Improving Diagnosis Resolution with Population Level Statistical Diagnosis

Kun Young Chung, Shaun Nicholson, Soumya Mittal, Martin Parley
Qualcomm Technologies, Inc., San Diego, CA, USA
kunyoung/shaunn/soumyam/mparley@qti.qualcomm.com

Gaurav Veda, Manish Sharma, Matt Knowles, Wu-Tung Cheng
Siemens Digital Industries Software, Wilsonville, OR, USA
gaurav.veda/manish.sharma/matthew.knowles/wu-tung.cheng@siemens.com

Abstract

In this paper, we present a diagnosis resolution improvement methodology for scan-based tests. We achieve 89% reduction in the number of suspect diagnosis locations and a 2.4X increase in the number of highly resolved diagnosis results. We suffer a loss in accuracy of 1.5%. These results were obtained from an extensive silicon study. We use data from pilot wafers and 11 other wafers at the leading-edge technology node and check against failure analysis results from 203 cases. This resolution improvement is achieved by considering the diagnosis problem at the level of a population (e.g. a wafer) of failing die instead of analyzing each failing die completely independently as has been done traditionally. Higher diagnosis resolution is critical for speeding up the yield learning from manufacturing test and failure analysis flows.

Introduction

With ever increasing time to market pressures, there is a constant requirement to shorten the yield ramp curve as much as possible for new technology nodes as well as new products on existing nodes. It is also well known that despite test chips, the best design for manufacturing practices during the product design phase, and various yield enhancement measures in the fab like OPC, in-line defect inspection, metrology, contaminant free manufacturing, process control, etc., some previously unknown systematic yield loss mechanisms still percolate into the final manufactured product. Such issues can then only be learned from manufacturing test failures at the end of the line before the product is shipped.

For digital logic, scan-based testing is an industry standard electrical test methodology [1-8]. It is particularly conducive to yield learning as automatic diagnosis of test failures is possible, and there are commercial software tools available to do so. These scan diagnosis tools can automatically analyze manufacturing scan test failures from a defective die and produce a list of potential defect sites within the logic under test. A defect site is a short or an open in a specified layer and bounding box. One of these sites most likely contains the actual defect. These potential defect sites are often referred to as *diagnosis suspects* or *candidates*. There are two main metrics that can be used to measure the quality of scan diagnosis results:

Diagnosis Accuracy: A diagnosis result is said to be accurate if the actual defect site matches one of the diagnosis suspects.

Diagnosis Resolution: It is defined as the number of diagnosis suspects in the diagnosis result. The more suspects, the less resolved is the result. Consequently, it is worse for downstream yield learning and failure analysis flows.

When diagnosing a single test failure in isolation, the diagnosis resolution gets limited by logic and electrical equivalencies. These equivalencies can appear due to only a small number of tests being run, as well as there being a limited number of observation points (output pins) in manufacturing test. Equivalencies can also appear due to our inability to exactly predict defect behavior. However, if we expand our view to look at the diagnosis problem at the level of a population of failing die, there is opportunity to improve diagnosis resolution beyond these limits. We need to make the key assumption that all the failing die are impacted by a common defect root cause distribution. Root causes are well defined failure mechanisms, such as opens or bridges on an interconnect metal layer, opens on a Via layer, opens or bridges on a cell internal layer, opens and bridges on specific layout patterns, etc. This paper presents *Statistical Diagnosis*, a new methodology that takes advantage of defect likelihoods learned from analyzing a population of failing die and using these to create diagnosis results with higher resolution than the current state-of-the-art techniques.

Previous work

Several past publications have addressed the problem of improving diagnosis resolution. One class of techniques [9] has focused on creating special test patterns for better resolution. However, such techniques are more focused on improving resolution of a specific diagnosis report, rather than in general improving the diagnosis resolution of a population of diagnosis reports as would be required for yield learning flows. Another class of techniques use supervised or semi-supervised machine learning methods to classify diagnosis suspects as good or bad candidates [10,11]. These techniques are reliant on there being a number of diagnosis reports which have just a single suspect. The suspects in such reports are assumed to be good and provide the training data for the machine learning classifier. The classifier considers each suspect in a diagnosis report

independently and classifies it without knowledge of the other suspects in the same report. In [11], a further step is taken which does consider the defect as a whole. Defects are clustered based on the feature similarity of their remaining suspects. The approach then again relies on there being reports with just a single remaining suspect, which are considered good suspects. Any suspect which is too far away from the good suspects are marked as bad. Such techniques do not take full advantage of the shared defect root cause mechanisms in a population of defective die.

There has also been extensive work done on learning common defect root causes from volume diagnosis [12,13]. However, up until this work, they have not combined the diagnosis step and the learning from diagnosis data into one combined optimization problem to determine the most likely diagnosis suspects. The work in this paper strives to do just that. Our work is based on the root cause deconvolution (RCD) [12] volume diagnosis learning engine.

Root Cause Deconvolution (RCD)

RCD analyzes a population of diagnosis reports to determine the most likely defect root cause distribution in that population. A root cause distribution ($\boldsymbol{R}$) is a vector with length equal to the number (M) of root causes, which sums up to 1. The entry in the vector associated with the root cause $\boldsymbol{r}$ denotes the probability of the root cause $\boldsymbol{r}$ given the root cause distribution $\boldsymbol{R}$: $\boldsymbol{P}(\boldsymbol{r}|\boldsymbol{R})$. We now proceed to give a brief description of RCD.

The main idea behind RCD is to find a small set of root causes that would provide a good explanation for all the failures being observed. The first step in RCD analysis is to obtain an estimate of how likely we are to observe this defect for a given root cause mechanism. We compute this for every defect and root cause pair. Formally, $P(d_i \mid r_j)$ is the probability of observing the failing behavior associated with defect d_i, given the underlying root cause is r_j. In order to compute this, we examine each of the suspects called out by a diagnosis report, and determine the probability of any of these suspects occurring, given a particular root cause. For instance, suppose one of the root causes is Open Metal 1, and it is critical area based. If the root cause for failure is Open Metal 1, the probability of observing a suspect is equal to the Open Metal 1 critical area associated with it, divided by the total Open Metal 1 critical area in the design. Since a defect has several suspects associated with it, the probability of the defect is a weighted sum of the probability of each suspect. The weight of a suspect is dependent on how much confidence we have in this particular suspect, and is computed by the diagnosis algorithm.

Once we have computed $P(\boldsymbol{d_i} \mid \boldsymbol{r_j})$ for every defect and root cause combination, we can compute the likelihood of a defect given a root cause distribution R, as follows:

$$P(d_i \mid R) = \sum_{j=1}^{M} P(r_j \mid R)\, P(d_i \mid r_j)$$

$$Objective\ Function = \prod_{i=1}^{N} P(d_i \mid R)$$

The objective function in RCD is the product of the likelihood of all the defects. In the above equation, M is the number of root causes and N is the number of defects. RCD uses the EM (Expectation Maximization) algorithm [14] to maximize the above function and obtain the root cause distribution, i.e., we want to determine the value of $\boldsymbol{P}(\boldsymbol{r}|\boldsymbol{R})$ for every value of r, that maximizes the objective function. The EM algorithm consists of two steps – the E-step and the M-step – which are called one after the other until convergence. We initialize with a random root cause distribution. Then, in the E-step, we compute the probability that a root cause is responsible for generating a defect, using the current estimate of the root cause distribution R.

$$P(r_j|d_i) = \frac{P(d_i|r_j)P(r_j|R)}{P(d_i)} \qquad \text{(E-Step)}$$

where the denominator $P(d_i)$ is the normalization constant (the sum of the numerator over all the root causes). In the M-step, we re-estimate the root cause distribution using the $P(r_j|d_i)$ computed in the E-step. We keep iterating until convergence, which is guaranteed.

$$P(r_j|R) = \frac{\sum_{i=1}^{N} P(r_j|d_i)}{N} \qquad \text{(M-Step)}$$

Statistical Diagnosis: A New Methodology for Diagnosis Resolution Improvement

Statistical Diagnosis has two main novel features:

1. It incorporates the decisions made by the diagnosis algorithm while analyzing an individual die, into the RCD learning model. This allows compensation for the fact that occasionally the diagnosis algorithm has to include suspects in the report due to the cost associated with simulating exact defect behavior.
2. It uses the learned defect root cause distribution to calculate a population-based probability of each diagnosis suspect in all diagnosis reports in the target population. Specifically, it plugs in the learned root cause distribution into the equation for the E-step mentioned in the previous section. This gives it a post RCD probability of each root cause for each suspect.

The suspect level probabilities are then used to drop any suspect assigned a zero probability to improve the diagnosis resolution for individual die.

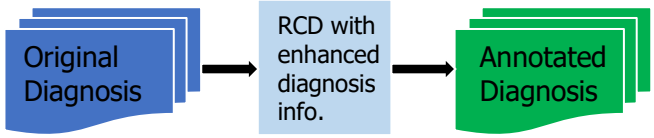

Figure 1. Flow using Statistical Diagnosis

We now illustrate *Statistical Diagnosis* using an example. Suppose we have a population of M diagnosis reports. Suppose the pareto R after running RCD analysis comprises of root causes R1, R2, R3, and R4, as shown in Figure 2. For example, R1 could be Cell Open Metal 1, i.e., opens on the Metal 1 layer inside the cells. Further, suppose there are 50 possible root causes from R1, R2, …, R50.

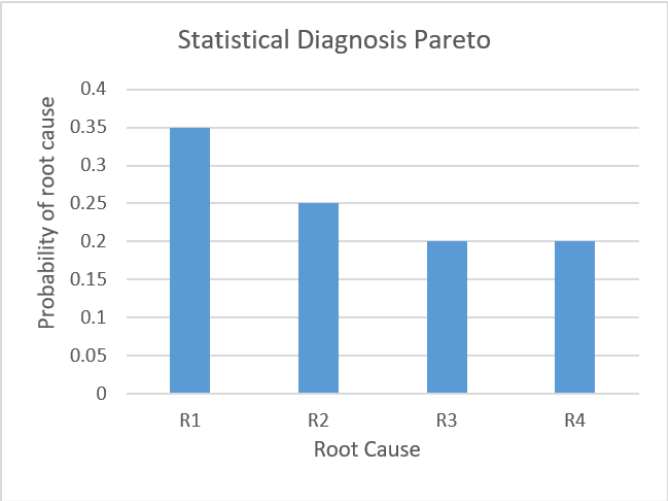

Figure 2. Example root cause distribution / pareto.

A typical diagnosis report will consist of a number of suspects. Suppose diagnosis report D1 comprises of 5 suspects. Each suspect is associated with a physical location on the chip and a number of defect root causes. Table 1 gives an example of the defect root causes associated with each of the 5 suspects contained in diagnosis report D1.

Table 1. The suspects associated with an example diagnosis report D1 and their possible root causes.

Suspect ID	Possible root causes associated with the suspect
S1	R6, R8, R9
S2	R6, R12, R18
S3	R1, R6
S4	R5
S5	R4, R7

Let us consider suspect S1. The physical location associated with S1 only contains faults which can be caused by root causes R6, R8, and R9. Based on the results of volume diagnosis, we have already ruled out R6, R8, and R9 as possible root causes. As shown in Figure 2, the joint analysis of all diagnosis reports in the population leads us to believe that only R1, R2, R3, and R4 are the possible root causes for this failing die population. Therefore, using *Statistical Diagnosis*, we are able to rule out S1 as a possible suspect.

Formally, the probability of root causes R6, R8, and R9 in the pareto R is 0, i.e.,

$$\boldsymbol{P(R6|\,R) = P(R8|\,R) = P(R9|\,R) = 0}$$

The suspect S1 can only be caused by root causes R6, R8, and R9. Therefore,

$$\begin{aligned}\boldsymbol{P(S1|D1,R)} = \;& \boldsymbol{P(S1|D1,R6) * P(R6|\,R)} \\ & + \boldsymbol{P(S1|D1,R8) * P(R8|\,R)} \\ & + \boldsymbol{P(S1|D1,R9) * P(R9|\,R) = 0}\end{aligned}$$

Similarly, we are able to rule out S2, and S4 as possible suspects, since none of the root causes that can cause the faults associated with S2 and S4 exist in the root cause distribution. Thus, after running *Statistical Diagnosis*, the diagnosis report will be left with just suspects S3 and S5.

Silicon Results

An extensive silicon study was undertaken to determine the efficacy of *statistical diagnosis*. The study was done on some pilot wafers and 11 failing wafer populations. We only have limited information available for the pilot wafers - the number of PFA done and whether *statistical* diagnosis led to any loss in accuracy.

Each population consisted of all hard (i.e., failing at all test voltages), stuck-at pattern failures from a single wafer. The number of diagnosis reports in a population ranged from 140 to 419. Note that each report could have multiple defects called out. Physical failure analysis was performed on 203 die to obtain the true defect locations. This was used as a golden reference to measure the impact to diagnosis accuracy.

Table 2. PFA results and no. of diagnosis reports per wafer.

	No. of diagnosis reports	No. of PFA cases
Pilot wafers	-	22
W01	213	22
W02	409	28
W03	163	28
W04	291	23
W05	142	21
W06	241	18
W07	240	15
W08	140	14
W09	249	13
W10	321	6
W11	419	2

Statistical diagnosis was set up to run on each of the populations and the following metrics were tracked:

1. Overall Diagnosis Resolution: Reduction in total number of diagnosis suspects over the entire set of diagnosis reports in each population.
2. High Quality Diagnosis Reports: Increase in the number of diagnosis reports with less than or equal to three diagnosis suspects. Based on communication with industry sources, reports with three or less diagnosis suspects are ideal candidates to perform PFA.
3. Diagnosis Accuracy Loss: For each die with confirmed defect location from PFA, check whether the matching diagnosis suspect was dropped from the result using statistical diagnosis.

The results over all the wafers are summarized in the table below.

Table 3. Summary of statistical diagnosis results on silicon.

Metric	Result
Average reduction in #diagnosis suspects/wafer	89.03%
Average increase in number of diagnosis results with <=3 suspects	2.4X
Percentage of FA cases with lost diagnosis accuracy	1.5% (3 out of 203)

Figure 3 shows detailed information on the diagnosis resolution improvement achieved for each wafer. The Y-axis represents the percentage reduction in the number of suspects. For example, for wafer W01, the diagnosis reports after annotation by *Statistical Diagnosis* contain 87% fewer suspects on average, as compared to the original diagnosis reports.

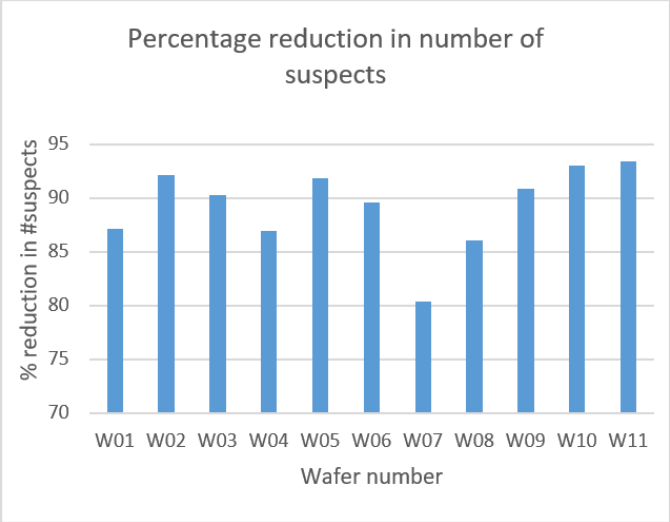

Figure 3. Suspect count reduction (%) per wafer.

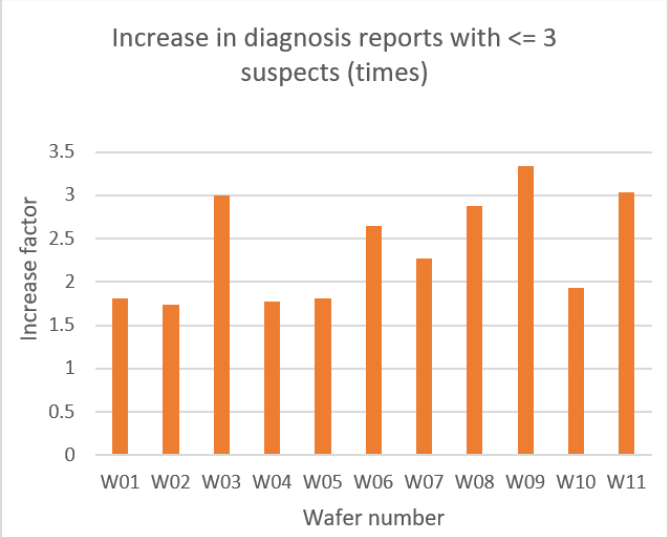

Figure 4. Increase in diagnosis reports with 3 or fewer suspects per wafer.

Figure 4 shows detailed information about the increase in the number of high quality diagnosis reports. As defined above, a high quality diagnosis report is one which contains 3 or fewer suspects. For example, for wafer W01, after annotation by *Statistical Diagnosis*, 69 reports contain 3 or fewer candidates. For the same wafer, only 38 original diagnosis reports had 3 or fewer candidates. Therefore, the increase in the number of high quality diagnosis reports is 69/38 = 1.8X. This (1.8X) is the number shown in Figure 4.

We investigated the 3 PFA cases for which *Statistical Diagnosis* had a loss in accuracy. One of these cases had 9 defects in the diagnosis report corresponding to the die and 23 suspects overall. Another case had 38 defects in the diagnosis report, and 143 suspects overall. These two cases correspond to poor diagnosis quality, likely due to very different defect behavior. As such, these reports are unlikely to be considered for PFA. The third case where we had a loss of accuracy was a reasonably good diagnosis report: 1 defect and 6 suspects. We are still investigating the cause for a loss of accuracy for this case.

As illustrated in Figure 1, RCD is a sub-step in statistical diagnosis that yields a defect root cause distribution. The defect root cause distribution (pareto) for each population in *Statistical Diagnosis* was also matched against the PFA defect pareto for 9 of the 11 wafers. The other 2 wafers and pilot wafers were excluded from this analysis since they had less than 10 PFA cases per wafer. We observe that:

1. For 5 out of the 9 wafers there is a strong correlation between the two pareto. The top 2 or 3 root causes and their ordering is identical for the ***Statistical Diagnosis*** pareto and the PFA defect pareto. We will shortly present an example illustrating this.
2. For 4 out of the 9 wafers there is very good correlation between the two pareto. The top root cause matches. The next three top root causes are the same, but have different ordering. Note that the PFA defect pareto is computed using a relatively small number of samples, as compared to the number of diagnosis reports. Therefore, while it is indicative of the root cause distribution, it might not be able to provide the exact ordering and proportion for each root cause. Therefore, even though the ordering between root causes 2-4 for RCD is different from that of the PFA defect pareto, it does not imply that the *Statistical Diagnosis* pareto is incorrect.

We now present an example of a wafer where the *Statistical Diagnosis* pareto and the PFA pareto match relatively well. Figure 5 shows the FA pareto for wafer W03. The top 3 root causes are Root Cause A, Root Cause B, and Root Cause C in that order. Figure 6 shows that these are also the top root causes pointed out by RCD for this wafer. The ordering is the same as well. We see that the fourth root cause is different for the FA and the *Statistical Diagnosis* pareto. However, the 5th and 6th root causes are again the same.

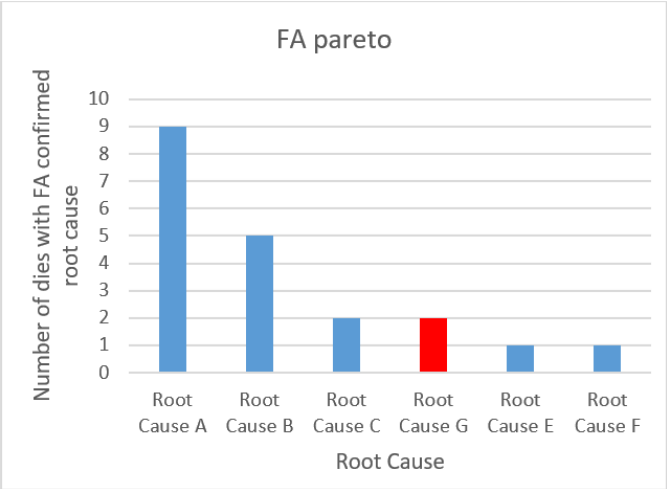

Figure 5. FA pareto for Wafer W03.

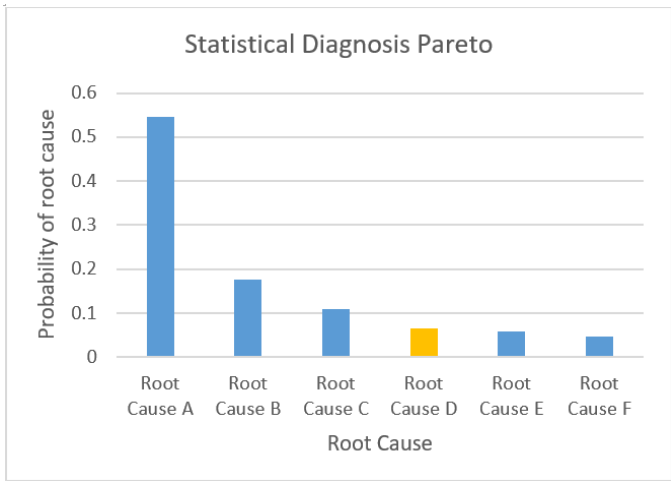

Figure 6. Statistical Diagnosis Pareto for Wafer W03.

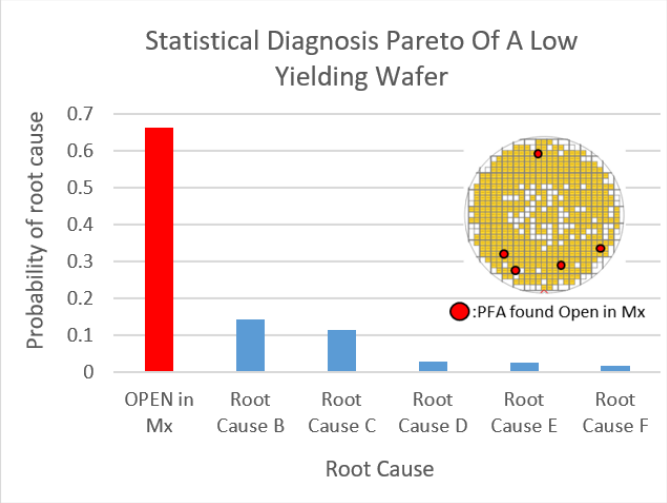

Figure 7. Statistical Diagnosis Pareto matching the major yield limiter.

Conclusion

Statistical Diagnosis pareto can also be used in predicting or identifying the dominant defect mechanism of low yielding wafers. Figure 7 shows one example case where interconnect open in a particular metal layer, say Mx, is highlighted in the *Statistical Diagnosis* pareto for the given wafer. In the wafer map in Figure 7, the yellow locations correspond to failing die, while the red locations indicate the dies on which PFA was performed. All 5 PFAs on this wafer found the same open defects in Mx, which matched to the weak point confirmed in a process setting by foundry. This approach can lead to faster yield learning by reducing the number of PFAs and even skip PFA if there is sufficient supporting evidence from other sources such as inline inspection data.

In this paper we propose a new statistical diagnosis methodology for diagnosis resolution improvement based on population level analysis. The methodology was validated in a large silicon study that showed significant improvement in diagnosis resolution using statistical diagnosis with very minimal impact to diagnosis accuracy. Such resolution enhancement can greatly speed up yield learning flows because this (i) shortens time to identify systematics with reduced noise in diagnostics, (ii) eliminates low-probability suspects to speed up target failure analysis, and (iii) enables characterizing and predicting underlying failure root-causes even before pursuing physical failure analysis.

References

[1] T. Bartenstein, D. Heaberlin, L.Huisman and D. Sliwinski, "Diagnosing Combinational Logic Designs using the Single Location At-ATime (SLAT) Paradigm," International Test Conference, 2001.

[2] S. Venkataraman and S. B. Drummonds, "POIROT: A Logic Fault Diagnosis Tool and Its Applications", in Proc. of Intl. Test Conf., pp. 253-262, 2000.

[3] R. Desineni, O. Poku, and R. D. Blanton, "A Logic Diagnosis Methodology for Improved Localization and Extraction of Accurate Defect Behavior," International Test Conference, 2006.

[4] M. Sharma, S. Schwarz, J. Schmerberg, K. Yang, T.-P. Tai, Y.-S. Chen, C.-Y. Chuang, F.-M. Kuo, M. Brennan, J. Yeh and A. Ma, "Layout-Aware Diagnosis Leads to Efficient and Effective Physical Failure Analysis," ISTFA 2011.

[5] E. Amyeen, D. Nayak, and S. Venkataraman, "Improving Precision Using Mixed-level Fault Diagnosis," International Test Conference, 2006.

[6] X. Fan, M. Sharma, W.-T. Cheng, and S. M. Reddy, "Diagnosis of Cell Internal Defects with Multi-cycle Test Patterns," Asian Test Symposium 2012

[7] F. Hapke, M. Reese, J. Rivers, A. Over, V. Ravikumar, W. Redemund, A. Glowatz, J. Schloeffel, and J. Rajski, "Cell-Aware Production Test Results from a 32-nm Notebook Process," International Test Conference, paper 1.1, 2012.

[8] H. Tang, B. Benware, M. Reese, J. Caroselli, T. Herrmann, F. Hapke, R. Tao, W.-T. Cheng, and M. Sharma,

"Diagnosing Cell Internal Defects Using Analog Simulation-based Fault Models," Asian Test Symposium, 2014

[9] H. Tang, A. Jain, S. K. Pillai, D. Joshi, S. Rao, "Using Cell Aware Diagnostic Patterns to Improve Diagnosis Resolution for Cell Internal Defects", Asian Test Symposium, 2017.

[10] Y. Xue, O. Poku, X. Li, R. D. Blanton, "PADRE: Physically-Aware Diagnosis Resolution Enhancement", International Test Conference, 2013.

[11] Q. Huang, C. Fang, R. D. Blanton, "LAIDAR: Learning for Accurate and Ideal Diagnostic Resolution", International Test Conference, 2020.

[12] B. Benware, C. Schuermyer, M. Sharma and T. Hermann, "Determining a Failure Root Cause Distribution from a Population of Layout-Aware Scan Diagnosis Results", in IEEE Design & Test of Computers, vol. 29, no. 1, 2012.

[13] C. Schuermyer; S. Palosh; P. Babighian; Y. Pan, "Application of Bayesian Machine Learning To Create A Low-Cost Silicon Failure Mechanism Pareto", Advanced Semiconductor Manufacturing Conference, 2019.

[14] C. Bishop. Pattern Recognition and Machine Learning, Springer-Verlag, 2006. Chapter 9.