Reducing the Complexity of Defect Level Modeling using the Clustering Effect

José T. de Sousa

IST/INESC
Technical University of Lisbon
R. Alves Redol, 1000 Lisboa, Portugal
jose.desousa@inesc.pt

Vishwani D. Agrawal

Bell Labs
Lucent Technologies
Murray Hill, NJ 07974, USA
va@research.bell-labs.com

Keywords: *defect clustering, defect level, fault clustering, fault coverage, reject ratio*

Abstract

Accounting for the clustering effect is fundamental to increase the accuracy of Defect Level (DL) modeling. This result has long been known in yield modeling but, as far as known, only one DL model directly accounts for it. In this paper we improve this model, reducing its number of parameters from three to two by noticing that multiple faults caused by a single defect can also be modeled as additional clustering. Our result is supported by test data from a real production line.

Nomenclature

α: defect clustering parameter.

λ: average number of defects per chip.

Ω: defect coverage.

DL: defect level or reject ratio.

R_i: number of new chips rejected by vector i.

T: single fault coverage.

Y: true yield.

Y_a: apparent yield.

Y_m: measured yield.

1 Introduction

Defect level, also known as *reject ratio*, is the fraction of faulty chips among the chips that pass the production test. These chips are taken as good devices and shipped to customers. They can later fail in the field, causing the manufacturer significant expense and adversely influencing important invisible parameters such as customer satisfaction, company prestige, etc. The economical importance of defect level hardly needs to be highlighted. The problem is how to predict and control its value.

The causes of field rejects are faults caused by manufacturing defects; the same defects that are responsible for reducing the yield. To cope with the complexities of physical defect phenomena, sophisticated yield models have been developed. It is important that these models account for the defect clustering phenomenon, for significantly more accurate yield estimates can be produced by doing so.

Despite the relevance of the clustering effect, it is rarely considered in defect level modeling. An earlier attempt to model clustering that uses the average number of faults in a faulty chip was made in [1]. About the same time, a theory of defect clustering using a negative binomial formulation for the number of defects in a chip was also developed in the context of yield modeling [2]. The negative binomial formulation was adapted to the testable-fault level modeling in the model proposed in [3], which provides the starting point for the present work.

The assumptions in this work are the same as in current yield theories — defects of random size and location, governed by specific probability distributions. Unlike the model in [3], we dispense with the parameter that models the number faults per defect, and let the remaining two parameters, the average number

0-7695-0537-6/00 $10.00 ©2000 IEEE

of defects per chip and clustering parameter, represent the tester yield. We observe that if one defect produces multiple faults, that simply corresponds to a higher degree of clustering. Consequently, the new DL model is simpler, and is able to produce a better fit to test data.

This paper is organized as follows. In Section 2 we discuss existing DL models, and provide the motivation for the new model. In Section 3 the new model is introduced and analyzed. In Section 4 an optimization procedure to find the parameters of the models is explained. Experimental results are reported in Section 5. Section 6 concludes the paper and gives directions for future developments.

2 Background

DL is the probability that a chip is faulty given it passed the test. This can be written as

$$DL = 1 - \frac{P(\text{chip good})}{P(\text{chip passed the test})} \tag{1}$$

The probability that a chip is good is the true yield Y. The chip is good if the *defect coverage* Ω of the test applied to it is $\Omega = 1$. Defect coverage is the ratio between the tested chip area and the total chip area. However, defect coverage is difficult to evaluate; what is normally used, is the *single fault coverage* T, which is defined as the fraction of faults from a given single fault model (e.g., the single stuck-at model) that can be detected by a given test set.

In defect level modeling one fundamental assumption must be made: that there exists a transformation between T and Ω, with the restriction that $T = 1$ implies $\Omega = 1$. We recognize that 100% fault coverage may not mean 100% *defect coverage*. Research on fault modeling should continue to ensure that this assumption is accurate. Nevertheless, this work is not on fault modeling and we must assume that satisfactory fault models are being used. If our fault models are not good enough, and we are getting too many field rejects, it is possible that the real value of DL is higher than what is predicted by the methodology described in this paper.

The probability that a chip passes the test for a given value of T will be denoted $Y_a(T)$ and called the *apparent yield for fault coverage* T. Therefore, from Equation (1), the definition of DL as a function of the fault coverage T becomes

$$DL = 1 - \frac{Y}{Y_a(T)} \tag{2}$$

In a lot containing N chips, if R_i new chips are rejected after application of test vector i, which corresponds to a fault coverage T, then a sample value $Y_m(T)$ for $Y_a(T)$ can be experimentally evaluated by

$$Y_m(T) = \frac{N - \sum_{i=1}^{n} R_i}{N} \tag{3}$$

To find a model for DL we need to find a model for $Y_a(T)$. We will require that a good $Y_a(T)$ model is such that it can fit the experimental points $Y_m(T)$ as accurately as possible. Note that N must be large so that the sample values become close to the real population values. Once a good $Y_a(T)$ model is found we can predict yield as $Y = Y_a(1)$ and DL by Equation (2). For T close to 1, experimentally observed $Y_a(T)$ is approximately linear. Thus, the derivative of $Y_a(T)$ at $T = 1$ is proportional to DL.

If, hypothetically, the relationship between Ω and T was known, we would be able to talk about the apparent yield $Y_a(\Omega)$ for defect coverage Ω. The key to any $Y_a(\Omega)$ model is the distribution $P(r)$ for the number r of defects in a chip, which we will call the *underlying defect model*:

$$Y_a(\Omega) = \sum_{r=0}^{\infty} (1-\Omega)^r P(r) \tag{4}$$

Note that the $(1-\Omega)^r$ is the probability that none of the r defects is detected. No transformation $\Omega \rightarrow T$ is being considered as yet.

With a set of assumptions equivalent to considering a Poisson distribution for the number of defects in a chip, Williams and Brown [5] derived a DL model, which implied the following apparent yield model

$$Y_a(\Omega) = e^{-\lambda\Omega} \tag{5}$$

where λ is the average number of defects per chip. Attempts to confirm the Williams-Brown model experimentally have not been very successful for two possible reasons. First, $\Omega = T$ is often assumed, while using the stuck-at fault model to determine T. Second, the model totally disregards defect clustering, which causes the distribution of the number of defects to be negative binomial rather than Poisson.

Considering a defect model with two parameters, the yield Y and the average number n_0 of faults (assumed a Poisson random variable) in a faulty chip, Agrawal, Seth and Agrawal [1] proposed the model:

$$Y_a(\Omega) = Y + (1-\Omega)(1-Y)e^{-(n_0-1)\Omega} \tag{6}$$

The fact that n_0 can be greater than 1 is a possible way of modeling the clustering effect. As a result Equation (6) proved much more accurate than

the Williams-Brown formula. Even assuming $\Omega = T$, specific relationships between T and Ω are implicitly accounted for when clustering is modeled. This idea is central to this paper and will be fully exploited in the new model. However, as we will show, this model has some difficulties in fitting the experimental data, which indicates that its underlying defect model may be inadequate.

De Sousa et al. [6] proposed a DL model based on the Williams-Brown formula of Equation (5), where Ω is substituted by the following function of T:

$$\Omega = 1 - (1 - T)^R \tag{7}$$

where R is parameter that was called the fault-to-defect susceptibility ratio. This model recognized that fault coverage must be mapped to defect coverage before it can be used in the Williams-Brown formulation. Compared to the Williams-Brown model it produces a substantially lower fitting error. However, its fatal flaw is that it predicts a zero DL derivative at $T = 1$, and consequently an erroneously low DL. This also disagrees with experimental data which tend to have a non zero derivative, even at high fault coverage T.

The clustering effect was finally given its due importance by Seth and Agrawal in their negative binomial model [3]. This model is derived from a formulation that also enables characterizing the yield equation using wafer test data. The resulting model is:

$$Y_a(T) = \left[1 + \frac{\lambda}{\alpha}\frac{(1 - e^{-cT})}{(1 - e^{-c})}\right]^{-\alpha} \tag{8}$$

where λ and α are the average number of defects in the chip and the defect clustering parameter, respectively, and c is the average number of faults per defect, assumed to be Poisson distributed. The three parameters are positive real numbers. Note that, according to the negative binomial formulation, $\alpha \to 0$ corresponds to strong clustering and $\alpha \to \infty$ corresponds to weak clustering. Also note that T, and not Ω, appears in the model as the parameter c takes care of mapping defects to faults.

Compared to the other models, this model produces the best fit to experimental data. However, modeling the clustering effect by means of the parameter α and again the number of logical faults per defect with c seems somewhat redundant. From a functional point of view it does not matter if multiple faults are being produced by a single defect or by a cluster of defects. This suggests that we do not need two clustering parameter. In other words, the questions we attempt to answer in the present work are: Can we use a single clustering formalism to also encapsulate multiple faults produced by a single defect? Could the overall fitting error be maintained or decreased if we did that? Also, if we can produce a model with two parameters this would show that the three parameters in the Seth-Agrawal model are possibly correlated. We find that the Seth-Agrawal model is undecisive about the relative values of the defect density and the faults per defect. Besides, nonlinear curve fitting is a computationally hard problem that explodes with the number of parameters. Hence, it is important to keep the number of parameters as low as possible.

3 The new defect level model

This section introduces the new DL model. Its key feature is the realization that defects causing multiple faults can be modeled by simply using the clustering effect. In this way, we do not model the number of faults produced by a single defect. We substitute each defect with a cluster of as many defects as the faults it produces. We assume a negative binomial distribution for $P(r)$ with parameters λ and α. Compared to the Seth-Agrawal model we expect a higher average number of defects per chip λ and a stronger clustering parameter α, since we no longer have the the average number of faults per defect c. We start by deriving the new model following the steps of [3].

The probability generating function (p.g.f.) $G(s)$ of a probability distribution $P(r)$ is defined as

$$G(s) = \sum_{r=0}^{\infty} s^r P(r) \tag{9}$$

The p.g.f. of the negative binomial distribution is known to be

$$G(s) = \left[1 + \frac{\lambda}{\alpha}(1 - s)\right]^{-\alpha} \tag{10}$$

We can safely substitute Ω with T since we assumed a one to one correspondence between faults and defects. Also, we assumed that $P(r)$ is a negative binomial distribution with parameters λ and α. Thus Equation (4) can be rewritten as

$$Y_a(T) = G(1 - T) = \left(1 + \frac{\lambda}{\alpha}T\right)^{-\alpha} \tag{11}$$

Hence, we have obtained a two parameter model, which incorporates both clustering and an implicit mapping of defects to faults, obtained by artificially assuming a one-to-one correspondence between faults and defects. As any apparent yield model, $Y_a(0) = 1$.

For $T = 1$ we get the negative binomial yield formula [2]:

$$Y = \left(1 + \frac{\lambda}{\alpha}\right)^{-\alpha} \quad (12)$$

Also, Equation (11) converges to Equation (5) as $\alpha \to \infty$, i.e., for low clustering it becomes the Williams-Brown model provided we replace T back to Ω again.

4 Optimization procedure

Having presented the most representative Y_a models, we must now address the problem of finding their parameters that best fit the given experimental data.

Our data consists of the results of applying a set of test vectors to a set of fabricated chips. For each applied test vector i we have the number of rejected chips R_i, and the cumulative stuck-at fault coverage T_i. We will use the stuck-at fault model to determine our fault coverage T. Again, it is not the objective of this paper to question the accuracy of the stuck-at fault model. The methodology is applicable to any fault model that is deemed appropriate. The experimental measured yield $Y_m(T)$ can be computed by Equation (3). Then, given a $Y_a(T)$ model we will determine its parameters so that the least square error with respect to the experimental points $Y_m(T)$ is minimized. Since, all our $Y_a(T)$ models are nonlinear, we have a nonlinear fitting problem. It is well known that the solution space for a nonlinear fit is far from being smooth, with multiple local minima. This forces us to use global search in a subspace where we know the solution exists. Lower complexity methods such as the steepest descent are not acceptable. Also, we are limited by a finite resolution when varying the parameters. We first consider the subproblem of finding a good starting point for the parameters and then the subproblem of optimizing from that point.

To determine starting points for the parameters we will use the new model as an example. The derivative of $Y_a(T)$ at $T = 0$ is $-\lambda$, according to Equation (11). Thus, by roughly measuring this derivative from the data we can determine a starting point λ_0 for λ. We can use the number of chips that pass all vectors and divide it by the total number of chips N to get a rough yield estimate Y_0. Using Y_0 and λ_0 we solve the nonlinear Equation (12) numerically and we get a starting point α_0 for α. Next, given this starting point (λ_0, α_0), we choose a vicinity beyond which local minima do not occur (e.g., $\alpha_0 \pm 1$ and $\lambda_0 \pm 2$, and fine enough resolution (e.g., 0.05 for both parameters). Then, for every point (λ, α) in the vicinity, we compute the fitting error which is the summation of the square error

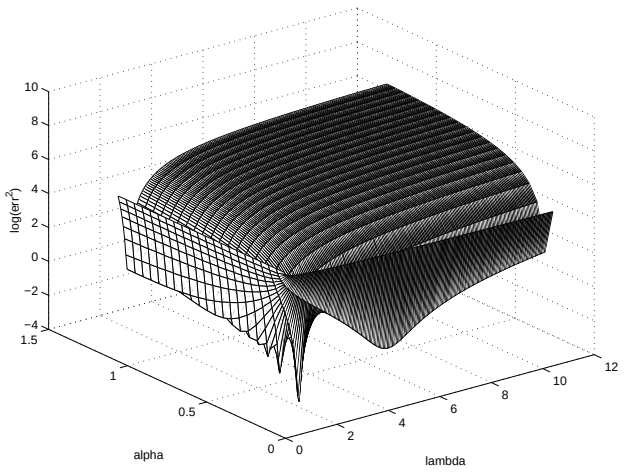


Figure 1: Logarithm of the squared fitting error as a function of α and λ for SEMATECH test data.

for every data point. The point with the least fitting error determines the best values for λ and α. To illustrate the procedure, we plot the square fitting error as function of λ and α for the new model in Figure 1.

The complexity of this method is related to the size of the grid that represents the solution subspace. Considering a square grid with $n \times n$ points the complexity of the algorithm is $O(n^2)$. Clearly, one more parameter, such as the parameter c in the Seth-Agrawal method, will increase the complexity to $O(n^3)$.

5 Experimental results

Our new model has been tested using the well known data provided by the SEMATECH consortium, also used in [4]. The contents of the data is as described in the previous section. Results were obtained for the 5 models featured in this paper: Williams-Brown, Agrawal et al., Seth-Agrawal, de Sousa et al., and the new de Sousa-Agrawal model. The optimization procedure was run to determine the parameters of the models. In Table 1 we give the minimized error and the estimates for Y and DL. We plot the yield $Y_m(T)$ data points in Figure 2, where the five fitted models are superimposed. The models can be differentiated by matching the corresponding yields from Table 1 with the value at $T = 1$. Williams-Brown model (dotted straight line) has an yield of 73.48%. Next higher yield of 76.15% for de Sousa-Agrawal (solid curve) almost matches the experimental data. Seth-Agrawal (dotted curve) has the yield of 76.45%. Agrawal et al. (dot-dash curve) and de Sousa et al. (dashed curve) have slightly above 77% yields.

As we can see the Williams-Brown model with just

Table 1: Yield and DL from SEMATECH test data

Model	Fit Error	Y(%)	DL(ppm)
Williams-Brown [5]	4.93	73.48	1900
de Sousa et al. [6]	0.34	77.11	3
Agrawal et al. [1]	0.25	77.01	245
Seth-Agrawal [3]	0.10	76.45	488
de Sousa-Agrawal	0.06	76.15	615

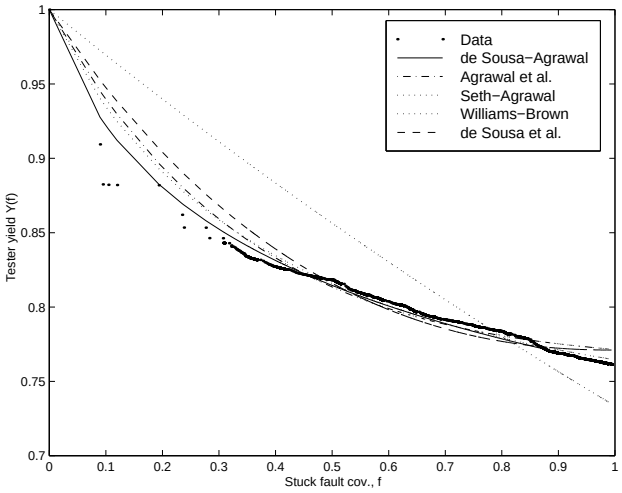


Figure 2: Fitting models to SEMATECH test data.

one parameter cannot follow the data well. Its fitting error is the largest, and its predictions of Y and DL are the least accurate. The accuracy of the Y and DL predictions depend on the ability to follow the data at T close to $T = 1$. The de Sousa et al. model produces a much better fit, but the its zero derivative at $T = 1$ kills the accuracy of the Y and DL predictions. The three models that account for clustering are the ones that are able to produce the best fits and the most accurate predictions. The Agrawal et al. model is the worst of the three, and we suppose that this is because its underlying Poisson defect model is not as good as the negative binomial formulation used in the other two. Finally, we can see that the new model performs better than the Seth-Agrawal model, both in terms of fitting error and accuracy of the Y and DL predictions. The logical explanation for this must be that modeling multiple faults via clustering is better than modeling the number of faults per defect with Poisson distribution. The optimizations took about a few seconds, a few minutes and several minutes for 1, 2 and 3 parameters, respectively, using a MATLAB script, running on a Linux PC.

6 Conclusion

We have modeled the VLSI tester yield with two parameters, the average number of faults per chip (λ) and a fault clustering parameter (α). This model is simpler than some of the previous models and yet closely represents the behavior of experimental data. As the results show, we can conveniently analyze the actual tester data to estimate the two parameters and predict the yield and defect level (DL). For the recent SEMATECH data, we determine a yield of 76.15% and a DL of 615 parts per million. A strong assumption in this work, which is also implicit in many previous analyses, is that DL drops to zero when the stuck-at fault coverage is 100%. To be precise, we should mean the coverage of the "realistic" (i.e., those that actually occur) faults. The justification for this assumption lies in the fact that the measured tester yield depends on the capability of the tests for detecting the actual, so-called realistic, faults. This argument needs further investigation. The paper only presents data on one chip and the applicability of the new model should be examined for a wide variety of chips.

References

[1] V. D. Agrawal, S. C. Seth, and P. Agrawal. "Fault Coverage Requirement in Production Testing of LSI Circuits". *IEEE Journal of Solid State Circuits*, SC-17(1):57–61, Feb. 1982.

[2] C. H. Stapper, F. Armstrong, and K. Saji. "Integrated Circuit Yield Statistics". *Proc. IEEE*, 71:453–470, Mar. 1983.

[3] S. C. Seth and V. D. Agrawal. "Characterizing the LSI Yield Equation from Wafer Test Data". *IEEE Trans. on CAD*, CAD-3(2):123–126, April 1984.

[4] P. Nigh, W. Needham, K. Butler, P. Maxwell, and R. Aitken. "An Experimental Study Comparing the Relative Effectiveness of Functional, Scan, IDDQ and Delay-fault Testing". In *Proc. of IEEE 15th VLSI Test Symp. (VTS)*, pages 459–464, 1997.

[5] T. W. Williams and N. C. Brown. "Defect Level as a Function of Fault Coverage". *IEEE Transactions on Computers*, C-30(12):987–988, Dec. 1981.

[6] J. T. deSousa, F. M. Goncalves, J. P. Teixeira, and T.W. Williams. "Fault Modeling and Defect Level Projections in Digital IC's". In *Proc. European Design and Test Conf. (EDTC)*, pages 436–442, March 1994.