From Tutorial Notes, 33rd Annual International Reliability Symposium, Las Vegas, Nevada, April 3, 1995, pp. 3.1 - 3.56.

A DEFECT MODEL OF RELIABILITY*

C. Glenn Shirley
Intel Corporation, MS: AL3-15
5200 N.E. Elam Young Parkway
Hillsboro, Oregon 97124 U.S.A
Phone: (503) 613-6050. FAX: (503) 613-5907. Email: Glenn_Shirley@ccm.ra.intel.com

ABSTRACT

The same defects that degrade device yield also affect device reliability. The complete theory is complicated and depends on factors such as die size, defect density, defect size distribution, circuit layout density, and environmental stress. We analyze the simplifying assumptions necessary for a practical model. Then we show how to use the practical model to extract process-specific reliability models, and thereby estimate failure rates of complex products, without reliance on full product-specific reliability data.

1. INTRODUCTION

Reliability characterization is an integral part of a wafer fabrication process development methodology. Specially designed test vehicles, such as static random-access memories (SRAMs) optimized for testability and failure analysis, are used to characterize the process. In the course of process development, a large amount of SRAM-specific reliability data and yield data is acquired. The reliability data include information about major failure modes, and their accelerations as a function of temperature and bias. The yield data include information about the major kinds of defects in the wafer fabrication process. The yield and reliability data acquired during the course of process development are called "baseline" reliability and yield data. Once the wafer fabrication process has reached certain yield and reliability goals (that is, is "certified"), the process is transferred to high-volume production.

Once in high-volume production, continuous improvement to improve yield is a ongoing activity. Also, at this stage of the life of a wafer fabrication process many new products which are to be produced using the process are introduced. Burn-in and life test is also performed on selected products, and this data is also added to the baseline data set.

The complexity of circuitry in typical microprocessor products make it prohibitively expensive and time-consuming to characterize directly the reliability of every new product introduced. This paper describes a method for making reliability estimates for new products produced on certified wafer fabrication processes without the necessity of gathering extensive reliability data for every new product introduced.

The model described in this paper is based on the premise that important failure mechanisms are *defect-related*, not intrinsic. The model exploits the relationship between yield and reliability, and uses the fact that the same defects which cause devices to fail at sort and raw class test, before any stress, are the same kind of defects which will eventually cause failure in subsequent stress. This relationship allows us to use *yield* data to make *reliability* predictions. There are a number of assumptions which are necessary in order to derive a practical model, and these are described in Section 3, after a brief review of basic reliability statistics in Section 2. In Section 4 we explain the formalism of acceleration models, and how to take into account burn-in.

By the end of Section 4, we have derived a complete defect reliability formalism. The model may be used in two ways: (1) To consolidate baseline data from disparate sources into a "reference" data set to which a single reference reliability model characteristic of the process may be fitted. This is described in Section 5. (2) To use reference reliability model to make predictions about "unknown" products. This is described in Section 6.

---

* Contains minor corrections from version published as IRPS tutorial.

2. RELIABILITY STATISTICS

This section introduces the terms and concepts needed to provide the mathematical framework for a statistical model for product reliability. The treatment follows standard reliability texts such as Tobias and Trindade[1].

Consider a population of units starting life at $t = 0$. As time progresses, members of the initial population will fail. There are several mathematical functions which are used to describe and analyze this process.

The cumulative distribution function, CDF, $F(t)$, is the probability that a unit drawn randomly from the initial population fails by time t. $F(t)$ has the properties: $F(t = 0) = 0$, $F(t = \infty) = 1$. $F(t)$ is undefined for $t < 0$, and increases monotonically as t increases. $F(t)$ is a probability, and can be manipulated using the mathematical rules of probability analysis. $F(t)$ is also called the "unreliability function" or "failure probability".

The survival function, , is defined by

$$S(t) = 1 - F(t) \quad \text{.....(1)}$$

$S(t)$ is also called the "reliability function" or "survival probability". $S(t)$ is the probability that a member of the initial population will survive to time t. $S(t)$ is a monotonically decreasing probability with properties: $S(t = 0) = 1$, $S(t = \infty) = 0$. The main objective of the reliability model described below is to relate $S(t)$ for a new "unknown" product to $S(t)$ for the "known" baseline data.

The probability density function, PDF, $f(t)$, is frequently used in theoretical discussions, but is rarely used directly in analysis of data or reliability predictions. $f(t)$ is defined as the rate of change of the fraction of the *initial population* which has failed:

$$f(t) = \frac{\text{Number of failures in } dt}{dt} \frac{1}{\text{Initial Population}}$$

or

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}, \quad \text{.....(2a)}$$

or

$$F(t) = \int_0^t f(t')dt'. \quad \text{.....(2b)}$$

A more fundamental function is the instantaneous failure rate, $h(t)$, defined as the rate of change of the fraction of the *population which has survived to time t*. $h(t)$ is defined as

$$h(t) = \frac{\text{Number of failures in } dt}{dt} \frac{1}{\text{Population at time } t}$$

or

$$h(t) = \frac{f(t)}{S(t)} = -\frac{1}{S(t)}\frac{dS(t)}{dt} = -\frac{d \ln S(t)}{dt}. \quad \text{.....(3)}$$

$h(t)$ is also known as the "instantaneous hazard". $h(t)$ can increase or decrease and can have any positive value, that is, $h(t)$ has the property $h(t) > 0$. Notice that, if the cumulative fraction failed is small ($F(t) << 1$), then $h(t) = f(t)/S(t) = f(t)/[1-F(t)] \approx f(t)$.

---

1 P. A. Tobias and D. C. Trindade, "Applied Reliability," Van Nostrand Reinhold (1986).

At early times, $h(t)$ is usually a decreasing function of time as weak parts are weeded out. At long times, $h(t)$ is an increasing function of time as the materials of the device degrade and the part wears out. This is the classical "bathtub" curve of reliability.

Reliability data for input to a reliability model is frequently not in a form which permits direct extraction of the survival function; hence the need for the cumulative hazard defined as

$$H(t) = \int_0^t h(t)dt \quad \text{.....(4)}$$

or, after substituting Eq. (3),

$$S(t) = \exp[-H(t)] \quad \text{.....(5)}$$

The cumulative hazard is useful because it is derivable directly from "censored" data (see Section 5 below), which is commonly encountered in practice. Equation (5) provides a relationship between the experimentally-determined cumulative hazard, and the survival probability which is of practical interest.

The functions $F(t)$, $S(t)$, $f(t)$, $h(t)$, and $H(t)$ are interrelated and given one, the others can be derived. The relationships in this section do not depend on the functional form of the probability distribution.

The average failure rate over various periods of product life is a commonly used indicator of reliability. The average failure rate between times t_1 and t_2 is defined as:

$$\text{AFR}(t_1, t_2) = \frac{\int_{t_1}^{t_2} h(t)dt}{t_2 - t_1}$$

or, from Eq. (5)

$$\text{AFR}(t_1, t_2) = \frac{\ln S(t_1) - \ln S(t_2)}{t_2 - t_1} \quad \text{.....(6)}$$

Thus, if $S(t)$ is known for a particular product, then the average failure rate between any two times can easily be calculated.

We extend the discussion by taking into account that integrated circuits will fail in service by any of several mechanisms. These mechanisms can be classified as either intrinsic or defect-related failure mechanisms. The intrinsic failure mechanisms are related to the materials and design rules of the manufacturing process. These mechanisms are due to wear-out, and cause rapidly increasing failure rates late in the lifetime of a device. Additionally, defects may be introduced during the manufacturing process. If these defects are not immediately fatal to the device, they may cause it to fail at early times. Defect-related failure mechanisms often cause rapidly decreasing failure rates early in the life of a device.

If a device can fail by any of several mechanisms, and if the earliest occurrence of any mechanism is fatal to the device, then the device is logically a chain.

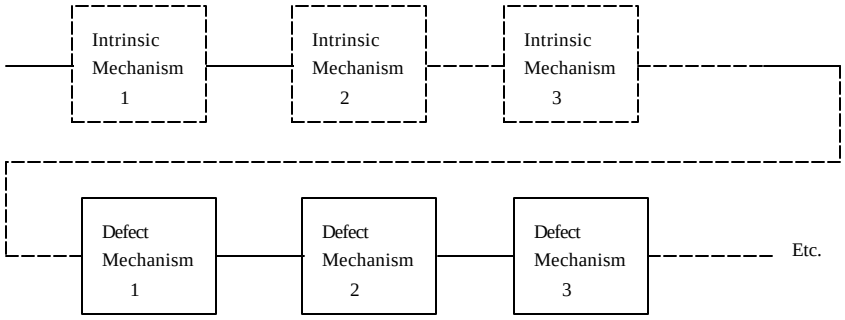

Fig. 1 Representation of reliability of device as reliability of a chain. Device will fail if any link fails. Some links represent intrinsic properties of materials, while others represent mechanisms caused by defects.

The survival probability of a chain is the product of survival probabilities of the links:

$$S(t) = \prod_j \hat{S}_j(t) \prod_i S_i(t). \quad \text{.....(7)}$$

where $\hat{S}_i(t)$ is the survival probability if intrinsic mechanism i were the only failure mechanism, and $S_i(t)$ is the survival probability if defect mechanism i were the only failure mechanism. From Eqs. (4) and (5), Eq. (7) is equivalent to:

$$h(t) = \sum_j \hat{h}_j(t) + \sum_i h_i(t). \quad \text{.....(8).}$$

In Eqs. (7) and (8) "hatted" functions refer to intrinsic mechanisms, whereas unmarked functions refer to defect-related mechanisms. Equation (8) shows that the failure rate of the component is the sum of the failure rates of the individual mechanisms. In the rest of this paper we shall ignore the intrinsic failure mechanisms since they are negligible over the service life of a well-designed process.

3. RELATIONSHIP BETWEEN YIELD AND RELIABILITY

Wafer fabrication is a batch-oriented manufacturing process. It is known that low-yielding lots usually have lower reliability, as evidenced by higher burn-in fallout. A source of "yield defects" fatal to the chip at sort (wafer-level) or raw class (packaged die before stress) is often also the source of "latent reliability defects" which are fatal early in the life of the chip. This section will quantify this relationship and show how to predict the reliability of an unknown product if the reliability of a "reference product" is known. In this section we shall confine our attention to prediction of reliability of an unknown product at the *same* environmental condition as the known "reference product". This restriction will be removed in Section 4.

A specific example of the relationship between yield and reliability will illustrate theoretical principles, and, more importantly, highlight the implicit assumptions made when invoking the relationship between yield and reliability. We use the specific example of line shorting by conductive particles. Consider an infinite pattern of metal lines of width w and spacing s, as shown in Fig. 2. Conductive defects, shown as circles, are distributed randomly across the pattern of metal lines. The centers of the circles will be distributed randomly with density D_{ave}. That is, on the average, there are D_{ave} circle-centers per unit area of the pattern. The circles have a distribution of sizes $D(x)$; that is, the number of circles (or circle-centers) per unit area with diameters between x and $x+dx$ is $D(x)dx$. Stapper[2] has suggested that a defect size distribution of the following form describes what is observed in the wafer fabrication plant:

$$\begin{aligned} D(x) &= D_{\text{ave}}(x / x_0^2) \quad \text{for} \quad x \le x_0 \\ &= D_{\text{ave}}(x_0^2 / x^3) \quad \text{for} \quad x > x_0. \end{aligned} \quad \text{.....(9)}$$

Stapper's distribution is shown schematically as the top graph in Fig. 3. x_0 is a characteristic length much less than the resolving power of the defect detection equipment. Since x_0 is much less than the characteristic dimensions of the pattern (w and s), we will see that the only part of the distribution of importance is the part for $x > x_0$. The exact shape of the distribution is not important to the discussion, and x_0 can be thought of as merely a normalizing constant. If a defect is sufficiently small it can *never* short metal lines. Also, if a defect is sufficiently large it will *always* short metal lines. Intermediate-sized defects can *sometimes* come within a very short distance δ of a metal line while simultaneously being in contact with an adjacent line - such defects are termed "latent reliability defects". Not all defects that are of a size to be possibly a latent reliability defect actually are reliability defects. δ is defined so that a defect contacting one line and lying within δ of an adjacent line will *eventually* fail within the service life of the component. This is illustrated in Fig. 2.

---

[2] C. H. Stapper, "Modelling of Integrated Circuit Sensitivities," IBM J. Res. Develop. Vol. 27, No. 6, pp 549-557 (November, 1983). The defect size distributions in Intel's factories do not necessarily have the form described by Stapper. We use Stapper's form for illustrative purposes only.

Assuming random distribution of defects, the proportion of defects of diameter x which short metal lines can easily be shown to be

$$P_{\text{yield}}(x) = 0, \qquad \text{for } x < s$$
$$= \frac{x - s}{s + w}, \quad \text{for } s \le x < 2s + w$$
$$= 1, \qquad \text{for } x \ge 2s + w \quad . \quad \text{.....(10)}$$

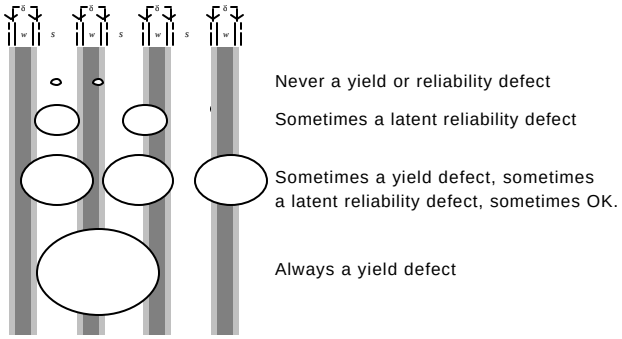

Fig. 2 Four lines of an infinite array of metal lines of width w and spacing s. Conductive defects of various sizes are superimposed on the array. If a defect shorts one or more lines it affects yield. If a defect lies on one line and comes within a small distance δ of an adjacent line (light shading), it is a "latent reliability defect". The value of δ is such that a latent defect will short lines within the service life of the unit.

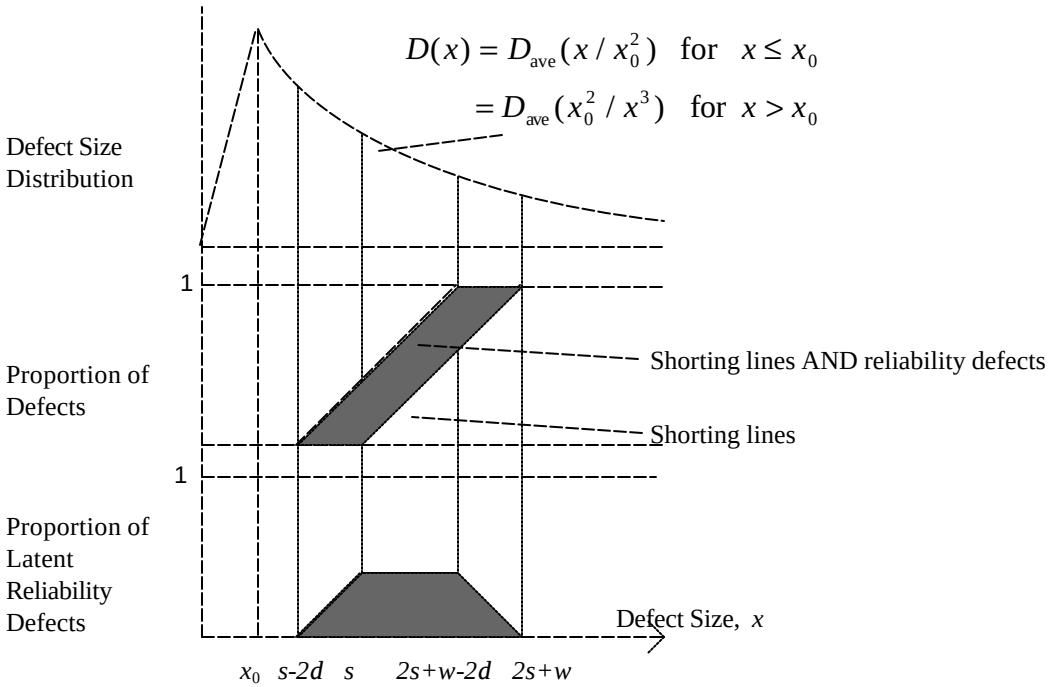

Fig. 3 Top: Defect size distribution. Middle: (1) Proportion of yield defects (immediately fatal), and (2) Proportion of yield and latent reliability defects combined. Bottom: Proportion of only latent reliability defects. All as functions of defect size.

By making the replacements, $s \to s - 2d$ and $w \to w + 2d$ in Eq. (10), we find the proportion of defects which are either yield or latent reliability defects:

$$\begin{aligned} P_{\text{yield \& latent rel.}}(x) &= 0, && \text{for } x < s - 2d \\ &= \frac{x - s + 2d}{s + w}, && \text{for } s - 2d \le x < 2s + w - 2d \\ &= 1, && \text{for } x \ge 2s + w - 2d \end{aligned} \qquad \text{.....(11)}$$

The proportions given by Eq. (10) and (11) are plotted in the middle graph in Fig. 3. The *difference* between these proportions, shown shaded and plotted at the bottom of Fig. 3, is the proportion of defects which are latent reliability defects, $P_{\text{rel}}(x)$.

The total density of "yield" defects (defects per unit area of the pattern) is

$$D_{\text{yield}} = \int_0^\infty D(x) P_{\text{yield}}(x) dx \,. \qquad \text{.....(12)}$$

After some algebra, we find for the specific example:

$$D_{\text{yield}} = \frac{D_{\text{ave}} x_0^2}{2s(w + 2s)}; \quad D_{\text{yield}} + D_{\text{rel}} = \frac{D_{\text{ave}} x_0^2}{2(s - 2d)(w + 2s - 2d)}$$

where the defect density of yield and reliability defects combined is obtained by substituting $s \to s - 2d$ and $w \to w + 2d$ in the expression for D_{yield}. So, taking the difference between the combined density and D_{yield}, we find

$$D_{\text{rel}} = \frac{D_{\text{ave}} x_0^2}{2} \times \left\{ \frac{1}{(s - 2d)(w + 2s - 2d)} - \frac{1}{s(w + 2s)} \right\}$$

so that

$$k \equiv \frac{D_{\text{rel}}}{D_{\text{yield}}} = \frac{s(w + 2s)}{(s - 2d)(w + 2s - 2d)} - 1 \approx d \times \frac{2(w + 3s)}{s(w + 2s)} + \text{ higher order terms in } d \,.$$

Therefore, for this example, the density of latent reliability defects is proportional to the density of yield defects (that is, k is constant) for

1. A fixed *shape* of the defect size distribution. (And fixed shape of the defects themselves.)

2. A fixed *pattern* on which the defects fall (w and s).

3. A fixed definition of "latent reliability defect" (value of δ)

4. Randomly distributed, non-interacting defects.

In principle, the value of k may be calculated theoretically for any defect (shorting or open-type defects) and pattern. For complex defect shapes and size distributions the calculation may be complicated; it is usually intractable in closed mathematical form. Monte-Carlo methods may be necessary to evaluate the constant. However, we shall show that for the reliability theory in this paper, an exact value of k is not needed. We shall only invoke the fact that k is a constant, if the above four conditions are satisfied. Constancy of k is not a law of nature: If for example a process excursion occurs which changes the shape of the defect distribution, k may vary. Finally, there is a value of k for each failure mechanism, so in general we have a value k_i for each mechanism i.[3] So, in general, if the above four conditions are satisfied, we may write for each mechanism i

[3] Formally, the specification of a mechanism depends not only on the type of mechanism (eg. metal slivers shorting metal 1), but also on the pattern on which it falls. Thus, metal slivers shorting a coarse pattern, and the

$$D_{\mathrm{rel}(i)} = k_i \times D_{\mathrm{yield}(i)} \quad \text{.....(13)}$$

To demonstrate that k exists for each mechanism of importance we compare the yield and reliability defect Paretos observed for SRAMs. It is apparent that the defects which are discovered most frequently at sort (low-yield-analysis, LYA, defects), are also seen most frequently after life test (RE defects), suggesting that RE defects are "escapee" LYA defects as modeled above.

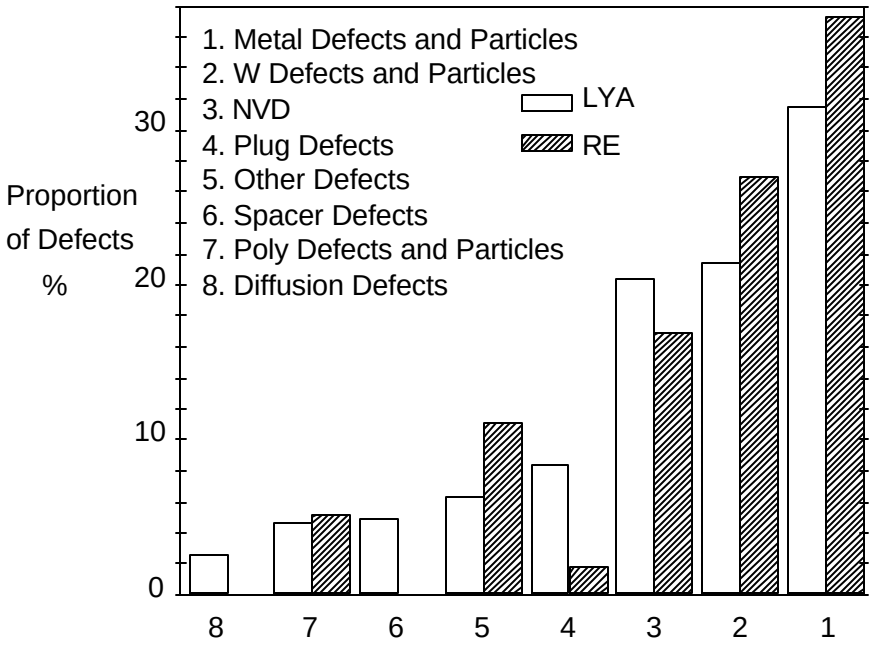

Fig. 4 Comparison of incidence of yield (low-yield analysis, LYA) and reliability (RE) defects observed for SRAMs. (NVD = no visible defect).

Strictly, k is not constant from product to product, because the pattern layout varies from product to product, violating assumption number 2 above. However, actual yield and reliability fallout is concentrated in a few mechanisms (see Fig. 4) and pattern variation does not appear to be an overriding factor for the patterns and defects that actually occur in products. This can be seen in Fig. 5 which shows that the total reliability defect density is proportional to the total yield defect density, even when different products are used to make the plot.

Each individual latent reliability defect has a survival probability, which is a function of time. We will denote this single-defect reliability probability by $s_i(t)$, for mechanism i. The survival probability of an "unknown" product p failing by mechanism i with n_i^p latent reliability defects of type i is

$$S_i^p(t) = [s_i(t)]^{n_i^p} = [s_i(t)]^{D^p_{\mathrm{rel}(i)} \times A_i^p} \quad \text{.....(14)}$$

where we have assumed that the latent reliability defects are independent, and so have invoked the "chain" rule illustrated in Fig. 1 and Eq. (7). We have used the fact that the number of reliability defects of type i on the product p is given by the product of the latent reliability defect density for defect i, and the sub-area of the die which has the appropriate pattern for mechanism i:

$$n_i^p = D^p_{\mathrm{rel}(i)} \times A_i^p \quad \text{.....(15)}$$

---

same metal slivers shorting a fine pattern would be taken account of in the model formalism by two distinct indexes, i.

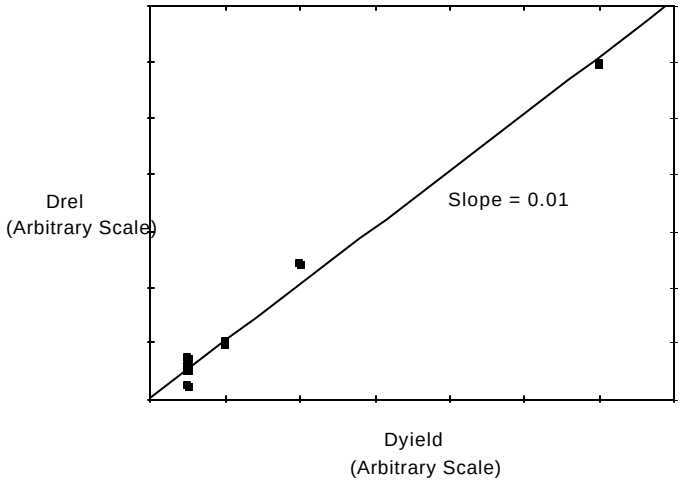

Fig. 5 D_{rel} vs D_{yield} for various designs produced on the same process as data in Fig. 4. D_{rel} is proportional to D_{yield} for all mechanisms combined. Defect densities are in arbitrary units. The reliability defect density is calculated from D_{rel} = -ln{cum. fail at 6 hours of burn-in}÷Area. Notice that reliability defect densities are 1/100 of the yield defect densities.

Let's assume that we know the survival probability $S_i^r(t)$ for a "reference" product r at the same environmental conditions as the unknown product p. $S_i^r(t)$ may be expressed in terms of $s_i(t)$ in exactly the same way as in Eqs. (14) and (15), except that p will be replaced by r. Elimination of the single defect survival function, $s_i(t)$ between the expressions for $S_i^r(t)$ and $S_i^p(t)$ gives the following:

$$S_i^p(t) = [S_i^r(t)]^{R_i(p|r)} \quad \text{.....(16)}$$

where, the "scaling ratio" is defined as

$$R_i(p|r) = \frac{D^p_{\text{rel}(i)} \times A_i^p}{D^r_{\text{rel}(i)} \times A_i^r}. \quad \text{.....(17)}$$

Using the chain rule (Fig. 1, Eq. (7)) for survival functions from multiple mechanisms, we can write the total survivor function for all defect mechanisms for an unknown product, p, as

$$S^p(t) = \prod_i S_i^p(t) = \prod_i \left[S_i^r(t)\right]^{R_i(p|r)}. \quad \text{.....(18)}$$

If the four conditions discussed above hold, then we may invoke Eq. (13) above, and obtain

$$R_i(p|r) \cong \frac{D^p_{\text{yield}(i)} \times A_i^p}{D^r_{\text{yield}(i)} \times A_i^r}. \quad \text{.....(19)}$$

The relationship Eq. (19) is a central concept because yield defect densities are measured and tracked in the wafer fabrication plants, whereas densities of latent reliability defects are much smaller and not as easily available.

Next we show how to express Eq. (19) in terms of commonly acquired yield indicators. One of the indicators is the mechanism Pareto ("LYA" in Fig. 4 is an example), that is, the proportion of all failures failing by each mechanism. To a good approximation (for small % yield fallout) the mechanism Pareto is defined as:

$$P_i^p = \frac{D^p_{\text{yield}(i)} \times A_i^p}{\sum_j D^p_{\text{yield}(j)} \times A_j^p}, \quad \text{.....(20)}$$

which is the proportion of sort and raw class units (of product p) failing due to mechanism i. Using Eq. (20), we may write

$$D^p_{\text{yield}\,(i)} \times A_i^p = P_i^p \times \sum_j (D^p_{\text{yield}(\,j)} \times A_j^p) = P_i^p \times D^p_{\text{yield}} \times A^p \quad \text{.....(21)}$$

where A^p and D^p_{yield} are given by

$$A^p \equiv \sum_j A_j^p \quad \text{(total die area)} \quad \text{.....(22)}$$

and

$$D^p_{\text{yield}} \equiv \frac{\sum_j D^p_{\text{yield}\,(j)} \times A_j^p}{A^p}. \quad \text{.....(23)}$$

D^p_{yield} is an average (weighted by sub area) defect density for all yield mechanisms. From Eqs. (19) and (21) we may write

$$R_i(p|r) = \frac{P_i^p \times D^p_{\text{yield}} \times A^p}{P_i^r \times D^r_{\text{yield}} \times A^r}. \quad \text{.....(24)}$$

Assuming Poisson statistics (random, non-interacting defects), the defect-related yield for a die with various identifiable sub-areas each with "stand-alone" defect-related yield Y^p_j may be written

$$Y^p = \prod_j Y_j^p = \prod_j \exp(-D^p_{\text{yield}(\,j)} \times A_j^p) = \exp\left(-\sum_j D^p_{\text{yield}\,(j)} \times A_j^p\right) \quad \text{.....(25)}$$
$$= \exp(-D^p_{\text{yield}} \times A^p)$$

where D^p_{yield} and A^p are given by Eqs. (23) and (22). From Eqs. (24) and (25) we may write

$$R_i(p|r) = \frac{P_i^p \times \ln(Y^p)}{P_i^r \times \ln(Y^r)} \quad \text{.....(26)}$$

So we may write Eq. (18) as

$$S^p(t) = \left[\prod_j S_j^r(t)^{\frac{P_j^p}{P_j^r}}\right]^{\frac{\ln(Y^p)}{\ln(Y^r)}}. \quad \text{.....(27)}$$

If we add the following condition to the above 4 conditions,

> **5.** The yield Pareto for "unknown" product *p* is the same as for the reference product, *r*.

then $P^p_{\,i} = P^r_{\,i}$ and

$$R_i(p|r) = \frac{D^p_{\text{yield}} \times A^p}{D^r_{\text{yield}} \times A^r} = \frac{\ln(Y^p)}{\ln(Y^r)}, \quad \text{.....(28)}$$

making the scaling ratio independent of mechanism i, so

$$S^p(t) = \left[S^r(t)\right]^{\frac{D^p_{\text{yield}} \times A^p}{D^r_{\text{yield}} \times A^r}} \quad \text{.....(29a)}$$

$$S^p(t) = \left[S^r(t)\right]^{\frac{\ln(Y^p)}{\ln(Y^r)}} \quad \text{.....(29b)}$$

Eq. (29a) is a simple equation which allows scaling of the survival function by area and defect density. The mathematical relationship between defect density and yield has been studied extensively[4]. The simplest and most pessimistic formula is the simple exponential formula assuming Poisson statistics. For simplicity, we have assumed Poisson statistics, but other more realistic (or proprietary) formulae could be used to extract the defect density from yield statistics. The defect density so extracted might not be the area-weighted defect density identifiable in Eq. (23), but it could still be used in Eq. (29a). Eq. (29b), however, is specific to Poisson statistics.

Depending on the level of approximation and information, expressions in Eq. (19), (24) or (28) may be used to estimate $R_i(p|r)$. The approximation we have used in a defect reliability prediction model is Eq. (29a), and its validity depends on the 5 conditions given in this section. In the usual application of Eq. (29a), we assume that yield and fully analyzed reliability data (that is $S_i^r(t)$) are available for a "known" or "reference" product, and a reliability prediction about an unknown product is desired. The yield for the factory or lot which produced the "unknown" product is also assumed to be known.

In this section we have assumed that the reference product survival function is known at the same environmental conditions as the unknown product. This condition is removed in the next section.

4. ACCELERATED STRESSING AND BURN IN

Baseline product reliability data is usually acquired at accelerated conditions of high temperature and high voltage, whereas reliability estimates for products are typically needed at the milder environmental conditions encountered in service. Also, products are often subjected to a preliminary period of accelerated "burn-in" stress to weed out weak units that are likely to fail early. In this section we shall extend the model given in Eq. (18) to allow prediction of product reliability at stress conditions different from the reference product data, and after burn-in.

To begin, we define the acceleration factor. If a certain proportion of a population stressed at a standard environmental stress with temperature T_1 and operating voltage V_1 fails by mechanism i in a time interval dt_1 and the same proportion fails by mechanism i in a time interval dt_2 at an accelerated, AF > 1 (or decelerated, AF < 1), stress condition with temperature T_2 and voltage V_2 , then the time-varying acceleration factor is defined by $dt_1/dt_2=\mathrm{AF}_i(2|1)$, or for constant acceleration:

$$t_1 = \mathrm{AF}_i(2|1)t_2 \quad \text{.....(30)}$$

where $\mathrm{AF}_i(2|1)$ is the acceleration of environmental condition 2 relative to environmental condition 1 for mechanism i.[5] For example, if environmental condition 2 is more accelerated than condition 1, then, $\mathrm{AF}_i(2|1) > 1$ and $t_2 < t_1$.

If we know the CDF, PDF, instantaneous failure rate, and cumulative hazard at the standard environmental condition 1, then corresponding functions at the accelerated or decelerated condition 2 are given by:

$$f_i(2|t) = \mathrm{AF}_i(2|1) \times f_i\{1|\mathrm{AF}_i(2|1)t\} \quad \text{.....(31a)}$$

$$h_i(2|t) = \mathrm{AF}_i(2|1) \times h_i\{1|\mathrm{AF}_i(2|1)t\} \quad \text{.....(31b)}$$

$$F_i(2|t) = F_i\{1|\mathrm{AF}_i(2|1)t\} \quad \text{.....(31c)}$$

$$S_i(2|t) = S_i\{1|\mathrm{AF}_i(2|1)t\} \quad \text{.....(31d)}$$

where we have introduced the notation that, for example, $S_i(2|t)$ is the survival function at environmental condition 2 as a function of t for mechanism i.

We choose to use the following acceleration factor function:

---

4 C. H. Stapper, "Integrated Circuit Yield Statistics," Proc. IEEE, Vol. 71 (1983), pp 453-470.

5 Acceleration factor functions have the general properties:

$$\mathrm{AF}_i(z|x) = \mathrm{AF}_i(z|y) \times \mathrm{AF}_i(y|x)$$

$$\mathrm{AF}_i(y|x) = 1/\mathrm{AF}_i(x|y)$$

$$\mathrm{AF}_i(x|x) = 1$$

$$\mathrm{AF}_i(2|1) = \exp\left\{\frac{Q_i}{k}\left[\frac{1}{T_1} - \frac{1}{T_2}\right] + C_i(V_2 - V_1)\right\} \quad \text{.....(32)}$$

where $k = 8.61\times10^{-5}$ eV/°K. The thermal activation energy and voltage acceleration constant for mechanism i are Q_i (eV) and C_i (volts^{-1}), respectively.

Thus, the survival probability of an unknown product p at temperature T_2 and V_2 (which might be the environmental condition in a service application) may be calculated from the survival probability of a reference product r at accelerated conditions T_1 and V_1 by taking into account the contribution from all mechanisms:

$$S^p(2|t) = \prod_i [S_i^r\{1|\mathrm{AF}_i(2|1)t\}]^{R_i(p|r)} \quad \text{.....(33)}$$

where we have used Eqs. (18), (30), and (31d). The time argument on the right-hand side of Eq. (33), $\mathrm{AF}_i(2|1)t$, may be interpreted as the time at conditions T_1 and V_1 which is equivalent to time t at conditions T_2 and V_2. Equation (33) permits calculation of the survival probability for any product at any environmental condition from baseline accelerated reliability and yield data, provided yield data is known for the unknown product.

Burn-in is often used to remove reliability defects. The proportion of the original population which survives burn in for a time t_B at T_B and V_B is

$$S^p(B|t_B) = \prod_i [S_i^r\{1|\mathrm{AF}_i(B|1)t_B\}]^{R_i(p|r)} \quad \text{.....(34)}$$

and after an additional time t at T_2 and V_2 the proportion surviving is

$$\tilde{S}^p(2|t) = \prod_i \left[S_i^r\{1|\mathrm{AF}_i(2|1)t + \mathrm{AF}_i(B|1)t_B\}\right]^{R_i(p|r)}$$

so that the probability that a unit will survive the additional stress for time t at T_2 and V_2, *given that a unit has survived burn-in*, is $S'^p(2|t) = \tilde{S}^p(2|t)/S^p(B|t_B)$ so

$$S'^p(2|t) = \prod_i \left[\frac{S_i^r\{1|\mathrm{AF}_i(2|1)t + \mathrm{AF}_i(B|1)t_B\}}{S_i^r\{1|\mathrm{AF}_i(B|1)t_B\}}\right]^{R_i(p|r)} . \quad \text{.....(35)}$$

When dealing with life-test data, the cumulative percentage of failures is typically less than 1%. If Eq. (1) is substituted into Eq. (35) and only leading terms in the CDF are retained, we find

$$F'^p(2|t) \cong \sum_i R_i(p|r)\left[F_i^r\{1|\mathrm{AF}_i(2|1)t + \mathrm{AF}_i(B|1)t_B\} - F_i^r\{1|\mathrm{AF}_i(B|1)t_B\}\right] + \mathrm{O}(F^2) \quad . \quad (36)$$

In this limit, it is apparent that the scaling ratio is a multiplicative factor for the net post burn-in fallout for each mechanism, and the acceleration factors compress the time axis, while burn-in gives a time-axis offset and vertical axis offset. The proportional error in Eq. (36) is of the order of the value of the CDF. That is, for example, if the CDF is 1%, then the error in the CDF is 1% of 1%, or ±0.01%.

Equation (35) is a complete model which gives the survival function of an unknown product in terms of a known baseline reliability model. It models the following effects by various expressions for the scaling ratio $R_i(p|r)$:

- Lot-to-lot or factory-to-factory yield variations via Y^p in Eq. (28).
- Product-to-product sensitivities to failure mechanisms via die areas A^p in Eq. (28).
- Acceleration factors between accelerated test conditions and use conditions via the acceleration factor given in Eq. (32).
- Burn-in time and conditions via Eq. (35)

5. ANALYSIS OF RELIABILITY DATA AND MODEL EXTRACTION

In the course of process development a large amount of reliability data is acquired using SRAMs design-optimized for reliability evaluation. Once a process is transferred to production, failures are analyzed and production burn-in data are acquired for a variety of products. Additionally, samples of various products are subjected to extended life-test at accelerated conditions of temperature and bias. This accumulated body of data is termed "baseline" reliability data. This data is acquired at a variety of times and temperatures, for different products (with various die sizes), and from factories and lots with varying yields. An example of one of the many data sets comprising the baseline lot data for a process is given in Table I[6]. In Table I the yield defect density is given in arbitrary units and is taken to be unity ($D_{\text{yield}} = 1$).

Table I. Example SRAM life test data at $D_{\text{yield}} = 1$ (arbitrary units).
Die area = $160 \times 226 = 36160$ mils^2.

	Hours at $V = 5.5$V and $T_j = 131$°C							
	6	12	24	48	168	500	1000	2000
Pass. Defect (PD)	No data	No data	1	0	0	0	0	1
Fab Defect (FD)	No data	No data	3	0	0	2	0	0
Bake Recov. (BR)	No Data	No Data	0	0	0	0	0	0
Junct. Spike (JS)	No Data	No Data	0	0	0	0	0	0
Sample Size (SS)	No Data	No Data	2748	2744	2743	2293	2290	2290

Table II. Acceleration model parameters for mechanisms indicated in Table I.

Mechanism	Q_i (eV)	C_i (1/volts)
PD	0.3	1.8
FD	0.5	2.0
BR	1.0	0.0
JS	1.0	0.6

The method of deriving a "reference" reliability model from all of the baseline data sets is summarized in Fig. 6. We show in Fig. 6 that many data sets comprising the baseline data (one of which is shown in Table I) are consolidated into a "reference" data set by scaling the data using the "scaling ratio" defined in Eq. (28) and the acceleration factor, Eq. (32), using acceleration parameters (assumed known from fundamental studies) in Table II.

The result of consolidation of many baseline data sets such as that in Table I, is the "Reference Lot" data in Table III. This may be thought of as a re-tabulation at *one* arbitrarily-chosen "reference" condition of temperature, bias, yield defect density, and die area, of data taken at *many* conditions of temperature, etc

Data in which the sample size changes from readout to readout is called "multi-censored" data. Normally, sample sizes will only decrease as stress time accumulates (because of failures and removals). However, in the case of Table III, sometimes sample sizes are seen to increase. This is a result of the way in which data which has incommensurate readout intervals must be combined. Multi-censored data may be analyzed by using hazard plotting, but this has the disadvantage of not providing confidence intervals. An equivalent method which provides upper and lower confidence intervals to the cumulative distribution function has been described by Nelson[7]. This method is the Kaplan-Meier method of calculating the survivor function, extended by Greenwood to provide estimates of the variance (the "KMG" algorithm). Data in Table III is plotted in Fig. 7 using the KMG algorithm.

Once the data in Table III has been plotted using the KMG algorithm, distribution functions may be fitted to the plotted data. We have chosen to use lognormal distributions, although other distribution functions such as Weibull or piece-wise exponential could be used. To fit a lognormal distribution through a set of cumulative proportion failing (F_i) versus time (t_i) data, we transform the data into (x,y) points according to

$$x_i = \ln(t_i); \quad y_i = \Phi^{-1}(F_i). \qquad \text{.....(37)}$$

6 The data and models in this report are examples only, and do not represent the reliability of any Intel product.

7 W. Nelson, "Accelerated Testing,", John Wiley & Sons (1989), pp145-151

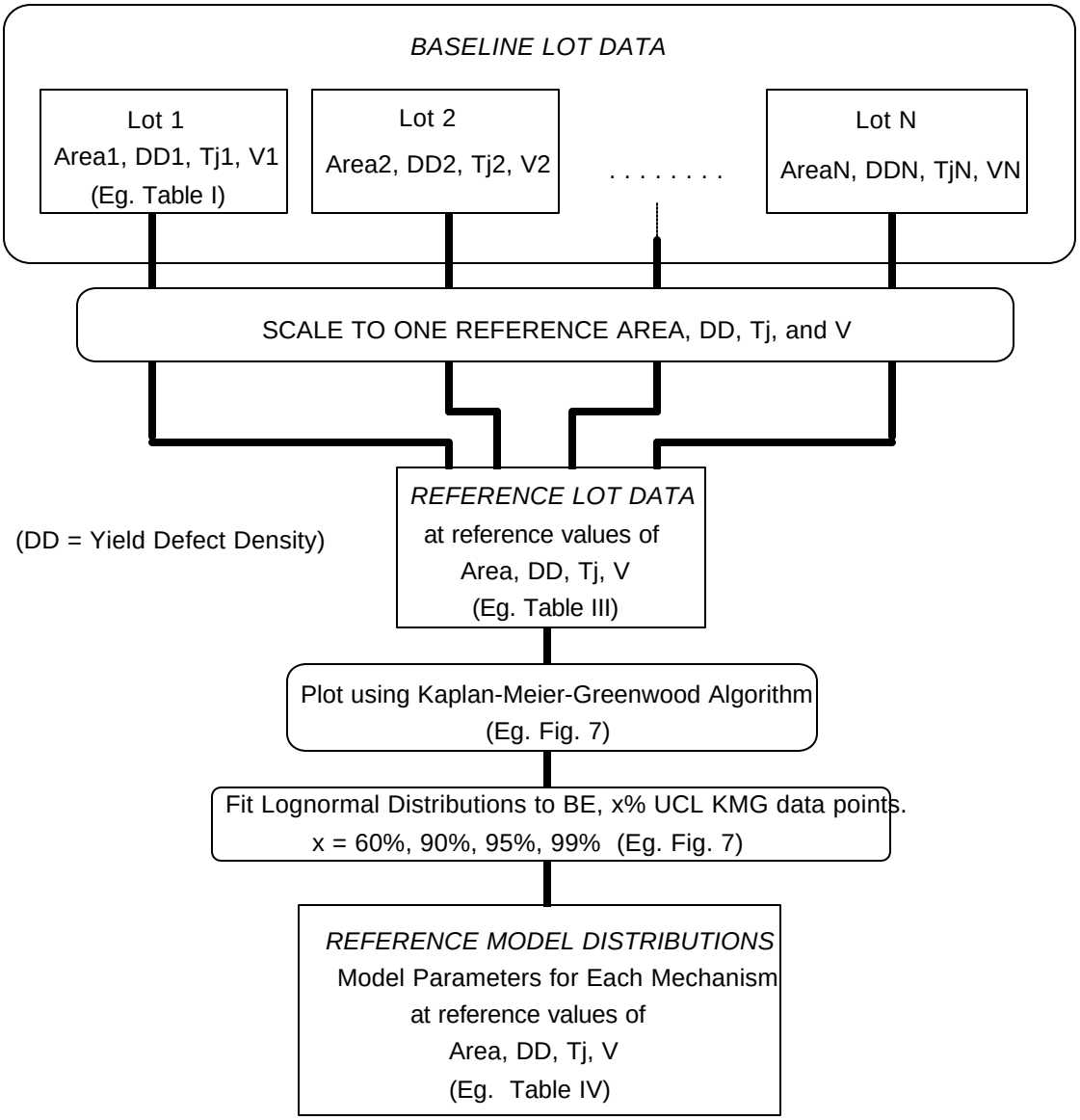

Fig. 6 Summary of the procedure for deriving a process reference model from baseline reliability data sets.

Table III. Data derived from a variety of products all produced on the same process as the SRAM data of Table I. Data has been scaled and consolidated to a virtual reference condition of V = 7 volts, T_j=160°C, Area = 268686 mils^2, and D_{yield} = 0.21 (arbitrary units). This is an example of "Reference Lot Data" in Fig. 6. These data are examples only, and are not representative of Intel products.

	Hours							
	6	12	24	48	168	500	1k	2k
PD	0	0	1.6	0	0	0	3.2	6.2
SS for PD	22642	1609	38305	51551	45212	5480	11808	5297
FD	105.7	0	18.6	54.0	53.9	19.1	24.8	20.4
SS for FD	21056	1407	34973	48604	42288	4304	10409	4207
BR	0	0	7.3	4.6	0	0	7.7	0
SS for BR	18281	1059	29629	47932	39302	3798	9383	3632
JS	0	0	2.9	0	27.7	7.5	0	9.6
SS for JS	18281	1059	29155	45472	37964	3015	8616	2958

where Φ^{-1} is the inverse normal probability function. Next we determine σ and μ from the slope of a line fitted through the (x,y) points from

$$s = 1/\text{slope}; \quad m = -s \times \text{intercept}. \qquad \text{.....(38)}$$

σ and μ are determined for the best estimate by a least-squares fit of straight lines through KMG ordinates on a lognormal plot (symbols in Fig. 7) for each mechanism. Only readouts with non-zero failures are used in the fit. If there is only one non-zero readout, a default value of σ based on experience with similar mechanisms is chosen. (This scenario occurs infrequently.) Distributions for 60%, 90%, 95%, and 99% one-sided upper confidence limits are derived by a slope-constrained least-squares fit of straight line distributions through the corresponding upper confidence KMG ordinates, using the best-estimate value of σ. This procedure gives a set of parameters, such as shown in Table IV, defining the "Reference Model Distributions" in Fig. 6.

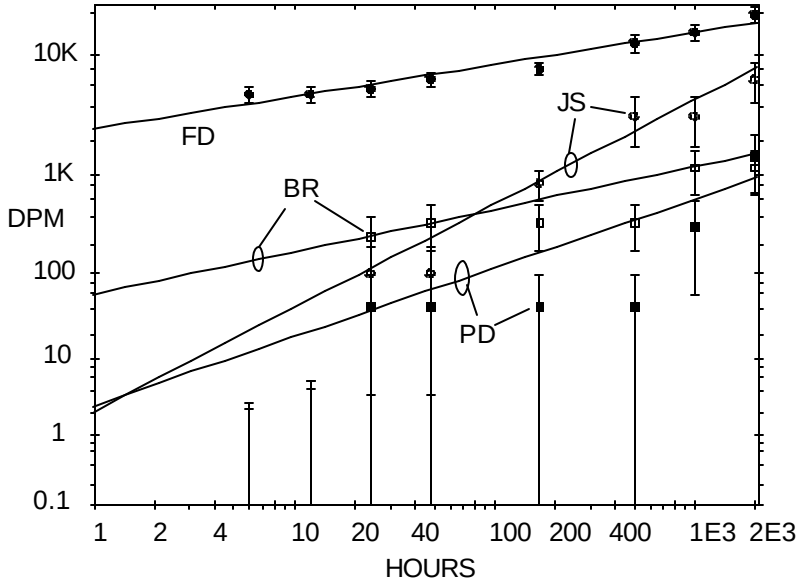

Fig. 7 Lognormal probability plot of cumulative distribution functions for data from Table III plotted using KMG algorithm. Error bars are 90% two-sided (95% one-sided) confidence limits. Superimposed on the plots are "Best Estimate" fitted distributions (sloped straight lines) using least-squares fitting through readouts with non-zero failures. The reference conditions for this plot are the same as in Table III: V = 7 volts, T_j = 160°C, A = 268686 mils^2, D_{yield} = 0.21 (arbitrary units).

Table IV. Reference Model Distributions for data of Table III at reference conditions of V = 7 volts, T_j = 160°C, A = 268686 mils^2, D_{yield} = 0.21 (arbitrary units). From fits to KMG ordinates plotted in Fig. 7. These models are examples only, and are not representative of Intel products.

Mechanism	σ	μ Best Est.	μ 60% UCL	μ 90% UCL	μ 95% UCL	μ 99% UCL
PD	5.24	23.94	23.76	23.20	23.05	22.79
FD	11.20	31.33	31.24	30.90	30.78	30.57
BR	8.51	32.81	32.63	32.00	31.81	31.49
JS	3.47	16.00	15.92	15.65	15.58	15.44

The reference model survival function for mechanism i at the reference conditions of bias, temperature, defect density, and die area are given by

$$S_i^r(t) = 1 - \Phi\left(\frac{\ln(t) - m_i}{s_i}\right). \quad \text{.....(39)}$$

where the appropriate values of σ and μ are selected from Table IV for the particular example used here.

The set of functions (one per mechanism) in Eq. (39) can be used in Eq. (35) to make model predictions for other products at other environmental conditions, and after various amounts of burn-in.

6. RELIABILITY PREDICTIONS

We demonstrate applications of the model by discussion of several specific examples. The reference model derived in Section 5, and summarized in Table IV and Table II, will be used in the examples.

Effect of Burn-In on SRAM Reliability.

We wish to calculate the survivor function at 85°C and 5 volts for an SRAM, with die area 36160 mils^2, with a lot defect density D_{yield} = 1 (arbitrary units), and with and without 10 hours of 125°C 5.5 volt burn in. The first step is to calculate $R_i(p|r)$ from Eq. (28):

$$R_i(\text{SRAM}|\text{reference}) = \frac{1 \times 36160}{0.214 \times 268686} = 0.628.$$

for each mechanism. This scaling ratio, and the reference model survivor function, Eq. (39), using the parameters in Table IV, is substituted into Eq. (35). The acceleration factors in Eq. (35), between the reference model and the SRAM, are calculated from Eq. (32) using acceleration parameters from Table II. The results of the calculation are shown in Fig. 8.

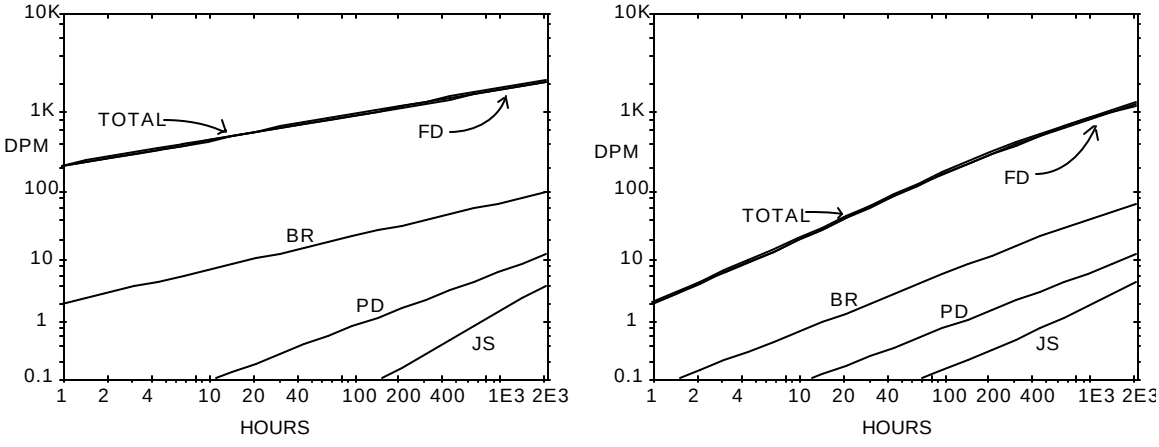

Fig. 8 Lognormal plots of cumulative distribution functions of SRAM (Area = 36160 mils^2, D_{yield} = 1, arbitrary units) failures at 85°C and 5 volts. Distributions for each mechanism, and the total, is shown. Notice that

the total is dominated by "fab defects" (FD). Left: No Burn-in. Right: After 10 hours of 125°C 5.5 volt burn in.

Fit of Model to Baseline Data

It is useful to compare the predictions from the reference model based on the entire baseline lot data set with one or two individual data sets selected from the baseline data set. In Fig. 9 we compare a large die microprocessor at low defectivity and high stress and a small die (SRAM) at higher defectivity and low stress with the model predictions.

The effect of area, yield defect density (given in arbitrary units), and stress on total cumulative failures (all mechanisms) are illustrated in Fig. 9 by changing the values from those for the microprocessor to those for the SRAM in the sequence:

1. Area = 268686 mils^2, $D_{\text{yield}} = 0.21$, $T_j = 160$°C, $V = 7$ volts. (Conditions of microprocessor lot data.)
2. Area = 36160 mils^2, $D_{\text{yield}} = 0.21$, $T_j = 160$°C, $V = 7$ volts.
3. Area = 36160 mils^2, $D_{\text{yield}} = 1$, $T_j = 160$°C, $V = 7$ volts.
4. Area = 36160 mils^2, $D_{\text{yield}} = 1$, $T_j = 125$°C, $V = 6$ volts. (Conditions of SRAM lot data.)

The fit of the SRAM data set to the model is very good, and the fit of the microprocessor data is quite good considering the magnitude of the area, yield defect density, and stress effects. Moreover, the model prediction is conservative for the microprocessor.

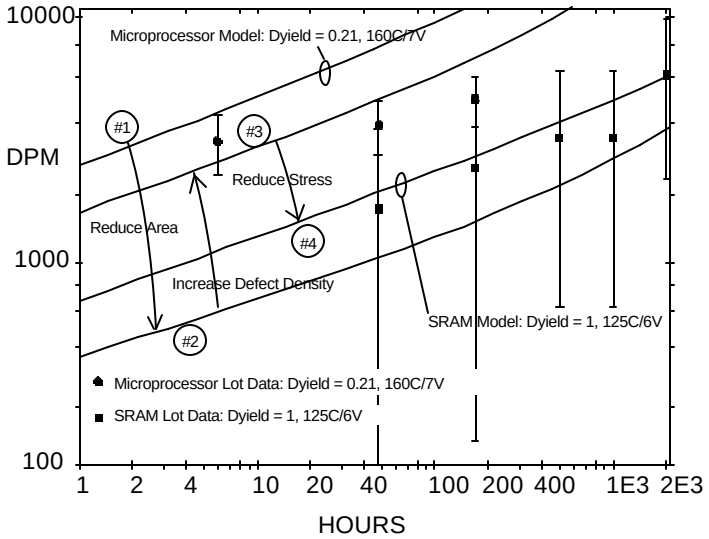

Fig. 9 Log-normal probability plot of best estimate model fits of total failures (all mechanisms) to two baseline data sets. One is a large die (414 x 649 mils) microprocessor at lower defectivity ($D_{\text{yield}} = 0.21$, arbitrary units) and high stress (160°C/7V); the other is a small die (160x226 mils) SRAM at higher defectivity ($D_{\text{yield}} = 1$, arbitrary units) and low stress (125°C/6V). Model predictions for the microprocessor (#1) and SRAM (#4) are shown with intermediate model curves showing the effect of reducing area (#2), increasing yield defectivity(#3), and finally, reducing stress (#4). Neither baseline data sets, nor model predictions have burn in. The data and models are examples only and are not representative of actual Intel products.

Calculation of Reliability Indicators

All reliability indicators can be calculated from the survival function, Eq. (35). In the following we indicate "post-burn-in" by the prime on the S (i.e. S'). Of course, the non-burned-in case can be derived by setting the burn-in time t_B in Eq. (35) to zero. All times in the following are "post-burn-in" times.

Intel's reliability indicators are defined at die conditions of 85°C and 5 volts[8] as follows:

---

[8] In newer technologies, a lower nominal voltage such as 3.3 volts is used.

- Infant Mortality[9]: 0 - 100 hours cumulative defects per million (DPM) at 85°C and 5 volts[8]

$$10^6 \times \{1 - S'(t = 100 \text{ hours})\}$$

- Early Life Average Failure Rate (AFR): 0 - 1 year average failure rate in FITs at 85°C and 5 volts[8], using Eq. (6),

$$-10^9 \times \ln[S'(t = 8760 \text{ hours})]/8760$$

- Long Term Average Failure Rate: 1 - 10 years average failure rate in FITs at 85°C and 5 volts[8], using Eq. (6),

$$10^9 \times \{\ln[S'(t = 8760 \text{ hours})] - \ln[S'(t = 87600 \text{ hours})]\}/78840$$

These indicators may be calculated as best estimates, or upper confidence limits at various percentage of significance, depending on the model parameters selected from Table IV. Typically, 60% one-sided UCL values are quoted.

Table V. Example of microprocessor 60% UCL reliability indicators calculated from reference model example used in this paper. Indicators have been calculated with and without burn-in, showing the effect of burn-in. These calculated indicators are examples only and are not representative of Intel products.

	Mech	0-100 hr DPM @ 85/5	0-1yr DPM @ 85/5	0-1 yr AFR (FIT) @ 85/5	1-10 yr AFR (FIT) @ 85/5
No Burn In	PD	2	69	8	4
	FD	1406	4827	552	48
	BR	39	305	35	6
	JS	0	42	5	6
	Total	1447	5241	600	65
Burn In = 168 hours @ 160°C and 7 volts	PD	0.4	35	4	3
	FD	1.6	133	15	13
	BR	0.5	45	5	4
	JS	0.6	52	6	6
	Total	3.1	266	30	25

7. SUMMARY

This paper shows how to use SRAM and product reliability data acquired during process development and early production to establish a baseline data set. The baseline data set is used to derive a reference model characteristic of each process. The reference model exploits the relationship between yield defects and latent reliability defects. This relationship makes it possible to predict the reliability of a product knowing only the sort yield data, die size, and, optionally, the yield defect Pareto. The reference model may therefore be used to predict the reliability characteristics of complex random logic VLSI products without the necessity of extensive (and expensive) reliability testing, including failure analysis, of each product.

It is necessary to make certain assumptions in the derivation of the model, and this paper enumerates them. However, the validity of the model really depends more on how well the single reference model actually fits the large number of data sets from many different products and conditions which comprise the baseline data set. In this paper we made a small demonstration of the fit using very different products; a microprocessor and an SRAM. Much more extensive comparisons of this type have been performed, giving us good confidence in the model.

The technique described in this paper provides methods to:

- Estimate any product's reliability characteristics, including the contributions of various mechanisms.
- Estimate failure rates of complex products without full reliance on failure analysis or complete data.

---

9 Intel now uses 0 - 50 hours as the defining period for infant mortality (6/16/95).

- Estimate the effect of die size, array area, etc. on the reliability characteristics of any proposed or new product using no or minimal data.
- Quantify the reliability benefits of process continuous improvement through defect density reduction.
- Calculate the effect of burn-in for any product.
- Calculate reliability indicators useful to customers at any desired level of confidence.

ACKNOWLEDGMENTS

Thanks are due to Ron Smith, Sonny Seaton, Chuck Fernald, Mary Beth Shew, and Babak Sabi for their contributions to the development of this model.