Confidence Estimation in Non-RF to RF Correlation-Based Specification Test Compaction

Nathan Kupp*, Petros Drineas†, Mustapha Slamani‡ and Yiorgos Makris*

*Department of Electrical Engineering, Yale University, 10 Hillhouse Avenue, New Haven, CT 06520, USA
†Department of Computer Science, Rensselaer Polytechnic Institute, 110 8th Street, NY 12180, USA
‡IBM, Wireless Test Development Group, 1000 River Street 863G, Essex Junction, VT 05452, USA

Abstract

Several existing methodologies have leveraged the correlation between the non-RF and the RF performances of a circuit in order to predict the latter from the former and, thus, reduce test cost. While this form of specification test compaction eliminates the need for expensive RF measurements, it also comes at the cost of reduced test accuracy, since the retained non-RF measurements and pertinent correlation models do not always suffice for adequately predicting the omitted RF measurements. To alleviate this problem, we develop a methodology that estimates the confidence in the obtained test outcome. Subsequently, devices for which this confidence is insufficient are retested through the complete specification test suite. As we demonstrate on production test data from a zero-IF down-converter fabricated at IBM, the proposed method outperforms previous defect filtering and guard banding methods and enables a more efficient exploration of the tradeoff between test accuracy and number of retested devices.

1. Introduction

The current industry practice in testing analog/RF integrated circuits relies on explicitly measuring all the performances of each fabricated device and comparing them to the specification limits. However, as the costs associated with this specification testing approach have been continuously escalating, a great incentive to reduce this burden by eliminating potentially redundant measurements has surfaced. This holds particularly true for RF circuits because the cost of pertinent Automatic Test Equipment (ATE) is significantly higher than that of their low-frequency mixed-signal counterparts. Such discrepancy has resulted in an intensified interest towards developing methods for accurately testing RF devices without explicitly measuring their RF performances. The underlying principle is to approximate these RF performances through correlation models based solely on non-RF performances (i.e. digital, DC, low-frequency), which can be explicitly measured through less expensive ATE. In essence, these non-RF to RF performance correlation models enable a form of specification test compaction and, ultimately, result in significant test cost reduction.

The framework of non-RF to RF correlation-based specification test compaction is depicted in Fig. 1. The learning phase relies on a training set of m devices, on which both the non-RF performances, NRF_i, $i = 1 \ldots s$, and the RF performances, RF_j, $j = 1 \ldots t$, are explicitly measured. Based on this information, statistical correlation models, $PRF_j = f(NRF_1, \ldots, NRF_s)$, $j = 1 \ldots t$, are learned, predicting each RF performance as a function of the non-RF performances of a device. Subsequently, for every new device in production, only the non-RF performances are explicitly measured, while the RF performances are predicted through the learned correlation models. A pass/fail decision is made by comparing the measured non-RF performances and the predicted RF performances to their specifications. Thus, an RF ATE is needed only for characterizing the small number of devices in the training set but is not necessary during production testing.

Unfortunately, while correlation-based specification test compaction promises great test cost reduction, the incurred test error prevents it from reaching the level of Defective Parts per Million parts shipped (DPM) typically sought by industry. Even when very elaborate models are used to learn the correlation between non-RF and RF performances, such error is bound to exist. Indeed, partly due to the limited size of the training set, which may not reflect accurately the statistics of the entire production, and partly due to the fact that the non-RF performances may not reflect the complete information spectrum of the RF performances, elimination of this test error is very unlikely. Instead, viability of this approach hinges upon accepting the fact that the performances predicted through the correlation models will not always yield correct pass/fail decisions and focusing on pinpointing the mispredicted devices.

To this end, providing a confidence level indication along with the predicted RF performances could go a long way. Devices for which this confidence level is low can then be identified and discarded at the expense of possible yield loss. Alternatively, these devices can be retested through the

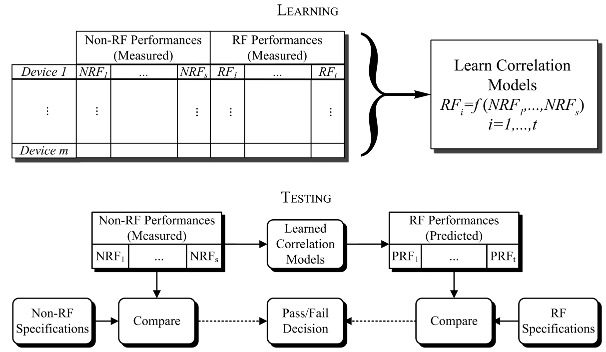


Fig. 1. Non-RF to RF Correlation-Based Testing

1530-1877/08 $25.00 © 2008 IEEE
DOI 10.1109/ETS.2008.31

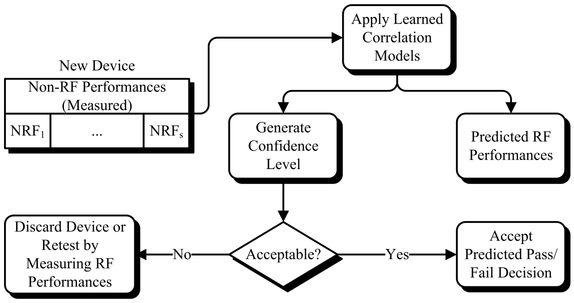


Fig. 2. Retesting when prediction confidence is low

complete specification test suite, as shown in the two-tier test approach of Fig. 2, at the expense of additional test cost. While the second tier requires additional handlers and RF ATE, if only a small fraction of devices goes to the second tier the overall cost savings can still be significant. Thus, successful deployment of non-RF to RF correlation-based specification test compaction calls not only for accurate correlation models but also for accurate assessment of the confidence in the corresponding test decisions, in order to explore effectively the trade-off between test error and test cost.

Two such approaches, generically termed *defect filtering* and *guard banding*, have been previously described in [1], [2], [3]. These, however, were not in the context of correlation-based specification test compaction but, rather, in the related field of alternate test [4], [5]. The key difference is that the former uses a low-cost subset of non-RF circuit performances to predict the dropped RF performances, while the latter relies on low-cost alternate measurements which constitute the response of the circuit to a carefully crafted and optimized stimulus. Thus, these alternate measurements may encompass more comprehensively the spectrum of information necessary to predict the circuit performances. Nevertheless, the accuracy boosting and trade-off exploration methods used therein are readily applicable to the non-RF to RF correlation-based specification test compaction problem and we examine them in detail.

In this paper, we introduce a novel *confidence estimation* method for deciding whether the pass/fail prediction yielded by the correlation models is sufficiently accurate or not. The proposed methods employs an additional learning phase, wherein a Support Vector Machine (SVM) [6] is trained to separate the hyperspace of the non-RF measurements into regions that are trusted or untrusted, with regards to the pass/fail decisions of the correlation models. The key advantage of the proposed confidence estimation method is that the outlined regions are created through highly non-linear separation hypersurfaces, rather than the hyperrectangular boundaries employed by defect filtering and guard banding. Furthermore, these regions are learned rather than set a priori based on the distribution of the training set. Thus, the proposed confidence estimation method promises lower test error and fewer retested devices.

The remainder of this paper is organized as follows. In section 2, we briefly discuss related efforts in analog specification test compaction. Then, in section 3, we describe in detail the aforementioned defect filtering and guard banding approaches, as well as the confidence estimation method proposed herein. Finally, in section 4, we provide experimental results comparing the three methods based on production test data from a zero-IF down-converter fabricated at IBM.

2. Related Work

Various other analog specification test compaction methods have been developed in the past. The linear error-mechanism model algorithm (LEMMA) [7] and various extensions thereof, aim to predict the complete vector of performance measurements by carrying out only a subset of cardinality which depends on the permitted measurement cost and the maximum tolerable prediction error. The selection process is performed through QR factorization [8] and minimizes the prediction variance. The effectiveness of the LEMMA method is limited by the use of a linear model to predict the behavior of a non-linear system, as well as the need for error mechanism models that are difficult to specify for complex circuits. In [9], a fault-driven test selection approach is proposed. Performance measurements are gradually added until a desired fault coverage level is reached. The disadvantage of this approach is its dependence on fault models, which have not been widely accepted in the analog/RF domain. In [10], the compaction problem is viewed as a binary pass/fail classification problem and an SVM is trained to separate the passing from the failing devices in the hyperspace of a subset of performances, eliminating one dimension at a time. In practice it is advantageous to consider subsets of performances, since combinations of performances can provide significant information which is not individually available in any of the performances. To this end, a genetic feature selection algorithm along with an Ontogenic Neural Network is described within the context of non-RF to RF specification test compaction in [11]. A guard-banding and a two-tier test method applicable to the latter is discussed in [12].

3. Accuracy Boosting and Trade-off Exploration

In this section, we discuss the previously proposed defect filtering and guard banding methods [1], [2], [3] and we introduce the proposed SVM-based confidence estimation method.

3.1. Prior Art: Defect Filtering

Defect filtering [1], [3] builds upon the well-known fact that accurate correlation models can only be learned through elements that belong to a distribution [13]. While passing devices and marginally failing devices are typically considered to belong to a distribution, grossly defective devices are not and, therefore, should be filtered out. In other words, such devices should not be used during the learning process and the learned correlations should not be used to predict the performances of such devices. To ensure this, defect filtering divides the devices into two sets, depending on whether they are considered to belong to the distribution or not.

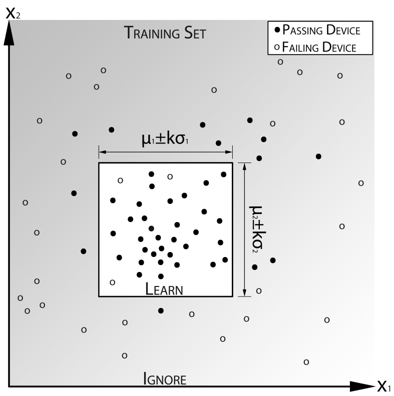


Fig. 3. Learning phase

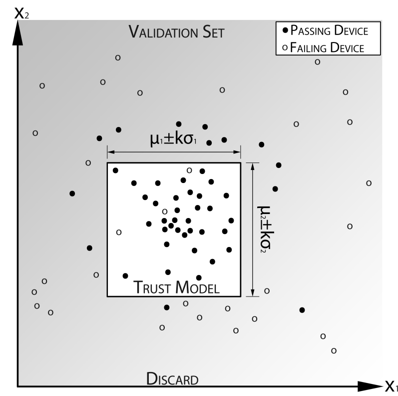


Fig. 4. Defect filtering - Testing phase

More specifically, in the context of specification test compaction, let us assume that we are trying to establish correlation models for predicting RF performances based on non-RF performances (predictor variables) X_i, $i = 1 \dots n$. Let us also assume that, in the training set, the mean of these n predictor variables is $\mu = \{\mu_1, \mu_2, \dots, \mu_n\}$ and the standard deviation is $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$. Then, a defect filter is defined as a hyperrectangle in the space of the training set:

$$H_k = \{\mu_1 \pm k \cdot \sigma_1, \mu_2 \pm k \cdot \sigma_2, \dots, \mu_n \pm k \cdot \sigma_n\}$$

where k is a positive real number. We will refer to this hyperrectangle as a k-filter.

The utilization of the k-filter during the learning phase is conceptually demonstrated for $n = 2$ in Fig. 3. During this phase, devices in the training set whose predictor variable vector falls outside the k-filter are ignored. Similarly, as shown in Fig. 4, during the testing phase devices whose predictor variable vector falls outside the k filter are rejected or retested through a second tier of complete specification testing. In essence, correlation models are only trusted when used for predicting the RF performances of devices within the k-filter.

Evidently, the choice of k is crucial since it affects both the number of retested devices and the accuracy of the learned correlation models. A strict k-filter may exclude many good devices during testing, resulting in high yield loss if they are rejected or high test cost if they are retested. A lenient k-filter may allow many devices that do not belong to the distribution to affect the accuracy of the correlation models during training, resulting in high test error. In essence, the choice of k facilitates exploration of the trade-off between test accuracy and test cost.

Note that assessment of candidate k-filters should not be performed using training set devices. Instead, a second set of devices, called the *hold-out set*, is used to drive the choice of k. The chosen k is then used along with the learned correlation models to calculate the figures of merit of defect filtering in a set of previously unseen devices, i.e. the *validation set*.

3.2. Prior Art: Guard Banding

While defect filtering offers a good first step towards boosting the accuracy of the correlation models, it suffers from two inherent limitations. *First*, due to the continuous nature of the predictor variables, the limited training set size, and the fact that non-RF performances may not reflect the complete information spectrum of RF performances, it is highly unlikely that correlation models perfectly separating the two populations of passing and failing devices will be learned. Therefore, despite the k-filtering approach, a test error is bound to exist, translating into yield loss and/or test escapes. *Second*, depending on the chosen course of action for devices outside the k-filter, defect filtering may incur unnecessary yield loss or test cost. Specifically, if all devices outside the k-filter are discarded, then passing devices often found just outside the k-filter will be thrown away. Similarly, if all devices outside the k-filter are retested, grossly defective devices typically found far away from the k-filter will waste test resources.

To alleviate this problem, guard banding [1], [2] complements the k-filter by an l-filter, $k \leq l$, thus dividing devices into three sets: (1) devices falling outside the l-filter are considered grossly defective and are discarded; (2) devices falling in between the k-filter and the l-filter are retested; (3) devices falling inside the k-filter are considered part of the distribution and the outcome of the correlation models is trusted for deciding whether they pass or fail. As in defect filtering, the correlation models are learned only from devices in the training set which fall inside the k-filter (see Fig. 3). The utilization of the k-filter and l-filter guard bands during the testing phase is conceptually demonstrated for $n = 2$ in Fig. 5.

The choice of k and l is instrumental in exploring the trade-off between retested devices and test error. As in defect filtering, selection of candidate k- and l-filters is performed in the hold-out set. In other words, k-filters are established and correlation models are learned in the training set, k and l are chosen by assessing the effectiveness of the k- and l-filters and the learned models in the hold-out set, and the chosen k and l are then used along with the correlation models to calculate the figures of merit of guard banding in the validation set.

3.3. Proposed Method: Confidence Estimation

Both defect filtering and guard banding rely on the mean and the standard deviation of the devices in the training set

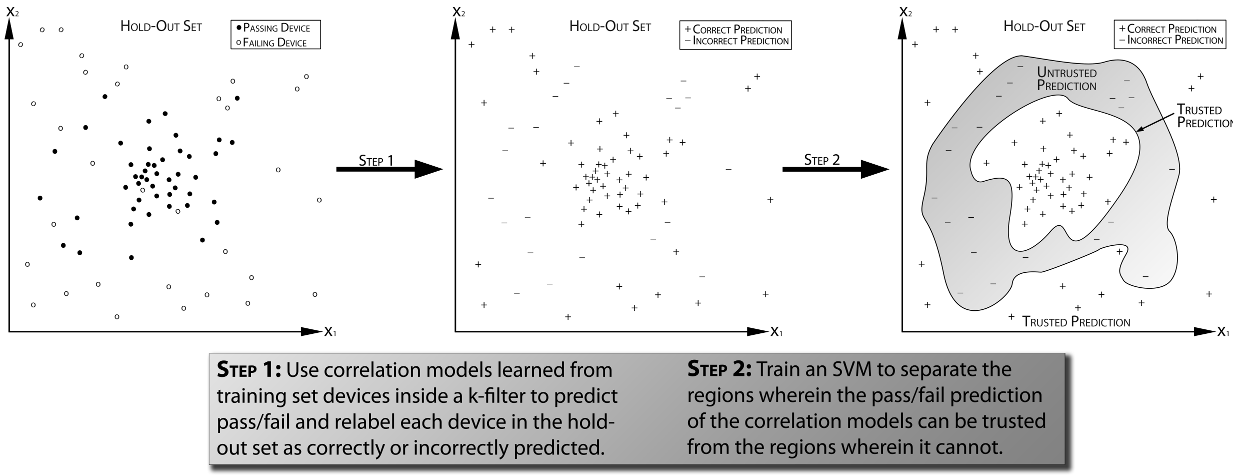


Fig. 6. Proposed confidence estimation method

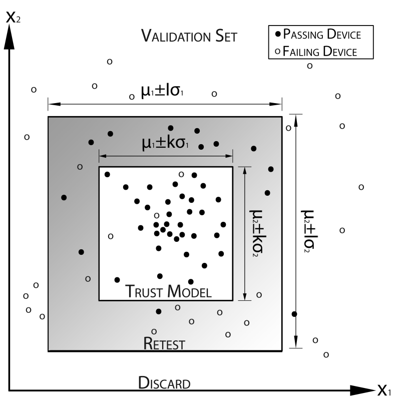


Fig. 5. Guard banding - Testing phase

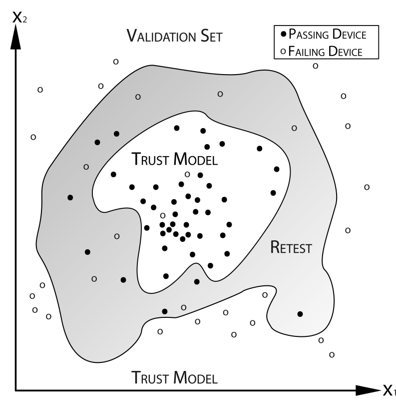
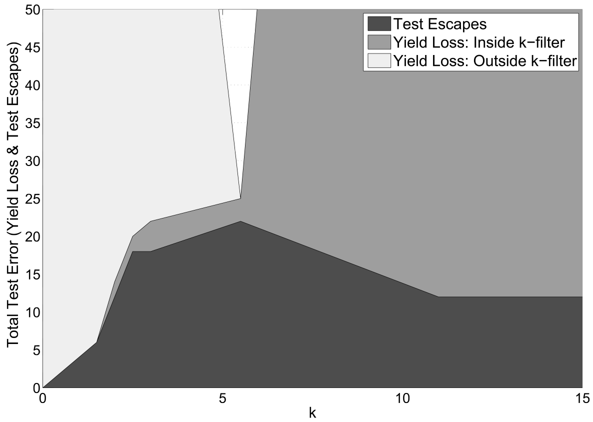


Fig. 7. Confidence estimation - Testing phase

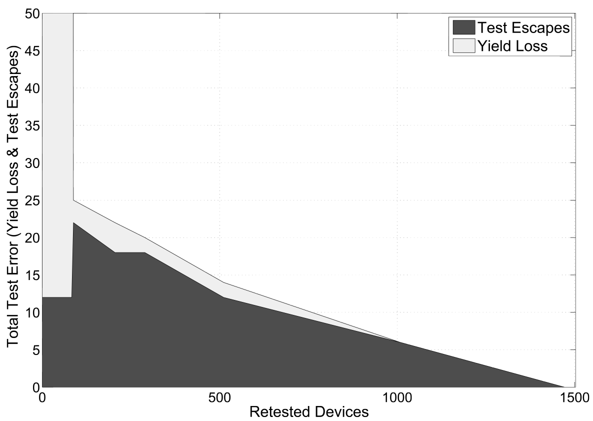
to establish the three regions wherein correlation models are trusted, devices are retested, or devices are discarded, respectively. These regions, however, are rather coarsely outlined through hyperrectangles, whereas the actual region in which the correlation models can yield a trusted prediction is likely to be more complex. Thus, a more refined division of the aforementioned regions holds promise for further improving the prediction accuracy of the correlation models and reducing the number of retested devices.

To this end, we propose a confidence estimation method which uses an SVM [6] to replace the coarse hyperrectangles with a detailed non-linear hypersurface. As previously, correlation models are initially learned from training set devices within a k-filter (see Fig. 3). Then, the 2-step procedure shown for $n = 2$ in Fig. 6 is applied to the devices in the hold-out set. In the first step, the learned correlation models are used to make pass/fail predictions and the devices in the hold-out set are relabeled as correctly or incorrectly predicted. In the second step, an SVM is trained to learn the boundary partitioning the predicted performance space into two subspaces: the area wherein correct predictions occurred (trusted), and the area wherein incorrect predictions occurred (untrusted). The choice of k is, again, crucial in establishing accurate separation boundaries via the SVM. Since the SVM is trained using the devices in the hold-out set, k has to be picked by examining the SVM performance on another set (e.g. the training set). The utilization of the SVM during the testing phase is conceptually demonstrated for $n = 2$ in Fig. 7. The pass/fail prediction of the correlation models is accepted only for devices with predictor variable vectors that the trained SVM classifies as trusted, while the rest of the devices are retested.

We note that the trusted area outlined in Fig. 6 and Fig. 7 is a simplification of the actual bounding done by an SVM, as the latter transforms the predictor variable hyperspace into a new hyperspace, wherein it learns the boundaries. This transformation (a.k.a. kernel) is what enables the SVM to draw highly non-linear surfaces in the original predictor variable hyperspace. We also note that the SVM marks the area of outliers (i.e. grossly defective devices) as “trusted” even though the correlation models perform poorly in estimating the performances of such devices. Indeed, while the performance prediction itself is inaccurate, it is still far off from the acceptable specification range and, thus, sufficient to ensure correct classification of these devices as failing. In this sense, the trusted/untrusted separation boundary established by the SVM replaces and refines both the k-filter and the l-filter.



(a) Discarding devices outside the k-filter



(b) Retesting devices outside the k-filter

Fig. 8. Defect filtering

4. Experimental Results

In order to compare the proposed confidence estimation method to the prior art methods of defect filtering and guard banding, we use production test data from a zero-IF down converter for cell-phone applications, designed in RFCMOS technology, fabricated at IBM and currently running in production. The device is characterized by 143 performances, 72 of which are non-RF (i.e. digital, DC, low frequency) and 71 are RF. The test dataset includes performance measurements for 4450 devices across 3 lots. Of these devices, 4141 pass all the specification tests while 309 fail one or more specification tests. The passing and failing devices are each randomly split into three subsets of equal size: P_1, P_2, P_3, and F_1, F_2, F_3. The sets $S_t = P_1 \cup F_1$, $S_h = P_2 \cup F_2$ and $S_v = P_3 \cup F_3$ are used as the training set, the hold-out set and the validation set, respectively. For all the experiments, correlation models are learned through MARS (Multiple Adaptive Regression Splines) [14]. The results for each of the three methods are reported below. We remind that the objective of non-RF to RF correlation-based specification test compaction is to predict pass/fail decisions by only measuring non-RF performances.

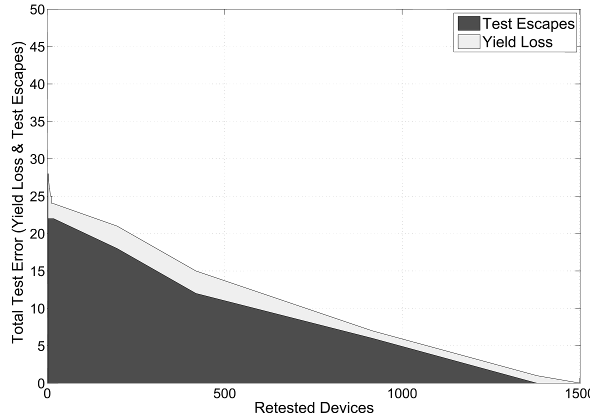


Fig. 9. Guard banding

4.1. Defect Filtering

As explained in section 3.1, the correlation models are learned in S_t, the parameter k is picked by assessing the behavior of the learned correlation models in S_h, and the final figures of merit of defect filtering are reported in S_v. The results when all devices outside the k-filter are discarded are shown in Fig. 8(a), where the number of test escapes, the yield loss inside the k-filter and the yield loss outside the k-filter are reported. As can be seen, the best trade-off point is found for $k = 6$, where 25 devices are misclassified (22 failing devices which are kept and 3 passing devices which are discarded). The results when all devices outside the k-filter are retested are shown in Fig. 8(b), where the test error is plotted against the number of retested devices as k decreases. As can be observed, for large values of k (near the y-axis), few devices are excluded by the k-filter and, therefore, retested. Yet the accuracy of the correlation models deteriorates as many devices not belonging to the distribution are included by the k-filter, resulting in high yield loss. As k is reduced, the number of retested circuits increases and the test error decreases.

4.2. Guard Banding

As explained in section 3.2, the correlation models are learned in S_t, the parameters k and l are picked by assessing the behavior of the learned correlation models in S_h, and the final figures of merit of guard banding are reported in S_v. The results are shown in Fig.9, where the test error is plotted against the number of retested devices for the Pareto front of (k, l) pairs. As expected, adding the l-filter slightly improves the results.

4.3. Proposed Method: Confidence Estimation

As explained in section 3.3, the correlation models are first learned in S_t. Then they are applied to the devices in S_h and a trusted/untrusted label is given to each device depending on whether the models predict its pass/fail label accurately or not. An SVM is, subsequently, trained to separate the trusted from the untrusted devices in S_h. The parameter k is picked by assessing the effectiveness of the trained SVM on an

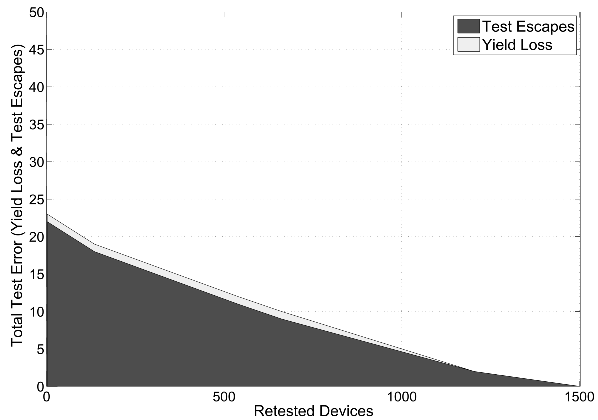


Fig. 10. Confidence estimation

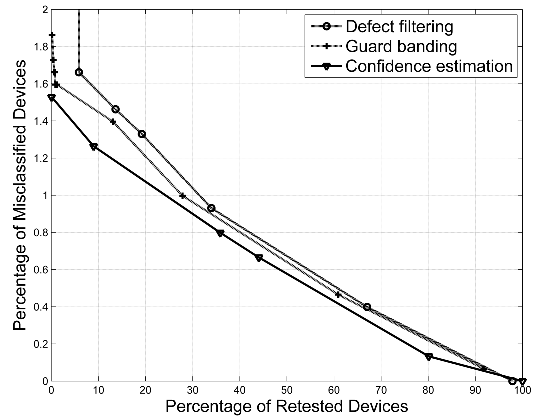


Fig. 11. Summary of results

independent set (we may use S_t for this purpose) and the final figures of merit of confidence estimation are reported in S_v. The results for the proposed SVM-based confidence estimation method are shown in Fig. 10. As can be observed, both the number of retested devices and the test error are reduced, as compared to defect filtering and guard banding.

4.4. Comparison & Future Work

The results for the three different methods described above are summarized and compared in Figure 11, where we plot the Pareto-optimal front of the percentage of test error vs. the percentage of retested devices. As can be observed, the proposed SVM-based confidence estimation method clearly improves upon the defect filtering and guard banding methods.

As a next step, we are currently in the process of obtaining a much larger dataset for the same IBM-fabricated device, as well as data from a different device, whereon we plan to repeat the experiment in order to corroborate our findings.

5. Conclusions

Specification test compaction through non-RF to RF performance correlation promises significant test cost reduction. Yet, in order to meet industry-level DPM standards, such compaction relies on efficient methods for boosting the accuracy of the correlation models and exploring the trade-off between the test error and the number of devices that need to be retested through complete specification testing. To this end, we developed a confidence estimation method which employs an SVM to decide whether the test outcome obtained through the learned correlation models can be trusted or not. As demonstrated experimentally using production test data from a zero-IF down-converter fabricated by IBM, the proposed method outperforms previously proposed defect filtering and guard banding methods, thus facilitating more accurate and less expensive production testing.

Acknowledgement

This research is partially supported through the Semiconductor Research Corporation (SRC Task 1632).

References

[1] J. B. Brockman and S. W. Director, "Predictive subset testing: Optimizing IC parametric performance testing for quality, cost, and yield," *IEEE Transactions on Semiconductor Manufacturing*, vol. 2, no. 3, pp. 104–113, 1989.

[2] R. Voorakaranam, A. Chatterjee, S. Cherubal, and D. Majernik, "Method for using an alternate performance test to reduce test time and improve manufacturing yield," Patent Application Publication #11/303,406, 2005.

[3] S. S. Akbay and A. Chatterjee, "Fault-based alternate test of RF components," in *Proc. International Conference on Computer Design*, 2007, pp. 517–525.

[4] P. N. Variyam and A. Chatterjee, "Specification-driven test generation for analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 10, pp. 1189–1201, 2000.

[5] P. N. Variyam, S. Cherubal, and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 3, pp. 349–361, 2002.

[6] N. Cristianini and J. Shawe-Taylor, *Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge, 2000.

[7] T. M. Souders and G. N. Stenbakken, "A comprehensive approach for modeling and testing analog and mixed-signal devices," in *IEEE International Test Conference*, 1990, pp. 169–176.

[8] G. N. Stenbakken and T. M. Souders, "Test-point selection and testability measures via QR factorzation of linear models," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-36, no. 2, pp. 406–410, 1987.

[9] L. Milor and A. L. Sangiovanni-Vincentelli, "Minimizing production test time to detect faults in analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, no. 6, pp. 796–813, 1994.

[10] S. Biswas, P. Li, R. D. (Shawn) Blanton, and L. Pileggi, "Specification test compaction for analog circuits and MEMS," in *Design, Automation and Test in Europe*, 2005, pp. 164–169.

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

[12] 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.

[13] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, John Wiley & Sons, 2003.

[14] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.