

# A Statistical Study of Defect Maps of Large Area VLSI IC's

Israel Koren, *Fellow, IEEE*, Zahava Koren, and Charles H. Stapper, *Fellow, IEEE*

**Abstract**—Defect maps of 57 wafers containing large area VLSI IC's were analyzed in order to find a good match between the empirical distribution of defects and a theoretical model. Our main result is that the commonly employed models, most notably, the large area clustering negative binomial distribution, do not provide a sufficiently good match for these large area IC's. Only the recently proposed medium size clustering model is close enough to the empirical distribution. An even better match can be obtained either by combining two theoretical distributions or by a “censoring” procedure in which the worst chips are ignored. Another goal of the study was to find out whether certain portions of either the chip or the wafer had more defects than the others.

**Index Terms**—Defect maps, large area clustering model, low quality IC's, medium area clustering model, yield models.

## I. INTRODUCTION

**D**EFFECTS MAPS of 57 wafers containing large area VLSI chips were analyzed with several purposes in mind: To find a good match between a theoretical yield model and the empirical spatial distribution of defects; to determine a criterion for distinguishing between “low quality” and “high quality” chips; and to determine whether certain areas on the wafer or on the chip are more prone to defects than others. See Fig. 1 for an example of a defect map. In this figure, the wafer is divided into  $12 \times 6 = 72$  chips, and the circles denote the locations of the defects found on the wafer.

The preliminary results of this study were reported in [5]. Unfortunately, the defect maps that were analyzed at that time were found later on to include what seemed like systematic defects along either horizontal or vertical lines, but were actually introduced by the measuring equipment. We had therefore, to remove these errors and repeat the analysis. This paper contains the analysis results using the corrected wafer maps, and includes new performance measures, graphs, and statistical tests which do not appear in [5].

## II. MATCHING A THEORETICAL DISTRIBUTION

The first objective of our study was to match a yield model to the empirical defect distribution. The goodness of

Manuscript received June 24, 1993; revised November 18, 1993. A preliminary version of this work was presented at the 1992 IEEE Workshop on Defect and Fault Tolerance in VLSI Systems, Dallas, TX, USA [5]. This work was supported in part by IBM under Contract 81775.

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

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

C. H. Stapper is with IBM Technology Products, Essex Junction, VT 05452 USA. He is now an independent consultant in Jericho, VT 05465.

IEEE Log Number 9400598.

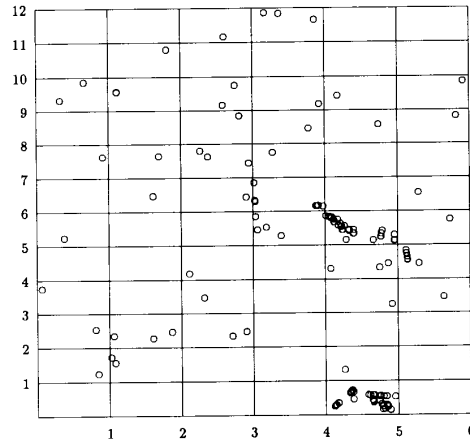


Fig. 1. An example wafer defect map with high defect density.

fit of a theoretical yield model to the empirical results can be measured in two different ways, based on the random variable of interest. One way is observing  $D_{\text{chip}}$ , the number of defects per chip, and comparing its theoretical and empirical distributions. An alternative method is through dividing the chip into equally-sized modules, and comparing the theoretical and empirical distributions of the random variable  $G_{\text{chip}}$ , defined as the number of defect-free (good) modules on a chip. This measure is of interest when the chips have some redundancy incorporated into them or can be used as partially good chips, since in those cases the yield of the chip is determined by the distribution of the number of its defect-free modules rather than by the defect distribution. The first method, i.e., that of comparing the defect distributions is more detailed, and therefore, more sensitive to small deviations. Comparing the good modules distributions is more robust and may in some cases be sufficient. Both methods of comparison are described and demonstrated next.

Denote by  $FD(k)$  the relative frequency and by  $PD(k)$  the theoretical probability (according to some yield model) of chips with  $k$  defects

$$PD(k) = \text{Prob}[D_{\text{chip}} = k], \quad (k = 0, 1, 2, \dots)$$

Let  $CFD(k)$  and  $CPD(k)$  denote the cumulative relative frequency and the cumulative probability, respectively, of  $k$  or

less defects in a chip

$$\text{CFD}(k) = \sum_{i=0}^k \text{FD}(i), \quad \text{CPD}(k) = \sum_{i=0}^k \text{PD}(i)$$

To test the goodness of fit between the two distributions we calculate the maximum absolute difference between  $\text{CFD}(k)$  and  $\text{CPD}(k)$ , denoted by DIFD.

$$\text{DIFD} = \max_{k \geq 0} |\text{CFD}(k) - \text{CPD}(k)| \quad (1)$$

According to the Kolmogorov-Smirnov test for goodness of fit [3], the fit between the two distributions is good if DIFD does not exceed 0.022 (the critical value for a sample size of 3876, which is the number of chips analyzed in this study, and for a significance level of 0.05).

Similarly, denote by  $\text{FG}(m)$  and  $\text{PG}(m)$  the relative frequency and the theoretical probability, respectively, of chips with  $m$  defect-free modules, and by  $\text{CFG}(m)$  and  $\text{CPG}(m)$  the cumulative relative frequency and cumulative probability, respectively, of chips with  $m$  or less defect-free modules.

$$\begin{aligned} \text{PG}(m) &= \text{Prob}[G_{\text{chip}} = m] \\ \text{CFG}(m) &= \sum_{i=0}^m \text{FG}(i), \quad \text{CPG}(m) = \sum_{i=0}^m \text{PG}(i) \end{aligned}$$

In this paper, as explained later, the chip is divided into 16 modules, and  $m$ , therefore, assumes the values 0, 1, ..., 16.

Using the same criterion as in (1), we define

$$\text{DIFG} = \max_{0 \leq m \leq 16} |\text{CFG}(m) - \text{CPG}(m)| \quad (2)$$

and accept the theoretical model (with a significance level of 0.05) if  $\text{DIFG} \leq 0.022$ .

In the past, the large area negative binomial distribution has been successfully used to describe the distribution of defects on a chip [1]. Denoting by  $D(A)$  the number of defects in an area of size  $A$  (contained within a chip), the probability function for this distribution is,

$$\begin{aligned} \text{Prob}[D(A) = k] &= \text{Prob}(k \text{ defects in area } A) \\ &= \frac{\Gamma(k + \alpha_A)}{k! \Gamma(\alpha_A)} \frac{(\lambda_A / \alpha_A)^k}{(1 + \lambda_A / \alpha_A)^{k + \alpha_A}} \end{aligned} \quad (3)$$

where  $\lambda_A$  and  $\alpha_A$  are the average number of defects and the clustering parameter for area  $A$ , respectively.

To test whether the same distribution can be used for the examined large area VLSI chips, we calculated the empirical frequency of chips with  $k$  defects,  $k = 0, 1, 2, \dots$ . We then tried to fit to it a large area negative binomial distribution, with  $\lambda_{\text{chip}} = 2.056$  (the average number of defects per chip in our sample) and an  $\alpha_{\text{chip}}$  which provided the best fit to the empirical results. The cumulative empirical frequency of chips with  $k$  defects or less and the cumulative theoretical probability of  $k$  defects or less are depicted in Fig. 2.

Since the fit did not prove good enough, we divided every chip into four equally sized quadrants, and divided each quadrant into four equally sized modules. The chip is thus divided into 16 modules. We compared the empirical cumulative frequency of quadrants with  $k$  defects and of modules with

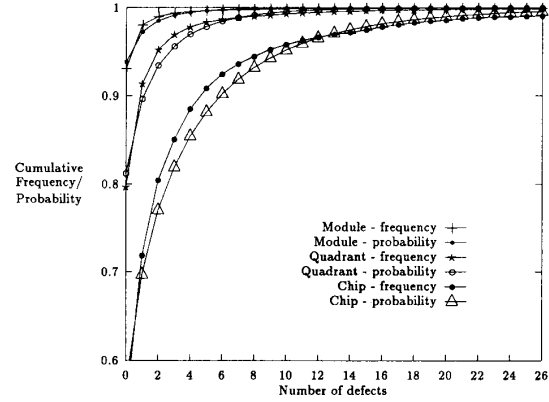


Fig. 2. Comparing the empirical cumulative frequency of the number of defects in a module/quadrant/chip to the large area negative binomial model.

$k$  defects to the corresponding large area negative binomial probabilities, under the constraints  $\lambda_{\text{quadrant}} = 0.514$  and  $\lambda_{\text{mod}} = 0.128$  and the best  $\alpha_{\text{quadrant}}$  and  $\alpha_{\text{mod}}$ . The results of these comparisons appear in Fig. 2. The conclusion that can be drawn from Fig. 2 is that the large area negative binomial distribution is not suitable for describing the defect distribution of the given large area chips. The smaller the area, the better the fit. In this case, the fit is best when the number of defects per module is concerned, but it is very poor when attempting to describe the defects on the chip as a whole.

We, therefore, tried to identify another theoretical yield model for describing the spatial distribution of the defects. We considered the following three probability distributions for  $D(A)$  (the number of defects in an area of size  $A$ ):

I. The Poisson distribution, according to which

$$\text{Prob}[D(A) = k] = e^{-\lambda_A} \frac{\lambda_A^k}{k!} \quad (4)$$

II. The small area negative binomial distribution [8]. Consider the module to be the basic unit of area and let any other area be measured in these units. According to this distribution, the number of defects per module, denoted by  $D_{\text{mod}}$ , is assumed to have a negative binomial distribution, and the number of defects in the different modules is assumed to be statistically independent. The probability function of  $D(A)$ , for an area of size  $A$  which is an integer multiple of a module, will, therefore, be the  $A$ -fold convolution of the probability function of  $D_{\text{mod}}$ , or

$$\begin{aligned} \text{Prob}[D(A) = k] &= \sum_{i_1 + \dots + i_A = k} \dots \sum_{i_1 + \dots + i_A = k} \prod_{j=1}^A \text{Prob}[D_{\text{mod}} = i_j] \\ &= \sum_{i_1 + \dots + i_A = k} \dots \sum_{i_1 + \dots + i_A = k} \prod_{j=1}^A \frac{\Gamma(i_j + \alpha_{\text{mod}})}{i_j! \Gamma(\alpha_{\text{mod}})} \\ &\quad \times \frac{(\lambda_{\text{mod}} / \alpha_{\text{mod}})^{i_j}}{(1 + \lambda_{\text{mod}} / \alpha_{\text{mod}})^{i_j + \alpha_{\text{mod}}}} \end{aligned} \quad (5)$$

III. The medium area negative binomial distribution [6] with different block sizes, namely,  $2 \times 2$ ,  $2 \times 3$ ,  $2 \times 4$ ,  $3 \times 3$  and  $4 \times 4$  modules. According to this distribution, the basic

unit is a block consisting of  $R \times C$  modules, the number of defects in a block has a negative binomial distribution, and the defects in distinct blocks are statistically independent. The distributions of  $D(A)$  can be calculated (similarly to (5)) as the convolution of the distribution of  $D_{\text{block}}$  (the number of defects per block). For a thorough discussion of the medium area negative binomial distribution see [6]. Note that the small and the large area negative binomial distributions are special cases of the medium area negative binomial distribution, with block sizes of  $1 \times 1$  and  $48 \times 24$  (the whole wafer), respectively.

Once a specific distribution out of the three mentioned above is selected, and since the area of a module is the unit area and the chip consists of 16 modules, the theoretical probability function of  $D_{\text{chip}}$  can be calculated as

$$PD(k) = \text{Prob}[D_{\text{chip}} = k] = \text{Prob}[D(16) = k] \quad (\text{for } k = 0, 1, 2, \dots)$$

The probability function of  $G_{\text{chip}}$  has been calculated using the inclusion and exclusion formula. The probability of  $m$  defect-free modules in a chip, denoted by  $PG(m)$ , is given by

$$PG(m) = \text{Prob}[G_{\text{chip}} = m] = \binom{16}{m} \sum_{j=0}^{16-m} (-1)^j \times \binom{16-m}{j} \text{Prob}[D(m+j) = 0] \quad (\text{for } m = 0, 1, \dots, 16) \quad (6)$$

Equation (6) assumes implicitly that the probability  $\text{Prob}[D(m+j) = 0]$  is a function of the area size  $m+j$  only. Note that although, for a general spatial defect distribution, the distribution of  $D(A)$  depends not only on the size  $A$  of the area but also on its shape and location, this is not the case for the Poisson, large area negative binomial, and small area negative binomial distributions. For these three distributions,  $\text{Prob}[D(A) = k]$  is a function of  $A$  only. When using the medium area negative binomial distribution, for which the probability function of  $D(A)$  does depend on the shape of the area, a slightly different version of (6) is used (as can be found in [6]).

The parameters for the different distributions have been estimated so as to minimize the maximal differences DIFD and DIFG under the constraint  $\lambda_{\text{mod}} = 0.128$  (the empirical average). The empirical and theoretical cumulative distributions for  $D_{\text{chip}}$  (defects per chip) and  $G_{\text{chip}}$  (good modules per chip) are depicted in Figs. 3 and 4 and the maximal differences appear in Tables I and II, respectively.

Note that when comparing the empirical distribution of defects per chip to the different probability distributions (Fig. 3), the curves of the large area and the small area negative binomial distributions almost coincide (although both are quite far from the empirical curve). This is an artifact of the numerical search for the "best" values of  $\alpha_{\text{chip}}$  and  $\alpha_{\text{mod}}$  and the similarity of the expression for  $\text{Prob}[D(A) = 0]$  in these two distributions. This, however, is not the case in Fig. 4, where the two distributions have distinctively different curves. Fig. 4 depicts the distribution of the number of defect-free modules per chip, and this number depends heavily on the

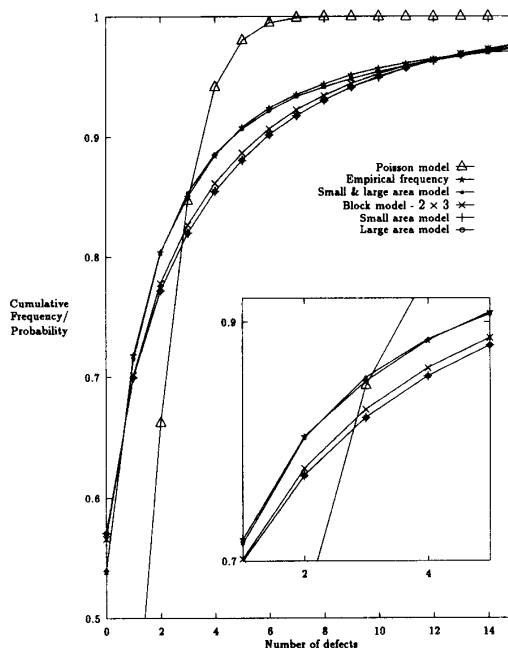


Fig. 3. Comparing the empirical cumulative frequency of the number of defects in a chip to theoretical distributions.(Inset shows detail).

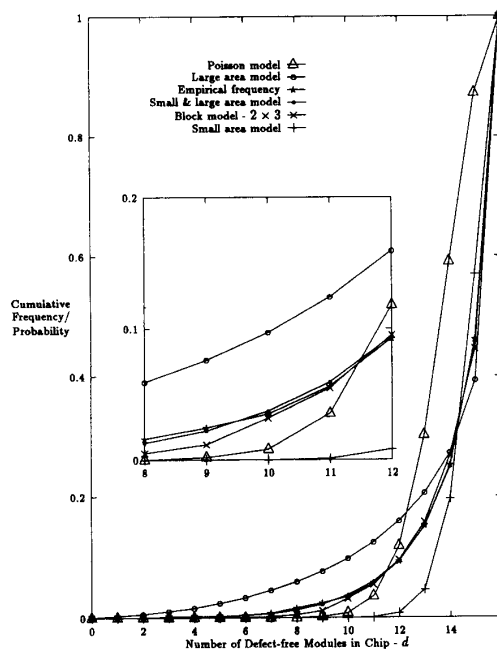


Fig. 4. Comparing the empirical cumulative frequency of the number of defect-free modules in a chip to theoretical distributions.(Inset shows detail).

clustering pattern and not only on the number of defects in the chip.

As can be seen in Tables I and II, out of the single probability distributions mentioned above, only the medium-

area negative binomial with a block size of  $2 \times 3$  had a satisfactory match with the empirical distribution of the number of defect-free modules per chip. For this probability model we obtain a maximal difference  $DIFG = 0.0164$ , which is below the limit of 0.022. (For  $DIFG = 0.0164$ , the theoretical model will be rejected only if the significance level chosen is greater than 0.25, which is usually considered too large). However, for the same distribution we obtain  $DIFD = 0.0269$  and the fit is not good when defect distribution is considered rather than good modules distribution (with a difference of 0.0269, the hypothesis of fit will be rejected at a significance level which is greater than or equal to 0.007). A different modeling approach should, therefore, be taken.

The two other approaches which we attempted were: to search for a combination of two defect distributions that will match the empirical distribution, or to remove those chips with a very large number of defects and try to fit a probability distribution to the remaining chips. These two approaches are described in the next two sections. Note that they are not equivalent, and each can be used for different purposes. The first approach should be used if the yield of all the manufactured chips is to be estimated, while the second is appropriate if the worst chips are discarded and we are interested in the yield of the remaining chips.

### III. COMBINING TWO PROBABILITY FUNCTIONS

Observing the defect maps indicates that the defects can be viewed as originating from two sources: systematic and random, or heavily clustered and less heavily clustered (see, for example, Fig. 1). To verify this observation, we calculated the clustering measure  $\alpha_A$  for the different area sizes  $A$  between  $1 \times 1$  and  $4 \times 4$  modules. For a given area size  $A$ ,  $\alpha_A$  was calculated as the solution of the equation

$$\hat{Y}_A = \left(1 + \frac{\lambda_A}{\alpha_A}\right)^{-\alpha_A}$$

where  $\hat{Y}_A$  is the empirical yield of an area of size  $A$  and  $\lambda_A = 0.128 \cdot A$ . A constant value of  $\alpha_A$  for every  $A$  would indicate large area clustering, while values linearly increasing with  $A$  would indicate small area clustering. The calculated  $\alpha_A$ 's were neither constant nor linearly increasing, but increased in a less than linear fashion. This again points to the presence of two independent defect sources.

In the past, the presence of systematic defects has been modeled by a "gross yield factor" [4]. This approach proved suitable when calculating the theoretical distribution of  $G_{\text{chip}}$ , and is equivalent to combining two distributions; one is the random defect distribution (usually the large area negative binomial) and the other models large defects which render the whole chip faulty. According to this second distribution,  $G_{\text{chip}}$  can only assume two values, namely, 0 (good modules) and 16 (good modules). When these two distributions are combined, the result is a bi-modal distribution of  $G_{\text{chip}}$ , with one of the modes at 0. This approach is suitable for small and medium area chips where the "gross yield" effect is noticeable. It cannot be applied to our data since the empirical distribution of  $G_{\text{chip}}$  did not have a mode at 0, probably

due to the large area of the chip. We, therefore, searched for some other combination of two probability distributions which would fit the data. We considered all combinations of two out of the following three distributions: the Poisson distribution, the large area negative binomial distribution, and the small area negative binomial distribution. The medium area negative binomial distribution has not been considered, since it by itself has 4 parameters, and combined with some other distribution we would have to search for 6 parameters.

The theoretical probabilities of  $D_{\text{chip}}$  (the number of defects per chip) and of  $G_{\text{chip}}$  (the number of defect-free modules per chip) for a combination of two distributions have been derived as follows. Denoting by  $D(A)$  the number of defects in an area of size  $A$ , we assume that  $D(A) = D_1(A) + D_2(A)$  where  $D_1(A)$  and  $D_2(A)$  are statistically independent random variables, denoting the number of defects of type 1 and of type 2, respectively, in the area of size  $A$ .

Since the chip is divided into 16 modules and assuming that a module has an area of 1,  $D_{\text{chip}} = D(16)$  and the probability distribution of  $D_{\text{chip}}$  is the convolution of the two probability distributions of  $D_1(16)$  and  $D_2(16)$ ,

$$\begin{aligned} PD(k) &= \text{Prob}[D_{\text{chip}} = k] = \\ &= \sum_{j=0}^k \text{Prob}[D_1(16) = j] \times \text{Prob}[D_2(16) = k - j] \end{aligned} \quad (7)$$

The expressions used for  $\text{Prob}[D_1(16) = j]$  and  $\text{Prob}[D_2(16) = k - j]$  appear in (3), (4), and (5) for the large area, Poisson, and small area distribution, respectively.

The probability distribution of  $G_{\text{chip}}$  has been calculated using the inclusion and exclusion formula, as given by Equation (6). Since  $D(A) = D_1(A) + D_2(A)$  and  $D_1(A)$  and  $D_2(A)$  are assumed to be statistically independent,  $\text{Prob}[D(A) = 0] = \text{Prob}[D_1(A) = 0] \cdot \text{Prob}[D_2(A) = 0]$ , and

$$\begin{aligned} PG(m) &= \binom{16}{m} \sum_{j=0}^{16-m} (-1)^j \binom{16-m}{j} \text{Prob}[D_1(m+j) = 0] \\ &\quad \times \text{Prob}[D_2(m+j) = 0] \end{aligned} \quad (8)$$

The expressions used for  $P[D_i(m+j) = 0]$  ( $i = 1, 2$ ) are special cases of (3), (4), and (5) with  $k = 0$ . Specifically,

$$\text{Prob}[D_i(m+j) = 0] = \left(1 + \frac{(m+j)\lambda_{\text{mod}}}{\alpha_{\text{chip}}}\right)^{-\alpha_{\text{chip}}} \quad (\text{large area})$$

$$\text{Prob}[D_i(m+j) = 0] = e^{-(m+j)\lambda_{\text{mod}}} \quad (\text{Poisson})$$

$$\text{Prob}[D_i(m+j) = 0] = \left[\left(1 + \frac{\lambda_{\text{mod}}}{\alpha_{\text{mod}}}\right)^{-\alpha_{\text{mod}}}\right]^{m+j} \quad (\text{small area})$$

Combining the Poisson distribution with the small area or with the large area negative binomial distribution [based on (7) and (8)] did not prove to be close enough to the empirical distribution of either  $D$  or  $G$ . The small area negative binomial, on the other hand, provided a very good fit when combined with the large area negative binomial ( $DIFD = 0.0032$  and  $DIFG = 0.0031$ ). The results of the  $D_{\text{chip}}$

TABLE I  
COMPARING THE EMPIRICAL CUMULATIVE FREQUENCY OF THE NUMBER OF DEFECTS IN A CHIP TO THEORETICAL DISTRIBUTIONS

	Poisson distribution	Large Area clustering 48 × 24 block	Small Area clustering 1 × 1 block	Medium Area clustering 2 × 2 block	Medium Area clustering 2 × 3 block	Combined Small Ar. & Large Ar. clustering
Maximal difference	0.4107	0.0323	0.0324	0.0297	0.0269	0.0031

TABLE II  
COMPARING THE EMPIRICAL CUMULATIVE FREQUENCY OF THE NUMBER OF DEFECT-FREE MODULES IN A CHIP TO THEORETICAL DISTRIBUTIONS

	Poisson distribution	Large Area clustering 48 × 24 block	Small Area clustering 1 × 1 block	Medium Area clustering 2 × 2 block	Medium Area clustering 2 × 3 block	Combined Small Ar. & Large Ar. clustering
Maximal difference	0.4107	0.0685	0.1087	0.0273	0.0164	0.0032

comparison appear in Fig. 3 and Table I, while the results of the  $G_{\text{chip}}$  comparison appear in Fig. 4 and in Table II.

As mentioned before, we estimated the parameters for each of the models by searching through the parameter space, so as to minimize the maximal differences DIFD and DIFG under the constraint  $\lambda_{\text{mod}} = 0.128$  (the empirical average). This method of estimation is similar, but not identical, to the maximum likelihood estimation method, and proved to be very time consuming. Other possible approaches for obtaining estimators for the parameters of the combined models are the moment method, the maximum likelihood method, or the “least squares” method in which the sum of the squares of the differences between the theoretical and empirical yields is minimized. The theoretical yields for every block size  $R \times C$  between  $1 \times 1$  and  $4 \times 4$  are:

$$Y_{RC} = \text{Prob}[D_1(RC) = 0] \cdot \text{Prob}[D_2(RC) = 0] \quad (9)$$

The empirical values of  $Y_{RC}$ ,  $\hat{Y}_{RC}$ , can be calculated for every  $(R, C)$ , and “least squares” estimates for the parameters can be obtained.

#### IV. SEPARATING THE “LOW QUALITY” ICs

A different method of modeling the defects in large area chips is separating the “low quality” from the “high quality” chips by determining a “cutoff” number of defects, such that a chip with this number of defects or more is considered of low quality and is ignored. This procedure can be justified in two different ways, practical and theoretical. Practically, some IC manufacturers actually throw away the heavily defective chips and deal with the rest. Theoretically, there exists a statistical method in which “outliers” (i.e., observations which seem to be very different from the rest) are ignored when the analysis is performed.

The candidates for being a cutoff point are: the average number of defects per chip, the median (thus throwing away the worst 50% of the chips), the 95th percentile (thus ignoring the worst 5% of the chips), or any other suitable percentile. In our sample, the median number of defects per chip is 0 and

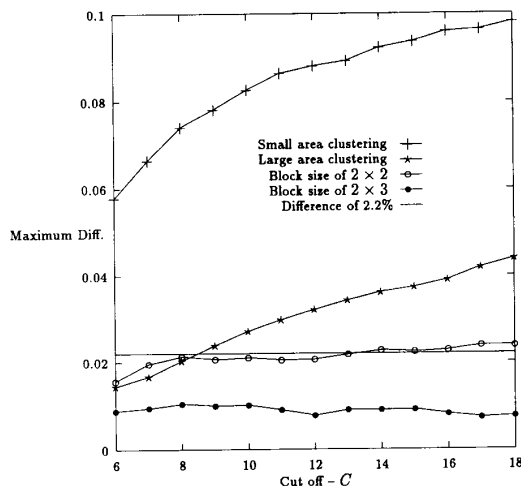


Fig. 5. Maximum difference between theoretical models and experimental results.

the average is 2.06, both too small to serve as cutoff points. The 90th percentile is 6, while the 97.5 percentile is 18. Any value between these two is a reasonable cutoff point.

For every cutoff point  $C$ , we “threw away” all the chips with  $C$  defects or more, and tried to fit a defect distribution to the remaining chips. The relative frequency of chips with  $m$  defect-free modules has been calculated (for  $m = 0, \dots, 16$ ) and compared to the theoretical probability obtained from four yield models: The large area negative binomial distribution, the small area negative binomial distribution, the medium area negative binomial distribution with a block size of  $2 \times 2$ , and the medium area negative binomial distribution with a block size of  $2 \times 3$ . The goodness of fit between the theoretical and empirical distribution has been measured by DIFG defined in (2). Fig. 5 depicts the maximal differences between the cumulative frequency and probability of  $G_{\text{chip}}$  for  $6 \leq C \leq 18$  and the four theoretical distributions. As can be seen from Fig. 5, the best fit has been obtained for the  $2 \times 3$  block, for all cutoff points. Note that the large area clustering distribution provides a reasonable approximation only for very low values of the cutoff point, while the small area clustering distribution provides a bad fit for all cutoff values.

Additional information that can be obtained by using the “censoring” technique is demonstrated in Table III. Table III shows the implications of choosing a specific cutoff point, namely, the percentage of chips that will be thrown away, the yield of the remaining chips (i.e.,  $\text{Prob}[D_{\text{chip}} = 0]$ ), and the parameters of the theoretical yield model of the remaining chips. The latter can be used to calculate the distribution of  $G_{\text{chip}}$  which is relevant when the chips have some incorporated redundancy, or can be used as partially good chips.

#### V. DEPENDENCE ON LOCATION

When analyzing very large area chips, two types of questions come to mind. One is whether the different areas on

the same chip are statistically dependent or independent with regard to defects, and the other is whether any specific area is more prone to defects than the other. The statistical tests which are to be used to answer the second question depend largely on the answer to the first question. If the different areas are independent, then tests for independent samples have to be used, while if they are dependent, then tests for paired samples must be used. Another problem is—can tests based on the normal distribution be used, or do we have to use nonparametric tests.

We were interested in comparing the top half and the bottom half of the chip, the left half and the right half of the chip, the boundaries and the center of the chip, and the boundaries and the center of the wafer, in order to detect any significant differences with regard to the number of defects.

For each of these comparisons, we performed four different tests, namely, the  $t$ -test for independent samples, the  $t$ -test for paired samples, the Mann-Whitney test for independent samples, and the Wilcoxon test for paired samples. The first two tests are based on the fact that although the number of defects per chip is not normally distributed, tests based on the normal distribution can be used with good accuracy when the sample size is very large (see [2]). The last two tests are nonparametric and do not assume any specific distribution of the defects. All four tests will be described next and their numerical results are summarized in Table IV.

We first describe the tests comparing the top half and the bottom half of a chip. Every chip was divided into four quadrants, and the number of defects in each quadrant was counted. Denote by  $x_i$  and  $y_i$  the number of defects in the top quadrant and in the bottom quadrant, respectively, for pair number  $i$  of adjacent quadrants. Since the sample consisted of 3876 chips, we had  $n = 7752$  pairs  $(x_i, y_i)$ .

Denote:

$$d_i = x_i - y_i; \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i;$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i;$$

$$S_x^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right);$$

$$S_y^2 = \frac{1}{n-1} \left( \sum_{i=1}^n y_i^2 - n\bar{y}^2 \right);$$

$$S_d^2 = \frac{1}{n-1} \left( \sum_{i=1}^n d_i^2 - n\bar{d}^2 \right)$$

where  $\bar{x}$ ,  $\bar{y}$ , and  $\bar{d}$  are the average number of defects per top quadrant, the average number of defects per bottom quadrant, and the average difference, respectively.  $S_x^2$ ,  $S_y^2$ , and  $S_d^2$  are the unbiased estimates of the corresponding variances.

The test statistic used in the  $t$ -test for independent samples is

$$t_{\text{indep.}} = \frac{\bar{x} - \bar{y}}{\sqrt{S_x^2/n + S_y^2/n}}$$

TABLE III  
IMPLICATIONS OF DIFFERENT CUTOFF VALUES

Cutoff value	% thrown away chips	Remaining yield	$\lambda$	$\alpha$
6	10.1%	59.3%	0.82	0.72
7	8.2%	58.3%	0.91	0.60
8	6.8%	57.6%	0.99	0.55
9	5.9%	57.0%	1.05	0.52
10	5.0%	56.6%	1.11	0.49
11	4.4%	56.2%	1.17	0.40
12	3.9%	56.0%	1.21	0.39
13	3.5%	55.8%	1.25	0.37
14	3.2%	55.6%	1.29	0.35
15	2.9%	55.4%	1.32	0.35
16	2.6%	55.3%	1.36	0.34
17	2.3%	55.1%	1.41	0.34
18	2.0%	54.9%	1.45	0.29

TABLE IV  
DEPENDENCE OF THE NUMBER OF DEFECTS ON LOCATION

Test		Wafer boundary vs. Wafer center	Chip boundary vs. Chip center	Upper half vs. Lower half	Left half vs. Right half
$t$ -test for Independent Samples	Average Difference	-0.021	-0.533	-0.074	-0.018
	Standard Deviation	0.110	0.073	0.032	0.032
	$t$	-0.195	-7.264	-2.282	-0.566
	Significant? (level=0.05)	No	Yes	Yes	No
$t$ -test for paired Samples	Average Difference	-0.021	-0.533	-0.074	-0.018
	Standard Deviation	0.075	0.051	0.024	0.029
	$t$	-0.287	-10.416	-3.040	-0.635
	Significant? (level=0.05)	No	Yes	Yes	No
Mann-Whitney test for Independent Samples	$W - E(W)$	-184.5	-785283	-987700	-70474
	Standard Deviation	176.45	98521	278537	278537
	$z$	-1.045	-7.971	-3.546	-0.253
	Significant? (level=0.05)	No	Yes	Yes	No
Wilcoxon test for paired Samples	$W - E(W)$	-116.5	-243876	-169681	-24242
	Standard Deviation	125.860	18415	31108	31479
	$z$	-0.926	-13.243	-5.455	-0.770
	Significant? (level=0.05)	No	Yes	Yes	No

The difference is significant if  $|t_{\text{indep.}}| > 1.96$  (at a significance level of 0.05).

If the two samples are assumed to be correlated, the  $t$  statistic to be used is

$$t_{\text{paired}} = \frac{\bar{d}}{\sqrt{S_d^2/n}}$$

and the difference is significant if  $|t_{\text{paired}}| > 1.96$ .

The two nonparametric tests are based on ranking the observations rather than on the observations themselves, and are therefore distribution-free [3]. In the Mann-Whitney test for independent samples, the  $2n = 15504$  observations  $x_1, \dots, x_n, y_1, \dots, y_n$  are combined, and ranked from smallest to largest.  $W$  is then defined as the sum of the ranks associated with the  $x_i$ 's. If there is no difference between the upper half

and the lower half of the chip then

$$E(W) = \frac{n(2n+1)}{2} \text{ and } \sigma^2(W) = \frac{n^2(2n+1)}{12}$$

Based on the central limit theory, the test statistic is

$$z = \frac{W - E(W)}{\sigma(W)}$$

and the difference is significant if  $|z| > 1.96$ .

In the equivalent of the Mann-Whitney test for paired samples, called the Wilcoxon test, the  $n$  differences  $d_1, \dots, d_n$  are ranked according to their absolute value from smallest to largest, and  $W$  is the sum of the ranks associated with the positive  $d_i$ 's. If there is no difference between the upper half and the lower half of the chip then

$$E(W) = \frac{n(n+1)}{4} \text{ and } \sigma^2(W) = \frac{n(n+1)(2n+1)}{24}$$

Based on the central limit theory, the test statistic is

$$z = \frac{W - E(W)}{\sigma(W)}$$

and the difference is significant if  $|z| > 1.96$ .

The comparison of the left half and the right half of a chip is done similarly, with  $x_1, \dots, x_n$  denoting the number of defects in the left quadrants and  $y_1, \dots, y_n$  denoting the number of defects in the right quadrants ( $n = 7752$ ).

To enable the comparison between the boundary of a chip and its center, each chip has been divided into an outer and an inner area of equal size. The number of defects in the outer areas of the different chips is denoted by  $x_1, \dots, x_n$ , and the number of defects in the inner areas is denoted by  $y_1, \dots, y_n$ , where  $n = 3876$  is the number of chips. The four statistical tests described earlier have now been performed using these two samples. The comparison of the wafer boundary to the wafer center is done by dividing each wafer into its outer layer, with a width equal to half the chip's width, and its inner area.  $x_1, \dots, x_n$  denotes the average number of defects per module in the outer areas, and  $y_1, \dots, y_n$  denotes the average number of defects per module in the inner areas ( $n = 57$ ).

The numerical results are summarized in Table IV. Based on these results, at a significance level of 5%, it is apparent that all four tests result in the same conclusion, for any of the four comparisons. Specifically, we conclude that the number of defects in the center of a chip is significantly larger than that on the boundary. In addition, the number of defects in the bottom half of the chip was found to be significantly larger than that of the top half, while there is no significant difference between the right half and the left half. These observations may prove to be of significance to the wafer fabrication process engineers. Another important conclusion that can be drawn is that there is no significant difference between the wafer boundary and its center. This is in contrast to the common belief that the number of defects per chip goes up as the distance from the center of the wafer increases.

The above conclusions can be reached whether the normal approximation to the number of defects is assumed or not. More accurate statistical tests must be based on the actual

defect distribution, which can only be obtained given a much larger sample of wafers.

## VI. CONCLUSION

Several conclusions can be drawn from this analysis of 57 defect maps of large area IC's. First, the large area negative binomial distribution breaks down as the area of the chip increases beyond a certain point and new theoretical models are needed for yield calculations. The medium-sized area negative binomial distribution can be used as a reasonably accurate approximation. Second, we could not determine the existence of any "gross yield" effect. Instead, it is necessary to either "throw away" the chips with the most defects or to find a combination of two probability distributions in order to get a good match with the empirical results.

In addition, statistical tests (both parametric and nonparametric) indicate that there are differences in the number of defects which are the result of the location within the chip. On the other hand, a "radial dependence" within the wafer has not been observed.

Some of the above mentioned conclusions are specific to the IC's that we have analyzed and cannot be generalized to other types of large area IC's. However, the methods and principles which we proposed here are universal and applicable to almost any large area IC. Based on this and previous studies [6] we believe that the defect distribution of most large area IC's has a medium size clustering factor in it. The block size, though, will probably differ from one type of IC to the other, and from one fabrication line to the other. Similarly, the exact nature of the differences among the different areas on the chip in their sensitivity to defects will depend on the specific chip type, fabrication line and wafer run.

## ACKNOWLEDGMENT

The assistance of S. Riley from IBM, Technology Products, who pointed out the existence of systematic defects in the preliminary wafer defect maps, is gratefully acknowledged.

## REFERENCES

- [1] J. A. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Trans. Semiconductor Manufact.*, vol. 3, no. 2, pp. 60-71, May 1990.
- [2] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Brooks and Cole, 1990, 3rd ed.
- [3] J. D. Gibbons, *Nonparametric Statistical Inference*. Marcel Dekker Inc., 1985, 2nd ed.
- [4] I. Koren and C. H. Stapper, "Yield models for defect tolerant VLSI circuits: A review." *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, "Analysis of defect maps of large area VLSI ICs," *1992 IEEE Wkshp. on Defect and Fault Tolerance in VLSI Syst.*, pp. 267-276, 1992.
- [6] I. Koren, Z. Koren, and C. H. Stapper, "A unified negative binomial distribution for yield analysis of defect tolerant circuits," *IEEE Trans. Comput.*, vol. 42, pp. 724-437, June 1993.
- [7] Z. Koren and I. Koren, "A unified approach for yield analysis of defect tolerant circuits," *Defect and Fault Tolerance in VLSI Syst.*, C. H. Stapper, V. K. Jain, and G. Saucier, Eds., vol. 2. New York: Plenum, 1990, pp. 33-45.
- [8] C. H. Stapper, "Small-area fault clusters and fault-tolerance in VLSI circuits," *IBM J. Res. Develop.*, vol. 33, Mar. 1989.



**Israel Koren** (S'72-M'76-SM'87-F'91) received the B.Sc., M.Sc., and D.Sc. degrees from the Technion-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-Israel Institute of Technology. He also held visiting positions with the University of California at Berkeley, University of Southern California, Los Angeles, and University of California, Santa Barbara. He has been a consultant to Intel, Digital Equipment Corp., National Semiconductor, Tolerant Systems, and ELTA—Electronics Industries. His current research interests are fault-tolerant VLSI architectures, models for yield and performance, floor-planning of VLSI chips, and computer arithmetic.

Dr. Koren has published extensively in the IEEE Transactions. He was a Co-Guest Editor for IEEE TRANSACTIONS ON COMPUTERS, Special Issue on High Yield VLSI Systems, April 1989. Since January 1992 he has served on the Editorial Board of these TRANSACTIONS. He also served as Program Committee member for numerous conferences. He has edited and co-authored the book, *Defect and Fault-Tolerance in VLSI Systems*, vol. 1, (Plenum, 1989). He is the author of the textbook *Computer Arithmetic Algorithms*, (Prentice-Hall, 1993).



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

She is currently with the Department of Industrial Engineering, University of Massachusetts, Amherst. Previously she has held positions with the Department of Statistics, University of Haifa, 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.



**Charles H. Stapper** (M'78-SM'82-F'92) received the B.S.E.E. and M.S.E.E. degrees from the Massachusetts Institute of Technology in 1959 and 1960, respectively. He received the Ph.D. degree from The University of Minnesota, Minneapolis, on an IBM resident study fellowship in 1967.

Currently, he is an independent industrial consultant who teaches a graduate course on integrated circuit yield and reliability at the University of Vermont. He retired from IBM where he worked on computer components from 1960 to 1993.

Dr. Stapper is a member of Sigma Xi, an editor for the *Journal of Electronic Testing: Theory and Applications* (JETTA), and a member of the board of directors of the Vermont State Mathematics Coalition. He is the principal author of papers that won the 1989-90 IEEE Solid-State Circuits Council Best Paper Award and the 1991 P. K. McElroy Award for best paper at the 1991 Annual Reliability and Maintainability Symposium.