



HAL
open science

Feature selection and feature design for machine learning indirect test: a tutorial review

Manuel J. Barragan, G. Leger

► To cite this version:

Manuel J. Barragan, G. Leger. Feature selection and feature design for machine learning indirect test: a tutorial review. International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD 2019), Jul 2019, Lausanne, Switzerland. 10.1109/SMACD.2019.8795292 . hal-02165859

HAL Id: hal-02165859

<https://hal.science/hal-02165859v1>

Submitted on 2 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Feature selection and feature design for machine learning indirect test: a tutorial review

Manuel J. Barragan*, Gildas Leger†

*Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMA, F-38000, Grenoble, France

†Instituto de Microelectrónica de Sevilla, CSIC-Universidad de Sevilla, 41092, Seville, Spain

Abstract—Machine learning indirect test replaces costly specification measurements by simpler signatures and use modern learning algorithms to map these signatures to specifications. Defining a set of relevant signatures that appropriately captures the circuit performance degradation mechanisms is then a key point for enabling machine learning indirect test. In this tutorial we review some methodologies for selecting and designing such a set of information rich signatures.

I. INTRODUCTION

Machine learning indirect test [1] has been used in the past decade to overcome the issues of traditional functional test of analog, mixed-signal and RF (AMS-RF) circuits. Extensive specification measurements are replaced by a set of simpler measurements, often called signatures or features, and machine-learning regression algorithms are then trained to map features to specifications.

A successful application of machine-learning indirect test relies on the definition of an appropriate set of signatures. Information-rich signatures that capture the major part of the circuit parametric variability are required. In practice, signatures are proposed based on expert knowledge but are usually suboptimal: they may contain redundant information, noisy signatures, non-relevant signatures, etc. The statistic field of feature selection comes at hand to clean up such an *ad hoc* signature set [2]. However, an obvious question remains unanswered: how can we propose the original set of features? Some generic approaches exist for providing an initial set of signatures, such as DC node probing, I_{DDQ} test, process monitoring [3]–[5], etc. Work has been also presented on optimizing a particular stimulus to define optimized signatures that minimize the prediction error [6], [7], but the choice and the parametrization of the input stimulus are still *ad hoc*. In essence, there is no guarantee that the resulting signatures may be actually relevant for the prediction of a given performance.

In this tutorial, we summarize our research on feature selection and design for indirect test applications. Throughout this tutorial, feature selection and design algorithms are targeted at minimizing the prediction error. It has to be noted that the presented algorithms can be easily extended to take into account other test criteria, e.g., the cost of measuring each feature. The rest of this paper is organized as follows. Section II presents methodologies for feature selection and design. Section III illustrate the practical application of these

This work has been partially funded by CNRS project IndieTEST PICS07703 and the Spanish Government project TEC2015-68448-R (co-funded by FEDER program).

methodologies to a case study. Finally, section IV summarizes the main contributions of this work.

II. THEORETICAL BASIS

A. Feature selection

Feature selection can be defined as the process of selecting a subset of relevant features for use in model construction. A variety of feature selection algorithms for indirect test have been explored in the recent years [8]–[14].

1) *The wrapper approach*: It considers feature selection as an optimization problem. Wrapper algorithms use the machine learning prediction model as a black box within an optimization loop. The model is used to evaluate the prediction error for a given signature subset, and the optimization loop minimizes this error. If the number of candidate features is small, a full search may be feasible, but as dimensionality rises more clever search strategies should be considered in the optimization loop. In this work we consider a simple wrapper implementation: the stepwise search with compound operators as described in [15]. 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. Then the search continues by exploring the combinations of the four children that give best results. This can significantly speed-up the exploration in the early stages of the search.

2) *The filtering approach*: Filtering algorithms rank a set of features according to some statistical criterion, before training any regression models. A wide variety of filters can be found in specialized statistics literature [2]. In this tutorial, we will compare two widely used strategies: correlation-based filters, and Principal Component Analysis (PCA).

Correlation to performance: A straightforward approach to supervised filtering would be to select the features that best correlate to the target specification. In [11], we proposed the use of Brownian distance correlation (dCor) to replace the standard Pearson's correlation. The distance correlation statistics is detailed in [16]. Compared to classical Pearson's correlation, the main advantage of Brownian distance correlation is its sensitivity to non-linear dependencies. Similarly, Kendall's Tau or rank correlation is also used to assess non-linear independence and has been used in the context of indirect test [12]. It tests the hypothesis that for two correlated variables, variations should evolve in the same direction. Another advantage of Brownian distance correlation is that

it is multidimensional. It can correlate an arbitrary subset of features to a given set of performances. This is not the case for Pearson's correlation and Kendall's Tau for which features are independently ranked, one by one, according to their correlation to one given performance.

Principal Component Analysis: The goal of PCA is to identify the directions in the feature space that best explain the variation observed in the data, supposing that the underlying structure in the data is linear. PCA performs an eigenvalue decomposition of the features covariance matrix. The Principal Components (PCs) are the eigenvectors that correspond to the largest eigenvalues. PCA does not necessarily reduce the number of features to be acquired. The PCs are linear combinations of the original features and they may require the complete set of features for their computation. In that case, features can be ranked according to their relative contributions to the PCs. Besides this practical aspect, one of the of the main conceptual drawbacks of PCA is the fact that it is a linear method, while the relation between features may be non-linear.

3) *The hybrid approach:* Filters and wrappers can be combined to obtain a trade-off between speed and precision. Our proposed hybrid approach first ranks the features with respect to their correlation to the target performance and trains an initial regression model with the best correlated feature. Then, we compute the residues of the fit (i.e. the regression errors for all the training samples). After that, the algorithm ranks the remaining features with respect to the correlation to the residues and select the best candidate to add it to the feature set. In each iteration the algorithm should add features that contains information that was previously missing. The idea behind this algorithm is that a given feature may explain the coarse behavior of the specification but fine details look like noise with respect to the main contribution. The search for new features is guided to the feature that best explains the prediction error. In that way, the next selected feature should always bring additional information.

B. Feature design

Feature design is aimed at proposing relevant features for a given machine learning application, either for completing the information in an initial feature set, or for proposing the initial set of features itself. In this tutorial we present two strategies in the context of machine learning indirect test applications.

1) *Stepwise addition of relevant features:* The first logical step for designing new signatures would be to identify the relevant information that is missing from a given set of signatures. The key idea behind indirect test is that signatures and specifications are tightly correlated because they are affected by the same underlying stochastic process: the fabrication process. Then, if process variations are the root cause of both signature and performance variations, they are the perfect candidates for diagnosis. In a real fabrication process we hardly have access to the underlying physical parameters. However, during the design stage we have access to these parameters at simulation level: this is the base of Monte Carlo (MC) process simulation. We propose to explore the set of MC

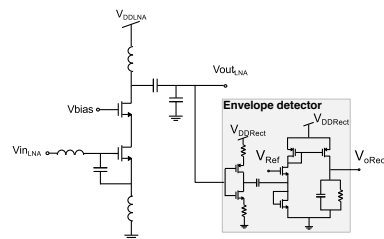


Fig. 1. Schematic view of the LNA with envelope detector

variables as if they were genuine signatures. Since the process parameters are independent random variables, we can simply perform stepwise addition. Starting from the best available signature subset determined by feature selection, we train a regression model and evaluate the model performance adding one MC process parameter at a time, include the best candidate in the test list, and iterate. This way, we can identify which MC variables bring relevant additional information with respect to the already available signatures. Finally, simple tests that target the selected MC parameters should be designed in order to add the identified missing information to the regression. Obviously, this requires design expertise but the diagnosis of the most important variation causes does bring significant information to guide the designer.

2) *Root cause analysis:* The previous algorithm can be generalized to the case where the initial set of signatures is empty. We can diagnose the main degradation causes by exploring the space of MC parameters, and then we can design appropriate signatures that target the identified degradation mechanisms. However, the space of MC parameters in current nanometric technologies is a high-dimensionality space so a full stepwise search may not be feasible. In this tutorial we introduce an efficient alternative for this exploration: the causal discovery analysis. Causal discovery algorithms, in contrast to previous feature selection algorithms, do not target to maximize the correlation between features and target variables. Instead, they are aimed at proposing a causal structure that represents the dynamic of the system and then identifying the minimum number of parameters that causally explain the state of a target variable (what is known as the Markov blanket of a variable). In this line, causal discovery algorithms have been proposed for inferring causal relationships from sets of observational data [17]. These algorithms identify the causal structure around the target variable and then they test conditional independence and dependence hypothesis between the identified variables to determine the Markov blanket of the target variable. Once the MC parameters in the Markov blanket of a given performance have been identified, simple tests targeting these parameters should be designed. This still requires design knowledge, but the identification of the performance degradation root causes enables a systematic design approach.

III. CASE STUDY

This tutorial illustrates the application of the described techniques to a 2.4 GHz Low Noise Amplifier (LNA) case study. Figure 1 shows the schematic of the LNA, which has been designed in a 90 nm CMOS technology. The envelope

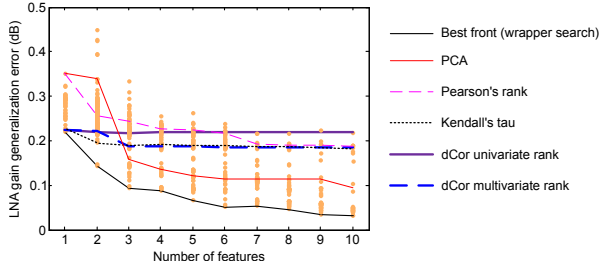


Fig. 2. Generalization error for the prediction of LNA gain: Fronts obtained from different feature filtering methods and scatterplot of the explored space through wrapper search.

detector at the output of the LNA has been included as a built-in test instrument [18].

Our initial set of signatures contains 42 features: the DC voltages in all the nodes of the LNA, the biasing current, and the output of the envelope detector. Additionally, all the features were measured under nominal and stressed power supplies since this may bring additional information. A population of 2000 instances of the LNA were generated using Monte Carlo simulation and 500 of them were set aside as validation set. In this case study we use a perceptron neural network in MATLAB to predict the gain of the LNA under test.

A. Comparison of filtering algorithms

For a direct comparison, the initial set of 42 features has been ranked according to: a) Pearson's correlation, b) Kendall's Tau, c) Univariate Brownian distance correlation, d) Multivariate Brownian distance correlation, and e) Principal Component Analysis (notice that in this case, all the features are actually used to compute each principal component).

Figure 2 displays the obtained results –up to a dimension of 10– together with the scatterplot of a wrapper search for reference. The generalization error is computed on the independent validation set. It appears that Kendall's Tau and Brownian distance correlation (both univariate and multivariate) accurately select the best first feature but fail to correctly identify the best additional features. Unsurprisingly, Pearson's correlation even fails to identify the best first feature, possibly due to non-linearity. For the PCA the X-axis in Fig. 2 must be understood as the number of PCs (which are linear combinations of all the features). Surprisingly, models trained with PCs are worse than the models trained with the best set of genuine features. Models trained with the first or the first two PCs are even worse than those produced by the studied filtering methods. For a dimensionality above 3, PCA-based models never reach the optimum front

B. Directed search with different correlation metrics

In this tutorial we explore the implementation of hybrid feature selection algorithms using different correlation metrics, namely, Pearson's correlation, Kendall's tau, and Brownian distance correlation.

Figure 3 shows the obtained fronts for our RF LNA case study. As expected, the algorithm based on Pearson's correlation performs much worse than the others. More surprisingly,

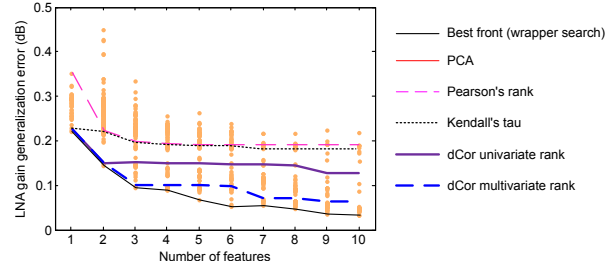


Fig. 3. Generalization error for the prediction of LNA gain: Fronts obtained for correlation-directed search with different correlations and scatterplot of the explored space through wrapper approach

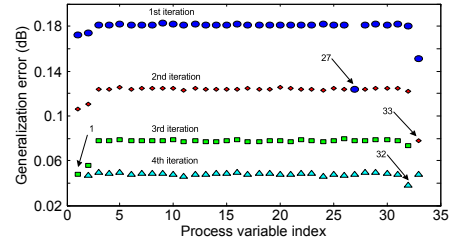


Fig. 4. Generalization error for stepwise addition of process parameters, starting from the DC signatures with supply stress.

the algorithm based on Kendall's Tau performs poorly. It manages to identify the first best feature but fails to find the appropriate ones in further iterations. This is probably because, conceptually, Kendall's tau is closer to an independence test than to a correlation metric. The version with univariate Brownian distance correlation works well initially but appears to saturate for models that use more than 3 features.

Noticeably, the directed search based on the multivariate Brownian distance correlation performs extremely well compared with the optimum wrapper search. Moreover, In terms of computation time, it offers a significant advantage with respect to the wrapper. Each iteration adds a relevant feature and requires to train a single regression model. For reference, in our case study the directed search took less than one minute to find the 10 most relevant features, while the wrapper search took about 5 hours to get an equivalent result.

C. Stepwise addition of relevant features

In order to illustrate the proposed procedure, we consider the regression of the RF LNA gain using as the initial set of features the set of DC voltages at all the nodes of the LNA, measured with and without power supply stress (a total of 32 features). A wrapper search reveals that the optimum set of features contains only 4 features and results in a generalization error around 0.2 dB in the prediction of the LNA gain. Improving this prediction error would require to design new features that bring new relevant information.

Following the proposed methodology, we perform a stepwise addition search using the 33 process variation parameters defined in the technology. Figure 4 shows the generalization errors obtained by adding one process parameter, for four iterations. At the fourth iteration, we see that the improvement is marginal. As a result, we identify three Monte Carlo

TABLE I
MARKOV BLANKET INFERENCE VS HYBRID FEATURE SELECTION IN THE
SPACE OF MONTE CARLO MODEL PARAMETERS

Markov blanket of LNA gain (electrical effect in PDK models)	Feature selection for LNA gain regression (electrical effect in PDK models)
MC ₁₀ (MOS operation point)	MC ₁₀ (MOS operation point)
MC ₁₁ (MOS operation point)	MC ₁₂ (MOS operation point)
MC ₁₂ (MOS operation point)	MC ₁₁ (MOS operation point)
MC ₁ (Inductance variation)	MC ₂₇ (Capacitance variation)
MC ₁₃ (MOS operation point)	MC ₁ (Inductance variation)
MC ₂₇ (Capacitance variation)	MC ₃₃ (Ohmic losses in passive components)
MC ₃₃ (Ohmic losses in passive components)	MC ₁₃ (MOS operation point)
MC ₃₂ (Ohmic losses in passive components)	MC ₃₂ (Ohmic losses in passive components)
MC ₂ (Inductance variation)	MC ₂ (Inductance variation)

process variables –labelled 1, 27, and 33 in Fig. 4– that bring relevant additional information. By tracing back these variables to the physical models in the design kit, we can identify metal-insulator-metal capacitors, polysilicon resistors, and RF inductors as non-modeled information. This result is coherent with an electrical analysis of the DUT: the initial set of signatures was composed by DC levels, so parametric variations of capacitors and inductors would not have been detected. This information can be used by the designer to develop new tests targeting these elements, for instance by adding non-intrusive process control monitors that target the identified parameters, as proposed in [4], [5], [8].

D. Root cause analysis

Let us consider again the prediction of the LNA gain, but without an initial set of signatures. As it was discussed above, exploring the space of MC parameters may guide the design of appropriate test features by determining the relevant MC parameters responsible for performance degradation. As an example, we propose to use causality discovery tools to explore the MC parameters space and find the Markov blanket of the LNA gain. For this experiment, we use the Incremental Association-Based Markov Blanket (IAMB) algorithm included in the Causal Explorer Toolbox [17].

The first column of Table I lists the Monte Carlo process parameters in the Markov blanket of the LNA gain. The Markov Blanket is composed of only 9 parameters from the total 33 candidates. For validation, a perceptron neural network model was trained for regressing the LNA gain from the identified 9 parameters. The generalization error in the validation set is of only 0.02 dB, which validates the relevance of the causal features. The second column of Table I shows also the 9 most relevant parameters as selected by distance correlation directed search in the same space [9]. As it could be expected, the obtained sets of parameters are almost identical (only the ranking order of some of the features is interchanged) for both causal and non-causal selection algorithms which further validates the results of the proposed causal analysis. Although the results seem equivalent, it should be remarked that the causal discovery algorithm is a supervised filtering algorithm, while the hybrid feature selection uses a wrapper. Causal discovery does not need model training, which saves significant processing resources and opens the door to high-dimensionality problems.

Additionally, given the topological simplicity of the case study, it is possible to interpret the obtained results from an electrical point of view. The parameters in the Markov blanket of the gain are related to the impedance of MOM capacitors and inductors, ohmic losses, and the operation point of the transistors. These causal relationships were obtained automatically, without the need of previous design knowledge of the LNA. Again, this information can be used by the designer to develop new tests targeting these elements.

IV. CONCLUSIONS

This tutorial presents a variety of feature selection and feature design algorithms in the context of AMS-RF machine learning indirect test. The basic concepts behind filters, wrappers and hybrid feature selection techniques have been laid out. Moreover, we have detailed methodologies for designing relevant features based on a search in the MC parameters space. Application examples have been presented using a 2.4 GHz LNA in a 90 nm CMOS technology.

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] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, Mar. 2003.
- [3] A. Zjajo, M. J. Barragan, and J. P. de Gyvez, “BIST Method for Die-Level Process Parameter Variation Monitoring in Analog/Mixed-Signal Integrated Circuits,” in *Design Automation and Test in Europe*, 2007.
- [4] L. Abdallah *et al.*, “Experiences with non-intrusive sensors for RF built-in test,” in *Proc. International Test Conference*, 2012.
- [5] F. Cilici *et al.*, “Assisted test design for non-intrusive machine learning indirect test of millimeter-wave circuits,” in *European Test Symposium*, May 2018, pp. 1–6.
- [6] R. Voorakaranam *et al.*, “Signature testing of analog and RF circuits: Algorithms and methodology,” *Trans. on Circ. and Syst. I: Regular Papers*, vol. 54, no. 5, pp. 1018–1031, 2007.
- [7] A. Halder *et al.*, “Automatic multitone alternate test generation for RF circuits using behavioral models,” in *Int. Test Conference*, 2003.
- [8] M. J. Barragan and G. Leger, “A procedure for alternate test feature design and selection,” *Design & Test*, vol. 32, no. 1, pp. 18–25, 2015.
- [9] G. Leger and M. J. Barragan, “Brownian distance correlation-directed search: A fast feature selection technique for alternate test,” *Integration, the VLSI Journal*, vol. 55, pp. 401–414, 2016.
- [10] H.-G. Stratigopoulos *et al.*, “RF Specification Test Compaction Using Learning Machines,” *Trans. on VLSI Syst.*, vol. 18, no. 6, pp. 998–1002, June 2010.
- [11] M. Barragan and G. Leger, “Efficient selection of signatures for analog/RF alternate test,” in *European Test Symposium*, May 2013, pp. 1–6.
- [12] A. Gomez-Pau *et al.*, “Efficient Production Binning Using Octree Tessellation in the Alternate Measurements Space,” *Trans. on Computer-Aided Design of Int. Circ. and Syst.*, vol. 35, no. 8, pp. 1386–1395, 2016.
- [13] H. Ayari *et al.*, “Smart selection of indirect parameters for DC-based alternate RF IC testing,” *VLSI Test Symposium*, pp. 19–24, 2012.
- [14] S. Larguech *et al.*, “Efficiency evaluation of analog/rf alternate test: Comparative study of indirect measurement selection strategies,” *Microelectronics Journal*, vol. 46, no. 11, pp. 1091–1102, 2015.
- [15] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial intelligence*, vol. 97, no. 1, 1997.
- [16] G. J. Szekely, “Brownian distance covariance,” *The Annals of Applied Statistics*, vol. 3, no. 4, pp. 1236–1265, Dec. 2009.
- [17] C. F. Aliferis *et al.*, “Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery,” in *Int. Conf. on Mathematics and Engineering Techniques in Medicine and Biological Sciences, METMBS*, vol. 3, 2003, pp. 371–376.
- [18] M. J. Barragan *et al.*, “Alternate Test of LNAs Through Ensemble Learning of On-Chip Digital Envelope Signatures,” *J. of Electronic Testing*, vol. 27, no. 3, pp. 277–288, 2011.