The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions

by C. H. Stapper

A method for modeling the variations in defect levels in circuits produced on modern integrated circuit manufacturing lines is described in this paper. The effects on defect and fault distributions are derived. A deficiency in some previous yield models is eliminated.

1. Introduction

Yield modelers have to take into account not only the wafer to wafer variations in defect densities, but also lot to lot, day to day, week to week, and month to month variations in defect levels that occur in integrated circuit fabrication. Models for these effects are described in this paper. All these models are based on the application of straightforward, elementary statistics. They are developed from fundamental random defect theory and adapted to actual data by deductive analysis.

2. A binomial model

Consider first one wafer with N random defects in an area S containing chips of area A. If the number of defects per chip is designated by the random variable X_D, then according to [1, 2] the probability of finding k defects on a chip is given by

$$P(X_D = k) = \frac{N!}{k!(N-k)!}(A/S)^k(1 - A/S)^{N-k}. \quad (1)$$

The yield is the probability of having zero defects on a chip so that

$$Y = P(X_D = 0)$$
$$= (1 - A/S)^N. \quad (2)$$

This is known as a binomial yield model.

3. A compound model

The nature of integrated circuit manufacturing lines is such that very seldom do all wafers have the same number of defects on them. In fact the number of defects N per wafer behaves like another random variable with its own probability distribution $P(N = i)$, where $i = 0, 1, 2, \cdots$. Note that this distribution does not depend on chip area. It can be combined with the distribution given in Eq. (1) by

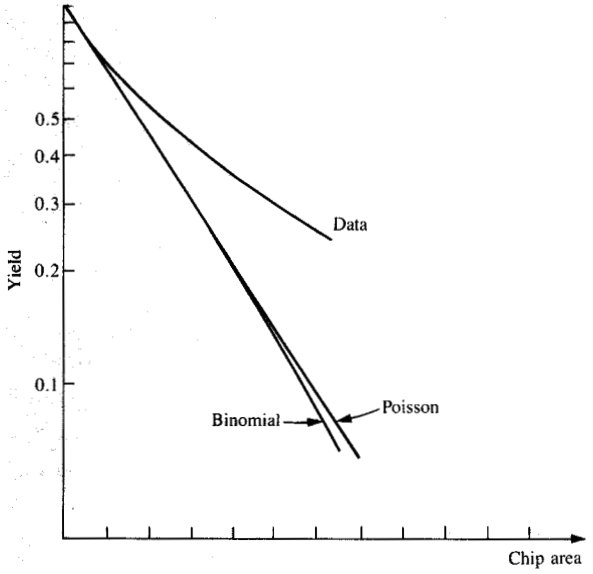

Figure 1

Yield as a function of chip area for the binomial and Poisson models.

the method of compounding [3–5]. This produces the defect distribution

$$P(X_D = k) = \sum_{i=k}^{\infty} P(N = i) \frac{i!}{k!(i-k)!} (A/S)^k (1 - A/S)^{i-k}. \quad (3)$$

It is possible to model $P(N = i)$ by the Poisson distribution

$$P(N = i) = \frac{e^{-\bar{N}} \bar{N}^i}{i!}, \quad (4)$$

where $\bar{N}$ is the average number of defects per wafer and the surface area of each wafer is equal to S. When this distribution is introduced into (3), the summation reduces to

$$P(X_D = k) = \frac{e^{-A\bar{N}/S} (A\bar{N}/S)^k}{k!}. \quad (5)$$

This is another Poisson distribution. The derivation of this result is shown in detail in Appendix 1.

4. Models versus reality

It would have been nice if the distribution described in Formula (5) had been the correct one for modeling defects in integrated circuit chips. All integrated circuit yield modeling could then be done with simple Poisson statistics. Unfortunately, actual defect distributions show invariably that the mean number of defects $E(X_D)$ is smaller than the variance $V(X_D)$ of the distribution [6, 7]. For the Poisson distribution in (5), these two quantities are equal, whereas for the binomial model in (1), the variance is smaller than the mean. The methods for modeling this are the topic of this paper.

The yield model associated with (5) can be written as

$$Y = P(X_D = 0) = e^{-A\bar{N}/S}. \quad (6)$$

The unsuitability of this function as a yield model can also be illustrated with yield versus area plots. The models in (2) and (6) give the results shown in the semilogarithmic plot in **Figure 1**. Plots based on actual data have been described in great detail by most authors in this field. The results give higher yields than those obtained with (2) and (6), as shown in Fig. 1.

5. Wafer partitioning

The discrepancy between actual data and (1) and (2) is caused by two effects. First of all the wafer to wafer defect distribution $P(N = i)$ cannot be correctly represented by the Poisson distribution in (4). The defect distribution data suggest that the correct model for $P(N = i)$ must be wider than the one given in Formula (4). This can be modeled by using compound Poisson distributions for $P(N = i)$ in Eq. (3). A general approach for doing this is given in Appendix 1. The historical approach, leading to the same results, is followed in the rest of this paper.

The second reason for the differences between data and models is defect clustering. Defects often congregate in areas such as the periphery of wafers. The model in (1) is therefore no longer valid. To get around this problem, wafers have been partitioned into regions. In some cases only two regions were used [8, 9]; others used more [10, 11]. In one IBM manufacturing plant defects are counted in five predetermined concentric regions.

In principle the defects in each region can be modeled with (1), although in practice it has been found more useful to model the defects per chip of area A with the Poisson distribution

$$P(X_D = k) = \frac{e^{-AD_i} (AD_i)^k}{k!}, \quad (7)$$

where D_i is the average defect density of region i. The rationale for the applicability of this distribution is addressed in [2] and [12].

6. Compounding partitioned models

Experimental data in [8, 9] have suggested that the average defect density in each region i varies from region to region and wafer to wafer. This can again be modeled with the method of compounding. In this case it takes on the form

$$P(X_D = k) = \sum_{i=0}^{M} P(D_i) e^{-AD_i} (AD_i)^k / k!, \quad (8)$$

where $P(D_i)$ is the probability distribution of average defect densities D_i for M regions. Note that $P(D_i)$ is not a function of chip area since it pertains to the average defect density in a region.

In practice the distribution given in Eq. (8) not only has to model the region to region variation but also the wafer to wafer variations of average defect densities for each region. At IBM the production of integrated circuits is planned by quarters. Yield projections for planning purposes are therefore made to cover a period of three months. The variations in defect levels during this period therefore have to be included. This makes the value of M very large, and for all practical purposes, it can be approximated by infinity.

If (8) becomes an infinite sum, it is again possible to model the compounder $P(D_i)$ by a discrete probability distribution. A Poisson distribution can be used. It takes on the form

$$P(aD_i = i) = \frac{e^{-a\bar{D}}(a\bar{D})^i}{i!}, \tag{9}$$

where a is a parameter and $\bar{D}$ an average defect density for all regions combined.

If (9) is substituted into (8), it results in

$$P(X_D = k) = \frac{e^{-a\bar{D}}(A/a)^k}{k!} \sum_{i=0}^{\infty} \frac{(a\bar{D}e^{-A/a})^i}{i!} i^k. \tag{10}$$

This is known as the Neyman type-A distribution. It has a mean and variance given by

$$E(X_D) = A\bar{D}, \tag{11a}$$

$$V(X_D) = A\bar{D}(1 + a\bar{D}). \tag{11b}$$

In this case the variance is larger than the mean. This distribution has indeed been used successfully by F. Armstrong and K. Saji to model actual particle, defect, and fault distributions in semiconductor processes [13].

The yield model associated with (10) has the form

$$Y = \exp[a\bar{D}(e^{-A/a} - 1)]. \tag{12}$$

This is an interesting function of chip area A, as is shown in **Figure 2**. For very large areas, the yield becomes asymptotic to a lower bound $e^{-a\bar{D}}$. This is independent of area A and therefore is constant. It is this property of the Neyman type-A yield model that restricts its usefulness to small chip areas. To alleviate this problem other defect density distributions have to be tried.

7. Transition to a continuous defect density distribution

The probability distribution of defect densities $P(D_i)$ is associated with M different defect densities D_i. Each one represents the average defect density for the regions indicated by the index i. The probability associated with each defect density depends on the area of the corresponding region

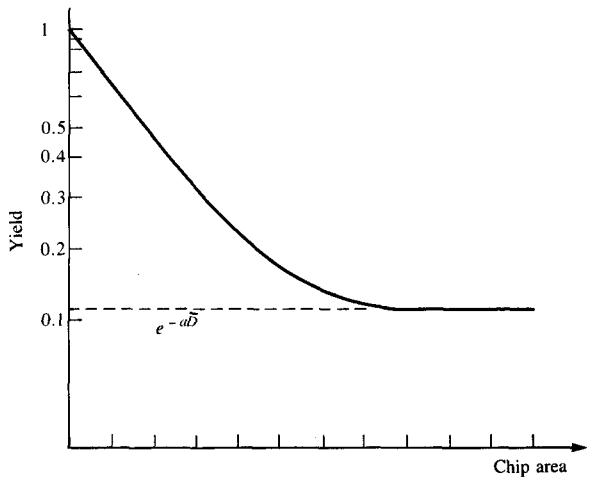

Figure 2

Yield as a function of chip area for a Neyman type-A model.

[2, 10, 12]. This probability is given by s_i/S, where s_i is the area of the ith partition and the total area S is obtained with the summation

$$S = \sum_{i=1}^{M} s_i. \tag{13}$$

The distribution $P(D_i)$ in (8) is therefore given by

$$P(D_i) = s_i/S. \tag{14}$$

Substitution of this result into (8) gives

$$P(X_D = k) = \sum_{i=1}^{M} \frac{s_i}{S} \frac{e^{-AD_i}(AD_i)^k}{k!}. \tag{15}$$

It is possible mathematically to express this result as

$$P(X_D = k) = \sum_{i=1}^{M} \frac{s_i}{S} \int_0^{\infty} \frac{e^{-AD}(AD)^k}{k!} \delta(D - D_i)dD, \tag{16}$$

where $\delta(D - D_i)$ is a unit impulse or delta function occurring at $D = D_i$. Rearrangement of (16) gives

$$P(X_D = k) = \int_0^{\infty} \frac{e^{-AD}(AD)^k}{k!} \sum_{i=0}^{M} \frac{s_i}{S} \delta(D - D_i)dD, \tag{17}$$

where it is possible to define a probability distribution of defect densities function

$$P(D) = \sum_{i=0}^{M} \frac{s_i}{S} \delta(D - D_i). \tag{18}$$

This is a string of delta functions occurring at the appropriate values of D. Combination of (18) and (17) gives

$$P(X_D = k) = \int_0^{\infty} \frac{e^{-AD}(AD)^k}{k!} P(D)dD, \tag{19}$$

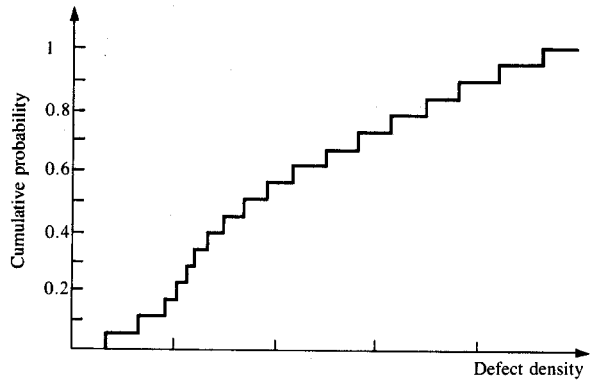

Figure 3

Cumulative distribution function for defect densities.

which is an integral form for compounding a Poisson distribution [4, 5].

Associated with the probability distribution of defect densities in Expression (18) is the cumulative distribution function

$$C(D) = \int_0^D P(D')dD'. \tag{20}$$

Since the integral of the delta function $\delta(D - D_i)$ is a unit step function $\sigma(D - D_i)$, it is possible to express the cumulative distribution by

$$C(D)_i = \sum_{i=1}^{D_i \le D} \frac{s_i}{S} \sigma(D - D_i). \tag{21}$$

This is a staircase function like the one shown in **Figure 3**. The summation is only over those values of i for which $D_i \le D$.

When we let the number M approach infinity, the sum

$$\lim_{M \to \infty} \sum_{i=1}^{M} s_i = \infty,$$

since the regional areas are finite. Therefore, the area S approaches infinity. Nevertheless the sum

$$\sum_{i=1}^{M} s_i/S = \left(\sum_{i=1}^{M} s_i\right) \Big/ \sum_{i=1}^{M} s_i \tag{22}$$

remains equal to one regardless of the value of M. We are therefore left with an infinite series of infinitely small steps that start at $C(0) = 0$ and end up at $C(\infty) = 1$. This can best be presented by a continuous cumulative distribution function $C(D)$.

The function $C(D)$ leads to the continuous probability distribution function

$$P(D) = \frac{dC(D)}{dD}. \tag{23}$$

This continuous probability distribution function can be used for compounding in the same way as the string of delta functions in (19).

A number of distribution functions have been tried for $P(D)$. Until now a gamma distribution has shown the greatest potential. It was discovered for use in wafer to wafer modeling in [14] and confirmed in [8] and [9] for regional and wafer to wafer modeling.

8. Compounding and chip area

We next investigate the effect of a change in chip area on compounding. The defect distribution

$$P(X_D = k) = \frac{e^{-A\bar{D}}(A\bar{D})^k}{k!} \tag{24}$$

remains unchanged when compounded with a delta function in the form

$$P(X_D = k) = \int_0^\infty \frac{e^{-AD}(AD)^k}{k!} \delta(D - \bar{D})dD. \tag{25}$$

The integral form of compounding was derived as the limit of a series of delta functions. Compounding (24) with a continuous distribution is identical to (19), as seen in the preceding section.

When we increase the chip area A by a factor n, the defect distribution is

$$P(X_D = k) = \frac{e^{-nA\bar{D}}(nA\bar{D})^k}{k!}. \tag{26}$$

This distribution differs from the one in (24). It also remains unchanged when compounded with a delta function to give

$$P(X_D = k) = \int_0^\infty \frac{e^{-nAD}(nAD)^k}{k!} \delta(D - \bar{D})dD. \tag{27}$$

The delta function clearly is not a function of the chip area. Compounding with a continuous probability distribution $P(D)$ derived in the limit from delta functions therefore gives

$$P(X_D = k) = \int_0^\infty \frac{e^{-nAD}(nAD)^k}{k!} P(D)dD. \tag{28}$$

It should be noted that the compounding of defect density distribution in (27) and (28) is not a function of the chip area.

9. Faults and defects

At IBM it has been found useful to distinguish between defects and faults. A fault is defined as a defect causing a failure. Not all defects cause chip failures. To do this they must occur in the areas where they interfere with the electrical operation of the chip. These are known as the critical areas [12]. In some cases failures are not detected until bias conditions are changed or the wave shapes of the

applied pulses are altered. If these changes and alterations do not exceed specifications, then the defects are causing faults. The fraction of defects that causes product failures under specified test conditions is indicated by θ. This has also been called the probability of failure.

The number of faults on a chip can never exceed the number of defects. The probability of finding x faults on a chip with k defects is a conditional probability designated as $P(X_F = x \mid X_D = k)$, where X_F is a random variable denoting the number of faults on a chip. The probability of finding x faults is given by θ^x. This implies that there are $(k - x)$ defects that will not cause faults. The probability of this happening can be calculated with $(1 - \theta)^{k-x}$. The number of combinations by which k defects can cause x faults is given by $k!/x!\,(k - x)!$. The probability of finding x faults on a chip that already has k defects is therefore given by

$$P(X_F = x \mid X_D = k) = \frac{k!}{x!(k-x)!}\theta^x(1-\theta)^{k-x}. \tag{29}$$

This is a binomial distribution and it clearly shows that the probability of failure is an applicable nomenclature for θ.

The probability of finding x faults on a chip can be caused by any number of defects $k > x$. To find the unconditional probability $P(X_F = x)$ therefore requires taking into account all probabilities $P(X_D = k)$ for $k > x$. This is done with

$$P(X_F = x) = \sum_{k=x}^{\infty} P(X_F = x \mid X_D = k)P(X_D = k). \tag{30}$$

Substitution of (24) and (29) into this expression gives

$$P(X_F = x) = \sum_{k=x}^{\infty} \frac{k!}{x!(k-x)!}\theta^x(1-\theta)^{k-x}\frac{e^{-AD}(A\bar{D})^k}{k!}. \tag{31}$$

Definition of the index $i = k - x$ and rearrangement of this expression give

$$P(X_F = x) = \frac{e^{-A\bar{D}}(\theta A\bar{D})^x}{x!}\sum_{i=0}^{\infty}\frac{[(1-\theta)A\bar{D}]^i}{i!}. \tag{32}$$

The summation on the right is equal to $\exp[(1 - \theta)AD]$, so that (32) becomes

$$P(X_F = x) = \frac{e^{-\theta AD}(\theta AD)^x}{x!}. \tag{33}$$

This is the Poisson distribution with the probability of failure included as a parameter.

The result in (33) can be compounded with a delta function to give

$$P(X_F = x) = \int_0^{\infty} \frac{e^{-\theta AD}(\theta AD)^x}{x!}\delta(D - \bar{D})dD. \tag{34}$$

The probability of failure in this case does not affect the compounding delta function. In the same way, it will not affect the continuous compounding derivable from (34). Thus

$$P(X_F = x) = \int_0^{\infty} \frac{e^{-\theta AD}(\theta AD)^x}{x!}P(D)dD \tag{35}$$

represents the general formula for compounding the fault distribution.

10. Measuring wafer to wafer defect density variations

The first application of (35) for modeling the wafer to wafer variations in defect densities occurred at IBM in 1972 and was reported in [14]. It was found experimentally that the number of failing defect monitors on test wafers were not distributed as either binomial or Poisson distributions. This could only happen if the defect densities were not the same from wafer to wafer. However, since this approach seems not to have been understood in recent papers [15–17], let us look into that venerable analysis in more detail.

The data in [14] consisted of distributions of the number of failing defect monitors per wafer. Some of these monitors had long serpentine lines to determine the defects that caused open circuit failures or faults. Others consisted of interdigitated fingers to detect the defects that caused short circuit faults. These monitors were made with diffusions, polysilicon and metal patterns. Each pattern was replicated 50 times on each wafer.

If the defect densities for a given defect type are the same from wafer to wafer, then the number of failing monitors X_{FM} is given by

$$P(X_{FM} = k) = \binom{50}{k}y^{50-k}(1-y)^k, \tag{36}$$

where y is the yield of the defect monitor and $k = 0, 1, 2, \cdots$. The mean and variance of this distribution are given by

$$E(X_{FM}) = 50(1 - y), \tag{37a}$$

$$V(X_{FM}) = 50y(1 - y). \tag{37b}$$

Here the variance is smaller than the mean. In the actual data the variance was found to be larger than the mean. An example of this is shown in **Figure 4**. This suggested a wafer to wafer variation in the value of y. To model this, (36) was compounded. Use of a beta function for the compounder occasionally gave a satisfactory model for these data. However, the resulting yield formulas, distributions, and estimators for the parameters were cumbersome. They were therefore not used as models and not reported in the literature. For completeness some of the compounded binomial distributions that were investigated at that time are given in Appendixes 2 and 3.

A less complex model was obtained by approximating (36) with the Poisson distribution

$$P(X_{FM} = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \tag{38}$$

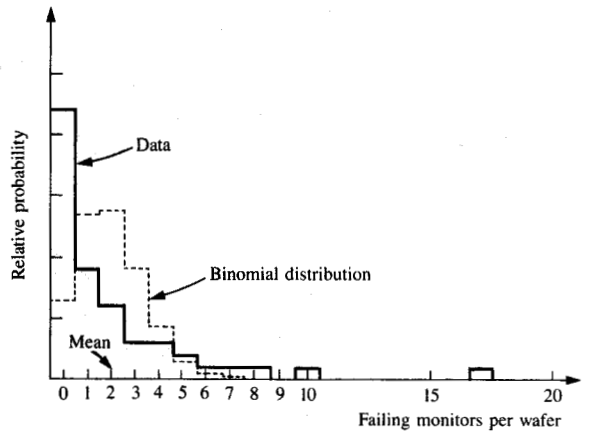

Figure 4

A distribution of the number of failing monitors per wafer. The data are shown in solid lines; a binomial distribution with the same average is shown in dashed lines. The mismatch between the two curves is caused by wafer to wafer variation in defect densities.

where λ is the average number of failing monitors per wafer. In this model the mean and variance are both equal to λ.

In [8] it is shown that any compounded Poisson distribution has a variance that is larger than the mean. Modeling the distribution of the number of failing monitors per wafer in [14] was therefore reduced to the search for a compounder that gave the correct distribution, mean, and variance. Use of the gamma distribution in the compounding formula

$$P(X_{FM} = k) = \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} P(\lambda)d\lambda \tag{39}$$

gave a very suitable fit to the data, as was duly reported in the literature [14].

The gamma distribution can be written as

$$P(\lambda) = \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\Gamma(\alpha)\beta^\alpha}, \tag{40}$$

where α and β are parameters. The mean and variance of (40) are given by

$$E(\lambda) = \alpha\beta, \tag{41a}$$

$$V(\lambda) = \alpha\beta^2. \tag{41b}$$

One useful property of this distribution is that

$$\alpha = E^2(\lambda)/V(\lambda). \tag{42}$$

This is the inverse of the square of the coefficient of variation, which is normalized with respect to the mean. The quantity α is therefore a single parameter that describes the width and the nature of the gamma distribution completely.

Substitution of (40) into (39) results in the negative binomial distribution

$$P(X_{FM} = k) = \frac{\Gamma(\alpha + k)\beta^k}{k!\Gamma(\alpha)(1 + \beta)^{\alpha+k}}. \tag{43}$$

It has been practical to use (41a) to define $\bar{\lambda} = E(\lambda) = \alpha\beta$, so that $\beta = \bar{\lambda}/\alpha$. It is therefore possible to rewrite Formula (43) as

$$P(X_{FM} = k) = \frac{\Gamma(\alpha + k)(\bar{\lambda}/\alpha)^k}{k!\Gamma(\alpha)(1 + \bar{\lambda}/\alpha)^{\alpha+k}}. \tag{44}$$

This distribution has the mean number of failing monitors per wafer

$$E(X_{FM}) = \bar{\lambda} \tag{45a}$$

and the variance

$$V(X_{FM}) = \bar{\lambda}(1 + \bar{\lambda}/\alpha). \tag{45b}$$

In this case the variance is larger than the mean.

It is possible to estimate $\bar{\lambda}$ and α from the average $\bar{k}$ and the standard deviation σ_k of the number of failing monitors per wafer of the data. This is done by using the formulas

$$\bar{k} = E(X_{FM}), \tag{46a}$$

$$\sigma_k^2 = V(X_{FM}), \tag{46b}$$

and solving them for λ and α to give

$$\bar{\lambda} = \bar{k}, \tag{47a}$$

$$\alpha = \frac{\bar{k}^2}{\sigma_k^2 - \bar{k}}. \tag{47b}$$

Techniques such as maximum likelihood estimators can also be used at this point, but (47a, b) have been found adequate in practice. For high monitor yields the average number of failing monitors is related to the defect density by

$$\bar{D} \simeq \bar{\lambda}/50\theta_m A_m, \tag{48}$$

where $\theta_m A_m$ is the critical area for each one of the 50 monitors per wafer. When $\bar{D}$ and $\bar{\lambda}$ are proportional, as in (48), the defect density distribution is also a gamma distribution. This has a mean and variance

$$E(D) = E(\lambda)/50\theta_m A_m, \tag{49a}$$

$$V(D) = V(\lambda)/(50\theta_m A_m)^2. \tag{49b}$$

The parameter α for this distribution is given by

$$\alpha = E^2(D)/V(D)$$
$$= E^2(\lambda)/V(\lambda). \tag{50}$$

It is therefore the same as the one in (42) and remains unchanged under this change of variables. The fact that α is independent of chip area makes it and the associated gamma distribution very suitable for modeling regional and

temporal variations in integrated circuit defect densities, especially when it fits the actual data.

The chip yield for each defect type can be calculated using Formula (35) with zero faults or $k = 0$. This results in

$$Y = P(X_{FM} = 0)$$

$$= \int_0^\infty e^{-\theta AD} P(D) dD, \tag{51}$$

where θA is the critical area of the product chip. Formula (51) is known as Murphy's yield model. When the gamma distribution is used for the wafer to wafer defect distribution $P(D)$, the integral in (51) evaluates to

$$Y = (1 + \theta A\bar{D}/\alpha)^{-\alpha}. \tag{52}$$

This has been referred to as the negative binomial yield model.

The derivation of (51) and the method for estimating the model parameters have been described in this paper with a great deal of care and in great detail. This has been done because of the far-reaching significance of this approach.

It is of interest to note here that the model in Eq. (52) is parsimonious with the number of variables that are required. Only a single variable α has been introduced to model the variations in manufacturing defect levels. When individual defect levels were measured with the methods of [8] in a pilot line, the values of α were observed to vary from lot to lot. This appears to have been caused by sample limitations, since long time averages showed considerable stability. For a long period of time the value of $\alpha = 1$ gave a good average approximation for most defect types in the IBM integrated circuit factories in Sindelfingen, West Germany, and Essex Junction, Vermont. As yields improved, however, the average values of α decreased. Values of α less than 0.5 have been found appropriate for a number of defect types in recent years.

11. Practical applications

In the preceding section it was shown how defect monitor data can be converted to an equivalent product yield for each defect type. This method was used in a pilot line to track defect yields during the years 1972 and 1973. The defect levels in that line became progressively lower during that period. As a result the yields increased. It was soon learned that yields projected with the model in (51) were lower than the actual ones obtained on the manufactured products.

The reason why (51) projected lower yields than the actual ones was easily determined. In the approach of the preceding section we assumed that the defects were distributed on each wafer like a simple Poisson defect model. This turned out to be the wrong assumption. Our data showed that the defects were clustered on each wafer, with more defects near the edge than near the center. Two methods for handling this effect evolved independently at IBM in the early 1970s.

The technique developed at the IBM Laboratory in East Fishkill, New York, has been described by Paz and Lawson in [9]. They used wafer maps to locate transistors that failed because of base collector shorts known as pipes. The map for each wafer was divided into an inner and an outer region or zone. The yield within each region was analyzed as a function of emitter area. This was done by combining transistor chain data into groups. Groups containing an equal number of adjacent transistors were randomly selected on each wafer map. The object was to determine the fraction of these groups that was defect free. This was the yield associated with each group size. Such yields were determined for each region on each wafer using eight different sizes of groups. Each group size contained a fixed amount of emitter area. The data consisted therefore of transistor pipe yield as a function of emitter area for each region on each wafer.

The object of the Paz and Lawson method was to obtain values for the yield model

$$Y = Y_0 e^{-AD}, \tag{53}$$

where Y_0 is a gross cluster yield and e^{-AD} represents the simple Poisson random defect yield in each region. The area A in (52) represents the total emitter area in a group, and D the random defect density. The values of Y_0 and D were determined with a linear regression technique using

$$\ln Y = -AD + \ln Y_0. \tag{54}$$

This was done for each region on each wafer. Results confirmed that the wafer to wafer defect densities could be modeled with a gamma distribution. The defect densities in the outer regions, furthermore, were found to be higher than those in the inner ones. It was also observed that the gross cluster yield Y_0 was lower in the outer regions. It furthermore appeared that the wafer to wafer distribution of Y_0 could be modeled with a beta distribution.

The model of Paz and Lawson is still used today to analyze defect data in the IBM East Fishkill manufacturing plant. It has withstood the test of time in an actual integrated circuit manufacturing environment.

Another way of estimating the same parameters was developed independently at the IBM laboratory in Essex Junction. This technique has been described in [8]. It also depended on wafer map analysis. In this case the wafer maps that were used showed the location of failing defect monitors. The maps were again divided into inner and outer regions, as was done by Paz and Lawson. The difference between the two methods lay in determining which failures were gross clusters and which ones were random. In the Essex Junction approach any group of three or more adjacent failing monitors was called a cluster. The total number of clustered monitors were counted to determine a cluster-limited yield. This was done independently for the inner and outer regions of each wafer. After completion of this tally, the clustered failures were discarded from the

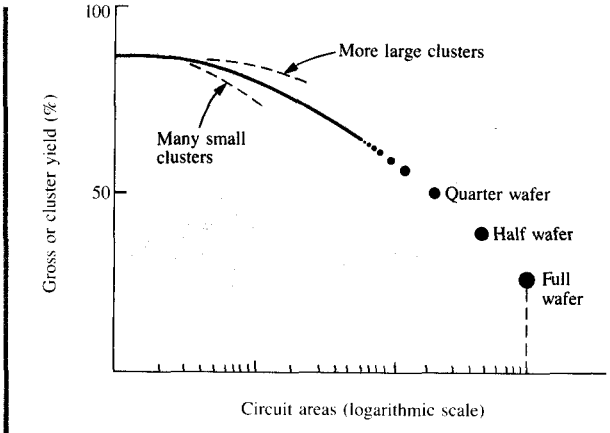

Figure 5

Cluster yield as a function of integrated circuit areas.

sample. This procedure is still referred to as "declustering." We therefore call this approach the declustering method.

After the clusters were removed, the remaining failing monitors were assumed to be randomly distributed according to simple Poisson statistics in each region. To determine the wafer to wafer variation, the distribution of the number of failing monitors in corresponding regions was obtained. This therefore is the same approach as the one described in the preceding section. The difference in this case is that smaller regions were used instead of full wafers.

There was another difference between the method described in the previous section and the one adopted in [8]. The samples in the latter paper were too small to determine whether the wafer to wafer variations of defect densities followed a gamma distribution. To circumvent this, the mean and the variance of the failing number of monitors in corresponding regions were calculated. When the mean was equal to or larger than the variance, simple Poisson statistics were assumed for the random defect model. When the mean was smaller than the variance, the negative binomial statistics in (44) were assumed to be appropriate. The parameters for these statistics were then determined with (47a, b) and the yield calculated with (52). Modified versions of this technique survive today in a number of integrated circuit manufacturing lines at IBM in Essex Junction.

Yield models with more than two regions have been used to calculate yields at IBM. At the development laboratory in East Fishkill, New York, K. Saji uses a model with three regions per wafer. A yield model used to control the photolithographic defects in one of the manufacturing lines in Essex Junction makes use of five regions per wafer.

The number of regions is usually determined by the nature of the defects. Serious clustering and regional variations require more regions. The shape of the regions also depends on the nature of the clusters. Most tools produce a radial defect density variation. Concentric regions are therefore applicable in many cases. Angular regions, however, also have been used in the past.

12. Large chips and wafer-scale integration

Implicit in the models derived and described in this paper is the assumption that the chip areas are smaller than the cluster and the regional areas. This does not mean that these models cannot be used to calculate yields for larger chips. This can best be demonstrated with an example.

A chip with area A that is larger than a regional area s_i must contain more than one region or a number of fractions of regions. The areas or fractional areas of the regions within the chip area A must be used individually to calculate the corresponding yield. In this calculation the appropriate form of the wafer to wafer yield formula (51) must be used. The resulting yields for these regions or fractional regions must be multiplied to obtain the total random defect yield for the chip. In the case of full wafers it consists simply of calculating the yield for each wafer region and multiplying the yields of all the regions on a wafer.

This is the entire random defect model. It has been in use in IBM Essex Junction since October 1972. The model was used by A. N. McLaren to estimate the yields of wafer-scale integrated circuit products for cost and productivity calculations. These estimates have aided IBM management in making the correct business decisions about the economic viability of such products.

Until now we have only looked at half the data that were collected with either the method of Paz and Lawson [9] or the declustering technique of [8]. The missing half is the gross clustering yield Y_0. As it turns out, this has a very profound effect on the yield of very large chips and full wafers.

The gross yields Y_0 in [9] consisted of cluster areas with extremely high density pipe defects. The data in [9] show an average yield for Y_0 of approximately 81% in the inner regions and about 75% in the outer regions. These are the gross yields that have to be used in the yield calculations for smaller chips. The values of Y_0 remain approximately constant for chip sizes that are smaller than the cluster areas. Chips that are the same size as, or larger than, the wafer regions must be completely free of such clusters. These large chips will therefore have a lower cluster yield than Y_0. A cumulative curve of Y_0 in [9] shows that this yield falls somewhere between 0 and 50% for the inner regions. No data are supplied for the outer regions.

The data in [8] are somewhat more specific. They show a cluster-limited yield of 97.3% for inner regions and 73.8% for the outer regions. The cluster-limited yield for the combined regions was given as 87.2%. All of these yields are pertinent for chips that are smaller than the cluster areas.

There is a distribution depicted in [8] for the number of monitors lost per wafer due to clustering. This distribution shows that only 30% of the wafers are without clusters. The cluster-limited yield for full wafers is therefore 30%. The

cluster yield consequently ranges between 87.2% for small chips and 30% for full wafers in these data. This includes open and short circuits among patterns in diffusions, polysilicon and metal only. It does not include the yield of oxide pinholes, diffusion leakage, or contact holes.

These cluster-limited yields have usually been described under the name "gross yield" by this author and sometimes as "area yield" by others. A discussion of the types of defects and manufacturing errors included in these yields is given in [12]. According to the same paper, these defects and errors cause parts of wafers or entire wafers to have no functioning chips. This implies that the chip size is smaller than the affected areas. For very large chips and full wafers this assumption no longer applies and other models have to be developed.

A yield versus area plot for cluster yields is shown in **Figure 5**. The two end points are those measured in [8]. The flat portion on the left indicates the constant Y_0 range for small chips. The extent of the flatness depends entirely on the area of the smallest clusters that are considered. The curvature of the yield plot for larger chips depends on the size distribution of the clusters. If there are a large number of small clusters, the roll-off will be sharp. For a distribution with more large clusters, the yield curve will fall off slowly. These conditions are shown with dashed lines in Fig. 5. This tells us that the distribution of cluster sizes has to be taken into account in the yield projections of large chips and full wafers. It is this effect that modulates the random defect yields, not the rejection of Formula (51) as was done in [16, 17].

A yield versus area plot in **Figure 6** indicates the range of yields that can be expected in the case of full wafers. The upper curve is the one that occurs in a manufacturing facility that succeeds in constantly making some defect-free wafers. The bottom curve is for a manufacturer who never makes a defect-free wafer. Manufacturing operations that do not produce 100% yield wafers in their current products all fall into this category.

Proponents of wafer-scale integration appear to advocate the use of simple Poisson statistics for calculating wafer yields, as was done, for instance, by Peltzer at Trilogy [18]. He mistakenly refers to Stapper as the source of that model. It is, however, the model proposed by single-wafer analysts and theoreticians, as for example in [11, 16, 17]. Their model clearly predicts too high a yield for factories that currently cannot manufacture defect-free wafers.

The cause of the zero yields in the lower curve in Fig. 6 can be determined experimentally. In data examined by this author it appears in the form of high defect levels in localized areas on wafers. These are the same areas that were modeled with the gross yields Y_0 by Paz and Lawson and in the declustering model. We must therefore conclude that in such a manufacturing environment the cluster yield curve in Fig. 5 goes to zero for circuit areas smaller than wafer size.

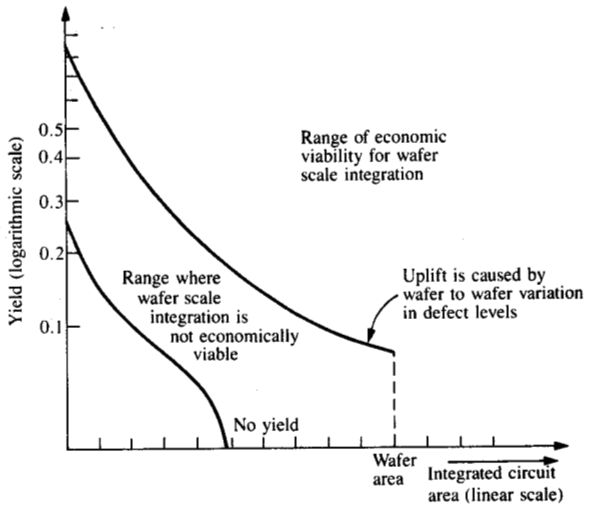

Figure 6

Yield as a function of integrated circuit area or circuit complexity. The yield axis is logarithmic.

It has been suggested [16] that the yield in the zone between the upper and lower curves of Fig. 6 can be modeled with simple Poisson statistics. The data from Paz and Lawson clearly show that this assertion is wrong.

In the method of Paz and Lawson [9], the introduction of the cluster yield Y_0 made it possible to model the remaining random defects with simple Poisson statistics. It was nevertheless an artifact, since the clusters did contain high density random defects. These, therefore, should have been included in a random defect model that does not follow Poisson statistics. An early model of this type has already been described in [19].

The fact that simple Poisson statistics is not applicable to wafer-scale integration has another important consequence. For a given yield, the Poisson model has a relatively low average number of faults. For the same yield, a negative binomial model has a much larger average number of faults. This is in agreement with actual data in [6, 7]. It is this effect that has a pronounced effect on the requirement of redundant circuits when redundancy is used to boost the yield.

Defect clusters within the chip or wafer area increase the need for redundant circuits even more. Advocates of Poisson statistics and simple Poisson yield models, to the contrary, predict the necessity for fewer redundant circuits. Miscalculations of this type can lead to serious product problems. Manufacturers of wafer-scale products who plan their redundancy requirements with Poisson models may be in for a nasty surprise.

A redundancy model that incorporates negative binomial statistics has been described in [6]. This model has been used at IBM to optimize the productivity of high density memory chips since 1975. Yield projections for 256K-bit dynamic random access memory chips currently in production at IBM were made with this same model.

13. Conclusions

It has been shown in this paper that wafer to wafer variations in defect levels can be modeled with techniques developed at IBM in the early 1970s. This eliminates the deficiencies that occur in a number of recently published yield models [16, 17].

On the basis of published data, it is also apparent that defect clusters have an important influence on the yield of very large chips and wafer-scale integration. It appears that yield models for such products have to include cluster sizes and cluster size distributions.

Appendix 1

In Section 3 it was shown that the number of defects X_D per chip are distributed as

$$P(X_D = k) = \sum_{i=k}^{\infty} P(N = i) \frac{i!}{k!(i-k)!} (A/S)^k (1 - A/S)^{i-k}. \quad \text{(A1)}$$

Let $P(N = i)$ be given by the compound distribution

$$P(N = i) = \int_0^{\infty} \frac{e^{-SD}(SD)^i}{i!} P(D)dD. \quad \text{(A2)}$$

Substituting this in (A1) and interchanging the sum and integration signs gives

$$P(X_D = k) = \int_0^{\infty} \left[\sum_{i=k}^{\infty} \frac{e^{-SD}(SD)^i}{i!} \frac{i!}{k!(i-k)!} \times (A/S)^k (1 - a/S)^{i-k} \right] P(D)dD. \quad \text{(A3)}$$

The sum within the brackets is the same one as in Section 3. When an index $j = i - k$ is substituted, this sum becomes

$$\begin{aligned} &\sum_{j=0}^{\infty} \frac{e^{-SD}(SD)^{k+j}}{(k+j)!} \frac{(k+j)!}{k!j!} (A/S)^k (1 - A/S)^j \\ &= \frac{e^{-SD}(AD)^k}{k!} \sum_{j=0}^{\infty} \frac{[SD(1 - A/S)]^j}{j!} \\ &= \frac{e^{-SD}(AD)^k}{k!} e^{(SD-AD)} \\ &= \frac{e^{-AD}(AD)^k}{k!}. \end{aligned} \quad \text{(A4)}$$

When this result is introduced into (A3), that distribution becomes

$$P(X_D = k) = \int_0^{\infty} \frac{e^{-AD}(AD)^k}{k!} P(D)dD. \quad \text{(A5)}$$

Similarly, if $P(N = i)$ is given by the compounded distribution

$$P(N = i) = \sum_{l=0}^{\infty} P_l \frac{e^{-SD_l}(SD_l)^i}{i!}, \quad \text{(A6)}$$

then substitution into (A1) results in

$$P(X_D = k) = \sum_{l=0}^{\infty} P_l \frac{e^{-AD_l}(AD_l)^k}{k!}. \quad \text{(A7)}$$

The preceding results are very general and show that compounded Poisson distributions can be used for the compounder in (A1) and result in a compound distribution of the same type for the defect distribution. Wafer to wafer variations in defect levels can therefore be modeled with compound Poisson statistics.

Appendix 2

The binomial distribution (36) in the more general form

$$P(X = k) = \binom{N}{k} y^{N-k}(1 - y)^k \quad \text{(A8)}$$

can be compounded with a distribution $P(y)$ to give

$$P(X = k) = \binom{N}{k} \int_0^1 y^{N-k}(1 - y)^k P(y)dy. \quad \text{(A9)}$$

For this distribution the mean and variance are given by

$$E(X) = N[1 - E(y)], \quad \text{(A10)}$$

$$V(X) = NE(y)[1 - E(y)] + (N^2 - N)V(y). \quad \text{(A11)}$$

Consider a compounder equal to a beta distribution:

$$P(y) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} y^{\mu-1}(1 - y)^{\nu}, \quad \text{(A12)}$$

for which

$$E(y) = \mu/(\mu + \nu), \quad \text{(A13)}$$

$$V(y) = \mu\nu/(\mu + \nu)^2(\mu + \nu + 1). \quad \text{(A14)}$$

When (A12) is introduced into (A9), it results in

$$P(X = k) = \binom{N}{k} \frac{\Gamma(\mu + \nu)\Gamma(\nu + k)\Gamma(\mu + N - k)}{\Gamma(\mu)\Gamma(\nu)\Gamma(\mu + \nu + N)}. \quad \text{(A15)}$$

This has the mean and variance given by

$$E(X) = N\nu/(\mu + \nu), \quad \text{(A16)}$$

$$V(X) = N\mu\nu(N + \mu + \nu)/(\mu + \nu)^2(\mu + \nu + 1). \quad \text{(A17)}$$

The parameters of the distribution (A15) can be estimated from the mean $\bar{k}$ and standard deviation σ_k of the data with

$$\nu = \bar{k}C, \tag{A18}$$

$$\mu = (N - \bar{k})C, \tag{A19}$$

where

$$C = \frac{(N - \bar{k})\bar{k} - \sigma_k^2}{(N - \bar{k})\bar{k} + N\sigma_k^2}. \tag{A20}$$

Appendix

The binomial distribution (36) can also be written as

$$P(X = k) = \binom{N}{k} e^{-(N-k)A_cD}(1 - e^{-A_cD})^k, \tag{A21}$$

where A_c is the critical area and D a defect density. The factor $(1 - e^{-A_cD})^k$ can be expanded with the binomial expansion so that

$$P(X = k) = \binom{N}{k} e^{-(N-k)A_cD} \sum_{i=0}^{k} (-1)\binom{k}{i} e^{-iA_cD}. \tag{A22}$$

This can be compounded with a defect density distribution $P(D)$ to give

$$P(X = k) = \binom{N}{k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} \int_0^\infty e^{-(N+i-k)A_cD} P(D)dD. \tag{A23}$$

When $P(D)$ is a gamma distribution with parameters α and β, as in (40), then

$$P(X = k) = \binom{N}{k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} [1 + (N + i - k)\beta]^{-\alpha}. \tag{A24}$$

It can be useful to substitute $\beta = \bar{D}/\alpha$ into (A24).

References

1. A. B. Glaser and G. E. Subak Sharpe, *Integrated Circuit Engineering*, Addison-Wesley Publishing Co., Reading, MA, 1977, Ch. 6.
2. C. H. Stapper, "Comments on 'Some Considerations in the Formulation of IC Yield Statistics,' " *Solid-State Electron.* **24,** 127–132 (February 1981).
3. J. Neyman, "On a New Class of 'Contagious' Distributions, Applicable in Entomology and Bacteriology," *Ann. Math. Stat.* **10,** 35–57 (1939).
4. W. Feller, "On a General Class of 'Contagious' Distributions," *Ann. Math. Stat.* **14,** 389–399 (1943).
5. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, John Wiley & Sons, Inc., New York, 1971, p. 57.
6. 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).
7. C. H. Stapper, "Modeling Redundancy in 64K to 16 Mb DRAMs," *Digest of Technical Papers*, 1983 IEEE International Solid State Circuit Conference, February 1983, pp. 86, 87.
8. C. H. Stapper, "LSI Yield Modeling and Process Monitoring," *IBM J. Res. Develop.* **20,** 228–234 (May 1976).
9. 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).
10. R. M. Warner, Jr., "Applying a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-9,** 86–95 (June 1974).
11. R. M. Warner, Jr., "A Note on IC Yield Statistics," *Solid-State Electron.* **24,** 1045–1047 (December 1981).
12. C. H. Stapper, F. M. Armstrong, and K. Saji, "Integrated Circuit Yield Statistics," *Proc. IEEE* **71,** 453–470 (April 1983).
13. F. M. Armstrong, IBM Data Systems Division, Poughkeepsie, NY, and K. Saji, IBM General Technology Division, East Fishkill, NY, private communication.
14. C. H. Stapper, "Defect Density Distribution for LSI Yield Calculations," *IEEE Trans. Electron Dev.* **ED-20,** 655–657 (July 1973).
15. S. M. Hu, "Some Considerations in the Formulation of IC Yield Statistics," *Solid-State Electron.* **22,** 205–211 (February 1979).
16. S. M. Hu, "On Yield Projection for VLSI and Beyond. I. Analysis of Yield Formulas," *Electron Devices*, No. 69, pp. 4–7 (March 1984). [Issued by the IEEE Electron Device Society; Order No. USPS 857-240.]
17. *Ibid.*
18. D. L. Peltzer, "Wafer Scale Integration: The Limits of VLSI?" *VLSI Design* **4,** 43–47 (September 1983).
19. C. H. Stapper, "Yield Model for Fault Clusters Within Integrated Circuits," *IBM J. Res. Develop.* **28,** 636–640 (September 1984).

Received May 29, 1984; revised August 21, 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.