Enhancing Confidence in Indirect Analog/RF Testing against the Lack of Correlation between Regular Parameters and Indirect Measurements
Haithem Ayari, Florence Azaïs, Serge Bernard, Mariane Comte, Vincent Kerzérho, Michel Renovell

To cite this version:

HAL Id: lirmm-00936443
https://hal-lirmm.ccsd.cnrs.fr/lirmm-00936443
Submitted on 27 Jan 2014
Enhancing Confidence in Indirect Analog/RF Testing against the Lack of Correlation between Regular Parameters and Indirect Measurements

Haithem Ayari, Florence Azaïs, Serge Bernard, Mariane Comte, Vincent Kerzérho and Michel Renovell
LIRMM - CNRS/Univ. Montpellier 2
161 rue Ada, 34392 Montpellier, France
{name}@lirmm.fr

Abstract—The greedy specification testing remains mandatory for analog and Radio Frequency (RF) integrated circuits because of the accuracy of the sorting based on these measurements. Unfortunately, to be implemented, this kind of testing method often incurs very high costs (expensive instruments, long test time...). A common approach, in the literature, is the so-called indirect/alternate test strategy. This strategy consists in deriving targeted specifications from low-cost Indirect Measurements (IMs). During the industrial test phase, the estimation of regular specifications using IMs is based on a correlation model that has been built previously, during a training phase. Despite the substantial test cost reduction offered by this strategy, its deployment in industry is limited, mainly because of a lack of confidence in the accuracy of estimations made by the correlation model. A solution to increase the confidence in the estimation of specifications using the indirect approach is to implement redundancy in the prediction phase. In this paper, we demonstrate that the redundancy implementation brings more than identifying rare misjudged circuits from a high-correlated model. Indeed redundancy massively increases the accuracy despite of the lack of accurate models that have been assumed in previous implementations of redundant indirect testing. This approach is illustrated on a real case study for which we have experimental measurements on a set of 10,000 devices.

Keywords- Test; Analog/RF integrated circuits, alternate test, indirect test, prediction error, machine-learning algorithm

I. INTRODUCTION

The efficiency of a mixed-signal electronics system relies on the performances of all the components. The global performance of such a system is close to the one of the component in the path with the lowest performance. Electronics system manufacturer also called in the industry Original Equipment Manufacturer (OEM) builds a system on the base of performances given in the datasheet of the components. As a consequence for the analog/RF Integrated Circuit (IC) provider, to ensure the confidence of the OEM in the performances of the components, it is mandatory to test these specifications after manufacturing. Historically, testing analog/RF integrated circuits aims at measuring the majority or totality of the circuit performances/specifications defined in its datasheet. Thanks to testing, the manufacturer is able to sort the manufactured circuits as good or bad regarding its datasheet. According to 2011 edition of the International Technology Roadmap for
Semiconductors (ITRS) report [1] on test and test equipment, the test cost for analog and mixed-signal is a great concern. “These trends of increasing ATE (Automated Test Equipment) instrument channel count, complexity, and performance are expected to continue, but at the same time the cost of test must be driven lower”.

In order to lessen the burden of specification testing, a promising solution is the so-called indirect/alternate testing, in which the results of specification testing are derived from a set of few Indirect Measurements (IMs) obtained with low-cost test equipment. The basic principle is, during a first phase, to use a limited set of devices in order to train a non-linear regression algorithm used to learn the mapping between some indirect measurements and the circuit performance parameters. For this first phase, it is mandatory to measure the circuit performance parameters and the IMs. The second phase is called the production testing phase. It aims at measuring the performances of a huge lot of new devices. Then during this phase, only the low-cost IMs are performed and used to feed the previously-trained regression algorithm in order to predict device specification or to classify the devices as good or bad. As a consequence thanks to indirect testing, it is possible to significantly decrease the number and complexity of measurement configurations. Despite the clear advantages of employing such approach and a number of convincing attempts to prove its efficiency [2-21], the deployment of the indirect test strategy in industry is limited. A solution to enhance the confidence in the test method is to implement a strengthened strategy based on prediction model redundancy [21]. According to [21], the combination of the predictions of three models with strong correlations between IMs and specifications, allows chasing the rare misjudged circuits. These rare misjudged circuits are commonly highlighted as the source of lack of confidence in the indirect strategy. The purpose of the publication is to demonstrate a new advantage of implementing redundancy into indirect test strategy. Indeed more than chasing rare misjudged circuits, redundancy can be efficient in the case that there are only available weak models in terms of correlation between IMs and specifications.

This paper is organized as follows. In section II, the general background of indirect testing is described and previous works are summarized. Section III introduces the problem of the evaluation of indirect test quality and shows that although a good average prediction error can be achieved, the maximal prediction error cannot be guaranteed. Section IV presents the proposed strategy to reduce the maximal prediction error based on the use of redundancy and a two-tier test scheme. This strategy is evaluated in section V on a real case study. Finally, section VI concludes the work.
II. BACKGROUND

A. Indirect/alternate test principle

The purpose of indirect testing is to relax constraints on the number and complexity of industrial test configurations by using low-cost indirect measurements in order to infer the results of specification testing. As presented by figure 1, the underlying idea of indirect testing is that manufacturing process variability that affect the conventional performance parameters of the device also affect non-conventional low-cost indirect parameters. If the correlation between the indirect parameter space and the performance parameter space can be established, then specifications may be verified using only the low-cost indirect signatures. Unfortunately the relation between these two sets of parameters is complex and cannot be simply identified with an analytic function. The solution commonly implemented uses the computing power of machine-learning algorithms.

![Figure 1. Underlying idea of indirect/alternate testing.](image)

The indirect test principle is split into two sequential steps, namely training and production testing phases. The underlying idea is to learn during the training phase the unknown dependency between the low-cost IMs and the conventional test ones. For this, both the specification tests and the low-cost measurements are performed on a training set of device instances. The mapping derived from the training phase is then used during the production testing phase, in which only the low-cost indirect measurements are performed.

Two main directions have been explored for the implementation of indirect testing, i.e. prediction-oriented strategy [2-14] or classification-oriented strategy [15-20] as illustrated in figure 2. In the first direction, the training set is used to derive functions that map the low-cost indirect measurements to the performance parameters (typically using statistical regression models or artificial neural networks). The objective is actually to predict the individual performance parameters of the device; subsequent test decisions can then be taken by comparing predicted values to specification tolerance limits. In the second
direction, the training set is used to derive decision boundaries that separate nominal and faulty circuits in the low-cost indirect measurement space (specification tolerance limits are therefore part of the learning algorithm). The objective is actually to perform the classification of each circuit as a good circuit or a faulty circuit, but without predicting its individual performance parameters.

![Diagram of test principle](image)

**Figure 2.** Alternate test principle: prediction-oriented strategy (a) and classification-oriented strategy (b).

### B. Previous works addressing analog/RF indirect testing

The fundamental principle of indirect testing, described in the previous sub-section, is common to all contributions found in the literature and have been applied to various types of devices, i.e. analog, mixed-signal and RF devices. Over this basic principle, there are several aspects of the strategy that have been considered in the literature.

At first the machine-learning algorithm should be defined. Different solutions have been proposed, based either on statistical regression models [2-4] or artificial neural networks (ANN) [12]. Then in order to feed the machine-learning algorithm there is a need of signatures giving strong information on the process variability. To gather relevant signatures, research developments have been done in order to define and optimize some stimuli: piece-wise linear signals [3, 4], multi-tone signal [6], digital bitstream [7, 8]. In case of complex signatures, some authors also propose to process these signatures to extract pertinent information using for instance Fast Fourier Transform (FFT) [4, 6, 7] or wavelet analysis [9]. Finally, another approach to gather pertinent information consists in inserting elements within the circuit such as sensors [5, 9-11] or DC probes associated with an analog test bus [11, 12], or exploiting different test conditions such as multiple Vdd conditions [12, 13].

Another aspect regarding the practical implementation of indirect testing deals with the selection of the appropriated indirect measurements in case of a large number of indirect measurements is available. This aspect is addressed in [10, 14] in case of prediction-oriented strategy and in [15, 16] in case of classification-oriented strategy.
Finally other essential points for the deployment of indirect testing deal with confidence in test predictions and efficiency evaluation. Regarding prediction confidence, a defect filter in presented in [17] that permits to screen out outliers, i.e. devices not consistent with the statistical distribution of manufactured devices and for which indirect test prediction should not apply. An interesting approach is also proposed in [18] in the context of classification-oriented strategy, that relies on the use of guard bands in the indirect measurement space in order to identify devices suspect to misclassification. Regarding efficiency evaluation, one of the main problems is that only a limited number of circuit instances are usually available for validation. Recent works have been reported in [19, 20] that permit to generate a large number of synthetic device instances from a comparatively small set of representative devices.

III. ON THE USE OF REDUNDANCY

As previously explained, two main strategies are possible for the implementation of indirect testing: prediction-oriented strategy or classification-oriented strategy. In this paper, we focus on the prediction-oriented strategy. The main advantage of this strategy is that it provides a prediction of the individual performance parameters. This information can then be used to monitor possible shift in process manufacturing, adjust test limits during production phase if necessary, or perform multi-binning. Model redundancy is complementary strategy to regular indirect testing technique. It has been firstly introduced in [21].

The problem highlighted in [21] is related to the efficiency evaluation of prediction-oriented indirect test methods. The efficiency is usually expressed in terms of prediction accuracy, and more particularly in terms of average prediction error. Many of the experiments reported in the literature on various devices demonstrate that very low average prediction error can be achieved, typically 2%. However, two main points limit the credit that can be given to this good accuracy. Firstly, low average prediction error does not guarantee low maximal prediction error, which is of crucial importance regarding the classification step which consists in comparing the predicted values to the specification limits indicated in the datasheet. Secondly, evaluation is usually performed on a small set of validation devices, typically ranging from few hundreds to one thousand instances, while the technique aims at predicting values for a large set of fabricated devices, typically one or several millions of devices. So even if low maximal prediction error can be observed on the small validation set, there is no guarantee that the maximal prediction error will remain in the same order of magnitude when considering the large set of devices under test.

Based on measurements made on 10,000 devices, and using the IMs selection process described in [14], it has been pointed out [21] an important weakness of the prediction-based indirect test method: although it provides accurate prediction results for most of the devices, aberrant predictions are observed for a very small number of devices. In addition aberrant predictions
happen regardless of the training and/or validation set size, and that such errors do not depend on the device itself since the same device may exhibit correct or aberrant prediction depending on the considered regression model. As a result, it seems extremely difficult to foresee such prediction errors. Obviously, this is a serious obstacle for the deployment of the strategy in an industrial context, where one or several millions of devices have to be processed by the test flow.

In this context, the strategy described in [21] consists in introducing an additional step in the indirect test flow in order to provide an indication on the confidence that can be placed on the prediction; if this confidence is low, the device is removed from the indirect test flow for further action to be taken. This approach is similar to the approach suggested in [18] in case of classification-oriented strategy, where guard-bands are allocated in the indirect measurement space in order to identify devices for which the indirect test decision is prone to error. Note that this solution does not apply in case of prediction-oriented strategy because test decisions are not taken in the indirect measurement space but in the performance parameter space by comparing predicted values to specification tolerance limits.

As it has been previously explained, different regression models may lead to aberrant predictions for different devices. As a consequence the indication of confidence in the estimation of a parameter for a given device is based on the crosschecking of the predictions obtained with different models. A device whose performance predictions are similar whatever the regression model used is likely to be properly predicted. On the contrary, when different models lead to different performance predictions for the same device, we can suspect that at least one of the models does not predict the performance correctly. Unfortunately, it is impossible to deduce which one of the predictions is correct and which is not. As a consequence the prediction for this device cannot be trusted, the prediction is considered suspect and the device is removed from the indirect test flow. In other words, model redundancy is used to distinguish reliable predictions from suspicious predictions.
This strategy, called two-tier indirect test scheme using model redundancy, is illustrated in Figure 5. In the classical implementation of the prediction-oriented indirect test, one regression model is built during the training phase for each specification; these models are then used during the testing phase to predict device specification values (see Figure 2.a). In the proposed new implementation, 3 regression models that involve different combinations of indirect measurements are built during the training phase, for each specification. During the testing phase for each specification, 3 predicted values are therefore computed using the 3 models derived in the previous phase; prediction confidence is then established by checking the consistency between these 3 predictions. More precisely for each pair of models, the difference between the predicted values is computed and checked against a threshold value $\varepsilon$. If all these differences are inferior to the threshold, it means that there is no discrepancy between the values predicted by the 3 models and the specification prediction is considered reliable. On the contrary, if one (or more) of these differences is superior to $\varepsilon$, the prediction is considered suspect. Two scenarios are then possible:

- Predictions are considered reliable for all specifications. In this case the device is directed to the first tier, where device performances are computed by averaging the 3 predicted values, for each specification.
• Prediction is suspect for one (or more) specification. In this case, the device is removed from the indirect test flow and directed to the second tier, where further testing may be applied to characterize the device (for instance standard specification testing).

Note that this strategy based on model redundancy assumes that it is possible to build several accurate models involving different combinations of indirect measurements.

Some experiments have been driven on a large number of devices for a given case study in order to demonstrate the efficiency of the model redundancy for indirect testing. Figure 6 and figure 7 present the results for two RF parameters (1dB compression point, CP1 and 3rd order intercept point, IP3) of the considered power amplifier. For each figure, is presented the scatter plot of the estimated performance versus the true value for the three prediction models and the scatter plot after removing the suspect estimations identified thanks to redundancy. The blue stars are the devices used for training, the red stars are the devices used for validation and the green stars the devices identified as suspect thanks to redundancy.

Figure 6. Correlation plots for CP1 prediction using 3 different models and after removal of suspect predictions (ε=1dBm)
IV. ACCURATE ESTIMATION USING REDUNDANCY DESPITE OF LOW-CORRELATION IMs

The previous section has introduced the concept of redundancy and demonstrated its efficiency to identify suspicious predictions. The circuits that have a suspicious prediction are not directly sorted as good or bad according to comparison between the value predicted and the test limits. The circuit should go through an additional test phase. This second test phase is not described here because it should be defined according to the targeted yield or the total test cost. The circuits can be dumped or tested using a conventional test methodology. The redundancy applied to indirect testing strategy improves considerably the confidence in the predicted RF specification value. As explained previously the strategy described in [21] and summarized in the previous section, that relies on model redundancy, improves considerably the confidence in the predicted RF specification value, but assumes that it is possible to build several accurate models involving different combinations of indirect measurements. This assumption can be justified by the potential wide range of IMs. Unfortunately there are cases for which the variety of potential IMs and the correlation between these IMs can be limited. Indeed the quality of an RF path is very sensitive to intrusion, and measuring IMs can require probing the RF path. Another reason is the potential additional cost of implementing circuitry for testing purpose. The limited number of IMs induces a lack of correlation between IMs and specifications and consequently models are not accurate enough.
In this section, we will demonstrate that redundancy also improves the accuracy and the prediction confidence in the case of non-accurate models. This will give more flexibility in IMs set selection and will give a kind of confidence even though we deal with non-accurate models.

To illustrate these points, we have performed some experiments on a case study for which we have experimental test data on a large number of instances. The case study is a Power Amplifier (PA) fabricated by NXP Semiconductors. The experimental test data, from 10,000 devices, include 37 low-cost Indirect Measurements (IMs) based on standard DC tests and 2 RF performance measurements, namely the 1dB compression point (CP1) and 3rd order intercept point (IP3). Because it is a case study, there is available a wide variety of IMs. Thanks to this variety, it is possible to study the impact of the range of the variety and the related correlation between IMs and RF specifications to estimate. These data are separate in two sets of 5,000 devices, one that can be used for training and the other for validation. All the models that are used in the following study have been trained using the same training set made of 5,000 devices. In [21] two optimum sets of 3 IMs from a global set of 37 IMs have been used to train the prediction models for the CP1 and the IP3. These two sets are selected using the method described in [14]. If we look at all possible solution using 3 IMs we will have 7,770 different models. The huge number of combinations of IMs induces a large variety of combinations in terms of prediction accuracy. The Figure 11 shows the histogram of the average prediction error observed on training set for all the 7,770 models. In these two figures, we identify some models having an average prediction error below the targeted value of 1%. In the two cases (i.e. CP1 and IP3) the accuracy of the majority of the models is above 1%. Also we have to notice that the targeted value of the average prediction error observed on training set is not guaranteed on the validation set. So, using models which have an average prediction error above the target is strongly not recommended. In our case we are fortunate to have models that fulfill the required target. But as explained previously, in many cases this could not be the case.

![Histogram of average prediction error observed on training set](image.png)

(a) case of CP1  
(b) case of IP3  

Figure 11. Histogram of average prediction error observed on training set
For the purpose of this study, we propose to select randomly three models which have an average prediction error observed on training set exceeding 1%. The selected models will be used according to the strategy proposed in the previous section. The same specification will be predicted using these models, the circuits having a difference between the three predictions exceeding a given threshold will be removed from the main test flow. For the other circuits the prediction of the specification will be the average of the three predictions. This operation will be repeated 5,000 times in order to obtain a set of 5,000 combinations of randomly-chosen models. In order to achieve a precise analysis, for every draw, the average prediction error and the maximum prediction error of the “Mean of models” (after removing suspect prediction) and the single models will be compared. Figure 12 illustrates the average prediction error observed on validation set for the used model and the “Mean of the models”. Here the illustrated single model is the best one: i.e. the model which has the lowest prediction error observed on training set of the three models used for redundancy.

![Figure 12. Average prediction error on validation set](image)

It clearly appears that over the 5,000 draws, there are a limited average prediction error observed on validation set when the redundancy strategy has been implemented. On the other side, the average prediction error of the best model is predominantly over 5% value and we can observe a large variance over the 5,000 draws. We can conclude that those “best” models (over the three used for redundancy), which have an average prediction error over 1% during training are not eligible for a regular indirect test because of a lack of correlation between the IMs and the wanted specifications.

To illustrate clearly this fact, we will rearrange the models regarding their ascending order of their average prediction error. The Figure 13 and Figure 14 respectively illustrate the average prediction error and the maximum estimation error of the worst models, the best models (over the three drawn randomly) and the Mean of models over the 5,000 iterations. The figures are zoomed in. This zoom is mandatory because of enormous average prediction error over #4,500 and enormous
maximum prediction error over #3,000. As the values are sorted in ascending order, we can expect that values that are not presented are over the maximum value of the ordinate. For the two single models, the average prediction error is usually greater than that observed for the Mean of models after removing suspect predictions. We can observe in the case of the CP1 that only 60% of time, for the best model, the average prediction error on the validation set is less than 2%. If we focus on the result of the Mean of models, we can identify clearly that 100% of time the average prediction error is bellow 2%. The same observation is done with the second specification i.e. IP3.

The benefit of the proposed method is clearly highlighted in this case for which the aberrant predictions are efficiently limited. In the case of the CP1, we see that in the best case only 30% of times the maximum prediction error is below 2 dBm. But we reach easily 99% of times with the “Mean of models”.

In the case of the IP3 specification the percentage of occurrences for which the maximum prediction error is below 2dBm is around 75% against the 20% in the best case when using only one model.
The implementation of redundancy with accurate models enables to provide strong confidence in accurate estimations. In addition the number of suspicious estimations remains very low. In the case study presented in [21], the number of suspicious estimations using different accurate models was between 0.05 and 0.1%. As demonstrated earlier, redundancy strategy provides the possibility to relax the constraint of cost of IMs to build prediction models with less correlation but with same confidence in the accurate estimations. The main consequence is the increased number of suspicious estimations. Graphic results are provided for CP1 and IP3 parameters respectively by Figure 15 a) and Figure 15 b). Numerical results are stated in Table 1.

![Figure 15: number of suspicious estimations over the 5000 iterations](image)

Figure 15 presents the number of suspicious estimations for the combinations sorted in ascending order of the averaged estimation error of the Mean of models (cf. blue line of Figure 13). Like in Figure 13 and 14, there is a small gain which is coherent with the constant increase of curves in Figures 13 and 14. But according to Figure 15 and to Table 1, there is only a small increase of the number of suspicious estimations compared to the case describes in [21], but it remains low. Indeed the average number of suspicious estimations for CP1 and IP3 are respectively 48.3 and 18.9 over 5,000 DUTs, which corresponds to 0.966%. Moreover for a maximum number of suspicious errors of 173 over 5,000 DUTs, it makes 3.46% of the circuits for which the sorting as good or bad according to the indirect approach cannot be done. A yield loss between 1 and 4% is a value usually accepted in the domain of industrial testing.
Table 1: Maximum, average and standard deviation of number of suspicious estimations over the 5000 iterations

<table>
<thead>
<tr>
<th></th>
<th>Maximum # of suspicious estimations</th>
<th>Average # of suspicious estimations</th>
<th>Standard deviation of # of suspicious estimations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP1</td>
<td>173</td>
<td>48.3</td>
<td>23.7</td>
</tr>
<tr>
<td>IP3</td>
<td>62</td>
<td>18.9</td>
<td>8.9</td>
</tr>
</tbody>
</table>

As explained previously, the objective of the redundant strategy is to sort the estimations as accurate or suspicious. Conventional sorting of DUTs can be achieved according to the accurate estimation of test parameters. The DUTs, for which the estimation of the test parameters is suspicious, will go through another test or will be dumped. The second step of the whole test strategy is particular for each product. The proposed approach will give great flexibility for test engineer to choose the set of used IMs to implement the indirect test. Moreover the results presented in this publication demonstrate that the redundant strategy makes the indirect test method still applicable to the test case for which there is a limited number of IMs or low correlation between IMs and conventional test parameters.

V. CONCLUSION

In this paper we have enhanced the confidence in using indirect testing combined with model redundancy in the context of a prediction-oriented strategy. It has been demonstrated that this strategy is efficient for any test case, even when the variety of IMs is limited for maintaining the quality of the RF function or simply for economic reasons aiming at limiting the amount of circuitry not dedicated to the core function. Indeed the low variety of IMs is an issue because it induces a lack of correlation between IMs and regular specifications; as a consequence it is impossible to find a combination of IMs enables to train an accurate prediction model. This problem is overcome by implementing model redundancy.

The proposed approach has been validated on a real case study for which we have experimental measurements on a set of 10,000 devices. Results show that most of the devices can be accurately predicted with the indirect testing strategy, with a maximal prediction error in the same range than the one observed on the training set. Only a very small number of the devices are directed to the second tier, therefore incurring very low test cost overhead. First we have used accurate models to implement the proposed method. Then, we have demonstrated that it can works also with less accurate models. This gives more flexibility for the best choice of IMs.

ACKNOWLEDGMENTS
The authors would like to thank C. Kelma from NXP Semiconductors for providing the PA production test data and for fruitful discussions.

REFERENCES


[12] Sofiane Ellouz; Patrice Gamand; Christophe Kelma; Bertrand Vandewiele; Bruno Allard; , "Combining Internal Probing with Artificial Neural Networks for Optimal RFIC Testing," IEEE Int’l Test Conference (ITC), pp.1-9, Oct. 2006


