EMPLOYING THE UNIFIED NEGATIVE BINOMIAL DISTRIBUTION FOR YIELD ANALYSIS OF EMPIRICAL DATA [1]

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

University of Massachusetts at Amherst
Department of Electrical and Computer Engineering
*Department of Industrial Engineering
Amherst, MA 01003

****IBM General Technology Division**
Essex Junction, VT 05452

ABSTRACT

The previously proposed unified negative binomial distribution for projecting the yield of defect tolerant integrated circuits is analyzed in this paper. The proposed model is first extended to a more general case. Then, the derived yield expression is validated through simulation. Finally, the new model is employed in the analysis of empirical data from twelve defect maps.

1. INTRODUCTION

The negative binomial distribution (with the two parameters λ and α) has been recognized by many researchers as adequately describing the distribution of defects in integrated circuits (e.g., [1]), since this distribution can model the clustering of defects which has been observed in practice. However, for actual yield calculations (especially when redundancy is involved), further assumptions such as large area clustering or small area clustering have been used [3].

In an earlier paper [2] we suggested the addition of a third parameter, namely the *block size,* to the existing two parameters of the negative binomial distribution. The purpose of this parameter is to unify the existing yield calculations as well as to add

[1]This work was supported in part by NSF under contract MIP–8805586.

a wide range of models (medium size clustering) that have not been dealt with before. The block size has been defined as the smallest area on the chip for which the defects in two adjacent areas are statistically independent. In this earlier paper we presented the formulas for calculating the yield of a chip as a function of the block size under the simplifying assumption that both the chip and the block are linear arrangements of modules. We showed that as the block size increases, the chip yield decreases, and the optimal redundancy (i.e., the redundancy which maximizes the yield per chip area) increases. Several methods of estimating the block size have been suggested as well.

In this paper we first extend the yield calculation to the two-dimensional case. We then validate the analytical model using both simulation and some actual wafer maps which have been analyzed using this model. Section 2 presents the yield equations and the parameter estimation. The model validation appears in Section 3 and Section 4 concludes the paper.

2. YIELD MODEL

This section generalizes the results in [2] to the more realistic case, where the chip and the block are two-dimensional. For completeness, we repeat here the model description and the main assumptions. The problem we deal with is that of calculating the yield of a chip consisting of N modules, out of which M are necessary for proper operation and $N - M$ are spares. The yield of such a chip is defined as the probability that at least M of the modules are defect-free. We use as our fault model the commonly used negative binomial distribution, i.e., the probability of x defects in an area of size A is

$$Prob\ (x\ \ defects) = \frac{\Gamma(\alpha_A + x)}{x!\ \Gamma(\alpha_A)}\ \frac{(\lambda_A/\alpha_A)^x}{(1 + \lambda_A/\alpha_A)^{\alpha_A + x}} \tag{1}$$

where λ_A and α_A are parameters.

Since λ_A is the average number of defects in an area of size A, it must be proportional to A and can be calculated by

$$\lambda_A = A \times \lambda$$

where λ is the average number of defects in a unit area. The parameter α_A is a measure of the size of the defect clusters and its dependence on the area A is not uniquely

determined. Two extreme cases were dealt with in the literature: One is large area clustering, where α_A is fixed for all A's,

$$\alpha_A = \alpha .$$

The second is small area clustering, for which α_A is linear in A, i.e.,

$$\alpha_A = A \times \alpha$$

where α denotes the clustering parameter corresponding to a unit area.

In order to unify the yield equations resulting from these two models and to add a large number of "medium size clustering" models, the addition of a third parameter, namely, the block size B, has been suggested in [2]. The block size is defined as the smallest area on the chip for which the defects in two adjacent areas are statistically independent, and inside which the defects are uniformly distributed. Intuitively, the blocks form a division of the wafer into sub-areas in such a way that distinct defect clusters are contained in distinct blocks.

We show next how the yield of a chip with redundancy can be calculated, assuming that the defects on the wafer have a negative binomial distribution with a block size B. For mathematical tractability we make the assumption that both the chip and the block are shaped as rectangles with an integer number of modules. $B_1 \times B_2 = B$ modules for the block and $C_1 \times C_2 = N$ modules for the chip. We choose as the unit area the area of a module, and as the basic parameters the module parameters λ_{mod} and α_{mod}. For the sake of simplicity we denote these by λ and α, respectively. Consequently, the block parameters are $\lambda_{block} = B_1 \cdot B_2 \cdot \lambda$ and $\alpha_{block} = \alpha$.

As was proven in [2], for an area consisting of D modules which is totally contained within one block, the probability of exactly k defect-free modules can be calculated using "large area clustering" as follows

$$a(k, D) = \binom{D}{k} \sum_{i=0}^{D-k} (-1)^i \binom{D-k}{i} \left(1 + \frac{(k+i)\lambda}{\alpha}\right)^{-\alpha} . \qquad (2)$$

Unless the block size is very large, the chip area will be divided into several sub-areas, each contained in a different block, and the number of defect-free modules in the entire chip can be described as the sum of the numbers of defect-free modules in

each of the parts. Since these sub-areas are in different blocks, they are assumed to be statistically independent with respect to the number of defects. The probability of k defect-free modules in the entire chip can, therefore, be calculated either by using the generating function technique or as a convolution. The details of the generating function technique are given in [2]. Since the convolution technique is computationally simpler, it is employed in this paper.

There are $min(B_1, C_1) \times min(B_2, C_2)$ possible placements of the chip relative to the block. Denote a placement by (R_1, R_2) where $1 \leq R_1 \leq min(B_1, C_1)$ and $1 \leq R_2 \leq min(B_2, C_2)$. The placement (R_1, R_2) determines the way the chip is divided into complete and partial blocks. For given values of R_1 and R_2 denote:

$$n_1 = \left[\frac{C_1 - R_1}{B_1}\right]$$

$$m_1 = (C_1 - R_1) \bmod B_1$$

$$n_2 = \left[\frac{C_2 - R_2}{B_2}\right]$$

$$m_2 = (C_2 - R_2) \bmod B_2$$

Note that $C_1 = R_1 + n_1 \times B_1 + m_1$, $C_2 = R_2 + n_2 \times B_2 + m_2$, and that for $C_1 \leq B_1$, $n_1 = 0$ and $m_1 = C_1 - R_1$. Similarly, for $C_2 \leq B_2$, $n_2 = 0$ and $m_2 = C_2 - R_2$.

Once R_1 and R_2 are determined, the chip is divided into (at most) 9 disjoint sub-areas in the following manner (see Figure 1):

(1) 1 (possibly) partial $R_1 \times R_2$ block.
(2) n_2 (possibly) partial $R_1 \times B_2$ blocks.
(3) 1 (possibly) partial $R_1 \times m_2$ block.
(4) n_1 (possibly) partial $B_1 \times R_2$ blocks.
(5) $n_1 \times n_2$ complete $B_1 \times B_2$ blocks.
(6) n_1 (possibly) partial $B_1 \times m_2$ blocks.
(7) 1 (possibly) partial $m_1 \times R_2$ block.
(8) n_2 (possibly) partial $m_1 \times B_2$ blocks.
(9) 1 (possibly) partial $m_1 \times m_2$ block.

For $C_1 \leq B_1$ and $C_2 \leq B_2$, only 4 sub-areas, namely, areas 1,3,7 and 9 exist.

To calculate the yield of a chip with redundancy, we need to find the probabilities $P(J = k)$, where J denotes the number of defect-free modules in a chip. These prob-

abilities are first calculated for a given placement (R_1, R_2), and then averaged over all possible placements.

Denoting by J_i the number of defect-free modules in sub-area i, $(i = 1, ..., 9)$, we have

$$J = \sum_{i=1}^{9} J_i \,.$$

Since the J_i's are statistically independent, the probability function of J is the convolution of the probability functions of $J_1, ..., J_9$, namely,

$$P^{(R_1,R_2)}(J = k) = \sum_{k_1+\cdots+k_9=k} P(J_1 = k_1)P(J_2 = k_2)\cdots P(J_9 = k_9)\,. \tag{3}$$

The superscript (R_1, R_2) indicates the dependence of the probabilities on the placement. For simplicity, it is omitted in the notation $P(J_i = k_i)$, although these probabilities clearly depend on (R_1, R_2).

To calculate the probability functions of J_i (for given (R_1, R_2)), we need to distinguish between $i = 1, 3, 7, 9$ and $i = 2, 4, 5, 6, 8$. For $i = 1, 3, 7, 9$ the sub-area is contained in one block, and

$$P(J_i = k_i) = a(k_i, D_i) \tag{4}$$

where $a(k, D)$ is defined as in equation (2) and D_i is the number of modules in sub-area i, namely,

$$D_1 = R_1R_2, \quad D_3 = R_1m_2, \quad D_7 = m_1R_2, \quad D_9 = m_1m_2\,.$$

For $i = 2, 4, 5, 6, 8$, sub-area i is itself divided into several parts, each contained in a different block. These parts have equal dimensions and are statistically independent. Denote by s_i the number of those parts, by D_i the number of modules in each of them and by k_{ij} the number of defect-free modules in part j of sub-area i $(j = 1, ..., s_i,$ $i = 2, 4, 5, 6, 8)$, then,

$$P(J_i = k_i) = \sum_{k_{i1}+k_{i2}+\ldots+k_{is_i}=k_i} a(k_{i1}, D_i)\; a(k_{i2}, D_i)\cdots a(k_{is_i}, D_i) \tag{5}$$

where $s_2 = n_2$, $D_2 = R_1B_2$, $s_4 = n_1$, $D_4 = B_1R_2$, $s_5 = n_1n_2$, $D_5 = B_1B_2$, $s_6 = n_1$, $D_6 = B_1m_2$, $s_8 = n_2$ and $D_8 = m_1B_2$.

Equations (4) and (5) are now substituted into (3), and (3) is averaged over all possible placements (R_1, R_2),

$$P(J=k) = \frac{1}{min(B_1,C_1) \times min(B_2,C_2)} \sum_{R_1=1}^{min(B_1,C_1)} \sum_{R_2=1}^{min(B_2,C_2)} P^{(R_1,R_2)}(J=k). \quad (6)$$

The yield of the chip can now be calculated using

$$Y(Chip) = P(J \geq M) = \sum_{k=M}^{N} P(J=k). \quad (7)$$

The calculation of the yield according to equation (7) requires the knowledge of the parameters λ, α, B_1 and B_2, and the correct estimation of these parameters is crucial to a precise yield prediction. Estimating λ and α is discussed in the literature (e.g., [5]). λ can be estimated from the average number of defects per module. Given W wafers with S modules each, let X_i denote the number of defects in module i $(i = 1, ..., S \times W)$, then,

$$\widehat{\lambda} = \frac{1}{S \times W} \sum_{i=1}^{S \times W} X_i \quad (8)$$

where $\widehat{\lambda}$ denotes the estimate of the parameter λ. The estimation of α is slightly more complicated since it depends on the assumed block size. For a given block size (B_1, B_2), $\alpha(B_1, B_2)$ can be estimated as follows. Let Y_{B_1,B_2} denote the fraction of defect-free $(B_1 \times B_2)$ blocks out of all given $\frac{S \times W}{B_1 \times B_2}$ blocks of this size. Then, $\widehat{\alpha}(B_1, B_2)$ is the solution of the equation

$$Y_{B_1,B_2} = \left(1 + \frac{\widehat{\lambda}}{\widehat{\alpha}(B_1, B_2)}\right)^{-\widehat{\alpha}(B_1,B_2)}. \quad (9)$$

We suggest two methods of estimating the block size (B_1, B_2), one is based on the values obtained for $\widehat{\alpha}(B_1, B_2)$, and the other is based on the Chi-square test for statistical independence.

The first method utilizes the fact (proven in [2]) that the parameter α remains constant within a block and increases when the area consists of several blocks. In this method, $\alpha(i, j)$ is estimated for every potential block size (i, j) and the values of $\widehat{\alpha}(i, j)$

are arranged in matrix form. We then search for the largest (i,j) for which $\widehat{\alpha}(i,j)$ is still close to $\widehat{\alpha}(1,1)$, and this (i,j) is used as an estimate of (B_1, B_2).

The second method is based on the assumption that the different blocks are statistically independent with respect to the number of defects. The block size in this method is determined in two steps, first B_1 and then B_2. For every potential block size (i,j), each wafer is divided into blocks of size (i,j). To determine the value of B_1, a Chi-square statistic is calculated, for every (i,j), to test independence between every two vertically neighboring blocks. The resulting values are then arranged in matrix form. The index i of the first row for which the Chi-square values are significantly lower than those of the other rows is chosen as B_1. Similarly, B_2 is determined by testing for independence between horizontally neighboring blocks, and by choosing the index j of the first column whose values are significantly low. The two methods for estimating (B_1, B_2) are demonstrated in the next section.

3. MODEL VALIDATION

This section is devoted to validating the above formulas, using simulated wafers first and then empirical data obtained from twelve defect maps of wafers manufactured by IBM [4].

As a first step, $10,000$ wafers of size 24×24 modules each were simulated, using the parameters: $\lambda = 0.1$, $\alpha = 0.25$, and a block size of 2×3 modules. The parameters λ, α, B_1 and B_2 were then estimated based on the simulated wafers.

The estimate obtained for λ was $\widehat{\lambda} = 0.1007$. The matrix of the α estimates for block sizes between (1×1) and (8×8) is given in Table 1. The two Chi-square tests (one for rows and one for columns) have been performed for the same block sizes, and the resulting statistics are presented in Tables 2 and 3. The block size (B_1, B_2) can be found either from the α matrix or from the Chi-square matrices. Observing the α matrix (Table 1), it is easily seen that the value 0.26 in the $(2,3)$ position is the farthest entry which is close to 0.28. We therefore deduce that the block size is (2×3). The same conclusion is reached by observing the Chi-square matrices. In the rows matrix (Table 2), line 2 is the first to have very small numbers (and so does every row whose index is a multiple of 2). In the columns matrix (Table 3), column 3 is the first with significantly low values (so is column 6, since 6 is a multiple of 3). The block size is,

therefore, estimated as $(2, 3)$.

The estimated parameters were then used for yield calculations. A chip size of 4×6 modules was selected, and the yield of this chip with d spare modules was calculated in four different ways. This yield is the cumulative probability of d or less defective modules, or the probability of $N - d$ or more defect-free modules, denoted by $P(J \geq N - d)$ in the previous section. First, we found the actual proportion of simulated 4×6 chips with d or less defective modules ($d = 0, ..., 24$). The theoretical probability was then calculated using three yield models: The large area clustering model, the small area clustering model, and the equations obtained from the block model. The results are depicted in Figure 2. As expected, the block model provides the best fit in this case. In addition, we can see that for $d \geq 2$, the large area clustering model underestimates the yield, while the small area clustering model overestimates it. For very small values of d, three of the graphs almost coincide but the small area clustering model underestimates the yield.

As a second step for validating the proposed yield equations, we analyzed 12 wafer maps manufactured by IBM [4], each consisting of 24×24 modules. We first estimated λ and obtained $\widehat{\lambda} = 0.1089$. We then estimated α for every possible block size between (1×1) and (12×12), and the results are presented in Table 4. (Block sizes larger than (12×12) have not been considered.) The Chi-square tests for independence were then performed, and the resulting statistics appear in Tables 5 and 6. The determination of the block size based on Tables 4, 5 and 6 is not as straightforward here as it was in the case of the simulated wafers. The empirical data include only 12 wafers, which is a very small number for statistical purposes. We, therefore, have to combine both methods of the block size estimation and consider all three tables simultaneously. By combining the information in Tables 4, 5 and 6 we estimated the block size (B_1, B_2) to be (10×8).

We then proceeded to compare the empirical and theoretical yield of a chip of size (10×11) modules. The yield of this chip (as a function of the number of spare modules, denoted by d) was calculated using both the large area clustering model and the block model with a (10×8) block. The results were then compared to the empirical proportion of chips with d or less defects in the actual wafer maps, and are depicted in Figure 3. As can be seen, the large area clustering model is more accurate for very small values of d. For $d \geq 2$, however, the block model with a (10×8) block provides

a much better fit to the empirical results.

To determine the sensitivity of the yield estimation to the exact choice of the block size, three other block sizes have been chosen: (10×7), (10×9) and (10×6), and the estimated yield of a (10×11) chip was calculated for each one of them. The results appear in Figure 3, and they demonstrate that if the deviation from the "correct" block size is small, so is the deviation of the predicted yield from the empirical results. This deviation increases as the error in estimating the appropriate block size increases.

4. CONCLUSION

The recently proposed generalization of the well-known negative binomial yield model has been analyzed in this paper. It has been demonstrated (using simulation and empirical data) that in certain situations the more general model provides a more accurate yield projection compared to the previously suggested models. Additional analysis of empirical data needs to be performed to gain better understanding of the circumstances under which the more general model has to be employed.

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," to appear, *Defect and Fault Tolerance in VLSI Systems*, Vol. 2, C.H. Stapper, V.K. Jain and G. Saucier (eds.), Plenum, 1990.

[3] C.H. Stapper, "Small-Area Fault Clusters and Fault-Tolerance in VLSI Circuits," *IBM J. Res. Develop.*, Vol. 33, March 1989.

[4] C.H. Stapper, "Correlation Analysis of Particle Clusters on Integrated Circuit Wafers," *IBM J. Res. Develop.*, Vol. 31, pp. 641-650, Nov. 1987.

[5] V. Foard Flack, "Estimating Variations in IC Yield Estimates," *IEEE J. of Solid-State Circuits*, Vol. SC-21, pp. 362-365, April 1986.

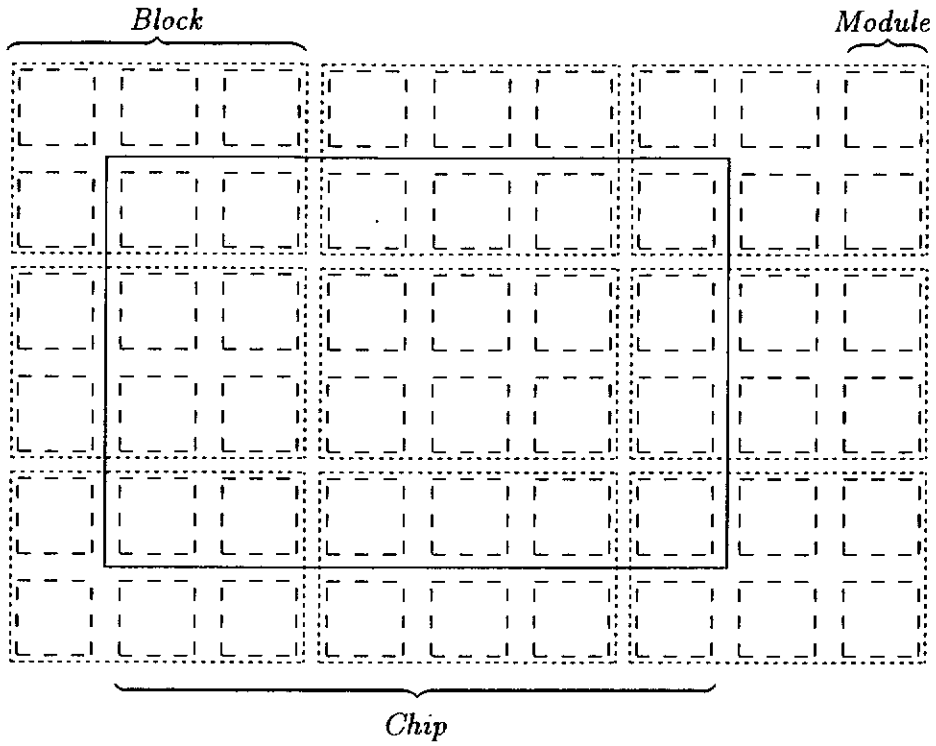

Figure 1: A placement of a 4×6 chip relative to 2×3 blocks, $R_1 = 1, R_2 = 2$.

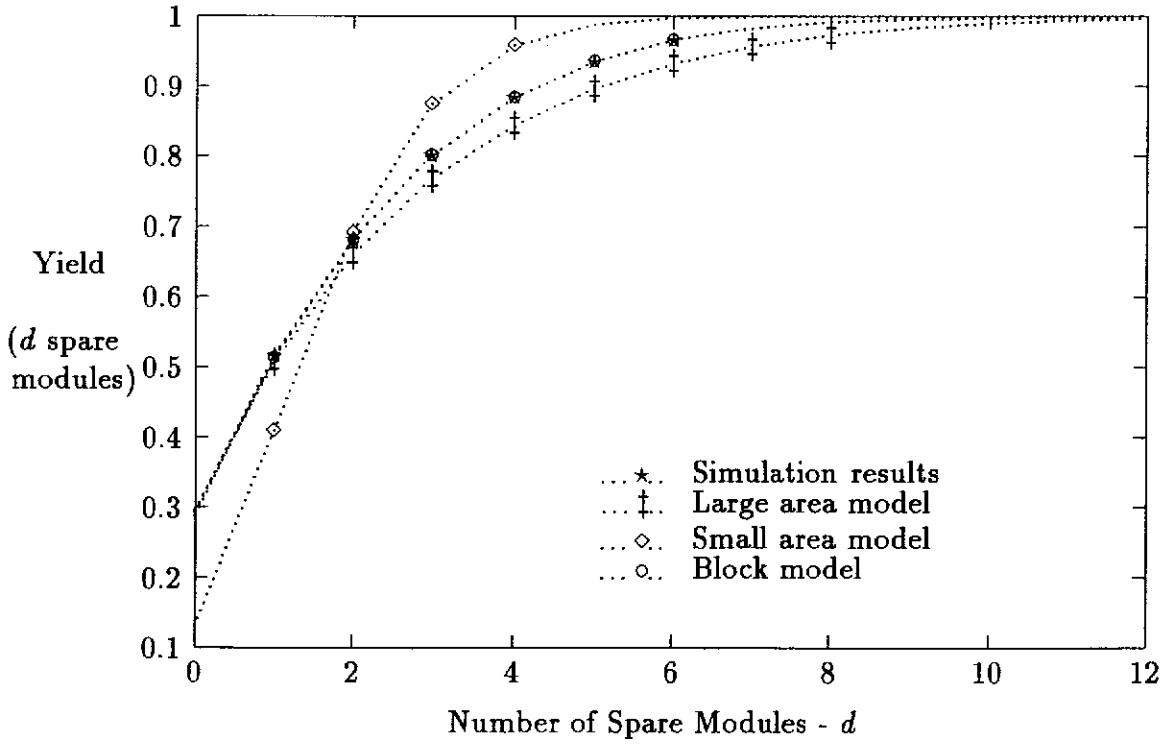

Figure 2: Comparing three theoretical yield models to simulation results.

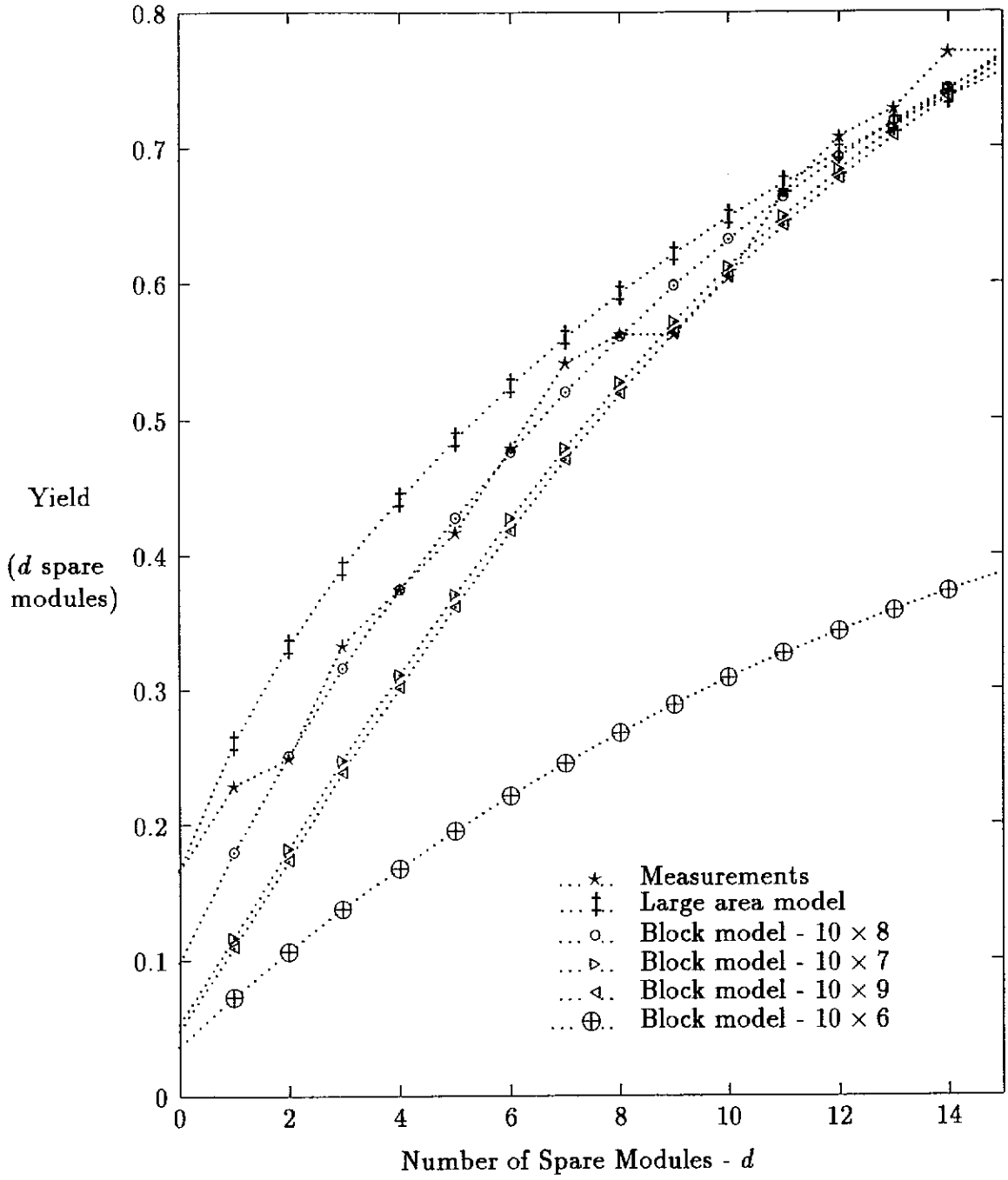

Figure 3: Comparing the block model (with four different sizes of the block) and the large area clustering model to the measurments for a 10×11 chip.

0.28	0.33	0.26	0.46	0.50	0.50	0.67	0.75
0.27	0.32	0.26	0.46	0.51	0.50	0.68	0.76
0.48	0.56	0.48	0.83	0.96	0.93	1.28	1.44
0.50	0.60	0.50	0.89	1.03	0.98	1.37	1.53
0.68	0.85	0.70	1.27	1.46	1.41	1.94	2.19
0.72	0.90	0.73	1.34	1.54	1.48	2.06	2.32
0.93	1.16	0.95	1.72	1.98	1.91	2.65	2.97
0.98	1.21	0.98	1.79	2.06	1.98	2.75	3.02

Table 1: The α matrix for the simulated wafers.

229683.34	203881.48	245000.22	136178.55	96182.90	115760.10	71963.05	71715.63
0.28	10.27	1.15	3.07	3.10	5.14	3.21	7.33
8050.98	7101.01	8798.31	4260.57	2917.43	3280.47	1825.31	1715.88
1.88	10.22	5.87	1.94	2.66	2.67	1.81	6.53
1234.57	1049.30	1191.06	550.32	302.30	347.95	158.85	171.62
4.44	10.70	0.71	0.78	2.76	1.14	0.43	5.22
300.75	289.79	278.46	111.66	73.70	72.53	40.11	25.26
5.97	10.63	4.01	0.45	1.03	2.63	5.61	3.56

Table 2: The Chi-Square matrix for the rows of the simulated defect maps.

100686.63	36679.00	0.07	4294.11	993.03	2.85	933.73	697.99
113273.83	35649.67	0.56	4179.13	938.29	1.41	865.55	620.23
58799.13	18425.02	1.43	1926.59	416.69	0.68	354.57	262.97
54168.29	16258.17	3.63	1691.40	359.87	1.06	244.62	150.38
30668.13	8739.49	1.22	836.14	181.54	2.22	99.11	66.78
34574.99	9604.63	0.65	880.34	125.98	3.37	84.79	55.49
22257.25	5940.09	2.30	544.87	104.04	5.95	70.72	32.58
24383.17	6195.51	3.18	518.31	78.54	2.67	40.53	23.38

Table 3: The Chi-Square matrix for the columns of the simulated defect maps.

1.10	0.56	0.52	0.40	0.91	0.49	0.57	0.48	0.72	0.74	0.56	0.48
0.53	0.44	0.44	0.35	0.61	0.46	0.55	0.47	0.66	0.67	0.59	0.48
0.48	0.44	0.44	0.36	0.56	0.46	0.48	0.43	0.60	0.60	0.54	0.45
0.41	0.42	0.42	0.36	0.54	0.49	0.52	0.43	0.62	0.65	0.59	0.49
0.38	0.39	0.38	0.33	0.44	0.43	0.48	0.39	0.64	0.57	0.49	0.45
0.46	0.45	0.47	0.42	0.64	0.53	0.57	0.51	0.70	0.69	0.62	0.53
0.46	0.48	0.48	0.38	0.57	0.53	0.58	0.47	0.79	0.73	0.65	0.54
0.44	0.47	0.48	0.43	0.65	0.55	0.58	0.50	0.72	0.80	0.70	0.58
0.43	0.43	0.44	0.37	0.53	0.45	0.49	0.46	0.68	0.64	0.55	0.49
0.45	0.47	0.49	0.41	0.54	0.51	0.56	0.44	0.71	0.68	0.58	0.56
0.48	0.51	0.51	0.42	0.62	0.53	0.61	0.47	0.77	0.72	0.62	0.60
0.52	0.51	0.52	0.45	0.72	0.59	0.62	0.51	0.83	0.91	0.87	0.72

Table 4: The α matrix for the twelve defect maps.

347.86	393.96	317.19	271.23	181.67	206.04	126.49	138.21	86.28	89.99	108.44	124.8
359.84	225.44	177.73	143.39	96.48	112.24	84.16	92.02	59.58	53.39	58.25	54.18
171.28	136.12	108.27	82.38	49.94	72.63	36.31	45.43	31.22	28.56	34.22	33.35
142.00	95.66	72.70	53.55	30.72	41.25	26.25	31.16	22.00	20.61	18.33	19.86
63.86	56.39	32.61	28.52	16.01	18.75	14.95	21.06	10.09	7.94	10.49	9.80
92.98	67.18	44.02	45.46	23.85	28.68	21.04	29.27	10.00	5.69	8.23	7.98
31.13	18.91	12.05	8.30	4.96	11.48	10.61	11.52	4.87	4.42	4.35	4.83
32.15	12.70	6.87	8.58	2.55	4.29	4.63	6.09	1.88	4.13	4.53	1.43
21.74	7.86	7.90	4.55	0.52	13.13	6.60	3.24	3.78	4.56	7.00	5.37
20.20	11.42	7.73	6.05	3.37	5.67	16.38	2.56	0.96	24.77	11.22	5.35
25.77	13.75	10.49	7.73	3.27	11.46	7.70	3.67	0.96	10.62	10.89	5.67
37.03	10.59	10.40	9.80	5.48	12.95	7.20	6.57	0.49	10.24	3.14	7.53

Table 5: The Chi-Square matrix for the rows of the defect maps.

682.08	448.34	209.07	100.24	123.47	137.51	41.95	67.89	47.57	48.19	36.93	35.79
449.74	359.78	129.08	52.37	57.48	87.67	32.70	31.36	24.53	30.38	24.94	15.34
276.16	262.93	106.83	50.36	45.50	70.77	18.72	20.26	20.94	25.59	21.05	17.06
202.53	220.85	78.51	40.58	33.66	54.05	16.97	14.12	17.63	22.21	17.38	13.58
168.06	170.58	67.68	36.02	24.14	51.42	19.39	18.95	25.86	20.97	15.70	13.47
184.45	143.17	70.18	33.07	28.80	40.49	12.16	11.12	14.60	14.01	12.83	7.98
141.92	149.14	56.52	21.81	19.59	30.76	10.89	13.60	15.47	15.24	12.48	7.29
126.23	125.89	35.34	23.05	22.66	29.23	14.54	13.19	18.83	16.85	13.93	9.43
123.96	85.27	37.17	18.22	17.61	20.58	7.09	9.58	15.87	15.87	15.01	12.15
95.49	82.81	32.48	22.22	22.48	19.18	10.14	9.29	16.66	16.72	11.99	12.17
99.84	97.34	43.10	19.97	10.56	16.31	6.69	8.61	13.13	11.63	7.74	7.16
104.05	80.36	32.66	20.99	7.20	14.89	3.73	5.27	2.69	5.19	7.53	3.14

Table 6: The Chi-Square matrix for the columns of the defect maps.