Correlation analysis of particle clusters on integrated circuit wafers

by C. H. Stapper

Defect clustering results in correlations between the numbers of defects or faults that occur on integrated circuit chips located adjacent to one another on semiconductor wafers. Until now, it has been believed that correlations of this type were not accounted for in existing yield models. It is shown in this paper that such correlations are present in yield models based on mixed or compound Poisson statistics. A quadrat analysis of particle distributions on semiconductor wafers is used to compare data and theory. The results show that the theoretical correlation coefficients are in agreement with the experimental ones. It was also determined from the particle data how these correlation coefficients vary as the distance between quadrats is increased. This variation provides a convenient method for determining the cluster dimensions.

1. Introduction

Defective integrated circuit chips often occur next to one another on semiconductor wafers. Similarly, functionally usable chips are also regularly found adjacent to one another. This is the result of defect clustering. Such clustering can lead to a correlation between the numbers of defects found in adjacent chips. Until now it has been assumed that such correlations did not occur in existing yield models. This assumption, however, is incorrect. Correlations of this nature are an integral part of any yield model based on mixed or compound Poisson statistics. It is shown here why this is so.

The need for using compound Poisson statistics originates from three effects. The first of these is the wafer-to-wafer variation of defect densities. It has been known for a long time that the distribution of the number of defects per wafer can best be represented with some form of compound defect statistics [1]. The second effect deals with the radial variation of chip yields. Defective chips are more likely to occur near the edges of wafers [2–6]. This effect has been modeled mathematically by partitioning the wafers into concentric circular zones. The use of two zones, an inner and an outer one, has been reported previously [7, 8].

Localized defect clustering is the third effect requiring the use of compound Poisson statistics. An example of this is found in the crystalline defects that occur in epitaxial silicon. This material is used in the manufacture of integrated circuits containing bipolar transistors. The defects cause short circuits between the emitters and collectors of such transistors. Another example of localized clustering is found in dirt particles that occasionally occur in the integrated circuit manufacturing processes. Such particles usually are the cause of defects in dielectric insulator materials or photolithographic patterns. Those defects therefore also have

a tendency to cluster. Models for this localized clustering have also made use of partitioned regions [9–11]. This partitioning, however, is more complex than in the case of radial variations of the yield.

Spatial distributions of defects and faults on semiconductor wafers are not readily available, which makes it difficult to verify correlations between adjacent chips. Correlations of this type can, however, be evaluated with maps showing the location of particles observed on integrated circuit wafers. This is done by overlaying the maps with a grid of squares known as quadrats. The number of particles per quadrat can then be counted and the statistics of these counts can be compared to those of existing theories. This technique is used here to study the correlations between the numbers of particles in adjacent quadrats. It results in correlation coefficients that agree with those calculated using the theory presented in this paper.

2. Random fault statistics

The key to the success of any integrated circuit manufacturing operation is the fraction of manufactured semiconductor chips that can be sold to the customers. This fraction is known as the yield. The yielding chips have to survive a battery of device tests. Manufacturers learned very early that chips which failed tests for one application could be perfectly usable for another one. This led to the practice of sorting chips.

The subtle differences between defects that affect final test sorts must be taken into account by theoreticians who want to help manufacturing lines achieve optimum production. A method that has proven very effective makes use of the terminology "faults." A fault is defined as a defect that causes an integrated circuit to fail when it is tested for a specific application. The statistics of such faults are of great importance to manufacturers of integrated circuits.

In this paper it is assumed that for purely random faults the distribution of the number of faults in integrated circuits follows Poisson's distribution. That distribution can be expressed in the form

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{1}$$

where X is a random variable designating the number of faults per chip, k an integer equal to 0, 1, 2, $\cdots$, etc., and λ a parameter.

Consider next two chips where the random number of faults on one is indicated by X and on the other by Y. Let the faults occur on these chips independently. The probability of finding $X = k$ faults on the first chip and $Y = l$ faults on the second one therefore is given by

$$\Pr(X = k,\ Y = l) = \frac{\lambda^{k+l}}{k!\ l!} e^{-2\lambda}. \tag{2}$$

This is a joint probability for two chips. It is valid even if these chips are not adjacent to each other. They must, however, occur as a pair in a region where the numbers of faults per chip are distributed with Poisson's distribution, having an average density of λ faults per chip. If the random variables X and Y are independent, it is not difficult to prove that their covariance and correlation coefficient are equal to zero.

3. Mixed and compound Poisson statistics

The wafer-to-wafer variation of the average number of faults per chip, and the spatial clustering of defects and faults, can both be modeled with compound Poisson statistics. Using the method of Feller [12], it was shown in Reference [13] that this leads to an integral of the form

$$\Pr(X = k) = \int_0^1 \frac{\lambda^k}{k!} e^{-\lambda} dF(\lambda). \tag{3}$$

The function $F(\lambda)$ is known as the compounder and satisfies the boundary conditions $F(0) = 0$ and $F(\infty) = 1$. It represents the cumulative distribution of the average number of faults per chip λ. Each one of these average values pertains to a wafer or a region on a wafer.

For a finite number of regions, the function $F(\lambda)$ is a staircase function. The number of steps in the staircase is equal to the number of regions. Use of such a function leads to a discrete form of compounding and results in the formula

$$\Pr(X = k) = \sum_{i=1}^{m} c_i \frac{\lambda_i^k}{k!} e^{-\lambda_i}, \tag{4}$$

where m designates the number of regions, λ_i the average number of faults per chip in region i, c_i a set of constants, and k an integer 0, 1, 2, 3, $\cdots$, etc.

The constants c_i in Equation (4) are a set of discrete compounders. They are equal to the probability of finding a randomly picked chip within a region associated with the index i. These probabilities are proportional to the areas of the partitioned regions, and their sum must satisfy

$$\sum_{i=1}^{m} c_i = 1. \tag{5}$$

This is the normalization requirement for the compounders.

It has also been observed that partitioning is necessary because of continuously varying manufacturing conditions. This results in wafer-to-wafer variations of fault densities. The fault and defect distributions described by Paz and Lawson [8] are the results of such effects. Wafer-to-wafer variations of the average number of faults per wafer have also been discussed by this author in Reference [14]. Such variations can be modeled by increasing the number of steps of the staircase function $F(\lambda)$ in the limit to infinity. Because of this, the function becomes continuous, so that

$$\Pr(X = k) = \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} f(\lambda) d\lambda, \tag{6}$$

where

$$f(\lambda) = \frac{\partial F(\lambda)}{\partial \lambda}.$$

This function $f(\lambda)$ is a normalized continuous probability distribution function of fault densities λ.

4. Covariance between the numbers of faults on chips

The method of mixing or compounding Poisson's distribution can be extended to pairs of chips. To begin with, assume that all chip pairs come from regions where Poisson's distribution is valid. This means that different pairs can come from different regions, but the two chips in a pair cannot. The method of compounding can then be applied directly to the joint distribution in Equation (2). This results in

$$\Pr(X = k,\ Y = l \mid X \text{ and } Y \text{ in the same region}) = \int_0^1 \frac{\lambda^{k+l}}{k!\ l!} e^{-2\lambda} dF(\lambda). \quad (7)$$

This is a bivariate distribution for the random variables X and Y. They are equal to k and l, which represent the integer values 0, 1, 2, 3, $\cdots$, etc. Equation (7) is a conditional probability to indicate that the faults X and Y occur jointly in the same region.

The cumulative distribution function $F(\lambda)$ in Equation (7) is not necessarily equal to the compounder in Equation (3). This depends on the distribution of the numbers of chip pairs per region and the distribution of the numbers of chips per region. If these two distributions are the same, the two compounders are equal to each other.

It is shown now how the covariance between the numbers of faults per chip can be determined. In general, the covariance between two random variables X and Y can be calculated with the formula

$$\operatorname{cov}(X,\ Y) = E(XY) - E(X)E(Y). \quad (8)$$

The key to this expression is the expectation $E(XY)$. It can be obtained by a double summation over distribution (7). This takes the form

$$E(XY \mid X \text{ and } Y \text{ in the same region}) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} kl \int_0^1 \frac{\lambda^{k+l}}{k!\ l!} e^{-2\lambda} dF(\lambda). \quad (9)$$

The expression is conditional to emphasize that both chips are in the same region. The summations can be evaluated and reduced to

$$E(XY \mid X \text{ and } Y \text{ in the same region}) = \int_0^1 \lambda^2 dF(\lambda). \quad (10)$$

The right-hand side of this expression is equal to the expectation $E(\lambda^2)$. In general, it therefore follows that $E(XY \mid X \text{ and } Y \text{ in the same region}) = E(\lambda^2)$.

In a similar way, the mathematical expectations $E(X)$ and $E(Y)$ are given by

$$E(X) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} k \int_0^1 \frac{\lambda^{k+l}}{k!\ l!} e^{-2\lambda} dF(\lambda) \quad (11)$$

and

$$E(Y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} l \int_0^1 \frac{\lambda^{k+l}}{k!\ l!} e^{-2\lambda} dF(\lambda). \quad (12)$$

In these equations the variables X and Y are treated individually. This eliminates the need for the conditional specification that was used in the previous formulas. When the sums in Equations (11) and (12) are evaluated, they both result in

$$E(X) = E(Y) = \int_0^1 \lambda dF(\lambda). \quad (13)$$

This is simply equal to the expectation $E(\lambda)$, and therefore represents the mean associated with the cumulative distribution function $F(\lambda)$.

The preceding results can be substituted into formula (8) for the covariance. This leads to

$$\operatorname{cov}(X,\ Y \mid X \text{ and } Y \text{ in the same region}) = E(\lambda^2) - E(\lambda)^2. \quad (14)$$

The right-hand side of this expression simplifies to the variance of λ, $V(\lambda)$. The result can therefore be expressed as

$$\operatorname{cov}(X,\ Y \mid X \text{ and } Y \text{ in the same region}) = V(\lambda). \quad (15)$$

This is the covariance between the number of faults occurring on one chip and the number occurring on another. It is an interesting result, since this quantity is simply equal to the variance associated with the compounding cumulative distribution function $F(\lambda)$.

The covariance obtained here is a two-dimensional spatial variant of the autocovariance used in stationary time series by Box and Jenkins [15]. These authors showed that such autocorrelations varied when the time lag between the correlated data was increased. Such an increase is akin to a larger separation between the chips that are used for the correlation. As long as the chips are in the same region, such increased separations have no effect on the correlation coefficient. When the distance between correlated chips causes the chips in the pairs to occur on both sides of regional boundaries, some interesting effects take place. These effects are discussed in subsequent sections.

5. The correlation coefficient

At this point it is possible to determine an expression for the correlation coefficient relevant to the compound Poisson statistics discussed in the preceding sections. In general, the

correlation coefficient ρ between two random variables X and Y is related to the covariance with the well-known formula

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}. \tag{16}$$

The variances $V(X)$ and $V(Y)$ in this expression are equal to one another because the random variables X and Y have the same distribution. Either one of these variances can therefore be calculated with the formula

$$V(X) = E(X^2) - E(X)^2. \tag{17}$$

The expectation $E(X^2)$ in this expression is obtained with the double sum

$$E(X^2) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} k^2 \int_0^1 \frac{\lambda^{k+l}}{k!\, l!} e^{-2\lambda}\, dF(\lambda). \tag{18}$$

These summations can be evaluated by the method described in Appendix A of Reference [7] to give

$$E(X^2) = E(\lambda^2) + E(\lambda). \tag{19}$$

Here $E(\lambda^2)$ and $E(\lambda)$ are, as they were in the preceding section, expectation values associated with the compounder $F(\lambda)$. These expectations can be introduced into Equation (17). Furthermore, according to Equation (13), the expectation value $E(X)$ is equal to the expectation $E(\lambda)$. Equation (17) therefore becomes

$$V(X) = E(\lambda^2) + E(\lambda) - E(\lambda)^2. \tag{20}$$

At this point it must be recognized that the variance of the compounder $F(\lambda)$ has the form

$$V(\lambda) = E(\lambda^2) - E(\lambda)^2. \tag{21}$$

When this is combined with Equation (20), it results in

$$V(X) = V(\lambda) + E(\lambda). \tag{22}$$

The variance $V(Y)$ is equal to the same formula.

The correlation coefficient is obtained by substitution of the preceding results into Equation (16). This produces

$$\rho(X, Y \mid X \text{ and } Y \text{ in the same region}) = \frac{V(\lambda)}{V(\lambda) + E(\lambda)}, \tag{23}$$

which shows that this correlation is completely dependent on the expectation $E(\lambda)$ and the variance $V(\lambda)$ associated with the compounder $F(\lambda)$. It should be noted that this correlation coefficient becomes larger when the ratio of the variance and the mean of the compounder increases. Any suitable distribution function can be used as compounder, thus illustrating the generality of this approach.

Suppose that wafers can be manufactured in such a way that the defects on them are purely random and follow Poisson's distribution. Adjacent chip pairs are then automatically constrained to the same regions, namely the wafers themselves. The wafer-to-wafer variation of defect densities then defines the compounding process. This compounding affects the chip pairs and the single chips equally. In that case, the compounders in Equations (3) and (7) are therefore the same. Furthermore, the conditionality constraining the chip pairs to the same region is implicitly satisfied and can be dropped from the equations of the preceding sections. Moreover, the covariance and correlation are completely dependent on the mean and variance of the wafer-to-wafer variation of defect densities.

Formula (23) is a novel result. It was believed previously that correlations of this type were not accounted for in existing yield models [16]. As is shown here, correlations between the number of faults or defects in adjacent chips are an inherent consequence of the compounding process. It does not require the forcing of correlations, as has been done elsewhere [16].

6. Wafer-to-wafer variation: An example

The wafer-to-wafer variations of defect densities were originally studied with electronic defect monitors [1]. The numbers of defects and faults in integrated circuits on actual wafers are, however, much more difficult to obtain. Until now, data for only half a wafer have appeared in the literature [17]. To circumvent this lack of data, statisticians have resorted to the use of water maps of particle locations [18]. Such maps can be obtained with electronic particle detectors which, by use of scattered light, detect the locations of particles on the surfaces of wafers. Particles often cause defects and faults in integrated circuits; particle distributions should therefore give an insight into the distributions of defects and faults.

A particularly useful set of particle maps were collected by Armstrong and Saji specifically for the evaluation of yield models [18]. The data consisted of the twelve square maps that are reproduced in **Figure 1**. These maps were analyzed with the so-called "quadrat" method. This method was developed by biologists for studying the spatial distributions of populations of species. The area to be studied is subdivided with a grid of squares known as "quadrats." The population within each square is then counted to obtain a population distribution.

The particle maps in Figure 1 contain, in ascending order, 9, 17, 22, 34, 39, 41, 42, 92, 102, 107, 108, and 140 particles. When an overlay grid of 6 × 6 quadrats is used for each map, the number of particles for each wafer can be divided by 36 to obtain an average number of particles per quadrat for each wafer. This results in a lowest particle density of 0.25 and a highest density of 3.8889 particles per quadrat. These values and those from the other wafers form a distribution. This distribution has a grand mean of 1.7431 particles per quadrat and a variance of 1.3700.

The wafer-to-wafer variation of particle densities corresponds to the wafer-to-wafer variation of the fault and

defect densities per chip that was discussed in the preceding sections. If we assume that the particles are distributed randomly on each wafer, we can then use formula (23) to calculate the theoretical correlation coefficient for the number of particles in adjacent quadrats. Doing so results in a value of 0.440.

It is also possible to determine the correlation coefficient from actual particle counts in each quadrat. These particle counts are tabulated as twelve arrays of numbers shown in **Figure 2**. Each one of these arrays represents a wafer, and each number in an array shows how many particles were counted within the quadrat at that location. The correlation coefficient for the number of particles in adjacent quadrats can be calculated from these data. This has to be done with a pairs analysis. The number of particles of each quadrat is to be correlated with the particle counts in the other quadrat of a pair. Double counting is not allowed. For example, if the number of particles in quadrat A is correlated with the number in quadrat B, then quadrat B should not be

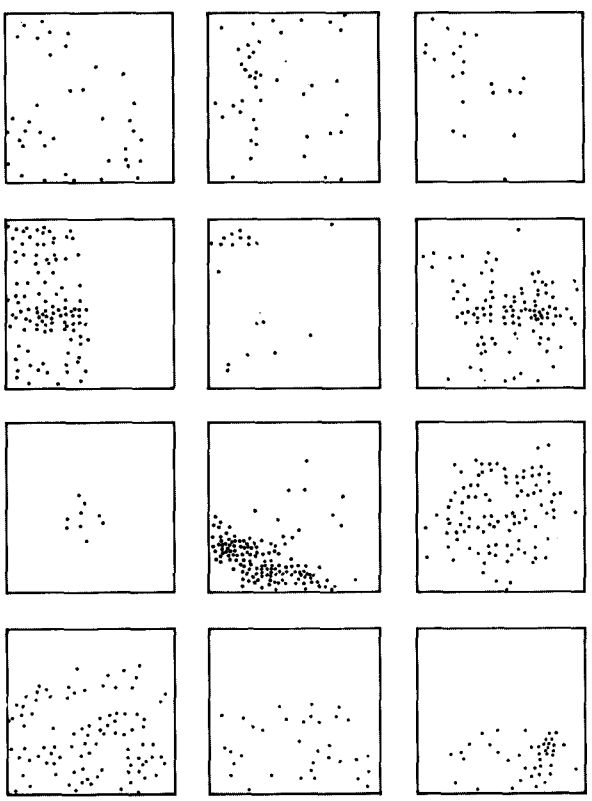

Figure 1

Particle maps obtained with an electronic particle detector.

2	3	1	0	0	0	1	1	2	1	3	0	1	2	1	0	0	0
0	1	1	1	0	0	0	6	0	0	0	0	3	3	0	0	0	0
0	0	2	0	1	0	0	4	1	2	1	0	0	2	2	3	0	0
2	1	0	1	2	0	3	4	0	0	1	0	0	1	0	0	0	0
4	3	0	0	3	0	0	3	0	1	2	0	0	2	0	1	0	0
2	1	2	2	4	0	1	1	0	1	2	0	0	0	0	1	0	0
7	8	3	0	0	0	5	5	0	0	1	0	0	0	0	1	0	0
3	8	2	0	0	0	1	0	0	0	0	0	3	2	5	0	2	0
6	4	4	0	0	0	0	0	0	0	0	0	0	5	4	6	9	2
8	18	15	0	0	0	0	1	0	0	0	0	0	4	9	12	17	4
3	0	6	0	0	0	0	1	1	1	0	0	0	0	7	2	4	1
4	4	5	0	0	0	1	0	0	0	0	0	0	1	2	3	2	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
0	0	0	0	0	0	0	0	0	1	0	0	0	2	7	5	5	0
0	0	2	0	0	0	0	0	1	1	1	0	0	7	4	5	7	1
0	0	4	2	0	0	6	1	1	1	2	0	2	9	7	8	4	1
0	0	1	0	0	0	28	20	10	0	0	0	1	4	6	6	5	0
0	0	0	0	0	0	1	20	22	19	4	1	0	0	2	2	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	4	2	1	2	0	0	1	1	1	0	0	0	0	0	0	0
5	1	4	8	6	0	1	3	1	3	2	0	0	0	2	1	2	0
4	7	4	1	7	4	3	1	0	3	3	2	0	4	2	2	14	1
3	1	6	2	6	1	1	1	1	3	0	3	0	1	1	3	8	1

Figure 2

Particle counts resulting from a quadrat analysis of the maps in Figure 1. Each number represents the number of particles found in one quadrat.

correlated again with quadrat A. For nearest-neighbor pairs this results in thirty horizontal and thirty vertical pairs on each map.

By using the maps of Figure 2, the correlation coefficient of the number of particles in adjacent quadrats was found to have a value of 0.5190. This result has 95% confidence limits of 0.42 and 0.56. The theoretical correlation coefficient of 0.44 falls comfortably between these two limits. Differences between the experimental and theoretical correlation coefficients can also be the result of the localized particle clustering. The maps in Figure 2 indicate that many quadrats with particles cover contiguous areas. In the same way, many quadrats without particles are adjacent to one another. The particles therefore appear to be clustered. As mentioned before, this is typical for particles and defects in

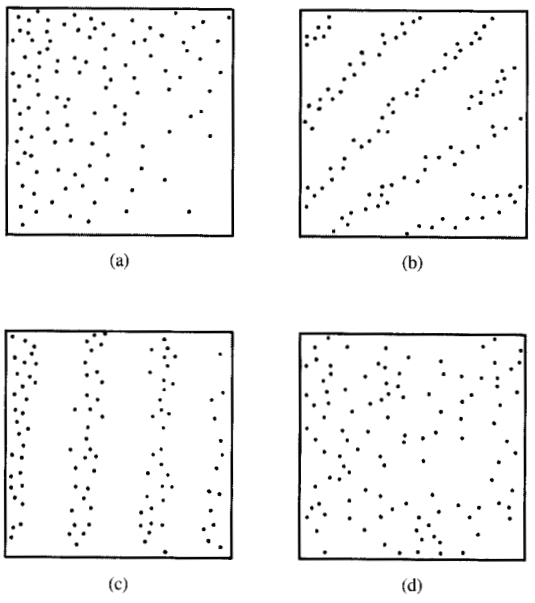

Figure 3

Contrived maps with dots, generated to illustrate various phenomena that affect the correlation coefficients for dot counts in adjacent quadrats. Each map contains the same number of dots.

Table 1 Frequency distribution of the number of dots per quadrat for the maps in Figure 3. The Poisson distribution with the same mean value is also shown.

Number of dots per quadrat	*Number of occurrences*	*Relative frequency*	*Poisson's distribution*
0	2	0.0556	0.0498
1	5	0.1389	0.1494
2	8	0.2222	0.2240
3	8	0.2222	0.2240
4	6	0.1667	0.1680
5	4	0.1111	0.1008
6	2	0.0556	0.0504
7	1	0.0278	0.0216
8			0.0081
9			0.0027
10			0.0008
11			0.0002
12			0.0001

integrated circuit manufacturing processes. Such localized clustering also results in a correlation between the numbers of particles in adjacent quadrats. An illustrative example of this effect is the subject of the next section.

7. Spatial clustering: An example

In Reference [18] the ratio of the variance and the mean of particle distributions was used to measure the deviation of data from theoretical Poisson distributions. This is a useful measurement when wafer-to-wafer variations dominate. It is, however, an insufficient condition for localized fault, defect, or particle clusters on the wafers themselves. This can best be illustrated with a contrived example.

Four square maps, each with 108 dots, are shown in **Figure 3**. The dots are arranged differently in each of the maps. The distributions of these dots were analyzed with the quadrat method by superimposing a grid of 6×6 squares over the maps. The number of dots found in each quadrat is shown in **Figure 4**. The maps were made up in such a way that the distribution of the number of dots per quadrat is the same for all four maps. The result is tabulated in **Table 1**. The mean of this distribution is equal to three dots per quadrat, and its variance is equal to 2.8333.

The distribution in Table 1 can be modeled with Poisson's distribution of formula (1). A value of three for λ provides a model that has a significance level >0.995 when analyzed with a chi-square test. The applicability of Poisson's distribution does not imply that the spatial distribution of the dots is homogeneously random. The human eye suggests that the dots in Figure 3 are clustered in different ways. In map 4(a) of Figure 4, the density of dots appears to be higher in the upper left-side corner than in the lower right-side corner. In map 4(b) the dots form striations diagonally across the map, while in many 4(c) the striations are vertical. The dots in map 4(d) can best be described as having a snowflake pattern.

The distribution of the number of dots per quadrat for these maps has a ratio of the variance and the mean that is smaller than one. Ratios greater than one are usually associated with clustering. Therefore, the ratio obtained here gives no indication of clustering. A better test than this ratio is necessary to measure the effect of these spatial clusters. The correlation between the numbers of dots in adjacent quadrats provides such a test. Correlation coefficients obtained in this way are shown in **Table 2**. These results were obtained from the data in Figures 3 and 4. The correlation coefficients for horizontally and vertically adjacent quadrats are shown separately in the two middle columns. The correlation coefficients for horizontal and vertical pairs combined are tabulated in the last column.

Completely random dots can be expected to have a correlation coefficient of zero for the number of dots in adjacent quadrats. The correlation coefficients for the snowflake pattern in map 4(d) come closest to this criterion. As shown in Table 2, for this case, the absolute values of the correlation coefficients for the horizontal, vertical, and combined pairs are all less than 0.1. In map 4(c), the coefficient obtained for the combined horizontally and vertically aligned pairs has a small negative value. In this

case, however, the horizontally aligned pairs produced a substantial negative correlation coefficient, while the vertically aligned pairs have a large positive one. These differences in sign are caused by the vertical striations of dots. When such striations run diagonally across the map, as they do in 4(b), the correlation coefficients are negative for both the horizontally and vertically aligned quadrat pairs. This contrasts with the results for map 4(a), for which all correlation coefficients are positive.

The maps in Figure 3 were made up to illustrate, by accentuation, the different forms of dot clustering that can affect the correlation coefficients of the number of dots in adjacent quadrats. Similar types of clustering have been observed for faults and defects in integrated circuit chips on semiconductor wafers, albeit with lower intensity (lower correlation coefficients). For example, the varying dot densities in map 4(a) are analogous to the radial variations of defect densities discussed by Yanagawa [2, 3] and Gupta et al. [4, 5]. Striation of defects can be caused by scratches [9], or by streaking of cleaning solutions, chemicals, or photoresist materials [19]. The occurrence of highly defective regions adjacent to defect-free or nearly defect-free regions can also be observed in the data published by Paz and Lawson [8].

The dots in Figure 3 were positioned in such a way that the correlations for a 6×6 grid of quadrats were maximized. A change to different quadrat sizes or different quadrat locations will change the results. The absolute magnitude of the correlation coefficients can be expected to decrease. In actual data, the locations of the clusters are usually random with respect to the quadrat locations. Similarly, the locations of fault clusters are random with respect to chip locations. This can therefore be expected to lower the correlation coefficients when this test is applied to real data.

Another effect that results in lower correlation coefficients is caused by large samples. In that case the positive and negative correlation coefficients from the localized clusters and striations tend to cancel each other. A somewhat artificial example of this is found with the dot maps of Figure 3. It occurs when the correlation coefficient is calculated for all the horizontally and vertically aligned pairs from all four maps combined. The correlation coefficient in that case is −0.02. This is low enough to indicate that there is no correlation. It illustrates, therefore, that low values for these correlation coefficients are not necessarily an indication of a lack of clustering. It also suggests that adding more quadrats to a sample can diffuse the local correlations that are caused by the localized clusters. This methodology therefore appears to be more effective when the correlation is performed on smaller samples of quadrats, preferably those that contain the localized clusters. Moreover, quadrat orientation and dimensions can be expected to influence results. Variable quadrat sizes and locations may prove to be useful for studying actual cluster data.

7	6	5	4	4	3
6	5	4	3	3	2
5	4	3	2	2	2
5	4	3	2	1	1
4	3	3	2	1	0
3	2	2	1	1	0

(a)

7	0	6	1	5	1
0	5	1	4	2	4
6	1	5	2	4	3
1	4	2	3	2	3
5	2	4	2	3	3
2	4	3	3	3	2

(b)

7	0	6	1	5	1
6	0	5	1	4	2
5	1	4	1	3	2
5	2	4	2	3	2
4	2	3	3	3	3
3	3	2	4	2	4

(c)

5	1	4	1	1	5
6	3	6	2	4	4
3	4	3	3	2	2
2	5	2	1	0	3
5	3	4	3	3	4
2	1	2	7	2	0

(d)

Figure 4

Dot counts resulting from a quadrat analysis of the dot patterns in Figure 3. For each map the dot counts per quadrat have identical frequency distributions, which can be approximated with Poisson's distribution.

Table 2 Correlation coefficients associated with the number of dots per quadrat for the maps in Figures 3 and 4.

Map	*Horizontal*	*Vertical*	*Combined*
4(a)	0.9608	0.9587	0.9595
4(b)	−0.9305	−0.9271	−0.9286
4(c)	−0.9290	0.9055	−0.0333
4(d)	−0.0833	0.0649	−0.0109

8. Increasing the distance between quadrats

The correlations obtained in the previous sections dealt with adjacent quadrats. It is also possible to perform this exercise with quadrats separated by a larger distance. Such an increase in distance results in a high chance of finding quadrat pairs that straddle different regions. As a consequence, the correlation coefficient can be expected to decrease when the distance between quadrats is increased. This is exactly what is found to be happening with the data of Figures 1 and 2. The results are shown in **Figure 5**, where experimentally determined correlation coefficients are plotted as a function of the space between quadrat centers. One of the curves represents the correlation for the quadrats

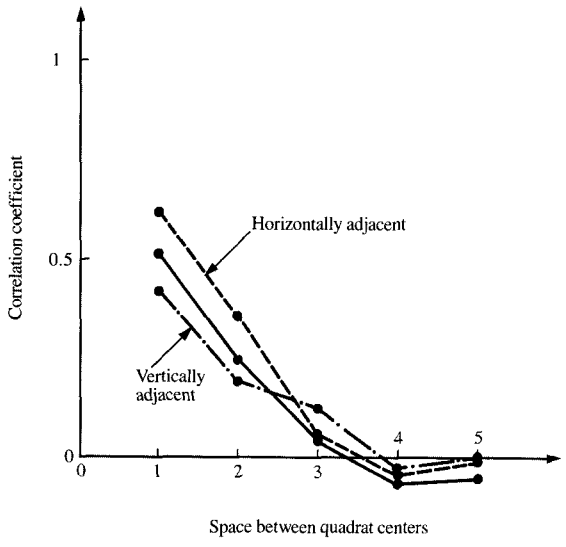

Figure 5

The effect of distance between quadrats. The correlation coefficent decreases as the distance between quadrats is increased. The difference in values for horizontally and vertically adjacent quadrat pairs could be an indication of anisotropic clustering. The correlation coefficients for the combination of horizontally and vertically aligned quadrat pairs are connected with a solid line.

that are aligned horizontally. A second curve results from vertically aligned quadrats. The third curve, plotted with a solid line, is for the combined horizontal and vertical data. These results indicate that positive correlation coefficients are observed to a distance of three quadrats. This implies that quadrat pairs separated by this distance, or less, have a chance of being located within the same region or cluster.

Correlation coefficients obtained in this fashion can also have negative values. As can be seen from the example with the dots, this happens when regions with high numbers of dots occur next to regions without dots. Such regions must be similar in shape and area. This did happen on several of the particle maps of Figures 1 and 2. Some of the negative correlation coefficients in Table 2 are a result of this. Quadrat pairs separated by a distance of four quadrat spaces or more are therefore more likely to have one quadrat inside a particle cluster while the other quadrat is located outside that cluster. These results suggest that the range of positive correlation coefficients is a measure of the average cluster dimension.

It is noteworthy that correlations as a function of distance also occur in models for distributions of galaxies in the universe. In the work of Peebles [20], such correlations are forced, with so-called "multi-point correlation functions." This differs from the approach followed in this paper, where nothing was forced. As mentioned before, the correlations here simply result from the use of compound Poisson statistics.

9. A simulation model

Correlations between the numbers of defects in adjacent chips are also taken into account in a simulation model described by V. Foard-Flack in Reference [16]. In that theory, defect distributions are generated with a background grid containing pseudorandom numbers. Although the numbers in the grid have a negative binomial distribution, they are completely independent of one another. At each gridpoint the random numbers of the neighboring gridpoints are added to the initial number. The rules used for picking the neighboring numbers can be changed. Each set of rules defines a different model. The resulting numbers at each grid point are assumed to represent fail counts for chips positioned at those points. It is shown in the paper that regardless of the model, the summed numbers also have a negative binomial distribution.

In the example given in [16], the random numbers of the four nearest adjacent grid points were added to the initial number. The number of defects at the location of each chip is therefore the sum of five random numbers. Two of these numbers are shared with a nearest neighbor, resulting in a dependence between the fail counts of adjacent chips. From the formulas in the paper, it can be determined that the theoretical correlation coefficient for this effect is equal to 0.4. This coefficient is constant. It does not change when the parameters defining the distribution of numbers in the background grid are changed.

The Foard-Flack model in essence smooths out the random numbers of the background grid. The resulting fail-count numbers therefore tend to undulate gradually. This is not necessarily what happens in actual integrated circuit manufacturing. Highly defective regions often butt up against defect-free regions. This phenomenon is evident in the maps of Figures 1 and 2, where quadrats with 18 and 19 particles are adjacent to particle-free quadrats. Another example of this can be found in the data published by Paz and Lawson [8].

The smallest defect pattern that can result from the example in Reference [16] is in the form of a cross. This occurs when the background grid has a single gridpoint with x defects in a field of zeros. From this, the model produces a cluster consisting of five chips arranged in a cross shaped like a + sign. Each chip within this cross contains exactly x defects. All other surrounding chips are defect-free.

It was mentioned in [16] that the rules used in the model could be modified to produce different patterns. This is done by changing the method for combining the random numbers of the initial reference grid. For example, let the faults at any gridpoint be the result of summing the initial random

number at that point only with the initial grid number to the left of it and the one below it. In that case the smallest cluster pattern consists of three chips grouped in the shape of a letter L. However, single isolated chip fails, or an isolated pair of chips adjacent to each other, cannot be generated with this model as it stands. Such patterns can be simulated only if more background grids, with pseudorandom numbers at each gridpoint, are added to the results of a Foard-Flack model. Such additions tend to lower the theoretical value of the correlation coefficients.

The example of [16] has a positive correlation coefficient only for adjacent chips. Chips that are further removed from each other have no correlations for the numbers of defects per chip. The data in Figure 3 suggest that these correlations should extend beyond the nearest neighbors. The rules for combining the random numbers of the reference grid can be modified to include points which are a given distance removed from each other. This does result in correlations between chips that are separated from each other by that distance. Until now such modifications have not resulted in simulated maps that resemble the data in Figure 1. This is unfortunate, because a good simulation model for clusters would be very useful in the field of yield modeling.

10. Conclusions

Compound or mixed Poisson statistics have been used for yield models since the early 1960s [21]. These models have been verified numerous times with data during the past twenty years. Yet, with all their successes, they remain poorly understood. It is hoped that the discourse in this paper is useful in clarifying some of the misunderstandings.

One such misunderstanding occurs when the distribution of the number of faults per chip is mistaken for the compounder used in the Poisson mixing process. As a result, the faults per chip are often confused with fault densities. To prevent such a mistake, the approach taken in this paper is somewhat tedious. However, only in this way could the fundamental assumptions be exposed in a clear manner. The model derived in this way predicts a nonvanishing covariance between the numbers of faults in adjacent chips. The value of this covariance is found to be equal to the variance of the compounder. This result leads to a formula for a correlation coefficient. Values calculated with that formula appear to be in reasonable agreement with correlation coefficients obtained from data of particle locations on semiconductor wafers. This was not anticipated in the literature, where it was assumed that correlations of this type could not be obtained with existing yield models [16].

The method of correlating the number of particles in adjacent quadrats provides a convenient test for clusters. This technique can be extended to quadrat pairs further removed from each other. In this way the correlation coefficients can be used as a statistical measurement of the size of clusters. Dimensions of clusters are critical in the design of very large chips that have redundant circuitry for yield enhancement. Such redundancy has little effect if the redundant circuits are located in the same fault cluster as the circuits that they have to replace.

References

1. C. H. Stapper, "Defect Density Distribution for LSI Yield Calculations," *IEEE Trans. Electron Devices* **ED-20,** 655–657 (July 1973).
2. T. Yanagawa, "Influence of Epitaxial Mounds on the Yield of Integrated Circuits," *Proc. IEEE* **57,** 1621–1628 (September 1969).
3. T. Yanagawa, "Yield Degradation of Integrated Circuits Due to Spot Defects," *IEEE Trans. Electron Devices* **ED-19,** 190–197 (February 1972).
4. A. Gupta and J. W. Lathrop, "Yield Analysis of Large Integrated-Circuit Chips," *IEEE J. Solid-State Circuits* **SC-7,** 389–395 (October 1972).
5. A. Gupta, W. A. Porter, and J. W. Lathrop, "Defect Analysis and Yield Degradation of Integrated Circuits," *IEEE J. Solid-State Circuits* **SC-9,** 96–103 (June 1974).
6. A. V. Ferris-Prabhu, L. D. Smith, H. A. Bonges, and J. K. Paulsen, "Radial Yield Variations in Semiconductor Wafers," *IEEE Circuits & Devices Mag.* **3,** 42–47 (March 1987).
7. C. H. Stapper, "LSI Yield Modeling and Process Monitoring," *IBM J. Res. Develop.* **20,** 228–234 (May 1976).
8. 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).
9. R. M. Warner, Jr., "Applying a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-9,** 86–95 (June 1974).
10. R. M. Warner, Jr., "A Note on IC Yield Statistics," *Solid-State Electron.* **24,** 1045–1047 (December 1981).
11. S. M. Hu, "Some Considerations in the Formulation of IC Yield Statistics," *Solid-State Electron.* **22,** 205–211 (February 1979).
12. W. Feller, *An Introduction to Probability Theory and Its Applications, II*, John Wiley & Sons, Inc., New York, 1971, p. 57.
13. C. H. Stapper, "Comments on 'Some Considerations in the Formulation of IC Yield Statistics,'" *Solid-State Electron.* **24,** 127–132 (February 1981).
14. C. H. Stapper, "The Effects of Wafer to Wafer Defect Density Variations on Integrated Circuit Defect and Fault Distributions," *IBM J. Res. Develop.* **29,** 87–97 (January 1985).
15. G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day Publishing Co., San Francisco, 1970, Ch. 2, pp. 23–45.
16. V. Foard-Flack "Introducing Dependency into IC Yield Models," *Solid-State Electron.* **28,** 555–559 (June 1985).
17. G. E. Moore, "What Level of LSI Is Best for You?" *Electronics* **43,** 126–130 (February 16, 1970).
18. C. H. Stapper, "On Yield, Fault Distributions, and Clustering of Particles," *IBM J. Res. Develop.* **30,** 326–338 (May 1986).
19. E. J. Nowak and W. M. Trickle, "Analysis and Solution of a Yield-Limiting Pattern-Fail Mechanism in a 1-Mbit DRAM," *1987 Symposium on VLSI Technology, Digest of Technical Papers*, May 1987, pp. 29–30.
20. P. J. E. Peebles, *The Large-Scale Structure of the Universe*, Princeton University Press, Princeton, NJ, 1980, pp. 141–148.
21. B. T. Murphy, "Cost-Size Optima of Monolithic Integrated Circuits," *Proc. IEEE* **52,** 1537–1545 (December 1964).

Received September 24, 1986; accepted for publication June 16, 1987

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 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, where he is now a senior engineer responsible for yield planning. During the early 1970s he developed a yield model for the analysis of defect monitor data which has been used ever since for line control and yield management. It has also been used extensively for productivity optimization of SRAMs and DRAMs with redundancy, as well as for planning the production of gate arrays, logic chips, and microprocessor chips. Dr. Stapper has published 14 papers on the subject of yield modeling and process monitoring. He is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.