A Dynamic Fault-Tolerant Mesh Architecture

Jyh-Ming Huang[1] and Ted C. Yang[2]

[1] Department of Information Engineering and Computer Sciences
Feng-Chia University
100, Wen-Hwa Rd., Sea-Tween
Taichung 407, Taiwan
jmhuang@pine.iecs.fcu.edu.tw
[2] Computer & Communication Research Laboratories
Industrial Technology Research Institute
Bldg. 51, 195-11 Sec. 4, Chung Hsing Rd.
Chutung, Hsinchu 310, Taiwan
tedyang@ccl.itri.org.tw

Abstract. A desired mesh architecture, based on connected-cycle modules, is constructed. To enhance the reliability, multiple bus sets and spare nodes are dynamically inserted to construct modular blocks. Two reconfiguration schemes are associated, and can eliminate the spare substitution domino effect. Simulations show that both schemes provide for increase in reliability over the interstitial redundancy scheme[11] and the multi-level fault tolerance mesh(MFTM)[6], at the same redundant spare ratio. Especially, with global reconfiguration, the reliability improvement ratio per spare (RIPS) can be at least twice of that of the MFTM scheme. Furthermore, the lower port complexity in spare nodes as compared to those in both of the aforementioned schemes, and versatility in reconfiguration capability are two additional merits of our proposed architecture.

1 Introduction

In general, two alternative strategies for the design of fault-tolerant multiprocessor systems are usually considered[2]. One is to gracefully degrade the performance of the system[9], the other is the so-called structure fault tolerance. In systems with structure fault tolerance, rigid topology is maintained through redundant spare element replacements caused by the detection of faults[3][9] [10] [11] [12] [13]. There are two extreme approaches of utilizing the spare during the system reconfiguration phase: global[8][13] and local reconfiguration[8].

Many researchers have been investigating into the reconfiguration schemes based on processor arrays[1][7]. Singh proposed an area efficient fault tolerance scheme, named interstitial redundancy[11]. The main attractive features are short PE interconnections and high utilization of failure-free PE’s. However, his scheme requires extensively redundant PE's with high number of ports. A multi-level fault-tolerant mesh design was proposed by Hwang[6]. When the redundancy level exceeds three, the design may not be practical because the area required for the interconnection of spare PEs may start dominating the area on the silicon, and the length of the interconnection links after reconfiguration may become objectionable.

A reliable cube-connected cycles structure, RECCC, was presented by Tzeng[12]. In that structure, buses and switches are inserted to serve as assistants for one-dimensional failed PE's reconfiguration, thus spare sharing can not be allowed between different dimensions. In addition, in some cases, when two faults are found within a window, spare substitution domino effect may take place.

To avoid the disadvantages of these schemes, we propose a compromised architecture to implement a dynamic fault tolerant scheme for mesh array. In our approach, four nodes are made into a connected-cycle module, and then multiple bus sets and soft switches are inserted for both inter-module normal connections and intra-module reconfiguration capability. To reduce the length of communication links after reconfiguration, spare nodes are inserted into the central position of a modular block.

2 Description of CCBM Structure

Our model, as shown in Fig. 1(a), consists of an m*n array of identical processing elements. For simplicity, we assume that m and n are all integer multiples of 2. To implement our architecture, we first connect four consecutive nodes in a counterclockwise direction to form a connected cycle, as shown in Fig. 1(b).

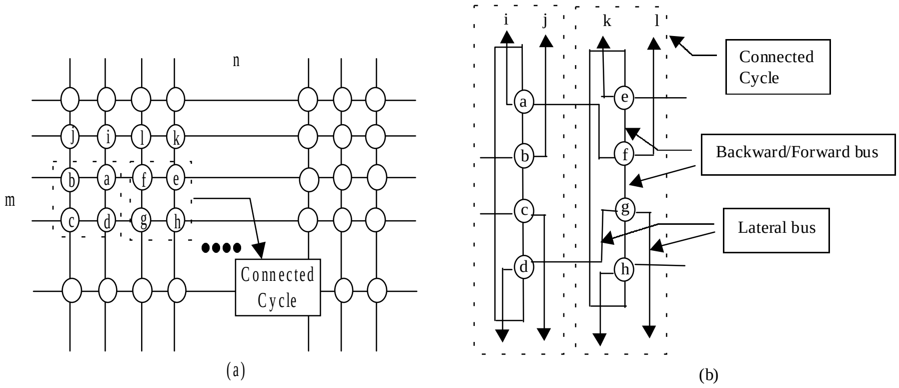

Fig. 1. (a) General model (b) The corresponding connections between two connected cycles

Fig. 2 depicts the detailed compact chip layout of the proposed fault-tolerant connected-cycle-based mesh(FT-CCBM), where spare nodes are denoted by darkened circles, and primary (active) nodes are represented by empty(white) circles filled with two numeric characters. The cb-*i*-bus, cf-*i*-bus, rl-*i*-bus and ll-*i*-bus indicates the *i*th cycle-connected backward bus, the *i*th cycle-connected forward bus, the *i*th right lateral-connected bus and the *i*th left lateral-connected bus, respectively. The cycle-connected buses are inserted to provide paths to spare, and the lateral-connected buses provide lateral connections to maintain the rigid FT-CCBM topology after reconfiguration. Switches on buses enable the buses to connect different pairs of nodes by appropriately making or breaking connections between bus segments, and between bus segments and node links. A switch can be set to one of the seven states shown in Fig. 3. To achieve multi-row fault tolerance, and avoid reconfiguration path conflict, multiple bus sets and extra switches located at the intersections of buses, and vertical reconfiguration buses that aside the spare connected cycle are needed.

Fig. 4 briefly shows the FT-CCBM structure of a conventional 2*n mesh with bus sets *i*=2. Each column represents a connected-cycle. Filled circles denote spares. By the number of given bus sets *i*, we can evenly divide the FT-CCBM into several *modular blocks*, such that each *modular block* consists of *2i*^{2} primary nodes plus *i* spare nodes. *Modular blocks* aligned in a horizontal line form a *group*.

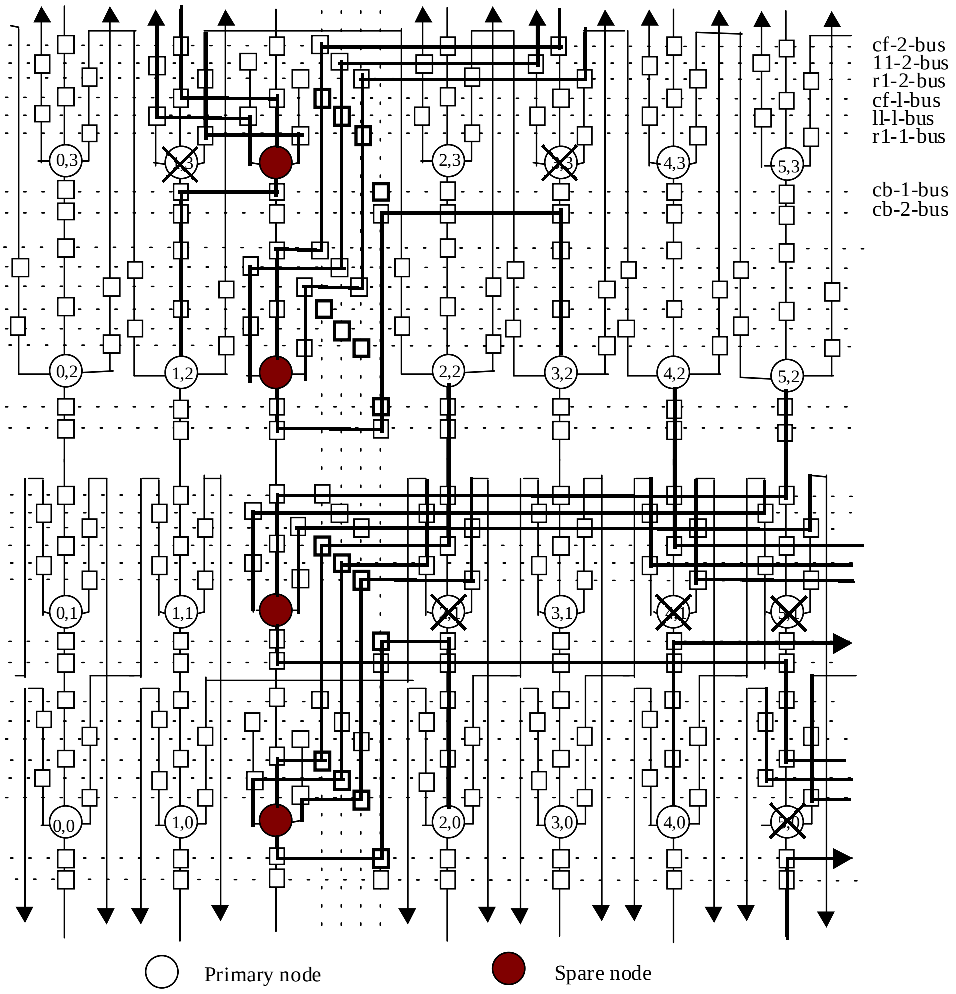

Fig. 2. A FT-CCBM compact chip layout and its reconfiguration scenarios

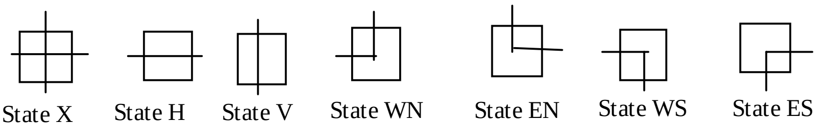

Fig. 3. Connecting states of a switch

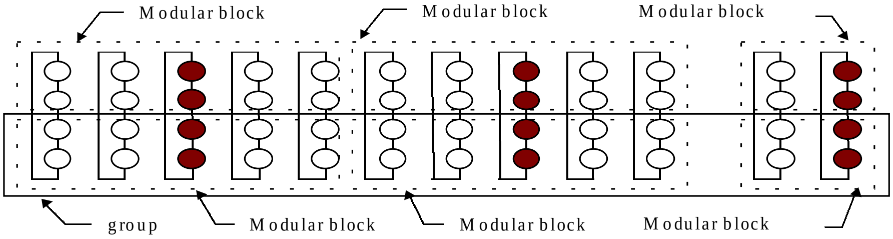

Fig. 4. The structure of 2*n FT-CCBM with *i*=2

3 Reconfiguration Schemes

Two schemes are proposed for system reconfiguration. In scheme-1, spare nodes can only replace faulty nodes in the same modular block. In scheme-2, in addition to local reconfiguration in scheme-1, partial global reconfiguration is also allowed if more than *i* nodes in a modular block are faulty.

The top half of Fig. 2 shows an example of scheme-1, with *i*=2 bus sets available. When a node(e.g. PE(1, 3)) in a modular block becomes faulty, scheme-1 first tries to replace the failed node with the spare node in the same row, by using the first bus set (i.e. cb-1-bus, cf-1-bus, rl-1-bus and ll-1-bus). If this is impossible, for example PE(3, 3), when the spare node in the same row has been used for a previous fault, then the second bus set along with the other row spare nodes are applied.

Scheme-2 is developed for overall system reliability improvement and better spare utilization. To achieve these goals, additional switches (as shown by the bolder boxes in Fig. 2) are needed. In this scheme, local reconfiguration is first performed as aforementioned. If the third faulty node is in the half modular block to the right(left) of the spare column, the available spares of the right(left) neighboring modular block will be borrowed to replace the faulty node. For instance, see the bottom half of Fig. 2, assume that the faults occur in PE(4, 1), PE(5, 0), PE(5, 1), and PE(2, 1) sequence, the faulty PE(4, 1) and PE(5, 0) are processed according to scheme-1, then when PE(5, 1) becomes faulty, because the spare nodes in the same modular block have all been used, the available spare in the left nearby modular block will be borrowed.

4 Reliability Analysis

The reliability of a single node at time t is $R_{pe}=e^{-\lambda t}$, given that the node is workable at time zero. For scheme-1, the block reliability R_{blk} can be expressed as:

$$R_{blk}=\left(\sum_{k=0}^{i}\binom{2i^2+i}{k}R_{pe}{}^{2i^2+i-k}(1-R_{pe})^{k}\right) \quad (1)$$

The following equations (2), (3) exploit the group reliability R_{g-1} and the system reliability R_{sys-1}, respectively.

$$R_{g-1}=R_{blk}^{\lfloor n/4i \rfloor} \quad (2)$$

$$R_{sys-1}=R_{g-1}^{\lfloor 2m/i \rfloor} \quad (3)$$

As for the analysis of scheme-2, we take an *i=2* group as an example, and logically rearrange the modular block boundary as region B_0, B_1, B_2, ... , B_m, and B_r, shown in Fig. 5. We can individually compute the reliability R_{b0} of region B_0, and the reliabilities of regions B_1, B_2, ... , B_m, and B_r, denoted as R_{b1}, R_{b2}, ..., R_{bm}, and R_{br}, respectively. The group reliability R_{g-2} of scheme-2 is the product of all the region reliabilities. After the reliabilities of all groups are computed, the system reliability R_{sys-2}, equation (4), can be evaluated.

$$R_{b0}=\sum_{k=0}^{i}\binom{i^2+i}{k}R_{pe}{}^{i^2+i-k}(1-R_{pe})^{k}$$ where *i* is the number of spares.

$R_{g-2}= R_{b0}*R_{b1}*R_{b2}* ....*R_{bm}*R_{br}$, and

$R_{sys-2}=R_{g-2}^{N_g}$ where N_g is the number of groups. **(4)**

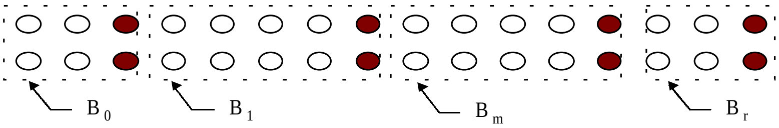

Fig. 5. A logical region view of scheme-2

5 Simulation and Comparison

A failure rate λ_{pe}=0.1 and some distinct values of the bus sets, *i*=2,3,4,…etc. are used to simulate on many different size FT-CCBM architecture. For simpilcity, however, only the simulation results for a 12*36 FT-CCBM are shown. Fig. 6 shows that the system reliability of scheme-2 is better than that of scheme-1 for the same number of bus sets. In addition, for a given redundancy ratio, maximum reliability can be achieved when the number of bus sets is 3 or 4, depending on whether a complete modular block is formed and whether spare nodes exist in the last region. Since when more bus sets are involved, the block redundant spare ratio will decrease swiftly, the system reliability will decrease if the number of bus sets exceeds 4.

We also made comparisons with the interstitial redundancy scheme[11] and the multi-level fault tolerant mesh(MFTM)[6]. Because the interstitial redundancy scheme performs only local reconfiguration and their redundant spare ratio is 1/4, therefore, we only compare its system reliability with our scheme-1. Results indicate our scheme always offers a much better reliability than the interstitial redundancy scheme.

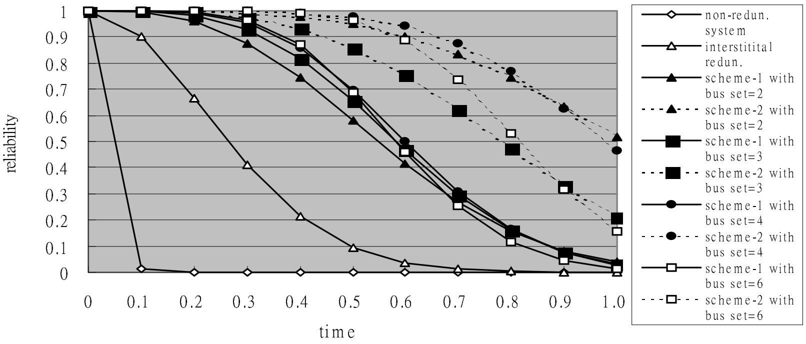

Fig. 6. System reliability of a 12*36 FT-CCBM

As for the MFTM, multi-level fault tolerance capability can be incorporated into the system. In order to have a fair comparison basis, we also evaluate the reliability improvement ratio per spare PE(RIPS) value as the MFTM did:

RIPS=(R_r-R_{non}) / The total number of spare PEs in the system,

where R_r and R_{non} are the reliabilities of the redundant and nonredundant system, respectively.

We compare our scheme-2 for systems with preferred bus sets=4 (denoted as FT-CCBM(2)) with a couple of two-level fault tolerance schemes of the MFTM, the first associates with k_1=1, k_2=1(denoted as MFTM(1,1)), and the second associates with k_1=2, k_2=1(denoted as MFTM(2,1)), where k_i is the number of spares in the ith level. Again, based on Fig. 7, our scheme in most cases, provides at least twice of the RIPS.

6 Conclusions

We have proposed a reliable structure with multi-row fault tolerance capability and spare substitution domino effect free as the two main merits. In addition, fewer ports in a spare node compared to both the interstitial redundancy scheme and the MFTM scheme, and versatile reconfiguration capabilities are two further merits of the proposed architecture.

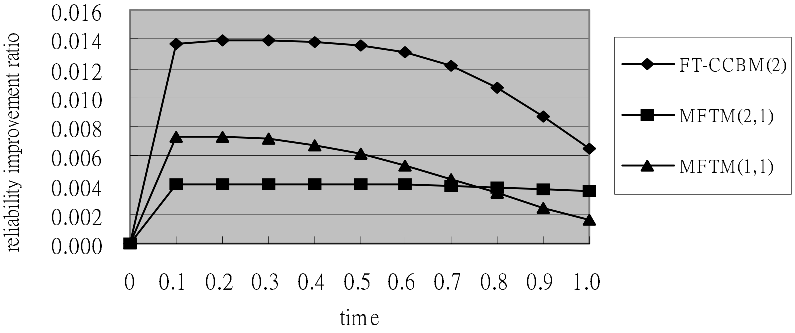

Fig. 7. The RIPS of a 12*36 mesh array with bus-sets=4

References

1. Chean, M., Fortes, Jose A. B.: A Taxonomy of Reconfiguration Techniques for Fault-Tolerant Processor Arrays. IEEE Computers, Vol. 23 (1990) 55-69
2. Chen, Y.Y., Upadhyaya, S. J.: Reliability, Reconfiguration, and Spare Allocation Issues in Binary-Tree Architectures Based on Multiple-Level Redundancy. IEEE Trans. on Computers, Vol. 42, no. 6 (1993) 713-723
3. Chen, Y.Y., Upadhyaya, S. J.: A comprehensive Reconfiguration Scheme for Fault Tolerant VLSI/WSI Array Processors. IEEE Trans. on Computers, Vol. 46, no. 12 (1997) 1363-1371
4. Chung, F.R.K., Leighton, F.T., Rosenberg, A.L.,: Diogenes: A Methodology for Designing Fault-Tolerant VLSI Processor Arrays. FTCS-13 (1983) 26-32
5. Huang, K.: Advanced Computer Architecture: Parallelism, Scalability, Programmability. Intl’s Edition, McGraw-Hill (1993)
6. Hwang, I-S.: A High Reliability Two-Level Fault Tolerant Mesh Design and Applications. Journal of Chinese Institute of Engineers, Vol. 19, no. 5 (1996) 607-613
7. Grosspietsch, K.E.: Fault Tolerance in Highly Parallel Hardware Systems. IEEE Micro, 2 (1994) 60-68
8. Kuo, S.Y., Fuchs, W.K.: Reconfigurable Cube-Connected Cycles Architecture. Journal of Parallel And Distributed Computing, 9 (1990) 1-10
9. Raghavendra, C.S., Avizienis, A., Ercegovac, M.D.: Fault Tolerance in Binary Tree Architectures. IEEE Trans. on Computers, Vol. C-33, no. 6 (1984) 568-572
10. Singh, A.D.: A Reconfigurable Modular Fault-Tolerant Binary Tree Architecture. FTCS-17 (1987) 298-304
11. Singh, A.D.: Interstitial Redundancy: An area Efficient Fault Tolerance Scheme for Large Area VLSI Processor Arrays. IEEE Trans. on Computers, Vol. 37, no. 11 (1988) 1398-1410
12. Tzeng, N.F.: A Reliable Cube-Connected Cycles Structure. Journal of Parallel and Distributed Computing, 19 (1993) 64-71.
13. Wang, M., Culter, M., Su, S.: Reconfiguration of VLSI/WSI Mesh Arrays Processors with Two-Level Redundancy. IEEE Trans. on Computers, Vol. C-38, no. 4 (1989) 547-554