Floorplanning of Memory ICs: Routing Complexity vs. Yield

Israel Koren and Zahava Koren

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

ABSTRACT

It has recently been shown that for very large chips, especially those with some incorporated redundancy, the chip's floorplan may affect its yield. When selecting a floorplan, the designer should, therefore, consider the expected yield in addition to the traditional objectives such as area, performance, and routing complexity.

This paper studies the two seemingly unrelated objectives of routing complexity minimization and yield maximization, and justifies the need for a trade-off analysis when determining the floorplan. We will focus on the analysis of large memory ICs with redundant modules, for which several alternative floorplans may exist.

Keywords: Floorplanning, memory, redundancy, yield, routing complexity.

1. INTRODUCTION

In the floorplanning stage of the chip design, the relative position of the building blocks (modules) of the chip is determined. In the case of a microprocessor, these building blocks include data cache, instruction cache, instruction decode unit, fixed-point arithmetic unit, and alike. In the case of memory ICs the building blocks include the basic memory sub-arrays and various redundant circuits.

The selected floorplan should depend on the preferred mutual position of the modules, which, if wiring length is concerned, is determined by the number of nets that connect them. Modules that have a large number of common nets should be placed as close as possible in order to reduce the total wiring length, which in turn may reduce the chip area. This would also result in a high performance design by reducing the communication delays among the modules. Unfortunately, the problem of finding the floorplan with the minimal wiring is, in the general case, NP-complete. Consequently, various heuristic algorithms for solving it have been proposed and are currently being used (see for example,[2,9]).

Until recently, the expected yield of the chip was rarely considered when selecting a floorplan for an IC. This is justified if the chip is either small or has a defect distribution which can be accurately described by the Compound Poisson yield model with large area clustering (i.e., the size of the defect clusters is larger than the size of the chip). For such chips, the selected floorplan has no effect on the projected yield of the designed chip.

With the recent introduction of integrated circuits with a total area of $2cm^2$ and up, ignoring the yield during the floorplanning stage is no longer justified. These large area ICs usually consist of different types of modules and may have some incorporated redundancy. We have shown[4] that if such chips are hit by medium sized defect clusters, then changes in the floorplan can greatly affect their projected yield. We have further shown that the optimal expected yield highly depends upon the position of the redundant modules relative to the primary modules. Since the relative position of the primary and redundant modules will undoubtedly affect the length of the wires connecting them, there is a need to carefully study the relationship between the objectives of wiring cost minimization and yield maximization, and explore the possible trade-offs between the two.

In the above mentioned paper[4] we have also shown that there are two types of ICs for which the projected yield may be affected by the floorplan. In the first type, the individual modules in the chip have substantially different device densities, leading to very different sensitivities to manufacturing defects. Most current microprocessors are of this type since they include very dense modules like data and instruction cache on one hand, and very sparse

Further author information: (Send correspondence to I. Koren)
E-mail: koren@euler.ecs.umass.edu

modules like instruction decoder, on the other hand. In the second type, the IC contains several spare modules to be used upon the failure (due to manufacturing defects) of a primary module. Memory ICs, especially very large ones, tend to contain redundant modules, and thus fall into the second category.

In another paper[7] we analyzed the relationship between the yield and wiring objectives for the first type of ICs and illustrated our results through an example of a 64-bit microprocessor. The most important conclusion drawn from this study was that even if the wiring cost is considered to be of utmost importance, the yield can still be maximized within the set of floorplans with the minimal wiring length or very close to it. The main goal of this paper is to study the same relationship between the yield and wiring objectives for the second type of ICs, which includes some of the more recent designs of memory chips.

This paper is organized as follows. In Section 2 we provide a simple example of the relationship between the chip floorplan, its projected yield, and its wiring cost. In Section 3 we analyze the trade-offs between the two objectives for an actual memory chip, and Section 4 concludes the paper.

2. A SIMPLE EXAMPLE

To illustrate the possible trade-offs between the yield and the wiring cost for a chip with redundancy, consider the following example of a chip which consists of nine modules, K_1, K_2, K_3, L_1, L_2, L_3, M_1, M_2 and M_3, where two modules of each type (K_i, L_i and M_i) are sufficient for the proper operation of the chip. The three topologically distinct floorplans for this chip are depicted in Figure 1.

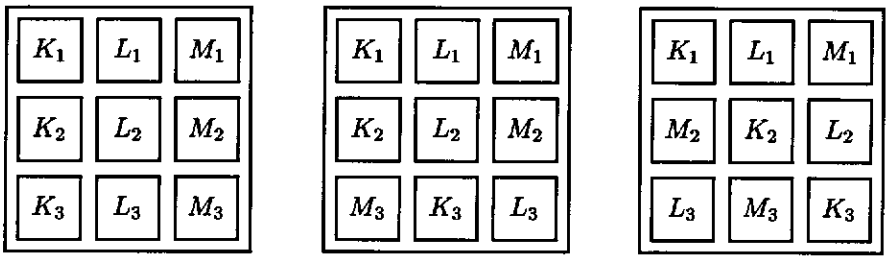

Figure 1. Three floorplans of a 3×3 array.

To estimate the expected yield of this chip, we use the negative binomial fault distribution with medium-sized clusters.[3] Under this model, the chip is divided into disjoint ares called "blocks", where the faults in each block have a negative binomial distribution and the faults in different blocks are statistically independent. Intuitively, the blocks are areas on the chip in which the defect clusters are enclosed. The size of the defect blocks is not known in advance, but can be estimated given a defect map.[3] Clearly, if the blocks are very small or very large relatively to the chip size, all three floorplans will have the same yield. Similarly, if the blocks are horizontal, the projected yield is the same for all floorplans. If, on the other hand, the blocks are vertical, the different floorplans will have significantly different yields, as shown below.

Suppose the chip consists of three vertical blocks of size three modules each, with an average of λ faults per module and a clustering parameter of α. For each of the three floorplans, the yield of the chip is calculated as the probability that at least two of the K_i, at least two of the L_i, and at least two of the M_i modules are fault-free. Using the negative binomial distribution, the probability that n modules located in the same block will be fault-free, denoted by y_n, is given by

$$y_n = [1 + n\lambda/\alpha]^{-\alpha} \tag{1}$$

Based on combinatorial considerations (i.e., the inclusion and exclusion formula), the yield of floorplan (a), $Y(a)$, is equal to[6]

$$Y(a) = (3y_2 - 2y_3)^3 \tag{2}$$

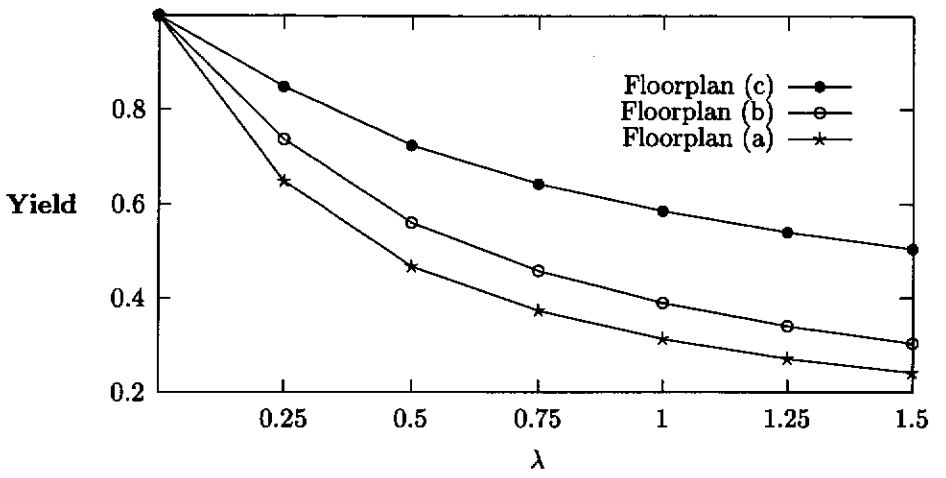

Figure 2. The yield of the three floorplans.

Again, based on combinatorics, the yield of floorplan (b), $Y(b)$, is

$$Y(b) = (2y_1 - 4y_2 + 2y_3)(2y_2 - y_3)y_3 + (2y_1 - 4y_2 + 2y_3)(y_2 - y_3)y_3$$
$$+(2y_2 - 2y_3)(2y_2 - y_3)^2 + (2y_1 - 4y_2 + 2y_3)(y_2 - y_3)y_3$$
$$+(2y_2 - y_3)(2y_1 - y_2)y_3 + (2y_1 - 3y_2 + y_3)y_2y_3$$

and after algebraic simplification

$$Y(b) = 18y_1y_2y_3 - 12y_1y_3^2 - 42y_2^2y_3 + 36y_2y_3^2 + 9y_2^3 - 8y_3^3 \tag{3}$$

Similarly, for the yield of floorplan (c), $Y(c)$, we obtain

$$Y(c) = y_3[y_3 + (3y_2 - 3y_3)y_1 + (3y_1 - 6y_2 + 3y_3)y_2 + (1 - 3y_1 + 3y_2 - y_3)y_3]$$
$$+(3y_2 - 3y_3)[y_1y_3 + (2y_2 - 2y_3)y_2 + (y_1 - 2y_2 + y_3)y_3]$$
$$+(y_1 - 2y_2 + y_3)[y_2y_3 + (y_2 - y_3)y_3] + (1 - 3y_1 + 3y_2 - y_3)y_3^2$$

which can be simplified to

$$Y(c) = 18y_1y_2y_3 - 18y_1y_3^2 - 36y_2^2y_3 + 36y_2y_3^2 + 6y_2^3 + 3y_3^2 - 8y_3^3 \tag{4}$$

It can be easily shown that for any values of λ and α, $Y(a) \leq Y(b) \leq Y(c)$. To demonstrate this point, we calculated the yield of the three floorplans, using the negative binomial model with $\alpha = 0.25$, several values of λ, and vertical defect clusters. The results are depicted in Figure 2, and the advantage of floorplan (c) is evident.

An intuitive justification for floorplan (c) having the highest yield is that it guarantees the separation between the primary modules and their spares for almost any size and shape of the defect clusters, whether they are horizontal or vertical. This results in a higher yield, since it is less likely that the same cluster will hit both the module and its spare, killing the chip. Thus, if maximizing the yield is the main consideration, floorplan (c) should be preferred over (a) and (b).

However, the three floorplans have different wiring lengths, and this criterion should be taken into account as well when selecting a floorplan. To calculate the total wiring length for a given floorplan we introduce the following notation. Denote a floorplan by a permutation $(n_1, ..., n_9)$ of the integers $(1, ..., 9)$. Denote by D_{ij} the distance

between modules i and j, and by N_{ij} the number of nets connecting modules i and j $(i,j = 1,...,9)$. Then W, the wiring cost for a given floorplan, is

$$W = \sum_{i=1}^{9} \sum_{j=1}^{9} D_{ij} N_{n_i n_j} \tag{5}$$

For D_{ij} we use the rectilinear (Manhattan) distance between the centers of the two modules i and j, i.e., $D_{ij} = (|x_i - x_j| + |y_i - y_j|)$ where (x_i, y_i) and (x_j, y_j) are the coordinates of the centers of i and j, respectively. Thus, for the $3x3$ chip,

$$D = \begin{pmatrix} 0 & 1 & 2 & 1 & 2 & 3 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 1 & 2 & 3 & 2 & 3 \\ 2 & 1 & 0 & 3 & 2 & 1 & 4 & 3 & 2 \\ 1 & 2 & 3 & 0 & 1 & 2 & 1 & 2 & 3 \\ 2 & 1 & 2 & 1 & 0 & 1 & 2 & 1 & 2 \\ 3 & 2 & 1 & 2 & 1 & 0 & 3 & 2 & 1 \\ 2 & 3 & 4 & 1 & 2 & 3 & 0 & 1 & 2 \\ 3 & 2 & 3 & 2 & 1 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 3 & 2 & 1 & 2 & 1 & 0 \end{pmatrix}$$

Suppose the matrix N is given by

$$N^{(1)} = \begin{pmatrix} 0 & 16 & 16 & 32 & 16 & 16 & 32 & 16 & 16 \\ 16 & 0 & 16 & 16 & 32 & 16 & 16 & 32 & 16 \\ 16 & 16 & 0 & 16 & 16 & 32 & 16 & 16 & 32 \\ 32 & 16 & 16 & 0 & 16 & 16 & 32 & 16 & 16 \\ 16 & 32 & 16 & 16 & 0 & 16 & 16 & 32 & 16 \\ 16 & 16 & 32 & 16 & 16 & 0 & 16 & 16 & 32 \\ 32 & 16 & 16 & 32 & 16 & 16 & 0 & 16 & 16 \\ 16 & 32 & 16 & 16 & 32 & 16 & 16 & 0 & 16 \\ 16 & 16 & 32 & 16 & 16 & 32 & 16 & 16 & 0 \end{pmatrix}$$

Then, based on Equation 5,

$$W(a) = 1344, \quad W(b) = 1472, \quad W(c) = 1536$$

Floorplan (a) has the lowest wiring cost, (b) is second, and (c) is the worst with regard to this criterion. Figure 3 depicts the three combinations of the yield and wiring cost, and there is clearly a conflict between the two. One of the points can be selected according to the relative importance of the two criteria.

If, on the other hand, the matrix N is given by

$$N^{(2)} = \begin{pmatrix} 0 & 32 & 32 & 16 & 32 & 32 & 16 & 32 & 32 \\ 32 & 0 & 32 & 32 & 16 & 32 & 32 & 16 & 32 \\ 32 & 32 & 0 & 32 & 32 & 16 & 32 & 32 & 16 \\ 16 & 32 & 32 & 0 & 32 & 32 & 16 & 32 & 32 \\ 32 & 16 & 32 & 32 & 0 & 32 & 32 & 16 & 32 \\ 32 & 32 & 16 & 32 & 32 & 0 & 32 & 32 & 16 \\ 16 & 32 & 32 & 16 & 32 & 32 & 0 & 32 & 32 \\ 32 & 16 & 32 & 32 & 16 & 32 & 32 & 0 & 32 \\ 32 & 32 & 16 & 32 & 32 & 16 & 32 & 32 & 0 \end{pmatrix}$$

Then,

$$W(a) = 2112, \quad W(b) = 1984, \quad W(c) = 1920$$

As can be seen in Figure 3, both criteria agree and floorplan (c) is the one to be selected.

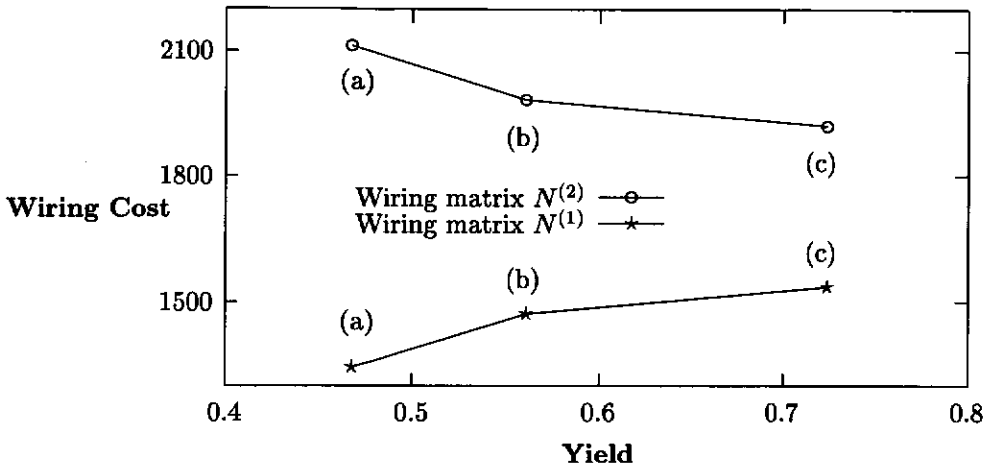

Figure 3. Wiring cost vs. yield for the three floorplans.

3. ANALYSIS OF A MEMORY IC

A new approach for incorporating defect-tolerance into memory ICs has been proposed and implemented at Samsung.[11] This is a hybrid design which combines the traditional row and column redundancy with several redundant sub-arrays. The main reason for this hybrid design is that the higher density of the new sub-micron memory ICs drastically increases the yield loss due to chip-kill defects, e.g., defects in core circuits like sense amplifiers and line drivers.[8,10,11] The conventional technique using spare rows and columns is incapable of dealing with such defects, and the entire sub-array must be replaced. The purpose of the redundant sub-arrays in the Samsung memory IC is to replace those primary memory sub-arrays hit by chip-kill faults.

The designed chip is a 1 Gbit memory which includes eight mats of size 128 Mbit each and eight redundant blocks of size 1 Mbit each (see Figure 4). Each redundant block consists of four basic 256 Kbit arrays and has additional eight spare rows and four spare columns (see Figure 5). The purpose of the spare rows and columns is to increase the probability that the redundant block is operational and can be used for replacing a block with chip-kill faults.

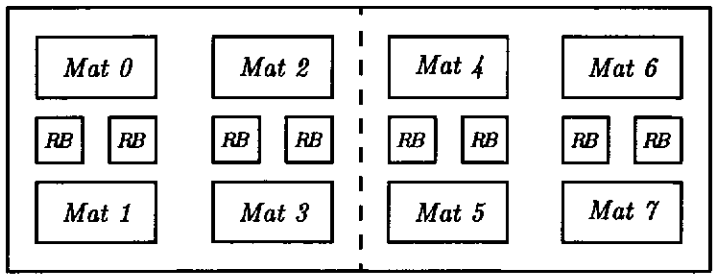

Figure 4. A 1 Gbit chip with eight mats of size 128Mbit each and eight redundant blocks (*RB*) of size 1Mbit each.

Each mat consists of 512 basic arrays of size 256 Kbit and has 32 spare rows and 32 spare columns. However, these spare rows and columns cannot be used to replace every defective row or column in the entire mat. Four spare rows are allocated to a 16 Mbit portion of the mat and eight spare columns are allocated to a 32 Mbit portion of the mat.

The yield of this new design of a memory chip was analyzed in[5] and compared to that of the traditional design with only row and column redundancy. This analysis demonstrated the benefits of some amount of block redundancy.

Figure 5. A redundant block including four 256 Kbit arrays, eight redundant rows and four redundant columns.

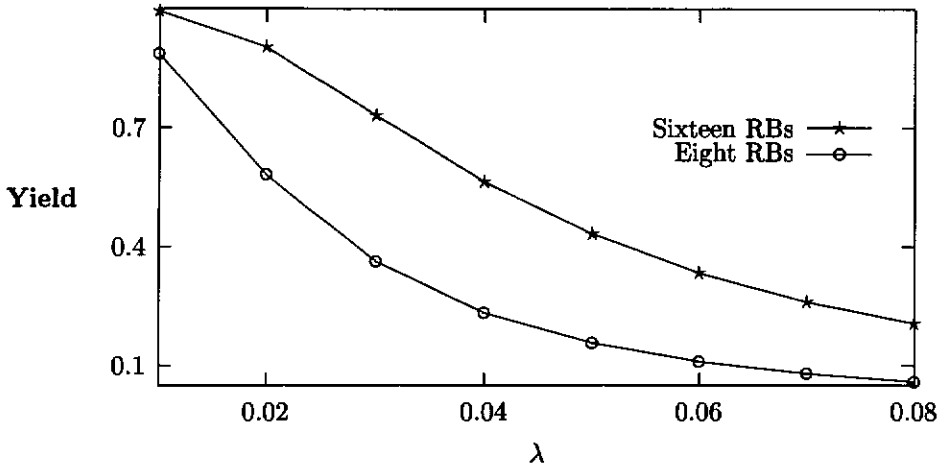

Figure 6. The yield of two redundancy schemes.

The increase in the yield is much higher than the 2% area increase required for the redundant blocks. Further analysis in[5] has shown that column redundancy is still beneficial even when redundant blocks are incorporated, and that the optimal number of such redundant columns is independent of the number of spare blocks.

The analysis in this paper concentrates on chip-kill faults, and as such, deals only with the placement of the redundant blocks. In the current design, the chip is divided into two halves, each with four mats and four redundant blocks (RBs), where a RB can replace any faulty block in the same half-chip. In an alternative floorplan, all four RBs in one half of the chip serve as spares for the mats in the second half. This is expected to increase the yield, since the mat and its spares will not be hit by the same cluster of defects, but the wiring length for this floorplan is clearly higher. Another possible solution which will increase both the yield and the wiring length by a smaller amount is to have two out of the four RBs as spares for one half-chip and the other two as spares for the other half. Other floorplans which can be considered are for a design with sixteen RBs. In this case, the alternatives are to identify 0, 2, 4 or 8 as spares for the same half in which they are located, and the rest as spares for the mats in the second half.

To compare the different floorplans, we used as yield model the negative binomial model, with an average of λ faults per block (1 Mbit), a clustering parameter α, and defect clusters with approximately the size of one eighth of a chip (128 Mbits). We first compared the yield, for several values of λ and $\alpha = 0.25$, of the designs with eight RBs and with sixteen RBs. The results appear in Figure 6, and as expected, the yield of the chip with sixteen spares is significantly higher. We then compared the yield of the three different floorplans with eight RBs, and the five different floorplans with sixteen RBs. The percentages of yield improvement, compared to the original design, are depicted in Figure 7. As this figure shows, the more spares are separated from the modules that they are replacing, the higher the yield. To determine whether the increase in yield is worth the increase in wiring cost, we first estimated the length of the wires connecting the redundant blocks to the mats for which they act as spares. Figure 8 shows the percentage of yield increase versus the cost of wiring the redundant blocks (which is only a small percentage of the total wiring cost for the entire chip). This figure shows by how much the yield can increase if the designer is willing to allow the wiring cost to increase slightly.

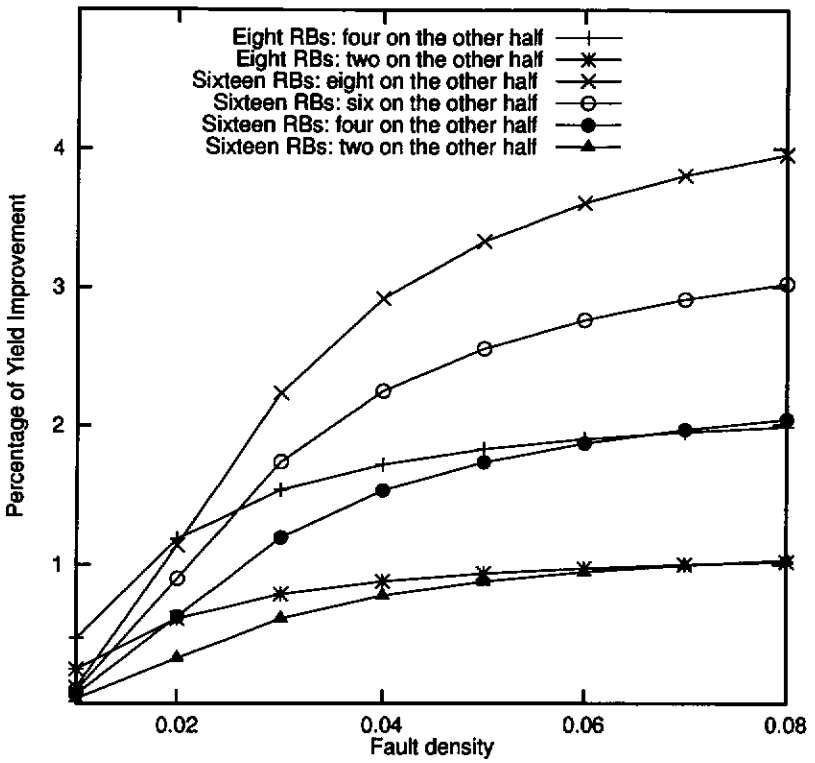

Figure 7. Percentage of yield improvement.

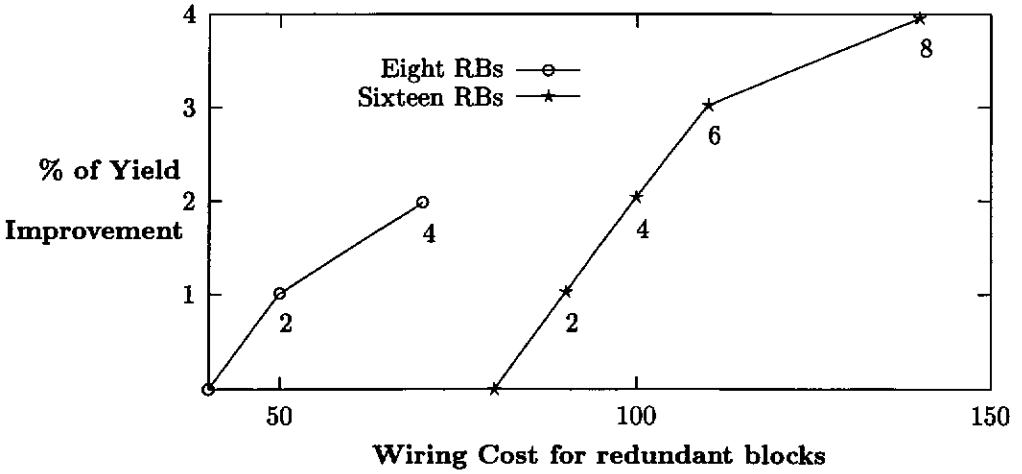

Figure 8. Yield improvement vs. wiring cost for the different floorplans (the number of RBs on the other half of the chip is marked).

4. CONCLUSION

We have analyzed in this paper two distinct objectives in floorplanning, namely, total wiring length minimization and yield maximization. We focused on designs which include redundant modules. For such designs, separating the spares from their primary modules is beneficial for yield enhancement, but may have an adverse effect on the wiring cost. Since in some cases such a separation may reduce the wiring costs, a detailed analysis should be performed for each design to determine the best floorplan for both criteria.

We demonstrated through an example of a memory IC how a tradeoff between the often conflicting objectives can be performed. We showed that an improvement in yield is possible with a limited increase in the wiring length.

ACKNOWLEDGMENTS

This work was supported in part by JPL, under contract 961294, and by NSF, under contract MIP-9710130.

REFERENCES

1. H.B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI,* Addison-Wesley, Reading, Massachusetts, 1990.
2. T. Lengauer, *Combinatorial Algorithms for Integrated Circuit Layout,* John Wiley & Sons, West Essex, England, 1990.
3. I. Koren, Z. Koren and C.H. Stapper, "A Unified Negative Binomial Distribution for Yield Analysis of Defect Tolerant Circuits," *IEEE Trans. on Computers*, vol. 42, pp. 724-437, June 1993.
4. Z. Koren and I. Koren, "On the Effect of Floorplanning on the Yield of Large Area Integrated Circuits," *IEEE Trans. on VLSI Systems*, Vol. 5, pp. 3-14, March 1997.
5. I. Koren and Z. Koren, "Analysis of a Hybrid Defect-Tolerance Scheme for High-Density Memory ICs," *Proc. of the 1997 IEEE Intern. Symp. on Defect and Fault Tolerance in VLSI Systems*, pp. 166-174, Oct. 1997.
6. I. Koren and Z. Koren, "Defect Tolerant VLSI Circuits: Techniques and Yield Analysis," *Proceedings of the IEEE*, Vol. 86, pp. 1817-1836, Sept. 1998.
7. I. Koren and Z. Koren, "Yield and Routing Objectives in Floorplanning," in *Proc. of the 1998 IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, pp. 28-36, November 1998.
8. T. Sugibayashi *et al.*, "A 1-Gb DRAM for File Applications," *IEEE J. of Solid-State Circuits*, vol. 30, pp. 1277-1280, Nov. 1995.
9. S. Wimer and I. Koren, "Analysis of Strategies for Constructive General Block Placement," *IEEE Trans. on Computer-Aided Design*, Vol. 7, pp. 371-377, March 1988.
10. T. Yamagata *et al.*, "A Distributed Globally Replaceable Redundancy Scheme for Sub-Half-micron ULSI Memories and Beyond," *IEEE J. of Solid-State Circuits*, vol. 31, pp. 195-201, Feb. 1996.
11. J-H. Yoo *et al.*, "A 32-Bank 1Gb Self-Strobing Synchronous DRAM with 1GB/s Bandwidth," *IEEE J. of Solid-State Circuits*, vol. 31, pp. 1635-1643, Nov. 1996.