Exploiting Multiple Mahalanobis Distance Metric to Screen Outliers from Analogue Product Manufacturing Test Responses

Shaji Krishnan∗ and Hans G. Kerkhoff†

∗Analytical Research Department, TNO, Zeist, The Netherlands
Email: shaji.krishnan@tno.nl
†Testable Design & Testing of Integrated Systems Group
University of Twente, CTIT
Enschede, The Netherlands
Email: h.g.kerkhoff@utwente.nl

Abstract—One of the commonly used multivariate metrics for classifying defective devices from non-defective ones is Mahalanobis distance. This metric faces two major application problems: the absence of a robust mean and covariance matrix of the test measurements. Since the sensitivity of the mean and the covariance matrix is high in the presence of outlying test measurements, the Mahalanobis distance becomes an unreliable metric for classification. Multiple Mahalanobis distances are calculated from selected sets of test-response measurements to circumvent this problem. The resulting multiple Mahalanobis distances are then suitably formulated to derive a metric that has less overlap among defective and non-defective devices and which is robust to measurement shifts. This paper proposes such a formulation to both qualitatively screen product outliers and quantitatively measure the reliability of the non-defective ones. The resulting formulation is called Principal Component Analysis Mahalanobis Distance Multivariate Reliability Classifier (PCA-MD-MRC) Model. The application of the model is exemplified by an industrial automobile product.

Index Terms—Test, Reliability, Analogue, Outliers

1 INTRODUCTION

Stringent quality requirements on final electronic products are continuously forcing semiconductor industries, especially the automobile industry, to insert additional reliability tests in their production flow. For this purpose, they subject their packaged devices to time consuming burn-in procedures and purchase expensive automatic test equipment. Essentially, the semiconductor companies are constantly searching for cheaper product manufacturing and reliability test methodologies that reduce the overall production costs, while leaving no compromise to the quality of their products.

One of the most commonly applied solutions to this problem is the early identification and elimination of latent devices in the production flow. Several statistical methods such as Parts Average Testing (PAT) help to identify latent devices during wafer-level testing. PAT was introduced by the Automobile Electronic Council to identify abnormal parts from a population [1]. Later on, methodologies such as die-level predictive models were introduced. They are capable of predicting the reliability of the device based on the failure rate of the neighboring dies [2]. Some other techniques that have been proposed to identify unreliable devices at wafer sort are based on test-response measurements from parametric tests like IDDQ/Delta IDDQ [3], [4], supply ramp [5] or functional tests [6]. Test-response measurements derived from such wafer-level tests are statistically post-processed to screen outliers.

Outliers are identified by analyzing data either in a univariate or multivariate space. In some situations, an inlier in the univariate space can be an outlier in the multivariate space. Multivariate unsupervised outlier identification methods like linear regressions models capitalize on known relationships among variables to calculate the error in the prediction of the dependent variable [3], [4], [6]. The distance is then the deviation of error from the population mean. The problem with linear regressions models for outlier identification in analogue and RF devices is that they do not generally account for manufacturing variability and test measurement shifts. Test measurements have less predictability and lead to less certainty in the population mean of the measurements. This affects the distance metric so that marginal outliers have a higher chance of being undetected.

To circumvent the problems associated with the calculation of a robust distance metric, the Mahalanobis distance (MD) [7] is employed as distance parameter in unsupervised outlier identification methods. Maha-
lanobis distance is invariant to measurement shifts and, most importantly, accounts for the relationships among data variables (covariance). Robustly estimated population mean and covariance are used in identifying outliers because the Mahalanobis distance metric itself is sensitive to outlier data [8]. To further reduce the variance in the distribution of the Mahalanobis distance metric, this paper proposes a novel way of using multiple sets of the Mahalanobis distance metric for reliability analysis of analogue electronic products. One of the main advantages of the proposed technique is the option to include sets of correlating and uncorrelating test variables into a single structure. Furthermore, the technique allows for both qualitative and quantitative reliability analyses of functionally qualified products.

2 MULTIVARIATE RELIABILITY CLASSIFIER (MRC) MODEL

The reliability classifier model for reliability analysis is formulated in the following section. The resulting formulation is of multivariate nature, hence the Multivariate Reliability Classifier (MRC) Model. The associated analysis is called multivariate analysis and comprises a set of data analysis techniques tailored to data sets with more than one variable. This means that the responses from several tests of electronic circuits can be combined in a multivariate space to describe the different response behaviors of the system. Qualitative and quantitative analysis of this multivariate response behavior, as formulated by the MRC Model, ultimately yield a measure of the reliability of the product. The following paragraphs briefly describe the steps for the formulation of the MRC Model. Detailed descriptions of the method can be found in a previous paper [9].

The multivariate reliability classifier function as a model for reliability analysis is formulated from multiple test sets. All tests within a test set are uncorrelated while the test sets are correlated with each other. Each test set contains a minimal number of tests corresponding only to the most significant tests. These tests are significant because they are capable to influence the results obtained with the reliability classifier model. The significant tests are found using Principal Component Analysis (PCA) variable reduction procedure and Scree plot [10]. After removal of all significant tests found in previous iterations, the initial set of tests, the iterative application of this procedure leads to a new test set. Once a suitable number of test sets, \( k \), is determined, the Mahalanobis distance metric for each sample, \( i \), is computed over each test-set, \( m = 1 : k \). The resulting Mahalanobis distance metrics, \( MD_{i,k} \), are then linearly regressed as shown in Equation 1 to form the PCA-MD-MRC Model equation, whereby the sample error and the offset are denoted by \( \epsilon_i \) and \( a_0 \) respectively. This model then uses the distribution of error to reduce the variance in measurement and qualitatively classify the reliability of the product.

\[
\text{MD}_{1,i} = a_0 + \sum_{m=1, m \neq 1}^{k} a_m \text{MD}_{i,m} + \epsilon_i \tag{1}
\]

Rewriting the PCA-MD-MRC Model equation from Equation 1 to 2 converts the error \( \epsilon_i \), from normal distribution to a chi-square (\( \chi^2 \)) distribution with one degree of freedom [11].

\[
\epsilon_i^2 = [\text{MD}_{1,i} - (a_0 + \sum_{m=1, m \neq 1}^{k} a_m \text{MD}_{i,m})]^2 \tag{2}
\]

The deviation of the squared error distribution from a Chi-square (\( \chi^2 \)) distribution is quantified in order to assign probabilistic values that reflect the reliability of the qualified product within the scope of the available test-response measurements.

3 EXPERIMENT AND RESULTS

The product is a single IC implementing a car radio tuner for AM and FM intended for microcontroller tuning with the I2C-bus. The number of functional tests at wafer level for this product is 175 and corresponds to several levels of DC tests, modes of AC and RF tests. Products that pass these functional tests are referred to as qualified products at wafer level. The samples chosen for demonstrating the PCA-MD-MRC are all qualified car radio tuner product instances chosen from a single wafer from the manufacturing lot. The number of qualified products, i.e. known good dies (KGDs) is 379, while the number of known defective dies (KDDs) is 99.

Functional test data from 175 tests and 379 qualified products were iteratively subjected to the PCA variable reduction procedure. During each iteration, a set of significant tests is being identified as belonging to one significant test set (T) based on the Scree plot. Nine significant tests were selected for the first two significant test sets (T1 and T2). The significant test set selection was terminated after the first two iterations (T1 and T2) due to lack of sufficient correlation (less than 0.7) of the third iteration (T3) with the selected significant tests of the two previous iterations (T1 and T2). Figure 1 shows the Scree plot of the test data for the two iterations (T1 and T2).

For each significant test set (T1 and T2), a Mahalanobis distance metric ((MD1 and MD2) is calculated for each of the 379 products. The test response data that corresponds to the tests within the significant test set is used for this purpose. The resulting Mahalanobis distance metrics for all products are collinear, i.e. each Mahalanobis distance metric can be expressed as function of another one. Hence a PCA-MD-MRC Model as shown in Equation 3 can be implemented whereby MD1 is the independent variable and MD2 is the dependent variable. The error in the PCA-MD-MRC Model is further analyzed to qualitatively and quantitatively classify the reliability of the qualified product.

\[
\text{MD}_2 = 0.97 \times \text{MD}_1 + 0.04267 \tag{3}
\]
3.1 Qualitative reliability analysis results

The goal of qualitative reliability analysis is to determine statistical outliers to the model e.g. by iteratively constructing the model and estimating the goodness of fit after every iteration. During each iteration, a set of data that does not seem to fit the model, i.e. the outliers, is removed and the model parameters are re-estimated until there is no further improvement in the fit is observed.

Following the criterion for the MRC Model described in Equation 3, the empirical distribution errors are used to qualitatively determine the outliers in the dataset. The error distribution of the fit will be close to a normal distribution since the linear regression fit follows a least square fit. The error distribution for all qualified products and four outliers are shown in Figure 2. The statistical outliers are thereby qualitatively classified as products that are potentially unreliable.

3.2 Quantitative reliability analysis results

Figure 3 shows the density function of squared errors derived from the MRC Model for all qualified products after having removed the outliers identified in the previous qualitative reliability analysis. A corresponding $\chi^2$ density function with one degree of freedom is overlayed to demonstrate the deviation from the density function of squared errors.

A $\chi^2$ test is conducted to determine the goodness of fit of the squared error with respect to the $\chi^2$ distribution [11]. To facilitate the $\chi^2$ test, the distributions are evenly sub-divided into classes. For each class between the second and fourth quartile of the $\chi^2$ distribution, the number of observed ($d$) and the number of expected ($e$) samples are determined from the squared error distribution and the $\chi^2$ distribution, respectively. The $\chi^2$ statistic is then computed for each of these classes from the observed and the expected samples (i.e., $(d - e)^2/e$). Two neighboring classes along with the current class are subjected to the $\chi^2$ test in order to accommodate the variations of adjacent classes. In other words, to determine the goodness of fit of a class from the squared error distribution, the $\chi^2$ statistic for that class includes two of its neighboring classes. Exceptions to this principle are the first and the last class, whereby only one neighboring class is included. Hence, the degree of freedom for all but the first and last class is three. For the first and the last class it is two. Whenever the expected class observation is less than two, the $\chi^2$ statistic is not computed since it will not provide any significant results. The reason for this is that an expected value of one observation is usually at the tail of the $\chi^2$ distribution. Any major deviation of the squared error distribution will be identified and removed by the qualitative reliability analysis procedure.

Table 1 shows the quantitative reliability analysis results for all qualified products other than the outliers identified in the qualitative analysis. For each class with a mid-value $Cls.,$ beginning from the second quartile,
In order to validate the invariance of the MRC Model to measurement shifts, the test data of all the qualified products were shifted (increased by 10% to its measured value) and subjected to the reliability classifier for qualitative analysis. Figure 4a shows the overlapped histogram of the measured (T13) and shifted test data

the observed (Obs.) and the expected (Exp.) number of observations are shown. The corresponding $(d - c)^2/c$ values and the $\chi^2_{df=3}$ statistic are calculated. The probability value ($p$ value), is determined for $\chi^2_{df=3}$ with according degrees of freedom $df$; two for the first and the last class and three for the remaining classes from the $\chi^2$ statistical table [12]. A probability value greater than 0.05 indicates that the deviation of the observed values from the expected is sufficiently small so that chance alone accounts for it, while a probability value less than or equal to 0.05 means that some factor other than chance is responsible for the deviation. Hence, a $p$ value equal to 0.05 for the qualified products belonging to the class interval with a mid-value 0.19 indicates that there is only 5% chance that those products belong to the $\chi^2$ distribution and are therefore less reliable. Since the $p$ values for the remaining classes are higher than 0.05, it can be concluded that the deviation is due to chance. Therefore, it can be probablistically considered reliable within the realm of the functional test conducted on the product.

### 3.3 Invariance to response measurement shifts

One of the major advantages of the MRC Model is its invariance to response measurement shifts because of the Mahalanobis distance is invariant to scaling [7]. For further reduction of the naturally occurring variance in test measurements, our model utilizes the error distribution of multiple collinear Mahalanobis distances. The MRC Model is therefore robust to both measurement variations and shifts.

In order to validate the invariance of the MRC Model to measurement shifts, the test data of all the qualified products were shifted (increased by 10% to its measured value) and subjected to the reliability classifier for qualitative analysis. Figure 4a shows the overlapped histogram of the measured (T13) and shifted test data

(T13*) in relation to the test variable (T13) from all qualified products. Figure 4b is the regression plot of the error distribution Error and Error* of the MRC Model constructed before and after shift of the test-response data. The regression line lying at 45° and passing through the origin indicates that the error distributions are equal to one another and asserts the invariance of the MRC Model to measurement shifts.

#### 3.4 Comparative results to other outlier detection methods

Table 2 shows the comparative results of the PCA-MD based MRC (PCA-MD-MRC) model to other multivariate outlier detection models like the Principal Components (PC), the Linear Regression (LR), and a univariate method like PAT. The labels denote the outliers identified by each of the methods. Although other outlier methods (LR and PC) produced comparable results, they did not identify the outliers exclusively like our PCA-MD-MRC model.

<table>
<thead>
<tr>
<th>Cls.</th>
<th>Obs.(d)</th>
<th>Exp.(c)</th>
<th>$(d - c)^2/c$</th>
<th>$\chi^2_{df=3}$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.11</td>
<td>10</td>
<td>14</td>
<td>1.14</td>
<td>1.24</td>
<td>0.35</td>
</tr>
<tr>
<td>0.13</td>
<td>9</td>
<td>10</td>
<td>0.10</td>
<td>1.74</td>
<td>0.4</td>
</tr>
<tr>
<td>0.15</td>
<td>6</td>
<td>8</td>
<td>0.50</td>
<td>4.76</td>
<td>0.1</td>
</tr>
<tr>
<td>0.17</td>
<td>1</td>
<td>6</td>
<td>4.16</td>
<td>4.66</td>
<td>0.1</td>
</tr>
<tr>
<td>0.19</td>
<td>4</td>
<td>4</td>
<td>0.00</td>
<td>5.16</td>
<td>0.05</td>
</tr>
<tr>
<td>0.21</td>
<td>6</td>
<td>4</td>
<td>1.00</td>
<td>2.33</td>
<td>0.3</td>
</tr>
<tr>
<td>0.23</td>
<td>5</td>
<td>3</td>
<td>1.33</td>
<td>4.33</td>
<td>0.1</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>2</td>
<td>2.00</td>
<td>3.83</td>
<td>0.25</td>
</tr>
<tr>
<td>0.27</td>
<td>3</td>
<td>2</td>
<td>0.50</td>
<td>3.50</td>
<td>0.20</td>
</tr>
<tr>
<td>0.29</td>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.31</td>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.33</td>
<td>1</td>
<td>1</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.35</td>
<td>0</td>
<td>1</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.37</td>
<td>1</td>
<td>0</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.39</td>
<td>1</td>
<td>1</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.41</td>
<td>2</td>
<td>0</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.43</td>
<td>0</td>
<td>0</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### 4 OTHER APPLICATIONS OF THE MULTIVARIATE RELIABILITY CLASSIFIER

This study indicates that the MRC that demands no information other than the existing univariate test response measurements seems to be a valuable tool for reliability analysis of functionally qualified electronic products. However, the scope of this MRC Model does not seem to be restricted to reliability screening purposes. Some other ways to use this model are described below and can be seen as suggestions for future research.

The sensitivity of a functional test to the outcome of the MRC is a suitable metric to re-condition the pass-fail boundary of a device. Determining those sets of tests...
that are sensitive to the results of the MRC is one of the problems that have to be addressed. Devising ways to determine the sensitivity of all tests and ordering the tests according to their sensitivity, are some of the implicit ways to grade the quality of functional tests. Subjecting a product to such an ordered set of tests, allows for classification of products according to varying quality levels.

Another application of the MRC Model may be the early detection of failing devices. The knowledge about the significant tests from the set of all functional tests is incurred during the application of the MRC model to a set of devices. Prioritizing these significant tests for the next set of devices may result in capturing the failing devices earlier in the test flow. The resulting economic benefits of capturing the failing devices are higher during immature stages of the production flow.

A robust and reliable in-line or post-processing statistical tool for reliability analysis is economically beneficial when compared to time-consuming industrial practices such as burn-in, and customer-return pareto analysis. The advantages and disadvantages for replacing these industrial practices by statistical test analysis have not been analyzed within the scope of this study. Future research is required to evaluate the proposed MRC Model from that point of view.

5 Conclusions
A MRC model that enables both qualitative and quantitative analysis regarding the reliability of analogue electronic products has been described and exemplified in this study. One of the major advantages of the model is that it allows the simultaneous use of correlating and uncorrelating test variables for the identification of outliers. A robust distance metric, the Mahalanobis Distance, has been added to the model for marginal outlier identification in order to avoid bias by commonly occurring shifts in test-response measurements. The MRC Model has been chosen in a way that the results of the qualitative analysis can be further analyzed quantitatively to measure the level of reliability of the product in a probabilistic sense. Other models such as the Linear Regression Model and the Principal Component Model do not identify the outliers exclusively like our MRC Model. Other applications of the reliability classifier such as the identification and conditioning of specific functional tests for product grading and early detection of failures remain to be investigated.

Acknowledgment
The authors would like to thank Anne Potzel for the language revision.

References