1991

Estimating The Quality Of Manufactured Digital Sequential Circuits

Dharam Vir Das
University of Nebraska-Lincoln

Sharad C. Seth
University of Nebraska-Lincoln, seth@cse.unl.edu

Vishwani Agrawal
AT&T Bell Laboratories

ESTIMATING THE QUALITY OF MANUFACTURED DIGITAL SEQUENTIAL CIRCUITS

Dharam Vir Das and Sharad C. Seth
University of Nebraska Lincoln, Nebraska 68588

Vishwani D. Agrawal
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

Abstract: Detection of a fault in a sequential circuit requires a sequence of test vectors. This sequence activates the fault and propagates the effect of the fault to a primary output. To accomplish this, the test sequence must set flip-flops through a series of states. Unlike a combinational circuit, many faults in a sequential circuit cannot be detected by a single vector. We propose a statistical model in which a fault is characterized by two parameters: a per-vector detection probability and an integer-valued latency. Irrespective of its detection probability, the fault cannot be detected by a vector sequence shorter than the latency. A joint distribution of the latency and detection probability over all the failed chips is thus obtained. Using the new model, an analysis of chip failure data to predict actual yield and reject ratio is given. For a large-volume CMOS chip, tested by vectors having 99.7% fault coverage, this analysis gives a reject ratio of 43 parts per million that is believed to be in close agreement with the field data.

1 Introduction

The detection of faults in sequential circuits differs from that in combinational circuits. A test vector in a combinational circuit activates the fault and propagates the fault effect to a primary output. Since there is no memory in the circuit, any given fault can be detected by a single vector. In a sequential circuit, the fault activation may require the control of several flip-flops. Depending on the levels (sequential depth) of flip-flops in the circuit, several test vectors will be needed. Further, the fault effect may have to be propagated through several levels of flip-flops, again requiring a sequence of vectors. This phenomenon is evident in the fault coverage versus vectors graphs. For sequential circuits, it is not uncommon to find no increase in coverage for several vectors followed by a sudden jump. Similar jumps are seen in the yield versus test-vector graph of sequential chips. This paper presents a statistical model of fault detection in sequential circuits.

In general for combinational circuits the coverage smoothly rises with vectors. However, jumps can still occur if the vectors are specifically generated to test different parts of the circuits.

Most existing models characterize faults by their detection probabilities. These are per-vector probabilities and result in a smoothly increasing cumulative probability of detection as more vectors are applied. Such a model cannot produce any jumps in the coverage or yield graphs. The distribution of detection probabilities is obtained by fitting the model to experimental data. This technique, however, predicts a pessimistically high reject ratio (low quality level). A study [1] of several previously proposed models [2] [3] [4] [5] [6] was recently published. For a VLSI chip, tested with vectors having a 99.7% coverage, the reject ratio was estimated by these techniques in the range of 480 to 5,320 parts per million (ppm). While it was hard to obtain statistically large amount of field data, these reject ratios were still considered too pessimistic. This conclusion was also supported by the less than satisfactory fit that all models provided for the experimental data.

The idea of fitting a model to experimental data is to eliminate random variations. However, the delayed detection of faults due to the sequential nature of the circuit cannot be considered a random phenomenon; the stepped decrease in yield is *real* and not random. Any attempt at fitting a smoothly rising curve will,

INTERNATIONAL TEST CONFERENCE 1991

therefore, lead to erroneous result.

We recognize that a fault embedded in a chain of flip-flops will require several vectors. Even if the fault is active, it is possible to have a zero probability of detection because there may be several flip-flops between the fault effect and a primary output. Therefore, in addition to the per-vector detection probability we introduce fault *latency*. The paper describes the theoretical development of this two-parameter model and its application to real VLSI test data.

2 Chip Failure Analysis

In the past, similar models have been used for analyzing fault coverage and yield as functions of the number of vectors [6] [7]. In this section we develop a new model for yield as a function of vectors.

A fault is characterized by a per-vector detection probability. Vectors are assumed to be independent in their fault detection capability. While such a model may be adequate for combinational logic, it does not account for the delayed detection problem in sequential circuits as discussed in the previous section. We, therefore, introduce an additional parameter in our model. Associated with each fault we have an integer called *latency* or the number of test vectors that must be applied before the fault is considered detectable by the subsequent vectors. The latency is zero for all detectable stuck type faults in a combinational circuit, even though the detection probability can be near-zero for a fault in one part of the circuit when the vectors are are biased to cover faults in another part. Nevertheless, the detection probability cannot be zero unless the fault is redundant. The latency of some faults in a sequential circuit may also be zero; such faults may be called *combinational* while the faults with non-zero latency will be called *sequential.*

The key assumption in all previous work is that associated with each collapsed fault class there is a fixed and independent per-vector detection probability, called *detectability.* This assumption is reasonable for a combinational fault but must be modified for a sequential fault. Under our model, a sequential fault can only be detected after its latency period is over, hence its detectability must be zero during the latency period.

We will assume that chip failure on an applied test vector is a random event. For a chip with a fault we can speak of the following random variables.

1. D: A random variable representing the latency of a fault. It takes values in the set $\{0, 1, 2, \ldots, \infty\}$.
2. X: A random variable representing the detection probability of the fault of latency d i.e. probabilty that a fault with latency d has occured and is detected by the vector. $0 \leq x \leq 1$.
3. G: $G = g_n(X, D)$. A function of two random variables. This represents the probability that a chip fails at the n^{th} vector.

$$g_n(x, d) = x(1-x)^{n-d-1} I_{\{d+1,\ldots,\infty\}}(n) \quad (1)$$

In the above expression $I_{\{d+1,\ldots,\infty\}}(n)$ is the indicator function[1] [8]. The expression indicates that a chip with a fault of latency d cannot fail for the first d vectors. Therefore its failure probability is determined by the per vector detection probability x.

We can associate a density function $f(x, d)$ with the random variables X and D. Then $f(x, d)\Delta x$ is the fraction of chips characterized by latency d that fail on a vector with probability between x and $x + \Delta x$. If y is the yield then $1 - y$ of the total chips can fail. Hence we can write

$$f(x, d) = y\delta(x, d) + p(x, d) \quad (2)$$

where $\delta(x, d)$ is the Kronecker delta function and $p(x, d)$ is the partial density function.

Since $f(x, d)$ is the density function we should have

$$\sum_{d=0}^{\infty} \int_0^1 f(x, d) = y + \sum_{d=0}^{\infty} \int_0^1 p(x, d) dx$$

[1] Indicator Function. Let Ω be any space with points ω and A any subset of Ω . The indicator function of A, denoted by $I_A(.)$ is the function with domain Ω and counterdomain the set $\{0, 1\}$, such that

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Suppose after application of n test vectors a certain fraction of chips has not failed. Then the expected value of this fraction is the yield of chips after n vectors and is denoted by y_n.

$Prob$(A chip does not fail after n vectors)

$$\begin{aligned} &= \begin{cases} 1 & n \le d \\ (1-x)^{n-d} & n > d \end{cases} \\ &= I_{\{0,\ldots,d\}}(n) + (1-x)^{n-d} I_{\{d+1,\ldots,\infty\}}(n) \end{aligned}$$

Therefore,

$$y_n = \sum_{d=0}^{\infty} \int_0^1 [I_{\{0,\ldots,d\}}(n) + (1-x)^{n-d} I_{\{d+1,\ldots,\infty\}}(n)]\, f(x,d)dx \quad (3)$$

Determination of $p(x,d)$. We will assume, *a priori*, that the random variables X and D are independent. The assumption is justified by considering that the detection probability and latency of a fault depend on rather independent circuit characteristics. Detection probability is strongly influenced by the functionality of the circuit while latency is affected more by the location of the fault site relative to the flip-flops in the circuit. Under the assumption, $p(X,D) = p_X(x)p_D(d)$.

From Equation(1) we know that the probability that a chip fails at vector number i is $x(1-x)^{i-d-1}$, $i > d$. Let N be the test length. Therefore i takes the value from 1 and N. Since X and D are random variables we can use Bayes' theorem, to write the probability that a chip fails at the i^{th} vector as,

$p_i(x,d) =$

$$\frac{x(1-x)^{i-d-1} I_{\{0,\ldots,i-1\}}(d) q_X(x) q_D(d)}{\sum_{d=0}^{\infty} \int_0^1 x(1-x)^{-d-1} I_{\{0,\ldots,i-1\}}(d) q_X(x) q_D(d) dx}$$

where, $q_X(x)$ and $q_D(d)$ are the *a priori* distributions of the detection probability and the latency. For simplicity we may assume uniform distribution for $q_X(x)$ and $q_D(d)$ with d taking integer values from 0 to N. Here we are assuming that test length is sufficiently long so as to include all the latencies.

$$\begin{aligned} q_X(x) &= 1 \qquad 0 \le x \le 1 \\ q_D(d) &= \frac{1}{N+1} \qquad 0 \le d \le N \end{aligned}$$

Hence,

$$\begin{aligned} p_i(x,d) &= \frac{x(1-x)^{i-d-1} I_{\{0,\ldots,i-1\}}(d)}{\sum_{d=0}^{N-1} \int_0^1 x(1-x)^{i-d-1} I_{\{0,\ldots,i-1\}}(d) dx} \\ &= \frac{i+1}{i} x(1-x)^{i-d-1} I_{\{0,\ldots,i-1\}}(d) \\ &= k_i x(1-x)^{i-d-1} I_{\{0,\ldots,i-1\}}(d) \qquad (4) \end{aligned}$$

where $k_i = \frac{i+1}{i}$.

The probability that a chip does not fail on the application of a test sequence is

$$(1-x)^{N-d} \qquad 0 \le d \le N$$

The corresponding Bayesian probability distribution is

$$\begin{aligned} p_0(x,d) &= \frac{(1-x)^{N-d} I_{\{0,\ldots,N\}}(d)}{\sum_{d=0}^{N} \int_0^1 (1-x)^{N-d} dx} \\ &= \frac{(1-x)^{N-d} I_{\{0,\ldots,N\}}(d)}{\sum_{i=1}^{N+1} \frac{1}{i}} \\ &= k_0 (1-x)^{N-d} I_{\{0,\ldots,N\}}(d) \qquad (5) \end{aligned}$$

where $k_0 = 1/\sum_{i=1}^{N+1} \frac{1}{i}$.

Let a sample of c chips be tested by a sequence of N vectors. As these vectors are applied we record the number of chips that fail for the first time on each vector. Let c_i denote the number of such chips for vector number i. If y is the true yield then $(1 - y - \frac{1}{c}\sum_{i=1}^{N} c_i)$ is the fraction of chips that are bad but did not fail on any of the vectors from 1 thru N. To determine the complete chip failure probability distribution $p_i(x,d)$ is weighted with $\frac{c_i}{c}$ and $p_0(x,d)$ is wieghted with $(1 - y - \frac{1}{c}\sum_{i=0}^{N} c_i)$.

$$p(x,d) = (1 - y - \frac{1}{c}\sum_{i=1}^{N} c_i) p_0(x,d) + \frac{1}{c}\sum_{i=1}^{N} c_i p_i(x,d)$$

Substituting this in Equation (2), after substitution for $p_i(x,d)$ and $p_0(x,d)$ from Equations (4) and (5),

we get,

$$f(x,d) = y\delta(x,d) + (1-y-\frac{1}{c}\sum_{i=1}^{N} c_i)k_0(1-x)^{N-d}I_{\{0,\ldots,N\}}(d) + \frac{1}{c}\sum_{i=d+1}^{N} c_i k_i x(1-x)^{i-d-1}$$

The second term in the above equation is the faulty chips not rejected by any of the N vectors. The third term groups the chips according to the vector number at which they failed.

Substituting for $f(x,d)$ in Equation (3) we get,

$$\begin{aligned} y_n = & \sum_{d=0}^{\infty}\int_0^1 \left[I_{\{0,\ldots,d\}}(n) + (1-x)^{n-d}I_{\{d+1,\ldots,\infty\}}(n)\right] \times \\ & \left[y\delta(x,d) + (1-y-\frac{1}{c}\sum_{i=1}^{N} c_i)k_0(1-x)^{N-d}I_{\{0,\ldots,N\}}(d) + \frac{1}{c}\sum_{i=d+1}^{N} c_i k_i x(1-x)^{i-d-1}\right] dx \\ = & \sum_{d=0}^{\infty}\int_0^1 y\delta(x,d)I_{\{0,\ldots,d\}}(n)dx \\ & + \sum_{d=0}^{\infty}\int_0^1 y\delta(x,d)(1-x)^{n-d}I_{\{d+1,\ldots,\infty\}}(n)dx \\ & + \sum_{d=0}^{\infty}\int_0^1 (1-y-\frac{1}{c}\sum_{i=1}^{N} c_i)k_0 \times (1-x)^{N-d}I_{\{0,\ldots,d\}}(n)I_{\{0,\ldots,N\}}(d)dx \\ & + \sum_{d=0}^{\infty}\int_0^1 (1-y-\frac{1}{c}\sum_{i=1}^{N} c_i)k_0 \times (1-x)^{N+n-2d}I_{\{d+1,\ldots,\infty\}}(n)I_{\{0,\ldots,N\}}(d)dx \\ & + \sum_{d=0}^{\infty}\int_0^1 \frac{1}{c}\sum_{i=d+1}^{N} c_i k_i x(1-x)^{i-d-1}I_{\{0,\ldots,d\}}(n)dx \\ & + \sum_{d=0}^{\infty}\int_0^1 \frac{1}{c}\sum_{i=d+1}^{N} c_i k_i x(1-x)^{n+i-2d-1} \times I_{\{d+1,\ldots,\infty\}}(n)dx \end{aligned}$$

$$\begin{aligned} = & y + (1-y-\frac{1}{c}\sum_{i=1}^{N} c_i)k_0 \times \left[\sum_{d=0}^{n-1}\frac{1}{N+n-2d+1} + \sum_{d=n}^{N}\frac{1}{N-d+1}\right] \\ & + \frac{1}{c}\sum_{d=n}^{N-1}\sum_{i=d+1}^{N}\frac{c_i k_i}{(i-d+1)(i-d)} \\ & + \frac{1}{c}\sum_{d=0}^{n-1}\sum_{i=d+1}^{N}\frac{c_i k_i}{(n+i-2d)(n+i-2d+1)} \end{aligned}$$

We also compute the measured yield for N vectors as,

$$y_N = 1 - \frac{1}{c}\sum_{i=1}^{N} c_i$$

Substituting $y_n = y_N$ and $n = N$ in the above equation, we get

$$\begin{aligned} y_N = & y + (y_N - y)k_0\left[\sum_{d=0}^{N-1}\frac{1}{2N-2d+1} + 1\right] \\ & + \frac{1}{c}\sum_{d=0}^{N-1}\sum_{i=d+1}^{N}\frac{c_i k_i}{(N+i-2d)(N+i-2d+1)} \end{aligned}$$

Solving for y we have,

$$y = y_N - \frac{e_2}{1-e_1}$$

where,

$$e_1 = k_0\left[\sum_{d=0}^{N-1}\frac{1}{2N-2d+1} + 1\right]$$

and

$$e_2 = \frac{1}{c}\sum_{d=0}^{N-1}\sum_{i=d+1}^{N}\frac{c_i k_i}{(N+i-2d)(N+i-2d+1)}$$

The above expressions can be used for estimating the true yield, the apparent yield after n vectors, and the reject ratio from the chip failure data. We will illustrate this in the next section for actual data.

3 Experiment

We use the experimental data for the VLSI chip from a previous paper [1] to illustrate the above analysis. The data is for a CMOS chip manufactured in a Class 100 clean room. We obtained wafer test data for a sample of chips. These tests included parametric, continuity, and functional tests. The functional test consisted of 12,188 clock steps. A fault simulator was used to determine the cumulative fault coverage after each test vector. The final fault coverage was found to be 99.7%. The reader is referred to the cited paper for further details about the data collection effort.

Since our analysis refers only to the functional tests, we eliminated the chips that failed the parametric and the continuity tests and considered the yield after the functional test as a fraction of the remaining chips. The estimated functional yield of the chip after the 12,188 clock steps was 0.712954 and the estimated true yield was 0.712923. These two quantities were computed from the equations for y_N and y above. The resulting reject ratio, computed as $(y_N - y)/y_N$, is 43 ppm. Figure 1 shows the best fit obtained for the experimental data. Figure 2 gives the same data between 1 through 500 vectors at an enlarged scale to show how well our model can fit the jumps.

The enhanced accuracy of the latency model over our earlier model [6] is evident in Figure 3. The figure shows the best fit to the same experimental data in the earlier model with an estimated reject ratio of 5,320 ppm. As the earlier model must approximate the jumps in the chip-failure data by a smooth curve it is seen that the projected true yield is smaller thus resulting in a pessimistic estimate.

Figure 4 illustrates the behavior of the new model vs. the old model close to the vector number after which no chip failures are recorded. The graphs for the two model meet at this point. This is so because in both the models the measured yield is equated to estimated yield in order to solve for true yield. The values of yield after this vector are the extrapolations of the two models. The new model approaches the asymptotic value of 0.712923 whereas the old model tends to approach the value of 0.709163. In the extrapolated region the behavior of both the models is governed by the per vector detection probability.

Figure 5 is a surface plot for the failure probability density function $p(x, d)$ derived from the experimental data. For clarity of illustration only a limited range of the latency d is shown in the figure. The surface plot is highly non-uniform and has peaks of high value at certain latencies due to jumps in the yield vs. vectors graph. Such surface plots provide a measure of chip testability: a concentration of the mass at higher values of x and lower values of d is indicative of higher testability.

4 Conclusion

The phenomenon of latency of faults in sequential circuits has been observed for a long time. However, this is the first time a model for this phenomenon is presented. The model with two parameters, namely, per-vector detection probability and latency, has the necessary degrees of freedom to provide a close fit to experimental test data. As a result, reject ratio predictions will be more realistic. The analysis of production data for a CMOS chip shows that a 43 ppm reject ratio can be expected when the fault coverage of test vectors is 99.7 percent. Based upon experience, this appears to be a realistic estimate. When latency is neglected, the fit to data becomes crude and the analysis would have predicted a much higher reject ratio.

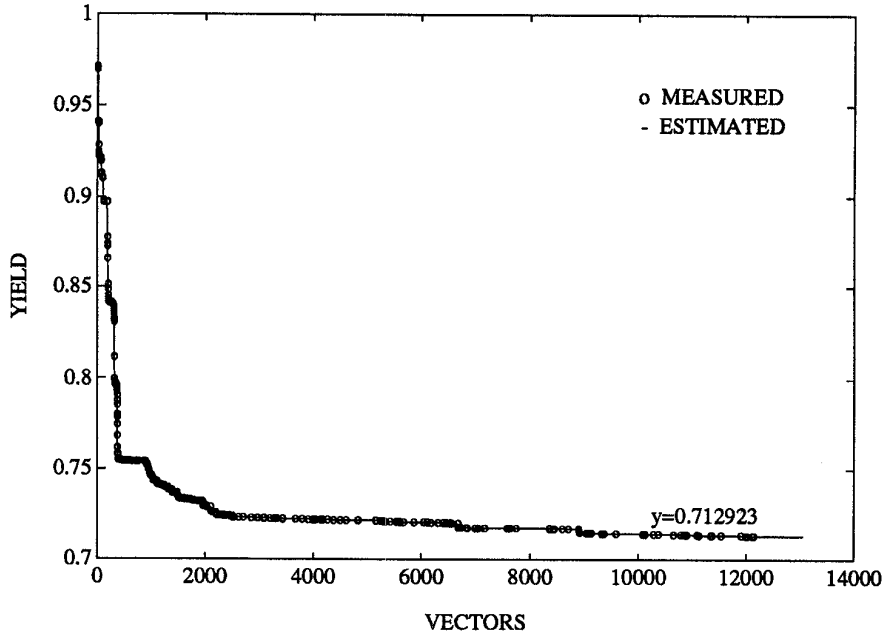

Figure 1: Fitting the Experimental Data

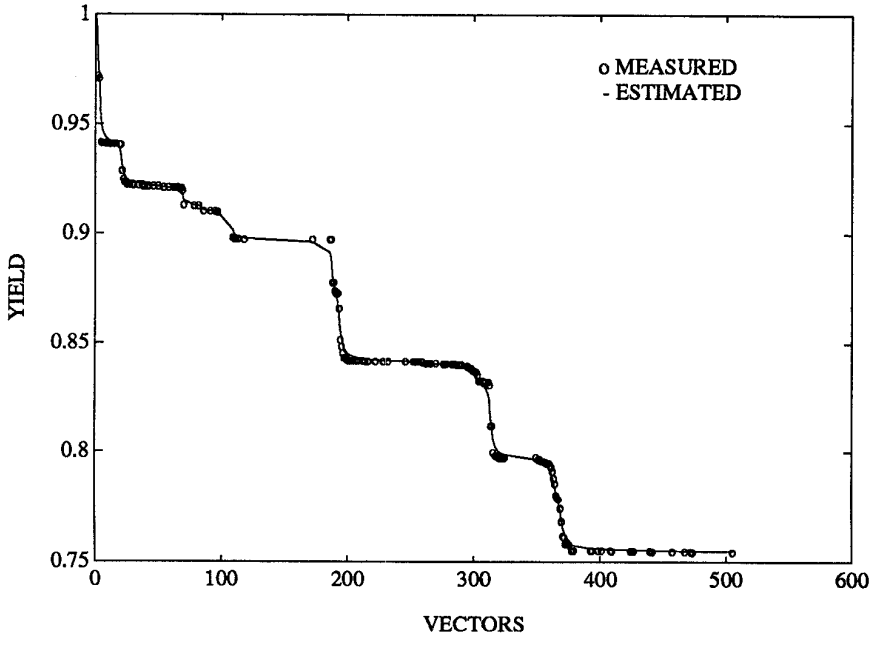

Figure 2: Efficacy of the model in fitting the jumps

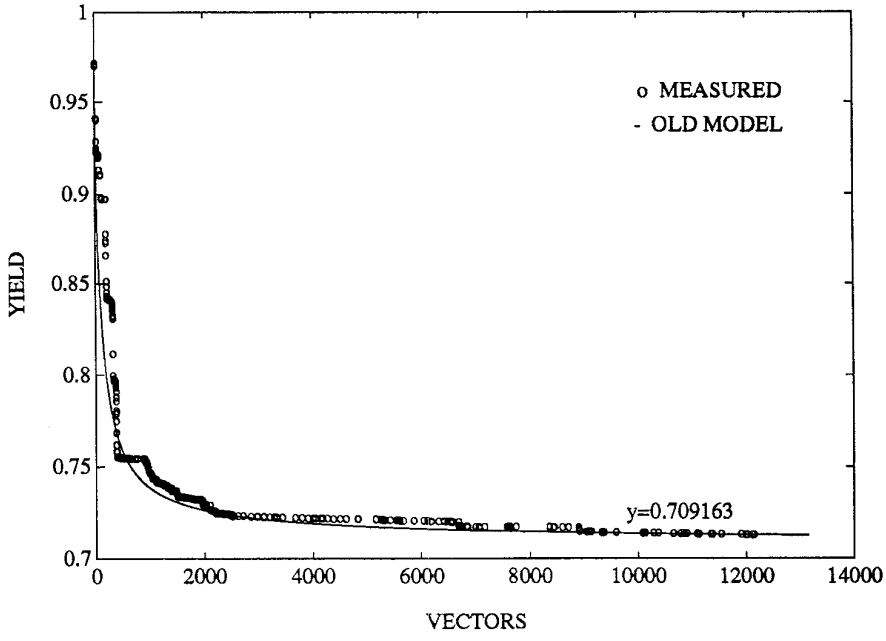

Figure 3: Estimation of true yield in the earlier model [6]

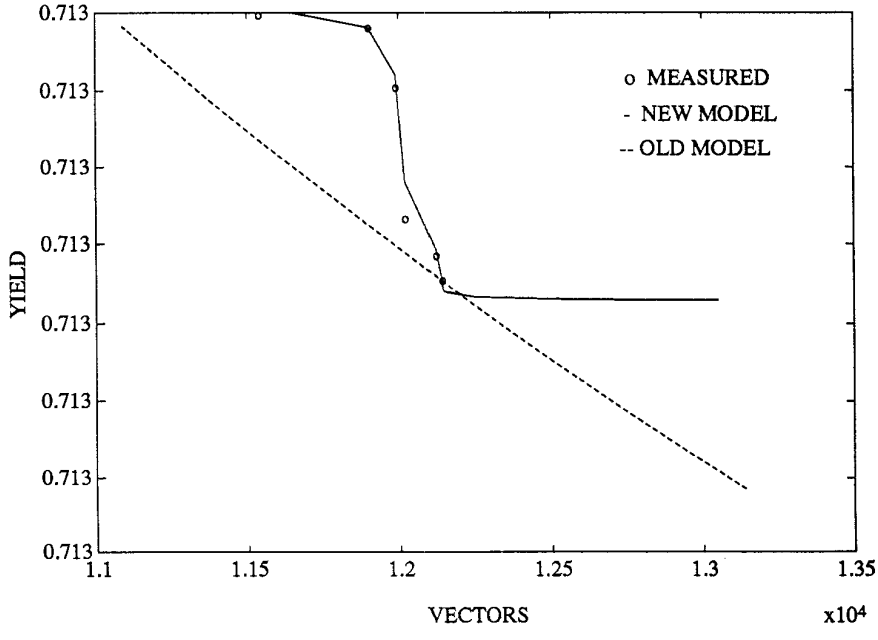

Figure 4: Extrapolated behavior of the two models

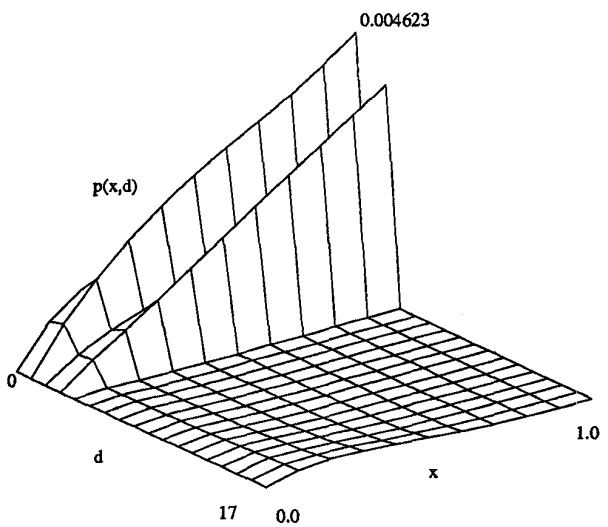

Figure 5: Density Distribution of Failed Chips

Bibliography

[1] D.V. Das, S.C. Seth, P.T. Wagner, J.C. Anderson, and V.D. Agrawal. An experimental study on reject ratio prediction for VLSI circuits: Kokomo revisited. In *Proc. International Test Conference*, pages 712–720, September 1990.

[2] R.L. Wadsack. Fault coverage in digital integrated circuits. *Bell Syst. Tech. J.*, 57:1475–1488, May-June 1978.

[3] T.W. Williams and N.C. Brown. Defect level as a function of fault coverage. *IEEE Trans. Comput.*, C-30:987–988, December 1981.

[4] V.D. Agrawal, S.C. Seth, and P. Agrawal. Fault coverage requirements in production testing of LSI circuits. *IEEE J. Sol. St. Circ.*, SC-17:57–61, February 1982.

[5] S.C. Seth and V.D. Agrawal. Characterizing the LSI yield equation from wafer test data. *IEEE Trans. CAD*, CAD-3:123–126, April 1984.

[6] S.C. Seth and V.D. Agrawal. On the probability of fault occurrence. In *Defect and Fault Tolerance in VLSI Systems, I. Koren (Editor)*, pages 47–52. Plenum, 1989.

[7] S.C. Seth, V.D. Agrawal, and H. Farhat. A statistical theory of digital circuit testability. *IEEE Trans. Comput.*, 39:582–586, April 1990.

[8] A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the theory of statistics.* McGraw-Hill, third edition, 1974.