Evaluation of a Two-Tier Adaptive Indirect Test Flow for a Front-End RF Circuit

Hassan El Badawi, Florence Azaïs, Serge Bernard, Mariane Comte, Vincent Kerzérho, Francoise Lefevre

Version submitted by the authors

Title:
Evaluation of a two-tier adaptive indirect test flow for a front-end RF circuit

Authors:
H. El Badawi [1,2], F. Azais [1], S. Bernard [1], M. Comte [1], V. Kerzerho [1], F. Lefevre [2]

Affiliations:
[1] University of Montpellier, CNRS, LIRMM, 161 rue Ada, 34095 Montpellier Cedex, France
[2] NXP Semiconductors, 2 Esplanade Anton Phillips, 14000 Caen, France

Corresponding author:
Florence Azais
LIRMM, University of Montpellier, CNRS, 161 rue Ada, 34095 Montpellier Cedex, France
Tel: +33 467 418 625
Email: florence.azais@lirmm.fr

Abstract:
This paper focuses on test cost reduction for RF circuits based on the use of an indirect test strategy. The implementation of a two-tier adaptive test flow is investigated, in which only circuits with a sufficient prediction confidence level are evaluated by the indirect test while others are re-evaluated by specification-based test. A methodology is presented that permits to explore different tradeoffs between test quality and test cost and to make pertinent choices for the efficient implementation of such a test flow. The methodology is applied to a front-end RF circuit designed for WLAN applications and results show that substantial test cost reduction can be achieved without compromising the test quality.

Keywords: Indirect test, RF integrated circuits, machine-learning algorithms, adaptive testing, test efficiency, test confidence

Version submitted by the authors

Evaluation of a two-tier adaptive indirect test flow for a front-end RF circuit

H. El Badawi [1,2], F. Azais [1], S. Bernard [1], M. Comte [1], V. Kerzerho [1], F. Lefevre [2]
[1] LIRMM, University of Montpellier, CNRS, 161 rue Ada, 34095 Montpellier Cedex, France
[2] NXP Semiconductors, 2 Esplanade Anton Phillips, 14000 Caen, France

***Abstract* — This paper focuses on test cost reduction for RF circuits based on the use of an indirect test strategy. The implementation of a two-tier adaptive test flow is investigated, in which only circuits with a sufficient prediction confidence level are evaluated by the indirect test while others are re-evaluated by specification-based test. A methodology is presented that permits to explore different tradeoffs between test quality and test cost and to make pertinent choices for the efficient implementation of such a test flow. The methodology is applied to a front-end RF circuit designed for WLAN applications and results show that substantial test cost reduction can be achieved without compromising the test quality.**

Keywords: Indirect test, RF integrated circuits, machine-learning algorithms, adaptive testing, test efficiency, test confidence

1 Introduction

Testing of Integrated Circuits (ICs) is a crucial step in the production process as it ensures the quality of manufactured devices and their compliance with the technical datasheet. In the case of digital parts, although their complexity has exploded in recent decades, the use of structural test techniques that rely on fault models as well as the implementation of Design-for-Testability (DfT) solutions have permitted a significant reduction of the test cost. On the other hand, for analog and Radio Frequency (RF) parts, although they have evolved less in terms of complexity than digital parts, their test cost still remains a major contributor to the total production cost. One of the main reasons is that there is no recognized fault model for analog and RF parts and therefore approaches based on fault detection are not applicable. Therefore, analog and RF circuits are tested using a functional approach, which is based on measuring the circuit performances and verifying that these performances meet the specifications guaranteed by the manufacturer. This approach ensures a good test quality, but its implementation requires expensive specific test equipment and long test times, resulting in a very high testing cost.

In this context, there is a strong demand for the development of alternative solutions to specification testing for analog and RF circuits. An interesting approach lies in an indirect testing strategy based on the use of machine-learning algorithms, in which the device performances are predicted from Indirect Measurements (IMs) that do not require specific test equipment. This approach has been first proposed for analog circuits using transient samples of the circuit response as alternate measurements [1,2], and then extended to RF circuits using DC voltages captured by internal probes as alternate measurements [3]. Since then, many works have addressed various aspects of the implementation of this strategy, such as the composition of the learning population [4], the use of embedded sensors [5,6] or multi-Vdd conditions [7], the selection of relevant indirect measurements [8-10], etc. A comprehensive review of works related to indirect test can be found in [11].

In this paper, we present on a practical case study how it is possible to benefit from the test cost reduction offered by indirect test without compromising the test quality brought by specification test by using a two-tier adaptive test flow. Indeed, test content, test flow and test limits are traditionally set statically, which means that all parts are tested in the same way, regardless of their individual performances. On the other hand, in adaptive testing, test content, test flow or test limits can be changed for each part based on manufacturing test data or statistical data analysis. This concept has emerged in the early nineties to optimize the tests applied on VLSI digital dies [12]. It has then been largely exploited in the context of Iddq-testing of digital circuits [13]. More recently, adaptive approaches have also been explored for analog and mixed-signal circuits [14], including in the context of indirect testing [15-19]. A novel implementation of a two-tier adaptive indirect test flow has been proposed in [20] and illustrated with preliminary results. This paper is an extension of this latter work, which includes new results using two different types of regression model and a specific study dedicated to the use of an optional filter during the data preparation phase.

The paper is organized as follows. Section II recalls the principle of indirect test and introduces the two-tier adaptive test flow together with its specific proposed implementation. The methodology for the practical elaboration of the test flow is then detailed in Section III. Finally, the case study is presented in Section IV, and results are presented and analyzed in section V. Section VI concludes the paper.

2 Indirect test strategy implementations

2.1 Classical implementation

The underlying assumption that supports the indirect test strategy is that manufacturing process variations that affect the RF performances of a device also affect some indirect parameters that can be measured at low cost. Therefore, if it is possible to establish a link between both parameter spaces, it should be possible to verify the RF performances of a device only by measuring the low-cost indirect parameters. However, the relationship between these two spaces is usually very complex and cannot be established with an analytical formulation; machine-learning techniques are then used to establish this link.

Based on this principle, the classical implementation of the indirect test strategy involves two distinct phases. The first one is the training phase in which both the conventional RF performance measurements and the low-cost indirect measurements (IMs) are performed on a set of training devices. A machine-learning algorithm is then used to train regression models that map the IM space to the RF performance space. Once the models are established, the actual production testing phase can start. In this phase, only the low-cost indirect measurements are performed for every new device. The RF performances are predicted using the mapping learned in the initial training phase. Device binning can then be accomplished by comparing the predicted performances with specification limits. This synopsis is illustrated in Figure 1.

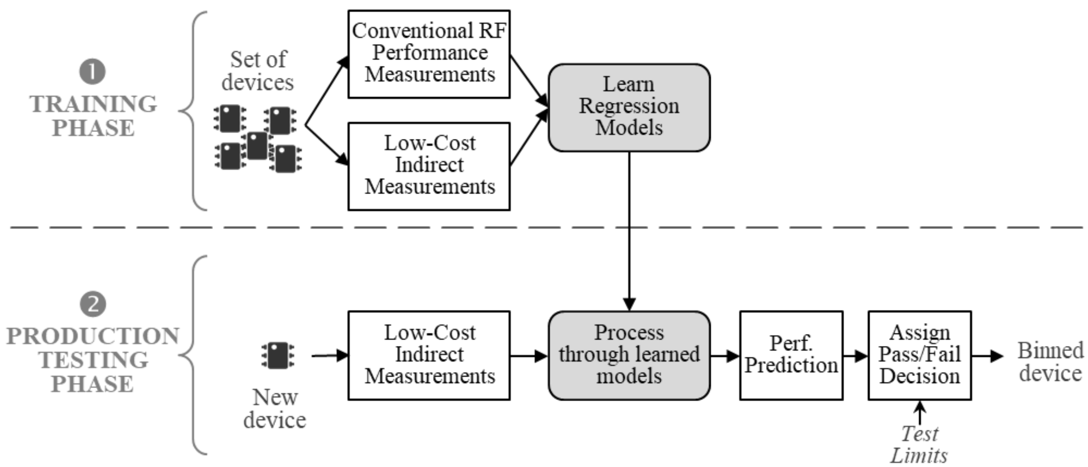


Fig.1. Indirect test synopsis

This strategy has the potential to lead to substantial test cost savings since the conventional costly RF measurements are performed only on a limited number of devices for the training phase (typically several hundreds to a few thousands). Then, all devices manufactured during the mass production (typically several hundreds of thousands) are evaluated using only the low-cost indirect measurements.

2.2 Two-tier adapative indirect test flow

2.2.1 Principle

One of the main issues that limit today the wide deployment of the indirect test strategy in industry is a problem of confidence in the predicted results. Indeed, the machine-learning algorithms used to build regression models are perceived as a black box and often induce a lack of confidence. To cope with this issue, an extension of the indirect test strategy has been proposed, called the two-tier adaptive test flow. The objective of this approach is to preserve the test quality of specification testing while leveraging the low cost of indirect testing.

The principle of the two-tier adaptive test flow is illustrated in Figure 2. As in the classical indirect test implementation, it involves two distinct phases: the training and production testing phases. The training phase is identical to that of the classical implementation, only the production testing phase differs. More precisely, the idea is first to process every device by the indirect test, i.e. to predict its performances based only on the low-cost indirect measurements using the models learned in the initial training phase. However, before assigning a pass/fail decision, the confidence in the achieved predictions is evaluated. If the confidence is high enough, predictions are considered reliable and the device is labeled according to the indirect test decision. This constitutes the first tier. If the confidence in the predictions is insufficient, the device is then directed to a second tier in which it is submitted to a standard specification test, i.e. the conventional RF measurements are performed and the device is labeled according to these measurements. The assumption underlying this approach is that the large majority of devices will be sorted by the first tier, and that only a small fraction of devices need to go to the second tier. This approach then permits to maintain a significant test cost reduction, but offers more confidence in the achieved test quality.

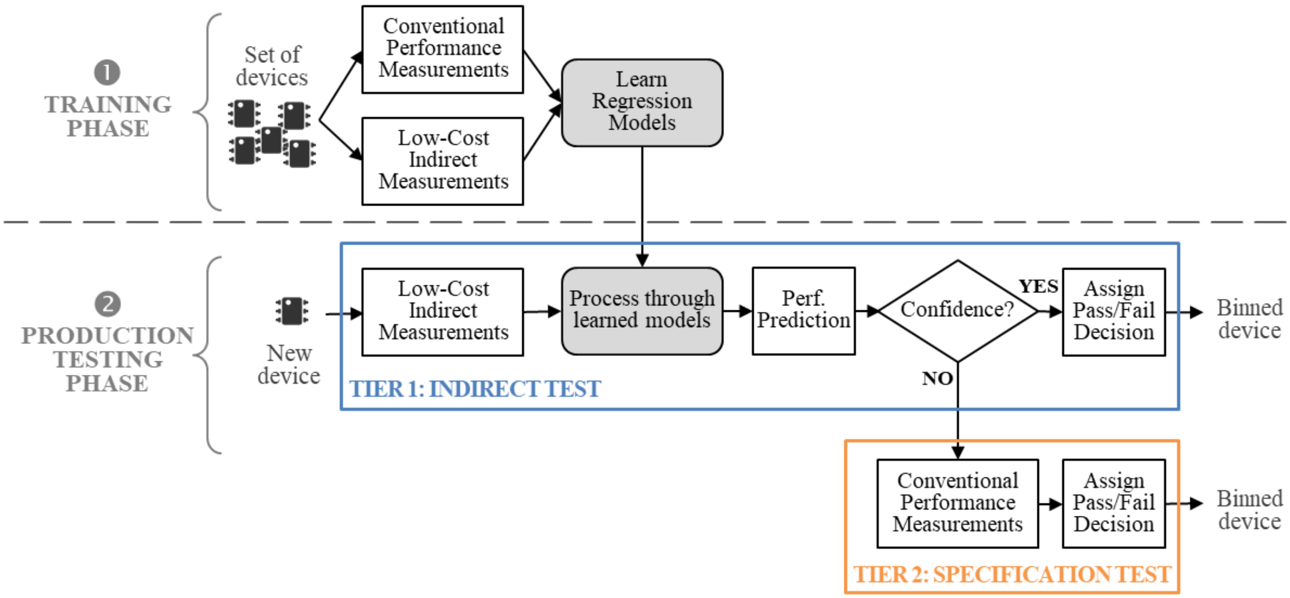


Fig.2. Two-tier adaptive test flow synopsis

2.2.2 State-of-the-art on adaptive indirect test

a) Classification-oriented indirect test

Adopting an adaptive test flow to improve the quality and accuracy of an indirect test strategy was firstly introduced in [15]. The authors have implemented an ontogenic neural network to create two guard-bands between faulty and good circuits in the indirect measurement space. During the production testing phase, each new device first goes through the indirect test tier. If the device falls outside the guard-bands, it is assigned to the dominant class (good or faulty) based solely on the low-cost indirect measurements. Oppositely, if the device falls within the constructed guard-bands, it is directed to the second tier where it is retested using the conventional specification test.

In another work, the authors of [16] have proposed a confidence estimation, by creating three different regions in the indirect measurement space, where the prediction of new devices can be trusted, discarded, or retested. The boundaries of these different regions are established using a SVM classifier between two different classes: trusted and untrusted predictions.

b) Prediction-oriented indirect test

The adaptive two-tier test approach has also been explored in the context of prediction-oriented indirect test. It was firstly introduced in [17], where the authors proposed a strategy based on model redundancy. The main idea is to apply various prediction models on the same training set, while using a different set of indirect measurements for each model. Whenever a lack of consistency in the different predictions for a particular instance is detected, the circuit is then re-directed for further thorough testing.

In another work, the authors of [18] have proposed two adjustable defect filters, in order to avoid the entailed risks in predicting the performances of marginal and outlier instances. The first one is a strict filter, where all the instances that are adequately represented in the training data will have their performance predicted by using the established regression function. On the other hand, all the suspicious instances are redirected to the more lenient filter, which will discard gross defects (outliers) and re-test marginal devices in a classical manner. Unlike the guard-bands introduced in [15], where a classifier is trained to create a buffer zone in the indirect measurement space, the defect filters are constructed based on Kernel Density Estimation of the indirect measurements, which has been introduced in [21].

c) Multi-site indirect test

Finally, in the case of multi-site testing, the authors in [19] have proposed an implementation of an adaptive test strategy to reduce test time. The main supposition is that the indirect test strategy, which has already defined the set of the most pertinent performance indicators, is capable of replacing the specification based testing and can replicate its accuracy. The adaptive element in this strategy is the number of indirect measurements included in the prediction model: the less is incorporated, the more time can be saved. Thus, the learning phase involves the ordering of the available features and the incremental training of different regression models by adding a new feature at each iteration. During the production test, the test program starts with an evaluation using a model with a low number of indirect measurements. If all the sites are predicted with an acceptable level of confidence, the circuits under test do not have to keep on exploring the remaining indirect measurements and the test program halts, leading to test time improvement.

2.3 Proposed implementation of the two-tier adapative indirect test flow

As discussed in the previous section, there are diverse ways of approaching an adaptive indirect test implementation. Our target is to develop an adaptive solution in the context of a prediction-oriented indirect test. Compared to the solution proposed in [18], where circuits are re-directed based on their distribution in the indirect measurement space, our preference is to base the adaptive strategy on the distribution of the predicted instances in the RF performance space. Our intention is also to limit as much as possible the number of used indirect measurements in order to maximize the cost reduction. Hence, the solution proposed in [17] is not suitable, since it relies on the building of redundant models that involve different indirect measurements, which necessarily increases the number of required indirect measurements.

In this work, we investigate a novel implementation of the two-tier adaptive test flow in the context of prediction-oriented indirect test. The idea is to evaluate the confidence in the predicted performance value based on a tolerance zone around test limits. Indeed, previous works have shown that almost all misclassified circuits are circuits with a predicted value close to a test limit, while correct decisions are taken for circuits with a predicted value far from test limits [22]. Therefore, the proposal is to establish confidence by looking at the location of the predicted value with respect to a tolerance zone defined around a test limit, as illustrated in Figure 3. More precisely, any device with a performance prediction that falls outside the tolerance zone will be directly classified as a good/bad device for this performance according to the indirect test tier, while any device with a prediction that falls within the tolerance zone will be directed to the second tier in order to be evaluated through conventional specification test.

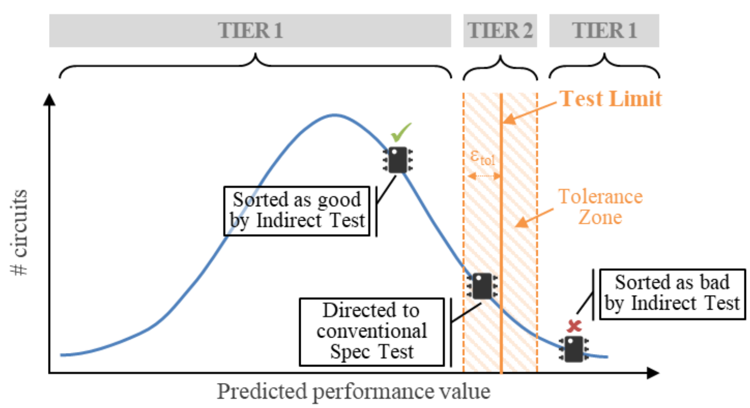


Fig.3. Principle of confidence estimation in the proposed two-tier adaptive test flow

A chief interest of this solution is its simplicity while it has the ability to adapt to different industrial constraints. Indeed, varying the size of the tolerance zone around the test limit permits to explore different trade-offs between test quality and test cost. With a tolerance zone set to zero, 100% of the devices are evaluated with the indirect test tier and the test costs are minimal; however, the test quality might not be sufficient to meet the industrial constraints. By creating and expanding the tolerance zone, we expect an improvement of the test quality but at the expense of a number of devices that need to be evaluated with the conventional specification test. Depending on the targeted industrial constraints, it is therefore essential to have an appropriate setting of this parameter during the initial learning phase in order to really benefit from the two-tier adaptive test approach.

3 Methodology for practical implementation

In the previous section, we have introduced the principle of the indirect test strategy and the two-tier adaptive test approach. The practical implementation implies several choices, such as the selection of pertinent IMs, the choice of the regression algorithm or the size of the tolerance zone. Obviously the achieved test quality depends on these choices. In this section, we describe the methodology that has been defined in order to assist the test engineer in the elaboration of the test flow. The general overview of this methodology is depicted in Figure 4 and the details on the different phases are given in the following subsections.

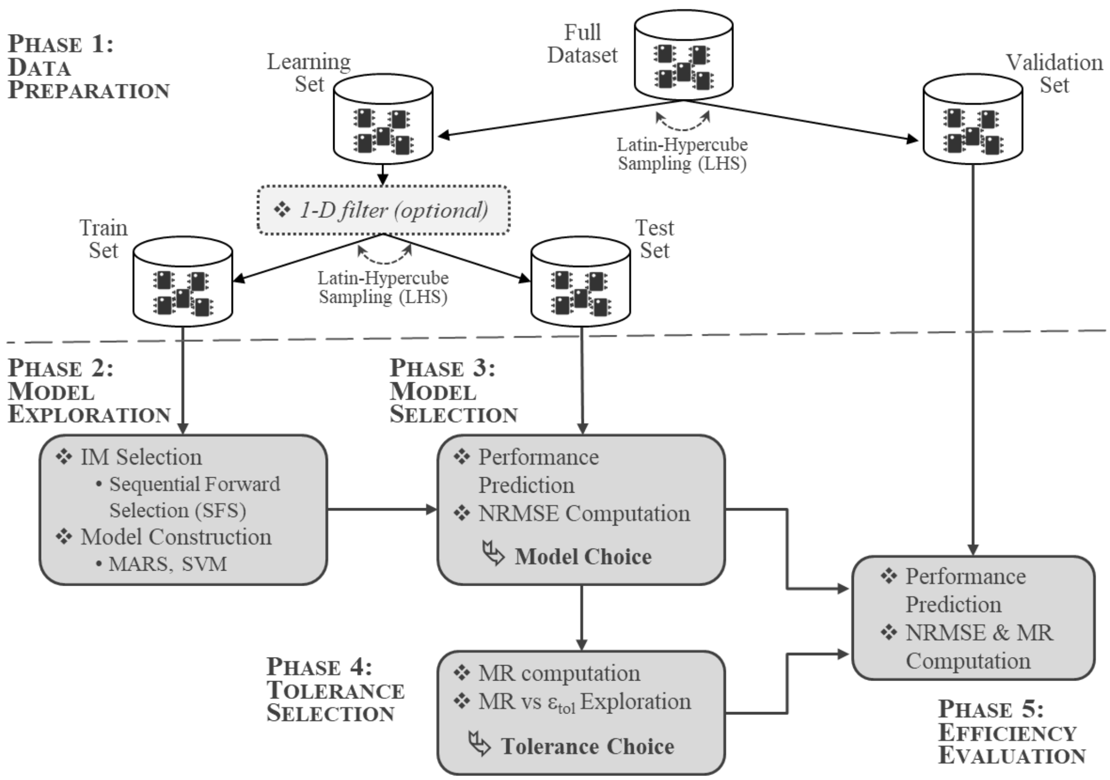


Fig.4. General overview of the adaptive test flow elaboration methodology

Data preparation

The first phase concerns data preparation. The initial dataset should contain the conventional performance measurements and a large variety of indirect measurements on a sufficient number of circuits (typically several thousands). This full dataset is first partitioned into two datasets, called learning and validation sets. The learning set will be used to explore the different possibilities regarding the test flow implementation and to identify the best options. The second set is dedicated to the validation of the retained options using an independent set of devices; it is intended to represent the production testing phase. Note that, although both sets are independent, it is essential that they present similar characteristics to ascertain the validity of results. Therefore, the partitioning is achieved using Latin-Hypercube Sampling (LHS), which is a sampling approach that preserves the statistical characteristics of the initial distribution in the sampled sub-datasets.

The learning set is in turn partitioned into two subsets, i.e. the train set and the test set. The first one will be used to train the prediction models and the second one to evaluate the accuracy of the constructed models. It is important to perform this evaluation on instances different from the ones used for training in order to verify model generalization ability and avoid issues related to overfitting.

Finally, note that it is often recommended in the literature to work with a dataset that does not contain outliers. Indeed, data outliers can spoil and mislead the training process resulting in longer training times, less accurate models and ultimately poorer results. Consequently, an optional pre-processing step is implemented to exclude these instances from the learning set by applying an iterative one-dimensional filter, which has been proposed in [23]. The basic principle of this filter is, for a given parameter, to remove all instances that have a measured value outside the interval $[\mu - k\sigma; \mu + k\sigma]$, where μ is the mean value of the population for the considered parameter, σ the standard deviation, and k a positive integer that permits to choose the strictness of the filter. The filter is applied individually on each RF performance and each indirect measurement. The final list of circuits excluded by the filtering process is the union of all circuits pruned by the filter over all RF performances and indirect measurements.

Model exploration

The second phase of the methodology is the model exploration. In this phase, a number of regression models will be built using different subsets of IMs. The problem of selecting a pertinent subset of IMs within a large set of candidates is a recurrent problem in the field of machine-learning, known as feature selection. In the context of indirect test, the common approach is a wrapper method called Sequential Forward Selection (SFS) [8]. For this study, we have implemented such a procedure, limiting the number of selected IMs to 15.

The next step is then to train regression models using the selected IMs. Many different algorithms exist to perform this task. Classical algorithms include Multiple Linear Regression (MLR), Multi-Adaptive Regression Splines (MARS), Support Vector Machine (SVM), or more elaborated algorithms that combine several models in an approach called ensemble learning. A comparative study of the efficiency of these different models in the context of indirect testing can be found in [22]. For this study, our focus is to explore the benefit that can be achieved by implementing a two-tier adaptive test flow, and not to perform a complete comparison of the different model types. Therefore, we have implemented only 2 types of regression models which are the most commonly used algorithms in the context of an indirect test strategy for their capabilities of depicting non-linear behaviors, i.e. MARS and SVM.

Model selection

The third phase of the methodology concerns model selection. In this phase, all models learned in the previous phase are used to perform prediction of the devices from the test set. The accuracy of these models is evaluated in terms of Normalized Root Mean Square Error (NRMSE), which is a normalized measure of the rms prediction error expressed in percentage:

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} / \bar{y} \tag{1}$$

where y_i is the actual performance value of the i^{th} instance, $\hat{y}_i$ the predicted performance value of the i^{th} instance, $\bar{y}$ the mean of the observed data and n the number of instances.

Models with the lowest NRMSE are then retained as the best solutions for each RF performance.

Tolerance selection

The following phase is specific to the implementation of a two-tier adaptive test flow. It is related to the exploration of the tradeoff that can be achieved between test quality, expressed in terms of Misclassification Rate (MR), and test cost, expressed in terms of percentage of devices that need to be retested with a conventional specification test. Practically for each selected model, the misclassification rate is first computed with a tolerance zone set to zero (only indirect test). The size of the tolerance zone is then progressively enlarged in order to study the evolution of the MR score versus the number of devices directed to the second tier. The appropriate size of the tolerance zone can be chosen for each RF performance with respect to a targeted test quality, i.e. the smallest size that does not overcome a predefined maximum MR.

Efficiency evaluation

Finally, the last phase of the methodology is dedicated to the evaluation of the two-tier adaptive test flow efficiency. All the options retained in the learning phase are evaluated on the devices from the validation set, by computing the MR score and the percentage of retested devices. Indeed, it is important to verify that the efficiency established on the test set during the learning phase is preserved on the validation set, which is intended to be representative of the realistic conditions encountered during the industrial testing phase.

4 Case Study

4.1 RF product

The case study is a front-end integrated circuit designed for WLAN applications. The three main specifications to be verified are the gain of the receiver chain (Rx-gain), the gain of the transmitter chain (Tx-gain) and the Error Vector Magnitude of the transmitter chain (Tx-EVM). The low-cost indirect measurements investigated for this product include standard DC measurements performed on external nodes of the device together with internal DC measurements (the device is equipped with an internal DC bus and internal probes that give access to the voltage at some specific nodes and signatures delivered by built-in process monitors). Overall, we have a total of 131 possible indirect measurements.

4.2 Measurement campaign

An extensive campaign of measurements has been carried out in the production test environment and test data have been collected on more than 26,700 circuits coming from different wafers fabricated under various extreme process conditions. The test data, which include both the conventional measurements of the three RF specification performances and the 131 indirect measurements, constitute the full dataset. The main characteristics of this dataset are summarized in Table I for the three RF performances.

TABLE I. CHARACTERICTICS OF THE FULL DATASET FOR THE 3 RF PERFORMANCES

		RF Performance		
		Tx-EVM	***Tx-gain***	***Rx-gain***
Full Dataset *26,706 instances*	Coef. of Variation	11.1%	3.0%	3.7%
	% of good circuits	76.4%	97.7%	100%
	% of bad circuits	23.7%	2.3%	0%

It can be noticed that the characteristics of the population significantly differ depending on the considered RF performance. For the Tx-EVM, we observe a quite large distribution with a dispersion around 11%; more than 76% of the circuits satisfy the targeted EVM requirement. For the Tx-gain, the distribution is tighter with a dispersion of only 3%; almost 98% of the circuits satisfy the targeted gain requirement. Finally, for the Rx-gain, we also observe a tight distribution with a dispersion around 3.7%; in this case the targeted requirement is sufficiently far away from the distribution so that 100% of the circuits satisfy the requirement. At this point, it is important to underline that a number of wafers are manufactured with corner process conditions, and their circuits have been included in the population on purpose. Therefore, the proportion of faulty circuits is not representative of what would be the actual production yield under normal process conditions.

4.3 Data preparation

4.3.1 Dataset partitioning

The first step of the data preparation is the partition of the full dataset in two sets, i.e. the learning and validation sets. In this work, we choose an equal repartition between the two sets, achieved using conditioned Latin Hypercube Sampling. The full dataset of 26,706 devices has therefore been partitioned into two sets of about 13,350 devices. The main characteristics of these two sets are summarized in Table II, which provides the coefficient of variation, the percentage of good circuits and the percentage of faulty circuits, for the three RF performances.

TABLE II. CHARACTERICTICS OF LEARNING AND VALIDATION SETS FOR THE 3 RF PERFORMANCES

	RF Performance

		Tx-EVM	***Tx-gain***	***Rx-gain***
Learning Set *13,354 instances*	Coef. of Variation	11.0%	3.0%	3.7%
	% of good circuits	77.2%	97.9%	100%
	% of bad circuits	22.8%	2.1%	0%
Validation Set *13,352 instances*	Coef. of Variation	11.3%	3.0%	3.7%
	% of good circuits	75.5%	97.6%	100%
	% of bad circuits	24.5%	2.4%	0%

From this table, it clearly appears that the learning and validation sets exhibit similar characteristics in terms of distribution dispersion and proportion of good or bad circuits for each RF performance, confirming that the use of Latin-hypercube sampling permits to obtain several sets with the same distribution characteristics as the initial population.

4.3.2 Use of the optional filter

The influence of using a filter during the learning phase has also been examined. In particular, a detailed analysis of the distribution and the properties of the circuits identified by the filtering process has been performed, in terms of number of circuits, repartition of these circuits with respect to their compliance with the RF specifications, and distribution of these circuits within the different subsets generated from the full dataset. Although this optional filter would be applied only on the learning set in the proposed methodology (Fig. 4), the preliminary analysis presented here has been performed on the full dataset and varying the strictness of the filter, i.e. varying the value of k. Concretely, we have considered two different filters, i.e. a strict one with a limit at $\pm 6\sigma$ and a more relaxed filter with a limit at $\pm 10\sigma$ (the influence of these two filters on the achieved test efficiency will be studied in the following sections). For the completeness of the current preliminary analysis, we have also included here more lenient filters, i.e. filters with a higher value of k (25, 50 and 100). Results are presented hereafter.

Figure 5 shows the evolution of the number of circuits identified by the filtering process with respect to the filter severity (note the log scale on the y-axis). As expected, the number of circuits with outlying values quickly diminishes as the filter becomes more relaxed. For the strict filter, 10,186 circuits are identified by the filtering process, which corresponds to 38.1% of the total population. This number reduces to 2,673 circuits for the relaxed filter, which corresponds to 10%. This number then rapidly falls down below 150 circuits for more lenient filters, which corresponds to a negligible portion of the population (less than 0.5%).

An important remark is that, whatever the filter severity, all circuits identified by the filtering process have outlying values because of one or several indirect measurements, but none because of the RF performances. Another remark is that, even for an extremely lenient filter with of value of k=100, 8 circuits present extreme IM values. Even small (only 0.03% of the population), this number is unexpected taking into account that none of the circuits present in the dataset has an RF performance value outside $\pm 6\sigma$ of the distribution.

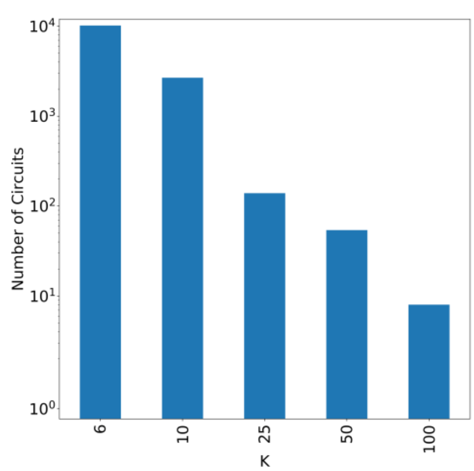
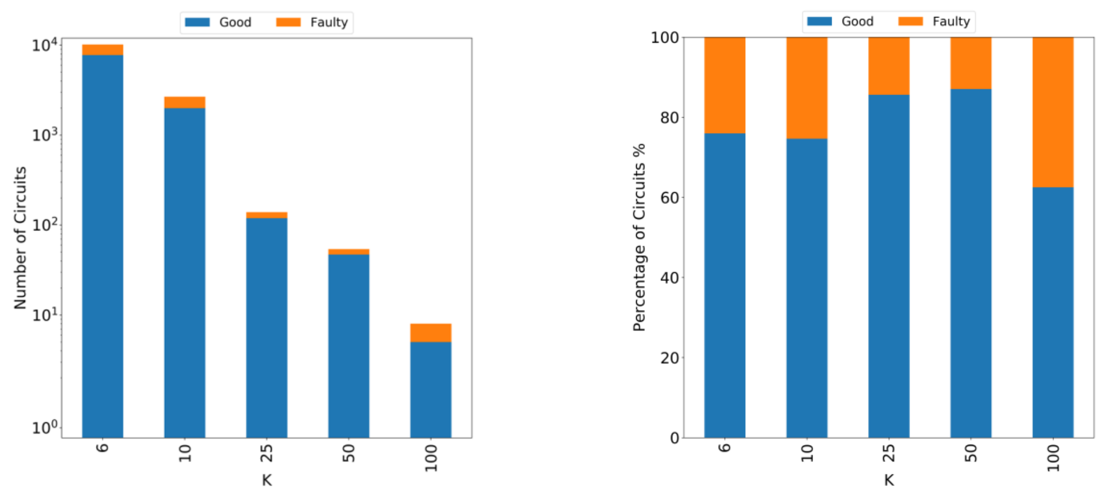


Fig.5. Number of circuits identified by the filtering process vs. filter severity

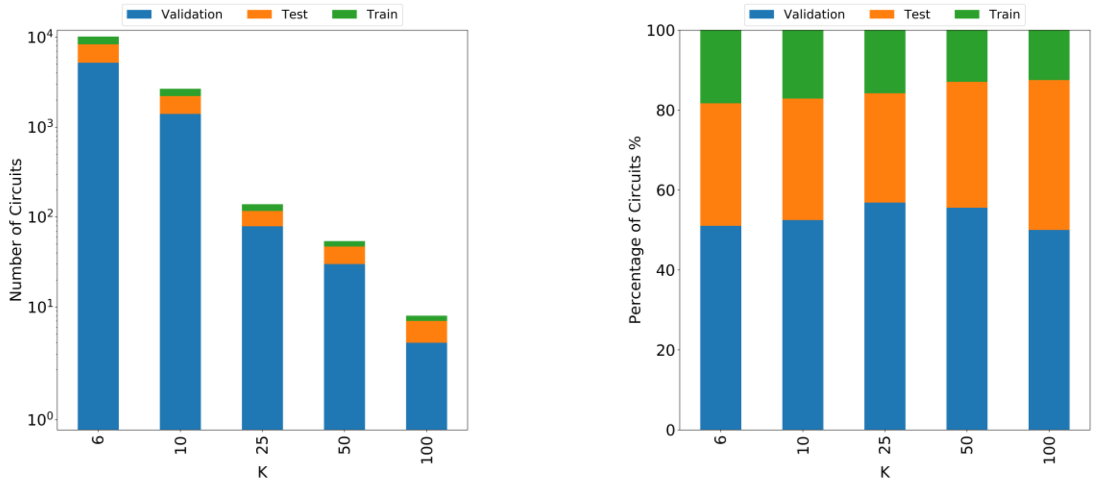
Then we have analyzed the repartition of circuits identified by the filtering process with respect to their compliance with the RF specifications. Results are reported in Figure 6. More precisely, in Figure 6.a, blue bars represent the number of circuits that have been flagged by the k-σ filter but which are compliant with respect to RF specifications (good circuits), while orange bars represent the number of circuits that have been flagged by the k-σ filter and which are not compliant with respect to at least one RF specification (faulty circuits). The same information is represented in Figure 6.b, but this time expressed in terms of percentage of the total number of circuits identified by the filtering process. It can be observed that the proportion of faulty circuits among the total number of circuits identified by the filtering process varies between 13% and 37% depending on the filter severity. Globally, this proportion is in relative good agreement with the proportion of faulty circuits within the full dataset, i.e. 24%. However surprisingly, we observe a non-monotonic variation. Indeed, the proportion of faulty circuits among the total number of circuits identified by the filtering process is around 25% for the strict and relaxed filters with k=6 and 10. This proportion falls down to 13% for the lenient filters with k=25 and 50. It then increases up to 37% for the extremely lenient filter with k=100. This reveals that there is no direct relation between the fact that a circuit exhibits extreme values for indirect measurements, and the fact that it is a good or faulty circuit with respect to its RF performances. This point is important because it indicates that the use of a one-dimensional filter applied on the IMs during the production testing phase would be totally ineffective since it does not help to discriminate between good and faulty circuits. Even worse, it would eliminate a number of good circuits, provoking yield loss.



(a) (b)

Fig.6. Representation of filtered circuits in terms of good and faulty circuits

Finally, we have analyzed the repartition of the circuits identified by the filtering process within the different subsets used in the learning and validation phases (Train, Test, and Validation). Indeed, the conditioned-LHS process used for the partitioning of the population is performed considering only the RF performances. To ensure that there is no bias coming from this partition, it is essential to verify that the generated subsets also reflect the original dataset with respect to the indirect measurements. In particular, the proportion of circuits identified by the filtering process in each subset should be in accordance with the achieved partitions, i.e. 18% in the train set, 32% in the test set, and 50% in the validation set. Results are reported in Figure 7. This figure shows that, even if the number of circuits identified by the filtering process decreases as the filter becomes more relaxed, the expected repartition is globally maintained.



(a) (b)

Fig.7. Repartition of filtered circuits over Train, Test and Validation subsets

To summarize, this analysis shows that the use of a one-dimensional filter does not significantly modify the characteristics of the initial population with respect to the RF performances. Moreover, the circuits flagged by the filter are distributed in the different subsets in accordance with the achieved partitions. It is therefore founded to investigate whether the use of such a filter can improve the accuracy of the models constructed during the learning phase, which might result in a better test efficiency during the production testing phase.

In the remaining of the paper, we will consider only the strict and relaxed filters (k=6 and k=10) and we will study how the use of these filters during the learning phase impacts the test efficiency achieved on the validation set. For the sake of clarity, we stress that in the proposed methodology the optional filter is applied only to the learning set (i.e. train and test sets) and not to the validation set, which should remain representative of a realistic production population. The main characteristics of the filtered learning sets are summarized in Table III. It can be clearly observed that, for each RF performance, the filtered learning sets exhibit a dispersion similar to the original learning set (maximum difference of 0.4%), and the proportion of good or faulty circuits is globally preserved (maximum difference of 1.2%).

TABLE III. CHARACTERICTICS OF THE FILTERED LEARNING SETS FOR THE 3 RF PERFORMANCES

		RF Performance		
		Tx-EVM	***Tx-gain***	***Rx-gain***
10σ-filtered Learning Set *12,067 instances*	Coef. of Variation	11.0%	2.9%	3.7%
	% of good circuits	77.4%	98.5%	100%
	% of bad circuits	22.6%	1.5%	0%
6σ-filtered Learning Set *8,295 instances*	Coef. of Variation	11.1%	2.6%	3.7%
	% of good circuits	77.8%	99.1%	100%
	% of bad circuits	22.2%	1.5%	0%

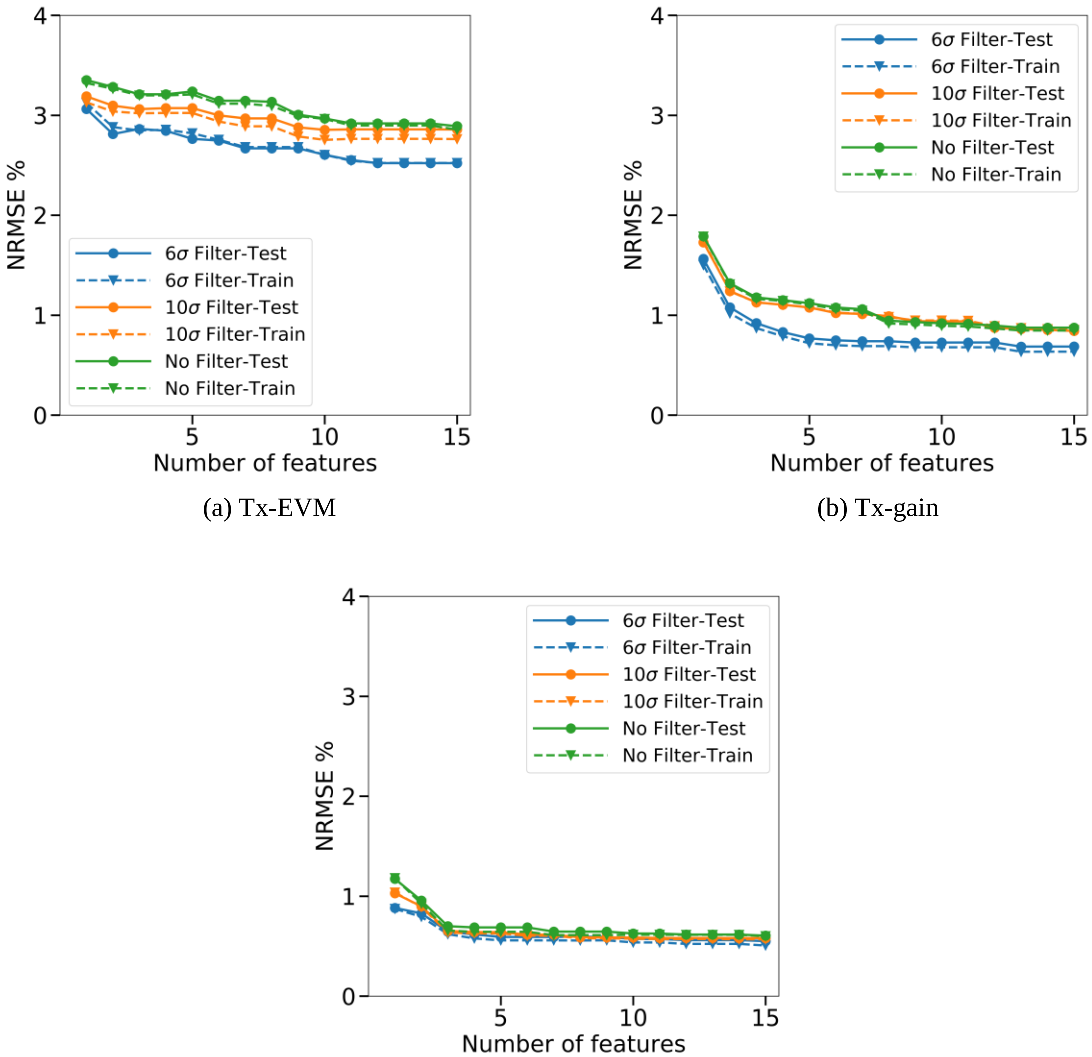
5 Results

The methodology presented in Section 3 has been applied to our case study. Results are commented in this section, first regarding the selection of pertinent models, then regarding the efficiency of a classical indirect test implementation, and finally regarding the efficiency of a two-tier adaptive test flow solution.

5.1 Model selection

Results regarding the accuracy that can be achieved in the prediction of the three RF performances are summarized in Figure 8 and Figure 9, for MARS and SVM models respectively. These figures report, for each RF performance, the evolution of the NRMSE score with respect to the number of IMs used in the construction of the model, considering either the original learning set or the filtered learning sets.

Several comments arise from the analysis of these graphs. A first general comment is that the different regression models built for each RF specification do not suffer from overfitting since there is no strong discrepancy between the NRMSE scores evaluated on train and test sets, for both types of regression models. Nevertheless, a slight advantage can be observed on this point for MARS models compared to SVM models. A second comment is that, whatever the model type, the level of accuracy differs over the different RF performances. Indeed, an NRMSE score below 1% can be achieved for the Tx-gain and Rx-gain performances, for both model types. The NRMSE score remains significantly higher for the Tx-EVM, with a best score around 2.5% in case of a MARS model and around 1.6% in case of a SVM model. Finally, the last comment concerns the influence of the learning population. Its impact is mostly visible on the prediction of the Tx-EVM performance. For both model types, we observe that the use of a filter leads to an improvement in the accuracy of the constructed models, especially in the case of a strict filter.

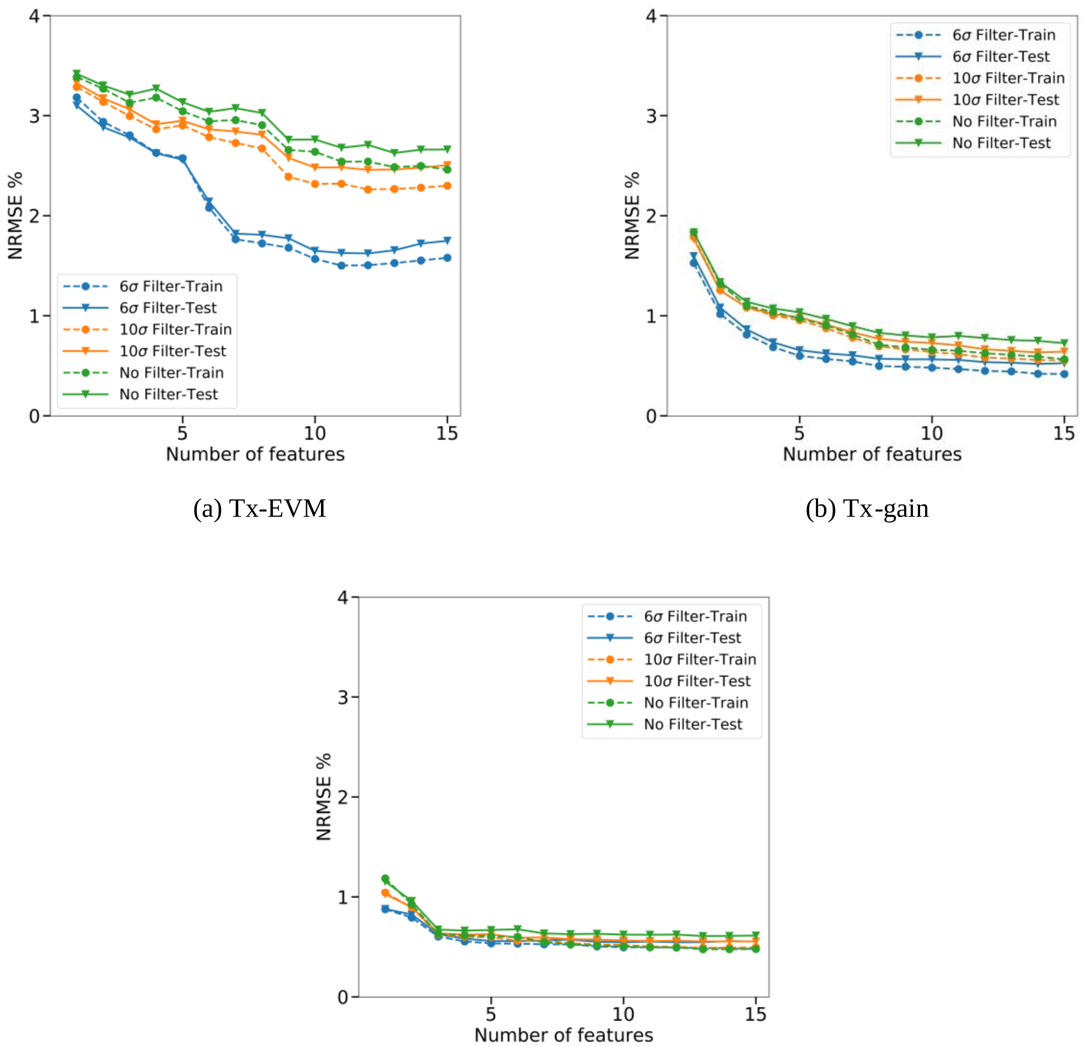


(a) Tx-EVM

(b) Tx-gain

(c) Rx-gain

Fig.8. NRMSE score achieved on train and test sets for the different scenarios of learning population - MARS model



(a) Tx-EVM

(b) Tx-gain

(c) Rx-gain

Fig.9. NRMSE score achieved on train and test sets for the different scenarios of learning population - SVM model

From this exploratory phase, the best model (i.e. the one with the lowest NRMSE score on the test set) can be selected for each RF performance and for the different scenarios. Results are summarized in Table IV, which reports for each model the number of selected IMs together with the NRMSE scores computed on train and test sets, for both model types.

These results confirm the general trends previously observed on the graphs. Indeed, we observe that whatever the learning set, the difference between the NRMSE scores computed on training and test sets never exceeds 0.2%, clearly indicating that there is no over-fitting. Regarding the improvement brought by the use of a filter, it can be considered as negligible in case of the relaxed filter with a reduction of the NRMSE score that remains inferior to 0.2% over the 3 RF performances for both model types. In case of the strict filter, it is also negligible regarding the Rx-gain and Tx-gain for both model types (NRMSE reduction less than 0.2%). It is more significant regarding the Tx-EVM, especially for the SVM model with an NRMSE reduction around 1% while it is only of 0.37% for the MARS model.

TABLE IV. SUMMARY OF BEST RESULTS ACHIEVED UNDER DIFFERENT SCENARIOS OF LEARNING POPULATION FOR THE 3 RF PERFORMANCES

		MARS			***SVM***		
		# IMs	***NRMSE*** *Train Set*	***NRMSE*** *Test Set*	***# IMs***	***NRMSE*** *Train Set*	***NRMSE*** *Test Set*
Original Learning Set	Tx-EVM	15	2.86%	2.89%	13	2.48%	2.63%
	Tx-gain	13	0.85%	0.87%	15	0.56%	0.76%
	Rx-gain	15	0.60%	0.60%	13	0.48%	0.61%
10σ-filtered Learning Set	Tx-EVM	10	2.76%	2.86%	12	2.26%	2.46%
	Tx-gain	15	0.85%	0.84%	14	0.56%	0.63%
	Rx-gain	14	0.57%	0.58%	15	0.49%	0.55%
6σ-filtered Learning Set	Tx-EVM	12	2.52%	2.52%	12	1.51%	1.62%
	Tx-gain	13	0.63%	0.68%	14	0.42%	0.52%
	Rx-gain	15	0.51%	0.55%	12	0.49%	0.54%

Globally, these results are positive for the implementation of the indirect test strategy since they show that it is possible to build quite accurate models for the three RF performances. The best solution is obtained using SVM models constructed on a learning population filtered with a strict filter. In this case, we obtain an accuracy of 0.54% for Rx-gain prediction, 0.52% for Tx-gain prediction and 1.62% for Tx-EVM prediction.

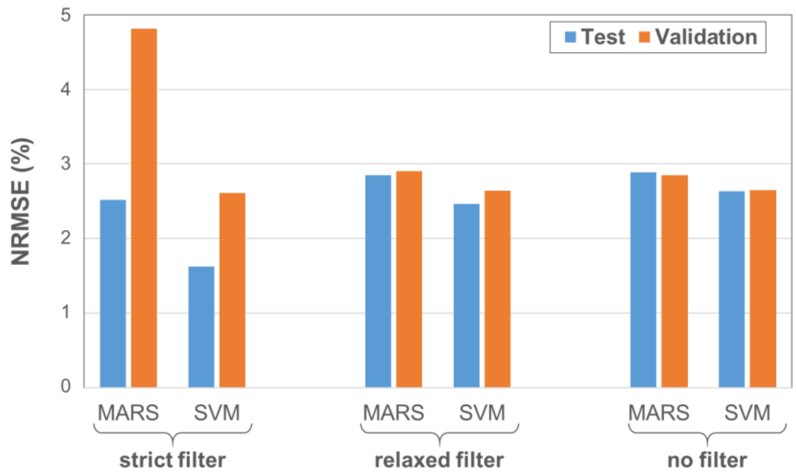
5.2 Efficiency of the classical indirect test implementation

In this part, we explore the efficiency that can be achieved with a classical indirect test implementation, i.e. all circuits are evaluated using only the indirect test flow and there is no circuit directed to a regular specification test flow (tolerance zone is set to zero). Additionally, we also investigate the influence of using (or not) a filter during the initial learning phase onto the fore-mentioned indirect test efficiency. Results presented in this section are obtained using the best MARS and SVM models constructed in the training phase for each scenario of learning population. Results are summarized in Figures 10, 11 and 12 for the three RF performances, in which the NRMSE and MR scores achieved on the test and validation sets are compared for the different scenarios of learning population.

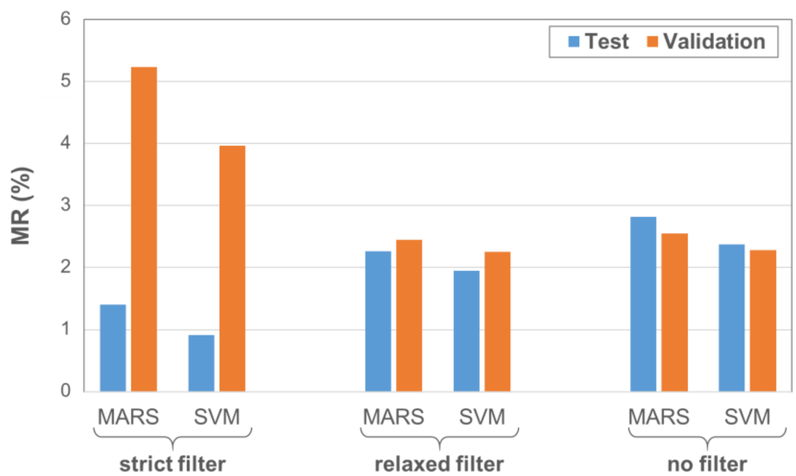
A first evident comment arises: although working with a filtered population during the initial learning phase permits to improve the quality of the constructed models, these models are nevertheless not able to correctly handle all circuits that can be encountered during the production testing phase, which are emulated by the validation set. Indeed, the best results in terms of NRMSE and MR scores achieved on the validation set are actually obtained when the models are built with the original learning population. Detailed comments on the influence of the filter are provided hereafter.

- When the strict filter is used, we can observe a significant degradation between the NRMSE and MR scores determined on the test set and the ones achieved on the validation set. The strongest difference is observed for the Tx-EVM (Figure 10). In this case, the NRMSE score increases by +2.3% for the MARS and +1% for the SVM model. Similarly, the MR score increases by +3.8% for the MARS model and +3.1% for the SVM model. For the Tx-gain (Figure 11), we observe a smaller increase of the NRMSE score by +0.4% for the MARS model and +0.5% for the SVM model, but still a significant increase of MR score by +2.0% for the MARS model and +1.6% for the SVM model. Finally, for the Rx-gain (Figure 12), the impact is mostly visible in case of the MARS model. Indeed, the NRMSE score increases by +1.3% and although a perfect MR score of 0% is expected from the result on the test set, 0.39% of the circuits of the validation set are misclassified. The degradation of the NRMSE score is more limited in case of the SVM model with an increase of only +0.4% and the perfect MR score of 0% is preserved. Globally over the three RF performances, SVM models built under this learning scenario outperform MARS models in terms of both NRMSE and MR scores achieved on the validation set.
- When the relaxed filter is used, the degradation of the NRMSE and MR scores between the test and validation sets lessens. Indeed in this case, the increase of NRMSE and MR scores does not exceed +0.3% and +0.7% respectively over the three RF performances, for both model types. Still, SVM models perform slightly better than MARS models.

- Finally when the learning is performed on the original population, the difference between the NRMSE and MR scores determined on the test set and the ones achieved on the validation set becomes negligible. Indeed, the maximum difference observed over the three RF performances is only of 0.06% for the NRMSE score and 0.3% for the MR score, for both model types (note that the difference in the scores between test and validation sets is positive is some cases and negative in other cases). SVM models present a slight advantage compared to MARS models, but not really significant.

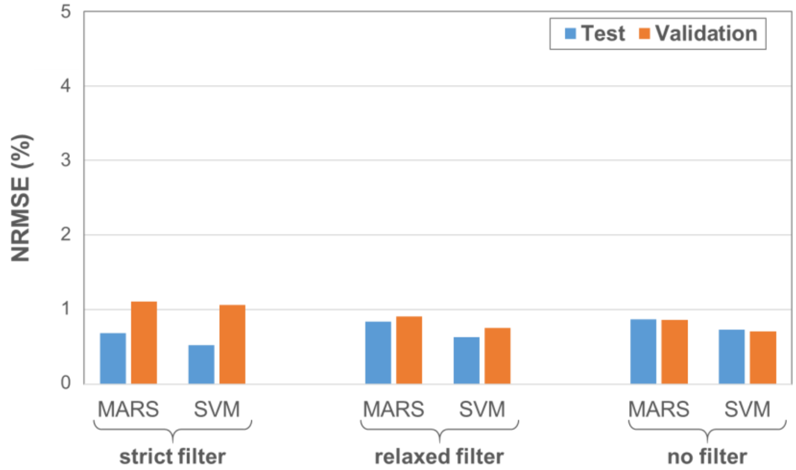


(a) NRMSE

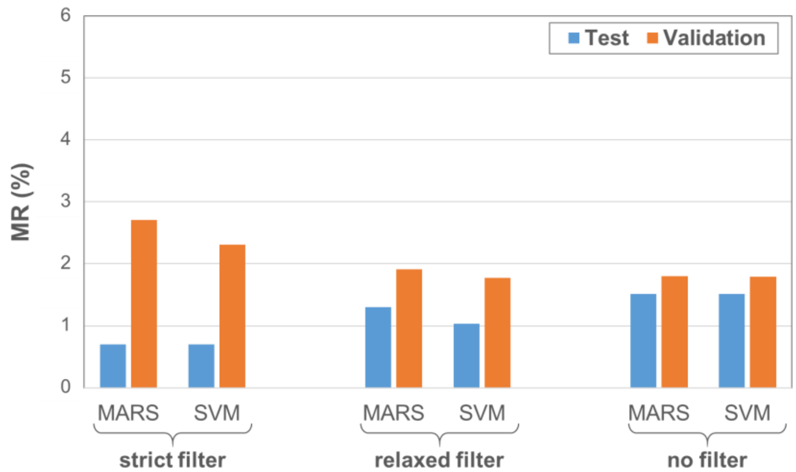


(b) MR

Fig.10. NRMSE and MR scores achieved on test and validation sets for the different scenarios of learning population - Tx-EVM

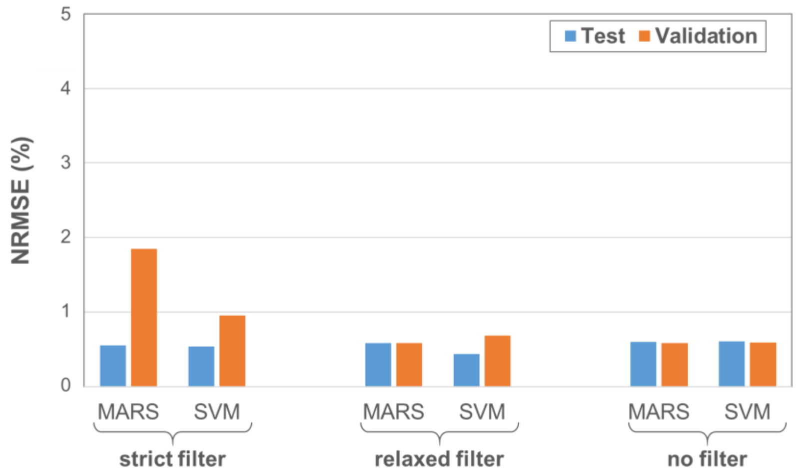


(a) NRMSE

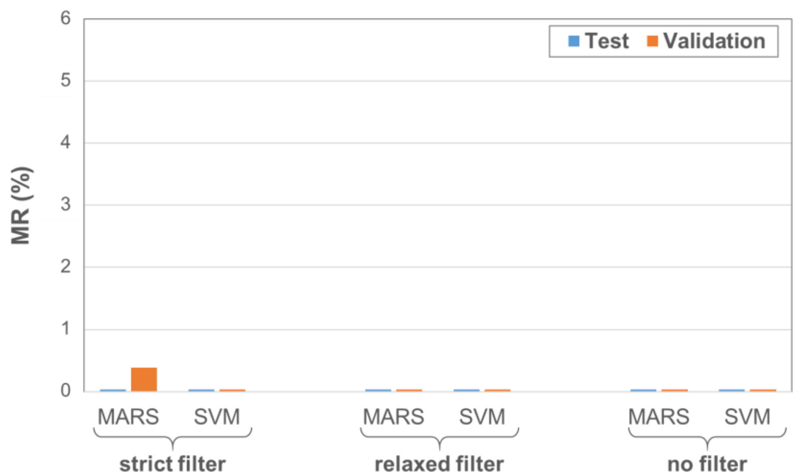


(b) MR

Fig.11. NRMSE and MR scores achieved on test and validation sets for the different scenarios of learning population - Tx-gain



(a) NRMSE



(b) MR

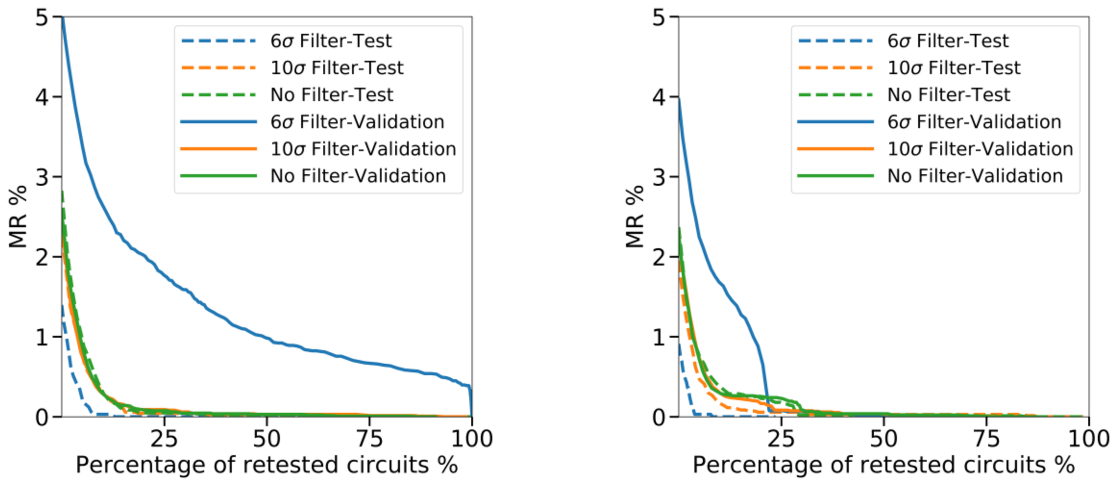
Fig.12. NRMSE and MR scores achieved on test and validation sets for the different scenarios of learning population - Rx-gain

To conclude on this study regarding the classical implementation of the indirect test strategy, a fairly good efficiency is attained for the practical case study investigated in this paper when the models are built on the original learning set (unfiltered). Indeed, evaluation on the validation set leads to a low MR around 2.5% and 1.8% for the Tx-EVM and Tx-gain respectively, and the ideal MR of 0% for the Rx-gain for both types of regression models. Globally over the three RF performances, the misclassification rate is around 4%. Note that this misclassification rate is higher than the one achieved on each individual RF performance because the circuits misclassified for a given performance are not necessarily the same as the ones misclassified for another performance. However, despite the drastic testing cost reduction offered by this solution where all the circuits are evaluated based only on low-cost indirect measurements, a misclassification rate around 4% might not be sufficient to comply with the industrial test quality constraints.

5.3 Efficiency of the two-tier adpative indirect test flow

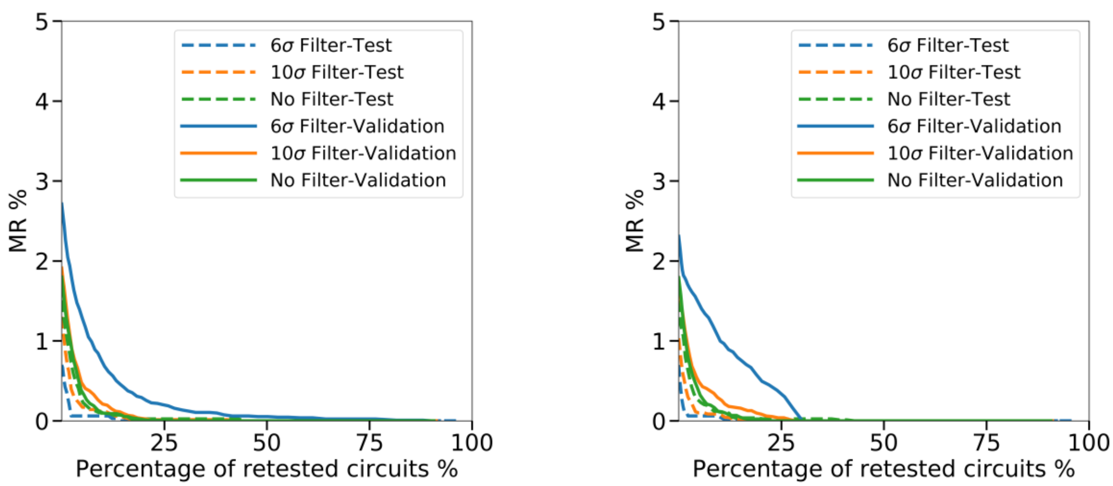
In this part, we present results regarding the benefit that can be brought by the implementation of a two-tier adaptive test flow, in particular regarding the tradeoff between test quality and test cost. As mentioned in Section II, this tradeoff depends on the size of the tolerance zone around the test limits.

Results are summarized in Figures 13 and 14, which report the trade-off curves between MR score and percentage of retested circuits obtained by varying the size of the tolerance zone, for both model types. Note that these curves are presented only for the Tx-EVM and Tx-gain performances since the ideal MR of 0% is achieved for the Rx-gain without the need of retesting any devices. These results indicate again that the use of a filter during the learning phase (especially the strict one) is not recommended, since there is a huge difference between the trade-off curve evaluated on the test set and the one observed on the validation set. Moreover, the decrease in the MR score observed on the validation set is much slower than the one obtained when the learning is performed on the original population. These results also clearly demonstrate that it is possible to significantly improve the test quality compared to a classical indirect test implementation. Indeed, with a learning performed on the original population, there is a rapid decrease of the MR score observed on the validation set, which means that the test quality improvement can be obtained with only a limited number of devices that need to be retested through a conventional specification test. In particular, it is possible to attain a very low MR score below few tenths of percent with a majority of devices that are tested using only the low-cost indirect measurements.



(a) MARS (b) SVM

Fig.13. Tradeoff curves between MR and percentage of retested devices –Tx-EVM



(a) MARS (b) SVM

Fig.14. Tradeoff curves between MR and percentage of retested devices –Tx-gain

For the sake of a concrete illustration, an arbitrary target of a MR score below 0.1% for each RF performance has been fixed. Based on devices from the test set, the size of the tolerance zone necessary to fulfill this constraint has been determined for each RF performance; the efficiency of the two-tier adaptive test flow has then been evaluated on the validation set. Results are summarized in Table V (results obtained under the best learning scenario with no filter applied on the learning population).

TABLE V. SUMMARY OF RESULTS ACHIEVED BY THE TWO-TIER ADAPTIVE TEST FLOW WITH A MR TARGET OF 0.1% FOR EACH RF PERFROMANCE

	MARS			SVM		
	MR		**Retested devices**	**MR**		**Retested devices**
	Test Set	Valid Set	Valid Set	Test Set	Valid Set	Valid Set
Tx-EVM	0.10%	0.14%	15.7%	0.10%	0.12%	29.0%
Tx-gain	0.10%	0.09%	12.0%	0.10%	0.08%	12.4%
Rx-gain	0%	0%	0%	0%	0%	0%
ALL RF Perf.	**0.19%**	**0.23%**	**23.8%**	**0.19%**	**0.20%**	**31.4%**

These results confirm that the two-tier adaptive test flow permits to reach a substantial reduction of the test cost while preserving a very good test quality. Indeed, the targeted MR score of 0.1% can be attained for each RF performance; the difference between the MR score anticipated on the test set and the one evaluated on the validation set remains inferior to 0.04%, for both model types. Only a limited number of devices need to be retested to ensure this quality, especially when considering a MARS model, i.e. around 16% for the Tx-EVM, 12% for the Tx-gain, and 0% for the Rx-gain. The percentage of retested devices is significantly higher for the Tx-EVM in case of a SVM model, with a value reaching 29%, while it remains around 12% for the Tx-gain, and 0% for the Rx-gain.

For both model types, the global misclassification rate achieved over the three RF performances is around 0.2%, so higher than the one targeted on each individual RF performance. The global MR score achieved over the three RF performances actually corresponds to the sum of the MR score achieved on each individual performance, indicating that circuits misclassified with respect to a given performance are different from the circuits misclassified for another performance. Regarding the percentage of retested devices, the global percentage over the three RF performances is also higher than the individual percentage on one performance, but inferior to the sum of the individual percentages. This indicates that among all circuits directed to the second tier of the test flow, a number of them present a low confidence for more than one RF performance.

To conclude on this study regarding the implementation of a two-tier adaptive indirect test flow, a very good test quality can be achieved for this practical case study, with only about 0.2% misclassified devices over the three RF performances while a majority of devices are processed using only the low-cost indirect measurements leading to substantial saving in the test costs. For this adaptive test flow, MARS models seem more performing than SVM models since, for the same level of misclassification rate, more than 76% of the devices are evaluated by the indirect test tier when using MARS models, and only 69% when using SVM models. Note that all these numbers correspond to worst-case results because they are established on a population manufactured with corner process conditions on purpose. We can expect lower numbers, especially the percentage of circuits that need to be retested, in the regular context of production testing where circuits are manufactured under normal process conditions.

Finally note that in this study, only the global misclassification rate has been considered without making a distinction between devices that lead to yield loss and devices that lead to test escape. However, in general, yield loss is preferred to test escapes. A possible adaptation of the methodology is to introduce this distinction and to make the choice of the tolerance zone by giving more weight to test escapes than yield loss, and to include the possibility of a non-symmetrical tolerance zone around the test limit. These refinements should permit to bias the adaptive indirect test framework towards zero test escape.

6 Conclusion

In this paper, we have investigated on a practical case study whether it is possible to benefit from the potential test cost reduction offered by the indirect test strategy without compromising the test quality. We have proposed an original implementation of a two-tier adaptive test flow that relies on the use of a tolerance zone around test limits in order to establish the confidence in the decision proposed by the indirect test; only devices with sufficient confidence are processed by the indirect test while others are directed to a second tier where they are evaluated by a standard specification test. A methodology has been defined in order to make the pertinent choices for the efficient implementation of this test flow.

Particular attention has been paid to the composition of the learning set, especially with regard to the presence of circuits that present outlying values. These circuits can be easily identified with a simple one-dimensional filter. In this study, we observed that it is not pertinent to exclude these circuits from the learning set. Indeed, although working with a filtered population improves the accuracy of the models built in the training phase, it results in a degradation of the test efficiency observed on the validation set, which is representative of the test efficiency that will be achieved during the production testing phase. Nevertheless, it should be highlighted that for this particular study, the circuits exhibit outlying values only with respect to the indirect measurements. The use of a filter might be relevant in the case of circuits that exhibit outlying values with respect to the RF performances.

Finally, results clearly demonstrated the value of the two-tier adaptive test flow, which allows attaining a very good test quality, while achieving a substantial reduction in the test costs. Indeed, in this case study, the misclassification rate attained by a classical implementation of the indirect test strategy remains above few percents, in the best conditions. Using the two-tier indirect test flow, a misclassification rate below few tenths of percent can be achieved with less than 25% of the devices that have to go through a standard specification test. Using the proposed methodology, test engineers have multiple choices at their disposal to ensure an efficient implementation of indirect testing.

Acknowledgment

This work has been carried out under the framework of PENTA-EUREKA project “HADES: Hierarchy-Aware and secure embedded test infrastructure for Dependability and performance Enhancement of integrated Systems”.

References

[1] P.N. Variyam, A. Chatterjee, “Enhancing test effectiveness for analog circuits using synthesized measurements”, Proc. IEEE VLSI Test Symposium (VTS), pp. 132-137, 1998.

[2] 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.

[3] S. Ellouz, P. Gamand, C. Kelma, B. Vandewiele and B. Allard, “Combining internal probing with artificial neural networks for optimal RFIC testing”, Proc. IEEE International Test Conference (ITC), p. 9, 2006.

[4] H. Stratigopoulos, S. Mir and Y. Makris, “Enrichment of limited training sets in machine-learning-based analog/RF test”, Proc. IEEE/ACM Design, Automation and Test in Europe Conference (DATE), pp. 1668–1673, 2009.

[5] L. Abdallah, H. Stratigopoulos, C. Kelma and S. Mir, “Sensors for built-in alternate RF test”, Proc. IEEE European Test Symposium (ETS), pp. 49-54, 2010.

[6] A. Dimakos, M. Andraud, L. Abdallah, H. Stratigopoulos, E. Simeu and S. Mir, “Test and Calibration of RF Circuits Using Built-in Non-intrusive Sensors”, Proc. IEEE Computer Society Annual Symp. on VLSI (ISVLSI), pp. 627-627, 2015.

[7] M. J. Barragan, R. Fiorelli, G. Leger, A. Rueda and J. L. Huertas, “Improving the Accuracy of RF Alternate Test Using Multi-VDD Conditions: Application to Envelope-Based Test of LNAs”, Proc. IEEE Asian Test Symposium (ATS), pp. 359-364, 2011.

[8] H. Ayari, F. Azais, S. Bernard, M. Comte, M. Renovell, V. Kerzerho, O. Potin and C. Kelma, “Smart selection of indirect parameters for DC based alternate RF IC testing”, Proc. IEEE VLSI Test Symposium (VTS), pp. 19-24, 2012.

[9] M. J. Barragan and G. Leger, “A Procedure for Alternate Test Feature Design and Selection”, IEEE Design & Test, Vol. 32, no. 1, pp. 18-25, 2015.

[10] M. J. Barragan, G. Leger, F. Cilici, E. Lauga-Larroze, S. Bourdel and S. Mir, "On the use of causal feature selection in the context of machine-learning indirect test", Proc. Design Automation Test Conference (DATE), pp. 276-279, 2019.

[11] H. Stratigopoulos, "Machine learning applications in IC testing", Proc. IEEE European Test Symposium (ETS), p. 10, 2018.

[12] A.D. Singh, C.M. Krishna, "On Optimizing VLSI Testing for Product Quality Using Die-Yield Prediction", IEEE Transactions On Computer-Aided Design of Integrated Circuits and Systems, Vol.12, no. 5, pp. 695-709, 1993.

[13] T. J. Powell, J. Pair, M. St. John and D. Counce, "Delta Iddq for Testing Reliability", Proc. IEEE VLSI Test Symposium (VTS), pp. 439-443, 2000.

[14] E. Yilmaz and S. Ozev, "Adaptive Test Flow for Mixed-Signal Circuits Using Learned Information from Device under Test", Proc. IEEE International Test Conference (ITC), paper 23.1, 2010.

[15] H. Stratigopoulos, 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.

[16] N. Kupp et al., "Confidence Estimation in Non-RF to RF Correlation-Based Specification Test Compaction", Proc. IEEE European Test Symposium (ETS), pp. 35-40, 2008.

[17] H. Ayari, F. Azais, S. Bernard, M. Comte, V. Kerzerho, O. Potin, M. Renovell, "Making predictive analog/RF alternate test strategy independent of training set size", Proc. IEEE International Test Conference. (ITC), p.9, 2012.

[18] H. Stratigopoulos, S. Mir, "Adaptive Alternate Analog Test", IEEE Design & Test of Computers, Vol. 29, no. 4, pp. 71-79, 2012.

[19] G. Leger, "Combining adaptive alternate test and multi-site", Proc. IEEE/ACM Design, Automation and Test in Europe Conference (DATE), pp. 1389–1394, 2015.

[20] H. El Badawi, F. Azais, S. Bernard, M. Comte, V. Kerzerho, F. Lefevre, I. Gorenflot, "Implementing indirect test of RF circuits without compromising test quality: a practical case study", Proc. IEEE Latin-American Test Symposium (LATS), pp. 1-6, 2020.

[21] H. Stratigopoulos, S. Mir, E. Acar and S. Ozev, "Defect filter for alternate RF test", Proc. IEEE European Test Symposium (ETS), pp. 265–270, 2010.

[22] H. El Badawi, F. Azais, S. Bernard, M. Comte, V. Kerzerho, F. Lefevre, "Investigations on the Use of Ensemble Methods for Specification-Oriented Indirect Test of RF Circuits", Journal of Electronic Testing (JETTA), Vol. 36, pp. 189–203, 2020.

[23] S. Larguech, F. Azais, S. Bernard, M. Comte, V. Kerzerho and M. Renovell, "A Generic Methodology for Building Efficient Prediction Models in the Context of Alternate Testing", Proc. International Mixed-Signal Test Workshop (IMSTW), p. 6, 2015.