How to Extract Defect Densities from Distributions

Alain Comeau

Reprinted from
IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING
Vol. 3, No. 2, May 1990

How to Extract Defect Densities from Distributions

ALAIN COMEAU

Abstract—**A novel technique is presented for representing a defect density, such as might be found on an integrated circuit. Distributions are a powerful tool for presenting defect distributions for test structures, and are convenient for yield analysis and modeling. How the concept of distribution can be extended to give more information on the defects themselves, is shown. The power of this approach derives from its invariability to test structure geometry, this enables direct comparison of data from different structures. An important difficulty in work on yield is relating defect distributions measured on test structures to those of other structures, i.e., integrated circuits. The work presented here helps understanding why and how it can be done. The mathematics are simple and lead to a defect density which can be expressed in units of defects per geometric factor per independent variable.**

Introduction

Scientists commonly use graphical statistical tools to present their work [1]–[6]. These tools are often powerful for synthesizing results. It is of prime importance that the tools help clarify the physical mechanisms involved. Sometimes authors make use of statistical tools in a manner that causes ambiguity [2], while others try to simplify the results of their work to the point where physical interpretation of the mechanisms is questionable [1].

Scientists have to learn to use tools that enable first and foremost an easy understanding of their work. Distributions have been such a tool for many of us. Distributions, however, sometimes do not go deeply enough to clearly show the physics involved. Using the concept of distributions a defect density closer to the physical mechanisms will be generated.

Theory

To begin, a distribution is defined as a discrete function $N(\varepsilon_i, \Delta\varepsilon)$ where ε_i is the independent variable and $\Delta\varepsilon$ is the interval between each value in the distribution. $N(\varepsilon_i, \Delta\varepsilon)$ is the number of discrete elements which passed a given test under the condition ε_i. For the sake of simplicity a continuous function $P(\varepsilon)$ will be used, being defined as

$$P(\varepsilon) = \lim_{\Delta\varepsilon \to 0} \frac{N(\varepsilon, \Delta\varepsilon)}{N_t \cdot \Delta\varepsilon} \tag{1}$$

where

$$\Delta\varepsilon = \varepsilon_{i+1} - \varepsilon_i \tag{2}$$

and N_t is the total number of elements tested.

From a formal point of view the use of $P(\varepsilon)$ to work out the following development may seem unjustified, however, its ease of use opens the way to a easier understanding. The conclusions remain the same in the discrete case.

This definition of $P(\varepsilon)$ is valid in the range of values of ε_i for which $N(\varepsilon_i, \Delta\varepsilon)$ is itself known, let this interval be $[\varepsilon_0, \varepsilon_f]$. The normalization condition in that interval for $P(\varepsilon)$ is

$$\int_{\varepsilon_0}^{\varepsilon_f} P(\varepsilon)\, d\varepsilon = 1. \tag{3}$$

Manuscript received September 9, 1989; revised November 13, 1989.

The author is with Mitel, S.C.C., 18 Boulevard de l'Aéroport, Bromont, PQ, JOE 110, Canada.

IEEE Log Number 8934137.

From this, one can extract the fraction of the sample for which $\varepsilon > \varepsilon_i$. This is the yield at ε_i ($Y(\varepsilon_i)$). It can be shown that

$$Y(\varepsilon_i) = \int_{\varepsilon_i}^{\varepsilon_f} P(\varepsilon)\, d\varepsilon = 1 - \int_{\varepsilon_0}^{\varepsilon_i} P(\varepsilon)\, d\varepsilon. \tag{4}$$

The last integral is obviously the fraction of the distribution for which $\varepsilon \leq \varepsilon_i$. This is commonly known as the percentile and is a concept also widely used in applied statistics for presenting distributions [2], [3]. We will note this as $\%(\varepsilon_i)$, a very explicit notation:

$$\%(\varepsilon_i) = \int_{\varepsilon_0}^{\varepsilon_i} P(\varepsilon)\, d\varepsilon. \tag{5}$$

One can note that $Y(\varepsilon_i)$, as previously defined, corresponds to the effective yield of the structure measured at ε_i in the case where the distribution is that of the number of test structures that failed at ε_i. This is a situation very often seen, for example, in the case of oxide integrity with breakdown distributions [4]–[6].

In this optic, one can define the average yield of the structure in the interval $[\varepsilon, \varepsilon + \Delta\varepsilon]$ as

$$Y(\varepsilon, \Delta\varepsilon) = e^{-AD(\varepsilon)\Delta\varepsilon} \tag{6}$$

where A is a constant geometric factor. $D(\varepsilon)$ is the defect density assuming a Poisson distribution. However, in most of today's yield analysis, clustering of the defects is considered. Clustering was neglected because in this case the distribution itself doesn't mean much. In the case where only one type of defect affects the distribution, clustering can be accounted for by using a clustering distribution to weight $D(\varepsilon)$. This case will not be discussed here.

Equation (6) means that a defect has a constant probability everywhere. Used as this, $Y(\varepsilon, \Delta\varepsilon)$ defines as being good whatever does not fall between ε and $\varepsilon + \Delta\varepsilon$. This includes values that are higher than ε.

Using this, one can extract the total yield of the structure $Y'(\varepsilon_i, \Delta\varepsilon)$ as being the product of all $Y(\varepsilon_j, \Delta\varepsilon)$ for which $\varepsilon_j \leq \varepsilon_i$. This makes a lot of sense; only the fraction that did not fail a previous test can fail the next one.

$$Y'(\varepsilon_i, \Delta\varepsilon) = Y(\varepsilon_0, \Delta\varepsilon) \cdot Y(\varepsilon_1, \Delta\varepsilon) \cdot Y(\varepsilon_2, \Delta\varepsilon) \cdots Y(\varepsilon_i, \Delta\varepsilon). \tag{7}$$

This product is more conveniently replaced by a summation through the use of the logarithm

$$\log\left(Y'(\varepsilon_i, \Delta\varepsilon)\right) = \sum_{j=0}^{i} \log\left(Y(\varepsilon_j, \Delta\varepsilon)\right) \tag{8}$$

and by using our previous definition of Y, see (6):

$$\log\left(Y'(\varepsilon_i, \Delta\varepsilon)\right) = -A \sum_{j=0}^{i} D\left(Y(\varepsilon_j, \Delta\varepsilon)\right). \tag{9}$$

This form is easily transformed into an integral by taking the limit when $\Delta\varepsilon \to 0$. One will note in this case that the total yield at ε_i becomes

$$\lim_{\Delta\varepsilon \to 0} Y'\left(\varepsilon_i, \Delta\varepsilon\right)\right) = Y(\varepsilon) \tag{10}$$

and therefore (9) becomes

$$\log\left(Y(\varepsilon)\right) = -A \int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\, d\varepsilon \tag{11}$$

or

$$Y(\varepsilon) = \exp\left(-A \int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\, d\varepsilon\right). \tag{12}$$

This is a from that closely resembles Poisson's distribution. Actually, it is the same, if one considers ε and ε_0 as being constant. Clearly this is not the case here. This form shows that $D(\varepsilon)$, as it is defined, represents an incremental defect density. In that sense, when one plots $D(\varepsilon)$ from an experimental data set, it gives the increase in defect density between ε and $\varepsilon + \Delta\varepsilon$.

Unfortunately (12) does not permit easy calculation of $D(\varepsilon)$. But this can be aleviated by some simple mathematics. As previously shown ((4) and (12)):

$$\log\left(Y(\varepsilon)\right) = \log\left(1 - \%(\varepsilon)\right) = -A\int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\,d\varepsilon. \quad (13)$$

Using the last two forms and differentiating with respect to ε one finds

$$\frac{d}{d\varepsilon}\left[\log\left(1 - \%(\varepsilon)\right)\right] = -AD(\varepsilon). \quad (14)$$

The former can be simplified to

$$\frac{1}{1 - \%(\varepsilon)}\frac{d\%(\varepsilon)}{d\varepsilon} = AD(\varepsilon). \quad (15)$$

By looking at the definition of $\%(\varepsilon)$ (5) one finds that

$$D(\varepsilon) = \frac{1}{A}\frac{P(\varepsilon)}{1 - \%(\varepsilon)}. \quad (16)$$

The final formula shown above states that for a given number of failures, the defect density is inversely proportional to the geometric factor A and to the fraction of the sample that had not previously failed $(1 - \%(\varepsilon))$. $D(\varepsilon)$ can be easily computed from any distribution $P(\varepsilon)$ using (16).

One can also do the opposite, that is compute the expected distribution for a given geometric factor. But to perform this, the latter form is not appropriate, one has to use a form that isolates $P(\varepsilon)$ and where $\%(\varepsilon)$ is absent. This can be done in the following manner: knowing that

$$1 - \%(\varepsilon) = \exp\left(-A\int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\,d\varepsilon\right). \quad (17)$$

Differentiating with respect to ε enables to isolate $P(\varepsilon)$:

$$P(\varepsilon) = \exp\left(-A\int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\,d\varepsilon\right)\frac{d}{d\varepsilon}\left(A\int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\,d\varepsilon\right) \quad (18)$$

which yields

$$P(\varepsilon) = AD(\varepsilon)\exp\left(-A\int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\,d\varepsilon\right). \quad (19)$$

In this way one can generate a distribution or evaluate distributions for any given geometric factor knowing $D(\varepsilon)$.

Of course this distribution, to be valid, must satisfy the condition of normalization over all values measured. $P(\varepsilon)$ calculated this way satisfies this condition as can be proven by evaluating $\%(\varepsilon_f)$ which must be 1 for this condition to be true. From (17):

$$\%(\varepsilon_f) = 1 - \exp\left(-A\int_{\varepsilon_0}^{\varepsilon_f} D(\varepsilon)\,d\varepsilon\right) \quad (20)$$

for $\%(\varepsilon_f)$ to equal 1 one must have

$$A\int_{\varepsilon_0}^{\varepsilon_f} D(\varepsilon)\,d\varepsilon \gg 1 \quad (21)$$

which means that the defect density summed over all values of ε must be much greater than $1/A$. In other words, all sites must have failed within the bounds $[\varepsilon, \varepsilon_0]$ when the distribution was constructed.

This actually implies that the distribution must contain all pertinent information, including the total number of samples. For this to happen all samples must be considered to have failed one way or another.

The size of the sample also comes into play, it must be large enough, so that the summed geometric factor of the test structures is at least as large as the geometric factor of the I.C. summed over one wafer. This condition insures a realistic evaluation of the yield.

When $D(\varepsilon)$ is computed with (16) one finds that fluctuations (or errors) in the evaluation of $D(\varepsilon)$ are greater near the end of the distribution (ε_f). The increased sensitivity is induced by the factor $(1 - \%(\varepsilon))^{-1}$. This factor implies that all sites must have failed at ε_f, thus a certain probability of failure equates an infinite defect density. This explains why $D(\varepsilon_f)$ diverges and $\%(\varepsilon_f)$ converges to 1.

Example

To show the use of this technique, the breakdown of junction diodes will be used as an example. Fig. 1 shows a distribution of breakdown (V_b) for $n+/p-$ diodes. The diodes each have an area of 186000 μm^2 (geometric factor) and in this special case crystalline defects are responsible for low breakdowns of the junctions. The breakdown was measured by applying a current of 1 μA and measuring the voltage. The peak which can be seen at about 23 V is the intrinsic breakdown of the junction. Whatever is seen beyond that peak, is bad data. It is mainly caused by bad contacts and missing implants near the edges of the wafer. Otherwise, the wafers showed no systematic spatial dependence.

Fig. 2 shows the defect density $D(V_b)$ computed with (16). One can see that the defect density of the low breakdowns rises sharply at about 10 V and then saturates to a nearly constant value between 12 and 22 V, this is a very wide range for a defect distribution to be constant and is a very valuable piece of information on the nature of the defects. This behavior is the signature of the defects, they begin turning on at about 10 V, and then they have a constant rate of appearance as a function of voltage. Next, at about 22 V the peak of the intrinsic breakdown of the junction (V_B) can be seen. The last data point in the defect density rises sharply to a very high value (actually limited to 10^{10} by calculation). This is a direct effect of the factor $(1 - \%(V_b))^{-1}$, which goes to infinity as the computations reach the last data point.

Now, if one was to use this defect density to compute the distribution of a structure with a smaller geometric factor, the results would be correct up to V_B, that is, for the defective sites. Because, with the smaller area, the probability of reaching a higher breakdown value is greater. However, the peak in $D(V_B)$, as shown in Fig. 2, does not represent the reality. This peak should continuously rise to infinity or at least to a value such that the probability that a site ends up with a value of $V > V_B$ is essentially zero.

In order to represent the reality in this case one must amend $D(V_b)$ with a model of the intrinsic breakdown. In Fig. 3, the use of a exponential defect density to model the actual distribution of intrinsic breakdown is shown. The rest of the distribution greater than V_B was deleted, as it became irrelevant. In the case where bad data is present, its importance is weighted as $(1 - \%(V_b))^{-1}$, therefore much more important near the end of the measurement range. One must keep in mind that by doing so, only the intrinsic breakdown is subject to alteration, not the low voltage region where the defects are.

Using this modified $D(V_b)$ one can now simulate the distribution for any practical size of diode. One can try, first, to regenerate the initial distribution, this can be seen in Fig. 4, the similarity is obvious, even with the alteration near the intrinsic breakdown.

From this point, one can evaluate distributions for different geometric factors. This is shown in Fig. 5 as a three dimensional graph. As the diode size moves from smaller areas to larger areas, the weight of the distribution gets shifted towards smaller voltages. This is a result of the increase in the probability of having a defect at any voltage. The same reasonning applies to an I.C., one merely

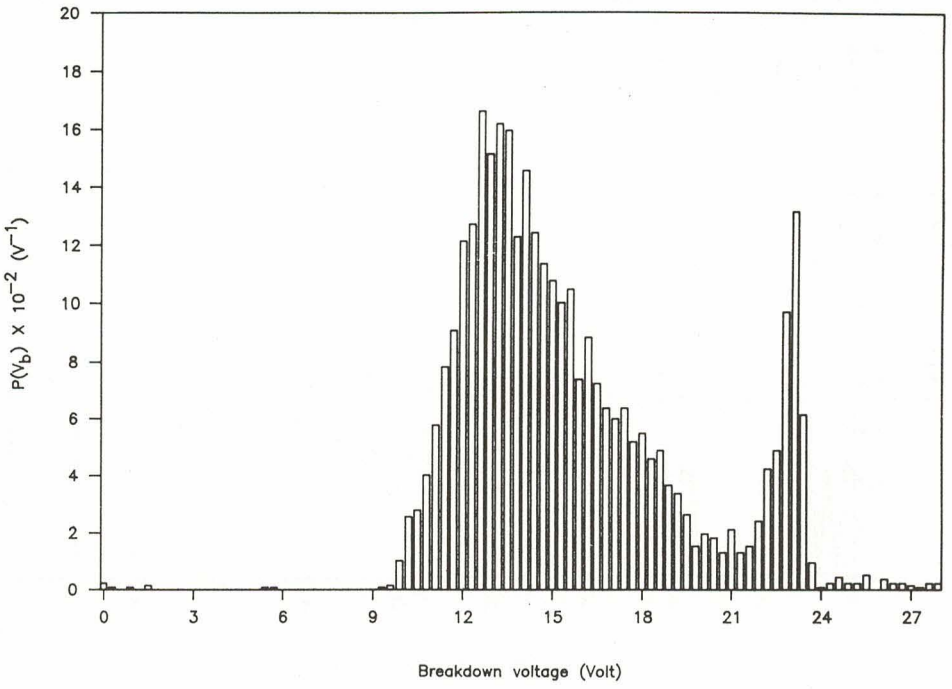

Fig. 1. Breakdown voltage distribution ($P(V_b)$) for $n+/p-$ diodes measured at 1 μA. The distribution is truncated at 28 V for the ease of comparison with Fig. 4. Raw data actually extended up to 40 V; only a few points are missing. Intrinsic breakdown is seen at about 23 V.

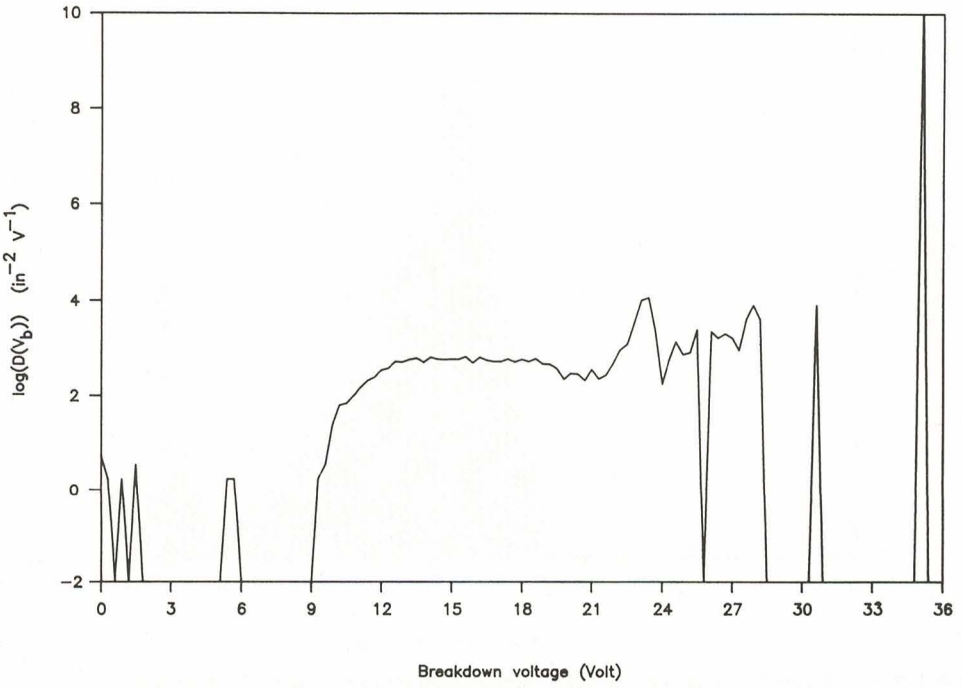

Fig. 2. Defect density $D(V_b)$ as calculated from (16). Large peak at last data point (35 V) is caused by $(1 - \%(\varepsilon))^{1}$ term in (16). Essentially all values above 24 V are caused by resistive effects.

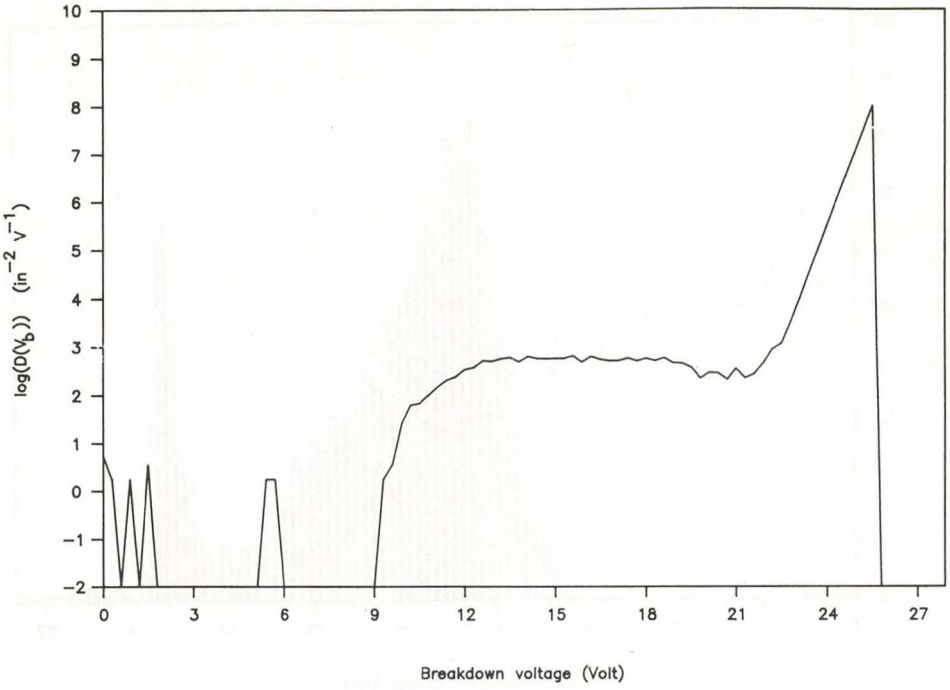

Fig. 3. Defect density of Fig. 2 corrected with exponential model of intrinsic breakdown. Defect density is limited to value such that probability of having sample above corresponding voltage is essentially zero. That value depends on geometric factor.

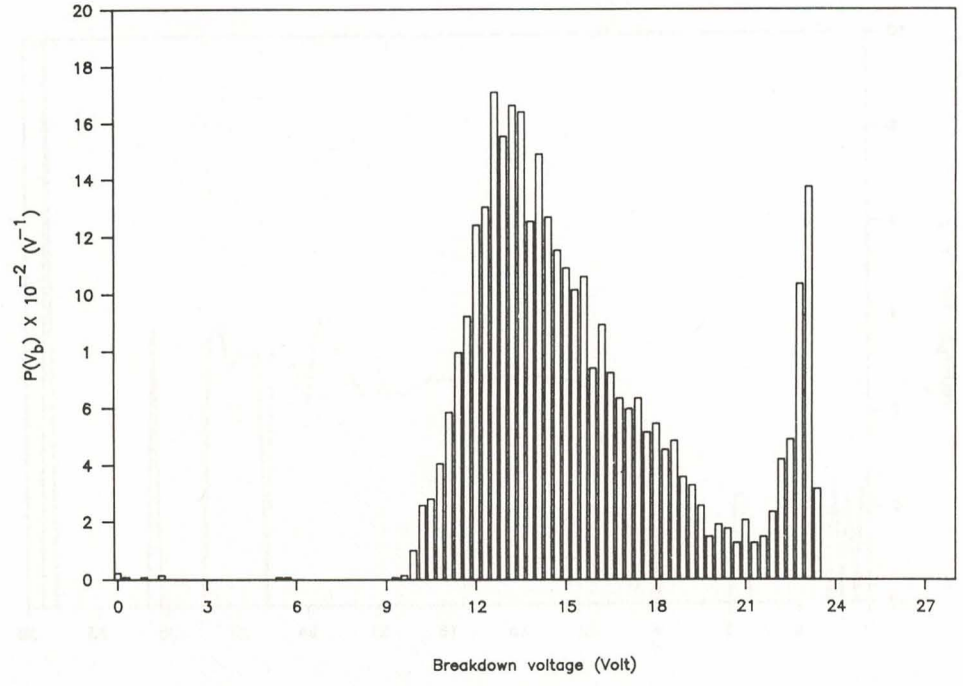

Fig. 4. Breakdown voltage distribution generated using defect distribution of Fig. 3 and (19). Area (geometric factor) was same than for raw data of Fig. 1. Note that in this case no data point exists above 24 V. Distribution is extremely similar to Fig. 1.

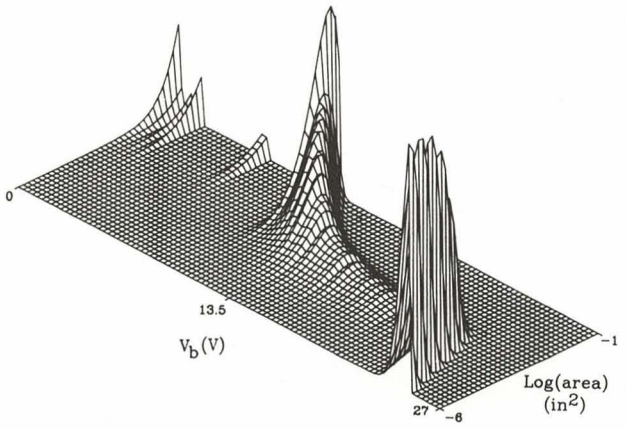

Fig. 5. Three dimensional graph of distributions expected from (19) using defect density of Fig. 3 and area varying from 10^{-6} to 10^{-1} in^2. Weight of distribution is shifted toward smaller voltage as area is increased.

extract the geometric factor for a given I.C. from the layout, and computes the yield (12) at any given voltage.

Multiple Geometrical Factors

In the case of a dependency on more than one geometrical factor, e.g., area(A) and perimeter(L), one needs to start with two distributions (P_1, P_2) having different geometrical factors (A_1, A_2, L_1, L_2) and consider two defect densities (D and σ). The formulation remains essentially the same, in this case (16) reduces to the equation set:

$$\frac{P_1(\varepsilon)}{1 - \%_1(\varepsilon)} = A_1 D(\varepsilon) + L_1 \sigma(\varepsilon) \tag{22}$$

$$\frac{P_2(\varepsilon)}{1 - \%_2(\varepsilon)} = A_2 D(\varepsilon) + L_2 \sigma(\varepsilon) \tag{23}$$

and one merely solve the set for all ε. From here, generalization to more than two independent defect densities should be obvious. It can also be shown that in this case

$$Y(\varepsilon) = \exp\left| -\left(A \int_{\varepsilon_0}^{\varepsilon} D(\varepsilon)\, d\varepsilon + L \int_{\varepsilon_0}^{\varepsilon} \sigma(\varepsilon)\, d\varepsilon \right) \right| \tag{24}$$

and

$$P(\varepsilon) = Y(\varepsilon)\left[AD(\varepsilon) + L\sigma(\varepsilon)\right]. \tag{25}$$

Conclusion

A method for calculating the defect density from a single set of data was presented. This defect density has the property of being invariant on the test structure, as expected, and also enables to compute the yield of any structure at any operating voltage (where voltage is used as the independent variable).

In fact, many process variables could act as independent variable in this case. Photoresist exposure time could be one variable for instance. This technique can be used to identify the physical mechanisms because it gives detailed information in a manner which is mostly independent of the structure used. The defect density, as defined in this work, gives a signature of the defect which is independent of the test structure.

In the case where the structure itself causes changes in the defect densities, these changes can be determined much more easily by comparing defect densities from structures with different geometric factors. This would have been nearly impossible with usual distributions.

References

[1] K. Yamabe and K. Taniguchi, "Time-dependent dielectric breakdown of thin thermally grown SiO_2 films," *IEEE Trans. Electron Devices*, ED-32, p. 423, Feb. 1985.

[2] F. Bryant, F.-T. Liou, Y. P. Han, and J. J. Barnes, "Thin dielectric quality/yield using a constant voltage ramp method," *J. Electrochem. Soc.*, vol. 136, p. 283, Jan. 1989.

[3] J. Lee, I.-C. Chen, and C. Hu., "Statistical modeling of silicon dioxide reliability," in *IEEE Int. Rel. Phys. Symp. Conf. Rec.*, 1988, p. 131.

[4] D. W. Ormond and J. R. Gardiner, "Reliability of SiO_2 gate dielectric with semi-recessed oxide isolation," *IBM J. Res. Develop.*, vol. 34, p. 353, May 1980.

[5] J. Laneuville, J. Marcoux, J. Orchard-Webb, and A. Comeau, "Defect characterization of a silicon gate CMOS process," *Semicond. Int.*, p. 250, May 1985.

[6] S. Gandemer, B. C. Tremintin, and J.-J. Charlot, "Critical area and critical levels calculation in I.C. yield modeling," *IEEE Trans. Electron Devices*, ED-35, p. 158, Feb. 1988.