Machine-Learning Based Overkill Reduction using Correlation within the Probe Tests

Vineeth Amriur Niranjan*, Deepika Neethirajan*, Dallas Webster†, Amit Nahar† and Yiorgos Makris*
*Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX USA, 75080
†Texas Instruments Inc., 12500 TI Boulevard, MS 8741, Dallas, TX 75243

***Abstract*—As the quality expectations increase, contemporary semiconductor manufacturing test solutions inevitably result in a non-negligible number of good devices being discarded which is termed as overkill. Overkill is the result of multiple factors affecting the manufacturing test process. Some of these factors stem from the test environment itself. During the testing procedures, stricter test program limits are used compared with actual product specs, resulting in overly conservative decisions and a considerable number of good devices being discarded. Additionally, test programs often include measurements that are not part of the specifications but established as part of the quality control process, for which limits are empirically defined, resulting in non-negligible yield loss of specification-compliant devices. Towards addressing the problem of overkill and improving the overall effectiveness of the manufacturing test process, in this project we propose to exploit multivariate statistical analysis methods. We do this by post processing the probe test measurements of the two different groups of tests – tests that are part of the specification list (Customer Promised) and tests with guard band limits that are not part of the specification. We take advantage of the correlation that exists between all the probe test measurements, irrespective of whether the limits of these tests are stricter or lenient. Using the probe test measurements of the set of devices which pass all the tests, we train a regression model to predict the probe test measurements of the devices that fail the tests that are designed for internal quality purposes. The effectiveness of the proposed methodology is demonstrated on an industrial dataset provided by Texas Instruments.**

***Index Terms*—post-silicon calibration, adaptive, test-cost reduction**

I. Introduction

As the complexity of devices increase with current new technologies, the testing procedures in semiconductor manufacturing are also complex and time consuming. These testing procedures are always comprehensive in order to combat the increasing process variations and exhaustive in order to satisfy the strict quality expectations in the market. The market expectations are higher than ever when it comes to defect tolerance in the range of Defective Parts Per Billion (DPPB) instead of Defective Parts Per Million (DPPM). The elaborate, expensive and stricter testing procedures aim to remove any of the defective chips from being shipped. During such testing procedures, stricter test program limits (guard bands) than the actual product specs are employed during test, resulting in overly conservative decisions and a considerable number of good devices being discarded. This results in yield loss (i.e., overkill). Additionally, test programs often include test measurements that are either listed as "typical" or are not part of the specifications. For these tests the limits are defined at the onset of production with limited set of device population and never revisited again to be changed. Because of these empirically defined limits, there is a non-negligible number of specification-compliant devices being removed as defective devices.

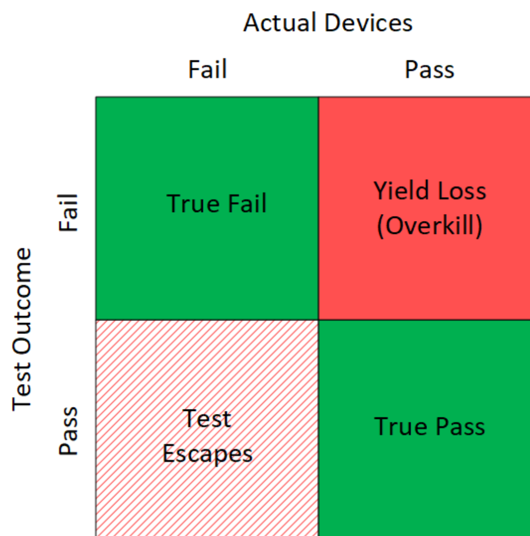

Fig. 1: Test Outcome Matrix

Figure 1 shows the overall outcome of the above mentioned testing procedure of fabricated ICs. The two green squares show the two masses of devices that comply with the design specifications and those that do not comply with the specification limits. These 2 groups will truly pass and fail the specification limits. However, as indicated by the solid red square at the top right, a number of chips will end up being classified as failing by the manufacturing test process because of the guard banded limits, resulting in overkill or yield loss. The striped red square in bottom left shows the marginally failing chips being identified as defective or underkill. The devices that are part of the solid red square comprising the yield loss factor implies the possible profit or money left on the table. Previous research shows multiple statistical solutions that take advantage of the relationship between the multiple test measurements of the product being used to rectify the situation at hand. Regardless of the developed solutions, there

is still existing yield loss and test escape problems that hamper the overall efficiency of the manufacturing industry.

		Tests with CP limits	
		Pass	Fail
Tests with NCP limits	Pass	Group 1	Group 2
	Fail	Group 3 (Our Focus)	Group 4

Fig. 2: Distribution Matrix for CP and NCP tests

Several components contribute towards the problem of overkill apart from the stricter test limits. Among those factors is the numerous test measurements that are designed specifically for internal process control. These test measurements are designed in such a way that passing them is not a requirement from the customer side and in most cases not even known to the customers. As mentioned earlier, in order to minimize customer returns, test limits are often set conservatively. Furthermore, the limit setting decisions are taken by considering only a limited set of devices and hence there is a chance of misinformed decision making. Nevertheless, chips failing these tests are still discarded as defective. Thus we have two sets of test measurements, one is the customer-promised tests (CP) that are part of the product specification and known to the customers; the other being the non-customer promised tests (NCP) that do not make the product specification and have empirically defined limits.

From the matrix shown in figure 2, we can see the test outcomes with respect to the two groups of tests. Group 1 devices are the ones that pass both CP and NCP tests and hence are considered good devices and can be shipped to customers. Group 2 devices are the ones that fail CP tests that are part of product specification and are discarded irrespective of passing the NCP tests. Group 3 comprises of devices that pass CP tests but fail one or more NCP tests and hence are discarded as failing devices. Group 4 devices fail; both CP and NCP tests and are discarded as defective devices. Our focus is on group 3 devices that pass the customer requested CP tests but fail one or more NCP tests. Since NCP tests have empirically defined limits at the onset of production, we strongly believe that there is some yield that is left to be recovered from this space. The goal of our work is to employ multivariate statistical models that takes full advantage of the available information from the testing procedure and the existing correlation between the different groups of tests to overcome the problem of yield loss.

II. Related Work

Several approaches that take advantage of the correlation between different test measurements have already been proposed with the aim of reducing the test costs and time associated with the complex testing procedures. Specifically, the above mentioned concepts are derived form *alternate test* [6].Prior work explored *alternate test* where the inter-test correlation is leveraged to skip exhaustive testing and reduce the overall testing costs as studied in [5] and [1].In [3] the authors explore the embedded correlation existing in the production test data and with the use of feature engineering, they are able to classify multidimensional spaces where the device is passing/failing to capture test escapes. The existing literature on yield recovery is based in post-calibration tuning of devices [4].

III. Proposed Methodology

Our methodology aims at predicting the NCP test measurements in place of the actual test measurements for group 3 devices. This is because we believe that the devices failing some of the NCP tests are mainly due to the empirically defined limits. Hence we propose to take the group 1 devices that pass both sets of tests and use the relationship between CP and NCP tests in Group 2 devices to predict the NCP test measurements of Group 3 devices. The model takes advantage of the statistical correlation that exists between the tests and leverages that relationship to predict the test outcomes and thereby reducing yield loss. One of the key components of building the model to predict the test outcome is the implementation of the Multivariate Adaptive Regression Splines (MARS) algorithm [2]. MARS is a powerful and flexible regression model that helps in representing relationships between a few variables in high-dimensional datasets. It takes advantage of additive and interactive relationships between variables, thereby resulting in using fewer variables to represent a high-dimensional dataset.

As shown in figure 3, in the machine learning based approach that we are proposing, there are two phases. During the training phase, we only make use of the devices that we trust, essentially the devices that belong to group 1. From these devices, we extract the probe tests measurements for both CP tests and NCP tests and train a MARS model. The MARS model learns the underlying statistical correlation between the CP tests and NCP tests from their probe test measurements. Once the model has been trained, we move to the testing phase, where we use the devices that belong to group 3 which pass all CP tests but fail one or more NCP tests. From these devices, we extract the probe test measurements from CP tests and pass them through the previously trained MARS model. This model based on its previous learning will be able to predict the test outcome of the NCP tests by correlating the test measurements of CP tests. Once the model predicts the test outcome for the NCP tests of group 3 devices, we can determine if the outcome will pass or fail the specific test limits.

The prediction results despite being good cannot be substantiated or verified because we do not have any ground truth. In order to substantiate the proposed regression-based approach, we have come up with a clustering-based approach. This is a step carried out using the predicted results from the regression

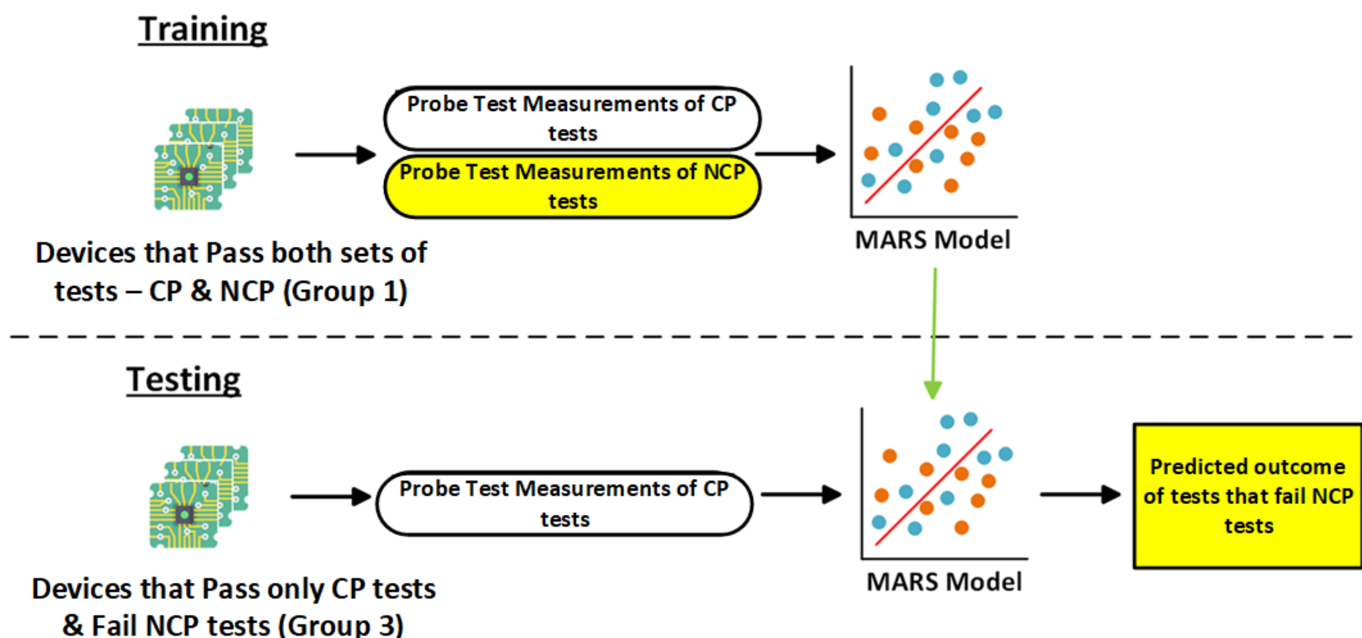

Fig. 3: Machine Learning Based Approach

approach. We take the predicted NCP test measurements of group 3 devices and cluster them using the unsupervised clustering algorithm like agglomerative hierarchical clustering to see if there is a distinct cluster containing the devices which have been predicted to pass and capable of being moved to group 1 area. Unlike the previous step's regression model, this unsupervised method is agnostic to limits and groups the devices only based on the probe test measurements' unique signature.

IV. Results

In order to implement the proposed methodology, we used an industrial dataset consisting of 92022 devices and their probe test measurements provided by our industry partners at Texas Instruments. The dataset consisted of probe test data along with the test limits for 66 customer promised tests and 241 non-customer promised tests. The overall test outcome among these tests for all devices under study is shown in figure 4. The 9.6% of the devices in the bottom left square are the set of devices which are to be studied to improve the yield and recover whatever is left on the table. The 87.4% of the devices are used in the training phase of the regression based approach. The regression model was trained to learnt he correlation between the CP and NCP test measurements of the group 1 devices. During the testing phase, the 9.6% devices and their CP test measurements were passed through the trained model and provides the predicted results for the group 3 NCP test measurements. It can be observed from the predicted regression results shown in figure 5 that we were able to recover 7953 devices out of 8840 of group 3 devices and move them from group 3 to group 1. This is because the 7953 devices were able to pass all the NCP tests base don our predicted test measurements. But as we mentioned earlier, there is no ground truth to verify if the 7953 devices can definitely be moved to group 1.

		Tests with CP limits	
		Pass	Fail
Tests with NCP limits	Pass	87.4% (Group 1)	0.78% (Group 2)
	Fail	9.6% (Group 3)	2.38% (Group 4)

Fig. 4: Distribution Matrix for the industrial dataset

In order to prove our results we take the predicted NCP test measurements and pass them through the unsupervised clustering algorithm. We employed agglomerative hierarchical clustering algorithm to cluster the predicted test measurements. The results of this clustering is shown in figure 6, where we observe 2 distinct yet slightly overlapping clusters. Based on the clustering algorithm output, we have 6446 devices in one cluster and 2394 devices in another cluster. With both sets of results, we can definitely see that there are 3 buckets of devices. The first bucket or the confirmed pass bucket consists of 6295 devices which includes devices that belong to one cluster and overlap with the regression model's recovered 7953 devices. The second bucket or the confirmed fail bucket

Actual

Tests with NCP limits	Tests with CP limits: Pass	Tests with CP limits: Fail
Pass	80261 (Group 1)	726 (Group 2)
Fail	8840 (Group 3)	2195 (Group 4)

Predicted

Tests with NCP limits	Tests with CP limits: Pass	Tests with CP limits: Fail
Pass	80261 + 7953 (Group 1)	726 (Group 2)
Fail	887 (Group 3)	2195 (Group 4)

Fig. 5: Results from Regression Model based Predictions

consists of 736 devices which is an overlap between the 887 devices that have not been recovered based on regression results and the second cluster from clustering results consisting of 2394 devices. The remaining 1809 devices fall into the third bucket or the gray area devices bucket for which we are unable to concretely prove if they can be recovered back to group 1 or not.

In order to make a decision on the gray area devices, we developed a risk metric that determines how many of the devices that are capable of being recovered to group 1. We came up with a euclidean distance based metric where we calculate the euclidean distance between gray area devices and the center of the group 1 devices cluster. The distribution of euclidean distances between the confirmed passing , gray area and confirmed failing devices and the group 1 devices cluster are shown in figure 7. As we can see the gray distribution corresponding to the gray area devices lie in between the confirmed passing and confirmed failing devices. This distribution will help the engineers to decide how many of the gray area devices are good enough to be recovered to group 1.

V. Conclusion

We presented a machine learning-based approach to recover the yield loss left on the table because of the empirically defined test limits. The presented approach takes advantage of the correlation between the CP and NCP test measurement groups. This is implemented by using the trusted group 1 devices and the correlation between the CP and NCP tests in this group of devices to train the machine learning model. In order to verify our predictions, we came up with a limits-agnostic clustering of predicted NCP test measurements to find out the overlap between the two steps to come up with three different buckets of devices. For the third bucket or gray area devices we proposed a euclidean distance based risk metric where we determine how many of the gray area devices can be recovered back to group 1.

Acknowledgment

This work was supported in part by Semiconductor Research Corporation (SRC).

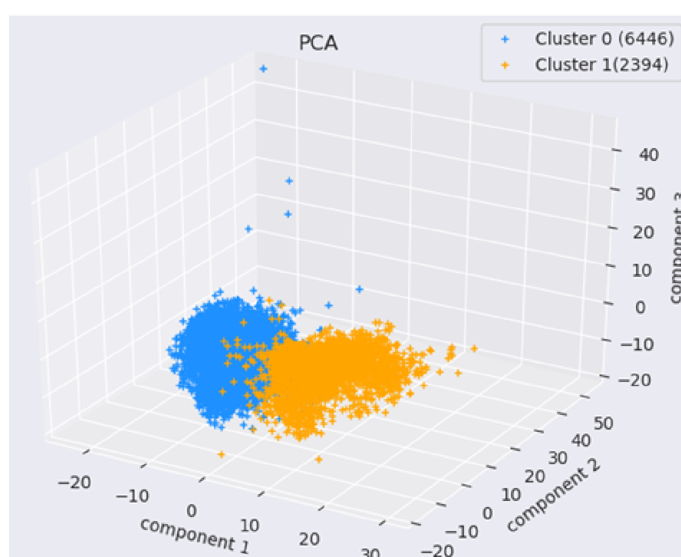

Fig. 6: Clustering of predicted NCP test measurements

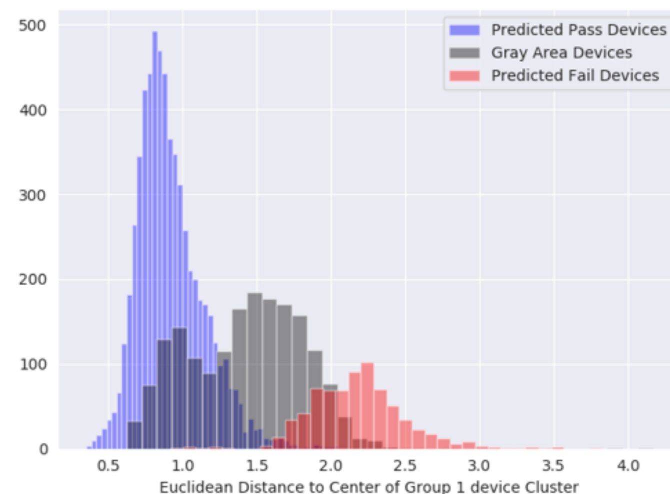

Fig. 7: Euclidean Distance based Risk Metric for Gray area devices

REFERENCES

[1] S. Biswas and R. D. Blanton. Reducing test execution cost of integrated, heterogeneous systems using continuous test data. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 30(1):148–158, 2011.

[2] J. H. Friedman. Multivariate adaptive regression spl@articleStr10, author=H.G. Stratigopoulos and P. Drineas and M. Slamani and Y. Makris, journal=IEEE Transactions on Very Large Scale Integration (VLSI) Systems, title=RF Specification Test Compaction Using Learning Machines, year=2010, volume=18, number=6, doi=10.1109/TVLSI.2009.2017196ines. *The Annals of Statistics*, 19(1):1–67, 1991.

[3] Fan Lin, Chun-Kai Hsu, and Kwang-Ting Cheng. Learning from production test data: Correlation exploration and feature engineering. In *2014 IEEE 23rd Asian Test Symposium*, pages 236–241, 2014.

[4] V. Natarajan, A. Banerjee, S. Sen, S. Devarakond, and A. Chatterjee. Yield recovery of RF transceiver systems using iterative tuning-driven power-conscious performance optimization. *IEEE Design Test*, 32(1):61–69, 2015.

[5] H.G. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris. Rf specification test compaction using learning machines. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 18(6), 2010.

[6] P. N. Variyam, S. Cherubal, and A. Chatterjee. Prediction of analog performance parameters using fast transient testing. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(3):349–361, March 2002.