Yield model for fault clusters within integrated circuits

by C. H. Stapper

Generalized negative binomial statistics turns out to be a model of the fault distribution in very large chips or wafers with internal defect clusters. This is expected to influence large chip and full wafer redundancy requirements. Furthermore, the yield appears to be affected by an experimental dependence of the average number of faults on chip area.

1. Introduction

With the advent of wafer scale integration the question of the adequacy of existing yield models has come up. Until now, clustering of faults within a chip or integrated circuit has not been addressed properly. It has been assumed that the defects causing faults on a wafer, or in some cases in given regions of a wafer, can be modeled with Poisson distributions [1-3]. It was observed experimentally that the defect densities of each region or wafer had to be modeled as an additional random variable. These variations in defect densities were modeled by using mixed or compounded Poisson statistics. Another approach described by Stapper, Armstrong, and Saji took clustering into account by considering the formation of defects on a chip during the manufacturing process [4]. It was fortuitous that all these approaches led to the same yield expression:

$$Y = (1 + \lambda/\alpha)^{-\alpha}, \tag{1}$$

where Y is the chip yield, λ the average number of faults (or the mean number of defects per chip causing failure), and α a cluster parameter.

The derivations of yield formula (1) did not deal with the clustering of defects within the chip. As chip sizes increase up to the limits of full wafers, the likelihood of having clusters within the chip area increases. The purpose of this paper is to provide the mathematical proof of formula (1), taking clustering within the chip into account. It is also shown that generalized negative binomial statistics associated with (1) is applicable to model within-chip clustering.

In a previous paper [5] a method was described for deriving Poisson statistics to model the fault distribution on chips. In this paper the same approach is taken, but it is extended to include clustering. The method is very similar to the one described in [4]. In that paper, however, clustering was the result of fault formation as a function of time. In this paper the probability of a defect causing a failure in a given area increases when a fault already occurs in an adjacent area.

2. Base for the model

The probability of x faults occurring within an integrated circuit area S can be denoted by $p(x,S)$. Let the random variable X_S designate the number of faults in S. Similarly, let $X_{\Delta S}$ be the random variable associated with the number of faults in a small area of integrated circuitry ΔS adjacent to and connected to S. The number of faults in the combined area $S + \Delta S$ is then given by the random variable $X_{S+\Delta S} =$

$X_S + X_{\Delta S}$. The objective of the derivation that follows is to determine an expression for the probability of finding $X_{S+\Delta S}$ faults in area $S + \Delta S$. This expression must be a function of the probabilities for finding faults in S.

There are $x + 1$ ways that x faults can occur in $S + \Delta S$. This includes combinations such as all x faults in S and none in ΔS; or $x - 1$ faults in S and one in ΔS; or $x - 2$ faults in S and two in ΔS; and so on. These are all independent events. The probability of finding $X_{S+\Delta S} = x$ faults in area $S + \Delta S$ is therefore

$$p(x,S + \Delta S) = \sum_{i=0}^{x} p(x - i,S)p(i,\Delta S \mid x - i,S), \tag{2}$$

where $p(i,\Delta S \mid x - i,S)$ is the probability of finding i faults in ΔS given that there are $x - i$ faults in area S. Expansion of the first two terms gives

$$\begin{aligned} p(x,S + \Delta S) = {} & p(x,S)p(0,\Delta S \mid x,S) \\ & + p(x - 1,S)p(1,\Delta S \mid x - 1,S) \\ & + \sum_{i=2}^{\infty} p(x - i,S)p(i,\Delta S \mid x - i,S). \end{aligned} \tag{3}$$

If the area of ΔS is small enough, we can assume that the probability of finding two or more faults in ΔS becomes negligibly small. In that case

$$\begin{aligned} p(x,S + \Delta S) \simeq {} & p(x,S)p(0,\Delta S \mid x,S) \\ & + p(x - 1,S)p(1,\Delta S \mid x - 1,S). \end{aligned} \tag{4}$$

The probability of finding no faults in ΔS is the complementary event to finding one or more faults. This can be written as

$$p(0,\Delta S \mid x,S) = 1 - \sum_{i=1}^{\infty} p(i,\Delta S \mid x,S). \tag{5}$$

Since it was assumed that the probability of finding two or more faults in ΔS is negligibly small, (5) can be approximated by

$$p(0,\Delta S) \simeq 1 - p(1,\Delta S \mid x,S). \tag{6}$$

Next an estimate must be made for the magnitude of the probability of one fault occurring in ΔS. In a previous paper a simple proportionality was assumed [5]:

$$p(1,\Delta S \mid x,S) = \phi \Delta S, \tag{7}$$

where ϕ was taken to be a constant. However, in this case the left-hand side depends on the number of faults in the adjacent areas. The more faults there are in this region, the more likely it becomes that one fault will be found in ΔS. Consequently the proportionality factor ϕ in (7) must be $\phi(x)$, a function of the number of faults $X_S = x$.

The combination of (6) and (7) gives

$$p(0,\Delta S \mid x,S) \simeq 1 - \phi \Delta S. \tag{8}$$

Substituting (7) and (8) into (4) results in

$$\begin{aligned} p(x,S + \Delta S) \simeq {} & p(x,S)[1 - \phi(x)\Delta S] \\ & + p(x - 1,S)\phi(x - 1)\Delta S. \end{aligned} \tag{9}$$

Rearrangement of this formula leads to the recursive difference relationship

$$\begin{aligned} \frac{p(x,S + \Delta S) - p(x,S)}{\Delta S} = {} & -\phi(x)p(x,S) \\ & + \phi(x - 1)p(x - 1,S). \end{aligned} \tag{10}$$

By letting the incremental area ΔS shrink to zero, the above becomes a differential recursive equation,

$$\frac{\partial}{\partial S} p(x,S) = -\phi(x)p(x,S) + \phi(x - 1)p(x - 1,S). \tag{11a}$$

Using the preceding approach, it is not difficult to show that the starting equation for $x = 0$ becomes [5]

$$\frac{\partial}{\partial S} p(0,S) = -\phi(0)p(0,S). \tag{11b}$$

These two equations must be solved to obtain the distribution $p(x,S)$. The nature of this distribution depends entirely on $\phi(x)$.

3. Yield formula

The solution of (11b) is straightforward and results in

$$p(0,S) = Ce^{-\phi(x=0)S}, \tag{12}$$

where C is a constant of the integration. The result looks like a straightforward exponential and results in the yield

$$Y = Ce^{-\phi(x=0)S}. \tag{13}$$

This has all the appearances of the yield model associated with Poisson statistics which is usually written in the form [4, 5]

$$Y = Y_0 e^{-\theta AD}, \tag{14}$$

where Y_0 is a gross yield and θ a defect sensitivity factor (often referred to as the probability of failure), A the chip area, and D a defect density.

Indeed the theory of Hu [6] and the data analysis of Warner [7] lead us to believe that for large chips an exponential yield model like (14) is sufficient. The above result in (13) appears to agree with those conclusions. However, appearances are deceiving and it is shown in this paper that (13) does not follow Poisson statistics. Instead (13) equals the model in (1) and is associated with generalized negative binomial statistics. To show this requires solving (11a) and (11b) for the actual fault distribution $p(x,S)$.

4. Solution of the fault distribution

The solution to Eqs. (11a) and (11b) can be found systematically. The approach is similar to one described in

[4]. In this case (11a) is multiplied by t^x and (11b) by $t^0 = 1$. Summing the results for all possible values of x results in the differential equation

$$\frac{\partial}{\partial S}\sum_{x=0}^{\infty} p(x,S)t^x = (t-1)\sum_{x=0}^{\infty} p(x,S)\phi(x)t^x. \tag{15}$$

The summation within the derivative on the left-hand side of the equation is the probability distribution generating function defined by

$$G(t;S) \equiv \sum_{x=0}^{\infty} p(x,S)t^x. \tag{16}$$

The fault distribution probabilities are obtained from the probability distribution generating function with

$$p(x,S) = \frac{1}{x!}\left.\frac{\partial^x G(t;S)}{\partial t^x}\right|_{t=0}. \tag{17}$$

The objective is to determine $G(t;S)$ from (15). When this is done, $p(x,S)$ can be calculated with (17).

When $\phi(x) = c$ instead of a function of x, Eq. (14) can be simplified to

$$\frac{dG}{dS} = (t-1)cG. \tag{18}$$

This has the solution

$$G(t;S) = e^{(t-1)cS}, \tag{19}$$

which by application of (17) results in

$$p(x,S) = \frac{e^{-cS}(cS)^x}{x!}, \tag{20}$$

better known as the Poisson distribution. This result is not surprising.

In the case of clustering $\phi(x)$ is dependent on the number of faults x already in S. Assume the linear relationship

$$\phi(x) = c + bx, \tag{21}$$

where c and b are constants. Introducing this into (15) and regrouping terms on the right-hand side of (15) results in

$$\frac{\partial G(t;S)}{\partial S} = (t-1)\left[cG(t;S) + bt\frac{\partial G(t;S)}{\partial t}\right]. \tag{22}$$

The solution to this equation is

$$G(t;S) = \{\exp(bS) - [\exp(bS) - 1]t\}^{-c/b}. \tag{23}$$

Again the probability distribution $p(x,S)$ can be obtained from this probability distribution generating function with (17). This gives

$$p(x,S) = \frac{\Gamma(x + c/b)}{x!\Gamma(c/b)} e^{-(x+c/b)bS}(e^{bS} - 1)^x, \tag{24}$$

which is a generalized negative binomial distribution, and not a Poisson distribution as might have been expected from (12) and (13).

The yield appears to be given by $Y = p(0,S)$, which results in

$$Y = e^{-cS}. \tag{25}$$

Except for the integration constant, this is the same result as the one obtained in (13), and is similar to the yield expression for Poisson statistics. This is often believed to show that when clusters are included in the chip area, the yield vs. area dependence should be like the Poisson yield model, as indeed it is here [6, 7]. What proponents of such theories usually neglect is the wafer-to-wafer, lot-to-lot, day-to-day, week-to-week, and month-to-month variations of defect levels in integrated circuit fabricators. This was taken care of by mixing or compounding Poisson statistics [1–3]. In exactly this same way the results in (24) and (25) still have to be compounded to take care of regional and long-time variations in the parameters c and b. When we do this, the resulting yield vs. area plots are still expected to show the familiar uplift that resulted from compounding Poisson statistics. Such data undoubtedly can still be approximated by a model of the type

$$Y = (1 + aS/\alpha)^{-\alpha},$$

with a and α as parameters. What the results of this paper imply is that the average number of faults is going to be higher than aS. The reason for this is described in the next section.

5. The average number of faults

It is instructive to determine the average number of faults $\lambda = E(X_S)$ that can be expected from the distribution in (24). Using

$$\lambda = E(X_S) = \sum_{x=0}^{\infty} xp(x,S), \tag{26}$$

rearranging some terms, and making the substitution $x = u + 1$ results in

$$\lambda = e^{-bS}(e^{bS} - 1)\left[c/b + \sum_{u=0}^{\infty} up(u,S)\right]. \tag{27}$$

The infinite sum on the right-hand side of this equation is the same as the one in (26), thus reducing (27) to

$$\lambda = e^{-bS}(e^{bS} - 1)(c/b + \lambda). \tag{28}$$

This can be solved to give

$$\lambda = (e^{bS} - 1)c/b. \tag{29}$$

The preceding result is most interesting. It shows that the average number of faults increases exponentially with the surface area S. This implies, for instance, that one half of the area has on the average fewer faults than the other half. This is all caused by the initial assumption in (21) and is precisely the type of situation that can be expected when clustering takes place within a chip.

It may be necessary in the future to restrict the variable S to a range $0 \le S \le A$, where A is a chip area. The value of c could then depend on this area, so that we get $c(A)$. This will then make (29) a function of A. Applicability of this type of condition will depend on the analysis of actual data.

6. Chip yield

The chip yield can be expressed in terms of the average number of faults. Rearrangement of (29) gives

$$S = \ln(1 + \lambda b/c)^{1/b}, \tag{30}$$

which when substituted into (25) results in

$$Y = (1 + \lambda b/c)^{-c/b}. \tag{31}$$

Substitution of $\alpha = c/b$ makes this result identical to (1). This completes the proof.

Using λ and α as the parameters in (24) results in the more familiar-looking generalized negative binomial distribution [4]

$$p(x,S) = \frac{\Gamma(x + \alpha)}{x!\Gamma(\alpha)} \frac{\lambda/\alpha}{(1 + \lambda/\alpha)^{x+\alpha}}. \tag{32}$$

The area relationship in this case is implied by the average number of faults λ, as it was in all previous theories. The only difference is that this area relationship is no longer linear but exponential.

7. Discussion

The real world of yield is very complex. Mathematics can only approximate that reality. Nevertheless, if properly used, mathematics can give us insight into the nature of the problems that have to be solved. This is the case here.

The results of this paper are extremely important for full wafer manufacturers. Peltzer [8] has already mentioned the use of Poisson statistics as a yield model for full wafer manufacture at Trilogy. He wrongly quotes this author as the source for that model. It is, however, the Poisson model in Reference [6]. This is wrong. The Poisson fault distribution is very narrow when compared to actual data [9]. Since full wafer manufacturers are using redundancy, they will be deceived if they use Poisson statistics. The theoretical calculation of yield with redundancy is too high when those statistics are used.

The fault distribution on actual products is very wide. This is caused by clustering and can be approximated with generalized negative binomial statistics as shown in Reference [9]. In this paper it has been shown that this approximation can be expected to hold for clustering within the chip. In previous papers it was shown that chip-to-chip variations due to large area clustering could be modeled similarly [1-5].

To correctly model yield it should be realized that Expression (1) as derived in this paper holds only for one chip or wafer. Variations among wafers and areas on wafers give different values of the average number of faults for each individual case. These effects have a far greater influence on the yield than internal clustering. To take them into account required that Expression (1) be compounded to give

$$Y = \sum_i g_i(1 + \lambda_i/\alpha_i)^{-\alpha_i}, \tag{33}$$

or

$$Y = \int_0^\infty \frac{g(\lambda)d\lambda}{(1 + \lambda/\alpha)^{\alpha}}. \tag{34}$$

Compounding like this is usually neglected in single wafer yield theories as in [6], therefore missing the major variable in yield modeling. This author has investigated several functions for $g(\lambda)$ in (34). Unfortunately, until now no easily usable results have been obtained.

The statistics associated with (33) and (34) will lead to fault distributions with very long tails. Data analyzed in the last few years by this author have indeed shown a propensity for such long tails. This is important for deciding the amount of redundancy that is needed to fix the faults expected in future products. Yield modelers who misjudge this effect can cause enormous capital losses in unsuccessful ventures if they do not rely on sufficient redundancy.

References

1. C. H. Stapper, "Defect Density Distribution for LSI Yield Calculations," *IEEE Trans. Electron Devices* **ED-20,** 655–657 (July 1973).
2. C. H. Stapper, "LSI Yield Modeling and Process Monitoring," *IBM J. Res. Develop.* **20,** 228–234 (May 1976).
3. O. Paz and T. R. Lawson, Jr., "Modification of Poisson Statistics: Modeling Defects Induced by Diffusion," *IEEE J. Solid-State Circuits* **SC-12,** 540–546 (October 1977).
4. C. H. Stapper, F. M. Armstrong, and K. Saji, "Integrated Circuit Yield Statistics," *Proc. IEEE* **71,** 453–470 (April 1983).
5. C. H. Stapper, "Comments on 'Some Considerations in the Formulations of IC Yield Statistics,' " *Solid-State Electron.* **24,** 127-132 (February 1981).
6. S. M. Hu, "Some Considerations in the Formulation of IC Yield Statistics," *Solid-State Electron.* **22,** 205-211 (February 1979).
7. R. M. Warner, "A Note on IC Yield Statistics," *Solid-State Electron.* **24,** 1045-1047 (December 1981).
8. D. L. Peltzer, "Wafer-Scale Integration: The Limits of VLSI?" *VLSI Design* **4,** 43–47 (September 1983).
9. C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product," *IBM J. Res. Develop.* **24,** 398–409 (May 1980).

Received March 27, 1984

Charles H. Stapper *IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452.* Dr. Stapper received his B.S. and M.S. in electrical engineering from the

Massachusetts Institute of Technology in 1959 and 1960. After completion of these studies, he joined IBM at the Poughkeepsie, New York, development laboratory, where he worked on magnetic recording and the application of tunnel diodes, magnetic thin films, electron beams, and lasers for digital memories. From 1965 to 1967, he studied at the University of Minnesota on an IBM fellowship. Upon receiving his Ph.D. in 1967, he joined the development laboratory in Essex Junction. His work there included magnetic thin-film array development, magnetic bubble testing and device theory, and bipolar and field effect transistor device theory. He is now a senior engineer in the development laboratory, where he works on mathematical models for yield and reliability management. Dr. Stapper is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.