On Defect Level Estimation and the Clustering Effect

José T. de Sousa
INESC/IST
Lisbon, Portugal *
jts@hobbes.inesc.pt

Abstract This paper presents a review of existing defect level models and introduces a new defect level model that accounts for the fault clustering effect. The model uses generalized negative binomial statistics to model the probability distribution of the number of faults in a chip. This analysis shows that clustering, in addition to naturally increasing the yield, also raises the detection probability and therefore lowers the defect level. By accounting for clustering, the new model predicts a less stringent fault coverage requirement than other models.

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

Nomenclature

α: fault clustering parameter.

α_d: defect clustering parameter.

λ: average number of faults per chip.

λ_d: average number of defects per chip.

Ω: fault coverage.

Ω_{max}: maximum attainable fault coverage.

$\bar{D}$: fault average area density.

* Work done while at Bell Labs, Murray Hill, NJ

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35498-9_57

L. M. Silveira et al. (eds.), *VLSI: Systems on a Chip*
© IFIP International Federation for Information Processing 2000

DL: defect level.

R: realistic to stuck-at fault detectability ratio.

T: stuck-at fault coverage.

Y: true yield.

Y_m: measured yield.

n_0: average number of faults in a faulty circuit.

r: number of faults in a circuit.

1. INTRODUCTION

Defect level (DL) is the fraction of faulty chips among the chips that passed production test. These chips are taken as good devices and shipped as so. Later they are likely to fail in the field, causing the manufacturer to incur in significant expenses. There are also important invisible costs 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 cause of field rejects are faults caused by manufacturing defects; the same defects that are responsible for yield loss. To cope with the complexities of physical defect phenomena, sophisticated yield models have been developed. These models take into account the non-equiprobability of physical defects, by considering a weighted variety of possible causes for yield loss. They also account for the defect clustering phenomenon, which produces significantly more accurate yield estimates than theories that assume defects that are probabilistically independent.

Despite the relevance of the clustering effect, many defect level models do not account for it [Williams and Brown, 1981, Agrawal et al., 1982, Sousa et al., 1996], but there are a few models that do [Seth and Agrawal, 1984, Singh and Krishna, 1996]. In [Seth and Agrawal, 1984] the clustering effect is implicitly taken into account by using the negative binomial distribution for the number of defects in a chip. In [Singh and Krishna, 1996] defect clustering is exploited to identify dice with different DL values in the same wafer. These dice are then placed in different bins, according to their quality level (the inverse of defect level). However, in none of these theories the overall effect of clustering on defect level has been investigated. Since it is well known how defect clustering affects the yield of integrated circuits, we now ask the question of how it affects DL. This is the problem addressed in this paper.

The assumptions of this work are the same as in current yield theories — defects of random size and location, governed by specific probability distributions. However, we directly model the probability that a chip will contain r

faults, rather than the probability that it will contain r defects. We observed that if one defect can produce multiple faults, this simply corresponds to a higher degree of fault clustering. In this way, there is no need to introduce a relationship between faults and defects as in [Seth and Agrawal, 1984]. Consequently, the new DL model is simpler, using one parameter less than the models in [Seth and Agrawal, 1984].

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. Section 4. concludes the paper and gives directions for future developments.

2. BACKGROUND

Assuming that faults in a circuit are probabilistically *independent* and have the same occurrence probability, Williams and Brown [Williams and Brown, 1981], in their seminal work, derived the following DL model:

$$DL = 1 - Y^{(1-\Omega)}. \tag{1.1}$$

This model constituted the first attempt to show how DL depends on the yield Y and the fault coverage Ω. Equating Ω to the stuck-at fault coverage T, causes the Williams-Brown model to spread generalized panic: unless 100% or very high fault coverage is obtained, the defect level will be unacceptable. For example, suppose $Y = 80\%$ and we want $DL = 100$ parts per million (p.p.m.). According to the Williams-Brown formula the fault coverage requirement is $\Omega = 99.55\%$. The belief that the stuck-at fault coverage T actually represents the real fault coverage Ω led test engineers to demand 100% stuck-at fault coverage no matter the cost. Paradoxically, the very same engineers would rest completely assured if 100% stuck-at fault coverage had in fact been achieved. In fact, reality is somewhat different: 100% stuck-at fault coverage may not be needed in practice; on the other hand, 100% stuck-at fault coverage does not prevent some other faults from escaping the test.

Despite its illustrative power, the Williams-Brown model has very low accuracy when used with the stuck-at fault coverage T. Some studies of defect level data from real manufacturing processes have demonstrated this fact [Maxwell and Aitken, 1991]. The most striking difference between the $DL(T)$ curves obtained with real data and the $DL(T)$ curves obtained with the Williams-Brown model is the type of curvature exhibited. Real $DL(T)$ plots show concave curvature (positive second derivative), whereas the Williams-Brown model shows convex curvature (negative second derivative).

It is also striking that Y, the yield parameter that appears in Equation (1.1) is usually computed under completely different assumptions. Y is obtained assuming that the number of faults in a chip has a generalized negative binomial

distribution [Stapper et al., 1983]. This distribution is given by the following equation:

$$P(r) = \frac{\Gamma(\alpha + r)}{r!\,\Gamma(\alpha)} \frac{(\frac{\lambda}{\alpha})^r}{(1 + \frac{\lambda}{\alpha})^{r+\alpha}}. \tag{1.2}$$

where λ is the average number of faults per chip, and α is the fault clustering parameter. The parameter λ is given by

$$\lambda = A\bar{D}, \tag{1.3}$$

where A is the chip area and $\bar{D}$ the average fault density. The clustering effect varies inversely with α: α close to zero indicates strong clustering, whereas α large indicates weak clustering. Equation (1.2) is defined for any positive real values of α and λ and any integer $r = 0,\ 1,\ 2,\ \ldots,\ \infty$. In reality, the number of faults in a chip is large but finite. Nevertheless, $P(r)$ being defined in the range $r = 0,\ 1,\ 2,\ \ldots,\ \infty$ is accurate, since $P(r)$ decreases fast with r. The yield Y is obviously given by $Y = P(0)$, which produces the well known negative binomial yield formula:

$$Y = \left(1 + \frac{\lambda}{\alpha}\right)^{-\alpha}. \tag{1.4}$$

The formula above accounts for fault interdependence due to clustering. In contrast, the Williams-Brown DL model assumes fault independence.

The assumptions of the Williams-Brown model imply that the number of faults r in a chip follows a Poisson distribution rather than a negative binomial distribution:

$$P(r) = \frac{\lambda^r}{r!} e^{-\lambda}. \tag{1.5}$$

For this distribution the yield $Y = P(0)$ is given by

$$Y = e^{-\lambda}. \tag{1.6}$$

The result above can also be obtained using Equation (1.4) and supposing weak clustering (large α). In fact, using Stirling's formula, the limit of Equation (1.4) when $\alpha \rightarrow \infty$ results in Equation (1.6):

$$\lim_{\alpha \rightarrow \infty} (1 + \frac{\lambda}{\alpha})^{-\alpha} = e^{-\lambda}. \tag{1.7}$$

Substituting Equation (1.6) in Equation (1.1), we obtain an expression for the Williams-Brown model which is preferred in this work:

$$DL = 1 - e^{-\lambda(1-\Omega)} \tag{1.8}$$

In [Sousa et al., 1996] it is suggested that the reason why the Williams-Brown formula combined with the stuck-at fault coverage T cannot track experimental fallout data is the fact that, unlike T, in the real fault coverage Ω each fault should be weighted with its probability of occurrence. In [Sousa et al., 1994] it is shown by extensive simulation of more accurate fault models that different weighting of the faults produce different non-linear relationships between Ω and T, which would explain the real shape of $DL(T)$ curves. The clustering effect is briefly mentioned in [Sousa et al., 1996] but its fundamental importance is not realized.

Agrawal et al. [Agrawal et al., 1982] supposed that the number of faults *in a faulty circuit* is Poisson distributed, with average $n_0 \geq 1$. This assumption produced the following model:

$$DL = \frac{(1-\Omega)(1-Y)e^{-(n_0-1)\Omega}}{Y + (1-\Omega)(1-Y)e^{-(n_0-1)\Omega}}. \tag{1.9}$$

This model provided a good fit to experimental DL data [Maxwell and Aitken, 1991] using the stuck-at fault coverage T as the fault coverage Ω. The value of n_0 can be determined using, for example, a least squares fitting method. The model realistically reproduces the concave curvature of the $DL(T)$ curve.

We regard the assumption that faults in a faulty circuit are Poisson distributed as a first attempt to incorporate the clustering effect on DL models. In fact, a higher value for n_0 merely indicates a higher propensity for multiple faults, which is basically what clustering is. As discussed for the Williams-Brown model, the assumptions underlying the Agrawal et al. model are not consistent with the assumptions for deriving Y, a parameter used in the model.

Seth and Agrawal proposed a defect level model based on negative binomial statistics [Seth and Agrawal, 1984]. This model is derived from a formulation that also enables characterizing the yield equation using wafer test data. The resulting model is the following:

$$DL = 1 - \left[\frac{\alpha_d + \lambda_d(1-e^{-c})}{\alpha_d + \lambda_d(1-e^{-c\Omega})}\right]^{-\alpha_d}, \tag{1.10}$$

where λ_d and α_d are the defect density and defect clustering parameter, respectively, and c is the average number of faults per defect, assumed to be Poisson distributed.

This model also reproduces the concave curvature of the $DL(T)$ curve in a realistic manner. Defect clustering is incorporated in the Seth-Agrawal model by means of the parameter α_d. However, the need to model the occurrence of defects and, separately, the relation between logical faults and defects is questionable. Note that the existence of more parameters than needed in a

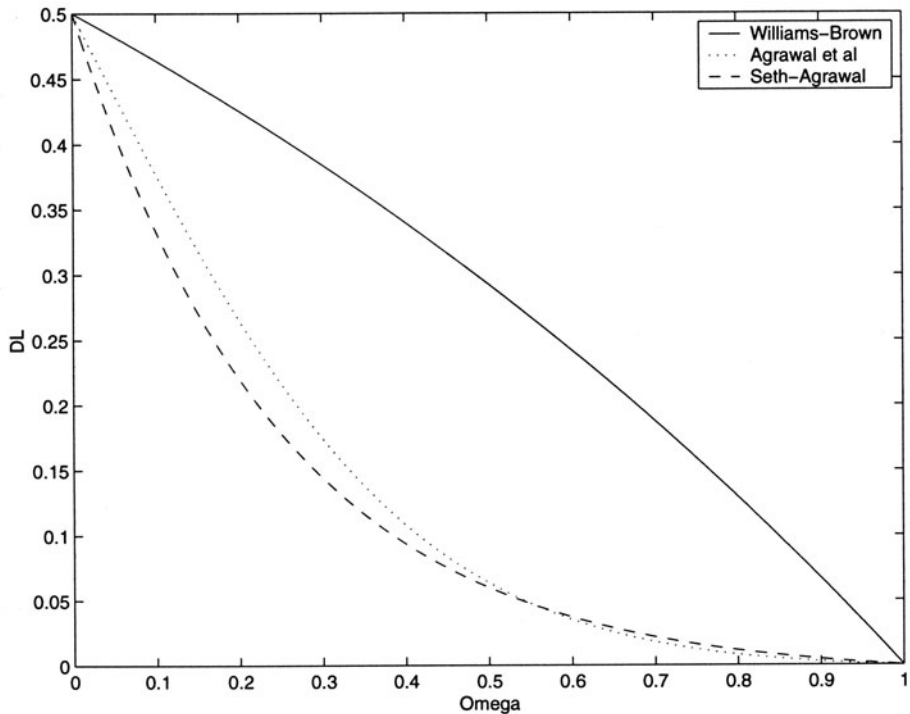

Figure 1 Comparison of DL models.

model leads to lack of decisiveness. That is, a good fit occurs for a family of values of the parameters, instead of for just one combination of values. This means that the parameters are related in a particular way; any combination of values that respects this relation will do. Also, albeit the clustering effect is accounted for, the effect of varying its intensity has not been studied.

The three models discussed are compared in Figure 1, which shows DL as a function of the fault coverage Ω, for a hypothetical circuit for which $Y = 0.5$ (note that $Y = 1 - DL$ at $\Omega = 0$). Since this paper presents a theoretical study, we have no need to relate Ω to a practical measure of fault coverage, such as the stuck-at fault coverage. All models are analyzed assuming the availability of a realistic fault coverage figure Ω.

For the Agrawal et al. model we used $n_0 = 5$. For the Agrawal-Seth model we chose the parameters $\alpha = 1$, $\lambda = 1.0187$ and $c = 4$. The values of the parameters were exaggerated to clarify the points made before, and illustrate what the models are capable of. The parameters are not chosen to track any particular data set.

While the other two models have second derivatives which can be tuned by their parameters, the Williams-Brown model exhibits negative second deriva-

tive for all values of Ω:

$$\frac{d^2DL}{d\Omega^2} = -\lambda^2 e^{-\lambda(1-\Omega)}. \tag{1.11}$$

Another interesting study is that of the first derivative at $\Omega = 1$. The following three equations give the first derivative at $\Omega = 1$ for the three models, in the same order they were presented:

$$\left.\frac{dDL}{d\Omega}\right|_{\Omega=1} = -\lambda \tag{1.12}$$

$$\left.\frac{dDL}{d\Omega}\right|_{\Omega=1} = -\frac{1-Y}{Y}e^{-(n_0-1)} \tag{1.13}$$

$$\left.\frac{dDL}{d\Omega}\right|_{\Omega=1} = -\frac{\alpha\lambda c e^{-c}}{\alpha + \lambda(1-e^{-c})} \tag{1.14}$$

The Williams-Brown model has only one parameter (λ) to match the slope at $\Omega = 1$, as well as the whole data set. The slope at $\Omega = 1$ is very important for computing DL at high fault coverage using a linear approximation. The other two models have more flexibility, since they have two and three parameters, respectively.

3. THE NEW DEFECT LEVEL MODEL

This section introduces the new DL model. The question of whether the clustering effect significantly affects DL is examined thoroughly. We start by deriving the model from a definition of DL, and then we present an analysis of the new model.

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

$$DL = P(\text{chip faulty} \mid \text{chip passed the test}), \tag{1.15}$$

which is equivalent to

$$DL = 1 - P(\text{chip good} \mid \text{chip passed the test}). \tag{1.16}$$

Applying Bayes' formula we can write

$$DL = 1 - \frac{P(\text{chip good AND chip passed the test})}{P(\text{chip passed the test})}. \tag{1.17}$$

Since all good chips pass the test

$$DL = 1 - \frac{P(\text{chip good})}{P(\text{chip passed the test})}. \tag{1.18}$$

The probability that a chip is good is the yield Y. The probability that a chip passes the test can be measured by counting such chips and dividing that number by the total number of chips. This quantity is obviously equivalent to the measured yield. Moreover, if the chips are counted after the application of each test vector, and if the fault coverage after application of each test vector is known, we can estimate the measured yield as a function of the fault coverage and denote this quantity $Y_m(\Omega)$. Therefore, the definition of DL as a function of the fault coverage Ω is

$$DL = 1 - \frac{Y}{Y_m(\Omega)}. \tag{1.19}$$

At this point we need to assume a distribution for the number of faults in a chip. In [Seth and Agrawal, 1984], a distribution for the number of defects (not faults) is considered first, and then another distribution for the number of faults per defect is postulated. In our method we directly consider faults, and assume that the number of faults has a distribution $P(r)$. Thus, we can obtain Y and $Y_m(\Omega)$ respectively by

$$Y = P(0), \tag{1.20}$$

$$Y_m(\Omega) = \sum_{r=0}^{\infty}(1-\Omega)^r P(r). \tag{1.21}$$

Note that the $(1-\Omega)^r$ is the probability that none of the r faults is detected.

Like in [Seth and Agrawal, 1984], we make use of the probability generating function (p.g.f.) method. The p.g.f. $G(s)$ of a probability distribution $P(r)$ is defined as

$$G(s) = \sum_{r=0}^{\infty} s^r P(r). \tag{1.22}$$

Since our objective is to study the effect of fault clustering, we will assume that $P(r)$ is a negative binomial distribution with parameters λ and α. The p.g.f. of the negative binomial distribution is known to be

$$G(s) = \left[1 + \frac{\lambda}{\alpha}(1-s)\right]^{-\alpha} \tag{1.23}$$

Thus, Equation (1.21) can be rewritten

$$Y_m(\Omega) = G(1-\Omega) = \left(1 + \frac{\lambda}{\alpha}\Omega\right)^{-\alpha}. \tag{1.24}$$

Comparing the equation above with the yield expression given by Equation (1.4), and bearing in mind the relationship between λ and the chip area A

as given by Equation (1.3), we conclude that Equation (1.24) gives the yield $Y_m(\Omega)$ of a chip whose area is $A\Omega$. This physical meaning agrees with that of expected measured yield as given before: a chip passes the test if it does not contain any faults in its tested area.

Replacing Equations (1.4) and (1.24) in Equation (1.19) results in the new defect level model, which can be written as

$$DL = 1 - \left(\frac{\alpha + \lambda}{\alpha + \lambda\Omega}\right)^{-\alpha}. \quad (1.25)$$

As a DL model, the new model gives $DL = 1 - Y$ at $\Omega = 0$ and $DL = 0$ at $\Omega = 1$. Moreover, it is interesting to note that

$$\lim_{\alpha \to \infty} 1 - \left(\frac{\alpha + \lambda}{\alpha + \lambda\Omega}\right)^{-\alpha} = 1 - e^{-\lambda(1-\Omega)}. \quad (1.26)$$

That is, as clustering weakens, the new DL model given by Equation (1.25) becomes equivalent to Williams-Brown model as given by Equation (1.8).

The new model resembles the Seth-Agrawal model. However, by directly considering faults, it does not need to model the number of logical faults caused by each defect. It is assumed that the fact that some defects may cause multiple faults is another form of fault clustering, which is subsumed by the clustering parameter α. In fact, it is possible to show that increasing parameter c (average number of faults per defect) in the Seth-Agrawal model has a similar effect on the DL curve as that of increasing the defect clustering parameter α_d. In our model, to study the effect of clustering we just need to vary the parameter α. That is done is Figure 2 for three values of α. It can be seen that as clustering increases (α decreases) the yield $Y = 1 - DL(0)$ increases and DL decreases for any fault coverage. It can also be seen that a low enough α can realistically reproduce the concave curvature of the DL curve. The curve for $\alpha = 0.1$ has positive second derivative.

The most important benefit of modeling clustering is the fact that we obtain a much more accurate fault coverage requirement, which is much easier to meet compared to the fault coverage required by the Williams-Brown model. To study the effect of clustering on fault coverage requirement we will use *test transparency* [McCluskey and Buelow, 1988] instead of fault coverage. The test transparency TT is defined by

$$TT = 1 - \Omega. \quad (1.27)$$

The maximum allowable test transparency TT_{max} is a better measure for the test effort because it tells us what is the fraction of the chip that we may afford to leave untested. Using a linear approximation in the neighborhood of $\Omega = 1$,

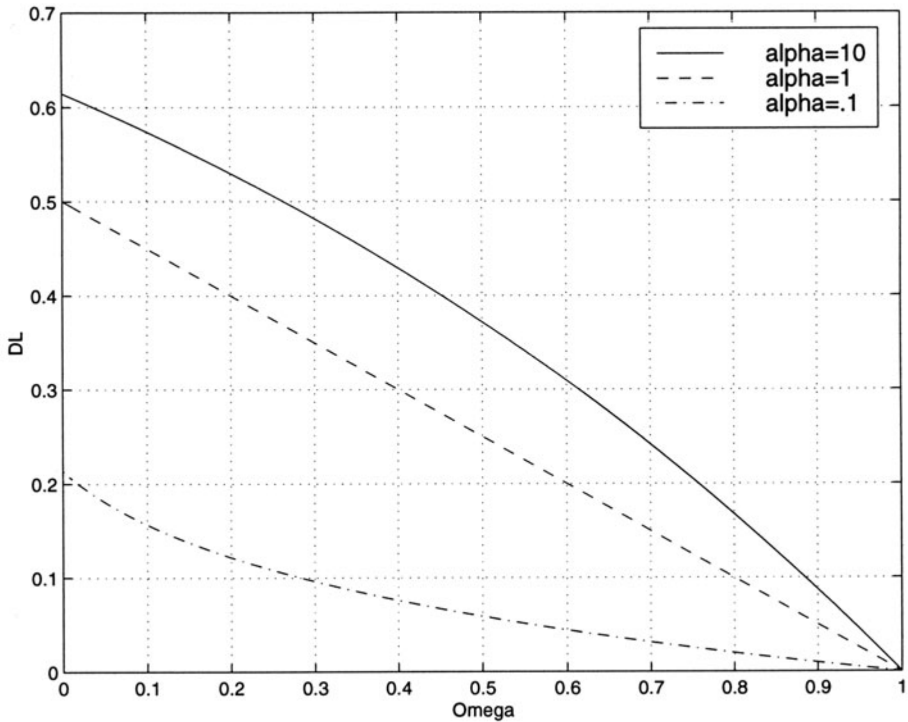

Figure 2 DL as a function of Ω with the new model for three values of α.

for a required DL_{max}, we need a TT_{max} given by

$$TT_{max} = \frac{DL_{max}}{\left|\frac{dDL}{d\Omega}\right|_{\Omega=1}}. \tag{1.28}$$

The first derivative at $\Omega = 1$ for the new DL is given by

$$\left.\frac{dDL}{d\Omega}\right|_{\Omega=1} = -\frac{\alpha\lambda}{\alpha+\lambda}. \tag{1.29}$$

The first derivative at $\Omega = 1$ for the Williams-Brown model is given by Equation (1.13). Then, comparing TT_{max} for the new model and TT_{max} for the Williams-Brown model, we obtain

$$\frac{TT_{max}(\text{new model})}{TT_{max}(\text{Williams-Brown})} = 1 + \frac{\lambda}{\alpha} \tag{1.30}$$

Typical values of α can be easily 10 times smaller than typical values of λ. Thence, the expression above can be further simplified to

$$\frac{TT_{max}(\text{new model})}{TT_{max}(\text{Williams-Brown})} \approx \frac{\lambda}{\alpha} \tag{1.31}$$

The equation above shows that the maximum allowable test transparency is inversely proportional to the clustering parameter α. This explains why the fault coverage requirement as predicted by the new model can be radically lower than that predicted by the Williams-Brown formula. For the example given before, if $DL = 100$ p.p.m. is required and $Y = 80\%$, the Williams-Brown model requires $\Omega = 99.55\%$, i.e., a maximum allowable test transparency $TT_{max} = .45\%$. If α is 10 times smaller than λ, $TT_{max} = 4.5\%$ for the new DL model. That is, instead of 99.55% fault coverage the new model requires only 95.5%.

4. CONCLUSION

In this paper a critical review of existing defect level theories has been presented, and a model that accounts for the clustering effect has been proposed. The original feature of the new model is that it assumes that the distribution of the number faults per chip is given by the generalized negative binomial distribution.

Other models in the literature either do not account for clustering or assume one distribution for the number of defects, and another distribution for the number of faults per defect. Our method of directly considering faults, eliminates possible overlapping between the roles of the parameters in the methods that consider two distributions. In this way, we were able to study the effect of clustering by varying a single clustering parameter.

Analysis of the new method revealed that the clustering effect is a very significant one, which cannot be ignored. Models such as the Williams-Brown model, that do not account for clustering, can easily underestimate the maximum allowable test transparency in one order of magnitude. In the case study presented in this paper, the Williams-Brown model required 99.5% fault coverage, while the new model required about 95.55%. This is much closer to what test engineers usually observe, and raises the optimism and confidence of manufacturers.

Directions for continuing this work are various. The newly derived model needs to be validated with actual DL data. The question of which fault models to use in order to represent the real faults remains open. Another question is whether the experimental DL curves contain or not any information about unmodeled faults. Should the *jumps* that appear on the measured yield versus fault coverage curve be modeled? In [Das et al., 1993], these jumps are not ignored and are treated as an integral part of the $Y_m(T)$ curve, but it is not known whether this is important.

5. ACKNOWLEDGMENTS

The author is grateful to Vishwani Agrawal for long and fruitful discussions on this theme.

6. REFERENCES

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

[Das et al., 1993] Das, D., Seth, S. C., and Agrawal, V. D. (1993). "Accurate Computation of Field Reject Ratio Based on Fault Latency". *IEEE Trans. on VLSI*, 1(4).

[Maxwell and Aitken, 1991] Maxwell, P. C. and Aitken, R. C. (1991). "The Effect of Different Tests Sets on Quality Level Prediction: When is 80% Better than 90%?". In *Proc. Int. Test Conference (ITC)*, pages 358–364.

[McCluskey and Buelow, 1988] McCluskey, E. J. and Buelow, F. (1988). "IC Quality and Test Transparency". In *Proc. Int. Test Conference (ITC)*, pages 295–301.

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

[Singh and Krishna, 1996] Singh, A. D. and Krishna, C. M. (1996). "On the effect of Defect Clustering on Test Transparency and IC Test Optimisation". *IEEE Transactions on Computers*, C-45(6):753–757.

[Sousa et al., 1996] Sousa, J. J. T., Goncalves, F. M., Teixeira, J. P., Marzocca, C., Corsi, F., and Williams, T. W. (1996). "Defect Level Evaluation in an IC Design Environment". *IEEE Trans. on CAD*, 15:1286–1293.

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

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

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