

# Feature selection for Alternate Test using wrappers: application to an RF LNA case study

Manuel J. Barragan

CNRS, TIMA, F-38000 Grenoble, France

Université Grenoble Alpes, TIMA, F-38000 Grenoble, France

e-mail: manuel.barragan@imag.fr

Gildas Leger

Instituto de Microelectrónica de Sevilla

CSIC-Universidad de Sevilla

Av. Américo Vespucio s/n, 41092 Sevilla, Spain.

e-mail: leger@imse-cnm.csic.es

**Abstract**—Testing analog, mixed-signal and RF circuits represents the main cost component for testing complex SoCs. A promising solution to alleviate this cost is the Alternate Test strategy. Alternate test is an indirect test approach that replaces costly specification measurements by simpler signatures. Machine learning techniques are then used to map signatures and performances. One key point that still remains as an open problem is the conception of adequate simple measurement candidates. This work presents efficient algorithms for selecting information rich signatures.

## I. INTRODUCTION

Nowadays, commercial trends of IC industry have forced the integration of complex Systems-on-Chip (SoCs) consisting of tightly integrated analog, mixed-signal, RF and digital circuitry onto a single IC substrate. This high integration level provides a significant reduction in production cost, but there is a simultaneous increase in the cost of testing and diagnosing these devices.

Generic cost-effective methods for testing the digital parts of these SoCs, based on standardized fault models, are already available. But testing Analog, Mixed-Signal, and RF circuits still relies on costly functional characterization.

Alternate test [1] is a promising strategy for overcoming this issue. The basic idea is to replace costly specification-based tests by a set of simpler (and cheaper) measurements, often called signatures, and then use machine learning algorithms to map these simple measurements onto the specification space.

The usual approach to perform this replacement is based on supervised machine learning algorithms. The process is developed in two stages: a learning stage, and a testing stage. During the learning stage both performance parameters and signatures are measured from a set of training devices. A machine learning algorithm is then trained over the two sets of measurements to build a mapping model. In the testing stage, signatures are measured for each Device Under Test (DUT), and performances are inferred by using the mapping model obtained in the previous stage.

However, finding an appropriate set of signatures to extract meaningful models is still a matter of creativity based on a precise knowledge of the DUT. Thus, signature sets are usually *ad hoc* and sub-optimum: they may contain redundant information, non-relevant data, noisy signatures, etc. These redundant or unreliable signatures should be removed from the

test program before deploying it. In our opinion, the definition of the input space of signatures is one of the key aspects that limits the adoption of Alternate Test as a methodology.

Some work has been presented on optimization and selection of signatures [2]–[6]. The most direct approach for selecting relevant signatures is to perform an iterative search. That is the procedure used in [2], [3]. A typical iterative search in a signature space starts by building machine-learning models for each individual signature. Then the signature that yields the best model is selected and the search is extended to all possible pairs of signatures that include the previously selected one. The process is then repeated and more signatures are added until some precision criterion is achieved. The works in [4], [5] also perform a similar iterative search, but in these cases signatures are first ranked based on a statistical metric. The proposal in [6], on the other hand, relies on training a genetic algorithm to guide signature selection.

This paper presents a generic methodology for guiding the selection of meaningful signatures to feed the statistical tools. The early identification of a small subset of appropriate signatures from a vast signature set enables test compaction, which leads directly to reduced test time, reduced interface needs, relaxed equipment requirements or even to a simplification of the DUT design due to not including some built-in test sensors.

The paper is organized as follows: Section II presents our proposed methodology for feature selection. Section III presents some relevant experimental results from a direct application to an alternate test strategy for RF LNAs. Finally, Section IV summarizes the main contributions of this work.

## II. FEATURE SELECTION

Feature selection can be defined as the process of selecting a subset of relevant features for use in model construction. It is a recurrent problem in machine learning, and has been addressed by numerous researchers. It is out of the scope of this paper to produce a full review, and interested readers can refer to [7] for an excellent introduction to the topic.

Statistics almost always address feature selection from the viewpoint of overfitting reduction. In the particular application case of IC testing, though, there is a specific additional concern: the feature cost. Any additional signature is a measurement that has to be performed in production.

### A. The filtering approach

The most direct approach to feature selection consists in pre-selecting a subset of features, based on some statistical observations, before training any regression models. This approach, widely used when the number of initial features is very high, is known as filtering.

Filtering can be unsupervised, that is to say that it uses only the signature information and not the specification information, or supervised, that uses the specification information to select the best signatures. The most well-known unsupervised filtering approach is the Principal Component Analysis (PCA). PCA considers a linear input space and performs an eigenvalue decomposition of the features covariance matrix to identify the principal directions, or Principal Components (PCs), of the variability in the space of signatures.

One of the main drawbacks of PCA is the fact that it is a linear method, while the relation between signatures may be non-linear. Instead of this linear unsupervised approach for signature selection, in this paper we propose a methodology that takes into account the intrinsic non-linearity of the space of signatures and the non-linear mapping to the performance space by incorporating model training to the selection algorithm. This approach is known as a wrapper approach [8].

### B. The wrapper approach

The wrapper approach consists in using the machine-learning prediction model as a black box within an optimization loop. The model is used to evaluate the prediction error (also called generalization error) for a given signature subset, and the optimization loop tries to minimize this error. More complex cost functions that take into account the test cost of each signature are also possible, but in this paper we are not interested in cost function modeling and we will focus on generalization error minimization. We will consider the number of features in a more qualitative manner.

If the number of variables is small, a full-search would be feasible. However, as dimensionality rises more clever approaches shall be implemented.

We opted for an optimization that is simple and easy to interpret: the stepwise search with compound operators, as described in [8]. This approach is based on a combination of two operations: stepwise forward addition and backward elimination of features. Starting from an initial subset of signatures, we explore all the possible children obtained by removing one existing signature or adding one new signature. Up to this point, this is very similar to what was proposed in [9]. But at the end of this one-change exploration, we also explore the combinations of the four children that have given best results. This can significantly speed-up the exploration in the early stages of the search.

Concerning the machine-learning model in the wrapper algorithm, we use a perceptron neural-network from the EN-TOOL MATLAB toolbox for ensemble learning [10]. The available data is split in 15 random fractions and 15 different models realizations are obtained by leaving one of the 15 fractions out of the actual training data. This left-out fraction

is used as a test set to derive the generalization error, in a classical cross-validation fashion. At the end of this process, an ensemble model is created: the 9 best performing models are selected and their outcome is averaged to get the ensemble model output. Apart from the total generalization error, we get access to the 15 individual errors. In this way, we can take the standard deviation of these errors to get a rough estimate of the confidence in the obtained generalization error.

As an arbitrary stopping criterion for the search, we consider that the generalization improvement for a new iteration should be higher than 25% of the standard deviation estimate. An important remark in this strategy, as in any other optimization algorithm, is that it is important to set aside a validation/verification set that is not involved in the optimization loop. We need an out-of-the-loop criterion to ensure/verify that we are not overfitting. We implemented a sanity check in this line consisting in monitoring the model performance on the out-of-the loop validation set, to verify that the prediction error for this set also improves.

### C. Hybrid approach

It is also possible to combine filtering and wrapper approaches to obtain a trade-off between speed and precision. In order to do that, we have to define a feature ranking criterion based on a filtering approach. In principle, it makes little sense to consider unsupervised candidates, since the training data is available. Instead of that, we propose to use as a simple metric for our ranking the correlation of the features with the specifications. Different definitions of this correlation can be used, such as Pearson's correlation, Brownian distance correlation, etc.

The proposed hybrid approach first ranks the features with respect to the original data and trains an initial model (hence using the wrapper approach). Then, we compute the residues of the fit (i.e. the individual errors for all the training samples). After that, the algorithm ranks the remaining features with respect to the residues and select the best candidate to add it to the feature set. Iterating the procedure, the algorithm should add features that add information that was previously missing.

Since only one model is trained for each new feature, a complete search may be feasible (though model training time increases with the number of features). In any case, a stopping criterion based on a minimum stepwise improvement can also be implemented.

## III. CASE STUDY: ALTERNATE TEST OF RF LNAs

### A. The Device Under Test

In order to exemplify the application of the proposed technique for feature selection, we have applied it to an Alternate Test strategy for predicting the gain of an RF LNA. LNAs are simple circuits with very few components, but that are quite sensitive to process variability. Both these aspects are valuable for our purpose since we need to perform many Monte Carlo simulations in order to build the training and testing sets of signatures and specifications. The variability is important to make the prediction more challenging. On

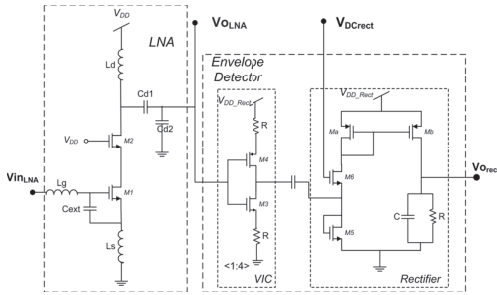


Fig. 1. Schematic view of the LNA with envelope detector

the other hand, limiting the discussion to the prediction of a single performance –the LNA’s gain in this case– facilitates the interpretation of the results, although the method is easily applicable to any number of target performances.

Figure 1 shows the schematic of the LNA, which has been designed in a 90nm CMOS technology. The envelope detector at the output of the LNA has been included as a built in test instrument [11].

### B. Simulation results

Our initial set of signatures in this case study contains 42 signatures, consisting in the DC voltages in all the nodes of the LNA, the output of the envelope detector, and all the previous signatures measured under power supply stress. The selection of these sets of signatures is based on our previous work on RF Alternate Test [5], [11], [12]. We have already demonstrated that these sets of signatures can be used separately for inferring RF performances. In this work, our goal is to optimize the complete set of signatures by using feature selection techniques.

A population of 2000 instances of the LNA was generated using Monte Carlo simulation and the complete set of 42 signatures was extracted for each instance to define our input space.

In order to explore the structure of the input space, we performed a PCA decomposition of this space. This decomposition allows us to rank our 42 signatures according to their contributions to the principal components. In this sense, Fig. 2 shows the Euclidean norm of the contributions of each signature to the principal components, for two truncated PCA decompositions: considering only the three most significant PCs (dashed-red line), and considering the ten most significant ones (solid-blue line).

It is quite clear that the main contributing signatures are roughly signatures 32 to 42, but an apparent result is that, even in the case of a 3 component decomposition, all the 42 signatures have significant contributions to the PCs, so a filtering decision turns out to be difficult if we base the choice only in the information offered by the PCA decomposition. On the other hand, our proposed wrapper approach brings additional information for the filtering decision by using machine learning to retrieve the non-linear relationship between the set of signatures and the target performance.

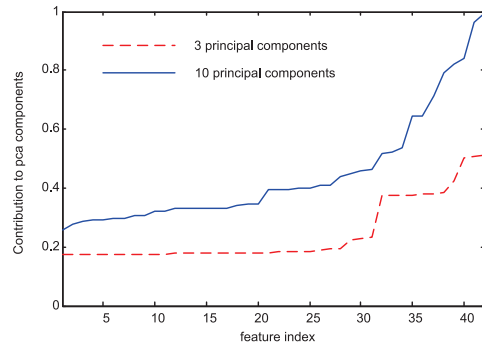


Fig. 2. Normalized contributions of the individual signatures to the Principal Components

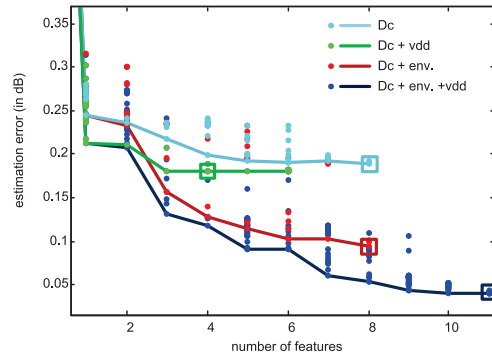


Fig. 3. Generalization error versus number of features, for all of the visited cases of the four different training scenarios

Thus, as a first application of the proposed wrapper feature selection method, we consider four different scenarios to predict the gain of the LNA:

- only DC signatures
- DC signatures and envelope detector
- DC signatures with power supply stress
- DC signatures and envelope detector with power supply stress

Again, we use the 2000 instances of the LNA generated using Monte Carlo simulation. In the four scenarios, we set aside 10% of the data for the verification set which represents 200 samples. We are left with 1800 samples for training and the cross-validation scheme leads to 15 partitions of 120 samples.

Figure 3 shows all the visited feature combinations for the four scenarios, in a scatterplot of the generalization error versus the number of features in the training set. It can be observed how the use of power supply stress to augment the information fails to improve significantly the model if only DC signatures are considered. However, this is not true when the envelope detector is considered. In this case, the gain error goes from 0.09dB to 0.04dB, approximately.

The optimum fronts for the different scenarios also provide interesting information. For instance, in the case of DC signatures with supply stress, the model with lowest error uses four

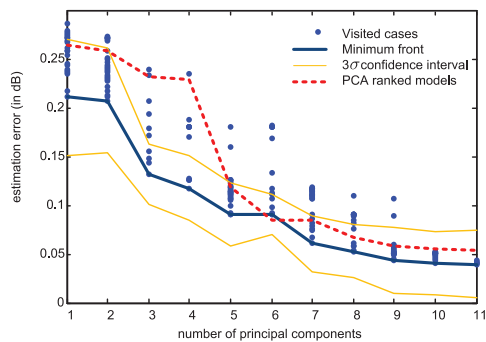


Fig. 4. Generalization error versus number of Principal Components used for model training

features, but it is clear that a model with three features gives almost the same prediction error, so the additional feature is probably not cost-effective. Similarly, for the most complete scenario the best model (with an error of 0.04dB) uses 11 features. However, the last steep improvement occurs when going from 6 to 7 features, with an error of 0.06dB. Is an error improvement of 0.02dB worth the introduction of 4 features? This is a matter of cost optimization, as commented above.

To further explore the benefits of the wrapper approach, we can make a comparison to the PCA decomposition. We considered the set of the 11 most significant principal components as our set of signatures for training machine learning models. Thus, Fig. 4 shows a scatterplot of the generalization error versus the number of PCs used in the training. Intuitively, it could be expected that the PCA decomposition would be competitive compared to the individual signatures, but as it is shown in Figures 3 and 4, that is not the case even for models trained with a low number of features. If we take into account that each principal component has contributions of all the 42 individual signatures, the advantages of the proposed feature selection technique based on the wrapper method is clear. Moreover, it is interesting to observe that the best PCA ranked models (red-dashed line in Fig. 4) are far from the minimum front for the estimation error (solid-blue line in Fig. 4).

Finally, Figure 5 displays the results obtained from three hybrid approaches using feature ranking on residues as explained above. For the first approach we use the traditional Pearson's correlation as a ranking method. For the second one, we use the individual Brownian distance correlation as introduced in [5]. And for the third one, we use the multivariate Brownian distance correlation, that is to say that the distance computation is performed in the complete feature space.

It can be seen that the multivariate distance achieves very good results compared to the wrapper search reference. It can even be seen that the optimum front of the wrapper approach is outperformed for some feature numbers. However, this may be explained by the fact that we introduced the compound operators and thus may have "skipped" some intermediate cases. Anyhow, this result suggests that the hybrid approach with multivariate distance correlation may be a good option if

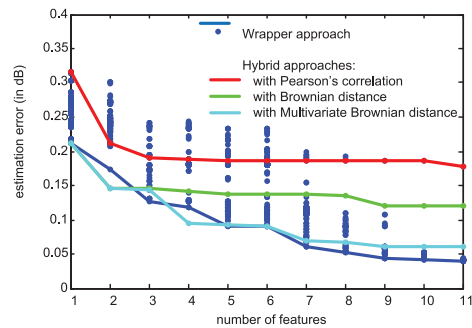


Fig. 5. Generalization error versus number of features, for all of the visited cases of the four different training scenarios

feature selection speed is critical.

#### IV. CONCLUSIONS

This work proposes two simple methodologies for the selection of information rich features. The first one is based on a wrapper technique, while the second one combines the wrapper with filtering decisions based on non-linear correlation coefficients. Both approaches have been validated in a RF LNA case study to demonstrate their feasibility. Obtained results show a clear potential of the proposed techniques to improve the quality of the input space of signatures.

#### ACKNOWLEDGMENT

This work has been partially funded by the Spanish Government project TEC2011-28302, co-funded by FEDER program.

#### REFERENCES

- [1] P. N. Variyam and A. Chatterjee, "Enhancing test effectiveness for analog circuits using synthesized measurements," in *VLSI Test Symposium*, 1998, pp. 132–137.
- [2] S. Bhattacharya and A. Chatterjee, "Use of embedded sensors for built-in-test RF circuits," *International Test Conference*, pp. 801–809, 2004.
- [3] L. Abdallah *et al.*, "Sensors for built-in alternate RF test," in *European Test Symposium*, 2010, pp. 49–54.
- [4] H. Ayari *et al.*, "Smart selection of indirect parameters for DC-based alternate RF IC testing," *VLSI Test Symposium*, pp. 19–24, 2012.
- [5] M. J. Barragan and G. Leger, "Efficient selection of signatures for analog/RF alternate test," in *European Test Symposium*, May 2013, pp. 1–6.
- [6] H. G. Stratigopoulos *et al.*, "RF Specification Test Compaction Using Learning Machines," *Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 6, pp. 998–1002, June 2010.
- [7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, Mar. 2003. [Online]. Available: <http://dl.acm.org/citation.cfm?id=944919.944968>
- [8] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1, 1997.
- [9] H.-G. Stratigopoulos and Y. Makris, "Nonlinear decision boundaries for testing analog circuits," *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 11, pp. 1760–1773, 2005.
- [10] J. D. Wichard, M. J. Ogorzalek, and C. Merkwirth, "Entool-a toolbox for ensemble modelling," in *Europhysics Conference Abstracts ECA*, vol. 27, 2003.
- [11] M. J. Barragan *et al.*, "Alternate test of Inas through ensemble learning of on-chip digital envelope signatures," *Journal of Electronic Testing*, vol. 27, no. 3, pp. 277–288, 2011.
- [12] M. J. Barragan *et al.*, "Improving the accuracy of rf alternate test using multi-vdd conditions: Application to envelope-based test of Inas," *Asian Test Symposium*, pp. 359–364, 2011.