# Harnessing Process Variations for Optimizing Wafer-level Probe-Test Flow

Ali Ahmadi*, Constantinos Xanthopoulos*, Amit Nahar†, Bob Orr†, Michael Pas† and Yiorgos Makris*

*Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080
†Texas Instruments Inc., 12500 TI Boulevard, MS 8741, Dallas, TX 75243

***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 widely used test cost reduction technique is to customize the list of tests according to their contributions to the overall test quality [1]–[3]. Specifically, the effectiveness of each test is monitored and the ones with little contribution to the overall test effectiveness are dropped from the test list. Such decisions are usually static and are easy to implement on the ATE by exclusion of the relevant portion of the test program. However, these methods have limited capability 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. To enhance this idea, several methods have been proposed that leverage statistical correlation between the dropped and retained tests and predict the outcome of the former from the latter [2], [4]–[6]. While additional ATE support or external resources are required to run the statistical models on-the-fly during test, these methods have demonstrated marked improvement in test quality over static test elimination technique. Still, the decision models remain static or only infrequently retrained to account for major events which can change the statistical profile of the production.

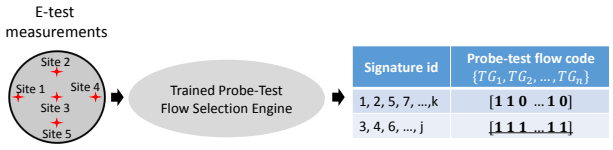
In order to address this issue, dynamic test adaptation is introduced in [7] to re-optimize the test list on a per-lot basis, using data collected from the first few wafers, on which the complete flow is applied. To take adaptation a step further, authors in [8] proposed a method that identifies, through sampling and clustering, wafer regions which have been affected similarly by process variations, and customizes the test list and test order to each such region. While this method was demonstrated in the context of final test, it could be readily applied at probe-test as well. However, this technique would complicate test floor logistics, as it would require two passes (for sampling and testing) and ATE support for applying different test programs to each region of the wafer. In fact, any solution that needs adaptation at a finer granularity than the wafer-level would require such support, which is often missing or cumbersome to implement in ATE platforms.

In [9], 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.

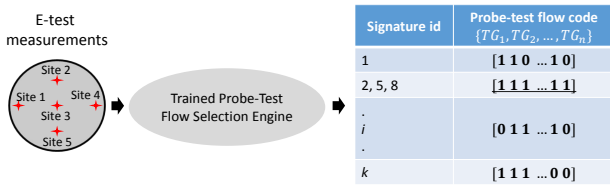
Figure 1 (a) depicts this adaptive method: a statistically trained entity examines the e-test data of a wafer and, depending on the extracted signature, it selects the appropriate

[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.


978-1-4673-8773-6/16/$31.00 ©2016 IEEE




(a) Adaptive method with two test flows [9]




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

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

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 [9], 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 [9], 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, extracts its process signature, and accordingly selects the most appropriate 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 whose e-test signatures have not been encountered in the past.

Accordingly, in this work we seek to develop a methodology which optimizes probe-test flow selection based on the following principles, in order to be readily deployable with minimal test operations support:

- The granularity at which test elimination decisions are made is at the test group level. The underlying assumption here is that the bulk of the cost incurred by a test group is related to switching into the appropriate test configuration. Accordingly, the incremental savings of eliminating a few measurements within a group are negligible.
- The granularity of the adaptation decision is at the wafer level, i.e., all die on a wafer are subjected to the same test flow, either the complete set of test groups or a subset thereof.
- Test has to be performed in one pass. In other words, solutions which first apply a reduced test flow and subsequently apply selectively more test items to die for which the decision confidence is low, such as the two-tier test method in [10], are not within scope.
- The decision has to be driven by a signature which reflects how process variations have affected a particular wafer. This is justified by historical evidence documenting that the necessity of a test group is strongly correlated with the operating point of the fabrication process.
- The decision has to be available prior to insertion of the wafer in the probe station and cannot be informed by measurements taken at probe. Inevitably, this leaves e-test as the only source available for capturing the impact of process variations on a particular wafer.
- The Automatic Test Equipment (ATE) supports multiple test flows, where test groups can be dynamically included or excluded based on an input provided before test commences for a wafer.

The remainder of this paper is organized as follows. In Section II, we discuss the steps required for evaluating potential reduced test flows and for extracting a wafer-level signature from e-test data. Then, in Section III, we describe the details of the proposed adaptive test flow method. Experimental results demonstrating the effectiveness of the proposed method on a large industrial dataset are presented in Section IV and conclusions are drawn in Section V.

## 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. In case of a large number of test groups, however, this approach

will not scale. In this case, heuristic search methods can be employed for effectively searching this space. The use of Genetic Algorithms has been popular in the literature and very successful when applied to this task [6], hence we can readily adopt it when exhaustive consideration is infeasible.

For each reduced flow, $j$, we consider the associated cost and the number of test escapes when this reduced flow is applied to all wafers in our training set, and we assign a fitness value defined as:

$$index_j = \frac{t_A - t_{B_j}}{t_A} * pctg_{B_j} \quad (1)$$

where $t_{B_j}$ denotes the test cost of the $j$-th reduced flow, $t_A$ denotes test cost of the complete test flow and $pctg_{B_j}$ represents the percentage of wafers that can be tested using the $j$-th reduced test flow, while keeping the total number of test escapes remains below a target DPPM level.

### 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 [11] which is the state-of-the-art non-linear transformation approach and which 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.

In Figure 2, we provide an example where we project a number of wafers to a 2-dimensional space after applying the t-SNE algorithm. The various markers used to represent each point indicate different test escape rates[2] when a randomly selected reduced test flow is applied to all wafers. Wafers with the same marker exhibit a similar level of test escapes. Two key observations can be made using this figure:

1) Projection of wafers on the e-test space is discontinuous, with most wafers being part of small clusters in this 2-dimensional space. This reflects the fact that the process jumps between a finite number of points.
2) Wafers within each cluster, i.e., with similar e-test signature, do not necessarily exhibit the same test escape rate. This implies that the correlation between device specifications and e-test parameters is complex and there is no simple boundary to separate wafers with high test escapes from wafers with low or zero test escapes. A more elaborate approach is, consequently, required for mapping e-test signatures to the appropriate test flow.

[2]The exact values of B-G are not important for this example.

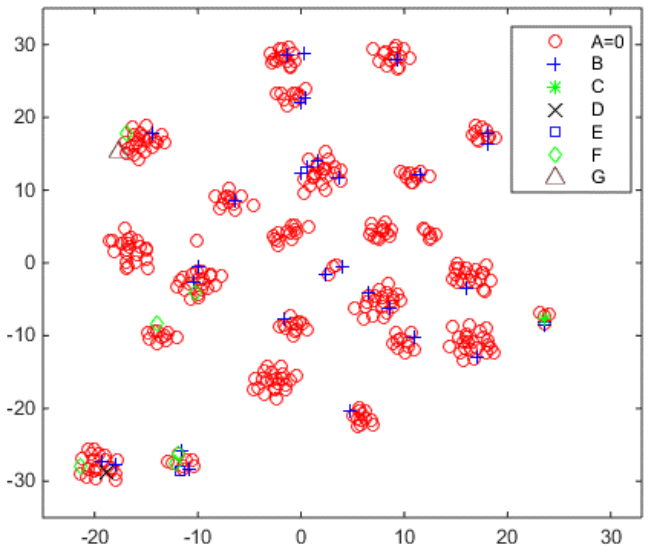



Fig. 2: Projection of e-test data onto two dimensions using the t-SNE algorithm

## 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 [9] 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 of Figure 2 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 wafers 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 the complete test flow. For the sake of simplicity, in the rest of the paper we refer to this approach as the *Bi-Flow* method.

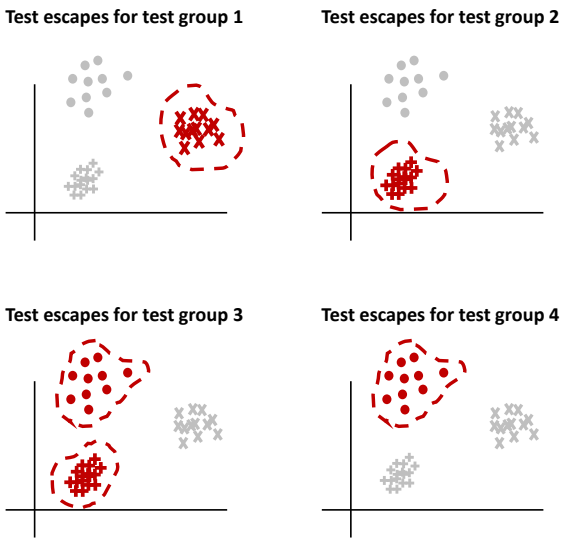
### B. Limits of The Bi-Flow Method

This method is simple and can be implemented easily. It is also very effective in finding the best reduced test flow and assigning either complete or reduced test flow to each cluster. However, its major limitation is the fact that the same reduced test flow is chosen for all process signatures (clusters) that will not undergo complete test. This simplifies test operations, as only two test flows are maintained, but it is also sub-optimal, since different clusters exhibit dissimilar failure patterns when a test group is removed from the test flow. Indeed, choosing a different test flow for each cluster holds promise for significantly higher test cost reduction.

To demonstrate this limitation, we select three clusters from Figure 2 and we compute the test escape rate of each cluster when a test group is removed from the test flow. Figure 3(a) demonstrates the test escape rate due to elimination of test groups 1-4 for these three clusters. Clusters in red color and enclosed in a red boundary reflect zero test escapes while gray color represents clusters with a non-zero test escape rate. In the table of Figure 3 (b), the test flow code of each cluster is represented using a binary vector where inclusion/exclusion of a test group is indicated by value 1/0 respectively. The third column of the table shows the two test flow codes which are generated by the Bi-Flow method [9], while the forth column contains the optimum test flow code if chosen individually per each cluster. As may be observed, the impact of skipping a test group is not identical for all clusters. The figure corroborates our conjecture that each cluster requires individualized test flow optimization. Therefore, a dynamic approach is required to generate the most appropriate probe-test flow per process signature, in order to maximize test cost reduction. In the next section, we introduce a methodology which addresses this limitation.

### C. 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.




(a) Test escape rate for selected clusters when test groups 1-4 were individually removed from the test flow

| Cluster symbol | Cluster id | Bi-Flow approach probe-test flow code $\{TG_1, TG_2, TG_3, TG_4\}$ | This work probe-test flow code $\{TG_1, TG_2, TG_3, TG_4\}$ |
|---|---|---|---|
| ● | 1 | [1 1 0 1] | [1 1 0 0] |
| ✚ | 2 | [1 1 0 1] | [1 0 0 1] |
| ✘ | *3* | [1 1 1 1] | [0 1 1 1] |

(b) Comparison of probe-test flow code corresponding to part (a) generated by technique [9] vs. optimum probe-test code for each cluster

Fig. 3: Limitations of Bi-Flow approach

*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 $\boldsymbol{TE} = [\boldsymbol{TE_1}, \cdots, \boldsymbol{TE_k}]^T$ and $\boldsymbol{TTR} = [\boldsymbol{TTR_1}, \cdots, \boldsymbol{TTR_k}]^T$ denote the test escape rate and test cost reduction matrices, where $\boldsymbol{TE_i}$ and $\boldsymbol{TTR_i}$ 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. Specifically, in our ILP, the variable $\alpha_{ij} = 1$ is used to indicate that the maximum acceptable test escapes for $i$-th cluster is $te_{ij}$, and therefore the test flow with index $j$ is selected for this cluster. Then, our binary ILP is defined as follows:

$$\begin{aligned} &\text{Maximize} \quad \sum_{i=1}^{k}\sum_{j=1}^{m} \alpha_{ij}.ttr_{ij} \\ &\text{subject to} \quad \sum_{i=1}^{k}\sum_{j=1}^{m} \alpha_{ij}.te_{ij} \leq DPPM_t \\ &\qquad\qquad \sum_{j=1}^{m} \alpha_{ij} = 1 \\ &\alpha_{ij} \in \{0,1\},\ i = 1,\ldots,k \ and \ j = 1,\ldots,DPPM_t \end{aligned} \tag{2}$$

The constraint $\sum_{j=1}^{m} \alpha_{ij} = 1$ is used to select only one test flow for each cluster.

An additional provision is also incorporated in the proposed methodology, in order to adapt to shifts in the process, which may result in previously unseen wafer signatures in the transformed e-test space. Specifically, as shown in Figure 4, for clusters where the ILP selects any probe-test flow other than complete test flow, we establish a boundary around the e-test signatures that belong to the cluster. 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 any reduced test flow, we perform one more check: if its signature is inside the boundary of that cluster, we follow the recommendation. Otherwise, we assume that despite being nearest to this cluster, the wafer is sufficiently different and we send it to the complete test flow. Based on this information, we periodically enhance the set of clusters and rerun the optimization algorithm to better track the process.

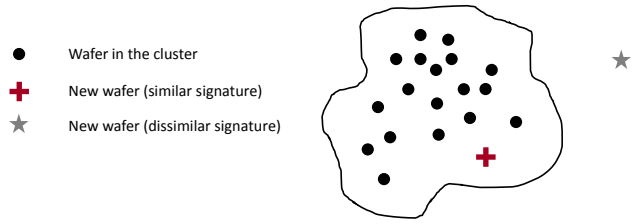



Fig. 4: Tracking process shifts: signatures of wafers belonging to cluster are enclosed by a boundary. New cluster members with signatures within the boundary are considered equivalent, and new members with signatures outside the boundary are considered outliers

## 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[3]. 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. The objective of our method is to find a subset of the 10 test groups as an optimized reduced test flow for each process signature and to train an intelligent system which will use the e-test measurements to select which test flow a wafer should undergo. In our experiments, we use 5-fold cross validation. Specifically, our data set is divided into 5 folds, where 4 folds are used for training and the remaining fold for validation. The procedure is repeated such that all folds are left out once as a validation set and, in the end, we report the average test escapes and test cost reduction across the five iterations. Using this dataset, our experiments seek to:

- Confirm that static test group elimination does not have the agility to support reduced test flows while maintaining a test escape rate in the very low DPPM region, therefore adaptivity is required to provide per wafer decision.
- 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. Limits of Static Test Elimination

Figure 5 reflects the number of defective die per million which are uniquely detected by each of the 10 test groups. In

[3]Details regarding the device and exact test escape numbers and DPPM levels may not be released due to an NDA under which this data has been provided to us.

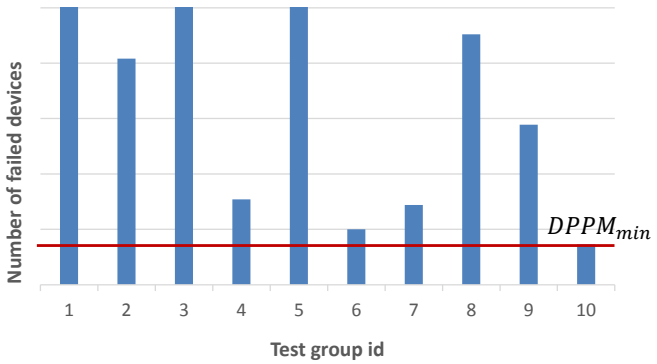



Fig. 5: Defective die per million which would escape detection if each of the 10 test groups were to be statically eliminated from the probe-test flow

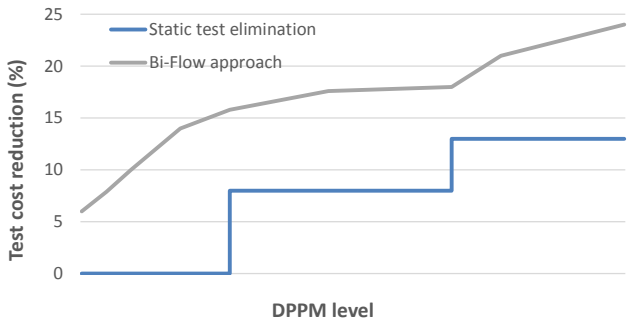



Fig. 6: Test cost reduction vs. test accuracy for static test elimination and Bi-Flow method for various DPPM levels

other words, this is the number of devices which would escape detection if each of these 10 test groups were to be statically eliminated from the probe-test flow. While we cannot reveal the exact number for $DPPM_{min}$, its order of magnitude is in the several tens. Accordingly, static test elimination cannot be used for test cost reduction when test quality expectations are set below this level. Therefore, exploration of the test cost vs. test quality trade-off in the sub-$DPPM_{min}$ realm requires adaptive test flow selection per wafer.

Figure 6 demonstrates the test cost vs. test quality trade-off for various DPPM levels. The two curves on this graph reflect solutions achievable by the static test elimination and Bi-Flow approach, which selects between the complete test flow and a single reduced test flow [9], respectively. Evidently, the 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.

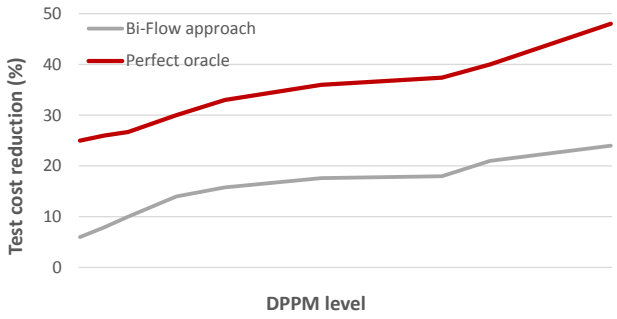



Fig. 7: Achieved test cost reduction using the Bi-Flow method vs. maximum possible test cost reduction for various DPPM levels

### *B. Limits of Bi-Flow Technique*

In this section, we examine the effectiveness of the Bi-Flow approach, which subjects a wafer either to a complete or a reduced test flow. To do so, in Figure 7, we compare its test cost reduction to the upper bound achievable when an oracle that can perfectly select the appropriate test flow (i.e., the complete or the single reduced test flow) for each wafer is used, for various target DPPM levels. As may be seen from the gap between the two curves, this approach leaves significant potential for test cost reduction on the table. To gain better insight, Figure 8 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 the target test escape rate is set to $DPPM_{min}$. The main disadvantage of this approach is the fact that the reduced test flow is generated collectively for all clusters rather than individually per cluster, based on the process signatures of wafers in a cluster.

To demonstrate the unique characteristics and test flow needs of each cluster, in Figure 8 we select clusters $C_1 - C_4$. Then, we compute the test escapes of each cluster when a test group is eliminated from the test flow (either complete or reduced test flow) which is assigned to that cluster. Figure 9 shows the number of test escapes for these clusters when test groups 1, 2, 8 and 10 were removed from the test flow individually. As it can be seen, the test escape profile of each cluster varies drastically. More specifically, based on this information, wafers in cluster $C_1$ can skip test group 1, while those in cluster $C_2$ require test groups 1 and 2, yet test groups 8 and 10 can be eliminated from their test flow. This experiment confirms that a dynamically optimized test flow generation per cluster is needed to maximize test cost reduction for any target DPPM level.

### *C. Dynamic Test Flow Optimization*

Figure 10 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

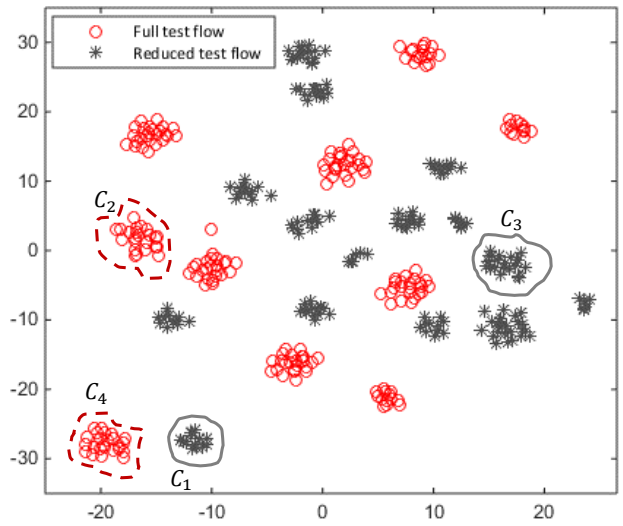



Fig. 8: Test flow assignment (either complete or reduced flow) for each cluster in the e-test space

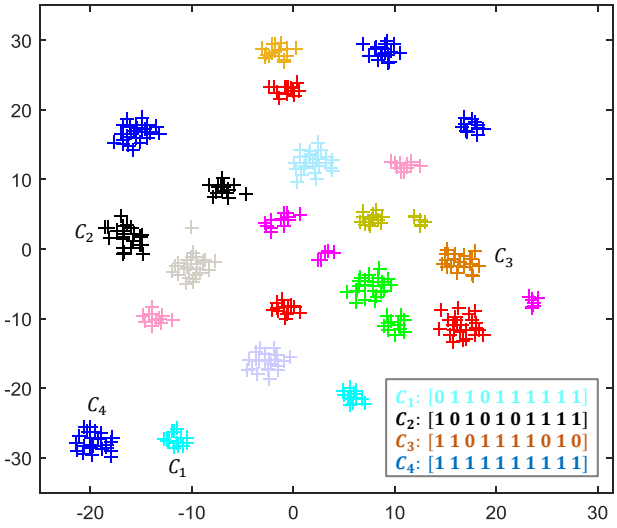



Fig. 10: Final assignment of the optimized probe-test flow code for each cluster (process signature)

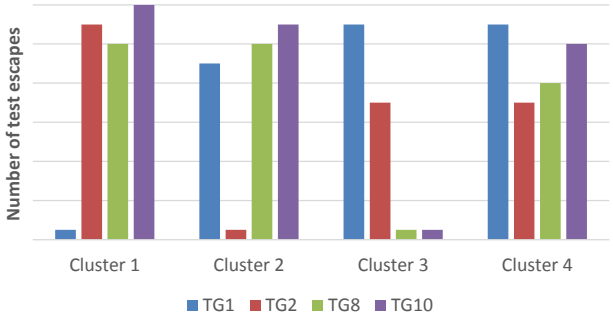



Fig. 9: Number of test escapes for selected clusters when each of the four test groups are eliminated from the probe-test flow of the cluster

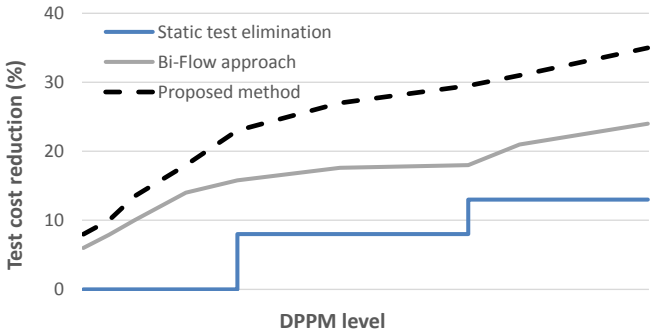



Fig. 11: Test cost reduction vs. test accuracy of three approaches for various DPPM levels

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 8, 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 and test quality, even in the region of very low DPPM, is demonstrated in Figure 11. 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.

Finally, to gain better insight as to how well our proposed method works, in Figure 12 we compare its test cost reduction to the upper bound achievable when an oracle is used, for various target DPPM levels. It should be noted that the maximum achievable test cost reduction, which is demonstrated in Figure 7 by the red curve, is the upper bound when we are allowed to have only two test flows (complete or reduced). However, in the new scenario where we can handle several test flows, the upper bound would be achieved by an oracle which can perfectly select the best test flow for each wafer. In Figure 12, the new upper bound is represented by the dotted line which is above all other curves.

We note that the gap between the proposed method and the upper bound shrinks as the targeted DPPM increases.

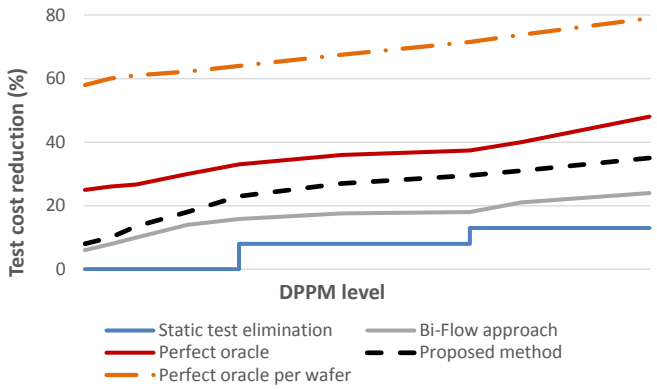



Fig. 12: Test cost reduction vs. test accuracy for three approaches and maximum possible test cost reduction for various DPPM levels

This is explained by the fact that, at very low DPPM levels, incorrectly channeling a wafer to a reduced instead of a complete flow can be detrimental and very difficult to recover from. In other words, very low DPPM leaves little room for error, hence the proposed method acts conservatively, selecting very few e-test signatures for reduced test flows and, thereby, limiting the achieved test cost reduction. We also note that this gap indicates what is still left on the table as possible further test cost reduction, which more advanced methods and better statistics may be able to potentially achieve. Therefore, our future research efforts will be directed towards further reducing this gap.

## 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 an 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.

## 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] S. Benner and O. Boroffice, "Optimal production test times through adaptive test programming," in *IEEE International Test Conference*, 2001, pp. 908–915.

[8] E. Yilmaz, S. Ozev, and K. M. Butler, "Efficient process shift detection and test realignment," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 12, pp. 1934–1942, 2013.

[9] 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.

[10] H.-G. D. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine-learning-based Analog/RF testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339–351, 2008.

[11] 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.