

ANALYSIS OF DEFECT MAPS OF LARGE AREA VLSI IC'S

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

Department of Electrical and Computer Engineering

*Department of Industrial Engineering and Operations Research

University of Massachusetts, Amherst, MA 01003

**IBM General Technology Division

Essex Junction, VT 05452

Abstract

Defect maps of 57 wafers containing large area VLSI ICs 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 ICs. Even the recently proposed medium size clustering model, although closer to the empirical distribution than other known distributions, is not good enough. To obtain a good match, either a combination of two theoretical distributions or a "censoring" procedure (i.e., ignoring the worst chips) is necessary.

1. Introduction

Defects 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 distribution of defects, to determine a criterion for distinguishing between "low quality" from "high quality" chips, and to determine whether certain areas on the wafer or on the chip are more prone to defects than others.

2. Matching a theoretical distribution

The first objective of the analysis was to match a yield model to the empirical defect distribution. In the past, the large area negative binomial distribution has been successfully used to describe the distribution of defects on a chip, i.e.,

$$\text{Prob}(x \text{ faults in area } A) = \frac{\Gamma(\alpha_A + x)}{x! \Gamma(\alpha_A)} \frac{(\lambda_A/\alpha_A)^x}{(1 + \lambda_A/\alpha_A)^{\alpha_A + x}} \quad (1)$$

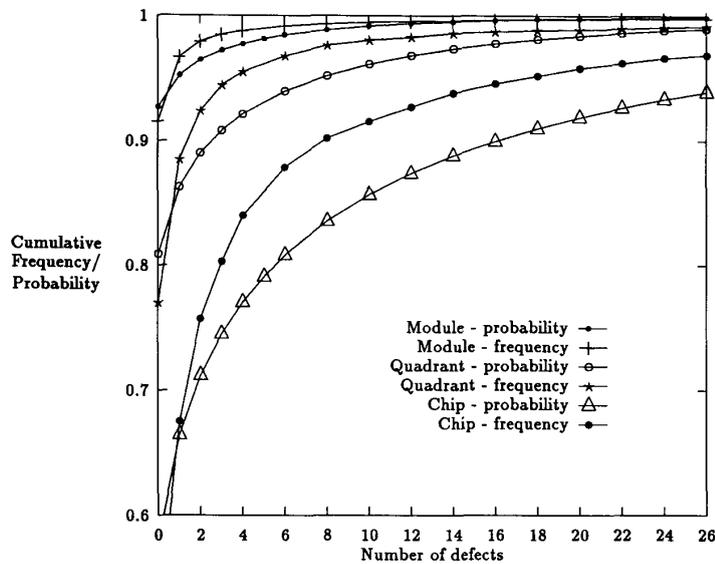


Figure 1: Comparing the empirical cumulative frequency of the number of defects in a module/quadrant/chip to the large area negative binomial model.

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, $k = 0, 1, 2, \dots$, and tried to fit to it a large area negative binomial distribution (with the constraint $\lambda_{chip} = 5.52$ which is the empirical average number of defects per chip in our sample). The cumulative empirical frequency of chips with k defects or less and the cumulative theoretical probability of k defects or less are depicted in Figure 1.

We then divided every chip into 4 equally sized quadrants, and divided each quadrant into 4 equally sized modules. The chip is thus divided into 16 modules. We next compared the empirical frequency of quadrants with k defects and of modules with k defects to the corresponding large area negative binomial probabilities (under the constraints: $\lambda_{quadrant} = 1.38$ and $\lambda_{module} = 0.345$). The cumulative frequencies and probabilities (for k defects or less) appear in Figure 1. The conclusion that can be drawn from Figure 1 is that the large area negative binomial distribution is not suitable for describing the defect distribution of

large area chips. The smaller the area, the better the fit. In this case, the fit is best when the area of a 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 which will be suitable for describing the defects on a chip. Since the chips may have some redundancy incorporated into them, or may be used as partially good chips, we chose the following criterion for goodness of fit of a model to the empirical results:

Denote by $freq(m)$ ($m = 0, \dots, 16$) the relative frequency of chips with m fault-free modules, and by $cfreq(m)$ the cumulative frequency of chips with m fault-free modules or less. Similarly, denote by $prob(m)$ the theoretical probability (according to some given yield model) of a chip with m fault-free modules, and by $cprob(m)$ the cumulative probability of m or less fault-free modules in a chip. We applied the Kolmogorov-Smirnov test for goodness of fit, and accordingly, searched for a yield model in which the maximum absolute difference between $cfreq(m)$ and $cprob(m)$, denoted by dif , does not exceed 0.02.

$$dif = \max_{0 \leq m \leq 16} |cfreq(m) - cprob(m)| \tag{2}$$

We first tried to match the empirical results with a single probability distribution. The candidate distributions were: The Poisson distribution; the small area negative binomial distribution [4]; the medium area negative binomial distribution with two different block sizes, namely, 2×2 and 4×4 [3]; and the large area negative binomial distribution. The parameters of these distributions are summarized in Table 1. The parameters have been

	λ_p	λ_s	λ_m	λ_l	α_s	α_m	α_l
Poisson Distribution	0.345	—	—	—	—	—	—
Large Area Clustering	—	—	—	0.345	—	—	0.123
Small Area Clustering	—	0.345	—	—	0.0200	—	—
Med. Area Clust. - 4×4 block	—	—	0.345	—	—	0.058	—
Med. Area Clust. - 2×2 block	—	—	0.345	—	—	0.032	—
Small Ar. & Large Ar.	—	0.260	—	0.085	0.0014	—	0.360

Table 1: Parameters of theoretical distributions.

selected so as to minimize the maximal difference dif under the constraint $\lambda_{module}=0.345$ (the empirical average).

The empirical and theoretical cumulative distributions are depicted in Figure 2 and in Table 2, and the maximal differences appear in Table 2. As can be seen, none of the five single probability distributions mentioned above had a satisfactory match with the empirical distribution of the number of fault-free modules per chip.

Number of fault-free modules	Empirical frequency	Poisson distribution	Large Area clustering	Small Area clustering	Medium Area clustering 4 x 4 block	Medium Area clustering 2 x 2 block	Combined Small Ar. & Large Ar. clustering
0	0.00000	0.00000	0.02738	0.00000	0.00273	0.00000	0.00002
1	0.00052	0.00000	0.04340	0.00000	0.00415	0.00000	0.00010
2	0.00155	0.00000	0.05723	0.00000	0.00561	0.00001	0.00034
3	0.00232	0.00002	0.07030	0.00000	0.00751	0.00001	0.00088
4	0.00284	0.00020	0.08321	0.00000	0.01789	0.00012	0.00194
5	0.00774	0.00121	0.09629	0.00000	0.02384	0.00026	0.00386
6	0.01213	0.00574	0.10983	0.00000	0.03073	0.00067	0.00708
7	0.01832	0.02141	0.12410	0.00000	0.04352	0.00138	0.01227
8	0.02838	0.06422	0.13938	0.00000	0.06389	0.00591	0.02032
9	0.04102	0.15656	0.15605	0.00001	0.08052	0.01106	0.03253
10	0.05908	0.31345	0.17465	0.00016	0.11042	0.02533	0.05076
11	0.08359	0.52114	0.19595	0.00148	0.13499	0.04493	0.07784
12	0.12590	0.73117	0.22129	0.01067	0.18216	0.12998	0.11883
13	0.18498	0.88802	0.25317	0.05796	0.24138	0.18914	0.18052
14	0.29128	0.96960	0.29752	0.22741	0.31975	0.33120	0.28364
15	0.48942	0.99600	0.37538	0.60519	0.43243	0.47434	0.49351
16	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Maximal difference		0.7030	0.1156	0.1270	0.0569	0.0399	0.0085

Table 2: Comparing the empirical cumulative frequency of fault-free modules to theoretical distributions.

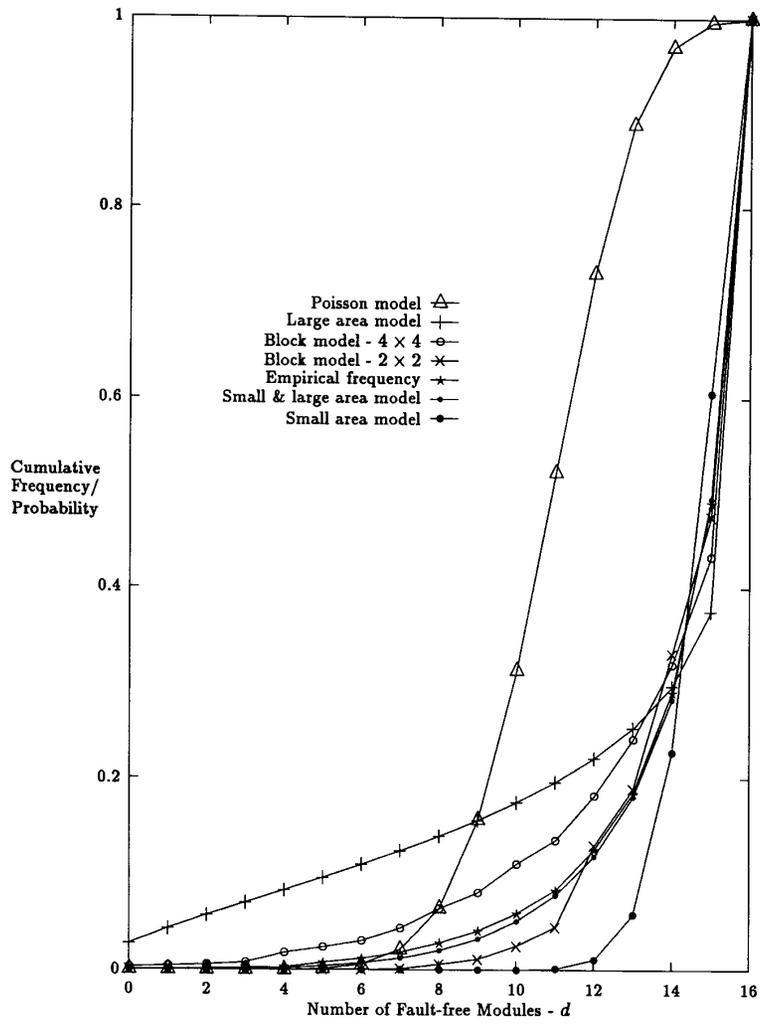


Figure 2: Comparing the empirical cumulative frequency of fault-free modules to theoretical frequency distributions.

Observing the defect maps indicates that there are two sources of defects, systematic and random, or heavily clustered and less heavily clustered. To verify this observation, we calculated the clustering measure α_A for the different sized areas A between 1×1 and 4×4 modules. For a given area size A , α_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 $\lambda_A = 0.345 \cdot A$. A constant value of α_A for every A would indicate large area clustering, while a value linearly increasing with A would indicate small or medium area clustering. The resulting values of α_A 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." This approach did not fit our data since the empirical distribution did not have a mode at 0, probably due to the large area of the chip. As a result, we attempted two other modeling approaches. One was to search for a combination of two defect distributions that will match the empirical distribution, and the other was 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.

3. Combining two probability functions

Denoting by $Z(A)$ the number of defects in area A , we assume that $Z(A) = X_1(A) + X_2(A)$ where $X_1(A)$ and $X_2(A)$ are independent random variables, denoting the number of defects of type 1 and of type 2, respectively, in area A ,

Given the two distributions of X_1 and X_2 , the theoretical probability of m fault-free modules in a chip (denoted by $prob(m)$) can be calculated as follows (we assume that a module has an area of 1),

$$\begin{aligned} prob(m) &= \binom{16}{m} \sum_{j=0}^{16-m} (-1)^j \binom{16-m}{j} P[Z(m+j) = 0] \\ &= \binom{16}{m} \sum_{j=0}^{16-m} (-1)^j \binom{16-m}{j} P[X_1(m+j) = 0] \cdot P[X_2(m+j) = 0] \end{aligned} \quad (3)$$

For the two types of defects we considered all combinations of two out of the following distributions: the Poisson distribution (with the parameter λ_p defects per module), the

large area negative binomial distribution (with the parameters λ_l defects per module and α_l), the medium area negative binomial distribution (with the parameters λ_m defects per module, α_m , and a block size that is equal to the size of the chip - 4×4 modules), and the small area negative binomial distribution (with the parameters λ_s defects per module and α_s).

Combining the Poisson distribution with the small area negative binomial distribution or with the medium area negative binomial distribution did not prove to be close enough to the empirical distribution. The small area negative binomial, on the other hand, provided a very good fit when combined with the large area negative binomial ($dif=0.0085$). The results appear in Figure 2 and in Table 2.

We estimated the parameters for each of the models by performing an extensive search in the parameter space (under the constraint $\lambda_{module} = 0.345$). The resulting parameters appear in Table 1. This method of estimation proved to be very time consuming, especially for the models combining two distributions. 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 1×1 and 4×4 are:

$$Y_{RC} = P[X_1(RC) = 0] \cdot P[X_2(RC) = 0] \quad (4)$$

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

4. 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 "cut-off" number of defects, such that a chip with this number of defects or more is considered of low quality and is ignored.

The candidates for being a cut-off point are: the average number of defects per chip, the 95th percentile (thus throwing away the worst 5% of the chips), or any value in between. Since the median in our sample has been 0, it is not suitable as a cutoff point. The average number of defects per chip has been 5.52, and 6 has, therefore, been selected as the first candidate for a cut-off point. The 95-th percentile is 17, and 18 is another possible cut-off point.

For every cut-off 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 fault-free modules has been calculated ($m = 0, \dots, 16$) and compared to the theoretical probability obtained from three yield models: The large area negative binomial distribution, the negative binomial distribution with a block size of 4×4 (equal to the chip size), and the negative binomial distribution with a block size of 2×2 (a quarter of a chip). The goodness of fit between the theoretical and empirical distribution has been measured by *dif* defined in equation (2). Figure 3 depicts the maximal differences between the cumulative frequency and probability, for $6 \leq C \leq 18$ and the three theoretical distributions. As can be seen from Figure 3, the best fit has been obtained for the 2×2 block, for all cut-off points. Table 3 shows the implications of choosing a specific cut-off point, namely, the percentage

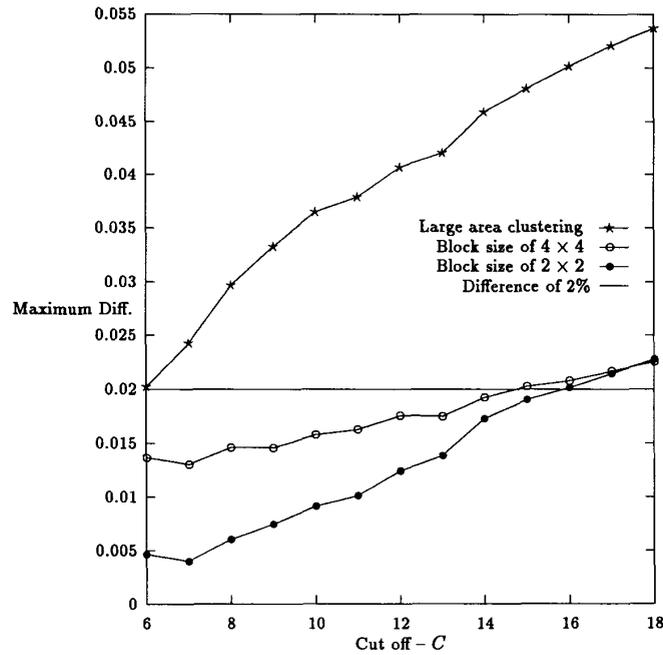


Figure 3: Maximum difference between theoretical models and experimental results.

of chips that will be thrown away, the yield of the remaining chips, and the parameters of the theoretical yield model of the remaining chips. The latter can be used to calculate the

yield of the remaining chips in case they have some incorporated redundancy, or can be used as partially good chips.

Cut-off value	% thrown away chips	Remaining yield	λ	α
6	0.1393	0.5932	0.8336	0.624
7	0.1215	0.5812	0.9383	0.552
8	0.1055	0.5708	1.0467	0.488
9	0.0978	0.5659	1.1064	0.448
10	0.0898	0.5609	1.1757	0.424
11	0.0844	0.5576	1.2280	0.408
12	0.0766	0.5529	1.3099	0.388
13	0.0728	0.5506	1.3545	0.376
14	0.0663	0.5468	1.4349	0.356
15	0.0617	0.5441	1.4971	0.344
16	0.0580	0.5420	1.5489	0.336
17	0.0542	0.5398	1.6080	0.324
18	0.0508	0.5379	1.6624	0.316

Table 3: Implications of different cut-off values.

5. Dependence on location

We proceeded to determine whether any part of the chip or any specific area on the wafer is more prone to defects than the rest. Specifically, we compared the top half and the bottom half of the chip, the right half and the left half of the chip, the center and the boundaries of the chip, and the center and the boundaries of the wafer.

To this end, we used statistical tests which are based on the Gaussian approximation and use averages and standard deviations of the number of defects per quadrant. The results are summarized in Table 4. We concluded that the number of defects on the boundary of a wafer is significantly larger than that in the inside area, and a similar conclusion can be reached for a chip. In addition, the number of defects in the top half of the chip was found to be significantly larger than that of the bottom half, while there is no significant difference between the right half and the left half. More exact statistical tests which will take into account the specific defect distribution need to be developed. The dependence of the number of defects on the location may require the modification of the yield models used.

Location	Average number of defects per quadrant	Standard Deviation	Statistical significance
Wafer boundary	2.154	24.60	
Wafer center	1.142	6.64	significant
Chip boundary	1.420	23.79	
Chip center	0.740	10.24	significant
Upper half of chip	1.860	17.82	
Lower half of chip	0.900	5.87	significant
Left half of chip	1.310	13.07	not
Right half of chip	1.450	13.48	significant

Table 4: Dependence of the number of defects on location.

6. Conclusions

Several conclusions can be drawn from this analysis of 57 defect maps of large area ICs. First, the large area negative binomial distribution breaks down as the area of the chips increases beyond a certain point and new theoretical models are needed for yield calculations. 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, there seem to be differences in the number of defects which are the result of the location, within the chip and within the wafer. More accurate statistical tests need to be developed to verify these differences.

Acknowledgment: This work was supported in part by IBM under contract 81775.

References

- [1] I. Koren and C.H. Stapper, "Yield Models for Defect Tolerant VLSI Circuit: A Review," *Defect and Fault Tolerance in VLSI Systems*, I. Koren (ed.), pp. 1-21, Plenum, 1989.
- [2] Z. Koren and I. Koren, "A Unified Approach for Yield Analysis of Defect Tolerant Circuits," *Defect and Fault Tolerance in VLSI Systems*, Vol. 2, C.H. Stapper, V.K. Jain and G. Saucier (eds.), pp. 33-45, Plenum, 1990.
- [3] I. Koren, Z. Koren and C.H. Stapper, "A Unified Negative Binomial Distribution for Yield Analysis of Defect Tolerant Circuits," to appear, *IEEE Trans. on Computers*, Vol. 41, 1992.
- [4] C.H. Stapper, "Small-Area Fault Clusters and Fault-Tolerance in VLSI Circuits," *IBM J. Res. Develop.*, Vol. 33, March 1989.