YIELD ENHANCEMENT DESIGNS FOR WSI CUBE CONNECTED CYCLES ¹

JIA-JYE SHEN and ISRAEL KOREN

Department of Electrical and Computer Engineering University of Massachusetts at Amherst, MA 01003

ABSTRACT

In this paper, we present and analyze yield enhancement designs for wafer scale Cube Connected Cycles (CCC). Improvements in yield can be achieved through silicon area reduction and/or through the incorporation of defect/fault tolerance into the architecture. Consequently, we first propose a new compact layout strategy for CCC. We then present a novel implementation of wafer scale CCC based on a universal building block. This implementation facilitates the introduction of redundancy to achieve defect-tolerance. Finally, we derive expressions for the yield of various yield enhancement designs and compare them numerically for several sizes of wafer scale CCC.

I. INTRODUCTION

With the recent advances of solid state technology, WSI becomes a promising way to implement parallel architectures. One of the major problems which still remains to be solved is the extremely low yield of a complete wafer. Major efforts to overcome this problem focused on fault tolerant designs for particular architectures. The common approach is to incorporate some redundancy in the target structure and employ an appropriate reconfiguration scheme. Since area considerations have an important impact on the yield [1][2], the area overhead due to the addition of redundant elements and the corresponding interconnection links and switches should be as small as possible. Also, more compact layout strategies are highly recommended for both non-redundant and redundant structures.

Many publications have discussed the design of fault tolerant linear arrays, 2D rectangular arrays and tree structures. Only few research works, however, have addressed fault tolerance in Cube Connected Cycles (CCC). The CCC topology was proposed by Preparata and Vuillemin [3] as a better substitute of the cube connected network in a VLSI environment. In it the N processing elements (PEs) are grouped into 2^s cycles with h PEs in each, satisfying $N = h \times 2^s$. The CCC can efficiently execute a large class of problems such as sorting and discrete Fourier transform and its structure meets the requirements of VLSI technology due to its modularity, ease of layout, simplicity of communication among processing elements, etc [3]. In order

CH2680-7/89/0000/0289\$01.00 © 1989 IEEE

289

¹This work is supported in part by NSF under Contract MIP-8805586

290 International Conference on Wafer Scale Integration

to make use of this powerful parallel architecture in the WSI environment, a defect tolerant design of the CCC network is required.

Several fault-tolerant designs of CCC were previously proposed in [4] and [5]. A key consideration in both publications was to achieve reliability improvement for the CCC network when implemented in VLSI or WSI. We believe, however, that in the WSI environment yield enhancement is a more important issue than reliability improvement. We therefore, focus our attention on yield enhancement designs for wafer scale CCC networks.

In the next section, we suggest a new layout strategy for CCC, which can also be applied to the designs in [4] and [5]. We then present in Section III a new implementation of CCC, which is based on universal building blocks. These building blocks allow the construction of CCCs of any size and simplify the incorporation of redundancy into the wafer scale CCC. The expected yield of our designs is evaluated in Section IV and compared with the yield of the designs in [4] and [5].

II. A NEW LAYOUT STRATEGY FOR CCC

The layout of the fault-tolerant CCC in [4] and [5] follows the layout strategy in [3], as shown in Fig.1. Although the layout area $O(N^2/log^2N)$ is optimal with respect to the area $\times (time)^g$ measure of complexity in the VLSI grid model [3], any constant factor improvement in layout area is important when yield and propagation delay are considered. The layout strategy that we suggest for CCC as depicted in Fig.2, is similar to that in [6], but we go one step further and fold the layout. The expressions for the area of the non-redundant CCC with the original layout strategy (CCC-0) and ours (CCC-1) are shown in Eq.(1) and Eq.(2), respectively.

$$A_{CCC-0} = (M_{pe} + 1) \times 2^{s} \times (2^{s} + h - s - 1) \times M_{pe}$$
(1)

$$A_{CCC-1} = (M_{pe} + 1) \times 2^{s-1} \times (2h \times M_{pe} + 2^s - 3)$$
(2)

where M_{pe} is the width ratio between a single PE and a link. The rate of the area reduction due to our scheme depends on the size of the CCC network as well as M_{pe} . For example, for $M_{pe} = 50$ and CCC size of (h,s)=(4,3), (4,4) and (8,4), the area ratios between CCC-0 and CCC-1 are 1.975, 3.874 and 2.337, respectively. The new layout strategy also reduces the length of internal links. The maximum lateral link, which is the longest link between neighboring PEs in different cycles, in the new layout is about half of that in the original layout.

The same layout strategy can be applied to the redundant CCC with local reconfiguration scheme in [5] after performing some rearrangement of PEs in each cycle. For the redundant CCC in [4], we can use the same strategy except no folding is needed. Expressions for the area of these layouts are presented in Appendix A.

III. BUILDING BLOCK APPROACH

In this section, we propose the use of a universal building block to construct CCC of any size with/without redundancy. The main advantage of this approach is that one

can use the proposed universal building block to easily construct CCC of any size. Another advantage of the building block approach is that it can facilitate the design of fault-tolerant CCC. Two levels of redundancy can be efficiently introduced into a building block CCC without excessive area overhead and without changing the regular structure of the CCC.

However, the building block approach has its disadvantages. The main drawback of this approach is that it increases the length of the lateral links. The increase depends on the size of the building block and the location of the lateral link in the CCC network. Another drawback is the introduction of some critical components such as switches and links into the structure, although the probability that those critical components will fail is relatively low. Consequently, a more accurate analysis is needed to determine the yield enhancement capabilities of the building block approach.

Two types of building blocks have been investigated, one is a vertical block, the other is a horizontal block. Based on our study, the vertical one has less area overhead and achieves higher yield than the horizontal one for redundant building block CCC. Therefore, only the vertical building block is presented here. The vertical building block, shown in Fig.3, consists of PEs, switches and interconnection links. The structure of the building block is regular. Every PE has a degree of four and has three kinds of switches, S1, S2, and S3 associated with it. S1 and S2 connect PEs in the same cycle, and S3 is needed to connect PEs in different cycles. An example of a CCC constructed of building blocks is shown in Fig.4. For a redundant building block, spare PEs, links and switches are added to the building block but the degree of each PE remains four. Since PEs are more expensive in terms of silicon area, one of the objectives when determining the structure of the redundant building block is to utilize as many fault-free PEs as possible. The number of S1 (and S2) switches per PE in a redundant block remain one, but the number of S3 switches depends on the number of spare PEs in a redundant building block. We therefore, restrict in our study the number of spare PEs in each redundant building block to be no more than two to limit the number of S3 switches and keep the regularity and simplicity of the redundant building block's structure.

Three redundant CCC structures can be constructed by using the building block approach. One can add redundant elements to each building block and use a local reconfiguration scheme. Another possibility is to avoid internal redundancy and add instead, spare blocks requiring a global reconfiguration scheme. Such a global scheme requires the addition of some external switches and links to interconnect blocks (see Fig.5). The number of these external switches and links depends on the number of cycles in a CCC topology as well as the number of spare blocks incorporated into the building block CCC. To simplify the resulting structure of the redundant building block CCC, the number of spare cycles is restricted to be no more than two. The third possibility is to incorporate both local and global redundancy. Here, the reconfiguration process consists of two steps: first, a local reconfiguration is performed in each block if necessary, and then, after executing a diagnosis and renaming process similar to the one in [4], the spare blocks can replace malfunctioning blocks. If there are more faulty blocks than spare ones, a CCC with a smaller size can still be constructed. In the latter case, the size of the functioning CCC is no more than half of the original one. The selection of one out of the above three structures should depend on the expected number of manufacturing defects and their distribution on the wafer.

292 International Conference on Wafer Scale Integration

IV. EVALUATION OF YIELD ENHANCEMENT DESIGNS

In this section, we evaluate several yield enhancement designs for WSI CCC. We derive expressions for the yield of these designs and compare them numerically. We assume that the size of the target CCC structure fits into a single wafer and that the distribution of defects on a wafer follows the generalized negative binomial distribution presented in [7]. The derivation of a closed form expression for the yield of a defect-tolerant architecture where clustering of defects is allowed, is in general very difficult. We therefore, adopt the approach suggested in [8], according to which, an expression for the yield, y_{ccc} , of the redundant CCC is first derived based on the assumption that the defect distribution is the Poisson distribution. Then, by averaging λ , which is the expected number of defects per chip, with respect to the Gamma distribution shown in Eq.(3), the yield, Y_{CCC} , of the redundant CCC with clustering of defects is obtained, as shown in Eq.(4). Here, α denotes the clustering factor for the particular manufacturing process.

$$f(\lambda) = \frac{1}{\gamma^{\alpha} \times \Gamma(\alpha)} \times \lambda^{\alpha-1} \times e^{-\lambda/\gamma}$$
(3)

$$Y_{CCC} = \int_0^\infty y_{ecc} \times f(\lambda) d\lambda \tag{4}$$

Following the above approach, we need first to derive combinatorial expressions for the yield, y_{ecc} , of different structures with different amounts of redundancy. However, since the exact combinatorial expressions can be very complex, some of the expressions that we have developed are approximate ones and provide conservative estimates for the yield.

The expression for the yield of a building block CCC with two levels of redundancy (Block-LR-GR) is developed in what follows. To construct a Block-LR-GR, $([h/n_b]) \times (2^s + n_{orb})$ redundant building blocks are required with $(n_b + n_{rb})$ PEs in each redundant building block, where n_b and n_{rb} denote the number of normal and spare PEs in a redundant block, respectively. n_{grb} denotes the number of spare cycles and thus, $([h/n_b]) \times n_{grb}$ is the number of spare blocks in a Block-LR-GR. We first calculate the yield for a single PE, y_{pe} , a switch, y_{sw} , and a link, y_{link} , according to the area they occupy assuming the Poisson distribution for the defects. The expression for the yield of a single redundant building block, y_{block} , shown in Eq.(5), is then derived in accordance with the amount of tolerable failures of PEs, switches and links in a redundant block. In a redundant building block, the S1 and S2 switches and the links which interconnect the S1 switches are critical and have to be fault-free, which account for the first two terms in Eq.(5). The third term in Eq.(5) is the yield for the links which connect the S3 switches and the external switches. The fourth term accounts for the yield of the S3 switches. The term in the brackets is the probability that no more than n_{rb} PEs are faulty and that the corresponding links which are needed to route around the faulty PEs, and the S3 switches as well as links which route to lateral links, are fault free.

The expression for the yield of a Block-LR-GR, $y_{Block-LR-GR}$, is then derived based on the amount of tolerable malfunctioning blocks and it depends on the probability of a fault in the external links and switches which interconnect among the blocks.

Circuits and Technology-II 293

In Eq.(6), A_0 denotes the area of CCC-1 with size (h,s)=(4,3), based on which the average defect density and clustering parameter are obtained for a particular manufacturing process. $A_{Block-LR-GR}$ denotes the area of a Block-LR-GR and is shown in Appendix A. The first term in Eq.(6) is therefore the area increase ratio. The second term is the yield of operational blocks which constitute the CCC network. The third and fourth terms account for the yield of the switches and links which interconnect blocks to form logical cycles, if necessary. The last two terms represent the yield of external switches and links which correspond to the lateral links in the CCC topology.

$$y_{block} = y_{sw}^{2(n_b+n_{rb})+1} \times y_{link}^{n_b+n_{rb}-1} \times y_{link}^s \times y_{sw}^{n_b} \times \\ [\sum_{i=0}^{n_{rb}} C_i^{n_b+n_{rb}} \times y_{pe}^{(n_b+n_{rb}-i)} (1-y_{pe})^i (y_{link}y_{sw})^i \times y_{link}^{(n_b+i-1)}]$$
(5)

$$y_{Block-LR-GR} = A_0 / A_{Block-LR-GR} \times y_{level}^{[h/n_b]} \\ \times (y_{sw}^{n_{grb}}y_{link})^{2([h/n_b]-1)(2^*+n_{grb})} \\ \times (y_{link})^{2n_{grb}([h/n_b]-1)(2^*+n_{grb}-1)} \\ \times (y_{sw}y_{link})^{s \times (2^*+n_{grb})} \times y_{link}^{s2^{r-1}}$$
(6)

where

$$y_{level} = \sum_{i=0}^{n_{grb}} C_i^{2^e + n_{grb}} y_{block}^{2^e + n_{grb} - i} (1 - y_{block})^i$$
(7)

Similar expressions for all other schemes investigated have been derived, but are not presented in this paper for the sake of brevity. The expressions were then averaged using equation(4) and the resulting yield for all the above mentioned structures (with different amounts of redundancy) was numerically compared. We present here the results for only seven yield enhancement designs. These include :

- 1. Non-redundant CCC with the layout strategy described in Sec. II. (CCC-1).
- 2. Locally reconfigurable CCC in [5]. (LR-CCC-0).
- 3. Locally reconfigurable CCC with the layout strategy in Sec. II. (LR-CCC-1)
- 4. Cubical Ring Connected Cycles in [4]. (CRCC-0).
- 5. Cubical Ring Connected cycles with the layout strategy in Sec. II. (CRCC-1).
- 6. Building block CCC with local redundancy only. (Block-LR).
- 7. Building block CCC with both local and global redundancy. (Block-LR-GR).

The comparison was done in the following way. We first compared the yield for the schemes which do not employ the building block approach with their original layout strategy and the layout strategy presented in Sec II. This numerical comparison was done for three different sizes of a CCC, (h,s)=(4,3), (4,4) and (8,4). For these sizes, we found substantial improvements due to the new layout strategy and LR-CCC-1 proved to be the best scheme among these with the highest yield. Fig.6 shows this type of comparison for the size (h,s)=(4,3). We then compared the yield for the schemes with building blocks and with different amounts of redundancy. Some of the results are

294 International Conference on Wafer Scale Integration

depicted in Fig.7. Either Block-LR or Block-LR-GR is the best among them depending on the average defect density and the clustering parameter. Finally, we compared the best schemes found in the non-building block and building block approaches. Fig.8-10 illustrate this type of comparison for sizes (h,s)=(4,3),(4,4) and (8,4), respectively. An interesting phenomenon is the dependency of the optimal defect-tolerance technique on the size of the CCC network. When the structure sizes are (h,s)=(4,3) and (4,4), LR-CCC-1 is the best choice when defects are clustered or when the defect density is low, while Block-LR-GR is the best when defects are more evenly distributed and the defect density is high. For (h,s)=(8,4), Block-LR-GR is the best choice. In principle, a defect-tolerant scheme with two levels of redundancy (i.e., local and global redundancy) can tolerate a larger number of defects. However, it may lose its superiority if the area overhead dedicated to the interconnection of redundant elements is too high.

V. CONCLUSIONS

Yield enhancement designs for WSI CCC were presented and analyzed in this paper. We proposed a new layout strategy for CCC, which can reduce the layout area for WSI CCC network. With slight modifications, this layout strategy can be applied to the previously suggested fault-tolerant designs for CCC in [4] and [5]. This way, we achieve yield improvement through area reduction. We then presented a universal building block which can be used to construct CCC of any size. Based on this building block, a yield enhancement design for building block CCC was suggested. With the proper selection of design parameters, such as the amount of redundancy, and reconfiguration scheme, higher yields can be obtained. The main contribution of this research is the development of a methodology to analyze the yield improvement of a given defect-tolerant technique when applied to a given size of a CCC network.

APPENDIX A: LAYOUT AREA

For brevity, only expressions for the area of LR-CCC-1, CRCC-1 and Block-LR-GR are shown. This notation is defined in Section IV.

$$A_{LR-CCC-1} = (M_{pe} + 2M_{sw} + \lceil ng/2 \rceil + 1 + lg) \times (2(h + h/ng)(M_{pe} + 2M_{sw}) + 2^{s} - 3) \times 2^{s-1}$$
(8)

where ng is the size of group in LR-CCC [5] and M_{sw} is the width ratio between a switch and a link. lg is the amount of vertical tracks occupied by the links which interconnect groups in the same cycle in LR-CCC [5]. $lg = \lfloor h/ng \rfloor - 1$, if there are no more than two groups in a cycle; otherwise, lg = 2.

$$A_{CRCC-1} = ((M_{pe} + 2M_{sw} + 1)(2^{s} + 1) + 2) \times ((M_{pe} + 2M_{sw})h + 2(2^{s} - 1))(9)$$

$$\begin{array}{lll} A_{Block-LR-GR} &= & (\lceil h/n_b \rceil \times ((M_{pe} + M_{sw}) \times (n_b + n_{rb}) + M_{sw}) + 2^s - 1 + \\ & & (2(\lceil h/n_b \rceil - 1) \times n_{grb} + s) \times M_{sw}) \\ & & \times (M_{pe} + (2 + \lceil n_b/2 \rceil) \times M_{sw}) \times (2^s + n_{grb}) \end{array}$$
(10)

References

- I. Koren and M.A. Breuer, "On area and yield considerations for fault- tolerant VLSI processor arrays", *IEEE Trans. Comput.*, Vol. C-33, Jan. 1984, pp.21-27.
- [2] I. Koren and D.K. Pradhan," Modeling the effect of redundancy on yield and performance of VLSI systems", IEEE Trans. Comput., Vol. C-36, Mar. 1987, pp.344-355.
- [3] F.P. Preparata and J. Vuillemin," The Cube-Connected Cycles: A versatile network for parallel computation", Comm. Ass. Comput. Mach., May 1981, pp.300-309.
- [4] P. Banerjee, "The cubical connected cycles: A fault-tolerant parallel computation network", IEEE Trans Comput., Vol. 37, May 1988, pp.632-636.
- [5] P. Banerjee, S.Y. Kuo and W.K. Fuchs, "Reconfigurable Cube-Connected Cycles architectures", in Proc. 16th Ann. Symp. Fault-Tolerant Compt., June 1986, pp.286-291.
- [6] M.S. Krishnan and J.P Hayes, "An array layout methodology for VLSI circuits", *IEEE Trans Comput.*, Vol. C-35, Dec. 1986, pp.1055-1067.
- [7] C.H.Stapper, F.M. Armstrong and K. Saji, "Integrated circuit yield statistics", Proc. IEEE, Vol. 71, Apr. 1983, pp.453-470.
- [8] I. Koren, Z. Koren and D.K. Pradhan," Designing Interconnection buses in VLSI and WSI for maximum yield and minimum delay" *IEEE*, J. Solid-State Circuits, June 1988, pp.859-866.



Fig.1: A non-redundant CCC, $N = 4 \times 2^3$, with the layout strategy in [3].



Fig.2: A non-redundant CCC, $N = 4 \times 2^3$, with our layout strategy.



Fig.3: A vertical building block with 2 spare PEs.





Fig.4: A building block CCC with folding, $N = 8 \times 2^4$, block size=4.

Fig.5: Block-LR-GR, with 1 spare PE in each block, and 1 spare block; $N = (4+1) \times (2^3 + 1).$



Fig.6: The yield as a function of the average number of defect for (h,s)=(4,3) and the designs: CCC-1; LR-CCC-0(x); LR-CCC-1(x); CRCC-0; CRCC-1; where x is the size of a group= 4,2; and $\alpha = 1$.



Fig.7: The yield for (h,s)=(4,3) and the building block designs: Block-LR (n_{rb}) and Block-LR-GR (n_{rb},n_{grb}) , where $n_{rb}=1,2$ and $n_{grb}=2$; $\alpha = 1$.



average number of defects Fig.8: Comparing the yield of three yield enhancement designs: Block-LR(1), Block-LR-GR(1,1) and LR-CCC-1(4) for (h,s)=(4,3), $\alpha = 1$.



Fig.9: Comparing the yield of four designs, LR-CCC-1(x) with x=4,2, Block-LR-GR(1,2) and Block-LR-GR(2,2) for (h,s)=(4,4) and $\alpha = 4$.



Fig.10: Comparing the yield of four designs, LR-CCC-1(x) with x=4,2, Block-LR-GR(1,2) and Block-LR-GR(2,2) for (h,s)=(8,4) and $\alpha = 4$.