On the Effect of Floorplanning on the Yield of Large Area Integrated Circuits

Zahava Koren and Israel Koren, *Fellow, IEEE*

***Abstract*— Until recently, VLSI designers rarely considered yield issues when selecting a floorplan for a newly designed chip. This paper demonstrates that for large area VLSI chips, especially those that incorporate some fault tolerance, changes in the floorplan can affect the projected yield. We study several general floorplan structures, make some specific recommendations, and apply them to actual VLSI chips. We conclude that the floorplan of a chip can affect its projected yield in a nonnegligible way, for chips with or without fault-tolerance.**

***Index Terms*—Clustering, defects, fault-tolerant IC's, floorplan, large-area IC's, yield.**

I. Introduction

IN the process of designing a new chip, yield issues are rarely a factor in the choice of the floorplan. This is justified when the chip is relatively small and the defect distribution can be accurately described by either the Poisson or the compound Poisson yield models ([3]). In particular, in the most commonly used compound Poisson model, i.e., the negative binomial (NB) distribution with large-area clustering [4], the "size" of the defect clusters is assumed to be much larger than the size of the chip and thus selecting a different floorplan will not affect the projected yield of the designed chip.

This situation is now changing with the introduction of integrated circuits with a total area of 2 cm^2 and up. Recent studies of defect maps of very large area VLSI IC's [5] have shown that the large-area clustering NB distribution does not provide a sufficiently accurate yield model for such IC's. The newly proposed medium-area clustering model [6] provides a much better match to empirical data [5]. Our objective is to study the possible impact that the floorplan of a large area chip (with or without redundancy) has on its yield, using the new medium-area clustering NB yield model.

In this paper, we report on a detailed study of the relationship between floorplanning and yield. Partial preliminary results of this study were reported in [1] and [2]. In Section II, we describe the yield model used in the analysis. In Section III, we present several general chip layouts and make some theoretical recommendations regarding their optimal floorplan. These recommendations are then illustrated in Section IV through actual test cases, namely DEC's ECL microprocessor [7], Matsushita's ADENART microprocessor [8], Hitachi's SLSI chip [9], DEC's Alpha chip [10,11] and Hughes Research Laboratories' three-dimensional (3-D) computer [13]. In all of these cases we show that changing the floorplan results in a different chip yield. Final conclusions are presented in Section V.

Manuscript received September 1, 1995. This work was supported in part by the NSF under Contract MIP-9305912. This paper was presented in part at the 1993 IEEE Workshop on Defect and Fault Tolerance in VLSI Systems, Venice, Italy [1], and the 1995 IEEE Conference on Wafer Scale Integration, San Francisco, CA [2].

Z. Koren is with the Department of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA 01003 USA.

I. Koren is with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA.

Publisher Item Identifier S 1063-8210(97)00743-9.

II. The Yield Model

We distinguish in our analysis between manufacturing defects and logical faults. Defects are the result of unwanted chemical and airborne particles deposited during the manufacturing process while faults are actual circuit failures such as line breaks and short circuits. Only a fraction of the defects cause circuit faults, with the precise number depending on the layout and density of the circuit.

We first show the statistical model used to describe the defect distribution, and then explain how it can be applied to the fault distribution, which eventually determines the yield.

It is well known that manufacturing defects tend to cluster on the wafer, and are, therefore, better matched by a Negative Binomial (NB) distribution than by a Poisson distribution [4].

The negative binomial distribution has two parameters, λ and α. When it is used to model the spatial distribution of defects in an area of size A, both of the parameters depend on A, and

$$\text{Prob}(x \text{ defects in area } A) = \frac{\Gamma(\alpha_A + x)}{x!\Gamma(\alpha_A)} \frac{\left(\frac{\lambda_A}{\alpha_A}\right)^x}{\left(1 + \frac{\lambda_A}{\alpha_A}\right)^{\alpha_A + x}}. \quad (1)$$

In particular, (1) lets us calculate the probability of zero defects in area A which, if no redundancy exists, constitutes the yield of this area

$$\text{Yield} = \text{Prob}(0 \text{ defects in area } A) = \left(1 + \frac{\lambda_A}{\alpha_A}\right)^{-\alpha_A}. \quad (2)$$

The parameter λ_A denotes the expected number of defects in area A, satisfying $\lambda_A = \lambda A$, where λ is the expected number of defects per unit area. α_A is the clustering parameter for the considered area, measuring the deviation from the Poisson distribution. The smaller its value, the larger the deviation. α_A is a nondecreasing function of A, but the exact dependence of α_A on A is not uniquely defined and depends on the clustering

pattern within the area. Most of the literature dealing with yield issues assumes large area clustering, i.e., large defect clusters comparable in size to the chip or even the wafer size. As has been demonstrated in [5], the empirical defect distribution of large area chips has a better fit to a medium-area clustering than to a large-area clustering NB distribution. A detailed description of the medium-area NB distribution, including several suggested ways of estimating the block size, appears in [6]. For the sake of completeness, its underlying principles are briefly outlined below.

Under the medium-area NB model, we view the defect clustering as an empirical phenomenon which is the result of the wafer area being divided into subareas which we call *blocks*, such that the defects in distinct blocks are statistically independent. The number of defects in each block has an NB distribution, with a uniform distribution within the area of the block. The large-area NB distribution is a special case where the whole wafer constitutes one block, resulting in very large defect clusters and at the same time large areas which are defect-free. Mathematically, the medium-area NB distribution, similarly to the large area one, can be obtained as a compound Poisson distribution as shown below.

According to the Poisson distribution, the probability of no defects in an area of size A is

$$P(\text{no defects in an area of size } A) = e^{-lA} \tag{3}$$

where l is the expected number of defects per unit area. Assuming that l is a random variable with a $\Gamma(\alpha, \frac{\alpha}{\lambda})$ density function, i.e.,

$$f(l) = \frac{\alpha^\alpha}{\lambda^\alpha \Gamma(\alpha)} l^{\alpha-1} e^{-\frac{\alpha}{\lambda} l} \tag{4}$$

and integrating (3) with respect to this density yields

$$P(\text{no defects in an area of size } A) = \int_0^\infty e^{-lA} f(l) \cdot dl = \left(1 + \frac{\lambda A}{\alpha}\right)^{-\alpha} \tag{5}$$

which corresponds to the large-area NB distribution. If, on the other hand, A is divided into two disjoint areas of sizes A_1 and A_2 which are assumed to be statistically independent with respect to the number of defects, then l should be averaged independently over each of these areas. This results in

$$\begin{aligned} &P(\text{no defects in an area of size } A) \qquad (6)\\ &= P(\text{no defects in an area of size } A_1 + A_2)\\ &= P(\text{no defects in } A_1) \cdot P(\text{ no defects in } A_2)\\ &= \int_0^\infty e^{-lA_1} f(l) \cdot dl \int_0^\infty e^{-lA_2} f(l) \cdot dl\\ &= \left(1 + \frac{\lambda A_1}{\alpha}\right)^{-\alpha} \left(1 + \frac{\lambda A_2}{\alpha}\right)^{-\alpha}. \end{aligned}$$

The expression in (6) corresponds to the medium-area NB distribution, and has a different numerical value than the expression in (5).

Since yield losses are caused by logical faults rather than by defects, we next show how the fault distribution can be obtained from the defect distribution. Let the defect distribution be the spatial Poisson distribution with an average of l defects per unit area, and let each defect have a probability p of becoming a fault. According to a well-known theorem in probability theory, the number of faults is also Poisson distributed with an average of lp faults per unit area. Assuming that l is a random variable with the density function in (4), and integrating over l similarly to (5), we find that the number of faults in area A has a negative binomial distribution with the parameters λpA and α.

If, however, the area consists of two disjoint parts of sizes A_1 and A_2, each with a different "critical area" [4] and probabilities p_1 and p_2, respectively, of a defect becoming a fault, then

$$P(\text{no faults in an area of size } A_1 + A_2) = e^{-A_1 p_1 l - A_2 p_2 l}. \tag{7}$$

Integrating (7) over l with respect to the Γ density function (4) can result in two different expressions, depending on whether the two areas, A_1 and A_2, are assumed statistically independent or not. In the first case

$$P(\text{no faults in an area of size } A_1 + A_2) = \left(1 + \frac{\lambda p_1 A_1}{\alpha}\right)^{-\alpha} \left(1 + \frac{\lambda p_2 A_2}{\alpha}\right)^{-\alpha} \tag{8}$$

while in the second case

$$P(\text{no faults in an area of size } A_1 + A_2) = \left(1 + \frac{\lambda p_1 A_1 + \lambda p_2 A_2}{\alpha}\right)^{-\alpha}. \tag{9}$$

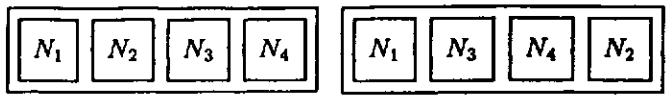
Note that although (8) and (9) are both obtained by integrating the same expression (7), their numerical values are not equal. If the second mode of integration is used, indicating large area clustering, then the resulting yield will not be affected by a change in the floorplan. If, on the other hand, the integration is done independently over disjoint areas, as dictated by the medium area clustering assumption and exemplified in (8), then different floorplans, i.e., different placements of the same logic modules, could result in different yields. This claim is demonstrated in the next two sections.

III. Theoretical Test Cases

To demonstrate the possible effect of the floorplan on the yield we consider several hypothetical chip layouts, with and without redundancy, and compare the projected yields of different possible floorplans for each.

A. Example 1—A One-Dimensional Floorplan Without Redundancy

In the first example, depicted in Fig. 1, the chip consists of four equal-area modules (functional units), N_1, N_2, N_3, and N_4. The chip has no incorporated redundancy, and all



floorplan (a) floorplan (b)

Fig. 1. Two floorplans for Example 1.

four modules are necessary for the proper operation of the chip.

We assume the medium-area NB distribution for the spatial distribution of the manufacturing defects, with parameters λ (per module) and α (per block). Suppose that N_1, N_2, N_3, and N_4 have different sensitivities to defects. A defect in N_i has a probability p_i of becoming a fault ($i = 1, 2, 3, 4$) with

$$p_1 \le p_2 \le p_3 \le p_4.$$

Denote

$$\lambda_i = \lambda p_i (i = 1, 2, 3, 4)$$

then

$$\lambda_1 \le \lambda_2 \le \lambda_3 \le \lambda_4.$$

This chip has $4! = 24$ possible floorplans, denoted by $(N_{i_1}, N_{i_2}, N_{i_3}, N_{i_4})$ where (i_1, i_2, i_3, i_4) is some permutation of $(1, 2, 3, 4)$. If small area clustering (clusters smaller than or comparable to the size of a module) or large area clustering (clusters larger than or equal to the chip area) are assumed, the projected yields of all possible floorplans will be the same. This is not the case, however, when medium area clustering (i.e., blocks of size 2 or 3 modules) is assumed.

The yield of floorplan $(N_{i_1}, N_{i_2}, N_{i_3}, N_{i_4})$, assuming blocks of size 2, is

$$\begin{aligned}&\text{Yield}\,(i_1, i_2, i_3, i_4)\\ &= \frac{1}{2}\left[\left(1+\frac{\lambda_{i_1}+\lambda_{i_2}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_3}+\lambda_{i_4}}{\alpha}\right)^{-\alpha}\right.\\ &\left.+\left(1+\frac{\lambda_{i_1}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_2}+\lambda_{i_3}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_4}}{\alpha}\right)^{-\alpha}\right]. \quad (10)\end{aligned}$$

It is clear from (10) that different permutations will result in different yields. The best permutation can be found by exchanging modules whenever this exchange increases the yield. It can be shown algebraically that exchanging i_1 and i_2 will increase the yield if and only if $\lambda_{i_1} > \lambda_{i_2}$. Similarly, exchanging i_3 and i_4 will increase the yield if and only if $\lambda_{i_4} > \lambda_{i_3}$. Exchanging i_2 and i_3 will increase the yield if and only if $\lambda_{i_1} > \lambda_{i_4}$ and $\lambda_{i_2} < \lambda_{i_3}$, or $\lambda_{i_1} < \lambda_{i_4}$ and $\lambda_{i_2} > \lambda_{i_3}$. Exchanging i_2 and i_4 will increase the yield if and only if $\lambda_{i_2} > \lambda_{i_4}$, and similarly, exchanging i_1 and i_3 will increase the yield if and only if $\lambda_{i_1} > \lambda_{i_3}$.

Taking into account all of the above inequalities, we conclude that the best permutation of the modules (under the condition $\lambda_1 \le \lambda_2 \le \lambda_3 \le \lambda_4$) is (N_1, N_3, N_4, N_2), which is shown in Fig. 1(b). The permutation (N_2, N_4, N_3, N_1) is equivalent and has the same yield.

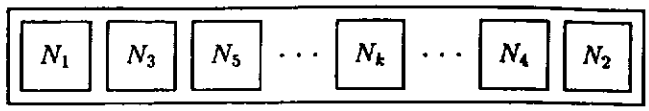
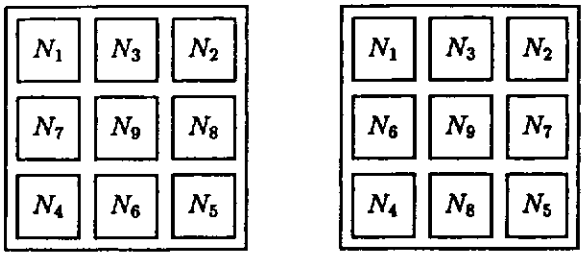


Fig. 2. A floorplan with k modules.



floorplan (a) floorplan (b)

Fig. 3. Two alternative floorplans for Example 2.

If the block size is assumed to be three modules, the projected yield for the floorplan $(N_{i_1}, N_{i_2}, N_{i_3}, N_{i_4})$ is

$$\begin{aligned}&\text{Yield}\,(i_1, i_2, i_3, i_4) =\\ &\frac{1}{3}\left[\left(1+\frac{\lambda_{i_1}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_2}+\lambda_{i_3}+\lambda_{i_4}}{\alpha}\right)^{-\alpha}\right.\\ &+\left(1+\frac{\lambda_{i_1}+\lambda_{i_2}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_3}+\lambda_{i_4}}{\alpha}\right)^{-\alpha}\\ &\left.+\left(1+\frac{\lambda_{i_1}+\lambda_{i_2}+\lambda_{i_3}}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_{i_4}}{\alpha}\right)^{-\alpha}\right]. \quad (11)\end{aligned}$$

The rules for selecting the best permutation happen to be the exact same rules as for a block of size 2. Our conclusion is that the floorplan (N_1, N_3, N_4, N_2) shown in Fig. 1(b) results in the highest yield, for any block size.

B. A Generalization of Example 1

The above conclusion can be generalized to a chip consisting of k modules $N_1, \cdots, N_k$ with different fault densities $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_k$. The possible floorplans can be represented by the permutations $(N_{i_1}, N_{i_2}, \cdots, N_{i_k})$. Similarly to the method used in Example 1, it can be shown that for any block size, the floorplan resulting in the highest yield is

$$(N_1, N_3, N_5, \cdots, N_k, \cdots, N_6, N_4, N_2)$$

as shown in Fig. 2.

C. Example 2—A Two-Dimensional Floorplan Without Redundancy

In this example, depicted in Fig. 3, the chip consists of nine equal-area modules, $N_1, \cdots, N_9$, arranged in the form of a square. All nine modules are necessary for the proper operation of the chip. We assume that the faults are distributed according

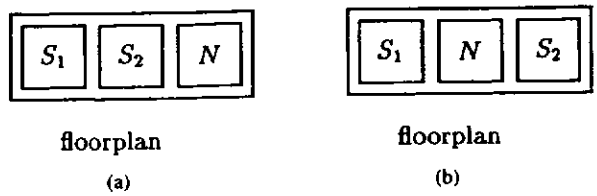


Fig. 4. Two alternative floorplans for Example 3.

to the medium-area NB distribution, with two-dimensional fault blocks. Let $\lambda_i (i = 1, \cdots, 9)$ denote the fault density of module i, and suppose that

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_9.$$

Clearly, if the block size is 1×1 (small area clustering) or 9×9 (large area clustering), all 9! possible floorplans have the same projected yield. The yields differ, however, for medium-sized fault clusters. Similar expressions to those in Example 1 show that the optimal floorplan depends on the size and shape of the fault blocks. For a 1×3 block the best floorplan is (a), while for a 2×3 block the best floorplan is (b). Although the optimal floorplan depends on the block size, some generalizations can be made. For all block sizes, the module with the highest defect density should be placed in the center of the chip, and similarly, each row or column should be rearranged so that the most sensitive module is in its center.

Based on the previous examples we draw the general conclusion that even in a floorplan with no redundancy (i.e, all faults are chip-kill faults [4]), if different modules have different fault densities then the relative position of the modules may have a significant impact on the yield. Our specific recommendation is to place the most sensitive modules in the center of the chip, the less sensitive ones in the boundaries, and the least sensitive modules at the corners. Note that we reached this conclusion without assuming that the boundaries of the chip are more prone to defects than its center. We illustrate this rule in Section IV.

In the next set of hypothetical examples, the chips have some redundancy incorporated in them for fault-tolerance.

D. Example 3—A One-Dimensional Chip with Redundancy

In this case, the chip consists of three circuits, S_1, S_2 and N. For proper operation, N and either S_1 or S_2 must be fault-free. The two possible floorplans for the chip are depicted in Fig. 4.

The difference between the two is that in floorplan (a), the module S_1 and its spare are placed next to each other, while in floorplan (b) they are separated by N. Clearly, there is no difference in the yield of the two floorplans when using a block size of one module or three modules.

Let the number of defects have a medium-area NB distribution with an average of λ defects per module, a clustering parameter of α (per block), and a block size of two modules. Suppose that a defect in S_i or N has a probability p_s or p_n, respectively, of becoming a fault, and denote: $\lambda_s = \lambda p_s$, $\lambda_n = \lambda p_n$. The yields of the two layouts will be

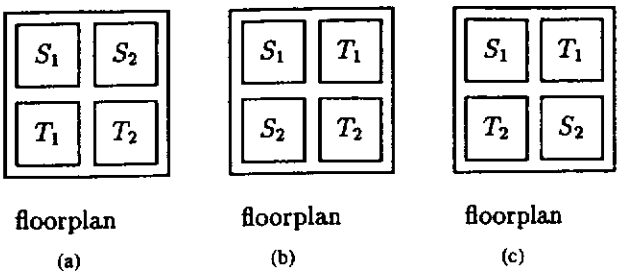


Fig. 5. Three alternative floorplans for Example 4.

$$\begin{aligned}\text{Yield}\,(a) &= \frac{1}{2} \cdot \left[\left(1+\frac{\lambda_n}{\alpha}\right)^{-\alpha} \left(2\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} - \left(1+\frac{2\lambda_s}{\alpha}\right)^{-\alpha}\right) \right. \\ &\quad + \left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} \left(1+\frac{\lambda_n}{\alpha}\right)^{-\alpha} + \left(1+\frac{\lambda_s+\lambda_n}{\alpha}\right)^{-\alpha} \\ &\quad \left. - \left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} \left(1+\frac{\lambda_s+\lambda_n}{\alpha}\right)^{-\alpha} \right] \end{aligned} \tag{12}$$

$$\begin{aligned}\text{Yield}\,(b) &= \left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} \left(1+\frac{\lambda_n}{\alpha}\right)^{-\alpha} + \left(1+\frac{\lambda_s+\lambda_n}{\alpha}\right)^{-\alpha} \\ &\quad - \left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} \left(1+\frac{\lambda_s+\lambda_n}{\alpha}\right)^{-\alpha}. \end{aligned} \tag{13}$$

It can be proven algebraically that for any λ_s and λ_n, Yield $(b) \geq$ Yield (a), with a strict inequality for $\lambda_s \neq \lambda_n$. Floorplan (b) should, therefore, be preferred over (a), with the practical implication being that the circuit and its spare should be separated rather than placed adjacent to each other.

E. Example 4—A Two-Dimensional Chip with Redundancy

The next example is that of a chip consisting of four modules, S_1, S_2, T_1, and T_2. For proper operation, one of S_1 and S_2, and one of T_1 and T_2 have to be fault-free. Three of the possible floorplans for this chip are depicted in Fig. 5. As before, let the number of defects have a medium-area NB distribution with an average of λ defects per module, a clustering parameter of α per block, and probabilities p_s and p_t of a defect becoming a fault in S_i and T_i, respectively. Denoting $\lambda_s = \lambda p_s$ and $\lambda_t = \lambda p_t$, and assuming that the chip consists of two horizontal blocks of two modules each, the yields of the three floorplans are

$$\begin{aligned}\text{Yield}\,(a) = \frac{1}{2}\left[\left(2\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} - \left(1+\frac{\lambda_s}{\alpha}\right)^{-2\alpha}\right) \right. \\ \left(2\left(1+\frac{\lambda_t}{\alpha}\right)^{-\alpha} - \left(1+\frac{\lambda_t}{\alpha}\right)^{-2\alpha}\right) \\ + \left(2\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha} - \left(1+\frac{2\lambda_s}{\alpha}\right)^{-\alpha}\right) \\ \left. \left(2\left(1+\frac{\lambda_t}{\alpha}\right)^{-\alpha} - \left(1+\frac{2\lambda_t}{\alpha}\right)^{-\alpha}\right)\right] \end{aligned} \tag{14}$$

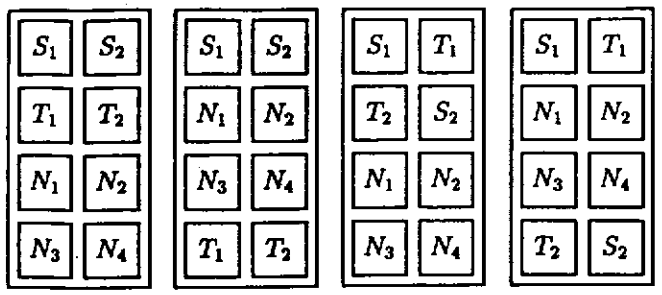


Fig. 6. Four alternative floorplans for Example 5.

and

$$\begin{aligned}\text{Yield}(b) &= \text{Yield}(c)\\ &= \frac{1}{2}\Bigg[\left(2\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha}-\left(1+\frac{\lambda_s}{\alpha}\right)^{-2\alpha}\right)\\ &\quad\left(2\left(1+\frac{\lambda_t}{\alpha}\right)^{-\alpha}-\left(1+\frac{\lambda_t}{\alpha}\right)^{-2\alpha}\right)\\ &\quad+2\left(1+\frac{\lambda_s+\lambda_t}{\alpha}\right)^{-\alpha}+2\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_t}{\alpha}\right)^{-\alpha}\\ &\quad-2\left(1+\frac{\lambda_s+\lambda_t}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_s}{\alpha}\right)^{-\alpha}\\ &\quad-2\left(1+\frac{\lambda_s+\lambda_t}{\alpha}\right)^{-\alpha}\left(1+\frac{\lambda_t}{\alpha}\right)^{-\alpha}\\ &\quad+\left(1+\frac{\lambda_s+\lambda_t}{\alpha}\right)^{-2\alpha}\Bigg].\end{aligned} \tag{15}$$

It can be easily proven that for any values of λ_s and λ_t, $\text{Yield}(b) = \text{Yield}(c) \geq \text{Yield}(a)$.

If, on the other hand, the chip consists of two vertical blocks, then clearly, $\text{Yield}(b)$ is given by (14) and $\text{Yield}(a)$ is equal to $\text{Yield}(c)$ and is given by (15). In this case, $\text{Yield}(a) = \text{Yield}(c) \geq \text{Yield}(b)$ for all values of λ_s and λ_t. Floorplan (c) should, therefore, be preferred over (a) and (b). An intuitive justification for the choice of floorplan (c) is that it guarantees the separation between the primary modules and their spares for any block size and shape. This results in a higher yield, since it is less likely that the same defect cluster will hit both the module and its spare.

F. Example 5—A Chip with Redundancy and Chip-Kill Area

The last example is that of a chip consisting of eight modules, N_1, N_2, N_3, N_4, S_1, S_2, T_1, and T_2. N_1, N_2, N_3, and N_4 have no incorporated redundancy, S_2 is a spare for S_1 and T_2 is a spare for T_1. For proper operation of the chip, either S_1 or S_2 and either T_1 or T_2 must be fault-free, and N_1, N_2, N_3 and N_4 all have to be fault-free. Four of the possible floorplans are depicted in Fig. 6.

Similarly to the previous examples, let λ_n, λ_s and λ_t denote the average number of faults in N_i, S_i, and T_i, respectively. Clearly, if small area clustering or large area clustering are assumed, the projected yields of all floorplans will be the same. For medium area clustering, on the other hand, the yield changes with the floorplan. Unfortunately, the optimal floorplan depends on the shape of the defect clusters. For vertical blocks (block sizes of 2×1, 3×1 or 4×1) floorplan (a) is optimal, while for horizontal blocks (1×2, 2×2 or 3×2) floorplan (d) is optimal (for any values of λ_n, λ_s and λ_t). Our recommendation, therefore, is to separate a module and its spare but to place together the spares of the different modules. This way, if a defect cluster hits all spares, the chip can still survive. These recommendations are illustrated in the next section.

IV. Practical Test Cases

We illustrate the general principles stated above through five practical test cases. Two of the chips, DEC's ECL microprocessor [7] and Matsushita's ADENART microprocessor [8] have no redundancy, while the other three, Hitachi's SLSI chip [9], DEC's Alpha chip [10,11] and Hughes Research Laboratories' 3-D computer [13], have some incorporated redundancy.

A. The ECL RISC Microprocessor

One of Digital Equipment Corporation's most recent IC's is a 300 MHz, 1.0 μm bipolar ECL RISC microprocessor. The 15.4 mm x 12.6 mm chip contains 468K bipolar transistors and implements a subset of MIPS R6000 architecture. It was designed in order to verify a new packaging technique, a new style of CAD tools and ECL circuit techniques. A simplified diagram of the chip's floorplan is shown in Fig. 7.

Fig. 7 shows that each of the two 2-kB cache units, the instruction cache and the data cache, has been separated into two halves, one at the top and the second at the bottom of the floorplan. Due to their small size, the two cache units do not include spare rows or columns for yield enhancement but do include byte parity bits. The middle section of the chip, occupying almost half of its total area, contains the processor's logic units, i.e., a register file, an integer execution unit, instruction decode, pipeline control, etc.

The density of bipolar transistors and resistors in the two cache units is more than double the corresponding density in the remaining logic units [7]. The two cache units are, therefore, expected to have a higher fault density than the logic units, and we used the ratio $\lambda_{\text{cache}}/\lambda_{\text{logic}} = 2.5$ in our analysis. Consequently, we might consider alternative floorplans in an attempt to improve the projected yield, following the principles outlined in Section III. Using the notation introduced in Section III-A, we partition the floorplan of DEC's ECL RISC processor into four, almost equal-sized, modules and denote them by N_3, N_1, N_2 and N_4 (as shown in Fig. 7) so that the relation $\lambda_1 = \lambda_2 < \lambda_3 = \lambda_4$ is satisfied with $\lambda_1 = \lambda_2 = \lambda_{\text{logic}}$ and $\lambda_3 = \lambda_4 = \lambda_{\text{cache}}$. The "optimal" floorplan according to the analysis in Section III-A would be (N_1, N_3, N_4, N_2). This however, implies that the processor will be divided into two almost equally sized modules and placed on both sides of the caches. If we insist on keeping all the logic functional units

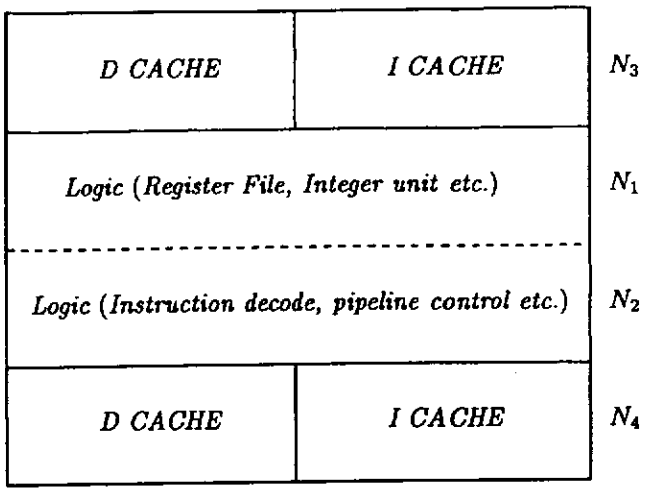


Fig. 7. The original floorplan (simplified) of DEC's ECL RISC microprocessor.

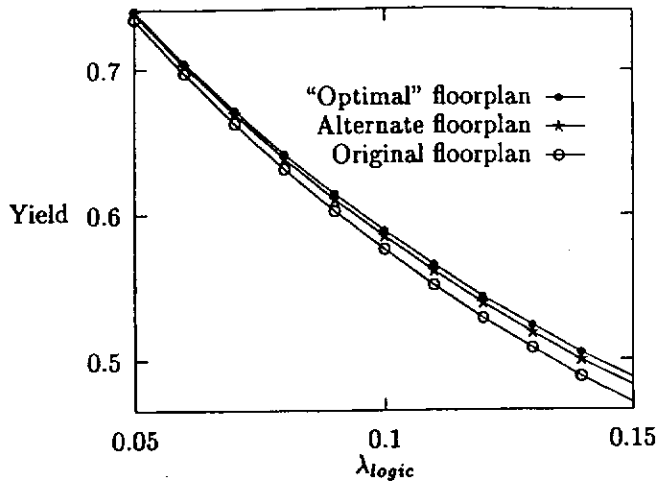


Fig. 8. The yield of the original, alternate and "optimal" floorplans of DEC's ECL RISC microprocessor as a function of λ_{logic} ($\alpha = 0.5$, $\lambda_{\text{cache}}/\lambda_{\text{logic}} = 2.5$).

adjacent to each other, then the floorplan (N_1, N_2, N_3, N_4) (i.e., with the two halves of the cache units placed next to each other) has been found to have a higher projected yield than that of the original floorplan. The optimal floorplan can, in principle, be implemented, but this might involve an extra routing penalty. There clearly is a tradeoff between yield improvement and increased routing overhead, a situation that we must expect to encounter in the general case.

Fig. 8 shows the projected yield of the original floorplan, the alternative floorplan where the two halves of the cache units are placed next to each other, and the "optimal" floorplan. The yield has been calculated using the medium-area NB distribution with a block size of two modules (although similar results were obtained for a block size of three modules). The alternative floorplan has a higher projected yield than the original one. The "optimal" floorplan has the highest projected yield but it is not clear whether the relatively small marginal improvement in yield (compared to the alternative floorplan) justifies the additional routing penalty.

B. Matsushita's ADENART Microprocessor

An 80 MFLOP's 64 b microprocessor has been developed by the Matsushita Company to serve as the basic processing element in their ADENART parallel computer with a target peak performance of 20 GFLOP's [8]. The microprocessor has a RISC superscalar architecture and contains an 8 Kbytes data cache (DCU) and a 2 Kbytes instruction cache (ICU). It has been implemented in a 0.8 μm CMOS technology and it contains 1300K transistors in a total area of 14.7×15.3 mm^2. A simplified diagram of the chip's floorplan is depicted in Fig. 9(a). The microprocessor includes five independent execution units, two register files (FR—floating-point registers and PR—pointer registers), an instruction decode and pipeline control unit (IDU), a data bus control unit (DBC) and a ROM which contains a look-up table for divisors' reciprocals. The five execution units include a floating-point add and subtract unit (FAU), a floating-point multiply and divide unit (FMU), a load address add unit (LDU), a pointer arithmetic and logic unit (PNU) and a flow control unit (FCU). The 12 blocks have six different transistor densities with the ROM having the highest density and the FCU and IDU, the lowest density. Assuming that the fault densities are linearly proportional to the transistor densities we define six fault densities which satisfy

$$\lambda_{lg\_1} < \lambda_{lg\_2} < \lambda_{lg\_3} < \lambda_{lg\_4} < \lambda_{lg\_5} < \lambda_{\text{cache}} < \lambda_{\text{rom}}.$$

These fault densities are assigned to the individual blocks as shown in Fig. 9. Based on the transistor densities reported in [8] we used in our analysis the ratio

$$\lambda_{\text{rom}} : \lambda_{\text{cache}} : \lambda_{lg\_5} : \lambda_{lg\_4} : \lambda_{lg\_3} : \lambda_{lg\_2} : \lambda_{lg\_1} = 8.88 : 7.69 : 3.27 : 2.42 : 2.27 : 1.69 : 1.$$

Clearly, this original plan does not follow the guidelines stated in Section III-C and is therefore not optimal with regard to yield. To demonstrate the effect of a different floorplan on the yield of the microprocessor, we selected two other floorplans. Floorplan (b), in which the modules with the higher fault density are moved to the boundaries and has a lower yield than the original, and floorplan (c) which follows the guidelines and is expected to have a higher yield than the original. All three floorplans are depicted in Fig. 9.

For yield calculation, we divided the chip area into a grid of 8×8, and calculated the projected yield for the three floorplans, using the medium-area NB distribution, for each of the block sizes between 1×1 and 8×8. As expected, we obtained

$$\text{Yield}\,(b) < \text{Yield}\,(a) < \text{Yield}\,(c)$$

for all block sizes, except small-area and large-area clustering (for which all floorplans have the same yield). Fig. 10 compares the yield of the three floorplans as a function of λ_{lg_1}, for two block sizes, namely 4×4 and 6×6. Fig. 10 demonstrates that the difference between the worst floorplan and the best floorplan can be significant.

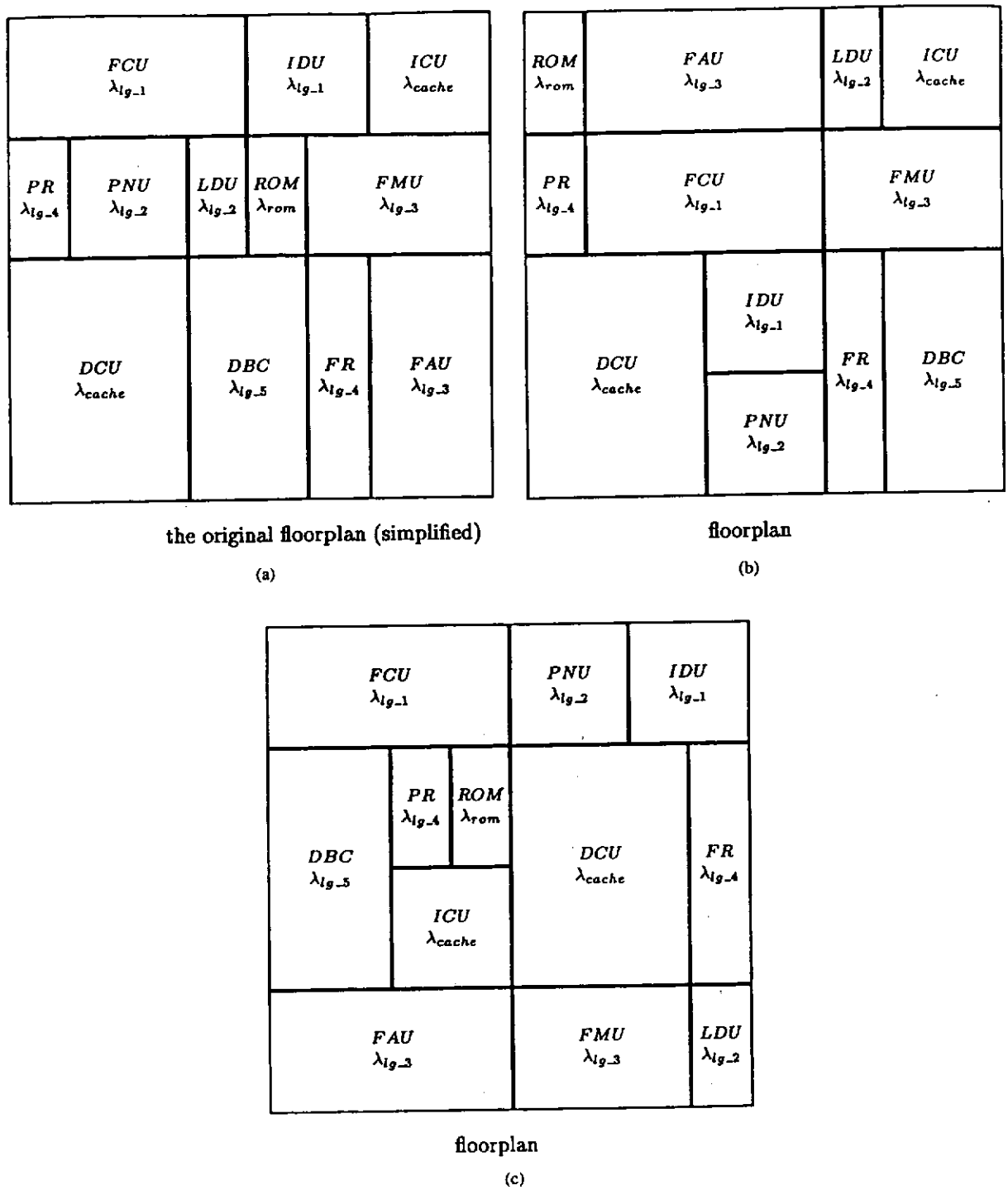


Fig. 9. The original and two alternative floorplans for the ADENART chip.

C. Hitachi's SLSI Chip

The floorplan of Hitachi's SLSI chip is depicted in Fig. 11. The chip size is $38.16 \times 50.4 = 1923.26$ mm^2 and could not fit the conventional reticle, which allows a maximum chip size of about 200 mm^2. Therefore, to manufacture the SLSI chip, the wafer underwent four separate fabrication steps in which the gate array, the SRAM's, the DRAM's and the interconnections were patterned. The 11 4Mb DRAM's which consume most of the chip area use a 0.8 μm process and incorporate redundancy for fault-tolerance. The remaining units in the chip use a 1.3 μm relaxed process and have no redundancy.

In what follows we will concentrate on the DRAM's, the only units in the chip that employ some fault-tolerance technique. Each DRAM is internally organized as a 1M word $\times$ 4 b memory. Thus, the DRAM consists of four identical parts, and each one of these four parts, in turn, is divided into eight sections which are called "mats." This is done to reduce resistance and stray capacitance on the bit lines. We estimate the DRAM size to be 13 mm $\times$ 5.5 mm and assume that the

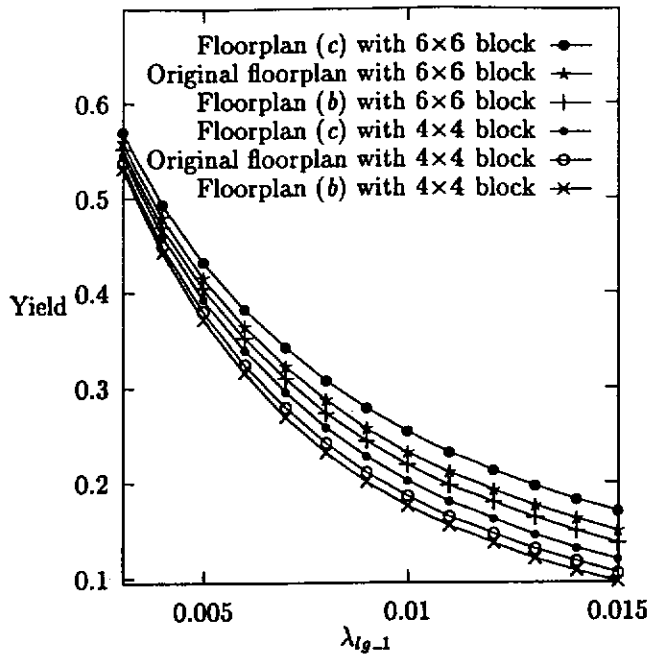


Fig. 10. The yield of the original and two alternate floorplans of Matsushita's ADENART microprocessor as a function of λ_{lg_1} ($\alpha = 0.5$).

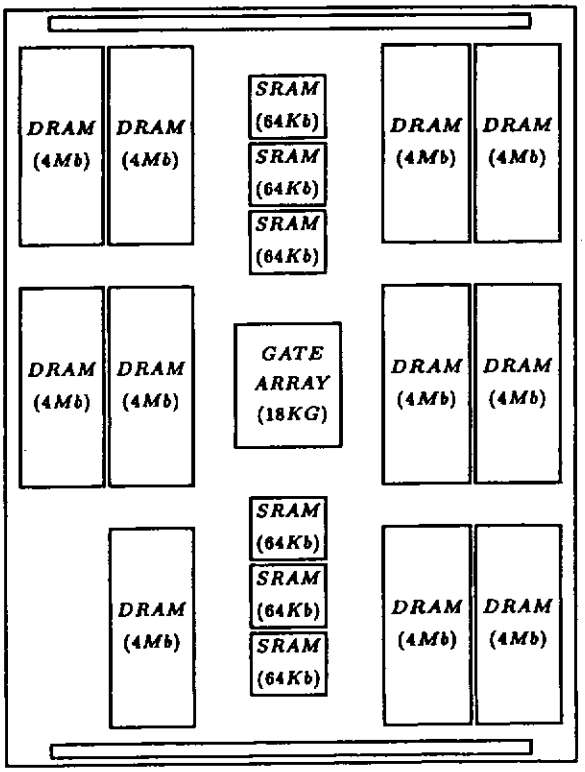


Fig. 11. The floorplan of Hitachi's SLSI chip is depicted in Fig. 11. The floorplan of the SLSI chip.

internal organization of each mat is 128 words × 1 Kb, i.e., there are 1 Kb lines in every mat with 128 memory cells per bit line. Such an internal organization is typical of most 4 Mb DRAM chips.

The DRAM employs two fault-tolerance techniques: adding spare lines, and using only six out of the 11 fabricated

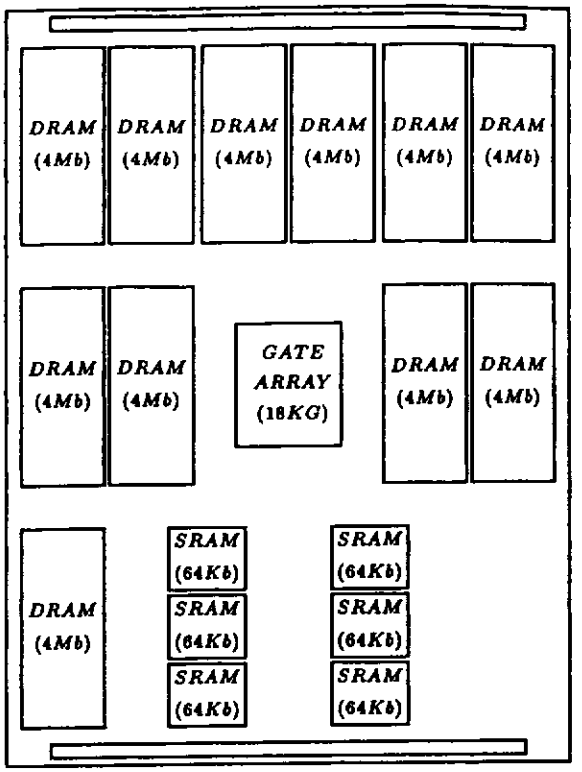


Fig. 12. An alternative floorplan for the SLSI chip.

DRAM's. The first technique is somewhat different from the traditional technique of adding spare lines. Instead of adding spare lines to every mat, all eight mats (in a quarter of the 4 Mb DRAM) share a set of redundant word lines that can be used to replace defective lines in each one of the eight mats. The traditional technique of adding spare lines to each mat separately would require a large number of spare lines since defects tend to cluster. The alternate technique of providing spare lines that are common to all eight mats requires fewer spares. Since each DRAM contains 32 mats, the requirement of having six operational DRAM's means that out of $32 \times 11 = 352$ mats, $32 \times 6 = 192$ acceptable mats are needed. Some of these 192 mats will be defect-free, while others will have a few defective lines which are replaced by spare lines.

In the original floorplan depicted in Fig. 11, the 18K gate array is positioned at the center of the chip. This equalizes the length of the communication links between the gate array and the SRAM's and DRAM's to eliminate timing mismatches. An alternative floorplan that will still keep the communication link equalized is shown in Fig. 12.

For yield calculation purposes, we divided the chip area into 36 modules, enabling the choice of block sizes between 1×1 and 6×6. As can be expected, there was no significant difference in yield between the two floorplans when either very small or very large block sizes have been selected. There were some noticeable differences, however, for medium size blocks, as can be seen in Fig. 13. Let λ_1, λ_2 denote the average number of faults per unit area in the DRAM's and in the remaining area, respectively, with $\lambda_2 < \lambda_1$. Fig. 13 depicts the yield of both the original and the alternate floorplans as

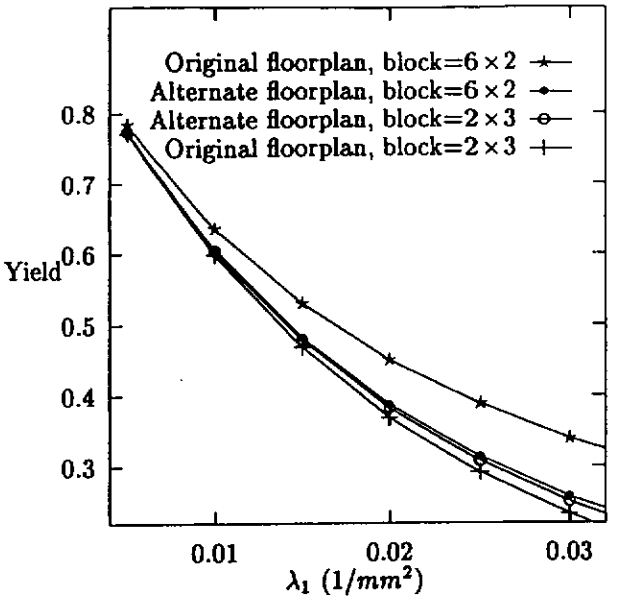


Fig. 13. The yield of the original and alternate floorplans of the SLSI chip as a function of λ_1 for two values of the block size (with $\lambda_2 = 0.3\lambda_1$ and $\alpha = 1.5$).

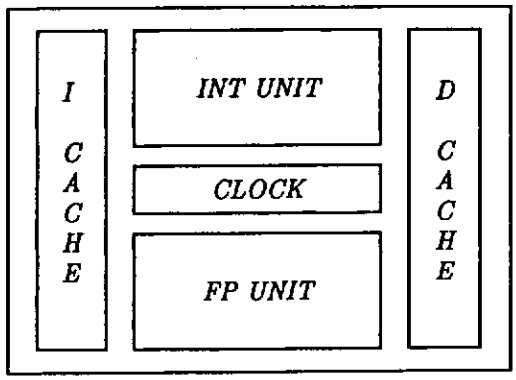


Fig. 14. The floorplan of the Alpha chip.

a function of λ_1 for $\lambda_2/\lambda_1 = 0.3$, $\alpha = 1.5$ and two block sizes, namely 6×2 and 2×3. As can be seen in Fig. 13, for vertical blocks of size 6×2 the yield of the original floorplan is much higher than that of the alternate, while for horizontal blocks of size 2×3, the alternate floorplan has a slightly higher yield than the original one. This phenomenon can be explained by the results of Section III. For vertical blocks, more of the DRAM's are placed in the same block in the original floorplan, while the opposite is true for horizontal blocks.

D. DEC's Alpha Chip

The floorplan of the Alpha chip is depicted in Fig. 14. The chip consists of the following functional units: the integer unit, the floating-point unit, the clock circuitry and two cache units. The two internal cache units are for data only (D-CACHE) or instructions only (I-CACHE). These two cache units have almost identical implementations. Each has 8 kilobytes (of data or instruction) organized as an array of 1024 cells wide by 66 cells tall. The top two rows (out of the 66 rows)

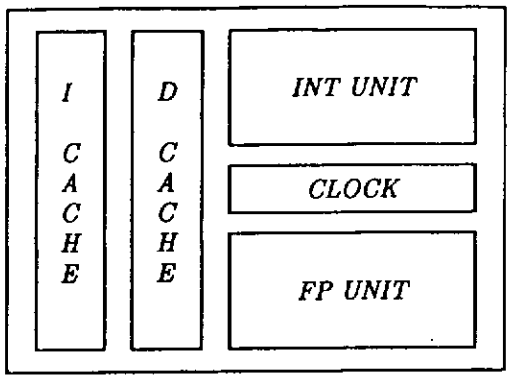


Fig. 15. An alternate floorplan for the Alpha chip.

constitute the redundant cells to replace either defective rows or defective individual cells. The remaining functional units have no redundancy incorporated into them and as such, each fault occurring in them is a chip-kill fault [4].

The current floorplan as shown in Fig. 14 has the D-CACHE on one side and the I-CACHE on the other side of the chip. For yield improvement, we recommended in Section III to separate the two spare rows from the original 64 rows and to place the two cache units together. The first half of the recommendation is impractical here, but the second half can be implemented in an alternate floorplan (depicted in Fig. 15) in which the two caches are located next to each other.

For the purpose of yield calculation, one row in a cache (1024×1 cells) is considered a module and its area has been chosen as the unit area. The total area of the chip, measured in these units, can be approximated by 1×396=396 area units. The average number of faults per unit area is denoted by λ_1 for the caches and by λ_2 for the remaining area. Usually, $\lambda_2 < \lambda_1$ since the layout of the random logic portion of the chip is less dense than that of the cache units. The fault distribution is the medium-area NB distribution, with a block size varying between 1×1 and 1×396. The projected yields of the two floorplans are depicted in Figs. 16 and 17.

Fig. 16 shows the yield of the original and alternate floorplans (depicted in Figs. 14 and 15, respectively) as a function of λ_1 for two values of the ratio λ_2/λ_1, namely, 0.1 and 0.9. The block size chosen has a height equal to that of the entire chip and a width of 198 cache rows (i.e., half the chip). The value of the clustering parameter is determined so that $\alpha_{\text{chip}} = 2.0$ [12]. We can see that there is no difference in the two yields for $\lambda_2/\lambda_1 = 0.9$, while for $\lambda_2/\lambda_1 = 0.1$ the alternate floorplan is slightly better, especially for the larger values of λ_1. However, for practical values of λ_1 (around $0.003/mm^2$ [12]) the difference is negligible.

Fig. 17 shows the yield of the original and alternate floorplans of the Alpha chip as a function of the block size for several values of λ_1 and the ratio λ_2/λ_1. The height of the block is fixed (set at the length of a single cache row) and the width varies. For very small or very large values of the block size (corresponding to the small-area NB distribution and the large-area NB distribution, respectively), the two floorplans have the same projected yield. For medium-sized blocks the

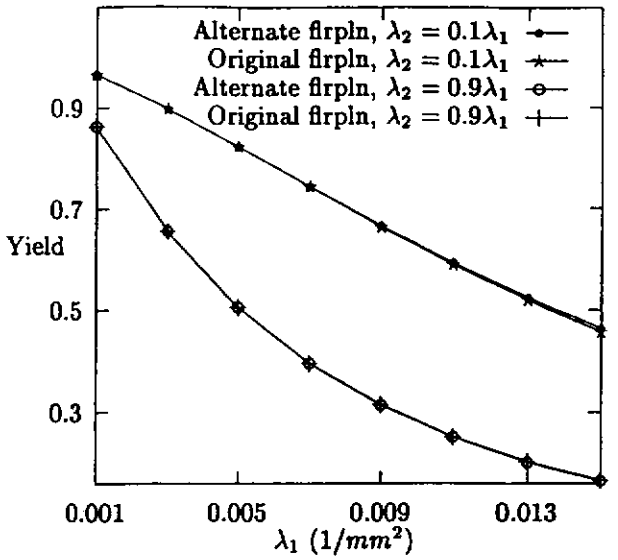


Fig. 16. The yield of the original and alternate floorplans of the Alpha chip as a function of λ_1 for two values of the ratio λ_2/λ_1 (with block size = 198 and $\alpha_{chip} = 2.0$).

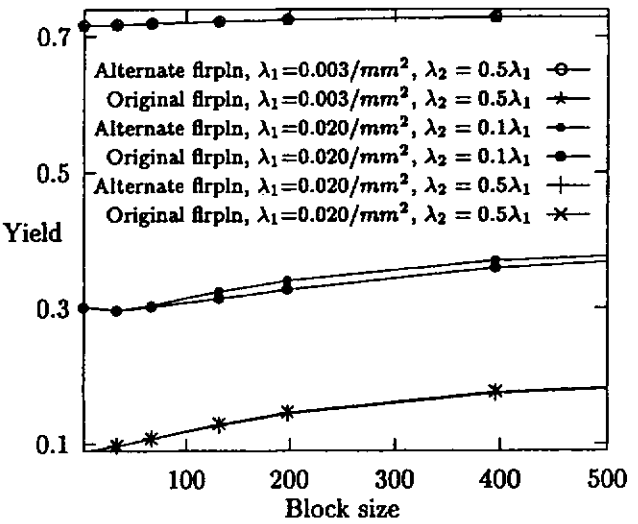


Fig. 17. The yield of the original and alternate floorplans of the Alpha chip as a function of the block size for several values of λ_1 and the ratio λ_2/λ_1 (with $\alpha_{chip} = 2.0$).

difference between the yield of the two floorplans is significant only when λ_1 is relatively high and λ_2 is relatively low. We conclude that although the alternate floorplan has a slightly higher yield, for practical purposes the two floorplans are equally good.

E. The 3-D Computer

The 3-D computer, designed by Hughes Research Laboratories [13], is a cellular array processor implemented in wafer scale integration (WSI) technology. The most unique feature of its implementation is its use of stacked wafers. The basic processing element is divided into five functional units, each of which is implemented on a different wafer. Thus, each wafer contains only one type of functional units and includes spares for yield enhancement as explained below. Units in different

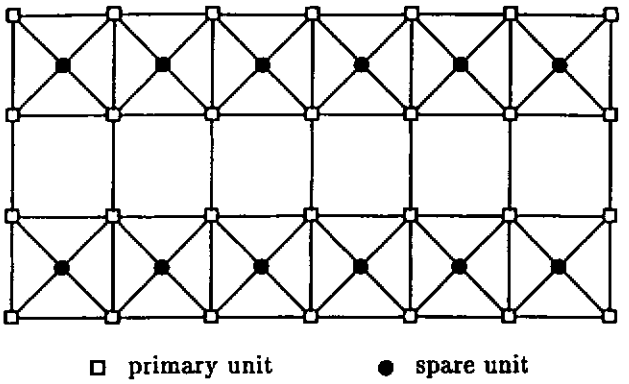


Fig. 18. The original floorplan of a wafer in the 3-D computer.

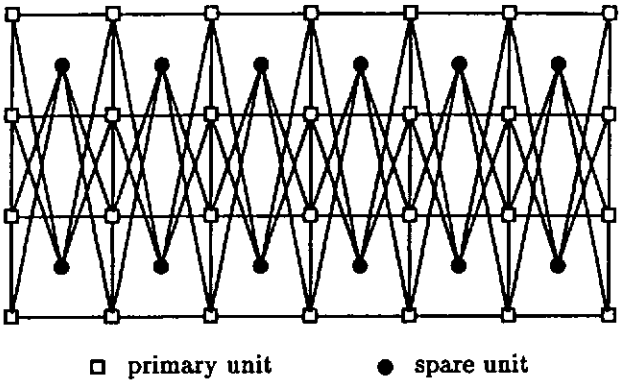


Fig. 19. An alternative floorplan of a wafer in the 3-D computer.

wafers are connected vertically through microbridges between adjacent wafers to form a complete processing element. The first working prototype of the 3-D computer, reported in [13], was of size 32 × 32. The current prototype includes 128 × 128 processing elements.

Fault-tolerance in each wafer is achieved through a (2,4) interstitial redundancy scheme [15]. In this scheme, each primary unit is connected to two spare units, and each spare unit is connected to four primary units, resulting in a redundancy of 50%. There are several ways in which the (2,4) scheme can be applied to two dimensional rectangular arrays [15]. The (2,4) structure that has been selected for implementation in the 3-D computer is shown in Fig. 18 [13].

The floorplan shown in Fig. 18 has every spare unit adjacent to the four primary units that it can replace. This layout has short interconnection links between the spare and any primary unit that it may replace and as a result, the performance degradation upon a failure of a primary unit is minimal. However, the close proximity of the spare and primary units may lead to a low yield in the presence of clustered defects since a single cluster may cover several of these units, as has been experienced in practice [14]. There are several alternative floorplans that place the spare farther apart from the primary units connected to it (as recommended in Section III). Two such floorplans are shown in Figs. 19 and 20.

The yields of the 128 × 128 array using the original floorplan (Fig. 18) or the alternative floorplan (Fig. 19) are shown

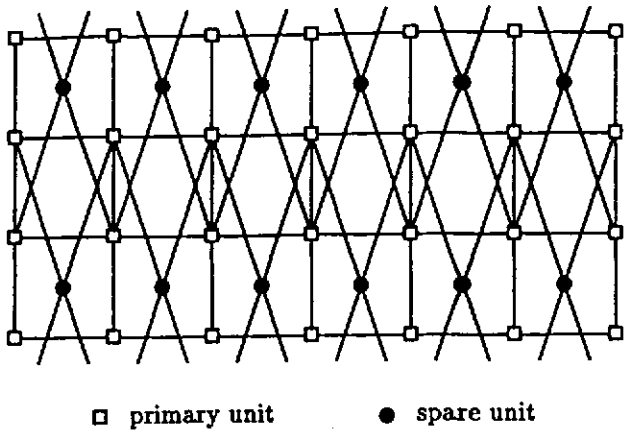


Fig. 20. Another alternative floorplan of a wafer in the 3-D computer.

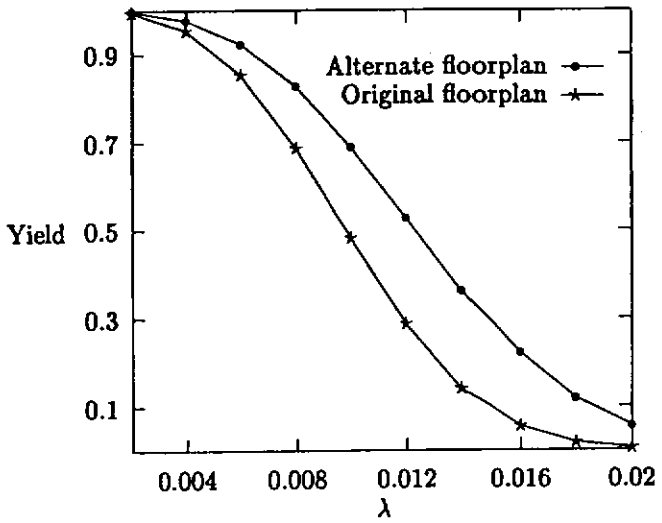


Fig. 21. The yield of the original and alternate floorplans, depicted in Figs. 18 and 19, respectively, as a function of λ ($\alpha = 2$).

in Figs. 21 and 22. The yield has been calculated using the medium-area NB distribution with a block size of two rows of primary units (see Fig. 18). Fig. 21 shows the yield of the original and alternate floorplans, depicted in Figs. 18 and 19, respectively, as a function of λ (the average number of faults per unit) with $\alpha = 2$. It clearly shows that the alternative floorplan, in which the spare unit is separated from the primary units that it can replace, has a higher projected yield. Fig. 22 shows the yield of the original and alternate floorplans (Figs. 18 and 19, respectively) as a function of α for two values of λ. We can see that for low values of the clustering parameter α, indicating heavy clustering, the chosen floorplan has a high impact on the yield, while as α increases and the fault distribution approaches the Poisson distribution, the impact of the particular floorplan becomes less important.

V. Conclusions

The impact of floorplanning on the yield of large area (both fault-tolerant and non fault-tolerant) IC's with medium size fault clusters has been analyzed in this paper. We have shown that under certain circumstances, the selected floorplan

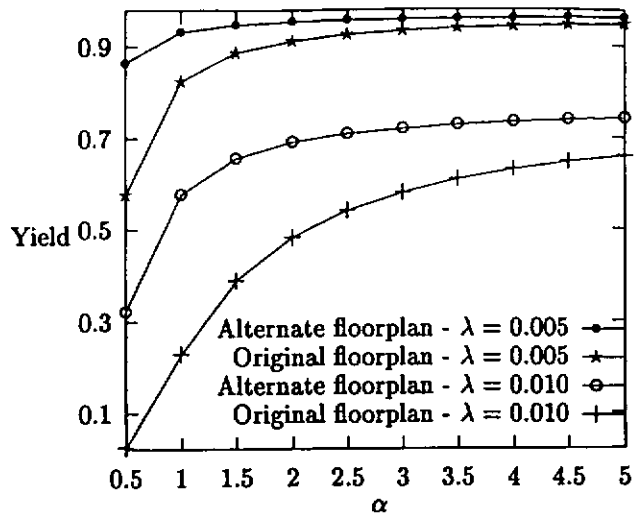


Fig. 22. The yield of the original and alternate floorplans, depicted in Figs. 18 and 19, respectively, as a function of α.

can significantly affect the projected yield. This has been demonstrated through several theoretical test cases and through five practical examples. In some cases, the exact size of the block (cluster of faults) has no effect on the resulting optimal floorplan, while in other cases, a different optimal floorplan emerges under different block sizes. In the latter case, the estimation of the block size is crucial to the floorplan selection, and several estimators for the block size have been suggested in [6]. We conclude that VLSI chip designers should take the yield into consideration, in addition to the more traditional factors like complexity of routing, when determining the floorplan of a new chip.

References

[1] Z. Koren and I. Koren, "Does the floorplan of a chip affect its yield?," in *Proc. 1993 IEEE Workshop Defect Fault Tolerance VLSI Systems*, Oct. 1993, pp. 159-166.

[2] ——, "The impact of floorplanning on the yield of large area IC's," in *Proc. Int. Conf. Wafer Scale Integr.*, Jan. 1995.

[3] F. Meyer and D. K. Pradhan, "Modeling defect spatial distribution," *IEEE Trans. Comput.*, vol. 38, pp. 538-546, Apr. 1989.

[4] I. Koren and C. H. Stapper, "Yield models for defect tolerant VLSI circuit: A review," in *Defect and Fault Tolerance in VLSI Systems*, I. Koren, Ed. New York: Plenum, 1989, pp. 1-21.

[5] I. Koren, Z. Koren, and C. H. Stapper, "A statistical study of defect maps of large area VLSI IC's," *IEEE Trans. VLSI Syst.*, vol. 2, pp. 249-256, June 1994.

[6] ——, "A unified negative binomial distribution for yield analysis of defect tolerant circuits," *IEEE Trans. Comput.*, vol. 42, pp. 724-437, June 1993.

[7] N. P. Jouppi *et al.*, "A 300-MHz 115-W 32-b bipolar ECL microprocessor," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1152-1166, Nov. 1993.

[8] H. Nakano *et al.*, "An 80-MFLOP's (peak) 64-b microprocessor for parallel computer," *IEEE J. Solid-State Circuits*, vol. 27, pp. 365-371, Mar. 1992.

[9] K. Sato *et al.*, "A wafer-scale-level system integrated LSI containing eleven 4-Mb DRAM's, six 64-Kb SRAM's, and an 18K-gate array," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1608-1613, Nov. 1992.

[10] D. W. Dobberpuhl *et al.*, "A 200-MHz 64-b dual-issue CMOS microprocessor," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1555-1565, Nov. 1992.

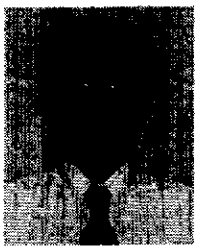
[11] R. S. Collica *et al.*, "A yield enhancement methodology for custom VLSI manufacturing," *Digital Technical J.*, vol. 4, no. 2, pp. 83-99, Spring 1992.

[12] R. S. Collica, private communication.

[13] M. Yung, M. Little *et al.*, "Redundancy for yield enhancement in the 3D computer," in *Proc. Int. Conf. Wafer Scale Integr.*, Jan. 1989, pp. 73-82.
[14] M. Yung, private communication.
[15] A.D. Singh, "Interstitial redundancy: An area efficient fault tolerance scheme for larger area VLSI processor array," *IEEE Trans. Comput.*, vol. 37, pp. 1398-1410, Nov. 1988.

Zahava Koren received the B.A. and M.A. degrees in mathematics and statistics from the Hebrew University, Jerusalem, Israel, in 1967 and 1969, respectively, and the D.Sc. degree in operations research from the Technion—Israel Institute of Technology, Haifa in 1976.

She is currently with the Department of Industrial Engineering at the University of Massachusetts, Amherst. Previously, she has held positions with the Department of Statistics, University of Haifa, the Departments of Industrial Engineering and Computer Science at the Technion, and the Department of Business and Economics, California State University in Los Angeles. Her main interests are stochastic analysis of computer networks, yield of integrated circuits, and reliability of computer systems.



Israel Koren (S'72–M'76–SM'87–F'91) received the B.Sc., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa in 1967, 1970, and 1975, respectively, all in electrical engineering.

He is currently a Professor of Electrical and Computer Engineering at the University of Massachusetts, Amherst. Previously, he was with the Departments of Electrical Engineering and Computer Science at the Technion. He also held visiting positions with the University of California at Berkeley, the University of Southern California, Los Angeles, and the University of California, Santa Barbara. He has been a consultant to several companies including Intel, Digital Equipment Corporation, National Semiconductor and Tolerant Systems. His research interests are fault-tolerant VLSI architectures, models for yield and performance, floorplanning of VLSI chips, and computer arithmetic. He has published extensively in several IEEE TRANSACTIONS and has over 100 publications in refereed journals and conferences. He has edited and coauthored the book, *Defect and Fault-Tolerance in VLSI Systems* (New York: Plenum, 1989, vol. 1). He is the author of the textbook *Computer Arithmetic Algorithms* (Englewood Cliffs, NJ: Prentice-Hall, 1993).

Dr. Koren has served on the Editorial Board of the IEEE TRANSACTIONS ON COMPUTERS. He has also served as program committee member for numerous conferences. He was Co-Guest Editor for the Special Issue on High Yield VLSI Systems, April 1989 for the IEEE TRANSACTIONS ON COMPUTERS.