J Electron Test (2009) 25:309–321
DOI 10.1007/s10836-009-5113-7

On Boosting the Accuracy of Non-RF to RF Correlation-Based Specification Test Compaction

Nathan Kupp · Petros Drineas · Mustapha Slamani · Yiorgos Makris

Received: 1 October 2008 / Accepted: 26 August 2009 / Published online: 10 September 2009
© Springer Science + Business Media, LLC 2009

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 explore several methodologies that estimate 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 methodologies overcome the inability of standard specification test compaction methods to reach industrially acceptable test quality levels, and enable efficient exploration of the tradeoff between test accuracy and test cost.

Keywords RF test · Specification test compaction · Machine learning

Responsible Editor: C. Metra

Preliminary results of the research described herein were presented at the 2008 IEEE European Test Symposium. This research is partially supported through the Semiconductor Research Corporation (SRC Task 1632) and the National Science Foundation (NSF ECS-0622081).

N. Kupp · Y. Makris (✉)
Department of Electrical Engineering, Yale University, 10 Hillhouse Avenue, New Haven, CT 06520, USA
e-mail: yiorgos.makris@yale.edu

N. Kupp
e-mail: nathan.kupp@yale.edu

P. Drineas
Department of Computer Science, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA
e-mail: drinep@cs.rpi.edu

M. Slamani
IBM, Wireless Test Development Group, 1000 River Street 863G, Essex Junction, VT 05452, USA
e-mail: slamanim@us.ibm.com

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 cost 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 s non-RF performances and the t RF performances are explicitly measured. Based on this information, statistical correlation models are learned, predicting each excluded performance as a function of the non-RF performances of a device, or a subset of those performances. Subsequently, for every new device in production, only the selected non-RF performances are explicitly measured, while the untested non-RF and RF performances are predicted through the learned correlation models. A pass/fail decision is made by comparing the explicitly measured non-RF performances and the predicted 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 selected 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 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 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,

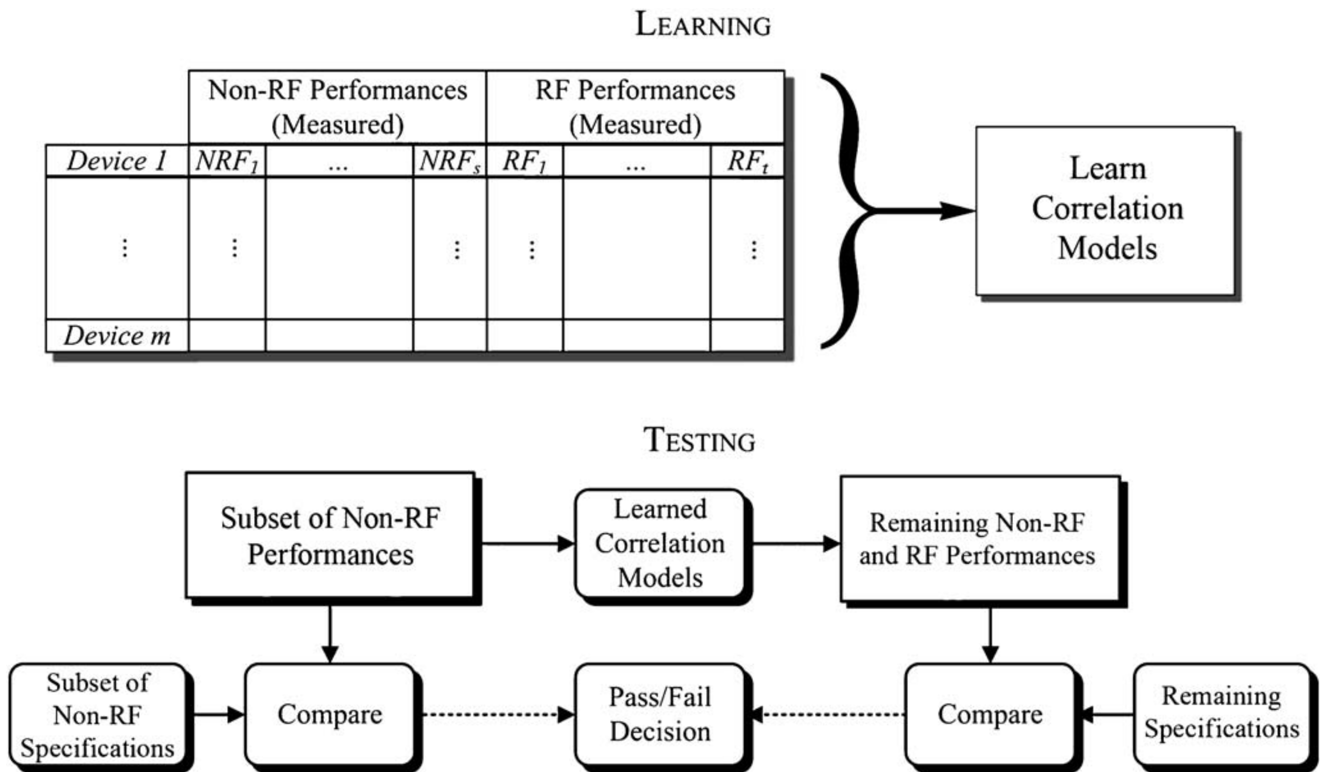


Fig. 1 Non-RF to RF correlation-based testing

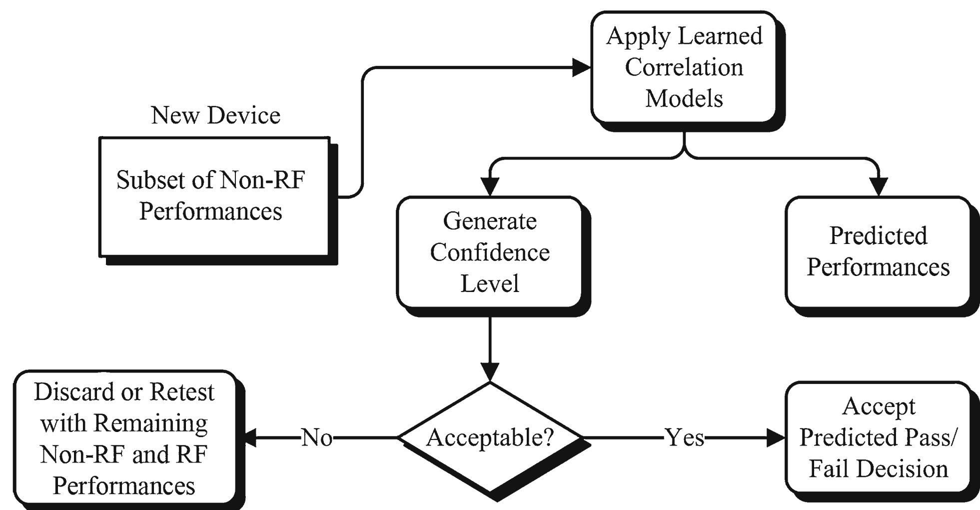


Fig. 2 Retesting when prediction confidence is low

successful deployment of 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 tradeoff between test error and test cost.

In this paper, we investigate three such methods. The first two methods, generically termed *defect filtering* and *guard banding*, are inspired by similar approaches that have been previously mentioned in passing in [1, 3, 15]. These, however, were not in the context of correlation-based specification test compaction but, rather, in the related field of alternate test [13, 14]. 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 tradeoff exploration methods used therein are relevant and can be adapted to the non-RF to RF correlation-based specification test compaction problem, so we examine them in detail. The third method, termed *confidence estimation*, is an entirely novel approach for deciding whether the pass/fail prediction yielded by the correlation models is sufficiently accurate or not. It employs an additional learning phase, wherein a Support Vector Machine (SVM) [4] 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 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. Overall, it is expected that these three methods will prove successively more powerful in reducing the cost of specification testing through the two-tier test approach of Fig. 2, without adversely impacting its effectiveness.

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 three aforementioned methods, namely defect filtering, guard banding, and confidence estimation. Finally, in Section 4, we provide experimental results based on production test data from a zero-IF down-converter fabricated at IBM, which demonstrate the comparative performance of these three methods and their ability to reduce test cost while maintaining industry-acceptable test quality levels.

2 Related Work

Various analog specification test compaction methods have been developed in the past. The linear error-mechanism model algorithm (LEMMA) [16] 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 [9] 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 [7], 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 [2], 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 [12]. A guard-banding and a two-tier test method applicable to the latter is discussed in [11].

3 Accuracy Boosting and Trade-off Exploration

In this section, we first discuss present the defect filtering and guard banding methods, as interpreted and adapted to the specification test compaction problem based on their brief descriptions provided in [1, 3, 15]. Then, we introduce the newly developed SVM-based confidence estimation method and we conclude the section by describing the mechanism that all three methods employ for selecting the subset of kept non-RF tests that are used as predictor variables for the omitted non-RF and RF tests.

3.1 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 [8]. 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.

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

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

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 omitted RF or non-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

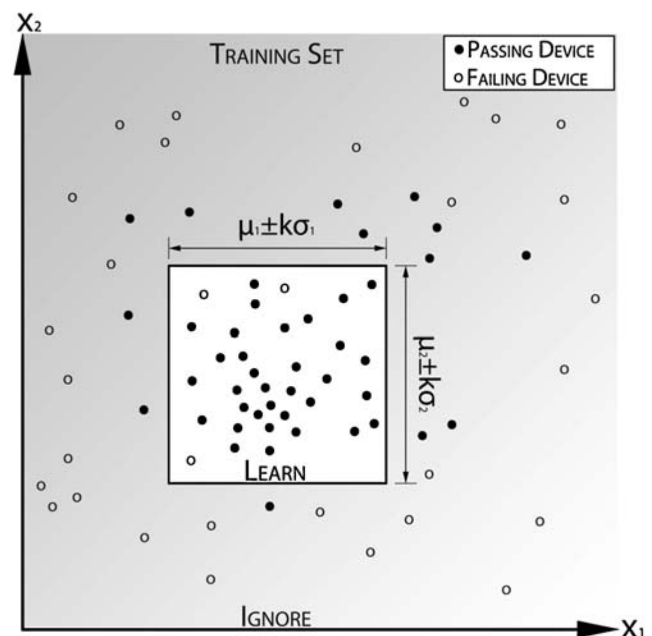


Fig. 3 Learning phase

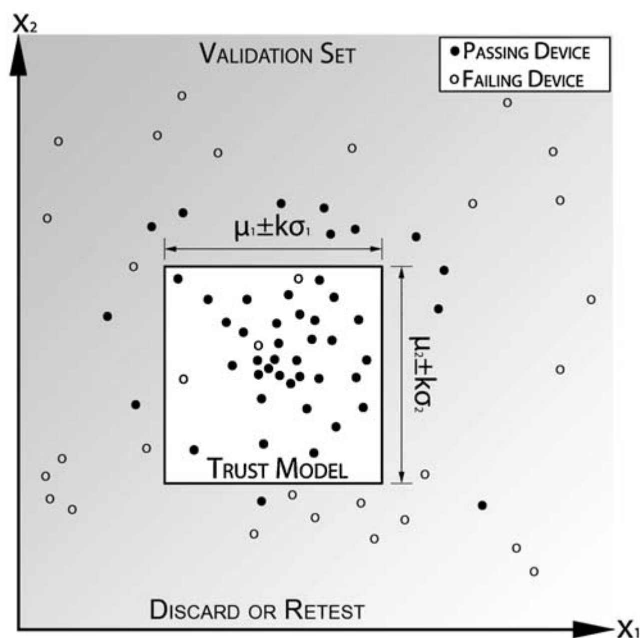


Fig. 4 Defect filtering—Testing phase

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 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 the kept non-RF performances may not reflect the complete information spectrum of the omitted non-RF and 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 [3, 15] 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.

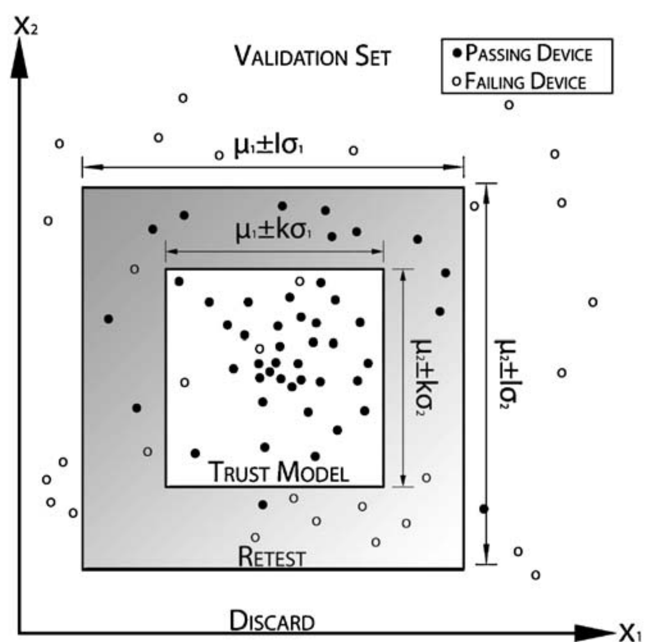


Fig. 5 Guard banding—Testing phase

3.3 Confidence Estimation

Both defect filtering and guard banding rely on the mean and the standard deviation of the devices in the training set 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 describe a confidence estimation method which uses an SVM [4] 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

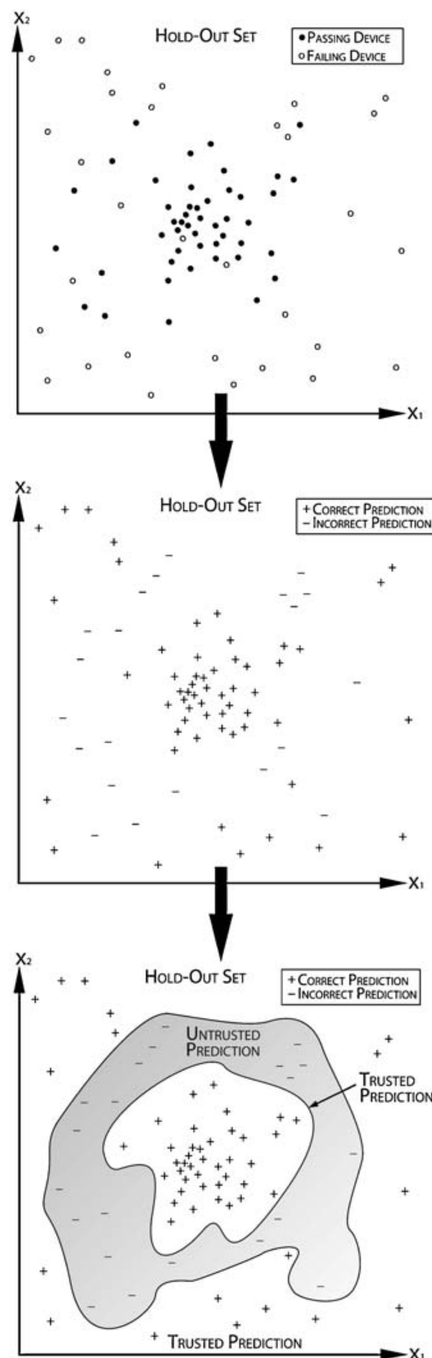


Fig. 6 Confidence estimation method

Step 1: Use correlation models learned from trained set devices inside a k-filter to predict pass/fail and relabel each device in the hold-out set as correctly or incorrectly predicted.

Step 2: Train an SVM to separate the regions wherein the pass/fail prediction of the correlation models can be trusted from the regions wherein it cannot.

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.

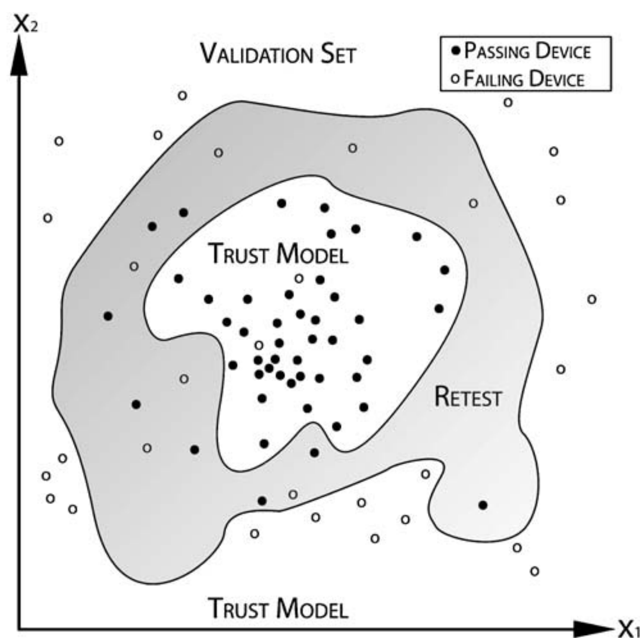


Fig. 7 Confidence estimation—Testing phase

3.4 Measurement Selection

Specification test compaction relies on keeping a subset of non-RF measurements and omitting the remaining non-RF and RF measurements. Therefore, an efficient mechanism for selecting the kept measurements is required. As described above, the three accuracy boosting methods appear to be exploring the tradeoff between mispredictions and retested devices. However, we can relate retests to a more concrete metric of success. Given a relative cost metric on the performance space, we can transform the trade off of mispredictions versus retests into one of mispredictions versus cost. Defining complete specification testing as a baseline cost, we can directly compare the cost of each method at a variety of solutions, each costing a fraction of the original specification test while introducing a small number of mispredictions, whether in yield loss or test escapes. To do this, we define a cost metric as follows.

Notation

C_T	Baseline test cost
T_T	Baseline test time
C_{RF}	Test cost per second for RF tester
M	Number of test sets
n_i	Number of tests in test set i
t_i	Relative test time contribution of test set i
c_i	Relative test cost contribution of test set i
x_{ik}	1 if test k in the test set i is present, 0 otherwise
C'_T	Test cost considering a reduced set of tests

Cost model The baseline test cost considering all tests is given by:

$$C_T = \sum_{i=1}^{M} (c_i C_{RF})(t_i T_T)$$
$$= C_{RF} T_T \sum_{i=1}^{M} (c_i t_i)$$

Assuming that the test time of any test k in the test set i is t_i/n_i, a first order approximation of the test cost of a reduced set of tests is given by:

$$C'_T = C_{RF}T_T \sum_{i=1}^{M} \left(c_i(1 - \overline{x_{i1}} \cdot \overline{x_{i2}} \cdot \ldots \cdot \overline{x_{in_i}}) \sum_{k=1}^{n_i} \frac{t_i}{n_i} x_{ik} \right)$$
$$= C_{RF}T_T \sum_{i=1}^{M} \left(\frac{c_i t_i}{n_i} \sum_{k=1}^{n_i} x_{ik} \right)$$

where the symbol $\cdot$ denotes the logic AND. Given the above expression for the test cost, our measurement selection algorithm utilizes the following normalized fitness function to assign a cost penalty to a selected test subset.

$$f(test\ subset) = \frac{C'_T}{C_T}$$
$$= \frac{C'_T}{C_T} = \frac{\sum_{i=1}^{M} \left(\frac{c_i t_i}{n_i} \sum_{k=1}^{n_i} x_{ik} \right)}{\sum_{i=1}^{M} c_i t_i}$$

Thus, we have a means of evaluating the cost level for any given subset of tests. This allows us to fold our retest count into the cost metric by computing it as follows. Suppose we have n devices on which we perform a reduced subset of the tests, and then use defect filtering, guard banding, or confidence estimation to flag a subset of those devices, $n_{retested} < n$ for retest.

We can then evaluate this relative cost metric for each method and use a single cost value to capture both the cost of the reduced set of specification tests, as well as the cost of retesting all devices flagged for retest:

$$cost = n \cdot f(retained\ tests) + n_{retested} \cdot f(excluded\ tests)$$

Genetic Algorithm With the cost function above, we proceed to automate measurement selection. With many non-RF tests, exhaustive search of the test set quickly becomes infeasible, as the number of possible test plans is $O(2^k)$, where k is the number of non-RF tests. However, the search space is well-suited for the application of a genetic algorithm. We have a pair of fitness functions, cost and mispredictions, which are quickly computable for any selected test subset and correlation-based specification test compaction method. Thus, by configuring a genetic algorithm to select a subset of the non-RF tests and k (defect filtering and confidence estimation) or k and l (guard banding) we can quickly find a set of non-RF tests producing tradeoff points between mispredictions and cost. In this work, we use the NSGA-II algorithm [5]. NSGA-II allows for chromosome elitism and preserves diversity, ensuring a good spread of Pareto-optimal solutions, and has been demonstrated previously [11, 12] to be effective in classification-based specification test compaction. The complete measurement selection procedure is depicted in Fig. 8. We note that the selection procedure is unbiased, in the sense that it minimizes mispredictions without showing preference to one of the two types of error (i.e. test escapes and yield loss). However, given relative costs of each of the above two types of error, it is possible to bias the selection procedure in order to favor one over the other in a weighted fashion. Details of how this can be done can be found in Section IV.C.3 of [10].

4 Experimental Results

In order to assess the relative effectiveness of defect filtering, guard banding, and confidence estimation, we use production test data from a zero-IF down

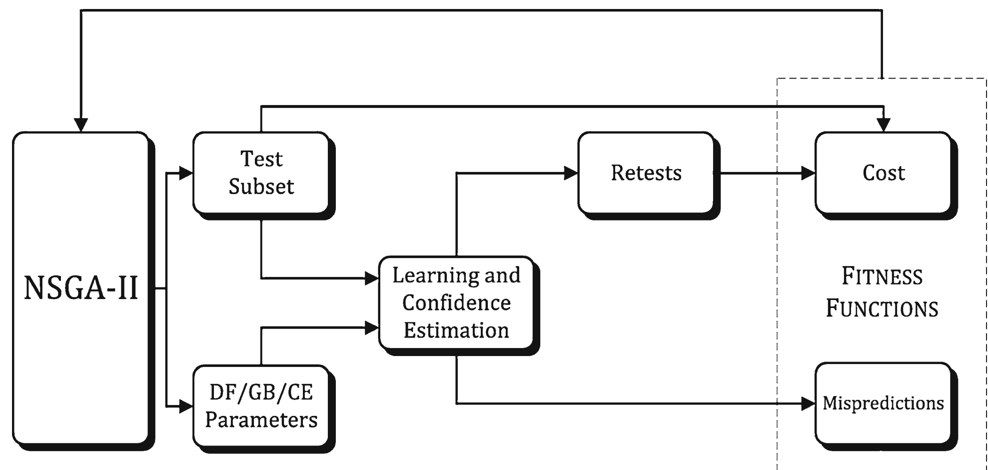


Fig. 8 Measurement selection algorithm

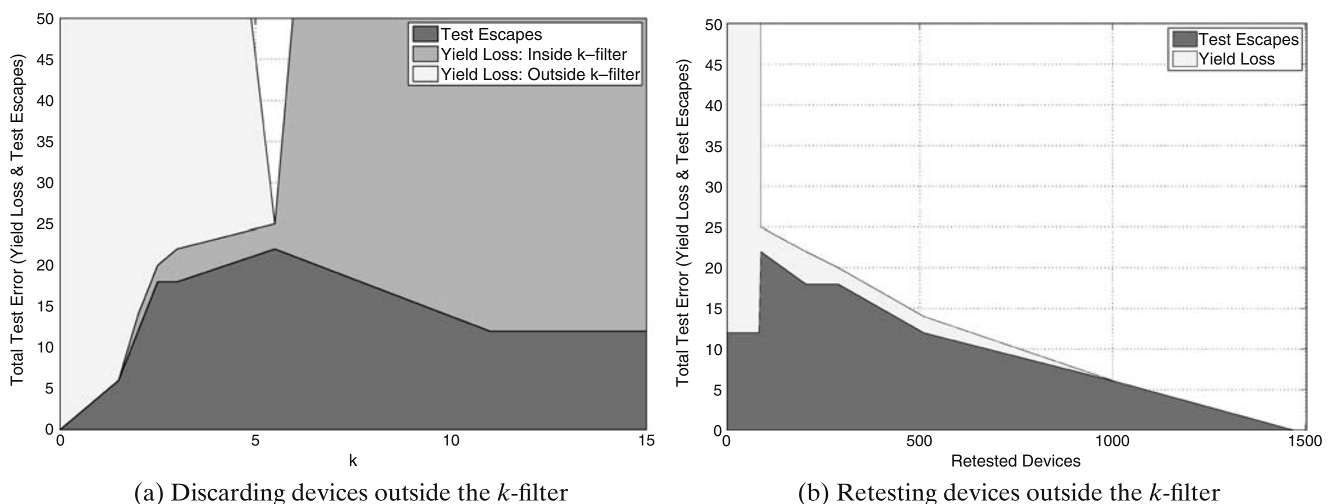
Table 1 Test cost data

Test Type	Test Group	# of Tests	Test % Time	Needed Tester Instrumentation	Tester Relative Cost
Digital	SBI tests	25	6.00%	Digital + DC	40%
DC	SBI tests/ Supply currents	34	13.30%		
Mixed Signal	Dac Tests	6	3.30%	Digital + DC + Mixed Signal	60%
	Lock VCO	6	13%		
	Lock VCO	1	1.10%		
RF	Filter tests	20	13.30%	Digital + DC + Mixed Signal + RF	100%
	Mixer tests	43	30%		
	LNA tests	8	20%		

converter for cell-phone applications, designed in RFCMOS technology and fabricated at IBM. The device is characterized by 143 performances, 72 of which are non-RF (i.e. digital, DC, low frequency) and 71 are RF. Table 1 displays a categorization of these tests, and also the relative costs and needed tester times for each test category. The test data set 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) [6]. 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 a subset of non-RF performances.

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. 9, 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. 9, 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



(a) Discarding devices outside the k-filter (b) Retesting devices outside the k-filter

Fig. 9 Defect filtering. **a** Discarding devices outside the k-filter. **b** Retesting devices outside the k-filter

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. 10, 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 improves the results.

4.3 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 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. 11. As can be observed in Fig. 12, where the results of all three methods are presented together, both the number of retested devices and the test error of the confidence estimation method are reduced, as compared to defect filtering and guard banding. While the improvement may seem relatively small, we emphasize that acceptance of specification test compaction methods hinges upon their ability to reach industrially acceptable test quality levels. Therefore, when trying to improve upon test accuracy levels below the 1% error rate, even the smallest improvement can play a very crucial role.

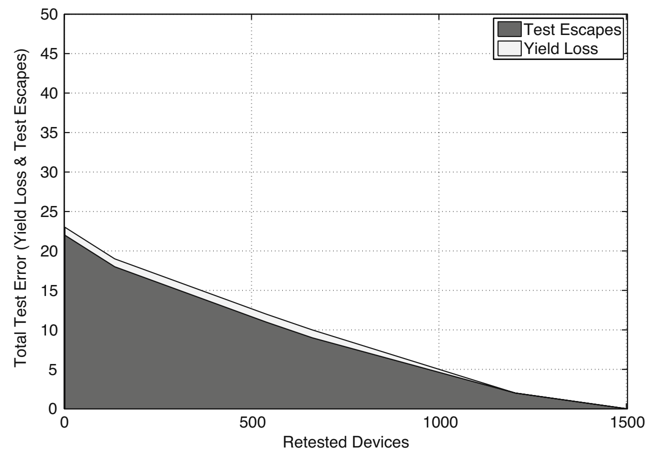


Fig. 11 Confidence estimation

4.4 Cost-Driven Measurement Selection

The preceding results explore the tradeoff between mispredicted devices and the number of retested devices. As described in Section 3.4, however, we can define a more accurate metric based on the cost of the included tests. This is achieved through defining a pertinent fitness function that drives the evolution of the genetic algorithm, which is wrapped around the correlation-based specification test compaction method and is used to select a subset of retained tests which optimizes both mispredictions and cost. In Fig. 13, we display the results from applying this technique for the

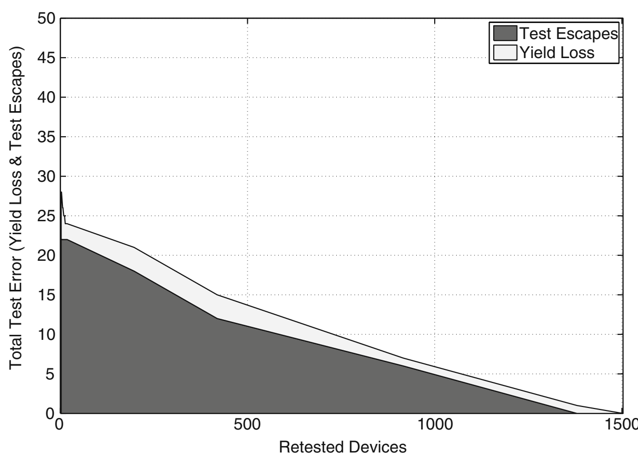


Fig. 10 Guard banding

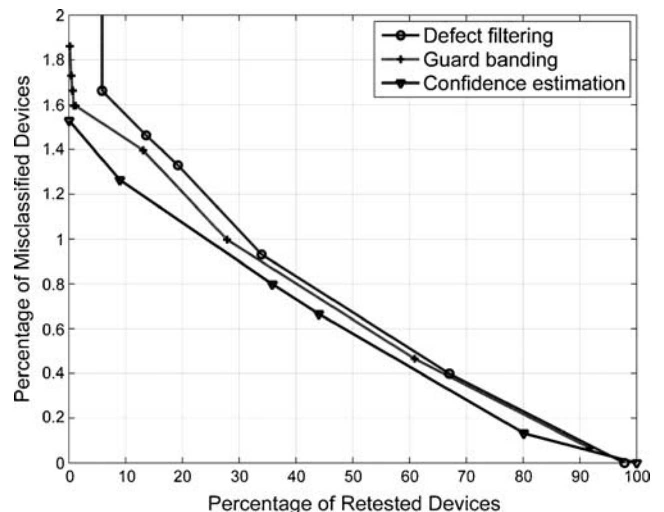


Fig. 12 Comparative results

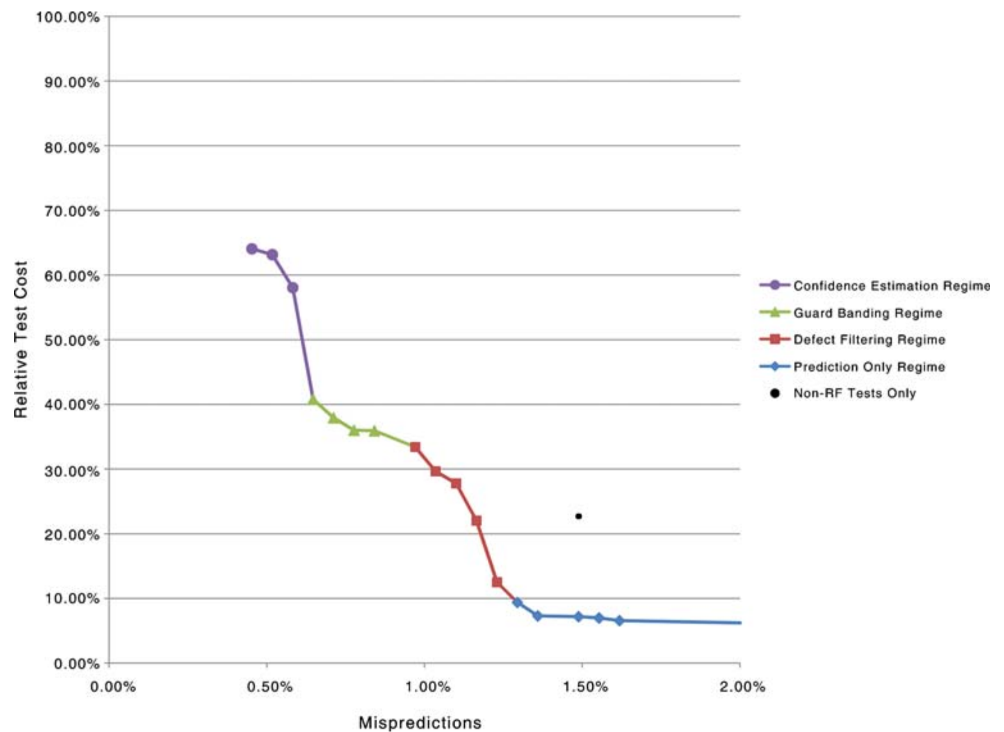


Fig. 13 Summary of cost-driven results

three methods described above. The vertical axis of this plot ranges from 0% to 100% of the cost of running the full specification test suite, while the horizontal axis represents the misprediction rate.

Along with results from the three methods described above, we also display the result of standard non-RF to RF specification test compaction (labeled "Non-RF tests only"), where all of the non-RF measurements are performed, based on which correlation models are built to predict all RF measurements. In addition, the plot labeled "Prediction only" provides the results of running the genetic algorithm for measurement selection without any of the defect filtering, guard banding, or confidence estimation methods. In other words, this plot reflects what non-RF to RF correlation-based specification test compaction can achieve without employing a second tier for retesting ambiguous devices.

Several observations can be made to assist in interpreting these results:

– First, we observe that correlation models built using *all* the non-RF measurements perform worse than correlation models that are built using a carefully selected subset of these measurements. Not only do the latter cost less, but they also result in a smaller percentage of mispredicted devices. While it may be counter-intuitive that the additional information available when involving all non-RF tests in the correlation models degrades their accuracy, this is a well-known fact referred to in pattern recognition terminology as the curse of dimensionality [4].
– Second, we observe that while specification test compaction without a second test tier incurs only 9.37% of the cost of complete specification testing, it cannot reduce test error below 1.29%, which is significantly higher than what industry is willing to accept. Therefore, methods that boost this accuracy, such as the one investigated herein, are particularly important.
– Third, we observe that the three developed methods are successively more powerful. Defect filtering enables us to move closer to acceptable test accuracy rates, dropping the test error to 0.97% at 33.4% of the cost of complete specification testing. Guard banding, which is a refined and more powerful method, is able to reduce the error even further to 0.65% by slightly increasing test cost to 40.8% of the cost of complete specification testing. Finally, confidence estimation, which is the most powerful of the three methods, enables us to limit the error to 0.453%, while cutting test cost to 64.05% of the cost of complete specification testing.
– Finally, we note that the lack of data points closer to the Y-axis in Fig. 13 is an artifact of the small size of the employed data set and not a limitation of the proposed method. Indeed, the best reported misprediction rate of 0.453% corresponds to only 7 devices; hence one should not directly extrapolate

to DPM levels from such a small data set, as that would artificially amplify the error.

5 Conclusion

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 investigated the effectiveness of three such methods. Two of these methods, namely defect filtering and guard banding, are adapted from pertinent ideas described within the context of alternate testing, while the third one, namely confidence estimation, is a new method which employs an SVM to decide whether the test outcome obtained through the learned correlation models can be trusted or not. All three methods employ a genetic algorithm and a test cost metric to perform measurement selection. As demonstrated experimentally using production test data from a zero-IF down-converter fabricated by IBM, these three methods facilitate an efficient exploration of the tradeoff between test cost and test accuracy, even in the region of very low DPM levels, wherein traditional specification test compaction is not able to reach.

Acknowledgments We would like to thank IBM for providing the production test data employed in this study, as well as Dr. Haralampos Stratigopoulos from TIMA Laboratory in Grenoble, France, for the useful discussions on the methodologies described herein and developing the cost metric described in Section 3.4.

References

1. Akbay SS, Chatterjee A (2007) Fault-based alternate test of RF components In: Proc. international conference on computer design, pp 517–525
2. Biswas S, Li P, (Shawn) Blanton RD, Pileggi L (2005) Specification test compaction for analog circuits and MEMS. In: Design, automation and test in Europe, pp 164–169
3. Brockman JB, Director SW (1989) Predictive subset testing: optimizing IC parametric performance testing for quality, cost, and yield. IEEE Trans Semicond Manuf 2(3):104–113
4. Cristianini N, Shawe-Taylor J (2000) Support vector machines and other kernel-based learning methods. Cambridge
5. Deb K, Pratap A, Agarwal A, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
6. Friedman JH (1991) Multivariate adaptive regression splines. Ann Stat 19(1):1–67
7. Milor L, Sangiovanni-Vincentelli AL (1994) Minimizing production test time to detect faults in analog circuits. IEEE Trans Comput-Aided Des Integr Circuits Syst 13(6):796–813
8. Seber GAF, Lee AJ (2003) Linear regression analysis. Wiley, New York
9. Stenbakken GN, Souders TM (1987) Test-point selection and testability measures via QR factorzation of linear models. IEEE Trans Instrum Meas IM-36(2):406–410
10. Stratigopoulos H-GD, Makris Y (2005) Non-linear decision boundaries for testing analog circuits. IEEE Trans Comput-Aided Des Integr Circuits Syst 24(11):1760–1773
11. Stratigopoulos H-GD, Makris Y (2008) Error moderation in low-cost machine learning-based analog/RF testing. IEEE Trans Comput-Aided Des Integr Circuits Syst 27(2):339–351
12. Stratigopoulos H-GD, Drineas P, Slamani M, Makris Y (2007) Non-RF to RF test correlation using learning machines: a case study. In: VLSI test symposium, pp 9–14
13. Variyam PN, Chatterjee A (2000) Specification-driven test generation for analog circuits. IEEE Trans Comput-Aided Des Integr Circuits Syst 19(10):1189–1201
14. Variyam PN, Cherubal S, Chatterjee A (2002) Prediction of analog performance parameters using fast transient testing. IEEE Trans Comput-Aided Des Integr Circuits Syst 21(3):349–361
15. Voorakaranam R, Chatterjee A, Cherubal S, Majernik D (2005) Method for using an alternate performance test to reduce test time and improve manufacturing yield. Patent Application Publication #11/303,406
16. Wrixon A, Kennedy MP (1999) A rigorous exposition of the LEMMA method for analog and mixed-signal testing. IEEE Trans Instrum Meas 48(5):978–985

Nathan Kupp received the B.S. in Electrical Engineering from The College of New Jersey in 2007 and is now a Ph.D. student at Yale University. His research interests are in the areas of analog and RF circuit testing.

Petros Drineas received the Diploma of Computer Engineering and Informatics from the University of Patras, Greece, in 1997, and the M.Sc. and Ph.D. in Computer Science from Yale University in 2001 and 2003, respectively. Since 2003, he has been an Assistant Professor of Computer Science at Rensselaer Polytechnic Institute. His research interests are in the areas of design and analysis of randomized and approximation algorithms for linear algebraic computations.

Mustapha Slamani joined IBM Microelectronics in 1999 as an RF test development engineer for SiGe/RFCOMS wireless communication products. He has a Ph.D. in mixed signal testing from Ecole Polytechnique de Montreal. His thesis results were used in two Fluence Technology, Inc. commercial analog and mixed signal test tools, the Fault-Maxx fault compiler, and TestMaxx test generator. Before joining IBM, he spent 4 years as a professor at the University of Quebec, Montreal, where he directed graduate student research in high-frequency testing, mixed signal testing, analog design, and signal integrity analysis. In 1998, he worked with Hyperchip Inc. in signal integrity analysis and design for test for high-speed digital circuits. Dr. Slamani is the co-founder and program chair of IEEE Workshop on Test of Wireless Circuits and Systems, topic coordinator of the RF session at International Test Conference and program committee member of Mixed Signal Test Workshop, Design Automation

and Test in Europe Conference, GHZ Test Workshop. He is a member of the industrial Die Product Consortium and ITRS RF chapter committee.

Yiorgos Makris received the Diploma of Computer Engineering and Informatics from the University of Patras, Greece, in 1995, and the M.Sc. and Ph.D. in Computer Science and Engineering from the University of California, San Diego, in 1997 and 2001, respectively. Since 2001, he has been on the faculty of Yale University, where he is now an Associate Professor of Electrical Engineering and of Computer Science. His research interests are in the area of test and reliability of analog, digital and asynchronous circuits and systems.