Harnessing Process Variation Signature for Dynamically Selection of Probe-Test Flow

Ali Ahmadi*, Constantinos Xanthopoulos* and Yiorgos Makris*
*Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080

***Abstract*—We propose a methodology for dynamically selecting an optimal probe-test flow which reduces test cost without jeopardizing test quality. The granularity of this decision is at the wafer-level and is made before the wafer reaches the probe station, based on an e-test signature which reflects how process variations have affected this particular wafer. The proposed method offers flexibility by optimizing test flow per process signature, and its implementation is simple and compatible with most commonly used Automatic Test Equipment. Furthermore, unlike static test elimination approaches, whose agility is limited by the relative importance of the permanently dropped tests, the proposed method is capable of exploring test cost reduction solutions which achieve very low test escape rates. Decisions are made by an intelligent system which maps every point in the e-test signature space to the most appropriate probe-test flow. Training of the system seeks to optimize the test flow of each process signature in order to maximize test cost reduction for a given target of test escapes, thereby enabling exploration of the trade-off between test cost reduction and test quality. The proposed method is demonstrated on an industrial dataset of a million devices from a 65nm Texas Instruments RF transceiver.**

I. Introduction

Continuous pressure for superior performance, along with intensified process variations and non-idealities in the latest semiconductor manufacturing technology nodes, have resulted in stringent limitations in the cost that can be devoted to testing each die, in order to ensure that it functions correctly before it is shipped to a customer. Especially in the analog/RF domain, where industrial practice still relies largely on lengthy test procedures and expensive instrumentation to explicitly measure the performances of a device and compare them to its specifications, test cost reduction has become a crucial requirement for maintaining profitability. Among the various directions which have been explored towards reducing test cost, significant effort has been invested in challenging the practice of subjecting every die in production to the exact same set of tests. Generally termed “adaptive test”, methods in this category seek to customize the test process to the needs of a target die, wafer, or lot, anticipating that the benefits from a reduced test flow will outweigh the effort and expenditure required for such customization.

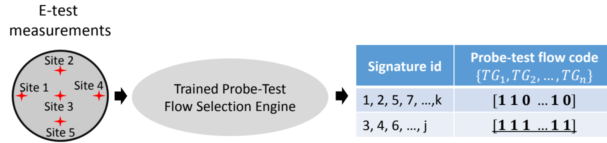
A very simple and commonly practiced approach to test cost reduction is to monitor the relative effectiveness of each test and drop the ones which contribute little or not at all to the overall test effectiveness [1]–[3]. Such decisions are usually static and are easy to implement on the ATE by exclusion of the relevant portion of the test program. However, the agility of such methods is insufficient to support solutions which offer savings yet maintain very low test escapes; essentially, they are bound by the percentage of faulty die that the dropped tests uniquely detect. Advanced versions of this idea, wherein statistical correlation between the dropped and retained tests is leveraged to predict the outcome of the former, have also been proposed [2], [4]–[6]. While additional ATE or external support is required to run the statistical models on-the-fly during test, these methods have demonstrated marked improvement in test quality. Still, the decision models remain static or only infrequently retrained to account for major events which can change the statistical profile of the production.

In [7], we proposed a new methodology for establishing an adaptive test flow which is deployable with minimum test operation support. For each wafer, this approach provides a decision as to whether to test it through the complete probe-test flow or a reduced test flow, in which some of the test groups are eliminated. This decision is made at an early stage, before the wafer reaches the probe station, driven through e-test[1] measurements.

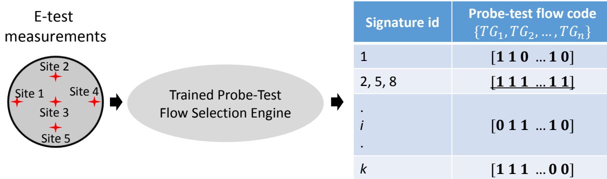
Figure 1 (a) depicts this adaptive method: a statistically trained entity examines the e-test data of a wafer and, depending on the extracted signature, it selects the appropriate test flow code. The test flow code is a vector, wherein each test group corresponds to a bit, with value ‘1’ signifying inclusion and value ‘0’ signifying exclusion of that test group. In the work described in [7], one of the choices of test flow code is the all ‘1’ vector while the other choice is *a single* carefully selected subset of test groups which maximizes test cost reduction for a target test escape rate.

However, a single reduced test flow (i.e., subset of all test groups) which is optimized across all process signatures is a restrictive and sub-optimal choice. Indeed, depending on how a wafer has been impacted by process variations, a different reduced test flow may offer the best option. Therefore, in this work we seek to investigate the utility of test flow optimization per process signature, towards achieving higher test cost reduction. To accomplish this, an optimization algorithm is employed to statistically select the best test flow for each signature such that test cost reduction is maximized while the required test quality is achieved. Figure 1 (b) depicts the proposed approach. Similar to [7], the decision is made before the wafer reaches the probe station and is driven by e-test measurements. The trained test flow selection engine processes the e-test measurements of a wafer, extract its process signature, and accordingly selects the most appropriate

[1]By the term e-test we refer to electrical measurements, which are typically performed on a few select locations across the wafer, using process control monitors (PCMs) included on the wafer scribe lines.



(a) Adaptive method with two test flows [7].



(b) Adaptive test flow selection per process signature (this work)

Fig. 1: Wafer-level probe-test flow selection

test flow for that signature during probe testing of this wafer. We note that the complete test flow remains one of the possible choices, especially for outlier wafers, i.e., those e-test signatures have not been encountered in the past.

II. Preprocessing

Before we address the problem of deciding an appropriate test flow for a wafer, we discuss the initial elements which are required prior to such a decision. These two elements are: (i) identifying an appropriate subset of test groups which could potentially be applied to a wafer, and (ii) crafting a wafer signature from its e-test measurement vector. In the following sections, we provide details of these two components.

A. Reduced Test Flow Selection

A reduced test flow is a subset of the complete flow, wherein one or more test groups are eliminated. The first challenge that naturally arises is the selection of the test groups which should be omitted in a reduced flow, such that the attained test cost reduction does not compromise test quality beyond a target level of acceptable test escapes. Since the granularity of elimination is at the test group rather than at the test item level, it may be possible to exhaustively search the space of solutions. For example, in our experiments we dealt with a set of 10 test groups, thus exhaustively searching in the power-set of 2^{10} subsets of the complete test flow to find the optimum subset was feasible and chosen due to its simplicity.

B. Wafer Signature Extraction from E-tests

E-test data contain many types of parameters, mainly focusing on simple physical/electrical characteristics reflecting the position of a wafer in the process space. For some of these measurements, there is no physical connection or reason why they should be correlated with probe-test outcomes or the necessity thereof. Accordingly, to avoid spurious autocorrelations and to gain better insight from our e-test data, prior to crafting a wafer signature based on the e-tests we apply a dimensionality reduction algorithm to transform the data onto a lower count of dimensions. Specifically, we use the *t-Distributed Stochastic Neighbor Embedding (t-SNE)* technique [8] which is the state-of-the-art non-linear transformation approach and is widely used in many applications for unsupervised dimensionality reduction. In general, t-SNE embeds wafers with similar signatures close to each other on a 2-dimensional map.

III. Test Flow Optimization

In Section II, we described the process of generating all potential reduced test flows as well as extracting a process signature from the e-test data of a wafer. Now, our objective is to assign the most appropriate reduced test flow to each process signature, such that we maximize test cost reduction while retaining the test escape rate below a target Defective Parts Per Million (DPPM) level.

A. Bi-Flow Method

This technique was proposed in [7] as an adaptive solution to reduce probe-test cost during wafer-level testing. In this approach, the objective was to subject a subset of wafers to a reduced probe-test flow in which some test groups were eliminated from the complete test flow. In summary, this approach comprises the following steps:

- First, the e-test space is partitioned into k clusters using the k-means clustering method.
- Then, a reduced test flow is selected from the list of all possible reduced flows which are generated as described in Section II-A. Subsequently, all the training wafers in every cluster are tested by the selected reduced test flow and the total test escapes of the cluster, te_i, which is the sum of test escapes of all wafer in the cluster, is computed.
- Finally, the last step is to decide for each cluster of wafers (i.e., based on their process signatures) whether to perform the complete or the reduced test flow, such that the test cost reduction is maximized while the total test escape rate is kept below a given DPPM level. To solve this optimization problem, a binary Integer Linear Program (ILP) is formulated. The ILP solution assigns a label to each cluster, indicating the appropriate probe-test flow for the wafers in that cluster.
- The above-mentioned procedure is repeated for all possible reduced test flows and the best candidate is selected, based on the criterion of maximizing test cost reduction while meeting the required test quality.

For a new wafer, the distance of its e-test signature from the centers of the clusters is first computed and the wafer is assigned to the nearest cluster. If the decision for this cluster is to apply the reduced test flow, the wafer will undergo only the preselected subset of test groups, otherwise it will be tested by complete test flow. For the sake of simplicity, in the rest of the paper we refer to this approach as the *Bi-Flow* method.

B. Dynamic Test Flow Generation

We now proceed to elaborate on how to optimize the test flow per cluster. Our methodology consists of two steps: (i) finding the best reduced test flow for each cluster individually for any target DPPM level, and (ii) determining the maximum test escape rate of each cluster through an optimization algorithm. Below we provide details of these two steps.

1) Test Flow Generation per Cluster: Let us consider cluster C_i, which includes a set of wafers, and let us assume that we are interested in finding the best reduced test flow among all n candidates which are generated using exhaustive search. Let $\boldsymbol{TE_i} = [te_1, \cdots, te_n]$ and $\boldsymbol{TTR_i} = [ttr_1, \cdots, ttr_n]$ denote the test escape rate and test cost reduction vectors of the i-th cluster, where te_j and ttr_j denote the number of test escapes and the amount of test cost reduction when all wafers in this cluster are tested by the j-th reduced test flow. For any DPPM level in the range $[0, DPPM_t]$, where $DPPM_t$ is the target DPPM level, a reduced test flow is selected such that its test escape rate for cluster C_i is lower than the DPPM level, while maximizing the test cost reduction. At the end of this step, each cluster has associated with it a table with multiple rows and three columns. Each row corresponds to a specific DPPM level and the three columns correspond to the test escape rate, test cost reduction and index of selected test flow, respectively.

2) Optimization Algorithm: The second part of our proposed method is an optimization algorithm, which selects the best probe-test flow for all k clusters while meeting the required test quality. Let $\boldsymbol{TE} = [TE_1, \cdots, TE_k]^T$ and $\boldsymbol{TTR} = [TTR_1, \cdots, TTR_k]^T$ denote the test escape rate and test cost reduction matrices, where TE_i and TTR_i represent the test escape rate and test cost reduction vectors for the i-th cluster, and te_{ij} denotes the test escape rate for the i-th cluster for the j-th DPPM level. Our objective is to distribute the target DPPM level among k clusters so as to maximize test cost reduction. Looked at from a different angle, we need to determine the maximum acceptable test escape rate for each cluster. To do so, we formulate this problem as an integer linear program (ILP). An ILP consists of a set of variables, which can only assume integer values, a set of linear constraints on these variables, and a cost function which is to be maximized or minimized. In our problem, our constraint is on the total number of test escapes, and our cost function is to maximize test cost reduction. Our ILP is actually a binary (0-1) version, where the value of each integer variable can only be either 0 or 1.

IV. Experimental Results

In order to experimentally evaluate the effectiveness of the proposed methodology, we use actual production data from a 65nm analog/RF transceiver currently in high volume manufacturing (HVM) production by Texas Instruments. The dataset comes from 400 wafers, each of which contains approximately 2500 die. E-test is performed on 9 sites across the wafers, with 250 measurements obtained from each site. On each die, 380 parametric probe-test measurements are obtained, organized in 10 groups. The percentage by which each group contributes to the total test cost is also provided.

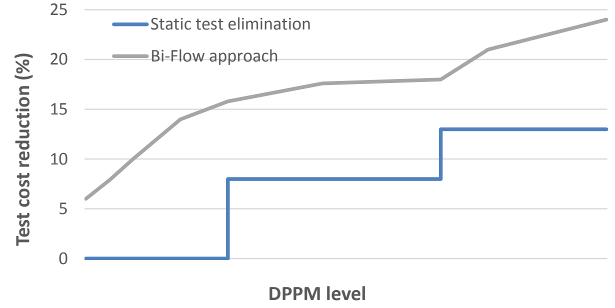


Fig. 2: Test cost reduction vs. test accuracy for static test elimination and Bi-Flow method for various DPPM levels

Using this dataset, our experiments seek to:

- Demonstrate that the effectiveness of the Bi-Flow method, which provides per wafer decision between a complete and a reduced test flow, is rather limited, thus a dynamic test flow generation with wafer-level granularity is required to optimize the test flow per process signature.
- Demonstrate that dynamic test flow generation per wafer based on e-test data can yield significant test cost reduction at realistic low DPPM levels.

A. Results of Bi-Flow Technique

Figure 2 demonstrates the test cost vs. test quality tradeoff for various DPPM levels. The two curves on this graph reflect solutions achievable by the static test elimination and an adaptive approach, which selects between the complete test flow and a single reduced test flow [7], respectively. Evidently, this Bi-Flow method outperforms static test elimination across the board. More importantly, it allows higher fidelity in the selection of a desirable point on this trade-off, starting from solutions with very low DPPM and small test cost reduction, and progressing at very fine-grained steps towards higher test cost reduction with higher test escape rates. Figure 3 depicts the outcome of the Bi-Flow approach in which the complete test flow assigned to a set of clusters (i.e., clusters with circle marker in red) and a reduced flow is selected for the remaining clusters, when target test escape rate is set to $DPPM_{min}$.

B. Dynamic Test Flow Optimization

Figure 4 depicts the outcome of the proposed dynamic test flow optimization technique when the target test escape rate is set to $DPPM_{min}$. In this graph, clusters with identical probe-test flow are represented by the same color; for example, clusters in blue such as C_4, require the complete test flow. On the bottom right of this graph, we present the optimized test flow code for clusters $C_1 - C_4$. In comparison to Figure 3, which shows the outcome of the Bi-Flow method for the same target DPPM level, the new method provides more flexibility for test cost savings.

The ability of the proposed dynamic test flow generation method to explore the trade-off between test cost reduction

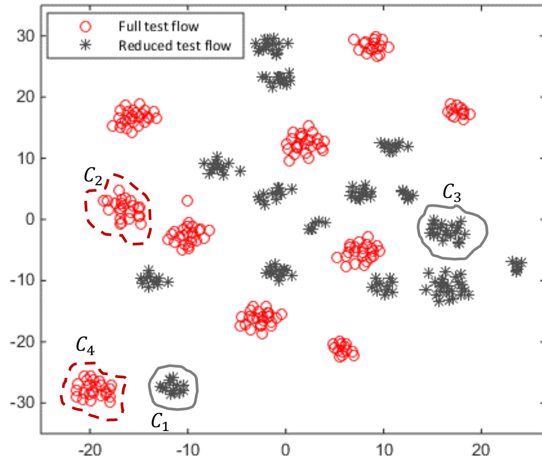


Fig. 3: Test flow assignment (either complete or reduced flow) for each cluster in the e-test space

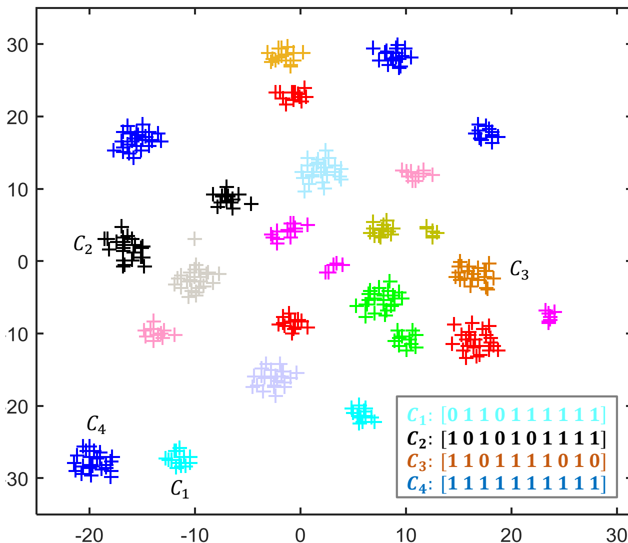


Fig. 4: Final assignment of the optimized probe-test flow code for each cluster (process signature)

and test quality, even in the region of very low DPPM, is demonstrated in Figure 5. The three curves on this graph reflect solutions achievable by static test elimination (blue curve), the Bi-Flow method (gray curve), and the proposed dynamic test flow generation (dotted black line) for various target DPPM levels. It is evident that the proposed dynamic test flow optimization approach significantly outperforms the other two approaches for any DPPM level. This is expected, since our dynamic approach successfully generates an optimized probe-test flow for each process signature.

V. Conclusion

Judicious harnessing of process variations in optimizing probe-test flow demonstrates great promise towards test cost reduction in analog/RF ICs. As we presented herein, each signature in the process space may require its own optimized test flow. The signature of a wafer can be obtained at early stage through e-test, reflecting how process variations have affected a given wafer. Deployment of the proposed method requires minimal test infrastructure support, yet is capable of identifying solutions with very low test escape rates, which is not possible through static test elimination. Experimental results using a large dataset of actual test measurements from a 65nm Texas Instruments RF transceiver confirmed the aptitude of the proposed method in effectively exploring the trade-off space between test quality and test cost.

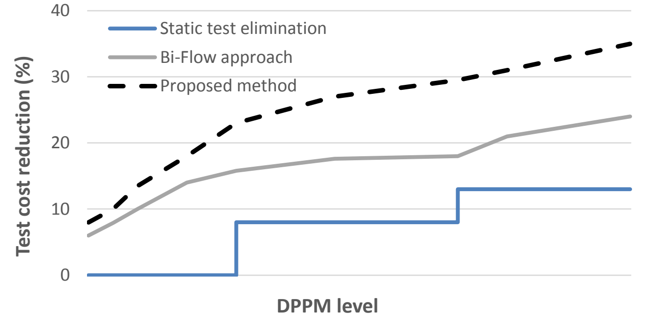


Fig. 5: Test cost reduction vs. test accuracy of three approaches for various DPPM levels

VI. Acknowledgment

This research has been partially supported by the Semiconductor Research Corporation (SRC) Task 1836.131.

References

[1] P. Drineas and Y. Makris, "Independent test sequence compaction through integer programming," in *IEEE International Conference on Computer-Aided Design*, 2003, pp. 380–386.

[2] H.-G. D. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris, "Non-RF to RF test correlation using learning machines: A case study," in *IEEE VLSI Test Symposium*, 2007, pp. 9–14.

[3] S. Biswas and R. D. Blanton, "Test compaction for mixed-signal circuits using pass-fail test data," in *IEEE VLSI Test Symposium*, 2008, pp. 299–308.

[4] S. Biswas, P. Li, R. Blanton, and L. T. Pileggi, "Specification test compaction for Analog circuits and MEMS," in *IEEE Design, Automation and Test in Europe*, 2005, pp. 164–169.

[5] H.-G. D. Stratigopoulos and Y. Makris, "Nonlinear decision boundaries for testing Analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 11, pp. 1760–1773, 2005.

[6] H.-G. D. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris, "RF specification test compaction using learning machines," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 18, no. 6, pp. 998–1002, 2010.

[7] A. Ahmadi, A. Nahar, B. Orr, M. Pas, and Y. Makris, "Wafer-level process variation-driven probe-test flow selection for test cost reduction in Analog/RF ICs," in *IEEE VLSI Test Symposium*, 2016, pp. 1–6.

[8] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 2579-2605, pp. 85, 2008.