Non-RF To RF Test Correlation Using Learning Machines: A Case Study

Haralampos-G. D. Stratigopoulos*, Petros Drineas†, Mustapha Slamani‡ and Yiorgos Makris§
*TIMA Laboratory, 46 Av. Félix Viallet, 38031 Grenoble, France
†Department of Computer Science, Rensselaer Polytechnic Institute, Lally Hall, 110 8th Street, NY 12180, USA
‡IBM, Wireless Test Development Group, 1000 River Street 863G, Essex Junction, VT 05452, USA
§Department of Electrical Engineering, Yale University, 51 Prospect Street, New Haven, CT 06520, USA

Abstract— **We present a case study that employs production test data from an RF device to assess the effectiveness of four different methods in predicting the pass/fail labels of fabricated devices based on a subset of performances and, thereby, in decreasing test cost. The device employed is a zero-IF down-converter for cell-phone applications and the four methods range from a simple maximum-cover algorithm to an advanced ontogenic neural network. The results indicate that a subset of non-RF performances suffice to predict correctly the pass/fail label for the vast majority of the devices and that the addition of a few select RF performances holds great potential for reducing misprediction to industrially acceptable levels. Based on these results, we then discuss enhancements and experiments that will further corroborate the utility of these methods within the cost realities of analog/RF production testing.**

I. Introduction

Specification testing, wherein the performances of the device are verified against the specification limits, still remains the only acceptable industrial practice for analog/RF devices. Yet the high cost of RF ATE and the lengthy test times involved have resulted in intensified efforts and interest in reducing the number and types of performances that are examined during production testing. A plausible direction towards decreasing cost, akin to test compaction practices in digital circuits, is to identify and eliminate information redundancy in the set of performances, thereby relying only on a subset of them in order to reach a pass/fail decision. Such redundancy is likely to exist since groups of performances refer to the same portion of the chip and are subject to similar process imperfections. Since it is not possible to express the relationship between performances in closed-form functions, the idea of identifying information redundancy through the specification test data logs has been pitched. Yet it is highly unlikely that such redundancy will manifest itself in a coarse and easily observable form of superfluous performances that can be summarily discarded. Instead, more advanced statistical analysis methods are likely to be required. In essence, these methods should entail two components, namely a *selection algorithm* for searching in the power-set of performances for a discriminative low-cost subset and a *prediction model* for making pass/fail decisions based solely on this subset.

In this paper, we present the results of a case study that examines the effectiveness of four such methods using production test data from an RF device. These four methods are listed in Table I. In the MAX-COVER formulation the prediction model is trivialized, i.e. pass/fail decision is reached by simply comparing the selected performances to their specifications and ignoring the missing ones. This simplistic approach serves mainly as a basis for comparison. In contrast, the other three methods implement *classifiers* that learn to map the selected set of performances directly to a pass/fail decision, thereby implicitly predicting conformance of the eliminated performances to the specifications. NN and LDA are standard machine learning approaches, while ONN can learn complex non-linear mappings. Using these four methods, we are interested in exploring the following two questions:

- How well can the pass/fail decision of the RF device be predicted through models constructed based solely on a select set of non-RF performances (i.e. digital, DC and low frequency)? The motivation for this question is the fact that by only relying on non-RF performances, the need for RF ATE is eliminated, thereby drastically reducing test cost.
- How does the prediction accuracy improve by selectively adding a few RF performances to the best non-RF performance subsets? The motivation for this question is that even if the cost of an RF tester cannot be completely eliminated, it may still be possible to decrease the time that each device spends on it and, by extension, the overall test cost, by reducing the number of RF performances that are explicitly tested.

The main conjecture drawn from this case study is that the machine learning approach to the specification test compaction problem shows great promise for reducing test cost. Indeed, a relatively small number of only non-RF performances are shown to suffice for predicting correctly the pass/fail decision of a very large percentage of devices (around 99% in our case study). Moreover, the addition of few RF performances, ameliorates this small prediction inaccuracy and results in very powerful prediction models, which enable test cost reduction while maintaining industrially acceptable test quality

TABLE I
LEARNING METHODS EMPLOYED IN THIS CASE STUDY.

Method	*Selection algorithm*	*Prediction model*
MAX-COVER	MAX-COVER formulation	Trivial
NN+QR	QR decomposition	Nearest Neighbors
LDA+QR	QR decomposition	Lincar Discriminant Analysis
ONN+GA	Genetic Algorithm	Ontogenic Neural Network

25th IEEE VLSI Test Symmposium (VTS'07)
0-7695-2812-0/07 $20.00 © 2007 IEEE

standards. The results of this case study also make evident that the more elaborate ONN+GA method outperforms the simpler MAX-COVER, NN+QR and LDA+QR methods; this eludes to the fact that the correlations between the kept and discarded performances are indeed intricate and justifies the use of advanced machine learning methods. While we acknowledge that the questions explored through this case study reflect a simplified model of test economics, they still reveal the underlying potential of machine learning-based analog/RF specification test compaction methods to reduce test cost. Thus, they encourage further experimentation and assessment of our methods using larger data sets and more complex cost models that reflect more accurately the realities of a production test environment.

II. Related work

In the linear error-mechanism model algorithm (LEMMA) [1], it is assumed that a model $y = Ax$ is available [2], where y is the $m \times 1$ measurement error vector, x is a $n \times 1$ circuit parameter error vector, A is a $m \times n$ sensitivity matrix, and m corresponds to the number of measurements required for an exhaustive test of performances. The method aims to predict the complete vector y by carrying out only a subset $\tilde{y}$. The cardinality p of $\tilde{y}$ ($p \geq n$) is a compromise between the permitted measurement cost and the maximum tolerable prediction error. The selection process is performed through QR factorization [3] and minimizes the prediction variance. In [4], an iterative selection approach is followed, which considers subsets rather than individual measurements. Next, the complete measurement vector is predicted by $y = \tilde{A}\left(\tilde{A}^T\tilde{A}\right)^{-1}\tilde{A}^T\tilde{y}$, where $\tilde{A}$ is the $p \times n$ reduced matrix A. A leisurely look at this approach and some refinements are provided in [5]. The LEMMA method has the following limitations: (a) it relies on a linear model to predict the behavior of a non-linear system, (b) the linear model is developed through simulation and (c) it requires error mechanism models that are difficult to specify for complex circuits.

In [6], a fault-driven test selection approach is proposed. The set of performances to be explicitly tested is cumulatively built by adding to the current set the performance p_s for which the yield of the set $\left\{P - \bar{P} - p_i\right\}$ is maximized, where $\bar{P}$ and P denote the current and the complete set of performances, respectively. The algorithm terminates when the desired fault coverage is reached. In [7], in addition to fault coverage, the selection is also driven by the degree to which faults are exposed. The disadvantage of these approaches is their dependence on fault models, which are incomplete and, thus may result in inadvertent yield loss and test escapes.

In [8], a data set is generated by measuring explicitly all performances for a representative set of devices. Here, it is not required to adopt a fault model since it is assumed that this set of devices reflects accurately the statistical mechanisms of the manufacturing process. Once a suitable subset of performances that need to be explicitly tested is identified, regression models are constructed for the untested performances using the data set, and test limits are assigned to the tested performances such that they guarantee the compliance of the untested performances to the specifications with the desired confidence levels. The authors, however, do not show how to select the subset of independent performances, and, moreover, do not show how to explore efficiently the trade-off between the number of independent performances and yield loss.

In [9], the compaction problem is viewed as a binary pass/fail classification problem. Similarly to [8], the method begins with generating a data set by measuring all performances for a set of devices. Then, starting with the complete set of performances, $P = \{p_1, p_2, ..., p_N\}$, one performance p_s is selected at each step for possible removal. The training data corresponding to the set $\left\{\bar{P} - p_s\right\}$, where $\bar{P}$ denotes the current set of performances, is used to train a support vector machine (SVM) for predicting pass/fail only by processing the values of performances in $\left\{\bar{P} - p_s\right\}$. If the prediction error is smaller than a user define threshold, ϵ_r, then p_s is considered redundant and is permanently excluded. This selection procedure is greedy since the result depends on the order in which performances are examined. In practice it is advantageous to consider subsets of performances since combinations of performances can provide significant information which is not available in any of the individual performances separately. The method is assessed on an operational amplifier and a MEMS accelerometer using simulation data.

III. Data Set

Our case study vehicle is a zero-IF down-converter for cell-phone applications that is designed in RFCMOS technology, fabricated at IBM, and currently running in production. The input is an RF signal and the output is an IQ baseband signal. In addition, the device has an integrated VCO, a baseband filter and a DC nulling DAC. The LNA output is connected to the mixer input using an external SAW filter. This device is characterized by 136 performances, 65 of which are non-RF and 71 are RF. The data set contains the measured performances for 944 devices. These performance values, combined with the specification limits promised in the data sheet, are used to assign to each device a status bit denoting whether it is functional or faulty. Overall, the data set contains 73 faulty and 871 functional devices.

Preprocessing: Let A be the $944 \times (65 + 71)$ matrix containing the given data. The performances have typical values which differ significantly. In order to avert skewing of the distance between two devices in the performance space, each column of A is individually normalized. More specifically, each column of A is divided by the maximal entry, in absolute value, in this column. This procedure scales all data in the range $[-1, 1]$. Then, the columns of A are mean-centered by subtracting from every entry in each column the mean of the column elements. Formally, let A_{ij} denote the (i, j)-th entry in A, and let $A^{(j)}$ denote the j-th column of A. Scaling the data amounts to getting a new matrix A' whose entries are

$$A'_{ij} = A_{ij} / \max \text{abs}\left(A^{(j)}\right)$$

25th IEEE VLSI Test Symmposium (VTS'07)
0-7695-2812-0/07 $20.00 © 2007 IEEE

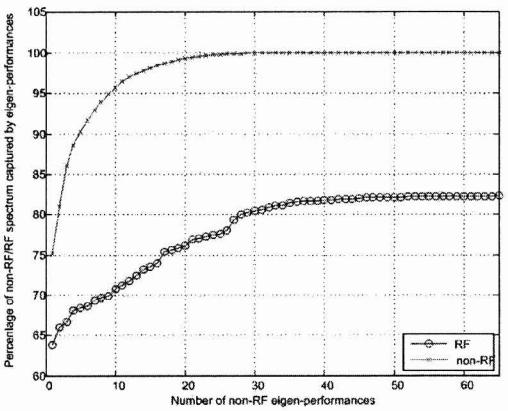

Fig. 1. Spectrum coverage plot for non-RF performances.

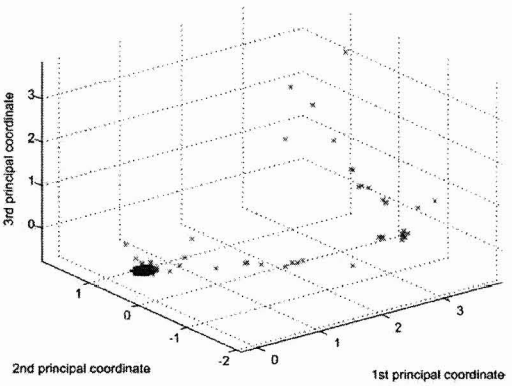

Fig. 2. "Best" 3D plot of non-RF performances.

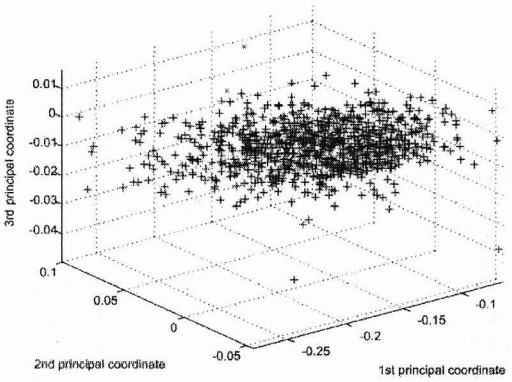

Fig. 3. Zoom in the core of functional devices in the 3D plot of Fig. 2

and mean-centering the data is equivalent to getting a matrix A'' whose entries are

$$A''_{ij} = A'_{ij} / \text{mean}\left(A'^{(j)}\right).$$

Visualization: To gain some intuitive understanding of the data, we first run Singular Value Decomposition (SVD), which returns the optimal ad hoc dimensions of the data set. Fig. 1 shows the spectrum coverage plot for the non-RF data. It can be observed that retaining about half of the principal dimensions suffices to capture all the information content in the non-RF data and over 80% of the information content in the RF data. This redundancy is key in accurately predicting pass/fail using only a subset of non-RF performances.

Fig. 2 plots the 944 devices on the coordinate system of the top three principal components of the non-RF data matrix. Blue '+' signs denote functional devices, whereas red 'x' signs denote faulty devices. Even in this rather primitive visualization, it is noticed that some faulty devices are easily detected since their patterns are very distant from the core of functional patterns. However, in Fig. 3, which zooms in on this core, we observe that there exist 10 faulty devices that are interwoven with the functional ones. Separating such devices in three dimensions seems difficult and one can only hope that this will be achievable by adding dimensions. We will monitor closely the effectiveness of the four methods on these devices.

IV. Learning Methods

A. MAX-COVER

A straight-forward algorithm for selecting a subset of performances and determining whether a device is functional or faulty by examining only this subset can be devised based on a simple maximum-cover formulation. Let T be an $m \times n$ matrix whose rows represent m devices and whose columns represent n performances. Let the (i,j)-th entry of T be set to 1 if and only if the j-th performance of the i-th device violates its specification; otherwise it is set to zero. We seek, among all possible subsets of c columns of T (performances), one that maximizes the number of faulty circuits detected by examining only those c performances. More formally,

$$\begin{aligned} \max & \|f\|_0 \\ \text{s.t.} \quad Tx &= f, \\ \sum_{i=1}^{n} x_i &\leq c, \\ x_i &\in \{0,1\}. \end{aligned}$$

The x_i's for $i = 1 \ldots n$ denote whether the i-th performance is selected or not. The i-th entry of f is non-zero if and only if the i-th circuit is faulty and detectable by the subset of the c selected performances. Notice that $\|f\|_0 = \sum_{i=1}^{n} (f_i)^0$ is exactly equal to the number of non-zeros in the vector f (by convention, $0^0 = 0$). Solving the above problem is NP-hard. A well-known approximation algorithm involves a straight-forward Integer Linear Programming (ILP) formulation of the problem and subsequently approximates the solution to the NP-hard ILP formulation by relaxing it to a Linear Program (LP), solving the LP, and using randomized rounding to map the resulting real values to integers. In [10], details of this procedure are given in the context of digital test compaction.

25th IEEE VLSI Test Symmposium (VTS'07)
0-7695-2812-0/07 $20.00 © 2007 IEEE

B. NN+QR

Nearest-Neighbor (NN) algorithms [11] constitute perhaps the simplest non-linear classifiers. Given a data point X_u whose label is unknown, NN examines its K labelled nearest neighbors (for a small odd value of K) and labels X_u by applying a majority vote on their labels. Picking the value of K is data-dependent. Typically, increasing K returns better results, until a point of diminishing returns is reached. Another relevant parameter for NN is the distance metric. In our experiments, we examine values of K between 1 and 11 and we use Euclidean distance to determine data point proximity.

In order to select subsets of performances, we use a variant of QR decomposition of matrices. Recall our discussion on the SVD of the non-RF performances: selecting a (small) number of eigen-performances (e.g., three) results to over 86% coverage of the performance spectrum. This procedure is equivalent to projecting all performances on a small number of eigen-performances, which capture most of the (linear) structure in the performance space. These eigen-performances are linear combinations of all performances, and thus do not represent any real, measurable quantity. An important question is whether we can select a small number of actual performances that behave in a similar manner, e.g., projecting all performances in the space spanned by the selected ones would result to capturing almost the same (linear) structure in the data that was captured by the eigen-performances. This research question has been answered affirmatively in [12], [13], and randomized algorithms were developed to perform subset selection. Here, we use a heuristic variant of the algorithm proposed in [13], originally developed in [14], which despite not having provable approximation guarantees of the form presented in [12], [13], it is deterministic and was empirically shown to perform well.

C. LDA+QR

Linear Discriminant Analysis (LDA) [11] is a more effective classification scheme, especially if the data originates from an (approximately) Gaussian distribution. LDA seeks a low-dimensional subspace, such that when the training data is projected on it, the ratio of the within-class scatter over the between-class scatter is minimized. Once the data is projected on the lower-dimensional space, the NN algorithm described in the previous method is applied to determine the label of new data points. LDA does not have any free parameters as the dimensionality of the subspace determined by LDA is always equal to the number of classes minus one; hence, in our case where we have two classes (functional and faulty), it is equal to one, i.e. the subspace collapses to a line. Performance subset selection is performed using the same QR decomposition algorithm as in the NN classifier.

D. ONN+GA

As a fourth technique, we use an ontogenic neural network (ONN) [15] to learn the position of the hypersurface separating the populations of functional and faulty devices in the training set, when these are projected on a subspace of the performances. This particular neural network is capable of allocating arbitrarily non-linear hypersurfaces, unlike SVMs that require the a priori definition of a kernel. The allocated hypersurface reflects all performances and constitutes a simple test criterion: the status bit of a new device is defined based on the position of the footprint of its reduced-size performance pattern with respect to the learned hypersurface.

In contrast to the previous methods, the stages of selecting performances and constructing the prediction model are optimized together. More specifically, we use a genetic algorithm (GA) to search in the space of performance subsets, assessing the fitness of each subset by training the ONN. The use of GAs for selecting features from a high-dimensional set is originally proposed in [16]. GAs start with a base population of chromosomes (bit strings, in our case, of length equal to the number of performances, where the k-th bit is set to 1 if the k-th performance is present in the subset and 0 otherwise), and use mutation and crossover operators to generate new off-spring populations. At the end of each generation, the fitness of chromosomes is evaluated and only the fittest chromosomes mate to produce off-springs. GAs evolve with the juxtaposition of bit templates, quickly optimizing the target fitness function. In this work, we use a multi-objective GA, called NSGA-II [17], to jointly optimize in one simulation run the prediction error of the ONN and the dimensionality of the selected subset of performances. For this purpose, NSGA-II has a diversity preserving mechanism that ensures a good spread of Pareto-optimal solutions.

V. Results

We split the data set equally in a training and a test set. As is common in evaluating machine learning techniques, we assume that the status bits are known only for the devices in the training set; status bits for the devices in the test set are assumed unknown and are only used to evaluate the prediction error of the learned models. Each experiment is repeated for 200 splits (devices are sampled uniformly at random) to reduce the variance of the reported results. The standard deviation of the prediction error is an order of magnitude smaller than its mean, so we can conclude that the mean of the prediction error is a statistically significant metric. As a comparison basis, the test set comprises 472 devices of which, on average, 37 are faulty; thus, a trivial classification algorithm always returning "functional" would achieve a prediction error of roughly 8%.

A. Selecting only non-RF Performances

Fig. 4 shows the experimental results for MAX-COVER. We run two experiments: in the first we use the complete set of available devices, while in the second we use the training/test set split formulation, we solve the maximum cover problem using only the training set, and we evaluate the prediction error on the test set. In the first experiment, a set of 6 non-RF performances suffices to cover 55 faulty devices, which is the maximum number that can be detected using only non-RF performances (i.e. the curve remains flat after adding more non-RF performances). This corresponds to an error (test

25th IEEE VLSI Test Symmposium (VTS'07)
0-7695-2812-0/07 $20.00 © 2007 IEEE

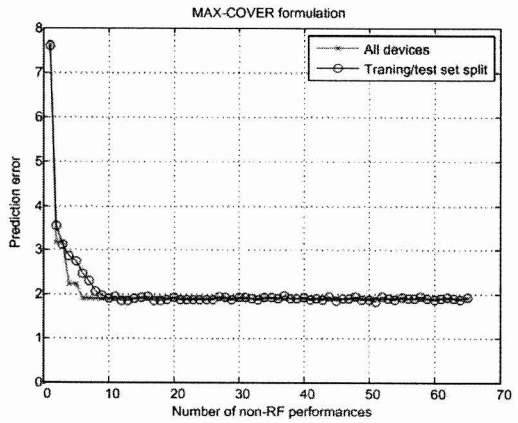

Fig. 4. Prediction error using MAX-COVER.

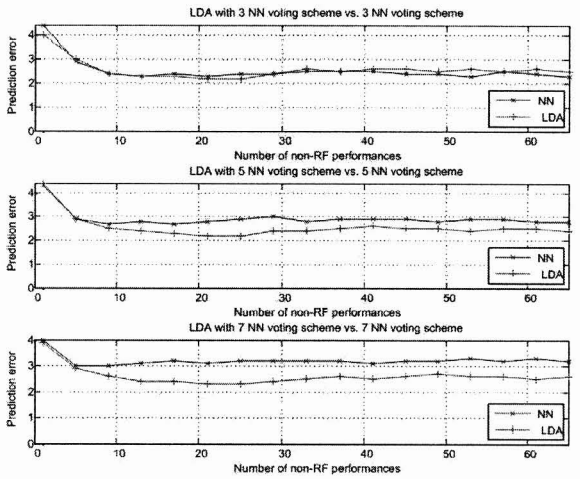

Fig. 5. Prediction error using NN+QR and LDA+QR.

escapes) slightly less than 2%. Almost identical results are obtained in the second experiment; the error converges very quickly to just below 2% for a set of 10 non-RF performances.

Fig. 5 shows experimental results of NN+QR and LDA+QR. We experimented with values of K (number of neighbors used by NN) between 1 and 11 and the results show that both classifiers achieve their best performances for $K = 3, 5$, or 7. LDA+QR returns slightly better results than NN, both of which are comparable to those of MAX-COVER. We also observed that all errors are test escapes and we examined how many correctly predicted devices belong to the set of the 10 faulty devices of Fig. 3. The average of this number over the 200 repetitions is zero, indicating that these "difficult" devices are always mispredicted.

Fig. 6 shows experimental results using ONN+GA. The 'o' points correspond to the optimal identified performance subsets for each cardinality. The dashed blue line runs along the Pareto-optimal points. A set of 16 performances achieves the lowest prediction error of 0.92%. This corresponds to 4 out of 472 devices being misclassified. Thus, ONN+GA misclassifies only the "difficult" faulty devices of Fig. 3, in contrast to NN+QR and LDA+QR, which misclassify, on average, 5 more devices whose performance pattern falls close to the core of functional devices.

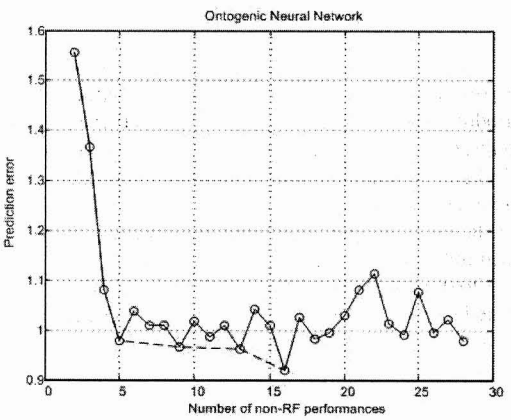

Fig. 6. Prediction error using ONN+GA.

B. Adding RF Performances

As a final experiment, we examine the prediction improvement that can be obtained by adding a few RF performances to the best subset of non-RF performances identified above. For example, for the ONN+GA, we start with the 16th-dimensional non-RF performance subset that results in the lowest prediction error of 0.92%. As seen in Fig. 7, by adding 3 RF performances to this set, the prediction error is reduced to 0.56%, and by adding 12 RF performances the error drops further to 0.38%. These rates correspond to only 1 misclassified device, which shows that there is great promise in using a subset of non-RF performances along with a very small number of RF performances to achieve accurate prediction. In the other three methods, the prediction error also is reduced but remains above 1% (plots are omitted due to space limitations).

VI. Future Work

While the results of this case study show great promise, the following enhancements targeting test cost reduction and test quality improvement would greatly support and expedite technology transfer to an industrial setting:

1) Regarding test cost, we aimed to minimize the cardinality of the set of selected performances, implicitly assuming that all performances incur the same cost. In practice, however, the cost varies for each performance due to differences in the corresponding test configuration and length. Furthermore, eliminating a performance does not necessarily save the cost of the corresponding configuration, since the latter may be shared across a group of performances. Thus, weighted versions of the compaction methods described herein should be

25th IEEE VLSI Test Symmposium (VTS'07)
0-7695-2812-0/07 $20.00 © 2007 IEEE

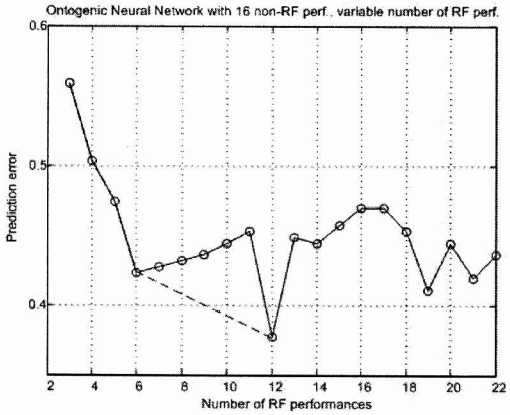

Fig. 7. Prediction error using ONN+GA when adding RF performances to the best identified non-RF performance subset.

explored, aiming to optimize a more complex function that better reflects the actual test cost.

2) Regarding test quality, the statistical nature of this specification test compaction approach entails a prediction error which, albeit small, may nevertheless be prohibitive for industrial standards, especially if it amounts to mostly test escapes. Thus, the use of some form of guard-banding [9], [18] should be explored, in order to deal with the devices that are prone to misprediction.
3) While prediction models are currently constructed using performances as inputs, greater prediction accuracy and elimination of more performances may be achievable by using, instead, the actual measurements obtained for computing the performances. In this case, further fine-grained cost reductions may be possible by eliminating individual measurements.

We also point out that the data set in this case study is fairly small and cannot provide a definitive answer as to whether the misprediction error really reflects a percentage (0.38%) due to underlying trends in the nominal and faulty distributions or whether it is an artifact of the data set and essentially reflects a constant (1 device) due to a peculiar outlier. Therefore, a continuation of this case study using many more devices and incorporating the above enhancements is currently under way.

VII. Conclusions

Analysis of production test data from an RF device reveals that performances comprise significant redundancy, which can be exploited to build prediction models for reaching pass/fail decisions based on a reduced-size set of performances. To this end, advanced machine learning and performance selection techniques, such as ONN+GA, achieve excellent results and demonstrate great potential for reducing test cost through specification test compaction. Further enhancements and evaluation of these methods on larger data sets is expected to confirm our findings and shape more research at the intersection of machine learning and analog/RF circuit testing.

References

[1] 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.

[2] G. N. Stenbakken and T. M. Souders, "Developing linear error models for analog devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 2, pp. 157–163, 1994.

[3] 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.

[4] J. Van Spaandonk and T. A. M. Kevenaar, "Iterative test-point selection for analog circuits," in *IEEE VLSI Test Symposium*, 1996, pp. 66–71.

[5] A. Wrixon and M. P. Kennedy, "A rigorous exposition of the LEMMA method for analog and mixed-signal testing," *IEEE Transactions on Instrumentation and Measurement*, vol. 48, no. 5, pp. 978–985, 1999.

[6] 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.

[7] G. Devarayanadurg, M. Soma, P. Goteti, and S. D. Huynh, "Test set selection for structural faults in analog IC's," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 7, pp. 1026–1039, 1999.

[8] 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.

[9] 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.

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

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining , Inference, and Prediction*, Springer, 2001.

[12] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Subspace sampling and relative-error matrix approximation: Column-row-based methods," in *European Symposium on Algorithms*, 2006, pp. 304–314.

[13] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Subspace sampling and relative-error matrix approximation: Column-based methods," in *APPROX-RANDOM*, 2006, pp. 316–326.

[14] M.W. Berry, S.A. Pulatova, and G.W. Stewart, "Computing sparse reduced-rank approximations to sparse matrices," Tech. Rep. UMIACS TR-2004-32 CMSC TR-4589, University of Maryland, College Park, MD, 2004.

[15] H.-G. D. Stratigopoulos and Y. Makris, "Constructive derivation of analog specification test criteria," in *IEEE VLSI Test Symposium*, 2005, pp. 395–400.

[16] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, pp. 335–347, 1989.

[17] K. Deb, A. Pratap, A. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[18] H.-G. D. Stratigopoulos and Y. Makris, "Bridging the accuracy of functional and machine-learning-based mixed-signal testing," in *IEEE VLSI Test Symposium*, 2006, pp. 406–411.