Topological Optimization of PLAs for Yield Enhancement [1]

Venkat K. R. Chiluvuri and Israel Koren
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst, MA 01003, USA

Abstract

Several topological optimization techniques have been developed to minimize the area of PLAs. Significant yield enhancement can also be achieved by minimizing the defect sensitivity of a design that is already optimized for area. In this paper, we propose a yield enhancement technique through which the defect sensitivity of the design will be minimized without increasing the area. This reduction in critical area is achieved primarily by minimizing the wire lengths in several layers of the layout. The yield enhancement results of the proposed technique on some benchmark PLA examples are presented.

1 Introduction

During the last decade, many structured design techniques have been developed to minimize the design cycle time of VLSI systems. PLAs, gate arrays and standard cell designs are some of the popular design styles. Two-level PLA-based logic synthesis is well developed and commercial automatic synthesis tools/silicon compilers are now available. One of the major drawbacks of these design styles is the large area overhead (due to sparsity of the personality matrices of most designs) and the attendant yield and performance degradation.

Several optimization techniques have been proposed to minimize the area of PLAs at various stages of the design, starting from functional design to physical design. PLA folding techniques and the corresponding software tools have been developed to optimize the topological representation of PLAs [3, 4, 10]. The primary objective of all these techniques is to reduce the area of the PLA. Significant yield enhancement can also be achieved by minimizing the defect sensitivity of the design that is already optimized for area. In this paper, we are proposing a yield enhancement technique through which the defect sensitivity of the design is minimized without increasing the area. We refer to this new approach as 'topological optimization for yield enhancement'. In this approach, the topological representation of the PLA is altered so that the critical area of the generated layout is minimized. This reduction in critical area is achieved primarily by minimizing the wire lengths in several layers of the layout.

[1] This work was supported in part by NSF, under contract MIP-9013013 and by IBM, General Technology Division, VT, under contract 81775.

0-8186-3502-9/93 $3.00 © 1993 IEEE

A variety of fault-tolerant techniques have been proposed for PLA-based designs in order to enhance yield [5, 8, 7, 13]. These fault-tolerant techniques have proved to be very effective in certain situations but involve a cost of additional area and design effort. In contrast, our approach does not require any additional area.

In a typical PLA implementation, the AND and OR planes consist of poly (for input lines), metal (product lines and ground) and diffusion layers [10]. Circuit yield can be improved by minimizing the critical area of some of these layers. In the module generators *octopus* & *panda* of the OCT and MAGIC systems [11, 12], input poly lines are truncated after the most distant gate to reduce capacitance. This reduction in wire length also results in a smaller critical area for short-circuit type faults. By modifying the topological representation of the PLA, wire lengths in the physical layout can be minimized without increasing the area, resulting in layers with reduced critical area. The amount of wire reduction in various layers depends on the folding level, technology and constraints imposed on the input/output positions, etc. For example, in simple column folding with constraints on the inputs, higher reduction in wire length is possible. We illustrate the PLA topological optimization with the help of some MCNC benchmark PLA examples [9].

2 Topological Optimization

The first example, *misex1*, consists of 8 input lines (16 lines with complements), 7 output lines and 12 product terms. An array optimizer (*genie*), module generator (*octopus*) and other support tools in OCT & MAGIC systems were employed to generate the CIF file. This CIF file is used to estimate the critical area of the layout using the yield analysis tool *Laser* [2]. First, logic minimization was performed using *espresso*. Then, simple column folding was executed by *genie*. *Octopus* and *sparcs* were used to generate the final compacted PLA layout. The original symbolic representation of the circuit is shown in Figure 1A. The inputs run from both the top and bottom. An input and its complement are constrained to lie adjacent to each other. The outputs run from the top. The primary objective of our topological optimization is to minimize the length of the poly input lines running from the top and bottom. The *genie* output file was modified to minimize the input wire length in the AND plane. Input wire length is reduced by 16.5%, from 115 to 96 units, after topological optimization. The optimization procedure is briefly explained below.

2.1 The Algorithm

The length of the input poly lines is reduced by permuting the product terms in row positions. First, column weights are calculated. Column weight is defined as the total number of transistors in that column. Then, a column with the least column weight is selected and, if all topological constraints are satisfied, the product terms of that column are placed in the row positions starting from the input buffer side. Column weights are then updated and another column with the least weight is selected. This is a greedy method and it is much faster than simulated annealing methods used for PLA folding. For non-folded and simple column-folded PLAs this greedy method guarantees optimum solution. A brief outline of the proof of optimality is given below.

Let us consider a very simple non-folded PLA with 3 inputs and an arbitrary number of product terms. Each input appears in several product terms of the PLA. The number of product terms in which an input will appear is defined as its column weight. The three input column weights are denoted by T_A, T_B and T_C. If all the product terms of input A are placed first from the input buffer side, then those of input B followed by input C, the total input wire length, denoted by I_{WL}, is given by

$$\begin{aligned} I_{WL} &= T_A + (T_A + T_B - T_{AB}) + (T_A + T_B - T_{AB}) + (T_C - T_{AC} - T_{BC} + T_{ABC}) \\ &= 3T_A + 2T_B + T_C - 2T_{AB} - T_{AC} - T_{BC} + T_{ABC} \end{aligned} \quad (1)$$

where T_{AB} is the number of product terms which include both A and B. The total input wire length for an N-input PLA can be written as

$$I_{WL} = N \cdot T_1 + (N-1)T_2 + (N-2)T_3 + ... + T_N - (N-1)T_{12} - ..... + T_{12...N} \quad (2)$$

To minimize the total input wire length, no column weight should be greater than the proceeding one. If two inputs appear in the same number of product terms, then the maximum number of common product terms with other inputs should be taken into consideration. The final optimal order of product terms is independent of their initial order. In a simple column folded case, product terms will be processed from either the top or the bottom. To calculate the column weight both inputs in that column will be considered. While the input column weight for one input will be the number of product terms which contain this input, the weight of the other input (in the opposite direction) will be the number of product terms which do not have this input. For example, in a PLA with 10 product terms, consider a folded column, where the bottom input appears in three product terms and the top input in two product terms. Then the column weight is $3 + (10 - 3 - 2)$. Once a product term with a top input is assigned a row position, then that column will not appear in further column weight calculations, since the input wire length will not change afterwards. A column will be considered only if the topological constraints are not violated. If the topological constraints are not satisfied then a column with the next smallest weight will be considered. The complete proof of optimality is omitted for the sake of brevity.

3 Critical Area Analysis

The rearrangement of product terms in row positions does not have any negative impact either on the performance or on the area. It may even improve them due to reduction in the input poly wire length. The output lines in the OR plane (metal-1), which run from either top or bottom, can also be included in the objective function. The optimized symbolic layout is shown in Figure 1B. The wire length and the critical area of open- and short-circuit type faults of polysilicon layer are shown in Table 1. The performance of the modified layout has been verified using *crystal*. There is no change in the maximum delay of $1.93ns$ in the *misex1* PLA circuit.

Due to this optimization, 19% wire length reduction is achieved in the polysilicon layer of the AND plane. Consequently, the critical area of this layer is reduced by 17%. It is interesting to observe the incidental reduction in the wire length of the other layers, e.g., metal-1 and diffusion layers as well as in the overall area. The critical areas of the metal-1

Parameter	Original layout	After topological optimization		After physical optimization		Total reduction
			reduc.		reduc.	
Poly wire length	$5.83mm$	$4.73mm$	19.0%	—	—	19.0%
A_c of short-ckt faults	$144\mu m^2$	$129\mu m^2$	10.4%	$145\mu m^2$	-11.0%	-0.7%
A_c of open-ckt faults	$665\mu m^2$	$542\mu m^2$	18.5%	$467\mu m^2$	13.8%	29.8%
A_c total	$809\mu m^2$	$671\mu m^2$	17.1%	$612\mu m^2$	8.8%	24.4%
Area	$0.024mm^2$	$0.021mm^2$	12.5%	—	—	12.5%
Delay	$1.93ns$	$1.93ns$	—	$1.93ns$	—	—

Table 1: Effect of topological optimization on the primary layer of *Misex1* PLA.

Fault type	Critical area (μm^2)					
	Original Layout	After topological optimization		After physical optimization		Total reduction
			reduc.		reduc.	
Metal-1 shorts	728	715	1.8%	752	-5.2%	-3.3%
Metal-1 opens	1021	964	5.6%	865	10.3%	15.3%
Diffusion shorts	241	230	4.6%	224	—	—
Diffusion opens	559	533	4.7%	524	—	—
Total	3358	3113	7.3%	2977	4.4%	11.4%

Table 2: Effect of topological optimization on other layers of *Misex1* PLA.

and the diffusion layers are shown in Table 2. The critical area is reduced by 7.3% in these two layers. The reduction in wire length of the individual layers may result in better compaction thereby reducing the overall area. In this example, the area is reduced by 12.5%. Therefore, the effective yield of the die also increases by approximately that percentage [1]. However, this much area reduction may not always be possible when performing yield optimization, if better folding tools are used [4]. In such a case, the percentage reduction in wire length and in the attendant critical area may be less. The reduction in the wire length of several layers facilitates the implementation of the physical layout design techniques for yield enhancement proposed in [1]. After implementing these changes in the layout (Figure 2), the critical area is further reduced by 8.8%. Therefore, the overall reduction in the critical area is about 24% in the polysilicon layer and about 11% in metal-1 and diffusion layers. In larger chips, these reductions in the critical areas and the chip area result in about 15% improvement in the yield [1].

The above modifications were restricted to the AND plane of this small PLA. Similar reduction in the poly lines of the OR plane is possible by permuting the output lines in the columns. In this case we may need to consider the positional constraints on the output lines. When the overall circuit is considered, the reduction in wire length of the polysilicon layer is only 7.2%. In PLAs of practical size, the AND and OR planes constitute more

Example				Symbolic level			Physical level					
							Critical area (μm^2)					
Name	I/P	O/P	Prod. terms	I/P wire length			Open-ckt faults			Short-ckt faults		
				original layout	optimal layout	% red.	original layout	optimal layout	% red.	original layout	optimal layout	% red.
clip	9	5	118	1973	1590	19	31465	26130	17	8680	5764	34
squar5	7	7	26	242	199	18	4065	3457	15	1236	805	35
inc	7	7	30	381	306	20	1602	1317	18	528	367	31

Table 3: Effect of topological optimization on critical area of polysilicon layer.

than 80% of the PLA area. Therefore, in larger PLAs the overall reduction will be close to that of the AND plane (and OR plane). Results of the topological optimization on larger PLA examples are shown in Table 3. In the *clip* PLA, there are 118 product terms. After topological optimization in the AND plane, the input wire length is reduced by 19%. This reduction in wire length at the topological level resulted in a 17% reduction in the critical area of open-circuit type faults and 34% reduction in the critical area of short-circuit type faults in the polysilicon layer of the physical layout. If the physical layout optimization techniques are applied on this layout, further reduction in the critical area can be achieved.

4 Conclusions

We have illustrated the significance of topological optimization for yield enhancement in PLA-based designs. This new approach for yield enhancement has many attractive features compared to conventional methods. Yield enhancement in PLAs through redundancy (spare product terms, input and output lines) and reconfiguration is proposed in [7, 8, 13]. In this approach, additional resources such as spares, testing aids, reconfiguration circuitry etc., require extra area up to 25%. Due to these additional area requirements, effective yield will go down beyond a certain optimal level. The marginal increase in yield for a spare will go down as the number of spares increases. In this approach, performance degradation is another concern. In the new approach proposed above, no additional resources are required to achieve the yield improvement. In a majority of the cases it may even result in better performance, due to reduction in wire length and overall area. We have considered wire length reduction only in the polysilicon layer. Similar yield optimization techniques can be applied for metal-1 (in the OR plane) and diffusion layers (in both planes) as well. The overall reduction in the critical area will then be even higher.

To achieve a higher yield for very complex VLSI/ULSI systems, they must be designed for yield. It is not enough just to look for area minimization. We should consider how various elements are arranged within the given area. Existing CAD systems need to be supplemented by yield optimization tools. We demonstrated through topological optimization for PLA-based designs that yield enhancement techniques can be applied even at a higher level of design abstraction. Designers do not have to wait until the final layout is generated to apply yield enhancement techniques. Similar yield enhancement techniques can be applied for other popular structured design styles. It is possible to achieve about 15% yield enhancement by applying these techniques at various stages of the design synthesis.

References

[1] V. K. R. Chiluvuri and I. Koren, "New Routing and Compaction Strategies for Yield Enhancement," *IEEE Int. Workshop on Defect and Fault Tolerance in VLSI Systems*, pp. 325-334, November 1992.

[2] J. P. Gyvez and Chennian Di, "IC Defect Sensitivity for Footprint-Type Spot Defects," *IEEE Trans. Computer-Aided Design*, Vol. 11, No. 5, pp. 638-658, May 1992.

[3] G. D. Hachtel, A. R. Newton and A. Sangiovanni-Vincentelli, "An Algorithm for Optimal PLA Folding," *IEEE Trans. Computer-Aided Design*, Vol. CAD-1, No. 2, pp. 63-76, April 1982.

[4] S. Y. Hwang, R. W. Dutton and T. Blank, "A Best-First Search Algorithm for Optimal PLA Folding,"*IEEE Trans. Computer-Aided Design*, Vol. CAD-5, No. 3, pp. 433-442, July 1986.

[5] I. Koren and A. D. Singh, "Fault Tolerance in VLSI Circuits," *Computer*, Special Issue on Fault-Tolerant Systems, Vol. 23, pp. 73-83, July 1990.

[6] I. Koren and C. H. Stapper, "Yield Models for Defect Tolerant VLSI Circuits: A Review," *Defect and Fault Tolerance in VLSI Systems*, Vol. 1, I. Koren (ed.), pp. 1-21, Plenum, New York, 1989.

[7] S. Kuo and W. K. Fuchs, "Fault Diagnosis and Spare Allocation for Yield Enhancement in Large Reconfigurable PLA's," *IEEE Transactions on Computers*, Vol. 41, No. 2, pp. 221-226, February 1992.

[8] R. Leveugle, M. Soueidan and N. Wehn, "Defect Tolerance in a 16 Bit Microprocessor," *Defect and Fault Tolerance in VLSI Systems*, Vol. 1, I. Koren (ed.), pp. 179-190, Plenum, New York, 1989.

[9] "MCNC Benchmark PLA Examples," *1992 International Workshop on Layout Synthesis*, North Carolina.

[10] G. D. Micheli, et al., "A Design System for PLA-Based Digital Circuits," *Advances in Computer-Aided Engineering Design*, Vol. 1, A. Sangiovanni-Vincentelli (ed.), Jai Press Inc, Greenwich, 1985.

[11] "OCT Tools Reference Manual," *M^2C CAD Tools Distribution*, Release 7.0, 1992.

[12] J. K. Ousterhout et al., "Magic: A VLSI Layout System," *21st IEEE Design Automation Conference*, pp. 152-159, 1984.

[13] C. Wey, "On Yield Consideration for the Design of Redundant Programmable Logic Arrays," *IEEE Trans. Computer-Aided Design*, Vol. 7, No. 4, pp. 528-535, April 1988.

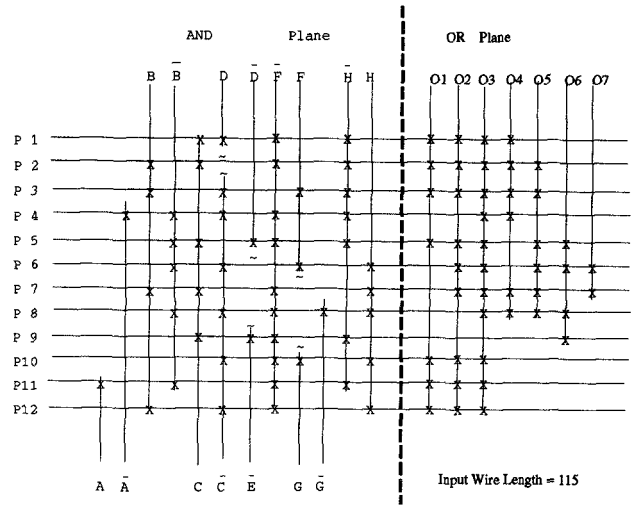

Figure 1a: Misex1 PLA Original Topological Representation

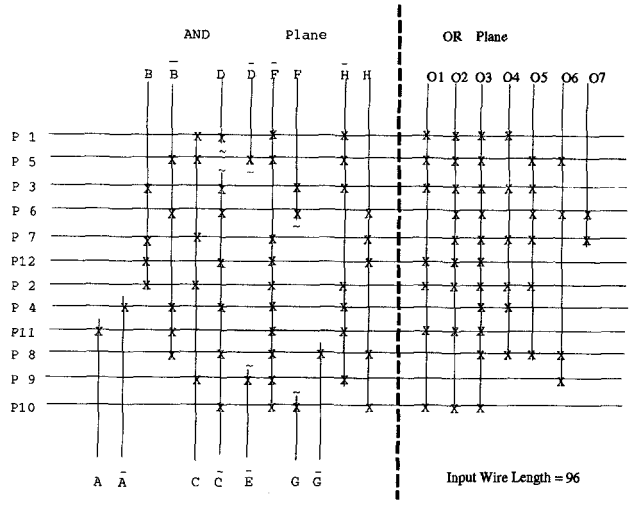

Figure 1b: Misex1 PLA Optimized Topological Representation

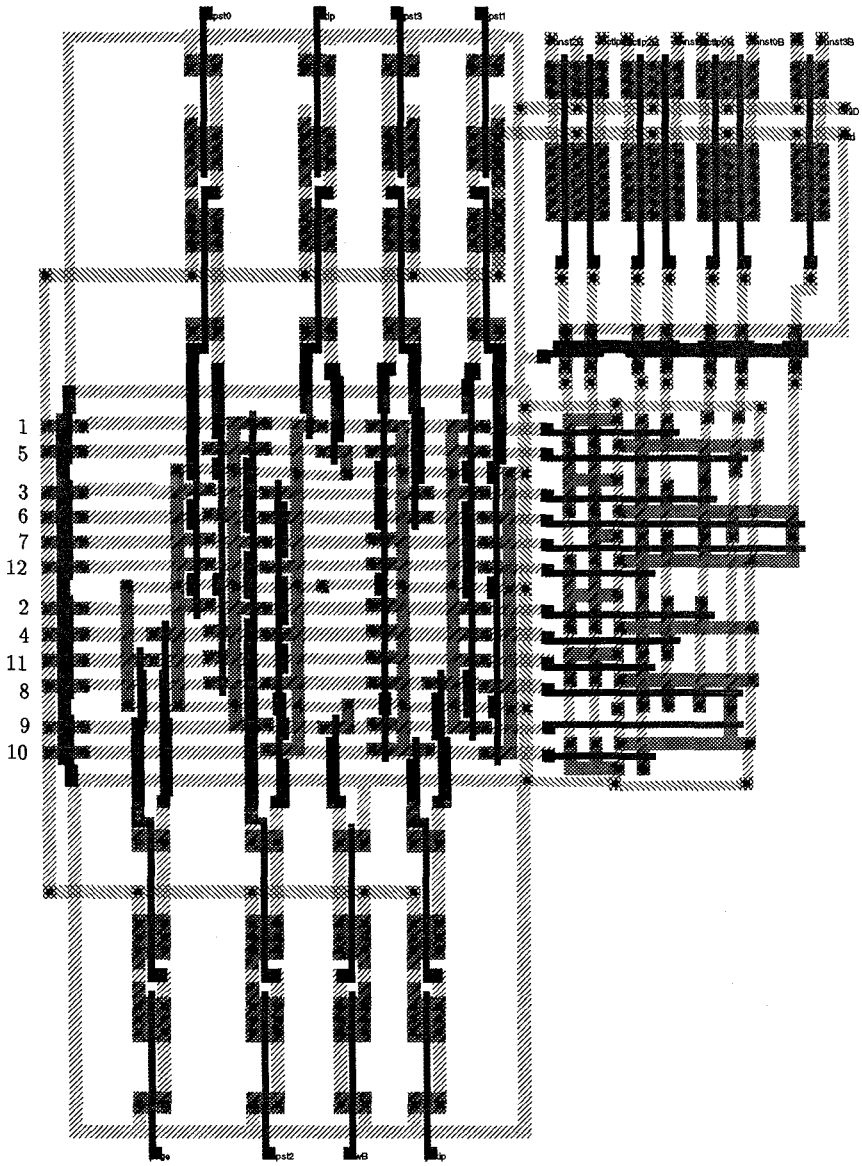

Figure 2: Final Layout of Misex1 PLA