are forwarded from a vertical link to a horizontal link at the same level. When they arrive at the destination BM, they are transferred to the destination PE. The deterministic, dimension-order routing of Level-*L* TESH networks can generally be classified into the following three phases [7].

Phase 1: Intra-BM transfer path from source node to the outlet PE of the BM.

(PEs which hold $n_1 =$ 0,3 or $n_0 = 0,3$)

Phase 2: Higher level transfer path.

Phase 3: Intra-BM transfer path from the outlet of the inter-BM transfer path to the destination PE.

Phase 2 is divided into the following sub-phases:

sub-phase 2.*i*.1: Intra-BM transfer to the outlet PE of Level (*L*-*i*) through the *y*-link/vertical link.

sub-phase 2.*i*.2: Inter-BM transfer of Level (*L*-*i*) through the *y*-link/vertical link.

sub-phase 2.*i*.3: Intra-BM transfer to the outlet PE of Level (*L*-*i*) through the *x*-link/horizontal link.

sub-phase 2.*i*.4: Inter-BM transfer of Level (*L*-*i*) through the *x*-link/horizontal link.

Here, $0 \leq i \leq L-2$.

We have considered the dimension order routing algorithm for the TESH network. We use the following strategy: at each level, vertical routing is performed first. Once the packet reaches the correct row, then horizontal routing is performed. Routing in the TESH network is strictly defined by the source node address and the destination node address. Let a source node address be $s = (s_{2L-1}\ s_{2L-2}) \cdots (s_3\ s_2)\ (s_1\ s_0)$ and a destination node address be

$d = (d_{2L-1}\ d_{2L-2}) \cdots (d_3\ d_2)\ (d_1\ d_0)$. The dimension-order of the TESH network is represented by R_1. Figure 2 shows the routing algorithm R_1 for the TESH network. The function *get_group_number* is the function to get group number. Arguments of this function are source PE address *s* destination PE address *d*, and direction. The function $\text{outlet}_x(g, l, d\delta)$ and $\text{outlet}_y(g, l, d\delta)$ are the function to get the value n_0 of *x* coordinate and n_1 of *y* coordinate of a PE *n* that has an inter-BM link. Variables $g, l, d\delta$ are group g $(1 \le g \le 2^q)$, level l $(1 \le l \le L)$, dimension d $(d \in \{vertical, horizontal\})$, and direction δ $(\delta \in \{+, -\})$, respectively[7].

```
/*Routing Algorithm for a Level-L TESH:*/

Routing(s, d)
s[2L]; /* source */
d[2L]; /* destination */
{

  group = get_group_number(s,d);


  for(i = 2L-1; i > 2; i--){

    if((d[i]-c[i]+2^m) mod 2^m <= (2^m)/2) dir = PLUS;
    else dir = MINUS;

    while(d[i] != c[i]){

      if(i is even number){
 outlet[0] = outlet_x(group,i/2+1,H,dir);
 outlet[1] = outlet_y(group,i/2+1,H,dir);}
      if(i is odd number){
 outlet[0] = outlet_x(group,i/2+1,V,dir);
 outlet[1] = outlet_y(group,i/2+1,V,dir);}
      if(outlet_node_x != c[0] or outlet_node_x != c[1])
        BM_routing(c, outlet);

      send_packet(dir);

  }

 }

  BM_routing(c, outlet);

}


BM_routing(c, outlet)
c[2];       /* current channel */
outlet[2]; /* outlet node */
{

  while(c[1] != outlet[1]){
    if(outlet[1] > c[1]) send_packet(UPPER);
    if(outlet[1] < c[1]) send_packet(LOWER);
 }
  while(c[0] != outlet[0]){
    if(outlet[0] > c[0]) send_packet(RIGHT);
    if(outlet[0] < c[0]) send_packet(LEFT);
 }

}
```

Figure 2: Routing algorithm of a TESH network.

4 Adaptive Routing Algorithm

Adaptive routing algorithms for TESH networks are classified into two groups: local and global algorithms. Local algorithms are defined as adaptive routing algorithms that run in one phase and global algorithms are defined as algorithms for which the order of phases can be changed.

In this section, we introduce two local adaptive routing algorithms called as channel select (CS) and link select (LS); and one global adaptive routing algorithm called dynamic dimension reversal (DDR)[17].

The deadlock in a k-ary n-cube network can be avoided using 2 virtual channels using following two conditions [14].

- Condition 1: Initially, first virtual channel (Channel-L) is used.
- Condition 2: Then the packet move to the second virtual channel (Channel-H) if the wraparound links is used for routing.

Local algorithms for TESH network are applied in sub-phases 2.*i*.2 or 2.*i*.4 in section 3. Because a ring network is formed using 4 outlet PEs (*m*=2) and inter-BM links. Local adaptive routing algorithms can be applied in this ring of TESH network. To discuss local adaptive routing algorithms, we allocate a local PE address to each of those four PEs. Let n_{local} be the local PE addresses of a ring network in the TESH. Then, n_{local} are addressed as follows:

$$n_{\text{local}} = \begin{cases} n_{2l-1}, & \text{Level} - l \text{ vertical link} \\ n_{2l-2}, & \text{Level} - l \text{ horizontal link.} \end{cases} \quad (3)$$

where n_{2l-1} and n_{2l-2} are the PE address defined in section ref{addrnd}. Below, we discuss two local algorithms for a 4-PE ring network in TESH by using n_{local}.

Since the higher-level links of TESH have *k*-ary *n*-cube network, adaptive routings of *k*-ary *n*-cube can be applied to TESH. Dynamic dimension reversal routing (DDR)[14] is proposed as adaptive routing algorithms of *k*-ary *n*-cube network. This algorithm has a lot of choice of the path and needs a few additional virtual channels. The DDR routing can also be applied to TESH network. However, unlike conventional *k*-ary *n*-cube network, the higher-level links are located in different PE in the TESH network. This is why, the choice of routing path in the TESH network is limited in comparison with *k*-ary *n*-cube network.

4.1 Channel Select (CS) Algorithm

To avoid deadlock in a ring network, two virtual channels (Channel-L and Channel-H) are needed in each direction. The CS algorithm is an adaptive routing algorithm that can use those channels freely. When wraparound channels are not used in routing, for example in the routing from PE($n_{\text{local}} = 0$) to PE($n_{\text{local}} = 2$), only Channel-L is used. In this case, because Channel-H is not used in dimension-order routing, it is possible to move from Channel-L to Channel-H or use Channel-H initially. When the routing is terminated at the output PE of a wraparound channel such as the routing from PE($n_{\text{local}} = 2$) to PE($n_{\text{local}} = 0$), only Channel-L is required. Therefore, it is possible to move from Channel-L to Channel-H or use Channel-H initially.

For dimension-order routing in a 4-PE ring network, the conditions for using only Channel-L are as follows:

- Wraparound channels are not used in routing.
- The routing is terminated at the output PE of a wraparound channel.

When the above conditions hold, the virtual channels are used according to the following order:

- Either Channel-L or Channel-H is used.
- The packet moves from Channel-L to Channel-H in the routing path.

The CS algorithm is an adaptive routing algorithm that can use virtual channels effectively. When the following three conditions hold in a 4-PE bidirectional ring network is deadlock-free[17]. The routing algorithm that applies this channel-selection principle is denoted as R_2.

Condition-1: Use Channel-L initially.

Condition-2: Use Channel-H when a wrap-around link exists in the higher level network.

Condition-3: When a packet is in a Channel-L satisfies either of the following conditions, it can move to Channel-H.

- Wrap-around links are not used in routing.
- The routing will be terminated at the output PE of a wraparound link.

4.2 Link Select (LS) Algorithm

Sub-phases 2.*i*.2 and 2.*i*.4 form a ring network. If the number of hops from the source PE to the destination PE is equal in the clockwise direction and in the counter-clockwise direction, then the packet can follow either of these two directions. The distance from $\mathrm{PE}_0(n_{\mathrm{local}}=0)$to $\mathrm{PE}_2(n_{\mathrm{local}}=2)$ in a 4-PE ring network is 2 in both the clockwise and counter-clockwise direction. Packet can follow path-a in the clockwise direction or path-b in the counter-clockwise direction.

If the following equation is satisfied, a packet can select from either a clockwise or counter-clockwise direction.

$$|s-d|=\frac{2^m}{2} \tag{4}$$

where *s* and *d* denote the source and destination PE addresses, respectively. The routing algorithm that applies this link-selection principle is denoted as R_3.

The CS algorithm is used to select a virtual channel in a physical link and the LS algorithm is used to select a physical link in a network. Therefore, both the CS and LS algorithms can be applied at the same time.

4.3 Dynamic Dimension Reversal (DDR) Algorithm

The dimension-order routing strictly maintain the restriction of routing dimension in an interconnection network, such as *k*-ary *n*-cubes. In the dimension-order routing for the TESH network the order of routing phases is fixed. However, an algorithm that can break the dimension order has already been proposed[14]. In this paper, the Dimension Reversal (DR) routing algorithm is applied in the TESH network. Dimension reversal routings of *k*-ary *n*-cubes are classified into two types: Static Dimension Reversal and Dynamic Dimension Reversal. Because Dynamic Dimension Reversal routing (DDR) can use channels efficiently, we apply DDR to a TESH. We called it as global adaptive routing algorithm. The DDR algorithm can be applied individually and simultaneously with the CS and LS algorithms.

In the DDR algorithm, each packet has a DR number, which is a count of the number of times that a packet has been routed from a channel in sub-phase 2.*p* to a channel in a lower-order sub-phase 2.*q*, *q* < *p*. Here, the format of *p* and *q* are $p_1.p_0$ and $q_1.q_0$. We assume that $p_1$and q_1 are the high-order digits and p_0 and q_0 are the low-order digits when *p* and *q* are compared. DR numbers are assigned as follows:

1. All packets are initialized with a DR of 0.
2. If a a packet routes from a channel c_i of sub-phase 2.*p* to a channel c_j of sub-phase 2.*q*, then its DR is incremented.

The DDR algorithm divides the virtual channels into two classes: adaptive and deterministic. Packets originate in the adaptive channels and while they are in adaptive channels, they may be routed by adaptive routing. Whenever a packet acquires a channel, it labels the channel with its DR number. To avoid deadlock, a packet with a DR of *p* need not to wait for a channel labeled with a DR of *q* if $p \geq q$. A packet that reaches a node where all output channels are occupied by packets with equal or lower DR numbers must switch to the deterministic channels. When a packet enters the deterministic channels, it must be routed by dimension-order routing and cannot re-enter the adaptive channels.

The adaptive routing algorithm for the adaptive channel is as follows. A packet with DDR routing of a *k*-ary *n*-cube network can be routed in any direction using adaptive channels. However, since a TESH is a hierarchical network, higher-level links of a *k*-ary *n*-cube network, where each node of this *k*-ary *n*-cube is not located in the same BM. Therefore, the packet cannot be routed freely.

There are four outlet PEs in an inter-BM links in each BM as shown in Figure 1. When the packet goes through those PEs during intra-BM routing, it can select a path from the following ones:

- Path 1: Interrupt the intra-BM routing and select the inter-BM link.
- Path 2: Continue the intra-BM routing.

When the above conditions hold, the packet selects path-1 first.

An example of the DDR algorithm applied to a TESH network is shown in Figure 3. The half-tone PE is the source PE, and the solid arrow in the BM is the deterministic routing path. In this example, we assume that a packet goes through first in a Level-3 vertical link and next in a Level-3 horizontal link. In the dimension-order routing of a TESH network, a packet is forwarded to the outlet PE of a Level-3 vertical link in phase 1 routing. However, in the DDR routing as shown in Figure 3 the packet goes through the outlet PE of a Level-3 horizontal link. When the packet reaches the outlet PE, it checks whether the Level-3 horizontal link is available or not. If it is available, the packet selects Path-1 and uses the Level-3 horizontal link before going to the outlet PE of the Level-3 vertical link. If it is not available, the packet selects Path-2 and continues with the phase 1 transfer.

The routing algorithm that applies DDR is denoted as R_4. The CS, LS, and DDR algorithms applies in the different resources of a network. Thus, these algorithms can be applied simultaneously to a network.

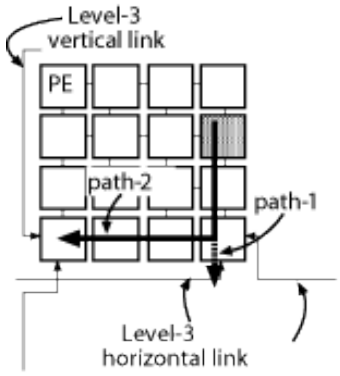

Figure 3: DDR algorithm in a TESH.

4.4 The Time Delay with Hardware Implementation

In our previous study, we have implemented the router by using VHDL language and Xilinx Project Navigator synthesis tool [18]-[20]. We have evaluated the time delay with hardware implementation

using 4 virtual channels as tabulated in Table 1. Dimension order routing (DOR) is widely used in the contemporary massively parallel computers. As we have seen from Table 1, the time delay of DDR is almost same as that of DOR. Thus, we can conclude that DDR is practically implementable.

According to the clock cycle driven in [20], we evaluated the dynamic communication performance and plotted in Figure 4. The clock time in Table 1 are used for the evaluation of each routing algorithms. One cycle time is the clock time of each algorithm in Table 1. The average transfer time as a function of network throughput is portrayed in Figure 4 using uniform traffic pattern for four channels. The horizontal axis indicate network throughput, i.e., the average number of flits delivered through the network per unit time. The throughput is the number of delivered flits per PE in 1 ns, i.e., Flits / PE·ns.

As shown in Figure4, the maximum throughput of the CS and LS algorithms on a TESH network is noticeably higher than that of the dimension-order routing. In the CS algorithm, there is no influence of channel selection circuit delay. In the LS algorithm, the delay is slightly high. However, the difference is trivial. The maximum throughput using normal implementation[20] of DDR algorithm is lower than that of other algorithms. Due to complicated routing principle of DDR algorithm, its link selection circuit delay is large, which in turns make the clock cycle time of DDR algorithm long. On the other hand, the maximum throughput of parallel-implemented[20] DDR algorithm is higher than that of other algorithms.

Table 1: The Time Delay of Routing Algorithms with 4 channels (ns).

Algorithm	Cycle Time (ns)
Dimension Order	11.86
CS	11.86
LS	12.89
DDR	11.87

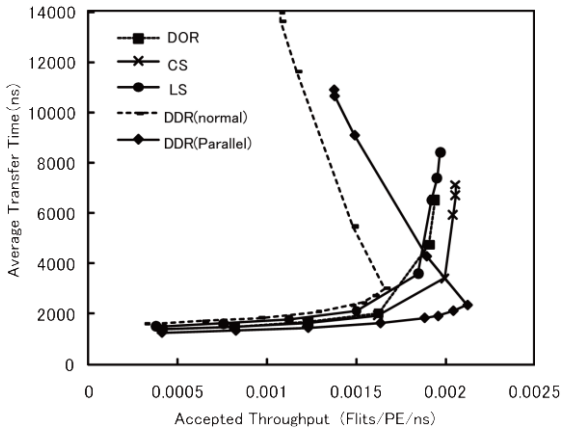

Figure 4: Comparison of dynamic communication performance of the TESH(2,3,0) network using hardware implemented router between dimension-order, CS, LS, and DDR algorithms with uniform traffic pattern: 4096 nodes, 4 VCs, and 16 flits.

5 Performance Evaluation

5.1 Simulation Environment

We have developed a wormhole routing simulator to evaluate dynamic communication performance of the TESH(2,3,0) network with 4096 PE. Dynamic communication performances are simulated for

dimension-order routing algorithm, CS algorithm, LS algorithm, DDR algorithm, and combinations of them.

Extensive simulations have been carried out for the following traffic patterns.

- uniform
- hotspot
- perfect shuffle
- complement
- local communication

The dynamic communication performance of an interconnection network is characterized by message latency and network throughput. Message latency refers to the time elapsed from the instant when the first flit is injected into the network from the source to the instant when the last flit of the message is received at the destination. Average transfer time is the average value of the latency for all packets. Network throughput refers to the maximum amount of information delivered per unit of time through the network. It is the average value of the number of flits which a PE receives in each clock cycle. In the evaluation of dynamic communication performance, flocks of messages are sent in the network to compete for the output channels. Packets are transmitted by the request-probability *r* during *T* clock cycles and the number of flits which reached at destination PE and its transfer time are recorded. Then the average transfer time and throughput are calculated and plotted as average transfer time in the horizontal axis and throughput in the vertical axis. The process of performance evaluation is carried out with changing the request-probability *r*.
The packet size is 16 flit and flits are transmitted for 20,000 cycles, i.e., *T*=20000. In each clock cycle, one flit is transferred from the input buffer to the output buffer, or from output to input if the corresponding buffer in the next node is empty. Therefore, transferring data between two nodes takes 2 clock cycles. Four virtual channels per physical channel are simulated, and they are arbitrated by a round-robin algorithm. The buffer length of each channel is 2 flits. In the DDR algorithm, the number of deterministic channel and adaptive channel are two respectively. In other algorithms, a pair of channels is used for deadlock avoidance, and two pairs are used to select channels. The transfer method of the packet is wormhole routing[23].

5.2 Uniform Traffic Pattern

In a uniform traffic, destinations are chosen randomly with equal probability among the nodes in the network. The result of uniform traffic was shown in [20].

Figure 5 shows the average transfer time as a function of network throughput[20]. The horizontal axis indicates network throughput and the vertical axis indicates transfer time. Also, the number inside the figure is the request-probability r. From Figure5, the throughput of CS, LS, and DDR algorithms and those combinations are higher than DOR.

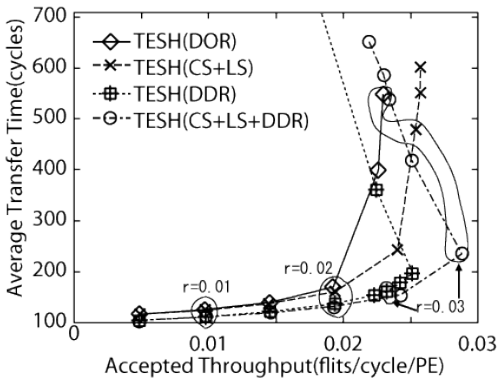

Figure 5: Comparison of dynamic communication performance of the TESH(2,3,0) network between dimension-order, CS+LS, DDR, CS+LS+DDR algorithms with uniform traffic pattern: 4096 nodes, 4 VCs, and 16 flits.

5.3 Hotspot Traffic Pattern

In a uniform traffic, destinations are chosen randomly with equal probability among the nodes in the network. The result of hotspot traffic was also shown in [20].

Figure 6 shows the average transfer time as a function of network throughput[20]. As depicted in Fig.6, the DDR algorithm has a low average transfer time in comparison with the other algorithms at zero-load like uniform traffic pattern. Also the maximum throughput of the DDR algorithm under hot-spot traffic pattern is higher than that of other algorithms. Because the choice of the inter-BM link is more than that of CS and LS algorithms. In this experiment, higher-level links of the neighborhood of destination PE are congested because hotspot packets use higher-level links intensively. Global adaptive routing algorithm such as DDR algorithm has an advantage in hot-spot traffic pattern because a lot of inter-BM links can be selected as compared with local adaptive routing. Therefore, with the hot-spot traffic pattern, the DDR algorithm yields better dynamic communication performance than that of dimension-order, LS, and CS algorithms.

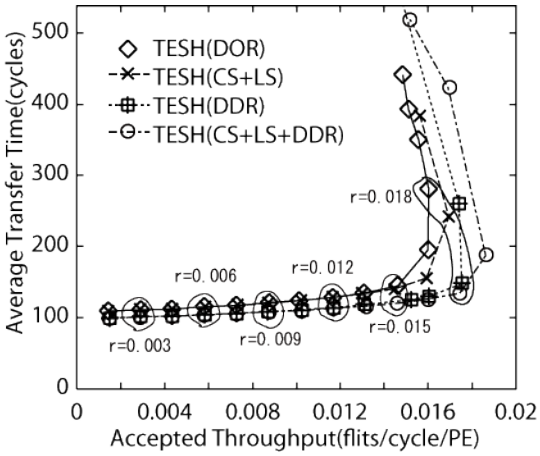

Figure 6 Comparison of dynamic communication performance of the TESH(2,3,0) network between dimension-order, CS+LS, DDR, CS+LS+DDR algorithms with hot-spot traffic pattern: 4096 nodes, 4 VCs, and 16 flits.

5.4 Bit Permutation and Communication (BPC) Traffic Pattern

Bit Permutation and Computation (BPC) [24] is a class of non-uniform traffic pattern, which are very common in scientific applications. BPC communication patterns take into account the permutations that are usually performed in parallel numerical algorithms [25][26]. These distributions achieve the maximum degree of temporal locality and are also considered as benchmarks for interconnection

networks. Among various BPC traffic patterns, in this paper, we have considered complement and perfect shuffle traffic pattern.

5.4.1 Complement Traffic Pattern

In the complement traffic pattern, the source PE $n_s = (n_{2L-1}\, n_{2L-2}) \cdots\cdots (n_3 n_2)(n_1 n_0)$ sends packet to the destination PE

$n_d = \overline{n_s} = (\overline{n_{2L-1}}\, \overline{n_{2L-2}}) \cdots\cdots (\overline{n_3}\, \overline{n_2})(\overline{n_1}\, \overline{n_0})$. In this traffic pattern, all the packets cross the bisection of the network. All communications are inter-level which creates the congestion in the inter-BM links.

Considering this congested scenario, we have evaluated by simulation the dynamic communication performance of the TESH network using DOR, LS, CS, and DDR algorithm and the result is plotted in Figure 7. From Figure 7, it is seen that dynamic communication performance of the CS and LS algorithm is almost similar to that of DOR algorithm. Due to the congestion in the middle of the network, the performance improvement by CS and LS algorithm is limited. It is also shown that the average transfer time and throughput under DDR algorithm is slightly improved than that of other algorithm.

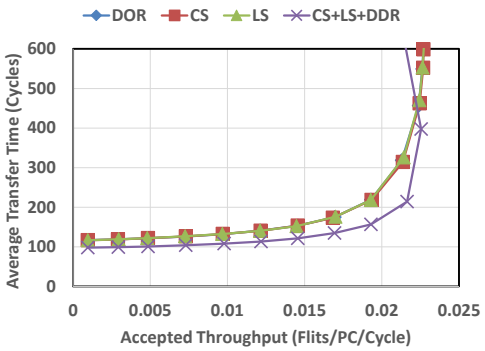

Figure 7: Comparison of dynamic communication performance of the complement traffic pattern on the DOR and adaptive routings for TESH(2,3,0) network: 4096 nodes, 4 VCs, and 16 flits.

5.4.2 Perfect Shuffle Traffic Pattern

Let the source PE address be n_s and the destination PE address be n_d. According to the perfect shuffle traffic pattern the destination PE n_d is determined as follows:

Let the source PE address be n_s and the destination PE address be n_d. According to the perfect shuffle traffic pattern the destination PE n_d is determined as follows:

$$n_d = \begin{cases} n_s \times 2 & (\ n_s < N/2) \\ (\ n_s - N/2) \times 2 + 1 & (\ n_s \geq N/2) \end{cases} \quad (5)$$

We have evaluated the dynamic communication performance of the TESH network using DOR, LS, CS, and DDR algorithm under perfect shuffle traffic pattern and the result is plotted in Figure 8. From Figure 8, it is shown that the throughput is considerably improved by DDR algorithm and slightly improved by LS algorithm. In the perfect shuffle traffic, the probability of packet reach to the destination in intra-BM is same as uniform traffic pattern. Therefore, it has the similar performance as that of uniform traffic.

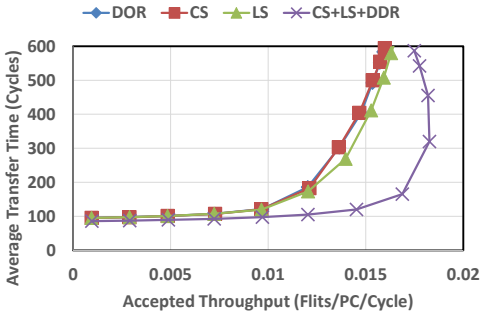

Figure 8: Comparison of dynamic communication performance of the perfect shuffle traffic pattern on the DOR and adaptive routings for TESH(2,3,0) network: 4096 nodes, 4 VCs, and 16 flits.

5.5 Local Communication Traffic Pattern

In some applications, there are a lot of communications between neighborhood nodes. To study this phenomenon in hierarchical interconnection network, we have evaluated the communication performance of a TESH network using local communication. In this traffic pattern, each node first generates a random number. If that number is less than a predefined threshold, the message will be sent to the destination PE in same BM. Otherwise, the message will be sent to any other nodes, with a uniform distribution. In this pattern, packet is sending from $\mathrm{PE}_s = (n_{s5}n_{s4})(n_{s3}n_{s2})(n_{s1}n_{s0})$ to $\mathrm{PE}_l = (n_5n_4)(n_3n_2)(n_{d1}n_{d0})$.

The local packet generation probability are assumed to be from $P_l = 0.0$ (all packets are sent to PE in other BMs) to $P_l = 1.0$ (all packets are local packet).

Figure 9 depicts the maximum throughput with respect to the local packet generation probability. Figure 10 portrays the ratio between maximum throughput of different adaptive routing algorithm and DOR ($Th_{\mathrm{Ad}}/Th_{\mathrm{DOR}}$) with respect to local packet generation probability. Here, Th_{DOR} is the maximum throughput using DOR and Th_{Ad} is the maximum throughput using adaptive routing algorithm. Thus the ratio $Th_{\mathrm{Ad}}/Th_{\mathrm{DOR}}$ is plotted in the *y*-axis and local packet generation probability is plotted in the *x*-axis.

As illustrated in these figures, the performance is improved using CS and LS algorithm when P_l is low and the performance is considerably improved using DDR algorithm when P_l is higher. Therefore, DDR is a suitable algorithm for local communication in the hierarchical interconnection network such as TESH because the performance is improved when P_l is higher.

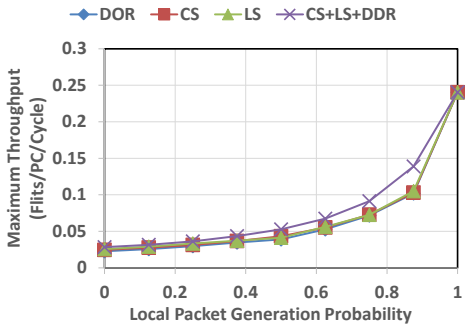

Figure 9: Comparison of maximum throughput of the local traffic pattern on the DOR and adaptive routings for TESH(2,3,0) network: 4096 nodes, 4 VCs, and 16 flits.

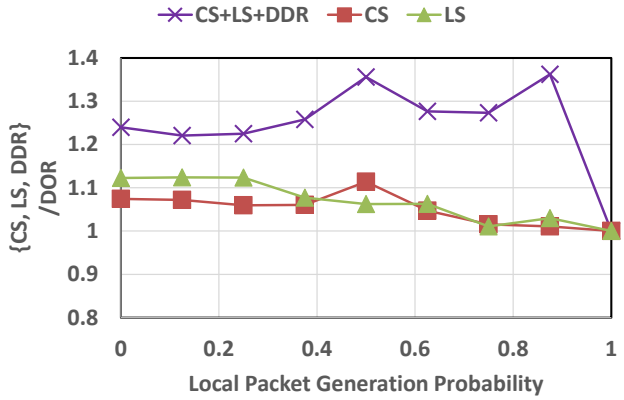

Figure 10: Comparison of maximum throughput with the DOR and adaptive routings of the local traffic pattern on the for TESH(2,3,0) network: 4096 nodes, 4 VCs, and 16 flits.

6 Conclusion

We have proposed three adaptive routing algorithms, CS, LS, DDR along with dimension-order routing with their hardware implementation for the TESH network. The proposed algorithms are simple and efficient for using the virtual channels, physical links, and direction of network to improve dynamic communication performance.

In this paper, we have evaluated the dynamic communication performance using different non-uniform traffic patterns named hotspot, complement, perfect shuffle and local communication traffic patterns. It was shown that the dynamic communication performance using DDR algorithm in a TESH network is slightly improved in the complement traffic and significant improved in the perfect shuffle traffic patterns. Also, in local communication traffic pattern, the dynamic communication performance is highly improved when inter-BM communications appear frequently.

In the application of hierarchical interconnection network, it is thought that data are laid out so that communication with neighborhood may become a lot. Therefore, DDR is suitable for the hierarchical interconnection network such as TESH.

REFERENCES

[1] W.J. Dally, Performance Analysis of *k*-ary *n*-cube Interconnection Networks, *IEEE Trans. on Computers*, vol. 39, No.6, pp.775--785, 1990.

[2] V.K. Jain, T. Ghirmai, and S. Horiguchi, TESH: A new hierarchical interconnection network for massively parallel computing, *IEICE Trans. on Inf. & Syst.*, Vol.E80-D, No.9, pp.837-846, 1997.

[3] V. K. Jain, T. Ghirmai and S. Horiguchi, Reconfiguration and Yield for TESH: A New Hierarchical Interconnection Network for 3-D Integration, *IEEE Proceedings of International Conference Wafer Scale Integration*, pp. 288-297, 1996.

[4] V.K. Jain and S. Horiguchi, VLSI Considerations for TESH: A New Hierarchical Interconnection Network for 3-D Integration, *IEEE Trans on VLSI Systems*, Vol.6, No. 3, pp. 346-353, 1998.

[5] S. Bhansali et al., 3D heterogeneous sensor system on a chip for defense and security applications, *Proceedings of the SPIE Defense and Security Symposium (DSS)*, pp.413-424, 2004.

[6] G. H. Chapman, V. K. Jain and S. Bhansali, Defect Avoidance in 3-D Heterogeneous sensor, Proceedings of the 19th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'04), pp.67-75, 2005.

[7] Y. Miura and S. Horiguchi, A Deadlock-Free Routing for Hierarchical Interconnection Network: TESH, *Proc. of the Fourth International Conference on High Performance Computing in Asia-Pacific Region*, pp.128-133, 2000.

[8] M.M. Hafizur Rahman, Y.Inoguchi, Y.Sato, Y.Miura and S.Horiguchi, On Hot-Spot Traffic Pattern of TESH Network, *11th International Conference on Computer and Information Technology(ICCIT 2008)*, 2008.12.

[9] M.M. Hafizur Rahman, Y.Inoguchi, Y.Sato, Y.Miura and S.Horiguchi, Dynamic Communication Performance of a TESH Network under the Nonuniform Traffic Patterns, *11th International Conference on Computer and Information Technology(ICCIT 2008)*, 2008.12.

[10] M.M. Hafizur Rahman, Y.Inoguchi, Y.Sato, Y.Miura, S.Horiguchi, Dynamic Communication Performance of the TESH Network under Nonuniform Traffic, *Journal of Networks*, Vol.4, No.10, pp.941-951, 2009.12.

[11] W.J. Dally, Virtual-Channel Flow Control, *IEEE Trans on Parallel and Destributed Systems*, Vol.3, No.2, pp.194-205, 1992.

[12] C. S. Yang and Y. M. Tsai, Adaptive Routing in k-ary n-cube Multicomputers, *Proc. of ICPADS '96*, pp.404-411, 1996.

[13] W.J. Dally and C.L.Seitz, Deadlock-Free Message Routing in Multiprocessor inter-connection Networks, *IEEE Trans. on Computers*, Vol.C-36, No.5, pp.547-553, 1987.

[14] W.J. Dally and H. Aoki, Deadlock-Free Adaptive Routing in Multicomputer Networks Using Virtual Channels, *IEEE Trans. on Parallel and Distributed Systems*, Vol. 4, No. 4, pp.466-475, 1993.

[15] C.J. Glass and L. M. Ni, Maximally Fully Adaptive Routing in 2D Meshes, *ISCA92*, pp.278-287, 1992.

[16] J. Duato, A New Theory of Deadlock-Free Adaptive Routing in Wormhole Networks, *IEEE Trans. on Parallel and Distributed Systems*, Vol.4, No.12, pp.1320-1331, 1993.

[17] Y. Miura and S. Horiguchi, An Adaptive Routing for Hierarchical Interconnection Network TESH, *Proc. of the Third International Conference on Parallel And Distributed Computing, Applications and Technologies*, pp. 335-342, 2002.

[18] Y. Miura, M. Kaneko and S. Horiguchi, Examination of Hardware Implementation on Adaptive Routing for Hierarchical Interconnection Network TESH, *Proc. of International Workshop on High Performance and Highly Survivable Routers and Networks (HPSRN 2008)*, 2008.

[19] Y.Miura, M.Kaneko, S.Watanabe, Adaptive Routing Algorithms and Implementation for Interconnection Network TESH for Parallel Processing, *The 35th IEEE Conference on Local Computer Networks (LCN)*，2010.10.

[20] Y.Miura, M.Kaneko, M.M.Hafizur Rahman and S.Watanabe, Adaptive Routing Algorithms and Implementation for TESH Network, *Communications and Network (CN)*, Vol.5, No.1, pp.34-49, 2013.02.

[21] Y. Miura, S. Watanabe, and M.M. Hafizur Rahman, The Communication Performance of Adaptive Routing for Hierarchical Interconnection Network for 3D VLSI, Proc. of 2015 International Conference on Information, Computer and Communication Engineering (ICC 2015) (Accepted).

[22] N.Tsuda, Hierarchical redundancy for array-structure WSIs, *Journal of Systems and Computers in Japan*, Vol.24, No.7, pp.13--30, 1993.

[23] L. M. Ni and P. K. McKinley, A Survey of Wormhole Routing Techniques in Direct Networks, *Computer*, Vol.26, No.2, pp.62-76, 1993.

[24] M. Grammatikakis, D.F. Hsu, M. Kratzel and J.F. Sibeyn, Packet routing in fixed connection networks: a survey, *Journal of Parallel and Distributed Computing*, Vol. 54, No. 2, pp.77–132, 1998.

[25] Andrew A. Chien and Jae H. Kim, Planer-Adaptive Routing:Low-cost Adaptive Networks for Multiprocessors, *Journal of the ACM*, Vol.42, No.1, pp.91–123, 1995.

[26] P.R. Miller, Efficient Communications for Fine-Grain Distributed Computers, *Ph.D. Dissertation, Southampton University*, U.K., 1991.