Feature Selection Methods for Analog and RF Test: A Case Study

Nathan Kupp* and Yiorgos Makris*
*Department of Electrical Engineering, Yale University, New Haven, CT 06511

***Abstract*—One of the biggest challenges in the analog and RF test domain is the high cost of traditional specification testing. In response, there is a great deal of interest in leveraging low-cost test methods as a drop-in replacement for specification test. With the correct choice of low-cost tests, substantial test cost savings can be achieved with only marginal effects on test error metrics. During characterization a large number of candidate tests are considered, though retaining all of these during production test would be impractical and inefficient. However, test selection can be quite challenging, and consequently, systematic selection of a low-cost test set to deploy in production is a recurring problem. In this work, we examine and compare several different feature selection techniques which address the test set selection problem for low-cost testing, and present the first use of Fast Function Extraction (FFX) in low-cost testing.**

I. Introduction

During semiconductor fabrication, every manufactured device must be thoroughly tested in order to guarantee that it meets the original design specifications. Such testing identifies latent defects that are due to the various sources of imperfection and variation in the fabrication process. Defects can present as either catastrophic or parametric. In the former case, they lead to a complete malfunction of the IC and, typically, can be detected by simple tests. In the latter case, they are caused by excessive process variation that may bring some or all of the specifications outside the allowable limits. Parametric defects are considerably harder to detect. To guard against test escapes, RF circuits are typically tested directly against the performances specified in the device data sheet. Although this approach is highly accurate, it comes at a very high cost, which can amount up to 50% of the overall production cost according to anecdotal evidence. Given that RF circuits typically occupy less than 5% of the die area, there is great industrial interest in the reduction of RF test cost [1], [2].

The high cost of RF test is due to the expense of automated test equipment that is required, and, on the other hand, the lengthy test times that result from a sequential measurement approach. Recently, there has been an intensified effort to develop alternative test approaches that relax the requirements on test equipment and/or reduce the associated test times. Among others, the built-in test solution is perhaps the most promising and advantageous [3]–[6]. It relies on extracting on-chip digital, DC or low-frequency test signatures that carry RF information. Thereafter, these test signatures can be transported off-chip and processed by an inexpensive tester with minimum requirements. Other work has focused on leveraging external DC, low-frequency, or other inexpensive measurements to measure device functionality without explicitly testing against expensive specification performances. Such low-cost tests are typically paired with learning algorithms which are trained to predict test outcomes [7]–[9].

For both built-in and external testing, a large population of candidate test sets can be devised to capture device performance. However, in production, we would like to only measure the test set that identifies device pass/fail labels as efficiently as possible. As a result, a recurring problem in low-cost testing is selecting appropriate test sets from the numerous possible candidate test sets. This is essentially a specific realization of a problem known as *feature selection*. Feature selection is the technique of retaining only a subset of measurements or independent variables from a full set of candidate measurements. This contradicts the intuition of "the more data, the better". However, spurious measurements are actually quite problematic, as data sparsity increases exponentially with the number of dimensions. Therefore, we retain only the most salient features to ensure stability of learning algorithms imposed on the low-cost measurement space. Beyond testing, feature selection is commonly used in many fields that attempt to construct models in high dimensional data, e.g. bioinformatics.

There are two reasons for pursuing feature selection in the analog and RF test space. First, test time is expensive, and it is not economical to expend test time collecting spurious measurements that provide little information about the device under test. Thus, we wish to only explicitly measure the minimal test set that properly identifies each device as passing or failing. Second, with low-cost testing we typically want to construct a model which employs the low-cost measurements to identify device pass/fail labels, or predict specification test outcomes. Limiting the number of features used in such models is crucial, as it permits us to sidestep the so-called "Curse of Dimensionality", which states that the predictive quality of a model decreases with the number of features retained.

In this work, we present a case study which compares various feature selection methods in the context of low-cost testing, and construct regression models on the retained features to evaluate the performance of each approach. Feature selection is not new to analog and RF test, although it has not been directly studied in this space before. In [10], the authors employ a genetic algorithm (GA) to identify optimal test stimuli. The work in [9] also employs a GA (specifically, NSGA-II [11]) for feature selection. In [12], we extended this approach to incorporate domain expert feedback directly into the GA training process. In [13], the authors use a greedy algorithm to retain tests. Indeed, the entire domain of test compaction [14], [15] is, at heart, an exercise in feature selection.

To evaluate the feature selection methods described in the following section, we employ a candidate set of low-cost tests from a Texas Instruments Bluetooth/Wireless LAN device. These measurements, known as low-cost on-chip built-in RF tests (ORBiTs) [6], form our set of candidate tests X_{ORBiT}. We then construct regression models to predict specification test outcomes P_i for the transceiver based on feature-selected subsets of X_{ORBiT}.

II. Feature Selection Methods

In this section we describe in detail each of the feature selection methods that were evaluated. Each of the methods has different characteristics, but ultimately the most important evaluation critieria is the error achieved. Thus, we first describe the error metrics we use as a basis of comparison.

A. Evaluation Criteria

To compare the various feature selection methods, we construct regression models of the following form:

$$\hat{P}_i = f(X) + \varepsilon \tag{1}$$

where P_i is an $n \times 1$ vector of measurements collected on n devices corresponding to the i-th specification performance, $X = [f_1, f_2, \ldots, f_m]$ is a $n \times m$ matrix of low-cost tests, where each column $f_j \in X$ is a candidate low-cost test, and f is a function mapping the low-cost tests to the specification test space $f\colon \mathbb{R}^m \to \mathbb{R}$. With feature selection, we aim to replace the matrix of low-cost tests X in Equation 1 with:

$$X' = [f_k, k \subseteq \{1, 2, \ldots, m\}] \tag{2}$$

that is, X' is a column-wise subset of X that retains only salient features of X.

By constructing such models for each feature selection approach, we can compare feature selection methods on (i) the normalized mean-square error (NMSE) achieved by each approach, (ii) the number of features retained, and (iii) the time required to perform the feature selection step.

B. Random Search

The most readily apparent approach to feature selection is to randomly select a subset of features, build and evaluate a regression model, and then use the observed NMSE to drive the search process. That is, we choose feature subsets X' as in Equation 2, with each feature subset containing m' number of features.

In theory, this is a poor approach, as the number of possible feature subsets is $2^m - 1$, where m is the total number of features. Even for reasonable datasets and evaluation times—say, 100 features and 1 second to evaluate a candidate feature set—exhaustive search of the entire space would take approximately 4×10^{22} years. In practice, if the search space is not too sparse (e.g. many features are moderately adequate predictors) then random search will find passable feature subsets.

C. Greedy Correlation Coefficient-Based Ranking

A more elegant approach is to employ pairwise Pearson correlation coefficients to rank the available features, and then choose the top-ranked features to retain as predictors in the regression model. The principle of this approach is that highly correlated features generally make suitable predictors. Using the previously defined notation, the Pearson correlation coefficients are computed as:

$$\rho_k = \frac{\text{cov}(f_k, P_i)}{\sigma_{f_k}\sigma_{P_i}} \tag{3}$$

Thus, for features $[f_1, f_2, \ldots, f_m]$ we can compute the correlation coefficients $[\rho_1, \rho_2, \ldots, \rho_3]$ and order them according to the induction rule:

$$\rho^{(1)} = \min_k \rho_k, k \in 1, 2, \ldots, m \tag{4}$$

$$\rho^{(i)} \geq \rho^{(i-1)},\ \forall i \tag{5}$$

Then, by defining a lower bound T_L on $\rho^{(i)}$, we retain all features that satisfy $\rho^{(i)} \geq T_L$.

D. NSGA-II

The non-dominated sorting genetic algorithm (NSGA-II) has become the de facto standard for GA problems. GAs, also known as evolutionary algorithms, attempt to emulate biological natural selection by creating seed "populations" of feature subsets, which subsequently undergo mating and mutation steps. These steps are repeatedly performed in phases known as "generations". The justification for such steps are intuitive: by mating two solutions, we may discover a better solution, and by perturbing our solutions via mutation we help avoid local optima. At each generation, every member of the population is evaluated via fitness/objective functions, and the best solutions of each generation are recorded. A Pareto-optimal subset of these solutions is then recorded at the termination of the GA.

To use NSGA-II for feature selection, we typically define a pair of minimization objectives: NMSE and number of features retained. With these objectives, the algorithm will produce a series of points that represent tradeoffs between NMSE and number of features.

E. FFX

Fast Function Extraction (FFX) is a recently proposed white-box regression method [16] which aims to perform massive basis expansion of the original features and then use state-of-the-art regularization methods to effectively perform feature selection on the new bases. Traditional least-squares regression solves the quadratic minimization problem:

$$\hat{\beta} = \arg\min_{\beta} ||y - X\beta||^2 \tag{6}$$

For low-dimensional problems, this formulation works fine—indeed, this is the regression method used with the other feature selection techniques described herein. However, feature selection can be built directly into the regression optimization problem by introducing regularization parameters, e.g. via elastic net [17]:

$$L(\lambda_1, \lambda_2, \beta) = ||y - X\beta||^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1 \quad (7)$$

$$\hat{\beta} = \arg\min_{\beta} L(\lambda_1, \lambda_2, \beta) \quad (8)$$

A very useful property of regularizing the regression problem in this fashion is that the number of features can actually be massively *expanded*, and allow the regularization to extract the appropriate parameters. FFX leverages this by combining elastic net regression with generalized linear models of the form:

$$\hat{P}_i = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot B_i(X) \quad (9)$$

where the $B_i(\cdot), i \in [1, 2, \ldots, p]$ are a large number of basis functions of X, and $p \gg m$. In other words, we regress P_i on a much larger number of features, and depend on elastic net regularization to identify the most salient features. The interested reader is directed to [16] for further details in the interest of space.

III. Experimental Results

To evaluate the feature selection techniques, we employed a Bluetooth/Wireless LAN device fabricated by Texas Instruments. 249 candidate low-cost tests were considered as features, and the data from one wafer (approximately 7,000 devices) was split 50/50 as training and test sets. Identical training and test sets were employed for all of the approaches to ensure a fair comparison. One particularly challenging-to-predict specification performance was selected as a target. Each of the feature selection approaches have various parameters, the settings for which we outline below. All execution times cited were achieved on a 2010 MacBook Pro.

A. Greedy Correlation Coefficient-Based Ranking

For the greedy correlation coefficient ranking, each of the 249 ORBiTs was ranked as described in Section II-C. A series of models were constructed with successively higher numbers of features retained, with up to 50 of the highest-ranked features retained. For each retained set of features X'_{ORBiT}, a least-squares regression model was consctructed and the NMSE was recorded. This method required less than 30 seconds to run, and the best NMSE achieved was 0.092371 with 7 retained parameters. The predicted vs. actual of this model is plotted in Figure 3.

B. NSGA-II

We configured NSGA-II with each chromosome C representing a 249-dimensional set of indicator variables, e.g. $C = [I_1, I_2, \ldots, I_{249}], I_k \in \{0, 1\}$, and $I_k = 1$ corresponding to retention of the k-th ORBiT feature, e.g.: $I_k = 1 \rightarrow f_k \in X'_{ORBiT}$. A population of 256 chromosomes was employed, and 500 generations were evolved for a total execution time of approximately 1 hour.

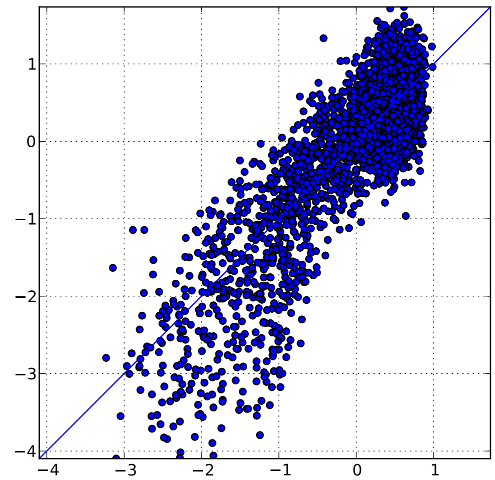

Fig. 1. Greedy CC Ranking: $\hat{P}_i$ vs. P_i

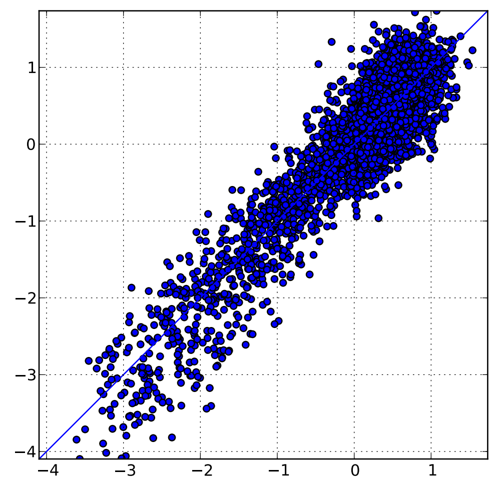

Fig. 2. NSGA-II: $\hat{P}_i$ vs. P_i

C. FFX

FFX has comparatively few parameters that the user must set. The elastic net pathwise learn algorithm has a maximum number of iterations, which we set to 1,000. FFX also has settings to enable or disable various types of basis functions, including interaction terms, rational functions, exponents, abs$(\cdot)$, $\log(\cdot)$, and hinge functions of the form $(x-b)_+$ similar to Multivariate Adaptive Regression Splines (MARS). The best NMSE achieved by FFX was 0.06198, with an execution time of approximately 1 hour.

D. All Results

Finally, we present the complete results of the experiment in Figure 4. Although we observed the lowest NMSE with FFX, NSGA-II generally dominated the performance of the other two methods, especially for small numbers of retained features. The variance of the target specification performance was generally explainable with a linear model, which mitigated the benefits of FFX versus NSGA-II. For specifications where

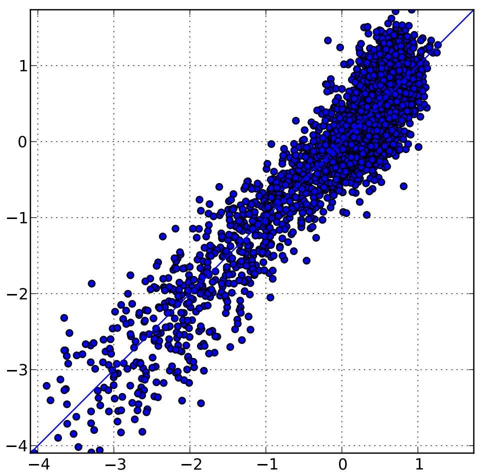

Fig. 3. FFX: $\hat{P}_i$ vs. P_i

the relationship is more complex/non-linear, FFX will likely dominate NSGA-II.

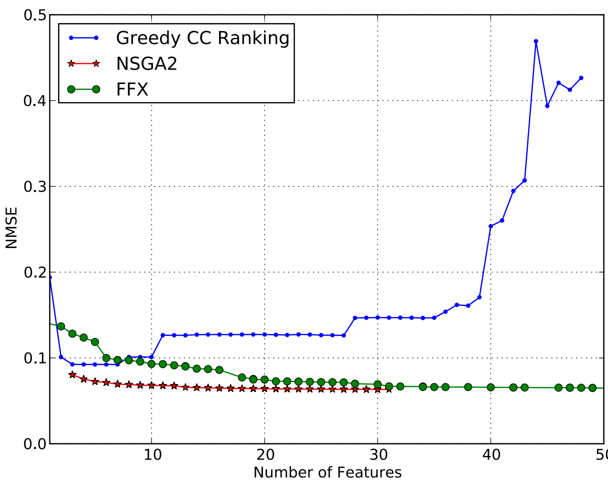

Fig. 4. All Methods: NMSE vs. # Features Retained

IV. Conclusion and Future Work

In this work, we have presented a case study of several feature selection methods for low-cost test. As noted previously, feature selection is a key step in deploying low-cost testing, and simply ad-hoc choice of features is likely inadequate to ensure good performance. As we have demonstrated, even ranking features based on correlation coefficients can have poor performance.

In the future we plan to extend this case study by examining additional feature selection methods, and include a more detailed analysis of each approach. In particular, we would like to establish confidence intervals on the performance of each feature selection technique. We would also like to explore datasets with more complex regression modeling problems, to validate our hypothesis that FFX will outperform NSGA-II in such cases.

Acknowledgement

This research has been carried out with the support of the National Science Foundation (NSF CCF-0916803 and CCF-0916415) and the Semiconductor Research Corporation (SRC-1836.073). The first author is supported by an IBM/GRC (Global Research Collaboration) graduate fellowship. The authors would also like to thank Elida de-Obaldia, John Carulli, and Ken Butler of Texas Instruments for their provision of (and technical support for) the data.

References

[1] J. Ferrario, R. Wolf, S. Moss, and M. Slamani, "A low-cost test solution for wireless phone RFICs," *IEEE Communications Magazine*, vol. 41, no. 9, pp. 82–88, 2003.

[2] A. Abdennadher and S. A. Shaikh, "Practices in mixed-signal and RF IC testing," *IEEE Design & Test of Computers*, vol. 24, no. 4, pp. 332–339, 2007.

[3] S. Bhattacharya and A. Chatterjee, "A DFT approach for testing embedded systems using DC sensors," *IEEE Design & Test of Computers*, vol. 23, no. 6, pp. 464–475, 2006.

[4] L. Rolindez, S. Mir, J.-L. Carbonero, D. Goguet, and N. Chouba, "A stereo $\Sigma\Delta$ ADC architecture with embedded SNDR self-test," in *IEEE International Test Conference*, 2007, Paper 32.1.

[5] M. Cimino, H. Lapuyade, Y. Deval, T. Taris, and J.-B. Bégueret, "Design of a 0.9V 2.45 GHz self-testable and reliability-enhanced CMOS LNA," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1187–1194, 2008.

[6] D. Mannath, D. Webster, V. Montano-Martinez, D. Cohen, S. Kush, T. Ganesan, and A. Sontakke, "Structural approach for built-in tests in RF devices," in *IEEE International Test Conference*, 2010, Paper 14.1.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] N. Kupp, , M. Slamani, and Y. Makris, "Correlating inline data with final test outcomes in analog/RF devices," *Design, Automation & Test in Europe*, 2011.

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

[14] E. Yilmaz and S. Ozev, "Adaptive test elimination for analog/RF circuits," in *Proceedings of the 46th Annual Design Automation Conference*, 2009, pp. 720–725.

[15] H.-G. D. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris, "RF specification test compaction using learning machines," *IEEE Transactions on Very Large Scale Integration*, vol. 18, no. 6, pp. 1002–1006, 2010.

[16] T. McConaghy, *Genetic Programming Theory and Practice IX*, chapter FFX: Fast, Scalable, Deterministic Symbolic Regression Technology, Springer, 2011.

[17] Hui Zou and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society*, vol. 67, no. 2, pp. 301–320, 2005.