

Title	Shifted Recursive Torus interconnection network for massively parallel computers
Author(s)	Inoguchi, Yasushi; Horiguchi, Susumu
Citation	Research report (School of Information Science, Japan Advanced Institute of Science and Technology), IS-RR-96-0008A: 1-19
Issue Date	1996-03-15
Type	Technical Report
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/8369">http://hdl.handle.net/10119/8369</a>
Rights	
Description	リサーチレポート (北陸先端科学技術大学院大学情報科学研究科)

**Shifted Recursive Torus  
Interconnection Network for  
Massively Parallel Computers**

Yasushi Inoguchi and Susumu Horiguchi  
1996, 3/15  
IS-RR-96-0008A

School of Information Science  
Japan Advanced Institute of Science and Technology, Hokuriku  
Asahidai 15, Tatsunokuchi  
Nomi, Ishikawa, 923-12, JAPAN  
inoguchi@jaist.ac.jp

©Yasushi Inoguchi, 1996

ISSN 0918-7553

## Abstract

A number of interconnection networks have been proposed for multiprocessor systems from view points of theoretical and industrial interests. In order to construct massively parallel computers in VLSI implementation, interconnection between processing elements is one of the critical problems. The number of links between processing elements are limited to reduce wiring area. On the other side, Scale Integration (WSI) has been attracted as a technology to implement a number of PEs on a silicon wafer. So defect and fault-tolerance schemes are also very important for massively parallel computers.

In this paper, we propose Shifted Recursive Torus (SRT) interconnection network based on shifting torus network to build multi-level interconnection recursively. The SRT interconnection network is designed by taking account of the limited number of links, easy implementation in VLSI and hierarchical structure for expandability. The SRT is defined as one dimensional interconnection network and is extended into two dimensional interconnection network. Network features such as diameter and average distance of SRT are discussed in detail. A reconfiguration strategy of the SRT proposed as a defect-tolerance scheme.

## 1 Introduction.

There are many types of interconnection networks for multiprocessor systems to achieve high communication performance between a large number of processing elements (PEs). A completely connected network has the shortest distance between PEs, but massively parallel systems never use it because it requires a large numbers of links. As one dimensional networks with limited number of links, Barrel Shifter [5], Chordal Ring [6] and Iliac Network [5] are well known. Barrel Shifter has a small average distance ( $\log_2 N/2$ ) where  $N$  is the number of PEs, but the number of links a node increases as  $2 \log_2 N - 1$ . Each node of Chordal Ring has also a small number of links ( $= 3$  and this number is fixed), but the average distance is large ( $\mathcal{O}\sqrt{N}$ ). Since Iliac Network is one type of Chordal Ring, it is not suitable for massively parallel computers.

On the other hand, a hypercube type network such as a binary cube and crossed cube network[1] is one of good candidates for massively parallel systems from the view point of interconnection complexity, but each node of the hypercube has  $\log_2 N$  links. Due to the number of links per PE node, it is difficult to implement a large scale hypercube in VLSI. Therefore, interconnection networks for massively parallel systems require the network property that the number of links of PE is small or fixed in the small number.

Several 2D-interconnection networks with the limited number of links such as Mesh network, Torus network, MANDALA network[2] and RDT[3] have been proposed. The mesh network and torus network have very simple structure, but their communication performance is lower than hypercube type networks. The MANDALA network is suitable for some specific problems because the network has diagonal links and a simple hierarchical structure. However, diameter and average distance of MANDALA network are not small. The RDT network is one of good candidates for massively parallel computers because it has small diameter and low average distance. The RDT network is a torus mesh with hierarchical diagonal links to improve network diameter.

Recently, Wafer Scale Integration (WSI) has been attracted as a technology to implement a number of PEs on a silicon wafer. The implementation of PEs and interconnection on the wafer achieves low signal delay, low power and small size of the system. However, defects and faults on the wafer cause some problems inherently. Some of PEs and links on a wafer become faulty. An interconnection network for WSI system must have fault-tolerance architecture by replacing faulty PEs by spare PEs.

Although the RDT has a simple structure and high communication performance, a wiring complexity of the RDT is not good and a fault-tolerance of the RDT is not easy due to diagonal links. To improve the wiring complexity of links and to achieve the fault-tolerance, we propose a new interconnection network refer to Shifted Recursive Torus (SRT)[4]. SRT consists of mesh-toruses and bypass links to shift torus recursively. The SRT is defined as one dimensional interconnection network and is extended into two dimensional interconnection



network. Since 2D-SRT has no diagonal links of the RDT, 2D-SRT easily implements a fault-tolerance scheme in WSI.

In this paper, we describe a definition of one dimensional SRT in the next section. Section 4 shows an extension of one dimensional SRT into two dimensional SRT. In section 4, we discuss the network feature of SRT by comparing diameter and average distance with those of RDT and Hypercube. Section 5 shows a fault-tolerance scheme for SRT to replace faulty PEs by spare PEs. Section 6 is conclusions.

## 2 Shifted Recursive Torus Interconnection.

### 2.1 One Dimensional Structure

The simplest torus interconnection is a ring network. To improve network performance of ring networks, chordal ring uses bypass links to connect between remote nodes with a long distance. However, chordal ring networks do not have enough network performance for a large number of PEs. To improve the network performance, we propose one dimensional SRT (1D-SRT) based on the ring network with bypass links under the network constraint that every node has a fixed number links. Let's define a basic structure of 1D-SRT in the following section.

**Definition 1 (Base Ring)** *The ring network consist of  $N$  nodes ( $N = 2^n$ ), and each node is addressed in the order. Node  $i$  is connected to two neighboring nodes ( $i \pm 1 \text{ mod } N$ ). Edge nodes are connected each other. like as torus network. Links consisting of the base ring network are named as level-0 link.  $\square$*

To improve a diameter and a average distance of the ring network, 1D-SRT adds bypass links to the base ring. The bypass links are added as the following scheme.

Let a node  $n_1$  among the base ring be as follows,

$$n_1(i) = 2i + 1 \quad (0 \leq i < 2^{n-1}) \quad (1)$$

Odd number nodes among the base ring are called as level-1 nodes. A sub-ring is made by connecting level-1 nodes with bypass links. The bypass links to compose sub-ring are called as level-1 links. Each level-1 node is connected with two neighboring nodes ( $n_1(i \pm 1 \text{ mod } 2^{n-1}) = 2i + 1 \pm 2 \text{ mod } 2^n$ ) on the sub-ring with level-1 bypass links.

Consider a node  $n_2$  as follow,

$$n_2(i) = 4i + 2 \quad (0 \leq i < 2^{n-2}) \quad (2)$$

Nodes having them addresses  $n_2$  are defined as level-2 nodes. Similarly, a sub-ring is made by connecting level-2 nodes with bypass links called as level-2 links.

Each level-2 node is connected to two neighboring nodes ( $n_2(i \pm 1 \bmod 2^{n-2}) = 4i + 1 \pm 4 \bmod 2^n$ ) on the sub-ring with level-2 bypass links.

The  $l$ -th level nodes and upper level links are defined recursively. Then we have the following definition.

**Definition 2 (Upper Level Link)** *Let  $l$  be the number of levels in one dimensional SRT networks, the nodes at the  $l$ -th level of SRT are defined as follows:*

$$n_l(i) = 2^l \cdot i + 2^{l-1} (0 \leq i < 2^{n-l}) \quad (3)$$

$n_l$  nodes are defined as level- $l$  nodes. Sub-rings are made by connecting level- $l$  nodes with bypass links called as level- $l$  links. Each level- $l$  node is connected to two neighboring nodes ( $n_l(i \pm 1 \bmod 2^{n-l}) = 2^l i + 1 \pm 2^l \bmod 2^n$ ) on the sub-ring with level- $l$  bypass links. Sub-rings consisting of nodes except for level-0 are called as upper level ring.  $\square$

To compose the  $l$ -th ring under a limited number of links, the 1D-SRT consists of  $N$  nodes having 4 links; 2 links for the base ring and 2 links for upper level rings except for nodes at  $0, \frac{1}{2}N, \frac{1}{4}N$  and  $\frac{3}{4}N$ . Each node has only one level, in 1D-SRT because nodes belonging to different level are not overlapped.

**Theorem 1** *Nodes in sub-rings at different levels are not overlapped.*

**Proof:** Considering the node  $n_k$  in the sub-ring at level- $k$ , and the node  $n_l$  in the sub-ring at level- $l$  ( $k > l$ ),  $n_k$  and  $n_l$  are given as follows,

$$\begin{aligned} n_k &= i \cdot 2^k + 2^{k-1} \\ n_l &= j \cdot 2^l + 2^{l-1}. \end{aligned}$$

Then the distance between  $n_k$  and  $n_l$  is

$$n_k - n_l = 2^{l-1} [2^{k-l-1} \cdot 2(2i-1) - (2j-1)] \neq 0.$$

Therefore two nodes at the different levels are not overlapped.  $\square$

Now, we show that nodes  $\frac{1}{4}N$  and  $\frac{3}{4}N$  have three links and nodes at  $0$  and  $N/2$  have two links. Levels of nodes  $\frac{1}{4}N$  and  $\frac{3}{4}N$  are  $n-1$ . A level- $(n-1)$  sub-ring consists of two nodes; node  $\frac{1}{4}N$  and node  $\frac{3}{4}N$ . Thus these nodes have three links; two links for base ring and one link to connect each other.

Nodes  $0$  and  $N/2$  have only two links for base ring and no links for upper level links.

**Corollary 1** *Node  $0$  and  $N/2$  have no upper level links.*

**Proof:** The number of the level- $l$  nodes is  $N_l = 2^{n-l}$ , and the highest level of 1D-SRT is  $n-1$ . Then the number of nodes belonging higher levels than level-1 is given by,

$$\sum_{l=1}^{l-1} N_l = \sum_{l=1}^{n-1} 2^{n-l} = 2^n - 2.$$

The equation shows us that two nodes at the level-0 has no upper links. It is obvious from equation (3) that the node 0 has no upper links.

Considering the node at  $N/2$ , we have the following equation,

$$N/2 - (2^l i + 2^{l-1}) = 2^{l-1} [2(2^{n-l-1} - i) - 1] \neq 0 \quad \text{for any } l, i$$

Therefore the nodes 0 and  $N/2$  have no upper links.

**Definition 3 (standard 1D-SRT)** *To compose 1D-SRT, sub-rings are constructed recursively from the level-1 to the level- $l_{max}$ , where  $l_{max} = n-1$  using the bypass links to the base ring.*  $\square$

Figure 1 shows the basic interconnection of standard 1D-SRT consisting of 32 nodes. In the standard 1D-SRT, the nodes 0 and  $N/2$  have only two links and the nodes  $\frac{1}{4}N$  and  $\frac{3}{4}N$  have only three links. Other nodes have four links. Adding the bypass links to the node 0 and the node  $N/2$ , we have Long Span 1D-SRT shown in figure 2 (a). Adding the bypass links to the nodes 0,  $\frac{1}{4}N$ ,  $\frac{1}{2}N$  and  $\frac{3}{4}N$ , we have Short Span 1D-SRT as shown in figure 2 (b).

**Definition 4 (Long Span 1D-SRT)** *To construct Long Span 1D-SRT (LS-1D-SRT), a link at the level- $l_{max}$  connects between the node 0 and the node  $N/2$ .*  $\square$

**Definition 5 (Short Span 1D-SRT)** *To construct Short Span 1D-SRT (SS-1D-SRT), remove a level- $l_{max}$  link connecting the node  $\frac{1}{4}N$  and the node  $\frac{3}{4}N$ . Then SS-1D-SRT is constructed by adding links at the level- $(l_{max}-1)$  connect between the node 0 and the node  $\frac{1}{4}N$ , and the node  $\frac{1}{2}N$  and the node  $\frac{3}{4}N$  of standard 1D-SRT.*  $\square$

Figure 2 shows configurations of the LS-1D-SRT and SS-1D-SRT with 32 nodes. LS-1D-SRT is expected to improve the communication performance between the node 0 or the node  $N/2$ . SS-1D-SRT is expected to improve average communication performance since it has links uniform by distributed in the standard SRT.

## 2.2 Planner Layout of 1D-SRTs.

In order to construct a multiprocessor system using the SRT interconnection, a planner layout is very important to reduce the wiring area to connect each node. We will discuss how much space for wiring of 1D-SRT needs.

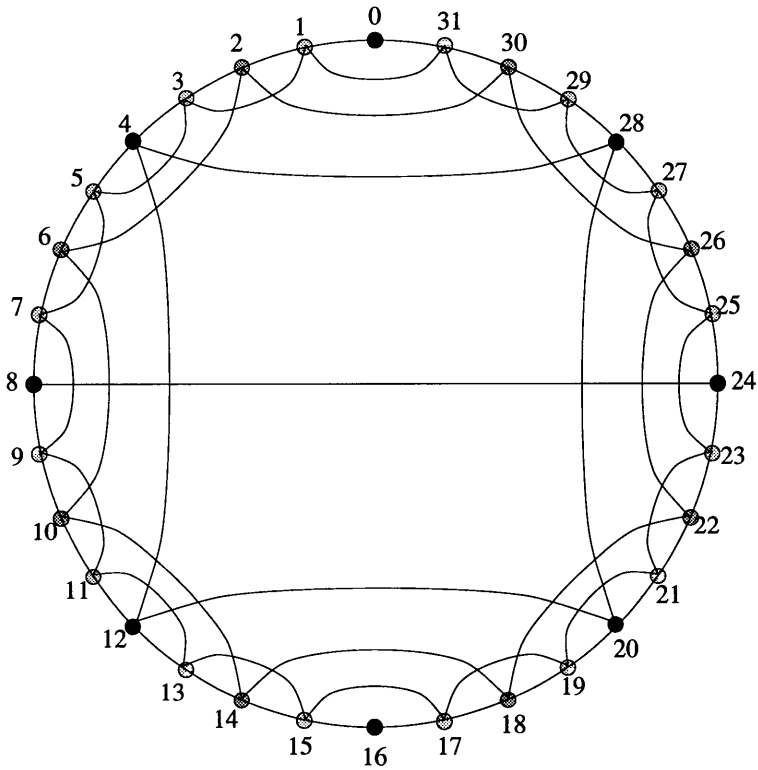


Figure 1: Standard 1D-SRT consist of 32 nodes.

Figure 3 shows a planner layout of the 1D-SRT in a straight line. We will show that 1D-SRT requires small space for wiring.

**Theorem 2** *The 1D-SRT requires  $2l_{max} + 1$  wiring-width in a straight line layout.*

**Proof:** In a standard 1D-SRT,  $l_{max} + 1$  wires are necessary for level-0 to level- $l_{max}$  to connect nodes at both ends.  $l_{max}$  wires are also necessary for level-0 to level- $(l_{max} - 1)$ . Since the same level wires are not layouted in parallel, standard SRT needs  $2l_{max} + 1$  wires in a straight line structure.  $\square$

**Corollary 2** *LS-1D-SRT requires  $2l_{max} + 2$  wiring-width wires area in a straight line layout.*

**Proof:** LS-1D-SRT has an extra wire of level- $l_{max}$  to connect between the node 0 and the node  $N/2$ . This wire can not overlap with another level- $l_{max}$

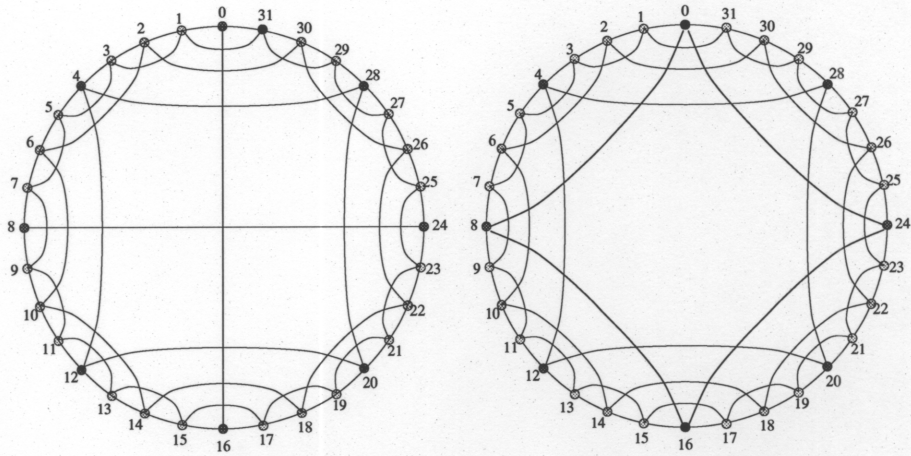


Figure 2: (a) Long Span 1D-SRT

(b) Short Span 1D-SRT

wire to connect between the node  $\frac{1}{4}N$  and the node  $\frac{3}{4}N$ . Thus, LS-1D-SRT requires the same wiring-width of standard 1D-SRT and one wiring-width for links between the node 0 and the node  $N/2$ , shown as follow,

$$(2l_{max} + 1) + 1 = 2l_{max} + 2.$$

**Corollary 3** *SS-1D-SRT requires  $2l_{max} + 2$  wiring-width in a straight line layout.*

**Proof:** To construct SS-1D-SRT, one wire of level- $l_{max}$  from standard 1D-SRT is removed, then level- $(l_{max} - 1)$  wires are added to connect nodes  $0, \frac{1}{4}N, N/2, \frac{3}{4}N$  at level- $(l_{max} - 1)$  sub-ring. In this case, note that level- $(l_{max} - 1)$  wires connecting nodes  $0, \frac{1}{4}N, \frac{1}{2}N$  and  $\frac{3}{4}N$  cross wires connecting nodes  $\frac{1}{8}N, \frac{3}{8}N, \frac{5}{8}N$  and  $\frac{7}{8}N$  since they belong to the same level. The level- $(l_{max} - 1)$  wires connecting nodes  $\frac{1}{8}N, \frac{3}{8}N, \frac{5}{8}N$  and  $\frac{7}{8}N$  additionally require two wiring-width: one to connect neighboring nodes and one to connect nodes of both edges. Therefore, SS-1D-SRT requires  $(2l_{max} + 1) - 1 + 2 = 2l_{max} + 2$  wiring-width in straight line layout.

In any case, 1D-SRT needs  $\mathcal{O}(\log_2 N)$  for the complexity of wiring-width because  $l_{max} = \log_2 N - 1$ . The wiring-width complexity of 1D-SRT than other 1D-networks such as Chordal Ring and Barrel Shifter. Chordal Ring requires

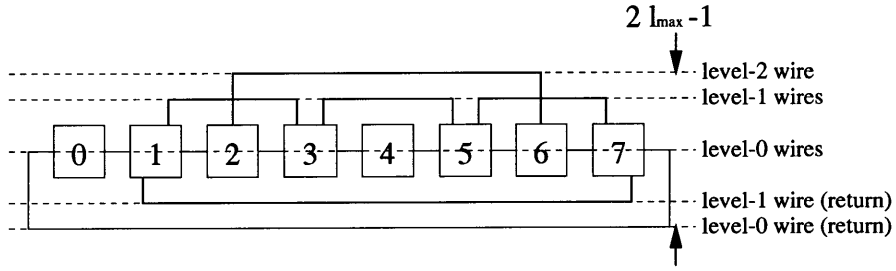


Figure 3: Straight placed 1D-SRT consist of 8 nodes.

$\mathcal{O}(\sqrt{N})$  wiring-width for a straight line layout. Barrel Shifter requires the larger wiring-width, since it has more links than Chordal Ring.

The 1D-SRT has an advantage that it requires wiring the smaller area than other 1D-networks.

### 3 Two Dimensional SRT.

#### 3.1 Extension of 1D-SRT into 2D-SRT.

In this section, we extend the 1D-SRT into two dimensional network named as 2D-SRT. A standard 2D-SRT has a two dimensional structure consisting of  $N \times N$  nodes ( $N = 2^n$ ).

We define a base torus for the 2D-SRT as same as the base ring of the 1D-SRT.

**Definition 6 (Base Torus)** Let  $(i, j)$  be the address of node in the 2D-SRT. The base torus of the 2D-SRT is shown as following orders,

$$\begin{array}{cccccc}
 (0, 0) & (1, 0) & (2, 0) & \dots & (N - 1, 0) \\
 (0, 1) & (1, 1) & (2, 1) & \dots & (N - 1, 1) \\
 (0, 2) & (1, 2) & (2, 2) & \dots & (N - 1, 2) \\
 \dots & \dots & \dots & \dots & \dots \\
 (0, N - 1) & (1, N - 1) & (2, N - 1) & \dots & (N - 1, N - 1)
 \end{array}$$

A node  $(x, y)$  is connected with neighboring nodes whose addresses are shown by  $(x \pm 1 \text{ mod } N, y \pm 1 \text{ mod } N)$ .  $\square$

To make sub-rings of the 1D-SRT, bypass links are added to nodes in the 1D-SRT. Interconnections of the 1D-SRT in the x-direction are shifted by a row location of 1D-SRT. Interconnections in the y-direction are also shifted by a column location of the 1D-SRT. We can define several types of 2D-SRT

according to shifting width in x-direction. Let's define a general structure of 2D-SRT.

**Definition 7 (2D-SRT)** Let  $(x_l, y_l)$  be a node address at the level- $l$  of 2D-SRT. Nodes  $(x_l, y_l)$  of the 2D-SRT are determined by following equations,

$$x_l = 2^l \cdot i + 2^{l-1} - s \cdot y_l \quad (0 \leq i < 2^{n-l}), \quad (4)$$

Where  $s$  is a shift width in the x-direction. The shift width  $s$  must satisfy following conditions,

1. Every row and column has only one origin node of 1D-SRT.
2. A transposed expression of node address must be defined.  
(i.e. we can find equivalent equation as like as  $y_l = 2^l \cdot j + 2^{l-1} - s' \cdot x_l$ )

The node  $(x_l, y_l)$  is connected with four neighboring nodes  $(x_l \pm 2^l \bmod 2^n, y_l \pm 2^l \bmod 2^n)$  using level- $l$  links.  $\square$

Figure 4 shows configuration of general form of 2D-SRT.

To connect nodes at the upper level, bypass links are added to nodes. A torus at the level- $l$  is composed by connecting nodes at the level- $l$ , each other. Nodes at the level- $l$  are connected with nodes  $(x_l \pm 2^l \bmod N, y_l \pm 2^l \bmod N)$  at the same level using vertical and horizontal bypass links whose span is  $2^l$ .

Let's discuss 1-shift 2D-SRT as the simplest structure of 2D-SRT.

**Definition 8 (1-shift 2D-SRT)** Nodes  $(x_l, y_l)$  at the level- $l$  are determined by following equations,

$$x_l = 2^l \cdot i + 2^{l-1} - y_l \quad (0 \leq i < 2^{n-l}). \quad (5)$$

A node  $(x_l, y_l)$  is connected with four nodes  $(x_l \pm 2^l \bmod 2^n, y_l \pm 2^l \bmod 2^n)$  using the level- $l$  links.

Each row and column of 2D-SRT have only one origin node of 1D-SRT. It is obvious from above definition that the interconnection in x-direction is the same as the standard 1D-SRT, and we show that the interconnection in y-direction is the same as the standard 1D-SRT too. The show transposed expression of equation (5) is given by

$$y_l = 2^l \cdot i + 2^{l-1} - x_l \quad (0 \leq i < 2^{n-l}). \quad (6)$$

Equation (6) is the same form of equation (5). Figure 5 shows a basic structure of 1-shift 2D-SRT consisting of  $8 \times 8$  nodes. The integer in the figure shows the level number of 2D-SRT.

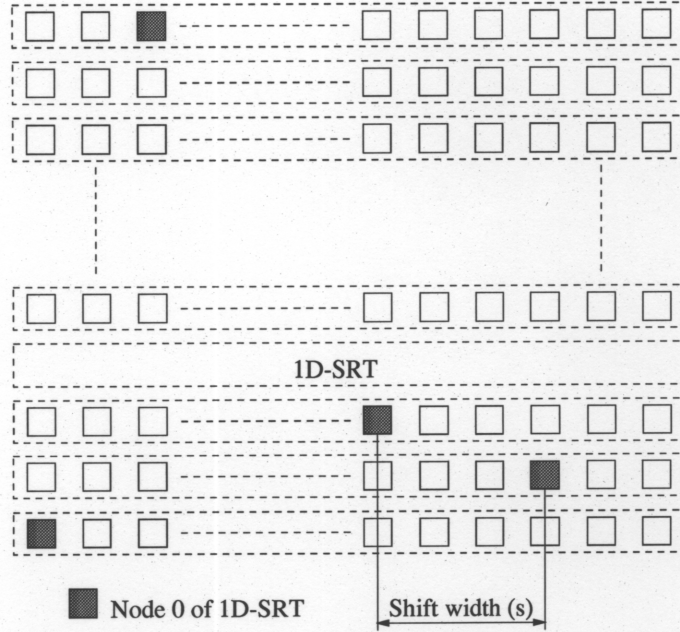


Figure 4: Configuration of general form of 2D-SRT.

### 3.2 Uniformed 2D-SRT

The 1-shift 2D-SRT has a structure so that nodes at the same level are placed in diagonal as shown in figure 5. In general, links at the lower level are used to interconnect the neighboring nodes and links at the higher level are used to interconnect nodes with long distance. To improve communication performance, interconnection scheme at the high level should be a uniform interconnection. Then, we propose uniform 2D-SRT by modifying the 1-shift 2D-SRT.

**Definition 9 (Uniform 2D-SRT)** *There are four types of uniform 2D-SRT. Node addresses at the level- $l$  are defined as follows,*

$$x_l = 2^l \cdot i + 2^{l-1} - y_l \left( 2^{\lceil l_{max}/2 \rceil} - 1 \right) \quad (7)$$

$$x_l = 2^l \cdot i + 2^{l-1} + y_l \left( 2^{\lceil l_{max}/2 \rceil} + 1 \right) \quad (8)$$

$$x_l = 2^l \cdot i + 2^{l-1} - y_l \left( 2^{\lceil l_{max}/2 \rceil} + 1 \right) \quad (9)$$

$$x_l = 2^l \cdot i + 2^{l-1} + y_l \left( 2^{\lceil l_{max}/2 \rceil} - 1 \right) \quad (10)$$

$$(0 \leq i < 2^{n-l}, l_{max} = n - 1)$$



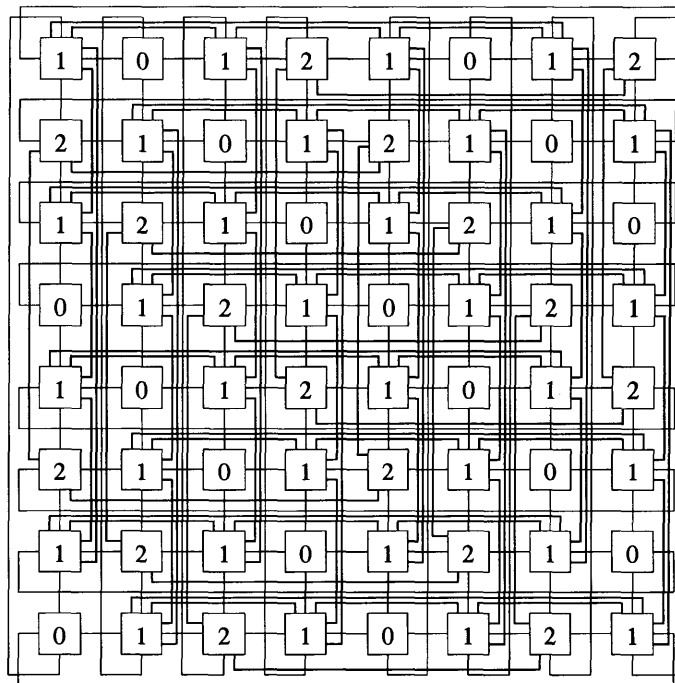


Figure 5: 1-shift 2D-SRT consist of  $8 \times 8$  of the 1D-SRTs.

*Sub-toruses at the  $l$ -th level are connected by bypass links. The level- $l$  torus is composed by connecting nodes at the level- $l$  each other. Nodes at the level- $l$  are connected with nodes  $(x_l \pm 2^l \bmod N, y_l \pm 2^l \bmod N)$  at the same level by using vertical and horizontal bypass links whose span is  $2^l$ .  $\square$*

It is obvious that the uniformed 2D-SRT includes  $N$  1D-SRTs in both of  $x$ -direction and  $y$ -direction. Equations (8) and (10) are transposed by equations (7) and (9). Equations (8) and (7) are symmetrical to equations (9) and (10). Figure 6 shows the level number of nodes for the uniformed 2D-SRT. Grayish squares in the figure show the placement of nodes at the level-0.

Using a standard 1D-SRT, LS-1D-SRT and SS-1D-SRT, we can define 6 types of 2D-SRT: 1-shift standard 2D-SRT, uniform standard 2D-SRT, 1-shift LS-2D-SRT, uniform LS-2D-SRT, 1-shift SS-2D-SRT and uniform SS-2D-SRT. Network performance of those schemes are discussed in the following section.

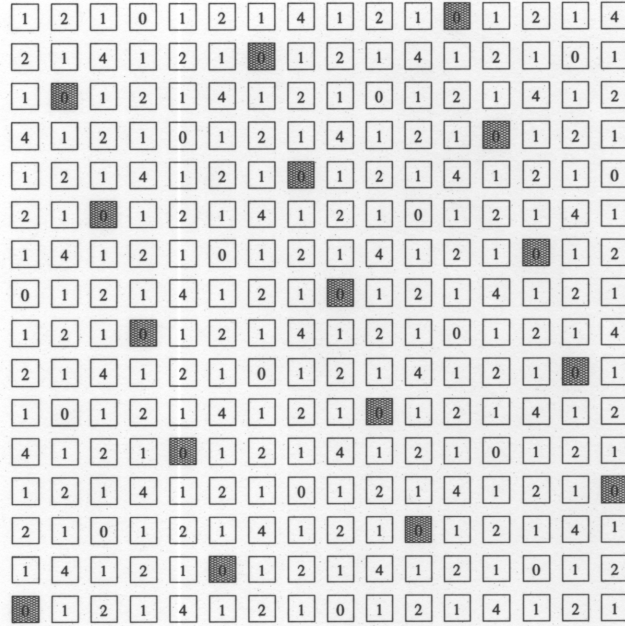


Figure 6: Level placement of the uniformed 2D-SRT consist of  $16 \times 16$  nodes.

## 4 Network Performance

We investigate diameters and average distances of 1D-SRT and 2D-SRT interconnections. Figure 7 shows average distances of 1D-SRT as a function of the number of nodes. It is shown that the average distance of the 1D-SRT becomes almost 1.2 times when the number of nodes becomes twice. The average distance does not increase much more when the number of nodes becomes a large number. Comparing the average distance of SS-1D-SRT with those of standard 1D-SRT and LS-1D-SRT, there is not much difference.

Figure 8 shows average distance of the 2D-SRT as a function of the number of nodes. The figure shows an increase of 30% on every four times of the number of nodes. There are few difference in the increasing ratio when the number of nodes are small. An increasing ratios of average distance of the 2D-SRT are low when the number of nodes is large. The difference of average distances between 1-shift 2D-SRT and uniform 2D-SRT increase when the number of nodes increases. The uniform SRTs are always better than 1-shift SRTs.

Table 1 shows diameters of SRT and other networks. For 64k nodes system, the number of links by node of uniform 2D-SRT is just half of crossed cube, but the diameters is smaller than two times of the diameter of crossed cube. The

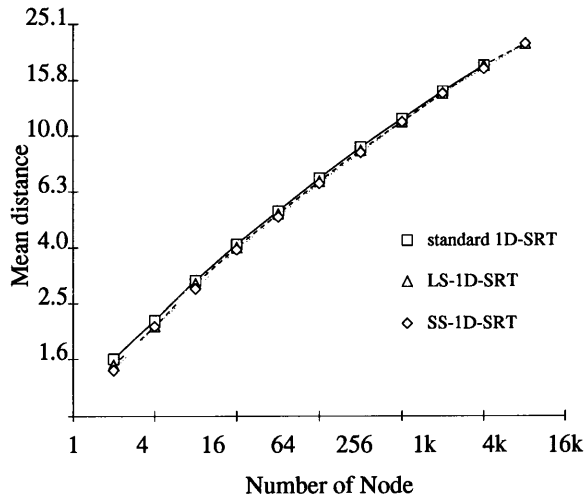


Figure 7: Average distances of 1D-SRTs as a function of the number of nodes.

diameter of SRT is little bit larger than that of RDT, but SRT has no diagonal links. It is an advantage when implementing the SRT into WSI system. Next section we discuss a reconfigurable scheme of SRT.

## 5 Fault-Tolerance Architecture for SRT

For mesh connected WSI system, defect avoidance algorithms has been proposed. The system has PE array at center of the wafer and spare PEs around them. Some of center PEs and spare PEs are faulted. To get fault free logical PE array from physical PE array with fault, available PE adjacent to the faulty PE execute the function of the faulty PE, and PE adjacent to PE just shifted function of the faulty PE execute the function of the adjacent PE. By shifting logical function of PEs recursively, we get a fault free PE array.

### 5.1 Fault-Tolerance scheme

A torus network has rap-round links between edge nodes. To achieve a reconfiguration of mesh array, Kung et al.[7] and Numata et al.[8] have proposed the reconfiguration algorithms for  $(N + R) \times (N + R)$  mesh-array. In this section, a reconfiguration scheme is proposed for torus networks. The fault-tolerance architecture is shown in figure 9. The fault-tolerance system of torus network of SRT has the processor array consisting of  $N \times N$  core PEs and spare PEs

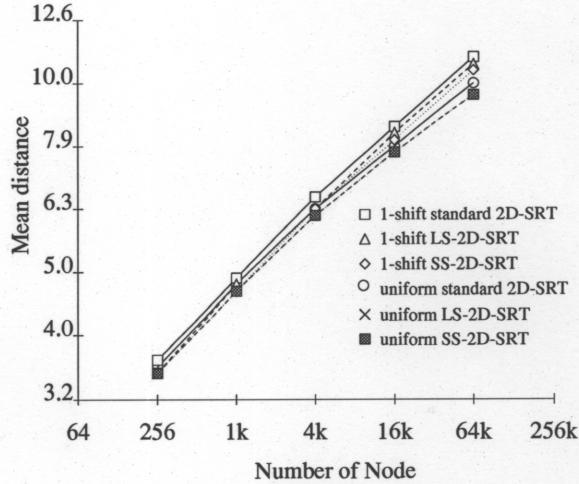


Figure 8: Average distances of 2D-SRTs as a function of the number of nodes.

around the array shown as figure 9. The facility for defect avoidance with shifting is simple, it requires one or two tracks between PEs and 4 switches around PE (note that number of switches is about  $N$  because a switch is shared with two neighboring PEs). Each switch has 4 states. To obtain  $N \times N$  mesh torus among  $(N + 2) \times (N + 2)$  PEs, status of switches are controlled by reconfiguration scheme. If all core  $N \times N$  PEs are fault free,  $N \times N$  mesh torus uses only core PEs. When some of core PEs faulty, the faulty PEs are replaced by neighboring PEs. Kung proposed an algorithm based on graph theory and Numata proposed an algorithm which use local information to determine shift direction. Figure 10 shows a reconfiguration of mesh torus network in the fault-tolerance architecture.

## 5.2 Reconfiguration for Bypass Links

To achieve a reconfiguring SRT by shifting scheme, levels of related nodes must be changed. Reconfiguration scheme has two steps to change a level of node using bypass links. First step is to shift levels of nodes on horizontal links. To achieve the level shift of horizontal links, we only consider a standard 1D-SRT as shown in figure 11. To shift the level of faulty PE at the 2nd-level to a right neighboring PE at the 1st-level, level of right neighboring PE must be changed from 1 to 2. To change the node level we introduce a bunch of links consisting of links of level- $l$  to level- $l_{max}$ . The bunch of links have switches to separate a link and to bypass link in horizontal directions. The switches on the

# of node	256	1k	4k	16k	64k
1D-SRT	17/4	25/4	41/4		
LS-1D-SRT	13/4	21/4	33/4		
SS-1D-SRT	12/4	20/4	30/4	45/4	
2D-SRT(1sft)	7/8	9/8	13/8	17/8	21/8
LS-2D-SRT(1sft)	6/8	8/8	10/8	14/8	18/8
SS-2D-SRT(1sft)	6/8	8/8	10/8	14/8	17/8
2D-SRT(uni)	6/8	8/8	11/8	13/8	16/8
LS-2D-SRT(uni)	6/8	7/8	9/8	12/8	14/8
SS-2D-SRT(uni)	6/8	8/8	10/8	12/8	15/8
RDT(2,4,1)	7/8	8/8	10/8	10/8	12/8
2D Torus	16/4	32/4	64/4	128/4	256/4
3D Torus			24/6		48/6
Hyper cube	8/8	10/10	12/12	14/14	16/16
Xed Cube	5/8	6/10	7/12	8/14	9/16

Table 1: Diameter/Links by node

bunch of links have two modes; a cross mode to bypass links and a connect mode to separate a horizontal links.

In figure 11, a node at level-2 is faulty and we try to shift right. The faulty PE is disconnected from the network to set all switches on the bunch to cross mode. Then the level of PE right neighboring of faulty PE is changed from 1 to 2. To achieve this procedure, change mode of level-2 switch on the bunch of the PE from cross mode to connect mode and set all other switches on the bunch of the PE to cross mode.

Next step is to shift a node level of vertical links. Figure 12 shows an architecture to achieve the level shift in vertical direction. To shift vertical links, we use switches and tracks between PEs as same as the reconfiguration of mesh torus in section 5.1. Switches has 4 status to connect a vertical link to a bunch of links. In figure 12, the status of bold-outlined switches on vertical links are changed to bend vertical links. To change the vertical level of PE right neighboring of faulty PE from 1 to 2, switch modes of bunch on vertical link of the PE are controlled so that level-2 switch are set to connect mode and other switches on the bunch are set to cross mode.

## 6 Conclusions

In this paper, we proposed SRT as a new interconnection network for massively parallel computers. Massively parallel systems require following network

features,

- Highly communication performance.
- The number of links of each node is small and fixed.
- The network is fault-tolerant.

To satisfy these network features, we proposed the SRT consisting of hierarchical shifting torus networks. Standard SRT is defined as one dimension network and Long Span SRT and Short Span SRT are derived. The two dimensional SRT are defined by expanding one dimensional SRT to two dimensional. Two dimensional SRTs are classified into six types interconnection by shift schemes: 1-shift standard 2D-SRT, uniform standard 2D-SRT, 1-shift Long Span 2D-SRT, uniform Long Span 2D-SRT, 1-shift Short Span 2D-SRT and uniform Short Span 2D-SRT. The average distance and diameter of 2D-SRT were compared with other interconnection networks. It is shown that Short Span SRT can reduce diameter and uniformed 2D-SRT can reduce average distance efficiently.

To achieve reconfiguration of SRT, we introduce bunch of links and switches as the fault-tolerance architecture. A level of node can be changed by changing switches on bunch of links and faulty PEs are replaced by neighboring PEs.

Advantages of SRT are summarized as follows,

- Number of links a node is fixed.(4 links for 1 dimension, 8 links for 2-dimension and 12 links for 3 dimension)
- The SRT has a hierarchical structure and an excellent scalability.
- We can easily extend 1-dimensional SRT to 2D-SRT.
- Diameter and average distance of SRT are much the same as RDT.
- The SRT has a fault-tolerant ability to avoid faulty PEs.

**Acknowledgment:** A part of this research was supported by a scientific grand-in-aid, Ministry of Education, Science and Culture of Japan.

## References

- [1] Kemal Efe: "The Crossed Cube Architecture for Parallel Computation" IEEE Trans. Parallel and Distributed System, Vol.3, No.5, pp.513-524 (1992)
- [2] Takuya Kanoh, Katsuhisa Hirota, Shigehiro Fujimoto, Andrew Flavell and Yoshizo Takahashi: "Design of Hierarchically Structure Destributed-Memory (in Japanese)" ISP of Japan, ARC-95-11 (Aug. 1992)

- [3] Yang Y., Amano H. Shibamura H. and Sueyoshi T.: "RDT:An Interconnection Network for Massively Parallel Computers (in Japanese)" Trans. of IEICE, Vol. J78-D-I No.2 pp.118-128 (Feb. 1995)
- [4] Yasushi Inoguchi and Susumu Horiguchi: "Shifted Recursive Torus: An Interconnection Network for Massively Parallel Computers (in Japanese)" Technical Report of IEICE, CPSY95-69 pp.25-30 (Oct. 1995)
- [5] Hwang K., Briggs F.A.: "Computer Architecture and Parallel Processing" pp. 339-342, 345-350 McGraw-Hill (1985)
- [6] Arden B.W., Lee H.: "Analysis of Chordal Ring Networks", IEEE Trans. Computers, Vol.C-30, No.4, pp.91-295 (1981)
- [7] Kung S. Y., Jean S. N. and Chan C. W.: "Fault-Tolerant Array Processors Using Single-Track Switches" IEEE Trans. Computers, 38, 4 (Apl. 1989)
- [8] Issei Numata and Susumu Horiguchi: "WSI Implementation of Mesh-Connected Multiprocessor Systems (in Japanese)" Trans. of IEICE, Vol. J77-D-I No.2 pp. 121-129 (Feb. 1994)

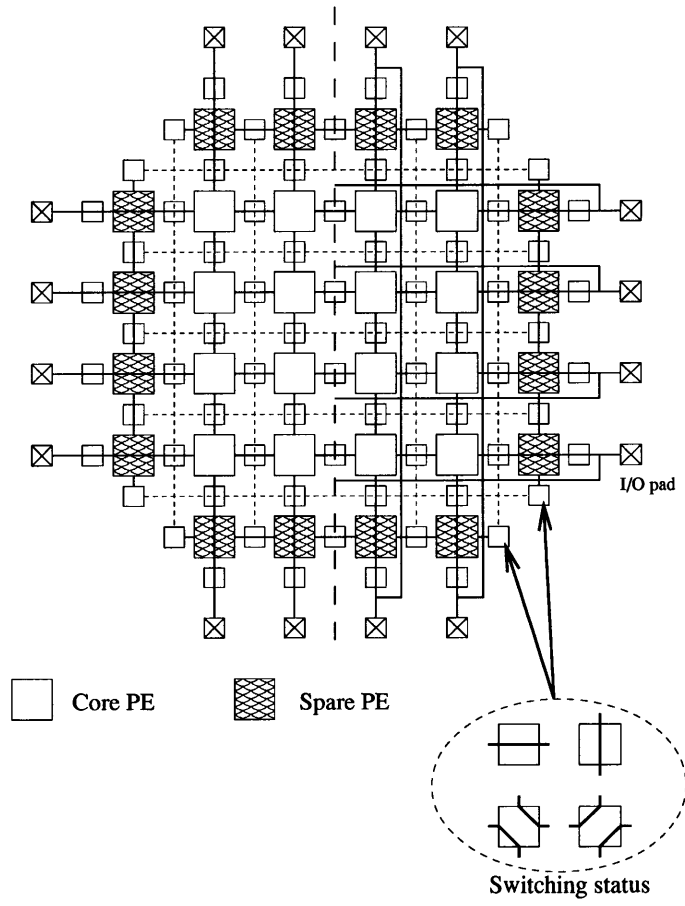


Figure 9: Fault-tolerance architecture torus network of 2D-SRT.



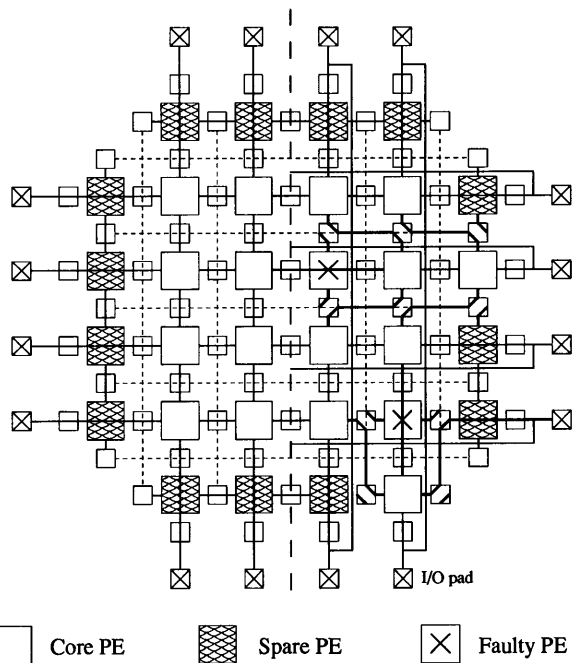


Figure 10: An example of reconfiguration mesh torus network.

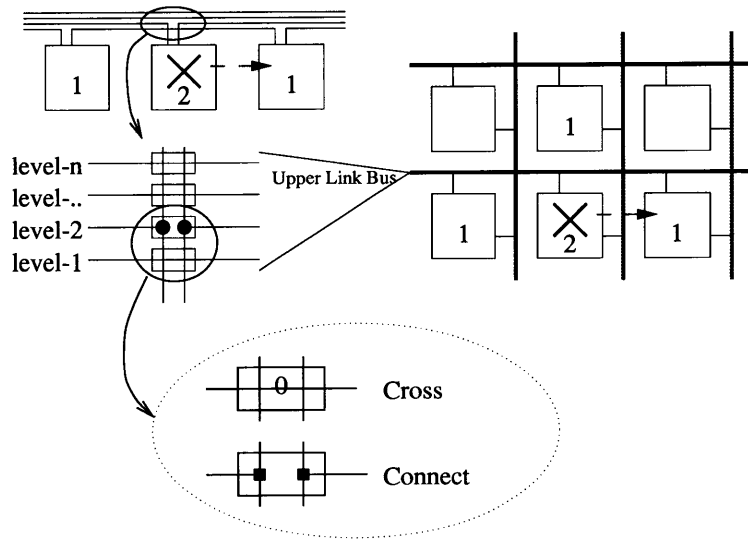


Figure 11: Horizontal reconfiguration of 2D-SRT.

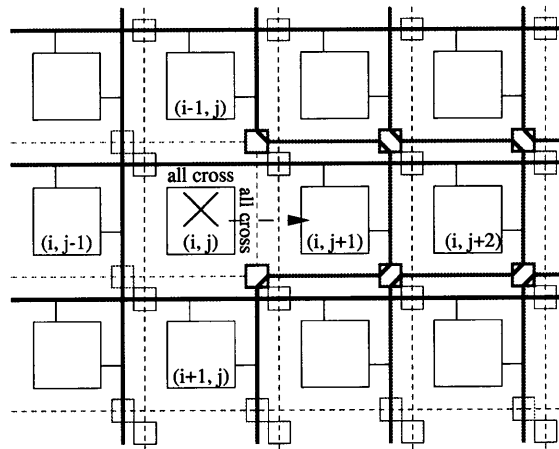


Figure 12: Vertical and horizontal reconfiguration of 2D-SRT.