

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:

The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

A PATTERN RECOGNITION BASED METHOD
FOR IC FAILURE ANALYSIS

by

A.J. Strojwas²x 3/./. Director

December, 1932

DRC-13-30-32

A PATTERN RECOGNITION BASED METHOD FOR IC FAILURE ANALYSIS

A. J. STROJWAS AND S. W. DIRECTOR

CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PA 15213

ABSTRACT

A new methodology for IC failure analysis has been developed. The approach is based on statistical pattern recognition concepts and is especially useful for the detection of parametric faults in monolithic ICs. This paper presents a set of algorithms which have been successfully applied to the analysis of the IC failure modes typical of those found in a commercial fabrication process.

1. INTRODUCTION

It is well known that the random fluctuations which are inherent in the IC manufacturing process cause yields to be significantly less than 100%. Therefore, an acceptable yield, Y_c , is usually specified for each different type of IC manufactured in a given fabrication process. The design process is said to be complete when the yield value from the test lots exceeds Y_c . If the value of yield in a given production lot exceeds Y_c , we will refer to the fabrication process in which this lot has been manufactured as a normal process. If the yield value is less than Y_c , we will refer to the fabrication process as a faulty process and we will say that a failure has occurred. A failure is caused by a fault in the fabrication process, i.e., an excessive variation in one or more fabrication step conditions. We will refer to the identification of a fault as failure analysis or fault detection.

In general faults can be classified as either catastrophic or parametric. Catastrophic faults are random defects which cause a hard failure of an element in a chip and, consequently, the whole chip fails. We say that a hard failure has occurred when the values of circuit performances are several orders of magnitude different from the tolerance limits. In particular, lithographic errors or oxide pinholes often cause undesired shorts or opens which result in a nonfunctioning IC. By parametric faults we mean excessive statistical variations in process conditions which cause a soft failure of an element in an IC chip. A soft failure is one which is not sufficient enough to make the IC inoperable, but sufficient enough to cause performance to deviate outside some allowable limits.

While catastrophic faults may be the main factor which limits yield in VLSI circuits, such faults are unavoidable. Thus, the only way to improve yield is to estimate the density of various defects and then use these estimates to establish a set of design rules. Parametric faults, which can

be caused by poor quality in processing equipment or materials, may affect many wafers or lots. By early detection and elimination of such faults, the yield in subsequent lots can be increased, thereby decreasing overall production cost. In this paper we describe a methodology for failure analysis; or more specifically, parametric fault identification.

Recently, there has been increased effort aimed at finding efficient fault diagnosis algorithms for integrated circuits. In the analog (linear) IC domain, algorithms for fault detection have been developed which determine faulty components or interconnections based upon node voltage measurements. For digital ICs, most of the effort has centered around the development of algorithms for automatic test generation. Such techniques are used to detect catastrophic "stuck-at" faults. These algorithms can be useful for hybrid ICs or Printed Circuit Boards (PCBs) which are composed of replaceable parts. However, with the exception of monolithic memories with built-in redundancy, these methods cannot be applied for failure analysis in monolithic ICs.

2. STOCHASTIC CHARACTER OF THE IC FABRICATION PROCESS

Fluctuations which occur in the steps of an IC manufacturing process have many causes, such as variations in the quality of the materials and processing equipment, or nonuniform conditions of the fabrication process for different IC chips or components within a single chip. These fluctuations cause variations in the electrical parameters of IC devices which, in turn, cause variations in IC performance. More precisely, device parameters and circuit performances are random variables and thus the IC manufacturing process has to be considered a stochastic process.

In order to model the random behavior of the IC fabrication process we observe that deviations in a particular process parameter usually affect some circuit performances more strongly than others. Moreover, deviations in several different process parameters may cause the same pattern of deviations in the performances. Each subset of the process parameters which affects the performances in the same way can be represented by a single random variable, called a process disturbance [1]. The entire fabrication process can be characterized by the n -dimensional random variable $O = (D_1, D_2, \dots, D_n)$ whose components are continuous independent random variables which represent the process disturbances. To describe local (U , within a chip) and

* This work was supported in part by the CMU/IBM Research Foundation grant. Order No. FC8203-U. © Hamam Research Corp.

** In what follows, if a capital letter represents a random variable, the corresponding lower case letter represents a particular value, or realization from a corresponding probability distribution.

global (ie. within a wafer) fluctuations inherent in the IC manufacturing process, we view D as having a multilevel structure [1]. Furthermore, it is reasonable to assume that the cumulative distribution of the components of D on a wafer is unimodal [2].

Now consider the effect of process disturbances on the joint probability distribution function (jpdf) associated with the parameters of an IC device and with the circuit performances. The electrical behavior of the circuit under consideration is characterized by a set of algebraic differential equations denoted by

$$M \dot{x} + x = D, \quad t_1 \leq t \leq t_2 \quad (D)$$

where I - is a vector of network variables, such as node and branch voltages, and branch currents; x - is a vector of parameters associated with the circuit element models (eg. MOSFET threshold voltage); t - is time; t_1 and t_2 - are the lower and upper bounds of the time interval of interest

Assume that circuit performances, Le[^] delay, power, etc. are represented by the n^{\wedge} -vector Z . The components of Z are typically expressed mathematically as

$$z_i = \int_{t_1}^{t_2} f_i(x, D) dt \quad (2)$$

where f is an appropriate function.

For nontrivial cases, the solution to (1) and (2) cannot be obtained analytically and we have to employ circuit simulation methods.

The relationship between the parameters of the device models X and parameters representing process disturbances D , is now in general explicitly known. Values for X may be found by solving a set of partial differential equations of the form [3]

$$M \nabla^2 p + p = d = 0 \quad (3)$$

where p - denotes the n -dimensional vector whose components represent the deterministic process controlling parameters, such as energy and dose of impurities in ion implantation; s - denotes the n^{\wedge} -vector of the deterministic circuit layout parameters, Le[^] the dimensions of the photolithographic masks of the devices and interconnections.

Usually, an analytical solution of (3) for z does not exist but must be computed numerically using fabrication process and semiconductor device simulation techniques.

In summary, the mappings $D \rightarrow X$ and $X \rightarrow Z$ are not explicitly known. However, as discussed in [2] these mapping functions are usually non-invertible over a wide range of realizations of D , and therefore the resulting jpdfs of the components of X and Z are non-unimodal. Generally, these jpdfs are not symmetrical due to the strong nonlinearity of the mapping functions. Clearly, the components of X (as well as the components of Z) are correlated since two different components are affected by the same process controlling parameters and a common subset of disturbances.

3. FORMULATION OF THE IC FAILURE ANALYSIS PROBLEM

The failure analysis method proposed in this paper employs the results of so called selective probe measurements which are performed at the end of wafer processing steps. Specifically circuit performances are measured in a sequence of static and dynamic tests on automatic testers controlled by computers. Other measurements which are performed during the IC fabrication process, Le. in-line and test pattern

measurements, are not satisfactory because of the small sample size.

We now make the following assumptions:

1. The design of both the process and the circuit is fixed. Le. the process controlling parameters and layout parameters are fixed.

2. The fabrication process is designed to produce a certain IC in large volume (thus the development of a failure analysis system is economically justifiable).

3. The probability distributions of D and Z associated with a normal process have been determined.

4. The yield value in a faulty process is greater than zero.

The IC failure analysis problem can be stated as:

If the yield in a certain production lot is less than Y , identify those components of the random variable D which are responsible for the decrease in yield, based upon an analysis of the joint probability density function characterizing the probability distribution of the circuit performances, Z .

It is evident from the above discussion that the relation between Z and D is not explicit and thus it is impossible to obtain a direct inverse mapping of Z into D . Therefore, analytic methods for fault detection cannot be applied.

The IC failure diagnosis problem, as formulated, can be viewed as a statistical pattern recognition problem. To see this assume that there exists a characteristic multivariate probability distribution of circuit performances corresponding to each fault (or combination of faults). This is equivalent to assuming that each class, which represents a particular fault type, has a specific "pattern" associated with it. This pattern is described in terms of the moments of the jpdf of Z . In the ideal case, patterns corresponding to different faults should be separable. In practice, however, it may happen that patterns described in terms of moments of the jpdf of Z corresponding to different faults are indistinguishable. Thus, the statistical pattern recognition problem can be decomposed into two subproblems:

1. Determining the most efficient representation of patterns for pattern classification purposes.

2. Development of a set of criteria which are used for pattern classification.

4. DEVELOPMENT OF THE IC FAILURE ANALYSIS SYSTEM

The following notation will facilitate our discussion. The n -dimensional vector of circuit performances, z , will be called a pattern vector. Each production lot which is subjected to failure analysis will be represented by an $N \times n$ pattern matrix B , composed of a sample of N pattern vectors $z_i, i=1 \rightarrow N$. The original pattern space is the n -dimensional Euclidean space containing the pattern vectors. A pattern class is a category determined by some given attributes of the pattern matrices. By a pattern classifier we mean a decision procedure which assigns a given pattern matrix to one of the available pattern classes (eg. Bayes classifier minimizes the total expected risk of misclassification).

For the problem under consideration, the description of pattern classes is not given a priori, and thus trainable

pattern classifiers have be be used. Representative pattern matrices corresponding to different faults are obtained from fault simulation experiments in which as artificially introduced fault (or combination of faults) causes a certain probability distribution of circuit performances.

The fault simulation experiments can be performed in three ways

1. experiments on a real fabrication process:
 - * expcimesB using a compimãöös of prãm asö event simulators?
- 1 establishing a nonlinear regression model relating **the components of Z to the components of D** and osig this model for generating patters matrices Δ STEtp- Δ ing to different pdf's of the **components of a**

Clearly, the latter technique for the fault simulation is the most efficient because the mapping of D into Z is approximated by as analytic function. As efficient algorithm for determining the structure of this regression model has bees developed and reported is [2]. This model is based upon data samples obtained by counting the statistical process simulator FABRICS [13 with a circuit simulator (e.g. SPICE).

The output from the fault simulation experiments consists of the samples of pattern vectors which are composed of circuit performances. Since the dimension of the vectors is high and the elements are sot statistically independent such a representation is far from optimal from the patters recognition viewpoint. Hence, this data has to be puiptiocitfrt to extract the relevant features. Two requirements are imposed on the feature extraction methods which are to be used is the pattern recognition system:

- L The features selected provide a better separability between pattern classes
2. The number of features selected is significantly kss than the dimensionality of the original pattern vectors.

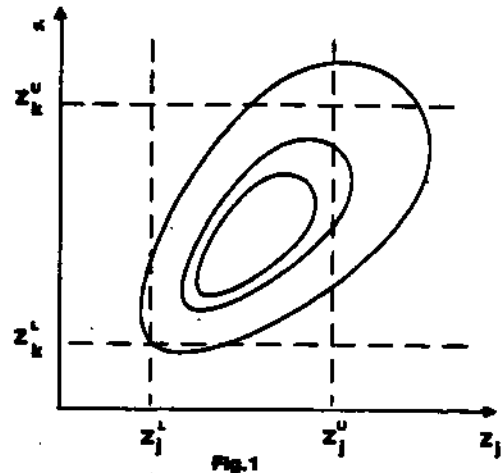
The most widely used techniques for dimensionality reduction are based upon transforming the original paucrn vectors into a lower-dimensional space (techniques for this include the principal component method by Hotelling [4] and the Karhunen-Loève expansion [4]). However, since the input to these methods consists of the covariance matrices of Z, these techniques are susceptible to local fluctuations which effect correlation coefficients among the components of Z. Therefore, a unique physical interpretation of the principal components becomes difficult (2). It should also be mentioned that these transformations, which involve covariance matrix factorization, are computationally expensive.

Note that the purpose of fault diagnosis is to find those factors responsible for a major chaste in the moments of the pdfs which characterize the probability distribution of the circuit performances (shifted mean value or significantly increased variance). Recall that we have postulated in Section 2 unimodality of the jpdf of Z, if we restrict consideration to the soft failures. Consider sow the level set $L_r(U)$ of the jpdf $f(z)$, which is defined m

$$L_r(U) = \{z | f(z) \geq r\} \quad (4)$$

To exclude the tail of $f(Z)$, we can assume that * is greater than some threshold $Wq. \ll \cdot .05$) and is tins case the level sets $L^A(U)$ win be connected and usually convex

[2]. The projection of the level sets $L_r(U)$ onto the plane z_j is r it shown m Fig. 1. The dashed haes represent the tolerance limits of circuit performances. Observe that analysis of the shape of $L_r(U)$ and its location with respect to the region of acceptability can provide useful information from the pattern recognition viewpoint. However, estimation of the level set contours is computationally expensive, so we employ a recognition scheme based upon scatterplots instead. Moreover, we can estimate the jpdf of Z by regkmtlization



of the Z space and construction of a multidimensional histogram. Is the two-dimensional case shown in Fig. 1, the space of circuit performances is divided into sine regions. This is equivalent to discretization of the components of Z which can be formulated as follows. Let the pattern matrix 1 contain a sample of N pattern vectors $r_i, i=1, 2, \dots, N$ and discretize the element* of B as follows:

$$b_{ij} = \begin{cases} 1 & \text{if } z_{it} > z_j^l \\ -1 & \text{if } r_{ij} < \wedge \\ 0 & \text{otherwise} \end{cases} \quad \begin{matrix} W- Z \dots N \\ j-1, 2, \dots, s_j \end{matrix} \quad (5)$$

where z_j is the j-th performance in the i-th chip.

Note that we can neglect for failure analysis purposes, the pattern vectors coresponding to good chips. (In other words we eliminate from 1 the rows containing only zero elements.) We can also restrict the analysis of UK pattern matrices to critical performances, which are defined as the performances which exceed their allowable limits most often in a gives lot. This is equivalent to deleting from the pattern matrix those columns is which the number of nonzero elements is less than some threshold value.

Now, recall that the objective of this pattern recognition system is to classify pattern matrices. Hence, further data reduction procedures are necessary to extract the relevant information from a pattern matrix. One pos«bk approach is io investigate similarity among pattern vectors and choose the swat typical osea to rptiaent the pattern matrix under consideration. Note precisely, the objective is to group the pattern vectors into subsets according to some similarity (or dHrinrlifiiy) measure, ie». perform a clusterint transformatiog. A distance function id., \wedge) is a particular measure of dhsimilanty which is often used is clustering. Generally, the number of dusters is sot gives a priori and therefore clustering is as iterative process is which the patters vectors are assigned to a number of arbitrarily

chosen clusters according to a certain distance measure from the cluster centers. In each iteration the cluster centers are updated and a new classification is performed. The partition should minimize the sum of intracenter distances between the pattern vectors and cluster centers and maximize the sum of intercluster distances between the cluster centers. Since in the pattern recognition problem under consideration the components of the pattern vectors are discrete-valued, Euclidean distance is not a meaningful measure of dissimilarity and other measures have to be used (e.g. Tanimoto measure [5]). The following alternate clustering schemes have been developed:

1. A Shortest Spanning Tree Algorithm.
2. A Pattern-Directed Algorithm.
3. A K-Means Algorithm [5] in the S-information Space [6].

For a detailed description of these algorithms see [2]. The largest clusters detected constitute pattern features which are then used for pattern classification.

Observe that each pattern matrix can be now represented in a form of a histogram of the clusters detected. A simple classifier can be constructed which performs pattern classification based upon histogram comparison (e.g. by Chi-square or Kolmogorov-Smirnov test for goodness of fit [7]). However, since in this approach features are treated as independent, this method is not accurate enough.

In [2] we have proposed two methods for pattern recognition which are based on the pattern deformational model [8]. An assumption is made that the patterns which are to be recognized can be considered deformed versions of the pure patterns which are obtained from fault simulation experiments. In the first approach a symbol is assigned to each cluster. The collection of clusters which defines a pattern is therefore represented by a string of symbols, called pattern primitives. Each pattern class can be then represented as a string with the additional attributes specifying the probability of occurrence of each symbol and the probability of co-occurrence of a certain group of symbols. These attributes are called stochastic production rules and can be inferred from fault simulation experiments. Two types of pattern deformations can be defined: structural deformations and local deformations. Pattern classification is performed in two stages. First, a particular pattern is assigned into a group of classes based upon a distance between the strings representing each class (e.g. Levenshtein distance [9]). This ensures that the pattern under consideration is structurally similar to any of the pattern classes in this group. Then we determine to which particular class the pattern under consideration is closest to using a Bayes distance [8] which is a measure of local syntactic and semantic deformations of the pattern primitives.

In the second method pattern matrices are represented by attributed relational graphs. The nodes denote pattern primitives, as described above, while the edges represent relations between pattern primitives (e.g. similarity measure between clusters). Pattern recognition is performed by transforming a pattern matrix into a structural representation using a relational graph and then matching this graph (i.e. searching for error-correcting graph isomorphisms [10]) with those which represent pure patterns. The above techniques for pattern classification are described in [2].

5. CONCLUSIONS

In this paper we presented a theoretical basis for an IC failure analysis system based upon statistical pattern recognition methods. The performance of this system has been verified on several examples of IC's which have been manufactured in different fabrication processes. In particular, we have considered such IC's as a bipolar operation amplifier, a CMOS analog switch and a NMOS DRAM. Several pattern classes corresponding to major failure modes have been created based upon data collected from commercial fabrication processes and from computer simulation. The pattern recognition algorithms presented above have proved to be efficient in the recognition of the reasons for yield drop in several production lots. However, lack of space does not allow us to give a detailed description of these results. This data can be found in [2] and will be summarized in the conference presentation.

REFERENCES

- [1] Maly, W. and A.J. Strojwas. Statistical Simulation of the IC Manufacturing Process. *IEEE Transactions on CAD*, July, 1982.
- [2] Strojwas, A.J. *A Pattern Recognition Based System for IC Failure Analysis*. PhD thesis, Carnegie-Mellon University, September, 1982.
- [3] Antoniadis, D.A. and R.W. Dutton. Models for Computer Simulation of Complete IC Fabrication Process. *IEEE J. Solid State Circuits* SC-14:412-422, April, 1979.
- [4] Fu, K.S. and A.B. Winston (editors). *Pattern Recognition Theory and Application*. Noordhoff-Leyden, 1977.
- [5] Tou, J.T., and Gonzalez, Rafael C. *Pattern Recognition Principles*. Addison-Wesley Publishing Company, 1974.
- [6] Wong, A.K.C. and Liu, T.S. Typicality, Diversity and Feature Pattern of an Ensemble. *IEEE Transactions on Computers* C-24(2):158-181, February, 1975.
- [7] Allen, A.O. *Probability, Statistics and Queueing Theory*. Academic Press, 1978.
- [8] Tsai, W.H. and K.S. Fu. Pattern Deformational Model and Bayes Error-Corresponding Recognition System. *IEEE Transactions on Systems, Man and Cybernetics* SMC-9(12), December, 1979.
- [9] Fu, K.S. *Syntactic Pattern Recognition and Applications*. Prentice-Hall, Inc., 1982.
- [10] Tsai, W.H. and K.S. Fu. Error-Correcting Isomorphisms of Attributed Relational Graphs for Pattern Analysis. *IEEE Transactions on Systems, Man and Cybernetics* SMC-9(12), December, 1979.