Reducing Underkill Using Unsupervised Machine-Learning Based Method in Analog/RF IC Testing

Vineeth Amritur 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*—Semiconductor manufacturers strive to maintain a balance between the reliability expectations of System On Chips (SOCs) and ensuring the overall cost of testing is low. Amongst manufacturers, there exists a strong desire to achieve minimal Defective Parts Per Billion (DPPB) as loss of yield is often exacerbated by the exponentially increasing complexity of devices in current technology nodes.Most semiconductor manufacturers balance the need for extensive testing and the additive overhead of test cost and test time associated with it, resulting in manufacturers having exceedingly optimistic device binning or insufficiently elaborate test programs. The aforementioned issues in high-volume manufacturing testing cause faulty or failure-prone ICs to be shipped out and increase the number of customer returns. We term this unfavorable test outcome as Underkill. To this end, we propose an unsupervised machine learning-based methodology to identify potential customer returns and thus reduce Underkill. Specifically, the proposed machine learning model captures any deviations in device behavior across different test insertions. Leveraging unsupervised machine learning models, we extract unique signatures from these devices and use Gaussian methods to learn from the distribution and identify one or more devices that may be a potential customer return(s). We employ our proposed approach on an industrial dataset provided to us by Texas Instruments. Our experimentation with the industrial dataset establishes effectiveness in correctly identifying devices with a higher probability of failure on-site.

***Index Terms*—post-silicon calibration, adaptive, alternate testing, outlier detection**

I. Introduction

The complexity of semiconductor devices has been growing exponential over the years, the increased integration and the functionality of devices necessitate comprehensive post-silicon testing. Semiconductor manufacturers design comprehensive test programs to ensure customer expectation of Defective Parts Per Million (DPPM) are met. The test procedures are often constrained by test time and test cost during High volume manufacturing, optimistic testing thresholds, and insufficient test coverage. This results in test escapes (Underkill). Additionally, several other factors such as process variations and environmental factors also contribute to test escapes, for our preliminary analysis we do not investigate these external factors. Semiconductor manufacturers intuitively implement advance process control techniques to alleviate process variation and environmental factors.

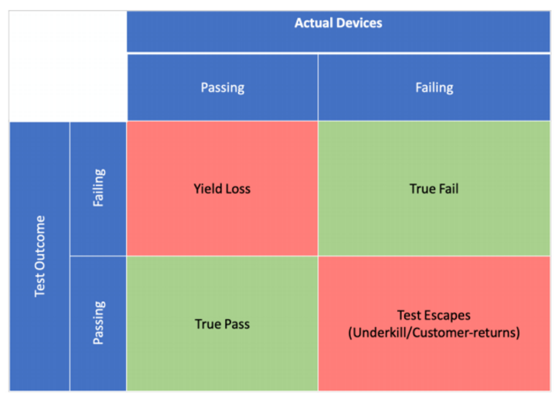

Figure 1: Device Classification Based on Test Outcome

Figure 1 shows the classification of devices based on their performance. The table categorizes devices depending on their performance in the manufacturer's test program and their actual performance on-site. The table classifies devices into the following four labels. If the devices pass the manufacturer's tests and it actually performs as per the design specification, the device is labeled as a "truly passing" device. Conversely, if the device fails the manufacturer's tests and does not perform

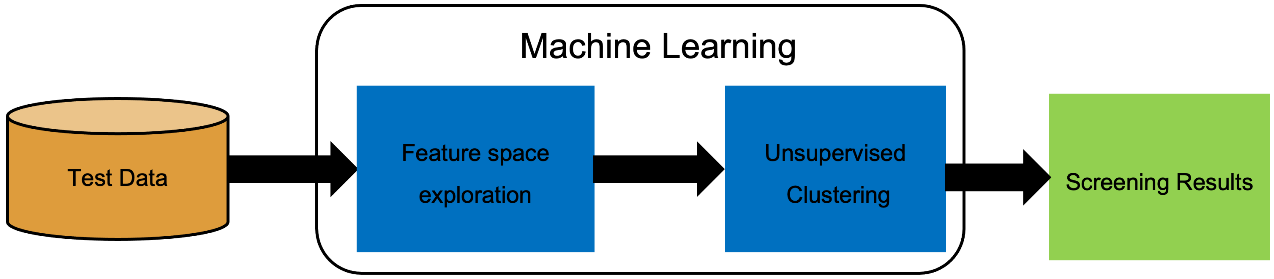

Figure 2: Machine Learning Based Approach

as per the design specification, the device is labeled as a "true failure" and is thus discarded. When the device fails the manufacturer's tests, its actual performance is as per the design specification, then such a device is termed yield loss. Finally, if the device passes the manufacturer's tests but fails to function as per design specification on-site, then the device is labeled as test escape caused due to Underkill.

Underkill is a rare unfavorable outcome for semiconductor manufacturers as this can damage their reputation, require extensive root cause analysis, and expensive retooling to modify the manufacturing and testing process. We perform statistical analysis on customer return devices or test escapes and alleviate Underkill.

II. Related Work

There are several efforts to reduce the instances of Underkill. In safety-critical chips, such as automotive IC, failure on-site can have a catastrophic impact. Ideally, semiconductor manufacturers want to achieve a zero customer return in order to cater to the strict demands of supplying automotive customers. To this end, most automotive device manufacturers include Dynamic Part Average Testing (DPAT) [1] in their test programs. Dynamic part average testing is derived from the concept of the six sigma test, where a given device is labeled as an outlier if the test measurements of the device are six standard deviations away from the mean test measurements. The prior works are broadly classified into univariate or multivariate machine learning models. Univariate machine learning models such as DPAT [1] or I_{DDQ} tests [12] [8] to reject yield based on statistical thresholds. Multivariate methods such as outlier detection using Principal Component Analysis (PCA) [6] [13] leverage the use of Principal Components from a high dimensional test measurement space to screen for defective chips. Other multivariate models include the use of a decision tree [2] to build predictive models to identify failing chips using test data from multiple test insertions and using a subset of tests that highly correlates with customer returns to identify potential failing devices [10].

In this work, we find the optimal feature space amongst the test measurements from different test insertions(measured at different temperatures) to isolate potential failures and extract device signatures. Once we extract this signature, we use unsupervised clustering as an extension of previously researched methods in [5] [13]. The novelty of our proposed approach is in the feature-space exploration without the loss of device information in conjecture with an unsupervised clustering technique.

III. Proposed Methodology

Our proposed methodology shown in figure 2 aims to identify devices that are statistically similar to faulty devices/customer returns and are distinct from passing devices. To this end, the test measurement data from test insertion recorded at high and room temperature is used as input to the two-stage unsupervised machine learning model.

The proposed machine learning model consists of the following two-stage :

- First, we explore the test measurements to generate an optimal feature space in order to separate a potential faulty chip from nominal device data
- We cluster devices using Gaussian Mixture Model (GMM) [7] based on the device measurements in the feature space obtained in the previous stage

A. *Feature Space Exploration*

The objective of the proposed step is to search for an n-dimensional feature space amongst the given test data, wherein a faulty device is isolated from the distribution of “truly passing’ devices in the above-defined feature space. The use of temperature cycle testing to perform device characterization is well documented, especially in analog/RF ICs [9]. The variations in a device’s test performance at high temperature and room temperature can help one infer the device’s individual performance signatures. The aforementioned signatures can be employed in grouping identically performing devices. Any anomalous variation in the performance measurements of devices at different temperature test insertions can indicate that they are probably marginal devices and are thus failure-prone. The above inference is leveraged in our methodology to obtain a feature space to isolate the faulty devices from the ”truly passing“ ones. We compute the test measurements’ deferential recorded across the test insertion at different temperatures. The above computed deltas, across all the test measurements, are our chosen n-dimensional feature space.

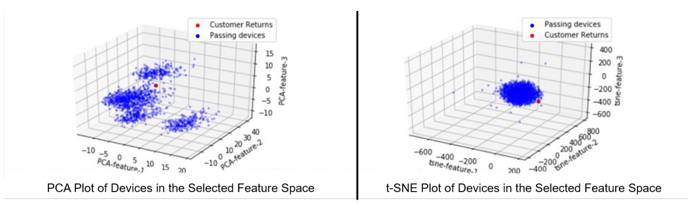

Figure 3: PCA and t-SNE Plots of Devices in the Selected Feature Space

B. Unsupervised Clustering

In the above n-dimensional feature space, we perform unsupervised clustering. Unsupervised clustering methods are often used to detect anomalies. These anomalies are points that fall outside of known good clusters or belong to clusters that are known to be associated with faults. The objective of clustering in our context is to identify the clusters of ”truly passing“ devices to separate them from faulty device distribution. The devices’ test measurements form a Gaussian distribution. Due to the Gaussian nature of our feature space, we chose Gaussian Mixture Model [7] to perform unsupervised clustering. GMM is a distribution-based clustering model. GMM assumes that a given distribution consists of several Gaussian distributions, and these distributions represent a cluster. Hence, GMM groups data points belonging to a single distribution together. The Gaussian cluster of the faulty device is statistically dissimilar with respect to the cluster of ”truly passing“ devices and hence, by using GMM clustering we seek to identify future customer returns.

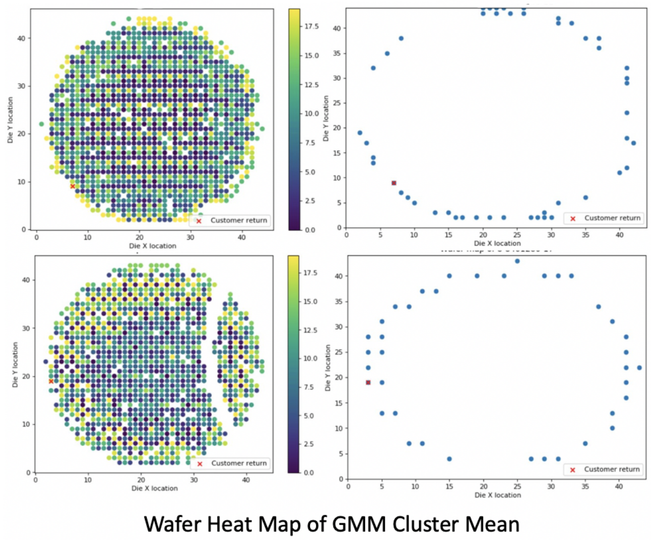

Figure 4: Wafer Heat Map of GMM Cluster Mean

Finally, in the last step, devices are screened using a confidence estimator that reverses or accepts the ”passing label“ assigned by the test program. The confidence estimator informs the test engineer of the probability of a device passing or failing on-site. The further a given device is from the center of ’truly passing’ devices distribution, the higher possibility of failure on-site. Devices belonging to a cluster associated with a known faulty device/customer return have a higher probability of failure. The confidence estimator builds on the above-mentioned theory.

IV. Results

To evaluate our proposed Overkill reduction methodology, we experimented with a dataset provided by our industrial collaborators at Texas Instruments. The following two sections will elaborate more on the dataset and the results obtained from our experimentation.

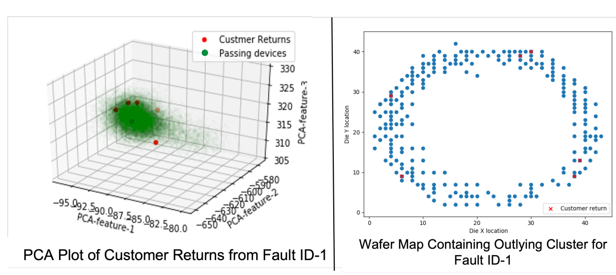

Figure 5: Analysis of Devices Belonging to Fault ID-1

A. Dataset Descriptions

The dataset provided by TI consists of test measurements of 24,000+ devices across two different test insertions. There are 19 wafers in the dataset that contains relevant wafer information such as wafer-id, die-id, and X-Y location of devices. There are 19 customer returns in our dataset, coincidentally one on every wafer. Further, the customer returns are analyzed upon return and categorized by the manufacturer into six unique fault-id types. Preliminary analysis of the dataset indicates that the occurrence of customer returns is rare, and subsequent categorization of customer returns gives us a limited number of devices to perform our statistical analysis. The first three fault-ids have sufficient customer returns to perform our analysis. We conduct our experimentation with the aforementioned fault-ids and use a resampling technique to generate additional customer returns using Adaptive synthetic sampling to build our confidence estimator. However, we record similar trends in customer return devices belonging to the remaining fault-ids and a simple extension.

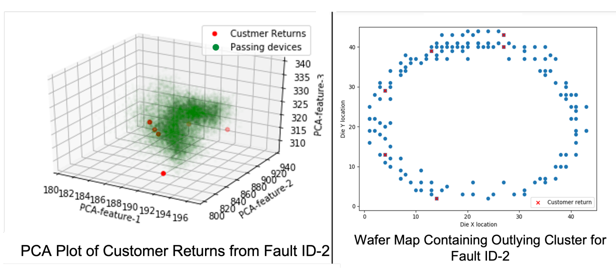

Figure 6: Analysis of Devices Belonging to Fault ID-2

B. Experimental Results

Our experiment starts with extracting the 711 test measurement from the test insertion data collected at high and room temperature. These test measurements are used to perform our feature space exploration and unsupervised clustering.

Feature Space Exploration: In this first step, we compute the difference in test measurements across the hot/room temperature test insertion. The resulting delta measurement is our chosen n-dimensional feature space where n=711. We observe that the n-dimensional feature space can effectively isolate the customer return devices from the "truly passing" device distribution. To visualize the isolation of customer return devices; all the devices for a given wafer are plotted in our chosen n-dimensional feature space. We compress our n-dimensional feature space to enable plotting in three dimensions. We use dimensionality reduction techniques such as PCA [4] and t-SNE [11] for our Visualization. For representation purposes the plot of a wafer is shown in figure 3.Similar patterns are seen across the 19 wafer with variability in degree of isolation. These results validate the effectiveness of our feature space.

Unsupervised Clustering using Gaussian Mixture Model: Next, we perform unsupervised clustering using GMM. The GMM clustering is performed on the n-dimensional feature space obtained in the previous stage. The number of Gaussian clusters is selected using the Bayesian Information Criterion (BIC) [3]. The clusters and their cluster means are obtained and recorded. The plot in figure 4 shows the clusters obtained from GMM clustering. A wafer heat map is generated based on the cluster mean values to evaluate our hypothesis. The cluster containing the customer return is highlighted, and we observe that the cluster is outlying with respect to the "truly passing" devices. Since we have sufficient

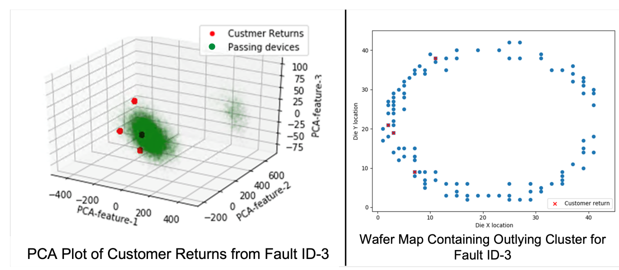

Figure 7: Analysis of Devices Belonging to Fault ID-3

customer return devices from three "Fault-IDs" in our dataset, we perform our analysis on them. The first two Fault-IDs have six customer return devices, and the last Fault-ID has four customer returns. We use adaptive synthetic (ADASYN) data generation to capture the distribution of customer returns and replicate them. This step creates sufficient samples to evaluate our proposed approach. We use the distribution of customer returns (actual and sampled) to build our GMM clusters. The results of our experimentation for the three fault-id types are presented in figures 5, 6 and 7. The left plot in the figures illustrates the distribution of devices from the wafer that contains customer returns of respective fault-id. The right plot shows the location of the faulty cluster in a wafer for each fault-id. To evaluate the effectiveness of clustering, we train the GMM with resampled customer returns and observe that the actual customer return belongs to the faulty cluster. We further validate the proposed methodology using leave one out validation using the resampled customer returns to observe that the outlying cluster contains the actual customer return.

V. Conclusion

The proposed methodology explored the ability to reduce the instances of test escapes (underkill) in analog/RF IC testing using multi-variate unsupervised machine learning models. The model extracts device signatures from its test measurements in high temperature and room temperature conditions, based on these signatures unsupervised clustering is performed to group devices that are statistically similar. The clusters that are outlying with respect to the distribution of passing device clusters have a higher probability of failure on-site. The model is evaluated using an industrial dataset provided to us by our collaborators at Texas Instruments. The results of our experimentation on the aforementioned dataset correctly identify the clusters of devices that contain known customer returns and establish that these devices are at the edges of passing device distribution.

Acknowledgment

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

References

[1] AEC. AEC - Q001 Rev-D. http://www.aecouncil.com/Documents/AEC_Q001_Rev_D.pdf.

[9] Z.D. Schwartz, A.N. Downey, S.A. Alterovitz, and G.E. Ponchak. High-temperature rf probe station for device characterization through 500/spl deg/c and 50 ghz. *IEEE Transactions on Instrumentation and Measurement*, 54(1):369–376, 2005.

[2] Sounil Biswas and Ronald D. Blanton. Statistical test compaction using binary decision trees. *IEEE Design Test of Computers*, 23(6):452–462, 2006.

[3] Arijit Chakrabarti and Jayanta K. Ghosh. Aic, bic and recent advances in model selection. In Prasanta S. Bandyopadhyay and Malcolm R. Forster, editors, *Philosophy of Statistics*, volume 7 of *Handbook of the Philosophy of Science*, pages 583–605. North-Holland, Amsterdam, 2011.

[4] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[5] A.E. Gattiker and W. Maly. Current signatures [vlsi circuit testing]. In *Proceedings of 14th VLSI Test Symposium*, pages 112–117, 1996.

[6] Peter M. O'Neill. Production multivariate outlier detection using principal components. In *2008 IEEE International Test Conference*, pages 1–10, 2008.

[7] Douglas A. Reynolds. Gaussian mixture models. In *Encyclopedia of Biometrics*, 2009.

[8] S.S. Sabade and D.M. Walker. Evaluation of effectiveness of median of absolute deviations outlier rejection-based i_{DDQ} testing for burn-in reduction. In *Proceedings 20th IEEE VLSI Test Symposium (VTS 2002)*, pages 81–86, 2002.

[10] Nik Sumikawa, Jeff Tikkanen, Li-C. Wang, LeRoy Winemberg, and Magdy S. Abadir. Screening customer returns with multivariate test analysis. In *2012 IEEE International Test Conference*, pages 1–10, 2012.

[11] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[12] T. W. Williams. i_{DDQ} test: Sensitivity analysis of scaling. *Proc. 1996 IEEE International Test Conference*.

[13] Jianbo Yu. Fault detection using principal components-based gaussian mixture model for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 24(3):432–444, 2011.