



HAL
open science

Indirect Analog / RF IC testing: Accuracy & Robustness improvements

Haithem Ayari

► **To cite this version:**

Haithem Ayari. Indirect Analog / RF IC testing: Accuracy & Robustness improvements. Micro and nanotechnologies/Microelectronics. Université Montpellier II - Sciences et Techniques du Languedoc, 2013. English. NNT: 2013MON20062 . tel-00998677

HAL Id: tel-00998677

<https://theses.hal.science/tel-00998677>

Submitted on 2 Jun 2014

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **Université Montpellier II**

Préparée au sein de l'école doctorale **Information,
Structure et Systèmes (I2S)**
Et de l'unité de recherche **Laboratoire d'Informatique,
de Robotique et de Microélectronique de Montpellier
(LIRMM)**

Spécialité : **Systèmes Automatiques et
Microélectronique (SYAM)**

Présentée par **Haithem Ayari**

**Test indirect des circuits analogique et RF :
Contribution pour une meilleur précision et
robustesse**

Soutenue le 12 décembre 2013 devant le jury composé de

Dr. Serge Bernard, Chargé de Recherche, CNRS-LIRMM	Directeur de these
Dr. Florence Azais, Chargé de Recherche, CNRS-LIRMM	Co-directrice de these
Dr. Salvador Mir, Directeur de Recherche, CNRS-TIMA	Rapporteur
Pr. Ian O'connor, Professeur, Université de Lyon-INL	Rapporteur
Dr. Michel Renovell, Directeur de Recherche, CNRS-LIRMM	Examineur
Dr. Christophe Kelma, Ingénieur, NXP Semiconductors	Examineur
Dr. Vincent kerzerho, Chargé de Recherche, CNRS-LIRMM	Invité
Dr. Mariane Comte, Maître de Conférences, Université Montpellier II- LIRMM	Invitée

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Microélectronique

Formation Doctorale : Systèmes Automatiques et Micro-électronique (SYAM)

Ecole Doctorale : Information, Structures, Systèmes (I2S)

Présentée et soutenue publiquement par :

Haithem Ayari

le 12 / 12 / 2013

***Test Indirect des circuits Analogique et RF: Contribution
pour une meilleure précision et robustesse***

Jury :

Serge Bernard	Chargé de recherche CNRS - LIRMM	Directeur de thèse
Florence Azaïs	Chargé de recherche CNRS - LIRMM	Codirectrice de thèse
Ian O'connor	Professeur Université de Lyon - INL	Rapporteur
Salvador Mir	Directeur de recherche CNRS - TIMA	Rapporteur
Michel Renovell	Directeur de recherche CNRS - LIRMM	Examineur
Christophe Kelma	Ingénieur Senior à NXP Semiconductors	Examineur
Dr. Vincent kerzerho	Chargé de Recherche, CNRS-LIRMM	Invité
Dr. Mariane Comte	Maître de Conférences, UM2-LIRMM	Invitée

***Indirect Analog/RF IC testing: Accuracy & Robustness
improvements***

©Copyright by Haithem Ayari 2013
All right reserved

To the memory of my grand-parents

To my parents & my two sisters

To my wife & my son

"We will either find a way, or make one"

Hannibal Barca, a Carthaginian general

This work is carried out within the framework of ISyTest (Institute for System Testing) a joint institute between NXP Semiconductors and LIRMM.

Acknowledgments

First, I would like to thank Dr. Salvador Mir, research director at CNRS and Pr. Ian O'connor, professor at Ecole Centrale de Lyon for their interest on my work and for accepting to be part of my dissertation committee.

There are several people who significantly influenced this dissertation and I would like to thank them here.

It's with my sincerest gratitude that I thank my advisors Dr. Serge Bernard and Dr Florence Azaïs for providing me the academic freedom to pursue research problems that truly interested me, and for that I am very grateful. Their genuine interests in my progress, technical insights and pursuit of perfection have largely been responsible for improving me.

I would like to thank Dr. Michel Renovell for his guidance and advice which has helped me shaping my career towards a positive direction of my life.

I would like also to express my thanks to Dr. Marianne Comte and Dr. Vincent kerzérho for the technical support that they provided me during the thesis.

My thanks are to Christophe Kelma from NXP Semiconductors for providing the test vehicles data and for fruitful discussions.

Finally, I would like to dedicate this work to my mother. She always believed in me, without her help and encouragement, I could have never achieved this milestone in my life.

Montpellier, July 31th, 2013

Haïthem Ayari

Abstract

The conventional approach for testing *RF* circuits is specification-based testing, which involves verifying sequentially all specification requirements that are promised in the data sheet. This approach is a long-time effective test approach but nowadays suffers from significant drawbacks. First, it requires generation and capture of test signals at the DUT operating frequency. As the operational frequencies of *DUT* are increasing, it becomes difficult to manage signal generation and capture using *ATE*. As a consequence, there is a need of expensive and specialized equipment. In addition, as conventional tests target several parameters, there is a need of several data captures and multiple test configurations. As a consequence, by adding setting time between each test and test application time, the whole test time becomes very long, and the test board very complex. Another challenge regarding *RF* circuit testing is wafer-level testing. Indeed, the implementation of specification-based tests at wafer level is extremely difficult due to probing issues and high parasitic effects on the test interface. Moreover, multi-site testing is usually not an option due to the small count of available *RF* test resources, which decreases test throughput. Hence, the current practice is often to verify the device specifications only after packaging. The problem with this solution is that defective dies are identified late in the manufacturing flow, which leads to packaging loss and decreases the global yield of the process. In order to reduce production costs, there is therefore a need to develop test solutions applicable at wafer level, so that faulty circuits can be removed very early in the production flow. This is particularly important for dies designed to be integrated in Systems-In-Package (*SiP*). In this context, a promising solution is to develop indirect test methods. Basically, it consists in using *DUT* signatures to non-conventional stimuli to predict the result of conventional tests. The underlying idea is to learn during an initial phase the unknown dependency between simple measurements and *RF* parameters. This dependency can then be modeled through prediction functions. During the testing phase, only the indirect measurements are performed and specifications are predicted using the prediction model built in the learning phase. Our work has been focused on two main directions. First, we have explored the implementation of the indirect test method based on *DC* measurements for *RF* circuits and we have proposed a methodology to select the most appropriate set of *DC* parameters. Results from two test vehicles (a *LNA* using electrical simulations and a *PA* using real production data) indicate that the proposed methodology allows precise estimation of the *DUT* performances while minimizing the number of *DC* measurements to be carried out (i.e. test solution cost). Second, we have proposed a novel implementation scheme of the indirect test strategy in order to improve confidence in predictions and to overcome the effect of limited training set sizes. The idea is to exploit model redundancy in order to identify, during the production testing phase, devices with suspect predictions; these devices are then removed from the alternate test tier and directed to a second tier where further testing may apply.

Résumé

Contrairement aux circuits numériques qui peuvent comporter plusieurs centaines de millions de transistors, les circuits analogiques et radio fréquence sont généralement constitués d'un nombre réduit d'éléments ne dépassant que rarement la centaine de transistors. Malgré tout, le test de ces circuits est un problème particulièrement critique. Conventionnellement, ce test se base sur la mesure des performances du circuit et la comparaison des résultats obtenus aux spécifications décrites dans le cahier des charges. Cette approche est considérée, de longue date, l'approche la plus efficace mais aujourd'hui elle souffre de sérieux inconvénients. Tout d'abord, elle nécessite la génération et la capture des signaux de test à la fréquence de fonctionnement du circuit sous test. Comme les fréquences de fonctionnement des circuits analogiques et radio fréquence sont en augmentation, il devient difficile de gérer la génération et la capture des signaux au niveau de l'équipement de test. En conséquence, il devient nécessaire d'utiliser des équipements de test spécialisés et extrêmement coûteux. En outre, comme cette approche conventionnelle cible le test de plusieurs paramètres du circuit, il est nécessaire d'effectuer plusieurs captures de données sous multiples configurations de test. En conséquence, en ajoutant le temps de configuration de chaque mesure au temps de l'application des signaux, le temps de test total devient très long et les cartes de test complexes. Un autre défi se présente vis-à-vis le test des circuits radio fréquence au niveau plaquettes (wafer). En effet, la mise en œuvre de tests basés sur les spécifications au niveau de la plaquette est extrêmement difficile en raison des parasites élevés sur les sondes de l'interface de test. En outre, le test multi-site n'est généralement pas une option envisageable en raison des ressources RF limitées sur les testeurs, ce qui réduit la cadence de test. En conséquence, la pratique actuelle est souvent limitée à la vérification des circuits après la mise en boîtier. Le problème d'une telle approche c'est qu'elle ne permet pas d'identifier les pièces défectueuses à un niveau tôt du processus de fabrication et génère une diminution du rendement global et augmente le coût de fabrication. Afin de réduire les coûts de production, il est donc nécessaire de développer des solutions de test applicables au niveau de la plaquette, de sorte que les circuits défectueux peuvent être retirés très tôt dans le flux de production. Ceci est particulièrement important dans le cadre des circuits intégrés de type systèmes en boîtier (SiP). Dans ce contexte, une solution prometteuse consiste à développer des méthodes de tests indirectes. Fondamentalement, elles consistent à utiliser des signatures et des stimuli non conventionnels pour prédire les performances du circuit sous test. L'idée sous-jacente est d'apprendre au cours d'une phase initiale de la dépendance inconnue entre des mesures simples et les paramètres RF. Cette dépendance peut alors être modélisée par des fonctions de prédiction. Au cours de la phase de test, seules les mesures simples, dites indirectes, sont effectuées et les performances du circuit sont estimées en utilisant le modèle de prédiction construit dans la phase d'apprentissage. Malgré que cette approche semble prometteuse pour l'industriel, elle ne suscite pas complètement confiance en ce genre d'approche. Ceci est dû au fait que les valeurs des performances du circuit sont prédites et peuvent parfois avoir des erreurs. Ce travail vise à améliorer la confiance en ce genre de test en améliorant la précision et la robustesse des valeurs prédites.

Table of Contents

Acknowledgments	vii
Abstract	viii
Table of Contents	x
Introduction	1
Chapter I	5
State of the Art of Analog/RF IC Testing	5
I.1 Introduction.....	5
I.2 Industrial test generalities	7
I.3 Analog/RF ICs testing specificity	9
I.3.1 Faults in analog/RF ICs.....	9
I.3.2 Analog/RF ICs test: current practice.....	9
I.4 Cost-reduced RF IC testing strategies.....	11
I.4.1 Integrated test solution	11
I.4.2 Loopback testing	12
I.4.3 Indirect testing.....	12
I.5 Conclusion	16
Chapter II	18
Analysis of indirect testing: case of specification prediction	18
II.1 Introduction.....	18
II.2 Classical implementation of prediction-oriented indirect test.....	19
II.3 DC-based indirect test strategy	20
II.4 Presentation of the test vehicles	21
II.4.1 The Low Noise Amplifier.....	21

II.4.2	Power Amplifier	22
II.4.3	<i>RF</i> Parameters	22
II.4.3.1.	Gain	22
II.4.3.2.	Noise Figure (<i>NF</i>).....	23
II.4.3.3.	Gain Compression	24
II.4.3.4.	Third Order Intercept Point	25
II.5	Benchmarking of some machine-learning regression algorithms	27
II.5.1	Multiple Linear Regression	27
II.5.2	Multivariate Adaptive Regression Splines	28
II.5.3	Artificial Neural Network.....	29
II.5.4	Regression Trees	29
II.5.5	Test case definition.....	30
II.5.6	Results and discussion.....	32
II.6	Limitations & bottleneck of the conventional indirect test scheme	36
II.6.1	Prediction confidence: flawed predictions	36
II.6.2	Dependency of model performances with respect to TSS.....	39
II.7	Conclusion.....	43
Chapter III		46
Strategies for IMs Selection		46
III.1	Introduction	46
III.2	Benchmark of some feature selection techniques.....	47
III.2.1	Variable subset selection techniques	47
III.2.1.1.	Filters.....	47
III.2.1.2.	Wrappers.....	49
III.2.2	Test case definition.....	51
III.2.3	Results and discussion.....	52
III.3	Proposed IM selection strategy	54
III.3.1	Dimensionality reduction of IM-space	55
III.3.2	Search space construction.....	56
III.3.3	Optimized IM subset selection	56
III.3.4	Results and discussion.....	57
III.4	Conclusion.....	59

Chapter IV	61
Strategy for increasing robustness of prediction-oriented indirect test	61
IV.1 Introduction.....	61
IV.2 Outlier definitions	62
IV.2.1 Process-based outlier.....	62
IV.2.2 Model-based outlier	62
IV.3 Redundancy-based filter for model-based outliers	62
IV.3.1 Principle	62
IV.3.2 Experimental results using accurate predictive models	64
IV.4 Making predictive indirect test strategy independent of training set size	69
IV.5 Model redundancy to balance the lack of correlation between IMs and RF performance parameters	74
IV.6 Different implementation schemes of redundancy	77
IV.6.1 Motivation.....	77
IV.6.2 Principle of augmented redundancy CPC schemes.....	78
IV.6.3 Experimental Results	81
IV.7 Conclusion	85
Conclusion	87
References	91
Related publications	95
List of Figures	96
List of Tables	98
List of acronyms	99

Introduction

The recent advances in fabrication and packaging technologies have enabled the development of high performance complex Radio Frequency (*RF*) chips for a wide range of applications. *RF* chips, which are the focus of this work, are on the leading edge of technological developments and rise a significant number of production problems. Considering first the global production costs of high volume production of Integrated Circuits (*ICs*), these costs include design, manufacturing and test costs. In the recent years it has been observed that i) design costs have grown significantly but not drastically due to improved design productivity, ii) manufacturing costs have remained reasonably flat because of technological advances, iii) test costs have dramatically increased because of ever demanding requirements on the test instrumentation. Note that this is true for today highly integrated digital System-on-Chip (*SoC*) and System-in-Package (*SiP*) products manufactured in nanometer technology for which an Automatic Test Equipment (*ATE*) equipped with high-speed digital resources are required. But it is even emphasized for analog and *RF* products for which not only high-speed but also high-precision analog and *RF* test resources are required.

Indeed the conventional approach for testing analog and Radio Frequency (*RF*) devices is specification-oriented testing, which consists in measuring the majority or totality of the circuit performance parameters defined in its datasheet and comparing these values to pre-defined tolerance limits in order to sort the fabricated circuits as good or bad circuits. Typical *RF* measurements include “Gain”, “Noise Figure” (*NF*), “Third-Order Intermodulation” (*IP3*), just to name a few... This strategy, summarized by **Fig. 1**, aims at adapting the test to each kind of circuit, according to its function and performances. Another strategy, called structural-oriented testing, has been developed over the years. This strategy relies on a list of fault models to be applied to any kind of circuit regardless of its function. The specification-oriented strategy was and continues to be predominantly adopted due to the lack of widely applicable fault models. The clear advantage of specification-oriented testing is that it obviously offers good test quality, but at extremely high cost due to the required sophisticated test equipment and long test time. In addition, testing is usually applied at two different levels of the manufacturing process, i.e. first at wafer-level after silicon wafers have been fabricated

and then at package-level once circuits have been encapsulated. For all these reasons, testing costs for *RF* products are becoming the largest part of the overall costs [1] [2].

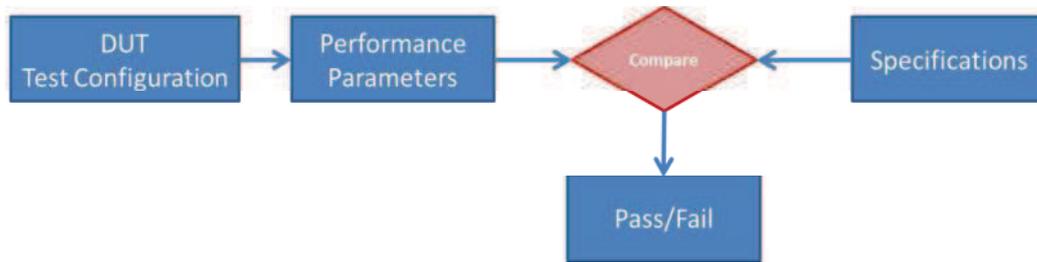


Fig. 1: Specification-based test strategy

Beyond the cost problems, technical measurement capabilities are also a challenge, especially at wafer level which is of crucial importance to guarantee Known-Good-Dies (*KGD*) [3]. Indeed *RF* measurements are severely impacted by the environment: parasitic elements, improperly calibrated equipment, external radiations, etc. Assuming a complex *RF* chip with limited access, it is very difficult if not impossible to get the correct *RF* parameters measurement even by using expensive *ATE* equipped with high-performance resources. Consequently due to the high cost and technical problems, specification-based *RF* IC testing is the major bottleneck to reduce the overall manufacturing cost in semiconductors industry.

In this context, industrials are continuously looking for novel low-cost test strategies for analog and *RF* devices to overcome the cost issue. Several techniques such as analog Built-In-Self-Test (*BIST*) and Design for Test (*DfT*), which are no longer based on *RF* specification measurements, have been investigated. They are based on signature measurements to classify good and bad devices. Another promising solution to lessen the burden of specification testing is indirect testing, also called in literature alternate testing. In this strategy, the results of specification testing are derived from a set of few Indirect Measurements (*IMs*) obtained with low-cost test equipment. The idea is to use a training set of devices in order to learn the mapping between the indirect measurements and the circuit performance parameters during a first phase; only the indirect measurements are then used during the production testing phase to perform device specification prediction and/or device classification. As a consequence, it is possible to significantly decrease the number and complexity of test configurations.

Despite the clear advantages of employing the indirect test approach and a number of convincing attempts to prove its efficiency [4] [5] [6], the deployment of this strategy in industry is limited. This is due to the fact that the *RF* parameters values are predicted and not actually measured; industrials have not sufficient confidence on the predicted *RF* parameters.

This lack of confidence is generated by the prediction model itself. Indeed it is very difficult to map all the interactions between an *RF* parameter to be predicted and the indirect measurements for the entire possible situations in a prediction model. Moreover, the prediction model is valid only on a set of devices having the same statistical properties of the set used to build the prediction model. These facts lead to inaccurate prediction of the *RF* parameter for some devices. Although the number of such devices is extremely small, the large error between actual and predicted values of these devices constitutes a serious obstacle for a large deployment of the indirect testing strategy in the industry. The objective of this work is to provide confidence in the indirect test strategy by improving prediction accuracy and ensuring robustness of the test procedure.

The first chapter is a quick overview on the analog and *RF ICs* testing state of the art. At this level, factors contributing to the cost of a given testing strategy are analyzed. Then we present the specificity of *RF IC* testing. Finally in this chapter, some cost-reduced *RF IC* testing strategies are presented.

In the second chapter, the efficiency of prediction-oriented indirect testing is deeply analyzed. First, we introduce the test vehicles which will be used for the experiments all over this work. These test vehicles are from *NXP Semiconductors* and they comprise a Low Noise Amplifier (*LNA*) and a Power Amplifier (*PA*) Then, we present results of experiments performed in order to compare some regression-fitting algorithms and highlight the limitations of the conventional implementation scheme.

The third chapter deals with the problem of selecting a pertinent set of indirect measurements that permits to accurately predict the device specifications. The efficiency of some commonly-used feature selection algorithms is investigated and an alternative selection strategy is developed with the objective to reduce the overall cost while maintaining the accuracy.

The fourth and last chapter is dedicated to study prediction confidence and model robustness versus some varying parameters. Here a strengthened implementation of the prediction-oriented indirect test is proposed. This new implementation is based on prediction model redundancy and it can be adopted to improve prediction accuracy and ensure robustness against learning conditions such as training set of limited size or use of indirect measurements with imperfect correlation with specifications.

Finally in the conclusion, the main contributions of this thesis are summarized and perspectives for future work are presented.

Chapter I

State of the Art of Analog/RF IC Testing

I.1 Introduction

The design of digital circuits follows, for decades, trends that nowadays potentially make them constituted of billions of transistors. On contrary analog/RF circuits remain made of a reduced number of basic elements, rarely exceeding a few hundreds.

From a testing point of view, the main issues stand on the side of analog/RF circuits. Indeed in parallel of the increasing number of transistors in digital circuits, the testing strategy that became predominant is called the structural testing. Thanks to the limited number states of a digital signal, it has been possible to develop a strategy that can be applied to any kind of digital circuit regardless of their function. In addition thanks to structural testing, it is possible to run parallel tests, in order to tackle the issue of the total time for testing. On the opposite, analog signals are continuous in time and amplitude, inducing infinity of possible values. Their characteristics strongly rely on the considered circuit and its function. In addition these signals are also sensitive to the using conditions such as the temperature and to the variations of manufacturing process variables. As a consequence, testing analog circuits requires the use of stimuli that are functional signals in order to measure the specifications of the considered circuit-under-test. As a consequence, testing methodologies for analog circuits are specific to each type of circuit (power amplifier, low noise amplifier, mixer...). For years, some researchers try to develop a structural test strategy for any kind of analog circuits, but it is very difficult to provide a relevant list of fault models affecting analog circuits like for digital circuits. As a consequence the specification-based test approach remains the main strategy used for testing analog circuits. Although, the specification based test strategy has the golden test quality, it is a very heavy procedure. It, due to the cost of the *RF* equipment, the calibration step, enabling multisite test and the large test time, industrials want to develop alternative analog RF test strategies to overcome these issues. This chapter provides an overview on the current practice in analog and *RF IC* test. In addition, we introduce several low-cost testing paradigms including the *DFT/BIST* based solution, the loopback testing, and

the indirect testing that offer the promise of significant test cost reduction with little or even no compromise in test quality.

I.2 Industrial test generalities

Testing integrated circuits or systems is a mandatory phase in the manufacturing process. As the functioning of an electronic device strongly relies on each circuit or systems that compose it, each sample of a circuit or system is consequently tested before commercialization. The test concern mainly two different phases in cycle of maturity of a semiconductor product: characterization phase and production phase. The objectives of product test at these two phases are radically different.

The first type is dedicated for the pre-series of a manufactured device. It is basically intended to validate the design of the circuit or system in terms of functionality and specifications. For this, a variety of tests involving different settings are performed to measure the characteristics of the manufactured circuit and most critical functional conditions. At this stage, sophisticated equipment is used; the test time and cost are not critical constraints. Once the characterization phase is complete, the high volume production of circuit is launched. In this level each manufactured device should be tested to ensure conformance to the datasheet specifications. Along the flow of *IC* manufacturing several tests are performed. We distinguish specially, to critical steps of the flow where the conformity of the devices should be checked. The first is before die dicing it called wafer sort where defective die are eliminated from the flow. The Second is after die packaging to ensure the conformance of the device to its datasheet requirements.

Given the large number of manufactured devise, typically several million per year, the test cost becomes very important criterion. The cost of a test solution results from many factors: test equipment capital, additional facilities capital (Handler, *RF* probes, etc.), operations overhead (operator, maintenance, building, etc.), Test development engineering [7] [8].

Fig. 2 emphasizes the relative importance of the test time on the overall test cost per device. The cost factors associated with testing a device at various test times are shown. Note that the primary contributor to the device test cost is the operations overhead cost followed by the test system capital cost.

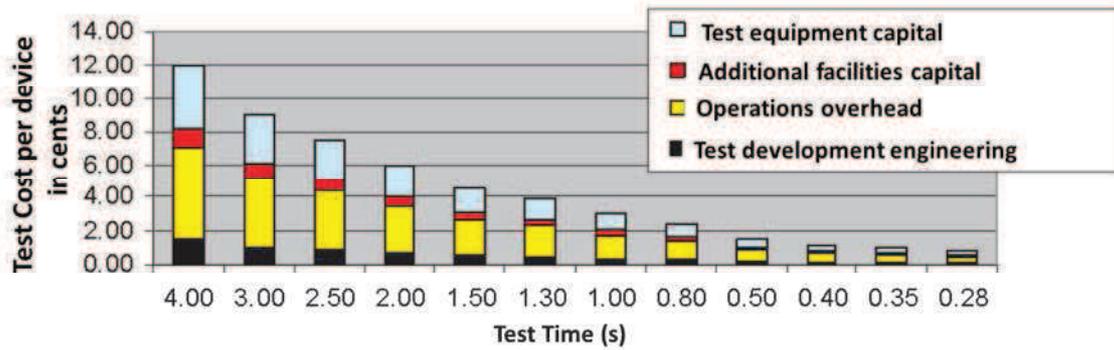


Fig. 2: Test cost versus test time.

According to this model the test time play a significant role to reduce the overall test cost per device. A total test time of 330ms leads to a one-cent-per-device total test cost. Based on the above observation, an *RF IC* test system is needed that could achieve a 300 ms (or less) per device test time to reach the targeted one-cent-per-device test cost. How to accomplish these two factors play a significant role starting with the hardware cost and then test time reduction.

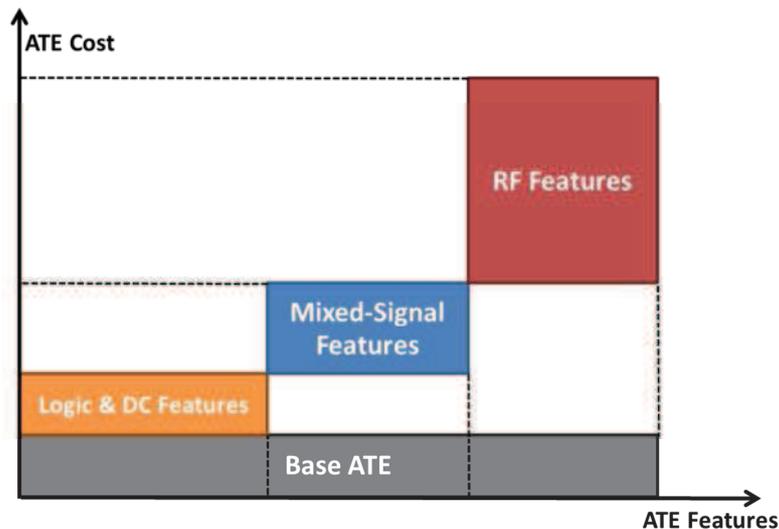


Fig. 3: ATE cost increase with additional added features

The **Fig. 3** shows the cost test equipment capital increase with adding additional features. Regarding to **Fig. 2** and **Fig. 3** *RF* devices have the highest test cost among all the types of circuits. This is due to the higher cost of the *RF* test features. This is a real obstacle for developing test solutions with a reduced cost per device. However, several techniques have been developed to relax the equipment constraints. The strategies presented in this chapter and in the next ones aim at reducing the amount due to the test equipment.

Concerning the test time, there are two main contributors: the time for the test program to run (which is linked to the time needed to set and operate each test) and how fast can the handler move the parts between the bins and the socket or move the wafer. This work focuses on the first contributor. One way to reduce the program test time is to simplify the test measurement required such as converting a test signal to a *DC* parameter instead of digitizing it. Further in this chapter, techniques for test time reduction and equipment cost reduction will be presented.

I.3 Analog/RF ICs testing specificity

I.3.1 Faults in analog/RF ICs

Faults in analog/RF ICs could be classified into two main categories: Catastrophic faults and parametric faults. Catastrophic faults (i.e. hard impact in the circuit) include generally shorts between nodes, open nodes and other hard changes in a circuit. Parametric faults (i.e. soft impacts in the circuit) are faults that do not affect the connectivity of the circuit; those are, for the most, variations of the dimensions of transistors and passive components due to a not-well controlled technology process. Moreover, the parametric faults are further categorized into global and local faults. The first one occurs when all active or passive area in the device are impacted while the second occurs only when these areas are affected locally in the circuit. Global defects usually result from fluctuation in the manufacturing environment, such as a systematic misalignment of masks or a problem which systematically affects the active areas of the transistors. The variation of manufacturing environment can also lead to local defects. In this case, it is not a systematic variation but a local variation generating slight random differences between two adjacent components: this is called mismatch. Other typical example of a local defect is constituted by a dust particle on a lithographic mask producing a disparity such as a local variation in the ratio w/l of a transistor.

I.3.2 Analog/RF ICs test: current practice

Usually, catastrophic faults are easy to detect. In this case the device has a severe dysfunction or it simply does not work. A simple continuity test is often enough to identify defective devices including this type of fault. In addition, defect models modeling shorts and opens in the circuit can be used to primary eliminate this kind of faulty circuits. The most problematic thing is how to detect devices including parametric faults. Insofar as the parametric faults affect the performance of the circuit, the device passes the continuity test

and there are no reliable fault models that can detect the faults. The only way to detect the devices affected by the parametric faults is to measure the performance of the circuits and to compare them to those defined in its datasheet.

In the context of analog/*RF ICs* the specification test is usually reserved to the final test (i.e. packaged device). For many years, the wafer sort is based on decimating devices only including catastrophic faults. As a result there was limited ability to reduce overall test costs. Another problematic fact that with novel *IC* integration methods like *SiP* (System in Package) and *3D* the industrials need for *KGD* (known good die) to develop reliable and cost efficient processes. The issue of package scrap is more problematic with these technologies. For the *RF* devices, *KGD* is synonym of specification test which performance of each device is measured. **Fig. 4** shows an example of wafer level *RF IC* test.

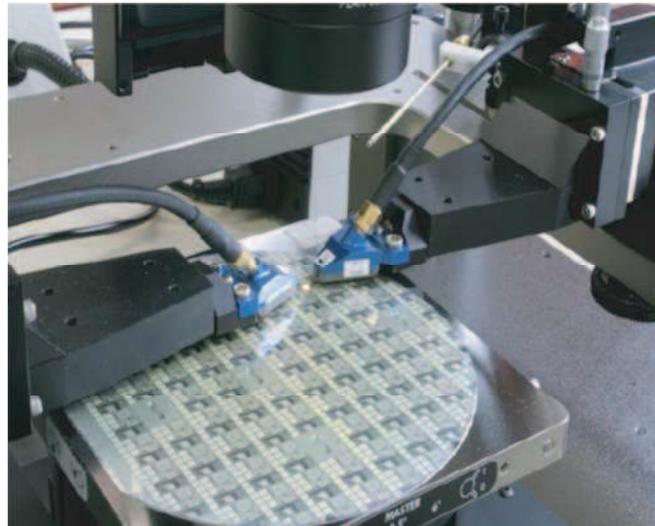


Fig. 4: Wafer-level *RF IC* test (*RF* prob)

It is clear that wafer level test in this case is very costly due to the need of expensive equipment that avoids the possibility for multi-site testing.

During many years test engineers should choose between two strategies for *RF IC* testing. The first favor a cost optimized test solution for which only basic tests are performed at wafer level and a specification oriented test is performed once the devices are packaged. Note that this strategy leads to high scrap and does not suite advanced *IC* integration technologies like *SiP* and *3D*. The second strategy favors a high quality test solution which is a costly and complex solution due to *RF* equipment needed to perform a specification based test at wafer level.

In the following some techniques for *RF IC* test cost reduction are discussed. Some of these techniques use the IC resources to relax the constraints on tester equipment. Others try at time to optimize test time and relax tester constraints by using *DC* stimulus.

I.4 Cost-reduced RF IC testing strategies

This section presents a state-of-the-art of strategies proposed in the literature to deal with the test cost reduction of the analog and *RF* circuits.

I.4.1 Integrated test solution

A classical approach to reduce the equipment cost required for the test procedure, is to embed all or a part of the test resources into the circuit itself. This approach is known as Design for Testability (*DfT*) or Built-In Self-Test (*BIST*) in case of self-testing. **Fig. 5** shows the principle of the integrated test.

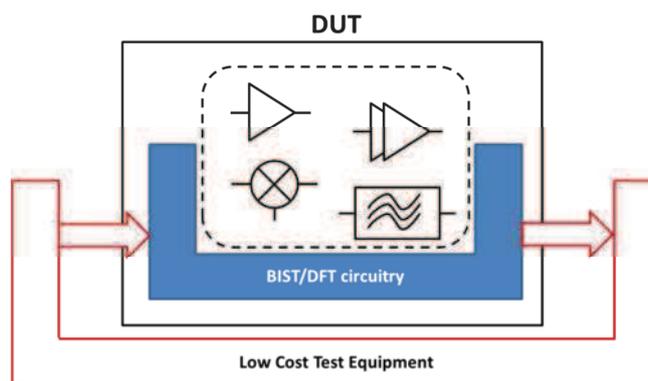


Fig. 5: Principle of integrated test solution

These techniques either directly measure the circuit performance on-chip or produce a signature that has strong correlation to the “health” of the circuit. To offer self-test capability, the *BIST* circuitry, which comprises a signal generator and a response analyzer, should be more robust than the *DUT*. These techniques imply adding additional circuitry that can dramatically increase the device area. In the context of small devices like Low Noise Amplifier (*LNA*), mixer or Power Amplifier (*PA*) these techniques are not efficient due to the amount of added area for test resources. Contrariwise, they could be very interesting test solution in the context of complex circuit like *SoC* and *SiP* systems which digital resources of the device could be used for the *RF* front ends test.

I.4.2 Loopback testing

The loopback testing is a low-cost solution for testing *RF* frontend modules and systems. Since the *RF* emitter and the receiver are integrated in the same device they are configured to test each other, the requirement for high-performance testers is alleviated [9]. The following figure shows the principle of the loopback testing.

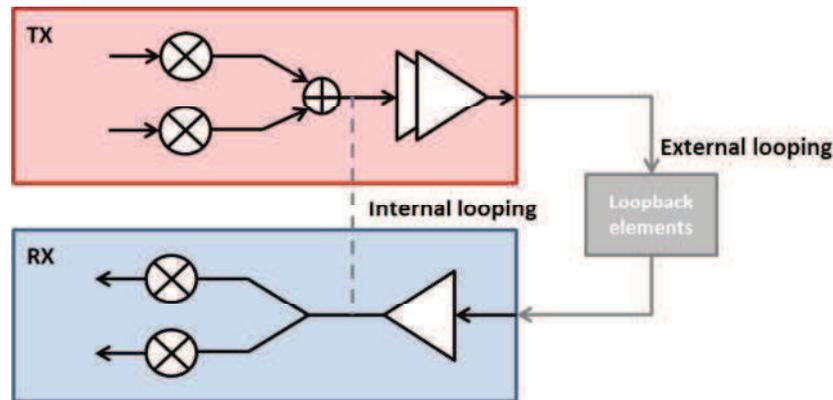


Fig. 6: Principle of the loopback testing

Actually, the loopback testing is a system-oriented strategy to the extent that we are interested in evaluating the performance of the entire system and not to evaluate the performance of each RF block. The main advantages of this approach are on the one hand, the relaxed constraints on the test equipment necessary since the application and analysis of test signals are in baseband domain. Secondly, the test time is reduced since the whole system is tested once [10]. However, this approach suffers from limited test coverage [11] and requires careful design elements inserted in the system to perform the loopback [12] [13].

I.4.3 Indirect testing

Also called alternate testing, the purpose of this strategy is to relax constraints on the number and complexity of industrial test configurations needed to perform the *RF* parameters evaluation. Instead of directly measuring the circuit performances, the approach predicts them based on a set of *DUT* signatures that are captured from cheaper and simpler test setups and measurements. As presented by the **Fig. 7**, the underlying idea of indirect testing is that process variations that affect the conventional performance parameters of the device also affect non-conventional low-cost indirect parameters in the same way. 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.

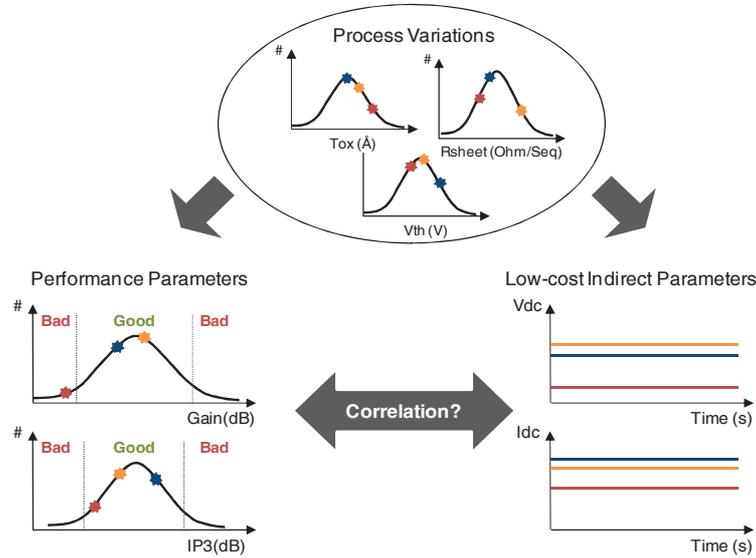


Fig. 7: Underlying idea of indirect testing.

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 indirect parameters 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.

The indirect testing is an interesting test solution for both package and wafer test levels. The non-complex indirect measurements are perfect for the wafer level test. Two main directions are explored for the implementation of the indirect testing, i.e. classification-oriented strategy [14] [15] [16] [17] [18] or prediction-oriented strategy [5] [19] [20] [21].

A. Classification-oriented strategy

As illustrated by the **Fig. 8** in the first direction, the Training Set (TS) 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.

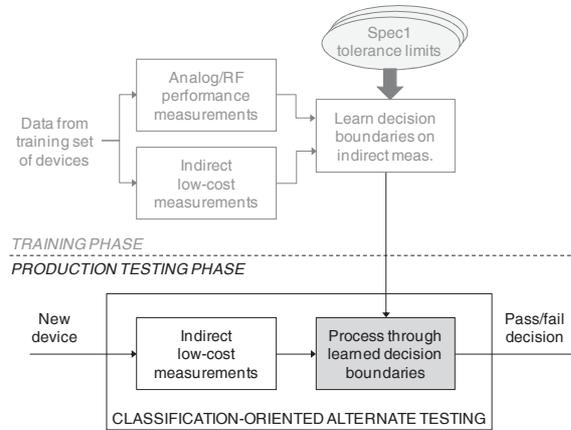


Fig. 8: Classification oriented Indirect Testing

B. Specification prediction-oriented strategy

As illustrated by the **Fig. 9** in the second variant, the training set is used to derive functions that map the low-cost indirect measurements to the performance parameters (typically using statistical regression models). 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

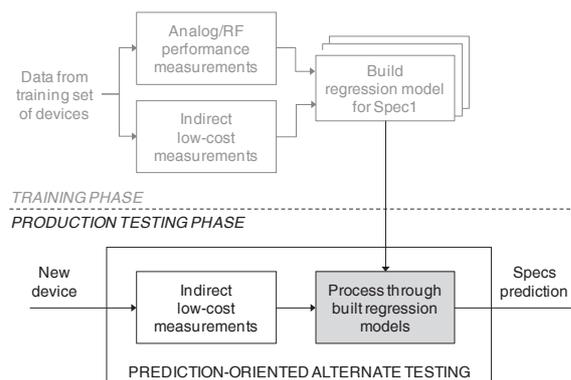


Fig. 9: Prediction oriented Indirect Testing

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. Because of these significant advantages of prediction-oriented approach on classification-oriented one, it is decided to focus improvement efforts on this approach. During next chapter this variant of the indirect testing will be studied in details.

Note that in case of DC-based indirect measurement this technique is practically suitable for wafer level testing.

C. Spatial correlation based testing

Another type of exploiting correlation in analog *RF IC* testing is the spatial correlation based approach. In this case, contrary to the indirect test approach, costly specification tests are not completely eliminated. Instead, they are only performed on a sparse subset of die on each wafer and, subsequently, used to build a spatial model, which is then used to predict performances at unobserved die locations in the wafer. The assumption made is that during the manufacturing process the neighboring dies are affected in the same way. Knowing the value of the performance of a die, we can predict the performance of its nearest neighbors. The **Fig. 10** shows the synopsis of the spatial correlation based testing.

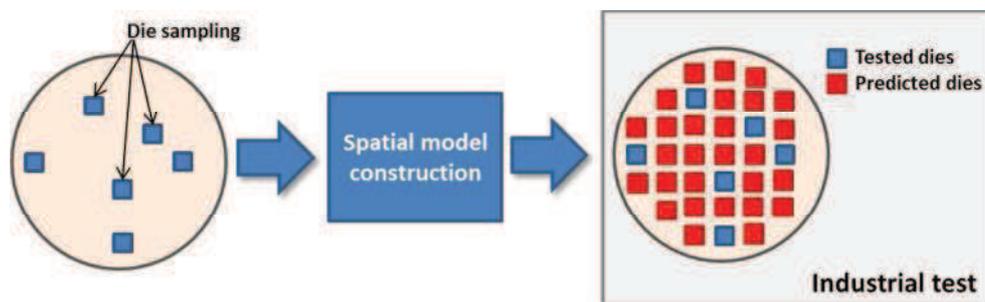


Fig. 10: Wafer level test cost reduction based on spatial correlation.

Several works investigate methods to develop variability decomposition method for spatial modeling. In [22] authors estimate the spatial wafer measurements using Expectation Maximization (*EM*) algorithm. The idea assumes that data is governed by a multivariate normal distribution. In case that the assumption is not true the authors use the Box-Cox transformation. Another way to model the spatial variation is the Virtual Probe (*VP*) which uses a Discrete Cosine Transform (*DCT*) to define the model [22]. Similarly, Gaussian Process (*GP*) models based on Generalized Least Square fitting can be used for the spatial interpolation of manufacturing data [23].

We note that good accuracy of spatial model is not usually guaranteed. In [24], authors combine the indirect testing with the spatial correlation for enhanced accuracy models.

I.5 Conclusion

The importance of developing high quality and cheaper *IC* is highlighted by the interest of semiconductor industrials to develop leading test techniques. Contrary to digital circuits the test of the *RF IC* is one of the problematic issues in semiconductors industry development. Due to the high cost of the *RF IC* test equipment. The need to develop cost efficient test strategies for the *RF IC* is expressed.

In this chapter, after introducing the general context, a brief recall of the specificity of the industrial test in particular *RF IC* test is presented. Factors impacting the cost of a given test solution are identified. It is pointed that the cost of the test solution can be easily reduced by acting on two factors: First, the test equipment capital, by relaxing its constraints using resources imbedded in the *DUT* for example. Second, test time which is reduced using low frequency and *DC* measurements. Then techniques allowing a cost reduced *RF IC* testing are presented. Emphasis is put in indirect testing which is a promising strategy to develop an extreme low cost RF test solution and a wafer-level as package-level suitable test solution. In the next chapter a case of study of the prediction-based indirect testing will be analyzed.

Chapter II

Analysis of indirect testing: case of specification prediction

II.1 Introduction

In this chapter, the author investigates the second approach of the indirect test, namely prediction-oriented indirect test. In literature, there are some contributions [4] [6] [25] [26] based on this approach; authors use different test vehicles, different regression algorithms and different efficiency metrics to implement and validate their test strategies. So, the analysis and the comparison between these works are not evident. The objective of this chapter is to analyze the efficiency of the classical implementation of prediction-oriented indirect testing and to define the framework that will be used all along the manuscript to implement and validate our proposals.

The first part of the chapter gives a brief overview of the classical implementation of prediction-oriented indirect testing and introduces the DC-based strategy we intend to use. The test vehicles that will be used for evaluation are then described and the RF parameters intended to be tested are defined. In the second part of the chapter, preliminary experiments are presented and discussed to compare some regression-fitting algorithms applied in the field of prediction-based indirect test and weaknesses of the conventional implementation scheme will be highlighted.

II.2 Classical implementation of prediction-oriented indirect test

The implementation of indirect testing involves two sequential phases, namely model building and production testing phases as illustrated in the **Fig. 11**.

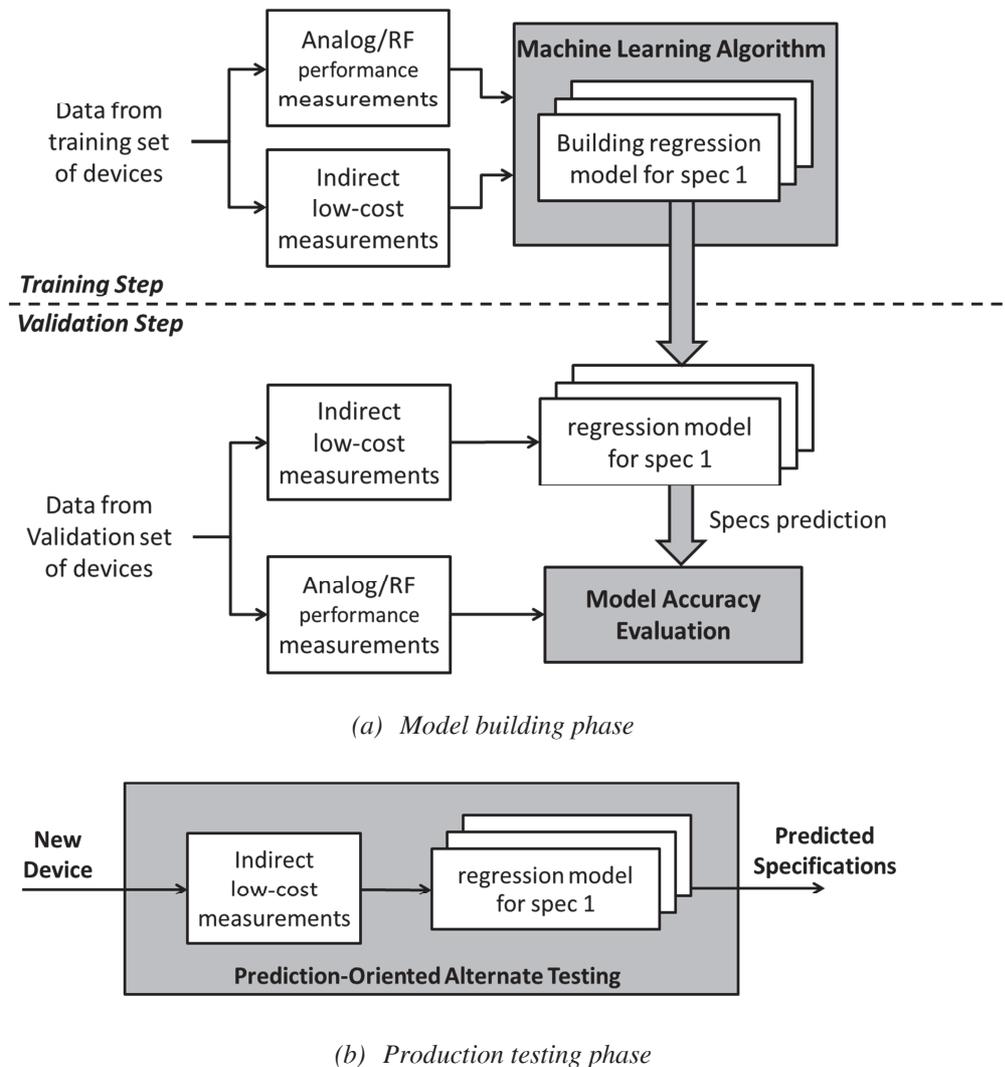


Fig. 11: Indirect Test: classical implementation

The first phase involves two steps: the training and the validation steps. During the training step, the unknown dependency between the low-cost indirect parameters and the *RF* performances is studied. For this, both specification tests and low-cost measurements are performed on a Training Set (*TS*) of devices and a machine-learning algorithm is used to build regression models that map the indirect measurements to the *RF* performances. There is then a validation step in which the accuracy of the derived models is evaluated

by comparing specification values predicted using the models to actual specification values. This evaluation is performed on a Validation Set (*VS*) of devices different from the training set, but for which both indirect measurements and specification measurements are available.

When the prediction accuracy meets the expectations, the production testing phase can start. In this phase, only the low-cost indirect measurements are performed and *RF* specifications are predicted using models developed in the previous phase.

II.3 DC-based indirect test strategy

A cornerstone of the efficient implementation of the indirect test approach is to find low-cost indirect measurements that are well correlated with the *RF* parameters. Chatterjee et al. first introduced this approach to reduce test time for analog and mixed-signal devices. They use stimuli like multi-tone or Piece Wise Linear (*PWL*) signals and they capture the transient output response in order to extract relevant signatures used to feed the machine-learning algorithm. The key idea to reduce test time is to predict all the circuit specifications from a single acquisition with a carefully optimized test stimulus, instead of using different dedicated test setups as usually required by conventional specification measurements. Regarding the use of the indirect test approach for *RF* circuits, the objective is also to relax the constraints on the required *ATE* resources, besides test time optimization. In this context, an attractive approach is to implement the indirect test strategy using only *DC* measurements. In this case, expensive *RF* options can be omitted from the *ATE* and only cheap *DC* resources are exploited. In addition because *DC* resources are usually widely available on a standard *ATE*, multi-site testing can be implemented to further reduce test time.

In this work, all experiments will be performed considering this *DC*-based indirect test strategy. Different types of *DC* measurements will be exploited including *DC* voltages on internal nodes (the circuit has to be equipped with simple *DfT* allowing to probe some internal nodes), standard *DC* measurements classically performed during production test (e.g. power supply current measurement, reference biasing voltage measurement...) or *DC* signatures extracted from embedded process sensors (e.g. *MIM*¹ capacitor, dummy transistors...). These measurements combined to different power, bias and temperature conditions could provide a huge number of indirect measurement

¹ *MIM* : Metal Insulator Metal capacitor

candidates. The choice of a pertinent set of *IMs* for predicting the *RF* specifications is an area of research that will be discussed in the next chapter.

II.4 Presentation of the test vehicles

In this section, we first present the two circuits from *NXP Semiconductors* that will be used as test vehicles: a Low-Noise Amplifier (*LNA*) and a Power Amplifier (*PA*). We then define the *RF* parameters to be predicted with the indirect test approach, together with the conventional method to measure these parameters.

II.4.1 The Low Noise Amplifier

The first test vehicle is a wideband variable-gain Low-Noise Amplifier (*LNA*) integrated in a hybrid tuner for analog and digital *TV*. **Fig. 12** shows the block diagram of the tuner where the *LNA* block is highlighted in red. This test vehicle has four different operating modes corresponding to four different gain settings: *6dB*, *9dB*, *12dB* and *15dB*.

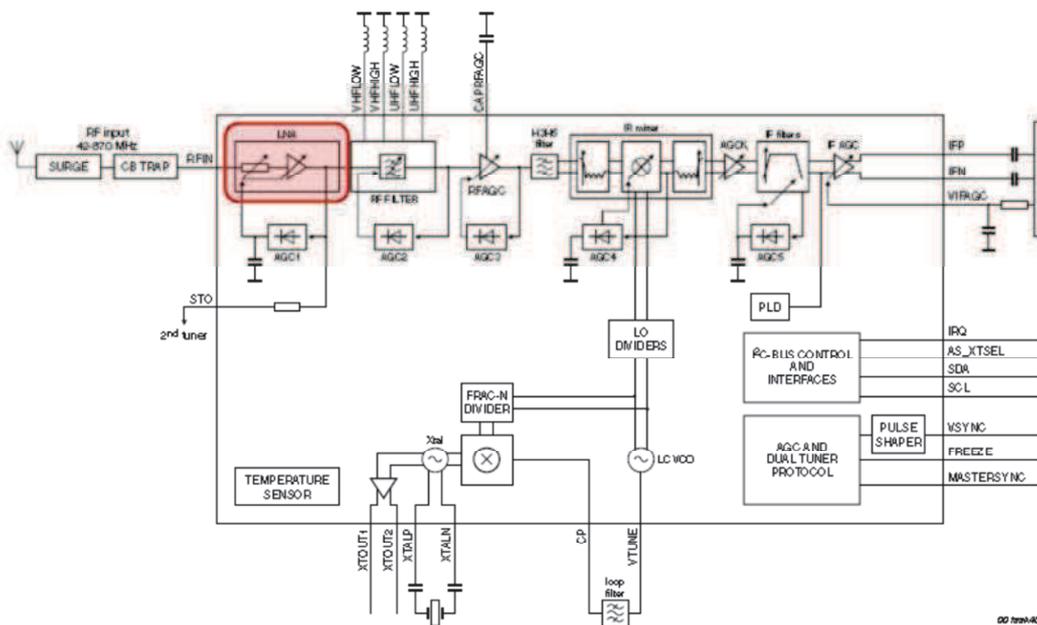


Fig. 12: Block diagram of the tuner

The objective is to estimate three different *RF* performances, namely Gain, Noise Figure (*NF*), and the 3rd order Intercept Point (*IP*₃) under the different operating modes, there are therefore twelve *RF* performances to predict.

This device is equipped with an analog bus that allows probing of six different internal nodes. *DC* voltages on these internal nodes are obvious *IM* candidates together

with *DC* voltage measurement on the conventional *RF* output. These measurements can be performed for the different functional modes of the device and for different values of power supply. Here five different values of the power supply are considered ranging from $3.0V$ to $3.6V$ (typical power supply voltage is $3.3V$). So the final set of *IM* candidates is composed of 120 elements. For this device, both *RF* performance measurements and indirect measurements are obtained from simulation of a population of 500 devices generated through Monte-Carlo simulations.

II.4.2 Power Amplifier

The second test vehicle is a Power Amplifier (*PA*) with high linearity (see **Fig. 13**). This *PA* is intended to be used in telecommunication applications. For this test vehicle, we have a large set of experimental (tester) data measured on $10,000$ devices, which includes 37 low-cost Indirect Measurements (*IM*) based on standard *DC* test and 2 *RF* performance measurements, namely the 1 -*dB* compression point (*CPI*) and the third order intercept point (*IP3*).

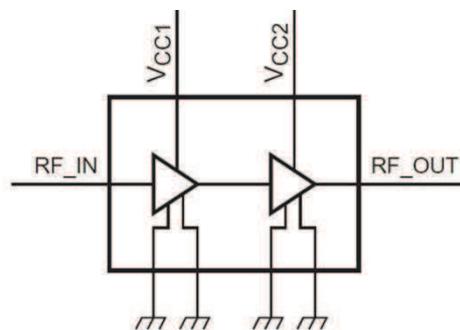


Fig. 13: Block diagram of the *PA*

II.4.3 *RF* Parameters

II.4.3.1. Gain

In *RF* devices, the power gain is more significant parameter than the voltage gain. The most common definition of power gain is the so-called transducer gain G defined by:

$$G_{dB} = 10 \log \left(\frac{P_{load}}{P_{available}} \right) \quad \text{Eq. (3.1)}$$

where P_{load} is the power at the load and $P_{available}$ is the power available from the source.

This definition assumes that matching at the input and the output ports of the *DUT* is optimized and that reflections at the input and the load could be neglected. However the gain exhibits a frequency-dependent behavior that should be characterized over all the functional frequency range of the device. Usually a network or spectrum analyzer equipment is used for this gain frequency response measurement. In volume production test, this technique is not preferred due to test cost and time considerations. Thus, in this context the gain is only evaluated at a given frequency in the functional range and only *RF* signal generator and power meter are required for the gain measurement.

II.4.3.2. Noise Figure (*NF*)

In telecommunication systems, especially those dealing with very weak signals, the signal-to-noise ratio (*SNR*) at the system output is a major criterion. The noise added by the system components might tend to obscure the useful signals and dramatically degrade *SNR*. The figure of merit that gives a measurable and objective qualification of this degradation is the Noise Figure (*NF*). **Fig. 14** illustrates the degradation. The basic definition of *NF* is the ratio of SNR_{in} at the input and the SNR_{out} at the output [27].

$$NF_{dB} = 10 \log \left(\frac{SNR_{in}}{SNR_{out}} \right) \quad \text{Eq. (3.2)}$$

There are two main techniques for the *NF* measurement: the “Y-factor” technique and the “Cold-source” technique [28] [29]. For the first technique (i.e. Y-factor) a noise source and two power measurements are required to calculate the *NF*. The first measurement is made with the noise source in its cold state: noise diode is off. Then the second measurement is made with the noise source in the hot state: noise diode is on. From these two measurements, and from knowing the ENR^2 values of the noise source, the *NF* can be calculated.

² ENR : Excess Noise Ratio

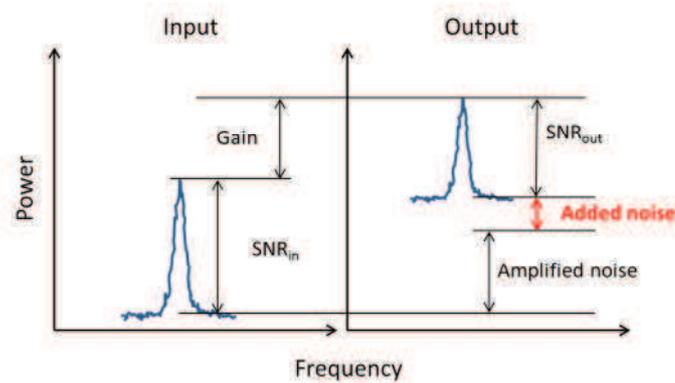


Fig. 14: SNR degradation of a signal passing through a semiconductor device

The cold-source technique consists in measuring the output power with the *DUT* placed at room temperature. The measured noise is the combination of the amplified input noise and the noise added by the device itself. If the amplification gain is accurately known, then the amplified input noise can be subtracted from the measurement, giving only the noise contribution of the *DUT*. From this the noise figure can be calculated. For this technique a vector network analyzer is required for doing measurements.

Note that for the two cited techniques, a calibration step is highly required to characterize and compensate the actual noise added by the circuitry of measurement equipment.

II.4.3.3. Gain Compression

The gain compression is a non-linear phenomenon due to the device saturation. In the linear region, when the input power increases, the output increases according to the device gain. As shown in **Fig. 15**, from a certain level of input the signal is not amplified as expected. This input level is said to be the compression point. Quite often, it is referred to the one-dB compression point (*CPI*) for amplifiers but two-dB or three-dB compression points could be defined for other devices or applications.

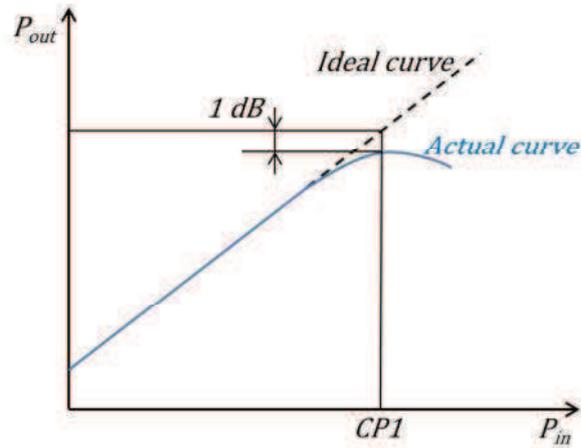


Fig. 15: Definition of the *1-dB* Compression Point

The measurement of the *CPI* is often performed in two steps. As the device gain response is not the same over frequencies, the frequency at which the *1-dB* gain compression first occurs is needed to be located. Then, a power sweep is applied to the device's input. The gain compression can be observed when the input power is increased by *2dB* while the output power increases by *1dB*. Note that at least an *RF* signal generator and a power meter are required for doing measurements.

II.4.3.4. Third Order Intercept Point

The third order intercept point (*IP3*) is an important parameter that defines the distortion caused by the nonlinearity of the device. This point is usually defined as the intercept point between the theoretical gain characteristic and the interpolation of the third order distortion characteristic as illustrated in **Fig. 16**.

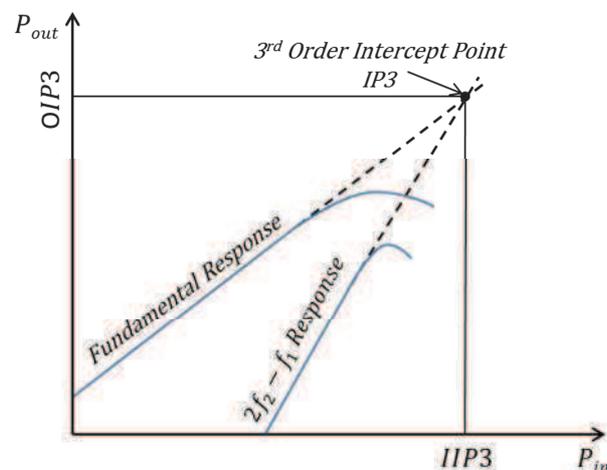


Fig. 16: Definition of the 3rd order intercept point

The measurement of the 3rd order intercept point is divided into two groups: in-band and out-of-band measurements. In-band measurements are used when the tones are not attenuated by filtering through the cascade.

For example, intercept point for a power amplifier is generally done with 2 tones that exhibit the same power throughout the system. Out-of-band measurements are used when they are attenuated like filtering in an Intermediate Frequency (*IF*). In the case of our study, only in-band measurements will be done. For the in-band measurements two tones, f_1 and f_2 are created by two signal generators and combined before entering the *DUT*.

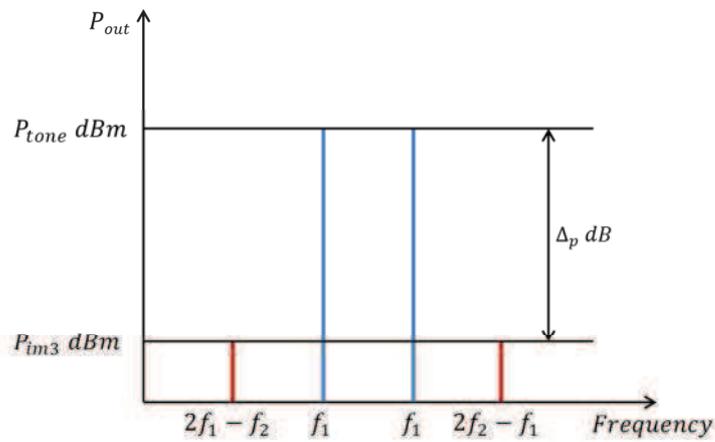


Fig. 17: In-band Intermodulation Measurement

The intercept point is determined from the measured power level of the two tones and the power levels of the intermodulation tones on a spectrum analyzer as shown in the **Fig. 17**. The Output third order Intercept Point (*OIP3*) is determined as follows:

$$OIP3_{dBm} = P_{tone} + \frac{\Delta_p}{2} \quad \text{Eq. (3.3)}$$

The Input third order Intercept Point is deduced as:

$$IIP3_{dBm} = OIP3_{dBm} + Gain_{dB} \quad \text{Eq. (3.4)}$$

II.5 Benchmarking of some machine-learning regression algorithms

There are several machine-learning algorithms used for regression mapping. In the field of indirect testing, people use different algorithms but they didn't discuss their choice in detail [5] [21]. In this section, we investigate the performance of four commonly-used machine-learning algorithms on a practical case study. The four algorithms are first briefly described. Then the case study is then presented. Finally results are analyzed and discussed.

II.5.1 Multiple Linear Regression

The most basic regression model consists in a linear relationship between the response variable to be evaluated (i.e. one analog/RF performance in our case) and one or more predictor variables (i.e. indirect measurements in particular case). The case of one predictor variable is called *simple linear regression* while for more than one predictor variable, it is called *multiple linear regression*. Given a dataset $\{Y_i, X_{i1}, X_{i2}, \dots, X_{ip}\}_{i=1}^N$ of N elements, where Y is the response variable to be predicted and $X_{1\dots p}$ are the p predictor variables, the multiple linear model takes the form:

$$Y = X\beta + \varepsilon \quad \text{Eq. (3.5)}$$

$$\text{With } Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}; X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NP} \end{pmatrix}; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

The β vector is usually estimated using the least mean square estimator as follows:

$$\hat{\beta} = (X^t X)^{-1} X^t Y \quad \text{Eq. (3.6)}$$

This equation assumes that $X^t X$ is invertible, which means that all variables are non-correlated. In the practical case of indirect test this assumption is not always verified, some measurements might be correlated.

II.5.2 Multivariate Adaptive Regression Splines

The Multivariate Adaptive Regression Splines (*MARS*) [30] is a form of regression model presented for the first time by J. Friedman in [31]. This technique can be seen as an extension of the linear regressions that modeled automatically interactions between variables and nonlinearities. The model tries to express the dependence between one response variable Y and one or more predictor variables $X_{1\dots p}$, on given realizations (data) $\{y_i, x_{i1}, \dots, x_{ip}\}_1^N$. The phenomenon that governs the data is presumed to be:

$$Y = f(X) + \varepsilon \quad \text{Eq. (3.7)}$$

The aim of algorithm is to use the data (learning technique) to build an estimated function $\hat{f}(X)$ that can serve as a reasonable approximation to $f(X)$ over the domain of the interest constituted by the predictors. The estimated function $\hat{f}(X)$ is built from bilateral truncated functions of predictors having the following form where the summation is over the non-constant M terms of the model:

$$\hat{f}(X) = \beta_0 + \sum_{m=1}^M \beta_m H_m(X) \quad \text{Eq. (3.8)}$$

This function is constituted by the term β_0 the value of Y where $X = (0, \dots, 0)$ and a weighted sum of one or many basis functions $H_m(X)$. Each basis function is simply a hinge function and takes one of the two following forms:

- A hinge function has the form of $\max(0, x - c)$ or $\max(0, c - x)$. Where c is a constant, called the knot. The hinge function is often represented:

$$(x - c)_+ = \begin{cases} x - c & x > c \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq.(3.9)}$$

- A product of two or more hinge functions which have the ability to model interactions between two or more predictors.

One might assume that only piecewise linear functions can be formed from hinge functions, but hinge functions can be multiplied together to form non-linear functions.

Note that the *MARS* model can treat classification problem as well as prediction problem. Therefore, it is used for both variants of indirect test, namely the classification-oriented test and the prediction-oriented test.

II.5.3 Artificial Neural Network

An Artificial Neural Network (ANN) is a mathematical model inspired by biological neural networks, which involves a network of simple processing elements (neurons) exhibiting complex global behavior determined by the connections between the processing elements and element parameters. Neural networks can be used for modeling complex relationships between inputs and outputs and they have been successfully implemented for prediction tasks related to statistical processes.

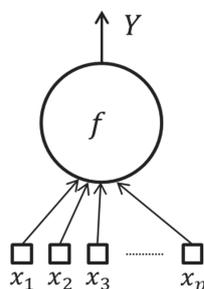


Fig. 18: Single Neuron

As presented in **Fig. 18** the basic processing element of a neural network, i.e. the neuron, computes some function f of the weighted sum of its inputs, where f is usually named the activation function. Many activation functions could be used: linear, z-shape, hyperbolic tangent, threshold, etc.

$$Y = f(a) \quad \text{with} \quad a = \sum_i w_i x_i \quad \text{Eq. (3.10)}$$

II.5.4 Regression Trees

Decision trees can be used to create a model that predicts the value of a target variable Y based on several input variables x_1, \dots, x_p . Each interior node corresponds to one of the input variables and each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner. The recursion is completed when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions [32]. **Fig. 19** illustrates the case of a

regression tree built for the prediction of the NF performance based on 4 indirect measurements (IM_i). Note that decision trees could fit with both variants of the indirect test strategy: prediction-oriented and classification-oriented strategies.

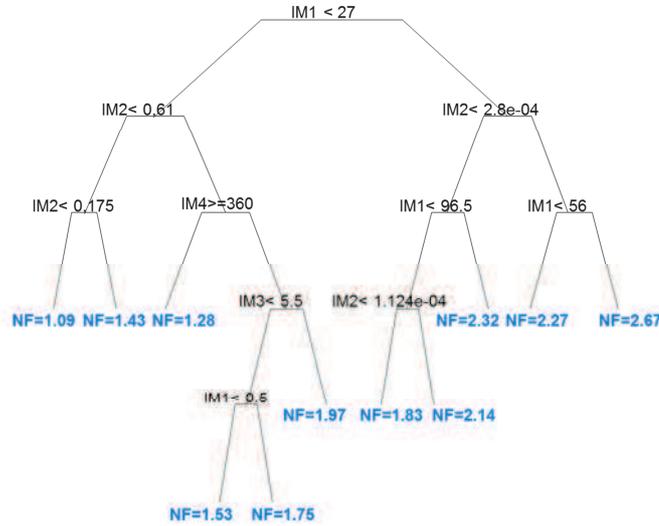
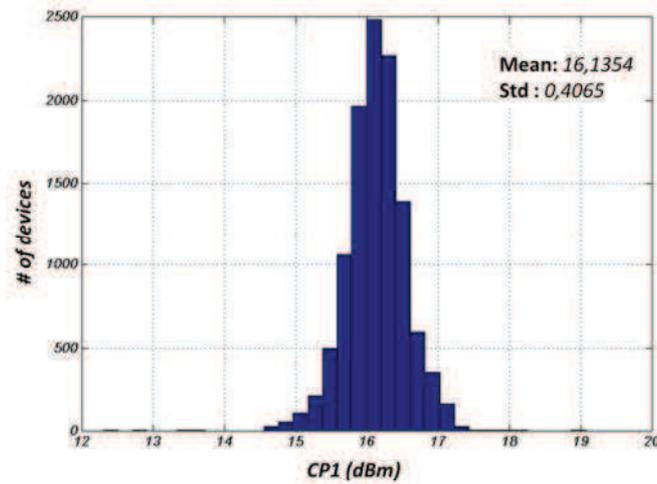


Fig. 19: The regression tree

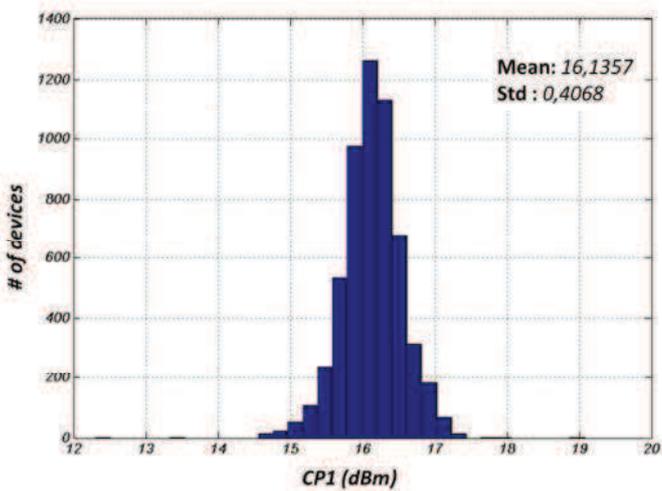
II.5.5 Test case definition

As previously mentioned, the choice of a given machine-learning algorithm is generally not discussed in the literature. In order to compare the performance of different algorithms, experiments have been performed on one case study that involves the prediction of the CPI specification for the PA test vehicle. Note that for meaningful comparison, exactly the same training and validation data will be used for the four algorithms.

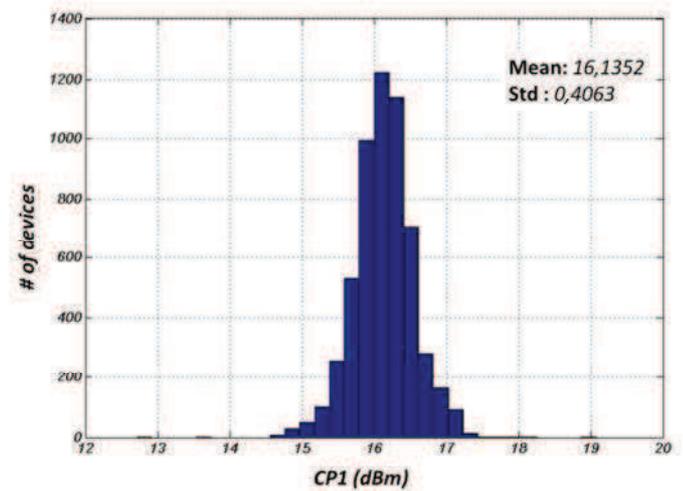
Practically, the experimental test data available from 10,000 devices are split into two subsets of 5,000 devices, one that will be used for training and the other for validation. A technique inspired from Latin Hypercube Sampling [33] [34] (LHS) is used to obtain two subsets with similar statistical properties regarding the CPI specification, as illustrated in **Fig. 20**.



(a) Whole set of devices



(b) Training Subset (TS)



(c) Validation Subset (VS)

Fig. 20: Statistical properties of Training and Validation subsets regarding CP1 specification

Data from the 5,000 training devices are fed into the different machine-learning algorithms and corresponding regression models are built (models are built with the same set of 4 pre-selected *IMs* in this experience). These four models are then used to perform *CP1* prediction for the 5,000 other devices of the validation set. Efficiency of the different algorithms can be evaluated qualitatively by comparing correlation graphs, i.e. graphs that plot predicted *CP1* values with respect to the actual *CP1* value. The closer the points to the first bisector, the better the model accuracy is. Efficiency can also be evaluated quantitatively by computing the Mean Squared Error (*MSE*) metric defined by:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{Eq.(3.11)}$$

where N is the number of devices, y_i and \hat{y}_i are the actual and predicted RF performances of the i^{th} device, respectively. The lower the MSE metric, the better the model accuracy is.

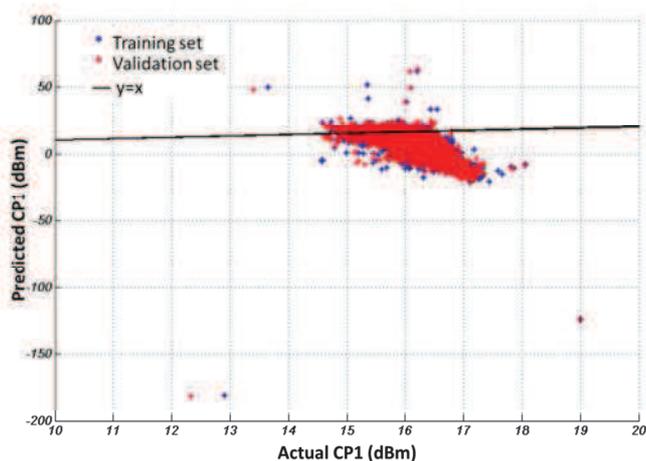
II.5.6 Results and discussion

Fig. 21 gives the correlation graphs obtained with the four different regression algorithms. For the sake of comparison, all graphs are presented with the same scale from 10 to 20dBm. It clearly appears that the regression models built using *MARS*, Regression tree (*M5P*) and *ANN* algorithms offer better performance than the model built using *MLR* algorithm, which could not fit the correlation between used *IMs* (i.e. predictors) and the *RF* specification. So the linear model will be ruled out for the rest of the study.

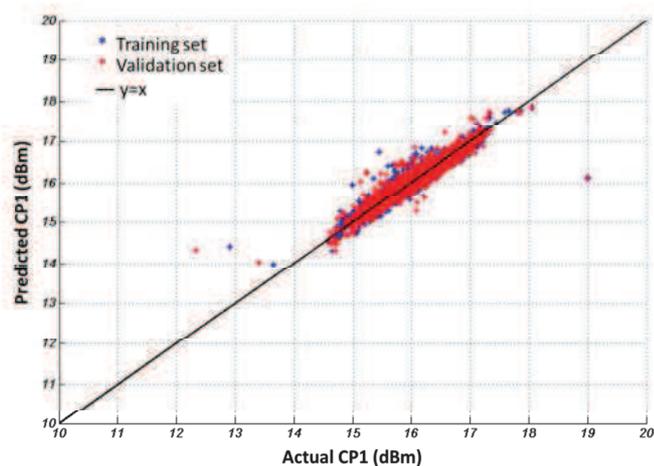
From the qualitative analysis of these graphs, there is no significant difference between *MARS*, *M5P* and *ANN* algorithms. In the three cases, most of the devices are correctly predicted with a good accuracy. This is confirmed by computing the *MSE* metric associated to each model, which exhibits similar value:

$$MSE|_{MARS} = 0.016; \quad MSE|_{M5P} = 0.02; \quad MSE|_{ANN} = 0.023$$

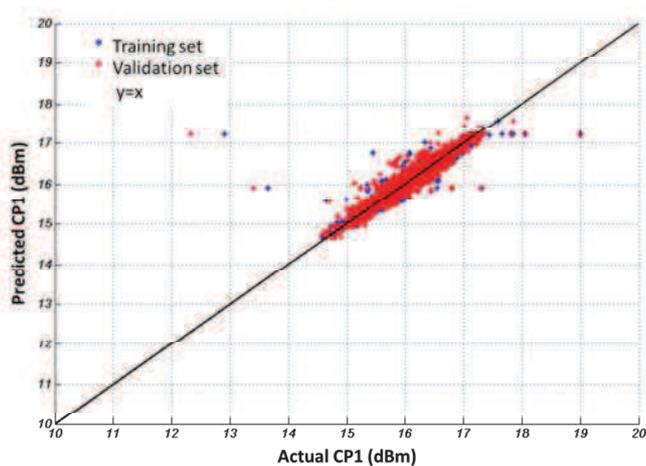
To further analyze the performance of these three regression algorithms, a more advanced evaluation is necessary. To do this, regression models are built for the three algorithms considering 1,000 different combinations of 3 *IMs* randomly chosen among the 37 low-cost *IMs* available for this case study. The *MSE* metric is then computed on both training and validation subsets, for each model and each regression algorithm.



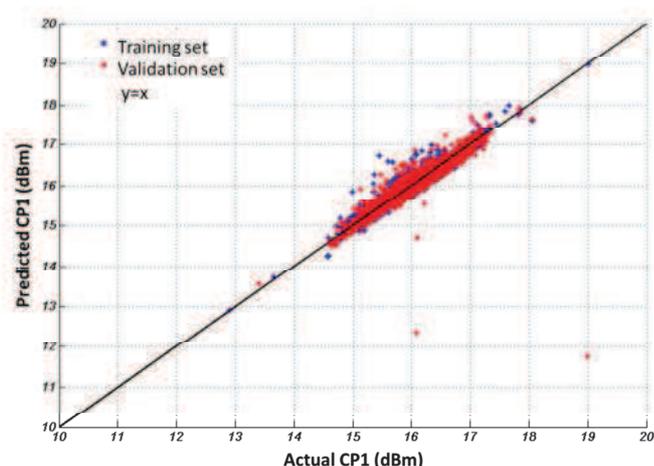
(a) Regression model built using MLR algorithm



(b) Regression model built using MARS algorithm



(c) Regression model built using MSP algorithm



(d) Regression model built using ANN algorithm

Fig. 21: Comparison of the correlation plots obtained with 4 different regression algorithms

During this experiment, it sometimes happens that the estimation of the *RF* performance is aberrant for a particular model and a particular device of the validation subset. As an illustration, **Fig. 22** shows the case of an aberrant *CPI* prediction for one device (device number 513) with a predicted value around $9.9 \cdot 10^{11} \text{ dBm}$, which of course never happens in real circuit. If this prediction is considered in the calculation of the *MSE*, it significantly affects the calculated value, which is $1.7639 \cdot 10^{20}$ for this example. If this aberrant prediction is removed for the computation, the non-biased value of the *MSE* is 0.0162 , which corresponds to a realistic value of the *MSE* for this example. Consequently for a faithful comparison of the model accuracy achieved using the different regression algorithms and the different *IM* combinations, such aberrant

predictions should be removed from the *MSE* calculation on the validation subset. Note that these aberrant predictions are easy to identify because they are totally out of the possible value range of the performance to be predicted (for the case of the *CPI* performance, a realistic range of the possible predicted values is from *0dBm* to *30dBm*). We denote these predictions Out-Of-Range (*OOR*) predictions.

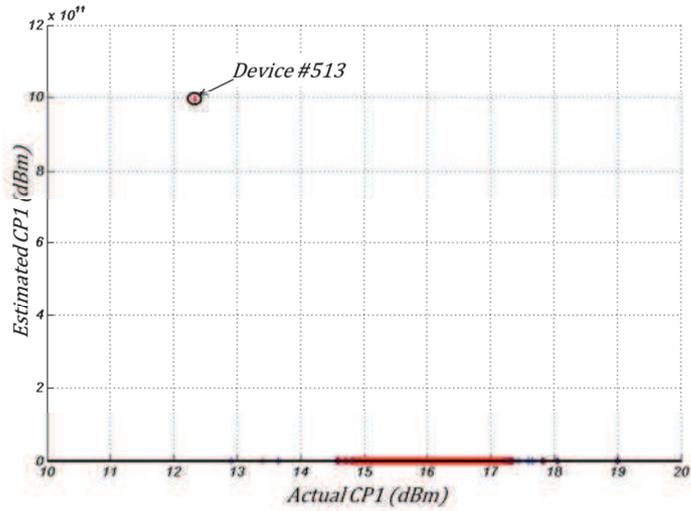
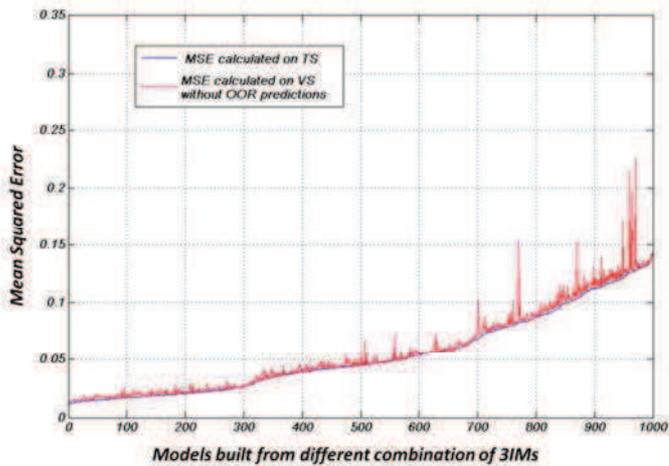
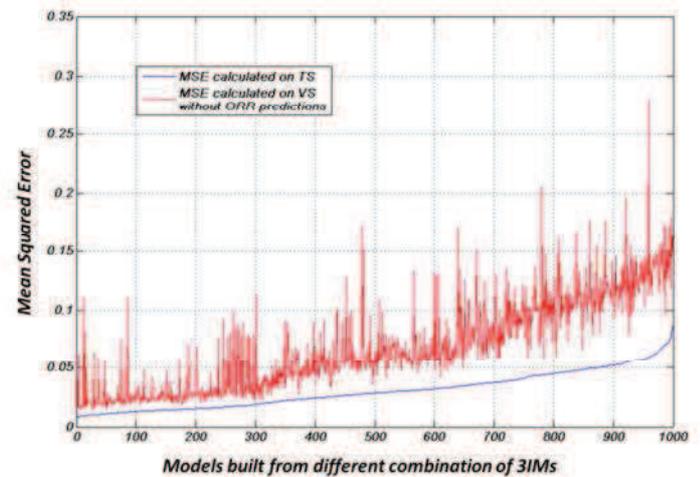


Fig. 22: Example of aberrant prediction for one particular device

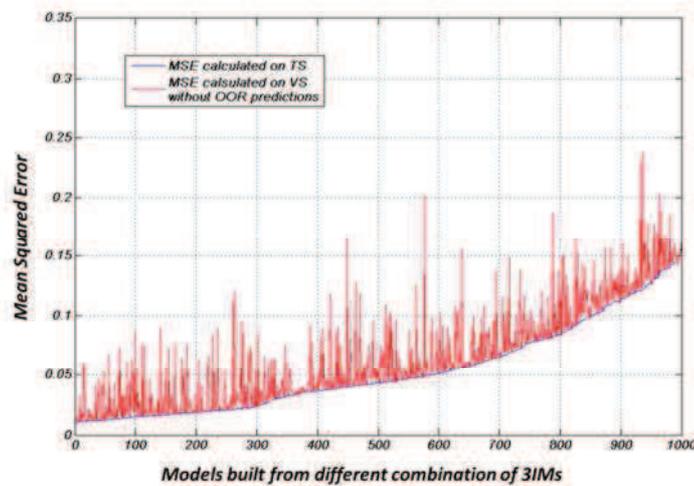
All the results presented in the following are computed after removing the *OOR* predictions from the validation subset. **Fig. 23** reports the *MSE* values corresponding to the 1,000 models built with different *IM* combinations, for the three different regression algorithms. In each graph, both the *MSE* calculated on the Training Subset (*TS*) and the *MSE* calculated on the Validation Subset (*VS*) are provided. The *MSE* calculated on *TS* translates the ability of the model to accurately represent the relation that links the selected *IMs* to the *RF* specification for the considered training devices, while the *MSE* calculated on *VS* translates the ability of the model to accurately predict the value of the *RF* specification from the selected *IMs* for new devices different from the training devices. So to ensure accurate performance prediction, a model should exhibit not only a low *MSE* value with respect to *TS*, but also an *MSE* value in the same range with respect to *VS*. Note that for the sake of clarity, models are sorted regarding their *MSE* computed on *TS*.



(a) Models built with MARS algorithm



(b) Models built with MSP algorithm



(c) Models built using ANN algorithm

Fig. 23: MSE variation over models built with different IM combinations for different regression algorithms

Analyzing these results, a slight advantage appears for the *MP5* regression algorithm regarding *MSE* values calculated on *TS*. However there is a significant discrepancy between *MSE* values calculated on *TS* and *VS*. In the same way for the *ANN* regression algorithm, a good *MSE* value on *TS* does not ensure a *MSE* value on *VS* in the same range. In contrast, the *MARS* regression algorithm appears much more robust since most of the models built with this algorithm give almost identical values for the *MSE* calculated on both *TS* and *VS*. For these reasons, the *MARS* algorithm will be chosen for regression model construction during the rest of the study.

II.6 Limitations & bottleneck of the conventional indirect test scheme

The main challenge of indirect test is the prediction confidence. The confidence is usually assimilated to two aspects: the first one is the model accuracy which is usually expressed in terms of average prediction error; the second one is the model robustness against anomalies which usually manifest themselves in the devices having a big prediction error. In the following some experiments which highlight the weaknesses of the classical implementation of the indirect test are presented.

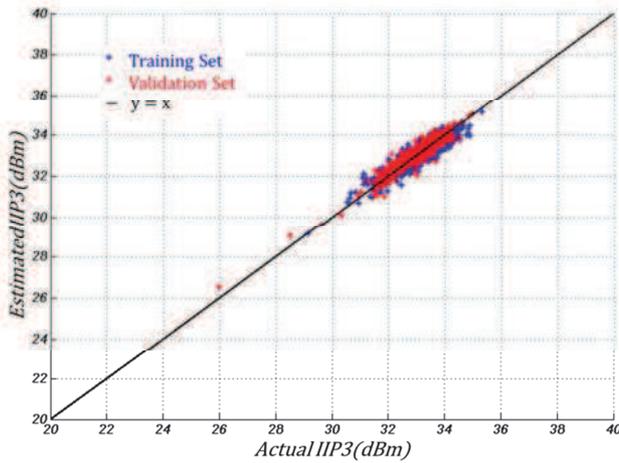
II.6.1 Prediction confidence: flawed predictions

Many of the experiments reported in the literature on various devices demonstrate that very low average prediction error can be achieved. However, two main points limit the credit we can give to this good accuracy. First, low average prediction error does not guaranty low maximal prediction error, which is of crucial importance regarding the classification step where the predicted values are compared to the specification limits promised in the datasheet.

The maximal prediction error is defined as follows:

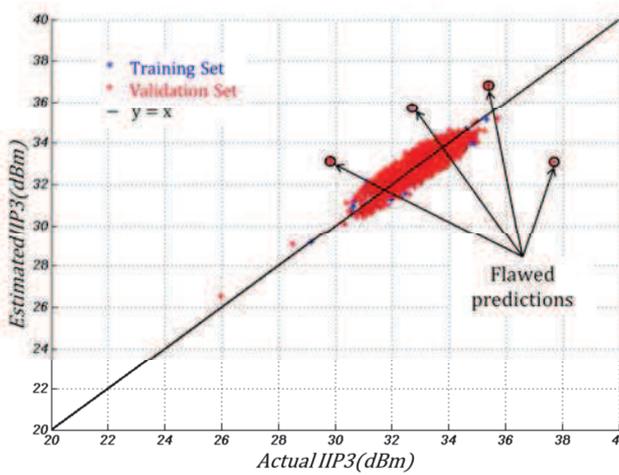
$$\varepsilon_{max} = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^N |y_i - \hat{y}_i|^p} = \sup_{1 \leq i \leq N} |y_i - \hat{y}_i| \quad \text{Eq. (3.12)}$$

Second, 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 on a large set of fabricated devices, typically one or several millions. 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 fabricated devices.



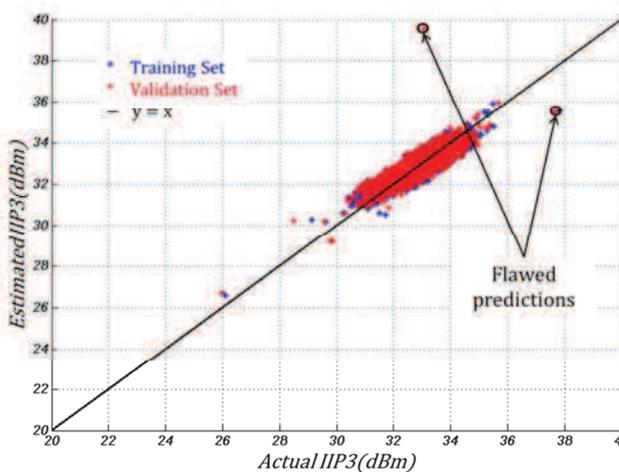
	<i>MSE</i>	ϵ_{max}
<i>TS</i>	0.098	1.012 dB
<i>VS</i>	0.094	0.998 dB

(a) *TS*: 1,000 devices – *VS*: 1,000 devices



	<i>MSE</i>	ϵ_{max}
<i>TS</i>	0.098	1.012 dB
<i>VS</i>	0.103	4.591 dB

(b) *TS*: 1,000 devices – *VS*: 5,000 devices



	<i>MSE</i>	ϵ_{max}
<i>TS</i>	0.101	1.251 dB
<i>VS</i>	0.111	6.518 dB

(c) *TS*: 5,000 devices – *VS*: 5,000 devices

Fig. 24: Illustration of prediction error for different sizes of *TS* and *VS*

To illustrate these points, some experiments have been performed on one case study that involves the prediction of the *IIP3* specification for the *PA* test vehicle. Here again,

the experimental test data available from $10,000$ devices are split into two subsets of $5,000$ devices with similar statistical properties, one used for training and the other for validation.

In the first experiment, $1,000$ devices are chosen randomly from the TS to build prediction models based on triplets of IMs and the prediction accuracy is evaluated with $1,000$ devices chosen randomly in the VS . As an illustration, **Fig. 24.a** presents an example of $IIP3$ prediction for one “good” model. In this case, MSE and ε_{max} values calculated on both TS and VS give coherent results, with both low average and maximal prediction errors.

Then, in the second experiment, the number of the devices used for validation is increased to $5,000$ devices, i.e. the obtained model is validated with a large VS . For most of the devices, the average prediction error is nearly preserved but we observe flawed predictions for some devices resulting in a large maximum prediction as illustrated in **Fig. 24.b**.

Such flawed predictions are usually attributed to the fact that the mapping obtained using finite-size TS is not fully representative of the actual complex mapping. So, in the third experiment, we use all the $5,000$ devices of the TS to build the regression model and the prediction accuracy is evaluated using all the $5,000$ devices of the VS . Unfortunately even with such a large training set, we still observe flawed predictions for some devices as illustrated in **Fig. 24.c**.

Note that one could reasonably think that these flawed predictions are due to some distinctive features of particular devices, for instance devices which are not consistent with the statistical distribution of manufactured devices. In this case, it would be possible to filter these devices prior to applying the alternate test procedure, as suggested in [15]. However this does not seem the case because different behaviors are observed depending on the model used to predict the performance, as illustrated in **Fig. 25**. Indeed even if both models are built with the same training set, flawed predictions are observed for devices 4,512 and 1,179 using model 1 while they are correctly predicted using model 2, and flawed prediction is observed for device 2,326 while it is correctly predicted using model 1.

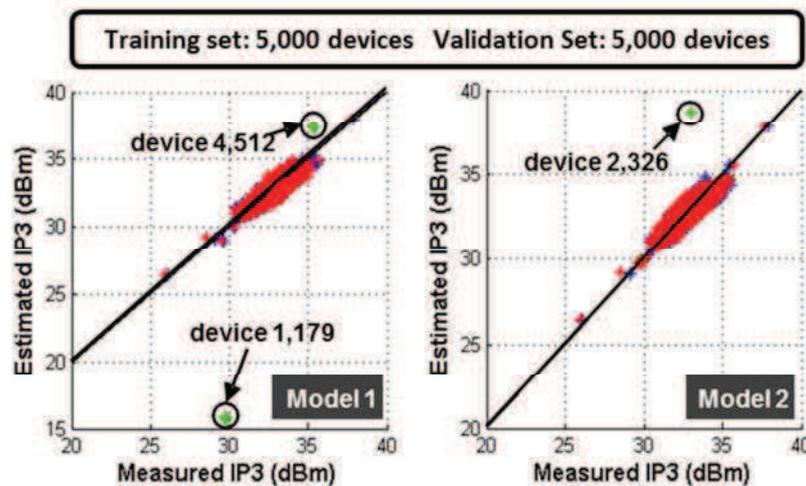


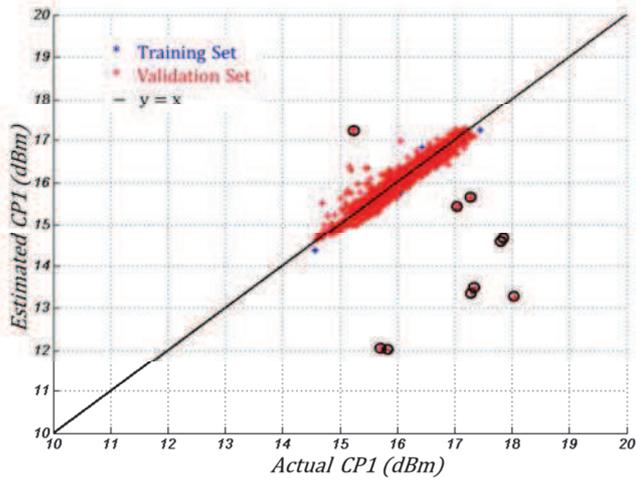
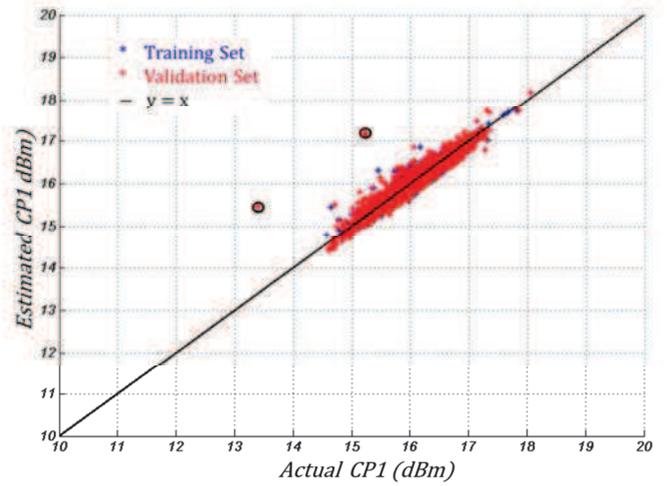
Fig. 25: Illustration of prediction error for different regression models
(with same training and validation subsets)

In summary, these experiments have pointed out an important weakness of the prediction-based alternate test method: although it provides accurate prediction results for most of the devices, flawed predictions are observed for a very small number of devices. 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.

II.6.2 Dependency of model performances with respect to TSS

The problem addressed in this subsection deals with alternate test robustness with respect to the Training Set Size (*TSS*). Note that there is no detailed study in the literature on this aspect. The main reason is that in most of the cases, data are available only for a limited number of devices, due to constraints on *RF* and indirect measurements. Indeed performing *RF* measurements is very costly since it requires high accuracy equipment and takes long time. So, the number of the measured devices hardly exceeds few thousands and people usually deal with few hundreds of devices. Still, the common assumption is that the larger the training set size, the better the prediction accuracy.

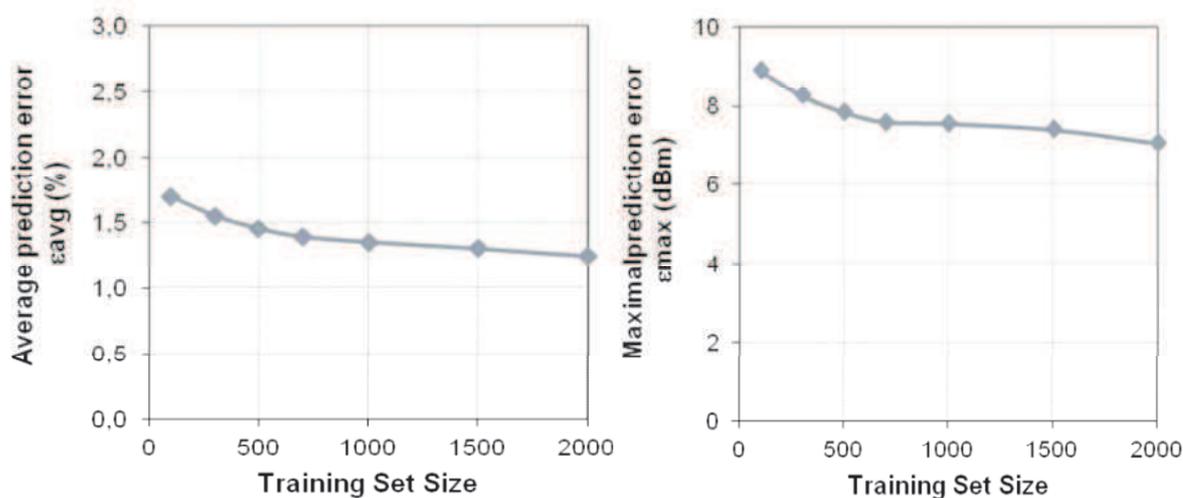
To illustrate the influence of the *TSS* on the regression model efficiency, experiments were performed on the *PA* case study, for which we have experimental test data on a large number of instances. The *CPI* is the specification considered in this study.

(a) $TSS = 300$ devices(b) $TSS = 2000$ devices**Fig. 26:** *CPI* prediction for two different *TSS*

In a first experiment, 300 devices randomly chosen in the *TS* are used to build prediction models based on triplets of *IMs* and we evaluate prediction accuracy using all 5,000 devices of the *VS*. Prediction accuracy is evaluated in terms of both average and maximal prediction errors. **Fig. 26.a** presents an example of *CPI* prediction for one “good” model, i.e. a model with low average and low maximal errors on the *TS*. On the validation set, although most of the devices are correctly predicted, some devices suffer from large prediction error. There are actually 2 circuits for which we observe an aberrant prediction with an error that exceeds several hundreds of dB (not represented on the graph for scale reasons). Such aberrant predictions correspond to *OOR* predictions, as discussed in the section II.5.6, and are easy to identify. More problematic is the case of circuits for which the predicted value is in the possible performance value range but with a significant prediction error. For the considered example, 10 circuits (represented with black circles) present a prediction error that exceeds 2.5dB, but is still in the possible operating range of manufactured circuits, i.e. [0dBm, 30dBm]. In these cases, there is no evident way to identify that the predicted value corresponds to a non-reliable prediction. Moreover, such predictions are a real concern in the context of production testing because they might generate additional yield loss or test escape. Note that the value of 2.5dBm has been arbitrarily chosen for illustration purpose but is of no particular significance.

In a second experiment, we perform the same study but using a much larger training set of 2,000 devices. Prediction accuracy is still evaluated using all 5,000 devices of the validation set. **Fig. 26.b** illustrates *CPI* prediction results for the regression model built

using the same triplet of *IMs*. Despite the larger *TS*, there are still some circuits (4 circuits) for which prediction error exceeds 2.5dB but is in the possible operating range of manufactured circuits. So although if increasing the size of the training set mitigates the number of problematic predictions, it does not solve the problem.



(a) average prediction error

(b) maximal prediction error

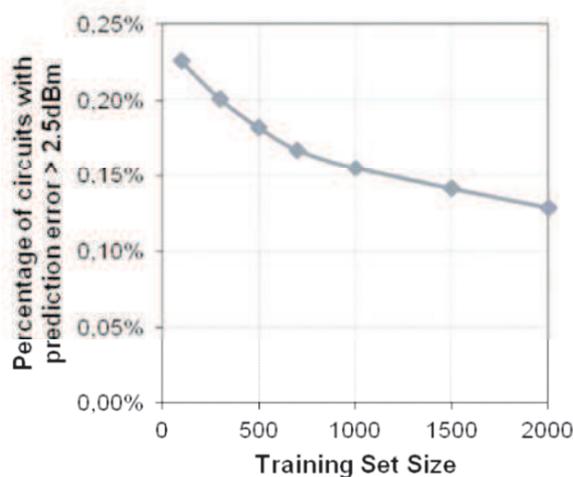
(c) percentage of circuits with prediction error larger than 2.5dBm

Fig. 27: Influence of Training Set Size (TSS) on prediction accuracy for *CPI* specification

To further corroborate these observations, we have carried out an extensive campaign of experiments considering different *TSS* from 100 to $2,000$ devices randomly chosen in the initial training set of $5,000$ devices. Each experiment is repeated 100 times and reported results correspond to the mean of results obtained over the 100 runs. Results are summarized in **Fig. 27** that reports average and maximal prediction errors together

with the number of instances for which prediction error exceeds $2.5dB$ with respect to TSS . It can be observed that the average prediction error is slightly reduced when using training sets of larger size. Still note that good accuracy is obtained whatever the size of the training set with an error that remains below 2% . In the same way, the maximal prediction error is slightly reduced when using training sets of larger size. However it remains high with an error that remains above $6dBm$, even when using large training sets. These results clearly translate the fact that accurate prediction results are obtained for most of the devices, but some circuits suffer from rather large prediction errors. The number of these circuits reduces when the size of the training set augments, as illustrated in **Fig. 27.c** which shows that the percentage of circuits with a prediction error larger than $2.5dB$ is divided by a factor 1.74 when TSS increases from 100 to 2,000 devices. It is worth noting that even with a small training set, this percentage remains very low, i.e. less than 0.25% of the circuits in case of $TSS=100$.

II.7 Conclusion

The indirect testing is a promising technique to overcome the prohibitive cost of conventional test methods. It is specially adapted to analog and *RF* devices testing which suffer from heavy testing strategies. The prediction-oriented indirect test is very interesting; it has the advantage to provide the value 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. The use of *DC* indirect measurements dramatically lessens the cost of the test strategy by removing the expensive *RF* equipment from the test process and allowing multi-site testing using only *DC* resources of the *ATE*.

Today there are many contributions in this field; however there is no clear benchmark to compare works: people use different specific *DUTs*, different learning conditions and different figures of merit for model accuracy evaluation. In this context, we have performed in this chapter a comparison of some commonly used machine-learning algorithms to build predictive models. Results show that models built with a simple multiple linear regression are not efficient to correctly represent the relationship between indirect measurements and an *RF* performance to be predicted, while models built with more refined algorithms such as Multivariate Adaptive Regression Splines, Artificial Neural Networks or Regression Trees indeed permit to accurately represent this relationship. Results also demonstrate that the *MARS* algorithm is more robust than the others regarding its ability to perform performance prediction. For this reason, only this algorithm will be used for the construction of predictive models in the rest of the manuscript.

Finally, we have performed some experiments in order to analyze the efficiency of the classical implementation of prediction-oriented indirect testing and highlight the weakness of such implementation. Two main aspects restrain the large deployment of such an approach by the industrials, namely confidence in the predicted values and robustness of the procedure against learning conditions and potential process shift. In chapter IV, we will propose novel implementation of the prediction-based indirect test which manages prediction confidence in order to improve the robustness of the procedure.

All results presented in this chapter are based on two test vehicles from *NXP Semiconductors*, which we have been described in the beginning of the chapter. These test vehicles will be also used in the rest of the manuscript.

Chapter III

Strategies for IMs Selection

III.1 Introduction

Typically, the used *IMs* are based on complex signals (multi-tone, *PWL*...) and people are focused on the optimization of the stimulus waveforms [4] [19] [35]. The need of developing efficient method to select the best set of *IMs* is first expressed when people begin to investigate the implementation of the alternate test method based on *DC* measurements. The main motivation for this *DC*-based strategy is that it offers the perspective of a really low-cost solution. Indeed as far as indirect measurements are performed in static conditions, no stimulus is applied to the functional input and obviously the cheapest stimulus is no stimulus. Discussions on stimulus optimization such as presented in [19] [35] are therefore not relevant in this context. In the same way, *DC* measurements are extremely simple signatures and there is no need of additional process to extract pertinent information such as presented in [20] [36].

As seen in chapter II, there is a large number of *DC*-based *IMs* that may be exploited to perform performance prediction for a given *DUT*. However as mentioned in the previous chapter, an analytic correlation between indirect measurements and *RF* specifications is extremely complex and it is even extremely difficult to forecast which indirect measurements are the most likely to have a strong correlation with the targeted specifications. Moreover because of the curse of dimensionality, it would not be efficient to build a predictive model based on all the possible *IMs*. Indeed it is well-known that for a fixed number of training samples, the predictive power of the model reduces as the dimensionality of the feature space increases. In this context, the selection of pertinent *IM* subsets for building predictive models is a keystone in the quality of the indirect test. The goal of this chapter is to investigate some solutions for the clever selection of a subset of *IMs* in the context of prediction-oriented indirect test.

III.2 Benchmark of some feature selection techniques

In the field of Information Technologies (*IT*), several methods for pertinent data selection are developed. There is an interdisciplinary subfield of computer science, namely data mining, which studies solution for efficient information extraction from large data. In this section, we investigate some commonly used methods for feature subset selection. Along the following subsections we will adopt terminology used by the specialists of this field. We denote by “feature selection” the algorithm that is used for subset selection. In our case, the term “feature” refers to the *IM* and the term “class” refers to the *RF* performance to be predicted.

III.2.1 Variable subset selection techniques

Variable subset selection methods are essentially divided into two categories, i.e. filters and wrappers. Filters select subsets of features as a pre-processing step independently of the chosen predictor, while wrappers utilize a learning machine as a black box to score subsets of features according to their predictive power. We choose to investigate three different filters and three different wrappers, which are briefly described in the following.

III.2.1.1. Filters

A. Correlation-based Feature Selector

The Correlation-based Feature Selection (*CFS*) is an algorithm that ranks feature subsets according to a correlation-based heuristic evaluation function. The evaluation function is toward favor the subsets that contain features (i.e. *IMs*) that are highly correlated with the class (i.e. *RF* specification to be predicted) and uncorrelated with each other. Irrelevant features should be ignored because they will have low correlation with the class. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The acceptance of a feature will depend on the extent to which it predicts class in areas of the instance space not already predicted by other features. The *CFS*'s feature subset evaluation function is defined here:

$$\text{Merit}_{x_k} = \frac{k\bar{r}_{x_c}}{\sqrt{k+(k-1)\bar{r}_{xy}}} ; x, y \in X_k \quad \text{Eq. (4.1)}$$

where $Merit_{X_k}$ is the heuristic ‘‘Figure of merit’’ of a feature subset X_k containing k features, $\overline{r_{xC}}$ corresponds to the mean feature-class correlation ($x \in X_k$), and $\overline{r_{xy}}$ corresponds to the average feature-feature cross-correlation. Details on the definition of $\overline{r_{xC}}$ and $\overline{r_{xy}}$ can be found in [37].

B. Relief algorithm

One of the most famous feature selection algorithms is the RELIEF algorithm based on the nearest-neighbor algorithm, which is described by Kira and Rendell in [38]. It uses instance-based learning to assign a relevance weight to each feature (i.e. *IMs* in our case). Each feature weight reflects its ability to distinguish among the class values (i.e. RF specification in our case). Features are ranked by weight and those that exceed a user-specified threshold are selected to form the final subset.

According to [39] [40], the RELIEF algorithm attempts to approximate the following difference of probabilities to compute the weight of a feature X :

$$w_X = \frac{\mathbf{P}(\text{different value of } X/\text{nearest instance of different class})}{\mathbf{P}(\text{different value of } X/\text{nearest instance of same class})} - 1$$

Eq. (4.2)

The previous equation can be reformulated as:

$$\mathbf{Relief}_X = \frac{\mathbf{Gini}' \sum_{x \in X} \mathbf{p}(x)^2}{(1 - \sum_{c \in C} \mathbf{p}(c)^2) \sum_{c \in C} \mathbf{p}(c)^2} \quad \text{Eq. (4.3)}$$

where C is the class variable (i.e. RF specification) and

$$\mathbf{Gini}' = \left[\sum_{c \in C} \mathbf{p}(c)(1 - \mathbf{p}(c)) \right] - \sum_{x \in X} \left(\frac{\mathbf{p}(x)^2}{\sum_{x \in X} \mathbf{p}(x)^2} \sum_{c \in C} \mathbf{p}(c|x)(1 - \mathbf{p}(c|x)) \right) \quad \text{Eq. (4.4)}$$

\mathbf{Gini}' is a modification of another attribute quality measure called the *Gini-index*³.

³ *Gini-Index* : a measure of statistical dispersion

C. Minimum-Redundancy-Maximum-Relevance

“Minimum-Redundancy-Maximum-Relevance” (*mRMR*) is a feature selection method first introduced by Pend and al. in 2005 [41] that can use either mutual information, correlation, distance/similarity scores to select features. The main idea is to select features with maximum relevance (*MR*) while minimizing the redundancy (*mR*) between selected features. For example, with mutual information, the relevance of a feature set F to predict the class C and the redundancy of all features in the set F are given by the following equations:

$$\text{Redundancy}(\mathbf{F}) = \frac{1}{P^2} \sum_{X, Y \in F} I(X, Y) \quad ; \quad \text{Relevance}(\mathbf{F}) = \frac{1}{P} \sum_{X \in F} I(X, C) \quad \text{Eq. (4.5)}$$

where F is the set of features and P its size. $I(X, Y)$ is the mutual-information between two different features from F . $I(X, C)$ is the mutual-information between a feature and the class. The following equation gives the mutual-information:

$$I(\mathbf{X}, \mathbf{Y}) = \sum_{x \in X} \sum_{y \in Y} \frac{p(x, y) \log(p(x, y))}{p(x)p(y)} \quad \text{Eq. (4.6)}$$

The *mRMR* score of a feature set F is a combination of both relevance and redundancy obtained either by a quotient or a difference:

$$\text{score}(\mathbf{F}) = \frac{\text{Relevance}(\mathbf{F})}{\text{Redundancy}(\mathbf{F})} \quad \text{or} \quad \text{score}(\mathbf{F}) = \text{Relevance}(\mathbf{F}) - \text{Redundancy}(\mathbf{F}) \quad \text{Eq. (4.7)}$$

The objective is then to select the set that have the maximum score.

III.2.1.2. Wrappers

The wrapper methodology offers a simple and powerful way to address the problem of feature selection. In its most general formulation, the wrapper methodology consists in using

the prediction performance of a given learning machine to assess the relative usefulness of subsets of features. In practice, one needs to define: (i) how to search the space of all possible feature subsets; (ii) how to assess the prediction performance of a learning machine to guide the search and halt it; and (iii) which predictor to use. An exhaustive search can conceivably be performed, if the number of variables is not too large. But the problem is known to be NP-hard and the search becomes quickly computationally intractable. In that case, a wide range of search strategies can be used, including best-first, branch-and-bound, simulated annealing, genetic algorithms or Greedy search.

In our case, we choose to use a simple iterative strategy to select for each class P_j , a subset S_j of features allowing to predict the class with a given accuracy constraint ϵ_{target} . This iterative strategy is described in the simplified diagram of **Fig. 28**.

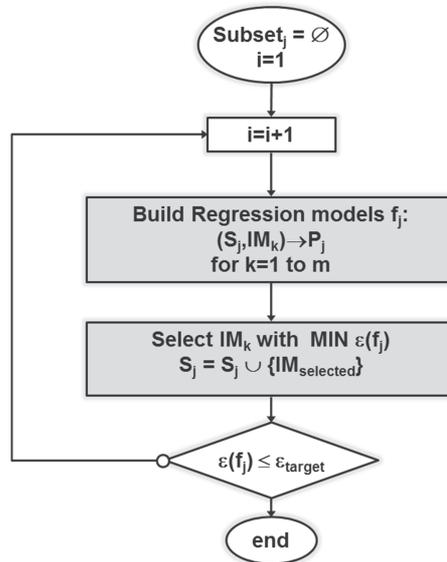


Fig. 28: Iterative search of IM subsets for specification prediction

The first iteration of the search algorithm consists in building a predictive model using each feature and selecting the feature giving the model with minimum average prediction error ϵ . In the second iteration, a predictive model is built using each pair of features including previously selected one, and the pair of features that gives the model with minimum average prediction error is selected. Then, we work with triplets of features, always keeping the features selected in the previous iterations and so on. The procedure stops when the obtained average prediction error is below a pre-defined target value $\epsilon \leq \epsilon_{target}$ (or when no further improvement is obtained by adding new feature to the model).

Note that the learning algorithm used to perform feature subset selection may be different from the learning algorithm used to predict a class. In our case, we will consider

both linear regression and multivariate adaptive regression splines (*MARS*) models for the evaluation of prediction performance during the iterative selection procedure.

Finally, the last method that will be investigated is the built-in variable selection procedure of the Multivariate Adaptive Regression Splines (*MARS*) algorithm itself. Indeed, a predictive model constructed by *MARS* consists of a linear combination of basis functions together with a constant ‘intercept’ term, where each basis function is associated to one or more input variables, i.e. Indirect Measurements in our case. During the construction of a model, there is a first forward phase in which a greedy algorithm is used to select basis functions that give the maximum reduction in sum-of-squares residual error. This process of adding terms continues until the change in residual error is too small to continue or until the maximum number of terms is reached. There is then a backward phase that prunes the model, i.e. it removes terms one by one, deleting the least effective term at each step until it finds the best sub-model (based on the Generalized Cross Validation criterion).

III.2.2 Test case definition

Experiments are performed on the *LNA* test vehicle. For this case study, we have data from Monte Carlo simulation of 500 ICs, which comprise 12 *RF* performances to be predicted and a large set of *IM* candidates composed of 152 DC-based measurements. So, this case of study illustrates well the difficulty to choose the best subset to predict each *RF* performance.

This dataset is actually divided into two distinct sets: the first one constituted of 300 devices that will be used both for *IMs* selection and for building of regression models (training step), and the second one is constituted of 200 devices that will be used to evaluate the accuracy of predicted specification values with respect to actual specification values (validation step). More precisely for each *RF* specification, we will use the previously presented methods to select a pertinent subset of *IMs* constituted from 7 elements and we will use these subsets to build regression models using *MARS* algorithm. Then the comparison between selection methods will be performed considering the 200 devices of the validation set, based on a modified version of the *MSE* metric. This metric will be denoted *RMSNE*⁴, it allows to express model accuracy regardless of the range of the *RF* performance. This metric is an image of the rms prediction error and is expressed in percentage:

⁴ Root Mean Squared Normalized Error

$$RMSNE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2} \quad (\text{Eq. 4.8})$$

where N is the number of devices, y_i and \hat{y}_i are the actual and predicted RF performances for the i^{th} device, respectively.

Note that we choose to build models with IM subsets of 7 elements because preliminary experiments have revealed that models built with a lower number of elements selected with procedures mentioned above exhibit a poor accuracy.

III.2.3 Results and discussion

Fig. 29 compares the different feature selection methods in terms of accuracy of the predictive model built for each RF specification. The subset of IMs used for each model is selected according to each feature selection method. Here we have six different methods, the three first methods corresponding to wrappers and the other ones corresponding to filters. Regarding wrappers, the first two methods are based on an iterative search strategy using either a linear regression model (Iter_1) or a $MARS$ model (Iter_2) during the selection procedure. The last method corresponds to the built-in variable selection procedure of the $MARS$ algorithm. Regarding filters, namely CFS , $mRMR$ and Relief, IM subset selection is performed with the corresponding algorithm.

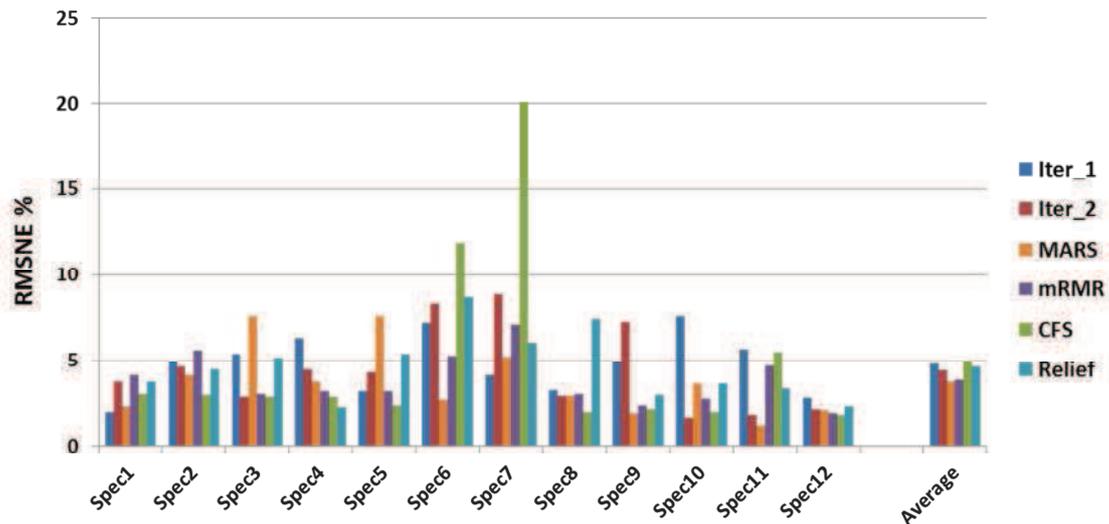


Fig. 29: Comparison of six different feature selection methods based on achieved model accuracy

Analyzing the results of **Fig. 29**, a first comment is that the achieved accuracy may significantly vary with the specification to be predicted and with the used feature selection method. However, there is no clear advantage in favor of one particular method and the average accuracy over the twelve *RF* performances is around 4% for the six methods.

A clever use of these different feature selection techniques consists in choosing the more appropriate method for each *RF* specification (i.e. the method that leads to the best model accuracy), instead of using the same feature selection method for all *RF* specifications. **Table 1** reports the average accuracy achieved over the twelve *RF* performances when using the six feature selection methods for all specifications, and when combining these methods depending on the specification to be predicted (Best*). It clearly appears that choosing the appropriate selection method for each *RF* specification leads to a significant accuracy improvement, with a reduction of the average prediction error by a factor of about 2.

Table 1: Average accuracy over the *RF* performances for the six feature selection methods and combined strategy according to specification to be predicted

	Iter_1	Iter_2	CFS	mRMR	Relief	Wrapper	Best*
Mean	4.77 %	4.40 %	4.94 %	3.84 %	4.61 %	3.75 %	2.31 %

Fig. 29 shows that there is no selection method suitable for all data and **Table 1** highlights that combining various selection methods can significantly improve the achieved prediction accuracy. But the question that must be asked is whether these selection strategies are giving solutions close to the optimum solution, which is accessible only using exhaustive subset evaluation. The exhaustive search is a brute-force- method, which is extremely time-consuming and that can be applied only when the problem size is limited. In our case, an exhaustive evaluation of subsets composed of 7 *IMs* would necessitate the construction of more than $32 \cdot 10^9$ predictive models for each *RF* specification, which is not feasible for obvious reason of computational time. Instead, we choose to perform an “exhaustive” evaluation of subsets constituted with only 3 *IMs*. Prior a post treatment on *IMs* set is done only 1/3 of *IMs* are kept to build the combination (i.e. 54 from 152 *IMs*). The operation of *IM* space reduction is discussed farther in the next section. In this case, 24 804 models are built for each one of the twelve *RF* specifications; the solution adopted as reference corresponds to the *IMs* subsets that provide the best prediction accuracy for each *RF* specification. **Fig. 30** shows that, whatever the *RF* specification to be predicted, the reference solution provides

much better accuracy than the Best* solution. As a result, the average prediction error reduces from 2.31% down to 1.18%.

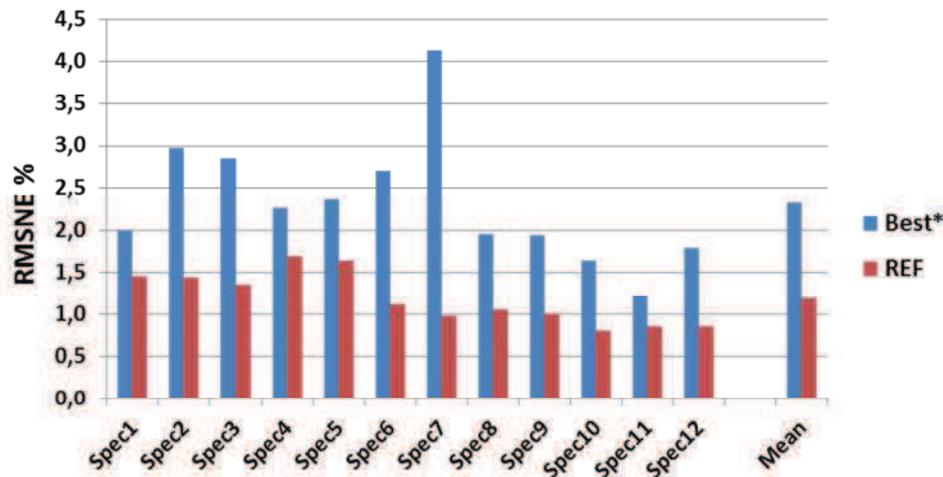


Fig. 30: Comparison of the Best* solution computed in the space of 7-IM subsets and the optimum solution in the reduced space of 3-IM subsets

From previous experiments, we can conclude that the combined feature selection strategy gives a solution with improved accuracy compared to solutions provided by a single selection method. However the achieved accuracy remains lower than the one obtained with an exhaustive search on a reduced *IM*-space. Besides, these experiments prove that there are solutions with a limited number of *IMs* that provide good model accuracy.

III.3 Proposed *IM* selection strategy

In the context of industrial testing, the most critical aspects of a test solution are accuracy and cost. The time to develop a test solution or strategy is only a secondary aspect, because it is shared by each tested device and therefore does not directly impact each device test cost. Hence, it is more interesting to spend time to define a test solution with optimized accuracy and cost rather than optimizing efforts to develop the test strategy. In this context, our objective is to define a strategy that permits to select a minimum number of *IMs*, which can then be used to build models allowing prediction of all specifications with a given accuracy. The idea is to reduce the number of indirect measurements that have to be performed during production test, and therefore the testing costs (time), but without

sacrificing accuracy. The problem to be solved is a classical coverage problem, which is known to be an NP-hard problem.

The problem is actually twofold. First, we have to identify, for each specification to be predicted, which are the *IM* subsets that give satisfying accuracy. Second, among all these possible *IM* subsets, we have to select one subset for each specification so that the total number of *IMs* required to predict all specifications is minimal. Although an exhaustive approach would permit to define the optimum test solution, it is clear that this approach is not feasible for *DUT* with several tens of *IM* candidates. We propose an alternative approach for smart *IM* selection based on three steps:

- Dimensionality reduction of the *IM*-space,
- Search space construction,
- Optimized *IM* subset selection.

III.3.1 Dimensionality reduction of IM-space

The first step of the procedure is to perform a pre-selection of a reduced number of *IMs* among all possible candidates; only these *IMs* will be used for the construction of the search space. The idea is that it is very likely that redundancy exists in the information contained by the different *IMs* and the objective is to keep only the *IMs* that contain valuable information. In this objective, our idea is to use the Principal Component Analysis (*PCA*) [42]. The *PCA* principle is to use an orthogonal transformation to convert a set of data of possibly correlated variables (in our case *IMs*) into a set of data of new uncorrelated variables (called Principal Components: *PCs*). These new data are arranged in order that the first components retain most of the information present in all the original data. The variance of the first one is as high as possible (maximal variability of the original data). The second component is built to have the highest variance possible but with the additional constraint to be orthogonal to the previous component and so on. It is clear that data variability is an important aspect for the implementation of an alternate test strategy. We can expect that the more the variations seen through *IMs*, the more efficient will be the prediction of *RF* specifications using these *IMs*.

In this context, our idea to preselect a reduced number of pertinent *IMs* is to consider only the first *PCs* and keep only *IMs* with a significant contribution (highest weighting

coefficients) to these *PCs*. The number of considered *PCs* and the thresholds on the *IM* weighting coefficients are defined empirically according to the considered application.

III.3.2 Search space construction

The second step of the procedure is to define the space of *IM* subsets and build the associated predictive models for each specification. This space has to be large enough so that the coverage problem has some solutions (i.e. it exists at least one model with satisfying accuracy for each specification), but small enough so that the resolution of the coverage problem is feasible.

For this, the idea is to iteratively construct the *IM* subset space based on the selection of a limited number k of *IM* subsets and increasing by 1 the considered *IM* subset size at each iteration. More precisely at a given iteration i and for each specification, the *IM* subsets are ranked according to the accuracy achieved by the associated predictive model and only the *IM* subsets corresponding to the k -best models are selected. These subsets composed of i *IMs* are then used in the following iteration to build models with subsets composed of $i+1$ *IMs* (i.e. subsets that include the preselected *IMs*). The procedure stops when it exists at least one model with satisfying accuracy for each specification.

III.3.3 Optimized IM subset selection

The last step of the procedure involves the resolution of the coverage problem, i.e. the determination of all *IM* subsets that permit to predict all specifications with a satisfying accuracy. From all these solutions, the one involving the best global cost is then selected. In the context of DC-based *IMs*, all indirect measurements have almost the same cost on *ATE*. Hence, the overall cost of the test solution could be assimilated to the number of used *IMs* to cover all the specifications. In the case of using complex *IMs*, different weights could be attributed to the *IMs*; the cost of a given solution then corresponds to the summation of these weights.

Note that the exact resolution of the coverage problem is feasible only for relatively small search spaces. In case of a *DUT* with many specifications to be predicted, it is likely that the search space will be too large to perform the exact resolution. In that case, heuristics may be employed to search for a satisfactory *IM* subset.

III.3.4 Results and discussion

The proposed optimized *IM* selection strategy has been applied to the LNA case study. The first step of this strategy is to reduce the set of *IM* candidates using *PCA*. In this experiment, we consider only the first 10 *PCs* and we only keep the 10 *IMs* with the highest weighting coefficients in the first *PC*, 9 *IMs* with the highest weighting coefficients in the second *PC*, and so on for the following *PCs*. With this preliminary process, the initial set of 152 *IMs* is reduced down to a set of 54 *IMs*. The second step of the strategy consists in the construction of the search space. Based on the previous observation that shows that only few *IMs* are actually sufficient to reach good accuracy, we arbitrarily choose to construct the search space considering all possible combinations of 3 *IMs* among the pre-selected 54 *IMs*, which corresponds to 24,804 predictive models built for each *RF* specification. Finally, the last step of the strategy is the selection of an optimized *IM* subset that allows to tradeoff test accuracy and global test cost. For this, we keep only 20 models with best accuracy (in terms of *RMSNE* evaluated on the training set of devices) for each specification. From the resolution of the coverage problem, we then select the solution that minimizes the number of *IMs* required to perform prediction over the global specification set.

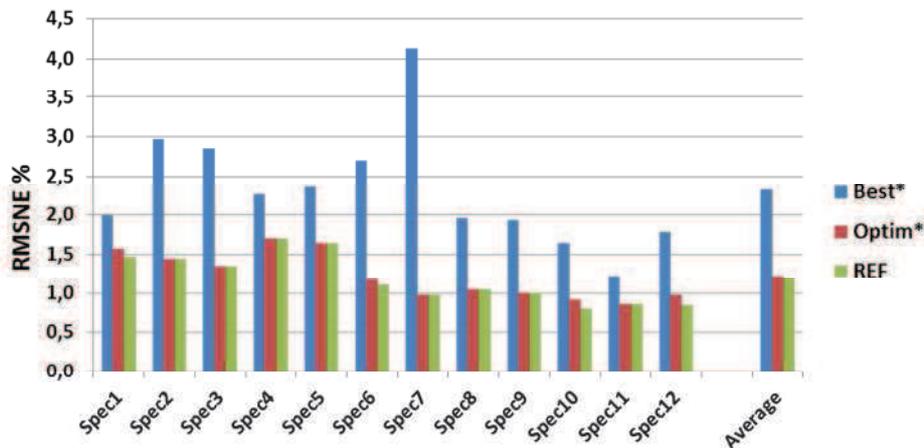


Fig. 31: Accuracy of the solutions given by Best*, Optim* and REF selection strategies

Fig. 30 shows the accuracy (in terms of *RMSNE* evaluated on the validation set of devices) achieved with the optimized *IM* selection strategy (Optim*) compared to the accuracy achieved with the combined feature selection strategy (Best*) and the reference solution (REF) obtained by selecting among the 24,804 models available for each *RF* specification, the one that provide the best prediction accuracy.

These results clearly demonstrate the superiority of the optimized *IM* selection strategy compared to the combined feature selection strategy, with an average prediction error over the twelve *RF* specifications of 1.22% instead of 2.31%. Moreover there is no significant degradation compared to the reference solution, with an accuracy difference of only 0.04%. In contrast, there is a significant gain in terms of number of *IMs* required to perform the prediction of all *RF* specifications, as illustrated in **Fig. 32**. Only 14 *IMs* are required with the optimized strategy instead of 18 *IMs* for the reference solution, which corresponds to a global test cost reduction of 22%.

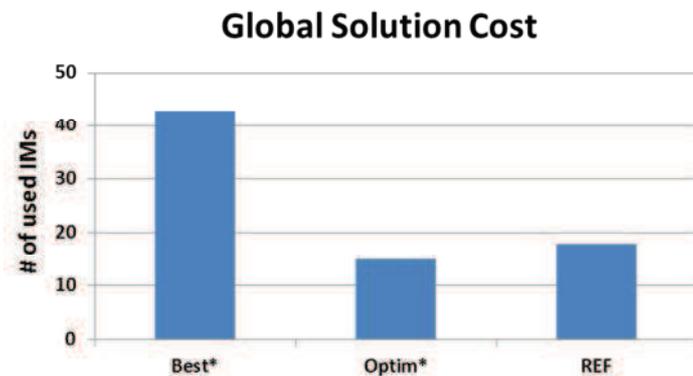


Fig. 32: Cost of the solutions given by Best*, Optim* and REF selection strategies

Finally to summarize, these results show that it is more interesting to use a method based on brute-force search for the pertinent selection of *IM* subsets rather than a strategy based on feature selection algorithms, even if the search space is not exhaustive. This permits to identify models built with few *IMs* that have good accuracy for each specification. Moreover by evaluating a large number of models, the overall cost of the test solution can be optimized without significant loss of accuracy.

III.4 Conclusion

In this chapter, we have explored different strategies for the selection of an appropriate subset of *DC* measurements to be used for *RF* performance prediction. In particular, the efficiency of some commonly-used feature selection algorithms are investigated and confronted to solution obtained with an exhaustive search. An alternative selection strategy that relies on the smart construction of a search space is developed with the objective to reduce the overall cost while maintaining the accuracy. The different strategies are evaluated and validated on the *LNA* test vehicle, for which a large set of 152 *IMs* is available to perform the prediction of twelve *RF* specifications. For this case study, results show that the accurate estimation of the *RF* performances can be achieved with only limited number of *DC* measurements using the proposed optimized selection strategy.

Chapter IV

Strategy for increasing robustness of prediction-oriented indirect test

IV.1 Introduction

As explained in chapter III, the main problem related to indirect testing, highlighted by industrial test engineers, is the lack of confidence in the efficiency of the method. The objective of this chapter is to tackle this issue in order to demonstrate the major breakthrough that the indirect testing strategy can make in the domain of low-cost testing of *RF ICs*. As shown in chapter III, the strategy is accurate for most of the tested devices but for few devices, large prediction errors are observed (these devices are commonly called outliers). From an industrial point of view, the misestimated *DUTs* might induce some test escapes or yield losses. They are therefore significant problems since one of the major challenges of production test is to guarantee low defective level (typically few *PPMs*). Additionally, the dependency of prediction model performance with respect to the data used during the training phase to build the model (both selected *IMs* and training devices) is also a concern. All these facts make industrial test engineers reluctant to implement the indirect test strategy for volume production testing of *RF ICs*.

In this chapter, we introduce a novel implementation of the prediction-oriented indirect test that permits to manage incorrect predictions and ensures a good robustness of the test procedure, even in case of models built with low correlation *IMs* and training set of reduced size. This solution is based on information redundancy, i.e. the same specification of a device will be predicted many times using different models built in different contexts; the resulting predicted values can then be cross-compared to identify anomalies in predictions. All along this chapter, we will illustrate the efficiency of this strategy using the *PA* test vehicle for which we have a large set of experimental data as a case study.

IV.2 Outlier definitions

IV.2.1 Process-based outlier

In the literature, some works investigate the problem of outliers which are usually related to devices affected by catastrophic faults occurring during the manufacturing process. In particular, a defect filter based on a kernel density estimation technique is presented in [15] that permits to screen out devices not consistent with the statistical distribution of manufactured devices. Obviously these devices should not be submitted to the indirect test procedure as they are likely to be incorrectly predicted. Note that this filter directly operates on indirect measurement data and therefore permits to identify devices qualified as outliers before any processing through a regression model.

IV.2.2 Model-based outlier

Let us now assume that we have a number of devices free of process-based outliers that are processed by the indirect test procedure. As pointed out in Chapter II, accurate prediction results can be achieved for most of the devices. However aberrant or flawed predictions are observed for a small number of devices. We define these devices as model-based outliers. Indeed such incorrect predictions do not depend on the device itself since the same device may exhibit correct or aberrant/flawed prediction depending on the considered regression model (cf. **Fig. 25**), but are generated by the predictive model. These model-based outliers are a real concern for the efficiency of the indirect test strategy since they might generate test escape or yield loss.

IV.3 Redundancy-based filter for model-based outliers

IV.3.1 Principle

In order to have an efficient implementation of the indirect test strategy, the procedure should be able to properly manage both process-based and model-based outliers. Process-based outliers can be easily detected using conventional statistical filters and removed from the test procedure. However to the best of our knowledge, there is no existing solution to identify model-based outliers.

In this context, our idea to identify model-based outliers is to introduce an additional step in the procedure that provides for each device an indication on the confidence that can be placed on the prediction. If this confidence is low, the device is identified as a potential model-based outlier and removed from the indirect test flow for further action to be taken.

This approach is similar to the approach suggested in [14] [17] in case of classification-oriented strategy, where guard-bands are allocated in the indirect measurement space in order to identify devices for which the alternate 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.

Our idea to provide an indication of confidence in the context of prediction-oriented indirect test is to exploit model redundancy [43]. Indeed as previously mentioned, different regression models may lead to either correct or aberrant/flawed predictions for the same device. Our proposal is therefore to use multiple models to predict one specification and to crosscheck the predictions obtained with the 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, we cannot know which one of the predictions is correct and which is not. We consequently consider that the prediction for this device cannot be trusted: the prediction is considered suspect and the device is removed from the alternate test flow. In other words, we use model redundancy to distinguish reliable predictions from suspect predictions. This operation will be denoted Check Prediction Consistency (*CPC*) procedure.

This strategy, called two-tier alternate test scheme using model redundancy, is illustrated in **Fig. 33**. In the conventional implementation of the prediction-oriented alternate 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. 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 ϵ . 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 ϵ , 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 alternate test flow and directed to the second tier, where further testing may be applied to characterize the device (for instance standard specification testing).

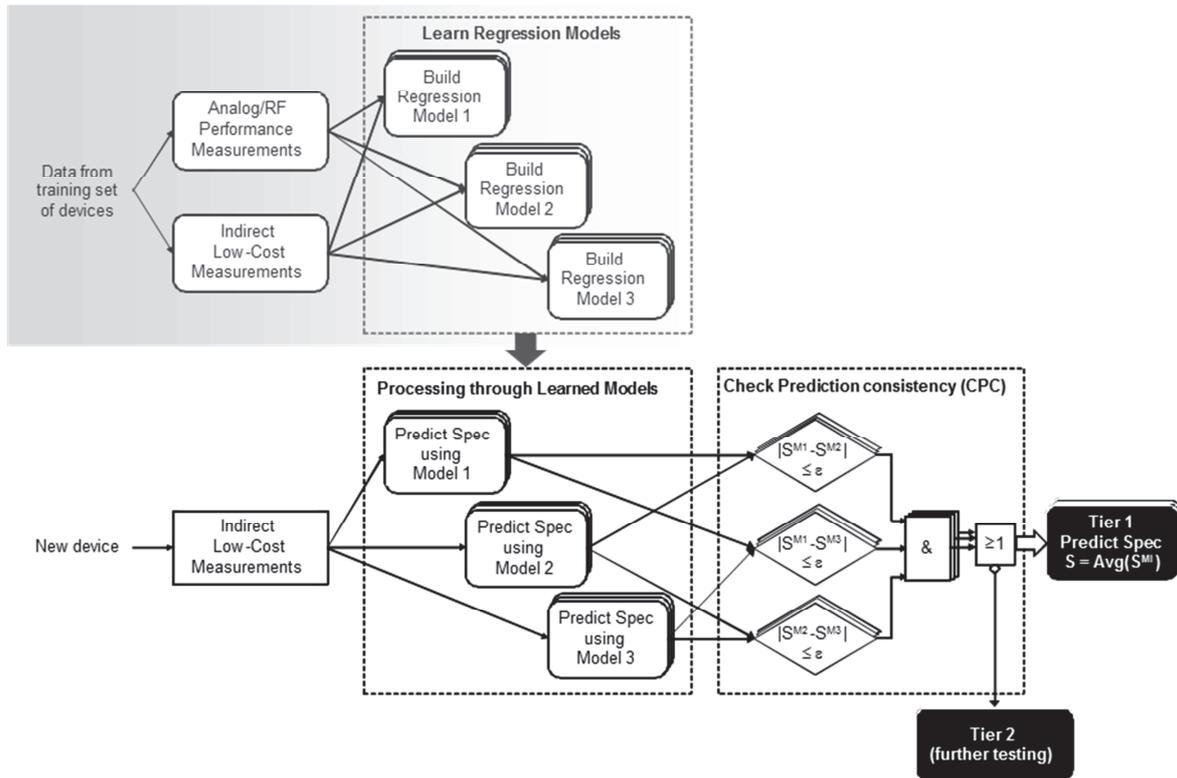


Fig. 33: Flow diagram of the proposed two-tier indirect test scheme using model redundancy.

IV.3.2 Experimental results using accurate predictive models

A number of experiments have been performed on the *PA* case study to validate the proposed strategy. The experimental test data available from *10,000* devices are split into two subsets of *5,000* devices with similar statistical properties, one for training and the other for validation. In all experiments presented in this section, models are built using the *5,000* devices of the training set and evaluation is conducted using the *5,000* devices of the validation set.

First in a preliminary step, we have chosen the combination of *IMs* that will be used to build the redundant models. For this, we have built all regression models based on triplets of

IMs, both for *CP1* and *IP3* specifications. For each specification, we have selected the 3 combinations offering the best performance in terms of average prediction error evaluated on the training set (see previous chapter). Results are summarized in the first sub-table of **Table 2**. It can be seen that the 3 models selected for each specification have almost equivalent performances: the average prediction error ranges between 0.62% and 0.69% for *CP1* prediction and between 0.92% and 1.19% for *IP3* prediction, and the maximal prediction error ranges between $0.81dB$ and $0.85dB$ for *CP1* prediction and between $1.10 dB$ and $1.24dB$ for *IP3* prediction.

Table 2: Average and maximal prediction errors for *CP1* and *IP3* specifications

	Training			
	CP1		IP3	
	Average prediction error	Maximal prediction error	Average prediction error	Maximal prediction error
Model 1	0.62%	$0.83 dB$	1.12%	$1.24 dB$
Model 2	0.62%	$0.85 dB$	0.92%	$1.10 dB$
Model 3	0.69%	$0.81 dB$	1.19%	$1.19 dB$

	Validation			
	CP1		IP3	
	Average prediction error	Maximal prediction error	Average prediction error	Maximal prediction error
Model 1	0.79%	$3.49 dB$	1.18%	$4.60 dB$
Model 2	0.71%	$2.04 dB$	1.29%	$14.01 dB$
Model 3	0.71%	$0.90 dB$	1.23%	$5.64 dB$
Mean of models	0.75%	$1.44 dB$	1.05%	$3.32 dB$

	After removing suspect predictions ($\epsilon_{CP1}=1dB, \epsilon_{IP3}=2dB$)			
	CP1		IP3	
	Average prediction error	Maximal prediction error	Average prediction error	Maximal prediction error
Mean of models	0.61%	$0.90 dB$	1.02%	$1.13 dB$

Then in a second step, these models are used to predict the performance parameter values for each device in the validation set. Results are summarized in the second sub-table of **Table 2**. Looking at the different models independently; the average prediction error is almost preserved, ranging between 0.71% and 0.79% for *CP1* prediction and between 1.18% and 1.29% for *IP3* prediction. However the maximal prediction error significantly increases compared to the maximal prediction error observed on the training set of devices: maximal prediction error reaches $3.49dB$ for *CP1* prediction and $14.01dB$ for *IP3* prediction.

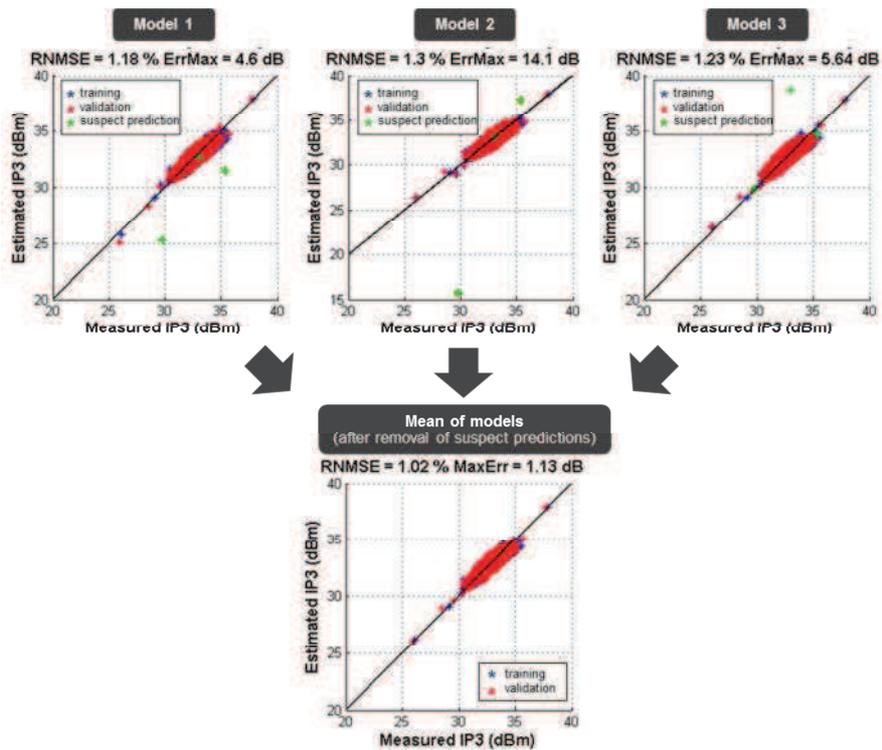
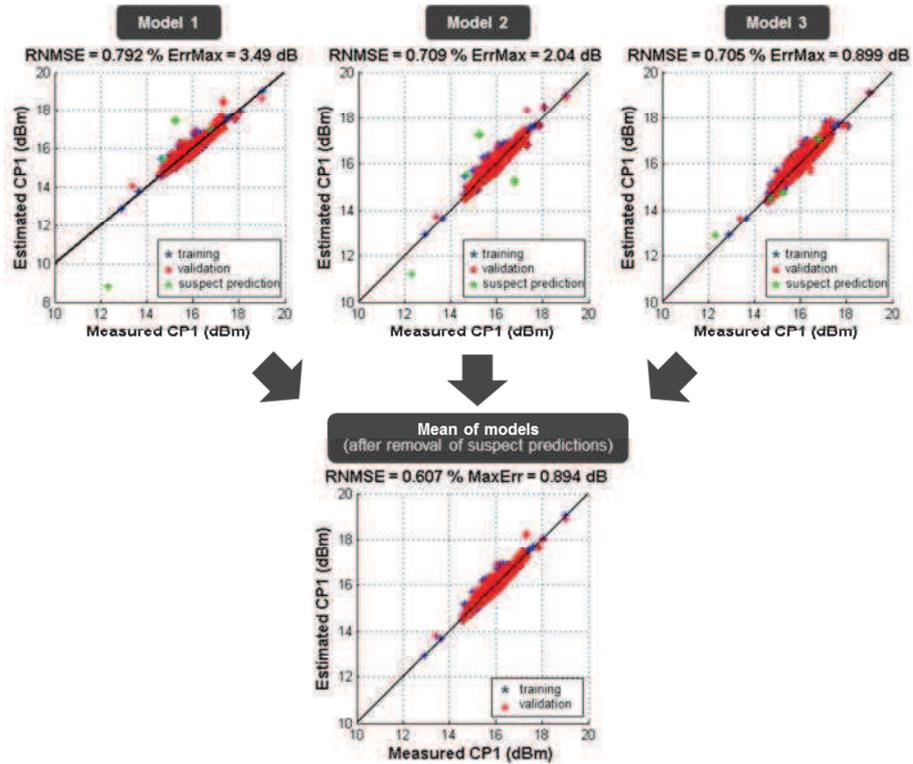


Fig. 34: Correlation plots for *CP1* and *IP3* predictions using 3 different models and after removal of suspect predictions (models built on 5,000 training devices)

This is due to the fact that some devices suffer from flawed predictions, as clearly illustrated on the scatter plots of **Fig. 34** that report the predicted value versus the actual value for *CP1* and *IP3* predictions, respectively. Note that averaging predictions of the 3 models permits to reduce the maximal prediction error, but the error is still much larger than the maximal prediction error expected from the training set.

In the last step, the prediction confidence procedure is applied considering a threshold value $\varepsilon_{CP1}=1dB$ for *CP1* prediction and $\varepsilon_{IP3}=2dB$ for *IP3* prediction. This procedure permits to identify 6 suspect predictions regarding *CP1* specification and 5 suspect predictions regarding *IP3* specification, represented by green marks in the scatter plots of **Fig. 34**. It can be seen that all flawed predictions are correctly identified by the prediction confidence procedure. Results in terms of average and maximal prediction errors obtained after removing devices with suspect prediction are reported in the last sub-table of **Table 2**: the average prediction is now 0.61% for *CP1* prediction and 1.02% for *IP3* prediction, and the maximal prediction error is $0.90dB$ for *CP1* prediction and $1.13dB$ for *IP3* prediction. These maximal prediction errors are in the same range than the maximal prediction errors expected from the training set, therefore demonstrating the efficiency of the proposed strategy.

In the previous experiment, the detection threshold was arbitrarily set to value $\varepsilon_{CP1}=1dB$ and $\varepsilon_{IP3}=2dB$. We have investigated how the performances of the proposed strategy are impacted by the value of this threshold. Results are summarized in **Fig. 35**, which give the evolution of the maximal prediction error and the percentage of devices directed to the second tier (i.e. devices with suspect predictions) with respect to the value of the detection threshold, regarding *CP1* and *IP3* specifications, respectively. It clearly appears that an adequate value of the detection threshold can be easily found that significantly reduces the maximal prediction error while only a limited number of devices are directed to Tier 2 due to suspect predictions.

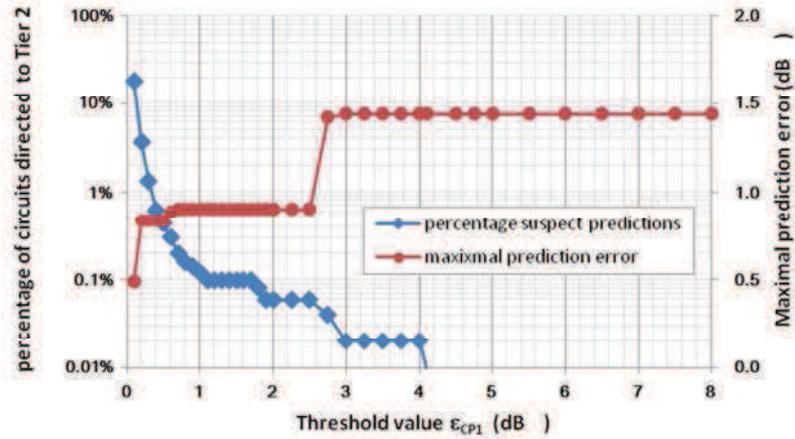
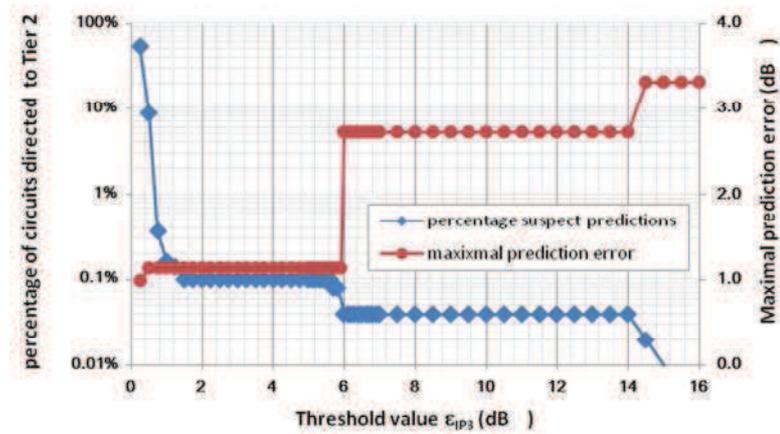
(c) *CP1 specification*(d) *IP3 specification*

Fig. 35: PA performance prediction with model redundancy: impact of the threshold value $\epsilon_{CP1/IP3}$ used during identification of suspect predictions

For *CP1* specification with a detection threshold ϵ_{CP1} between $1dB$ and $2.5dB$, the maximal prediction error observed over the training set of $5\ 000$ devices is limited to $0.90dB$ while less than 0.1% of the devices are directed to Tier 2. In the same way, for *IP3* specification with a detection threshold ϵ_{IP3} between $1.5dB$ and $5.9dB$, the maximal prediction error is limited to $1.13dB$ while less than 0.1% of the devices are directed to Tier 2. Note that there is not real benefit at lowering the detection threshold below these values because it results in a minor reduction of the maximal prediction error but at the price of a very strong increase in the number of devices directed to Tier 2.

IV.4 Making predictive indirect test strategy independent of training set size

The previous section has introduced the concept of model redundancy to establish prediction confidence and demonstrated its efficiency to identify suspicious predictions. As a consequence, aberrant or flawed predictions can be avoided and accurate predictions results are achieved for all circuits processed by the indirect test tier. It is worth noting that these results were obtained using predictive models built on a large training set of 5,000 devices. The objective of this section is to investigate the efficiency of the proposed strategy in case of training sets of reduced size. Indeed as highlighted in Chapter II, the Training Set Size (*TSS*) has a significant impact of the quality of the built regression models, in particular regarding they ability to correctly predict new devices different from the training devices (cf. **Fig. 27**), which shows that the number of circuits affected by large prediction errors augments when the size of the training set reduces). However using large training sets increases the overall cost of the test method since the conventional and expensive RF measurements have to be performed for all devices of the training set.

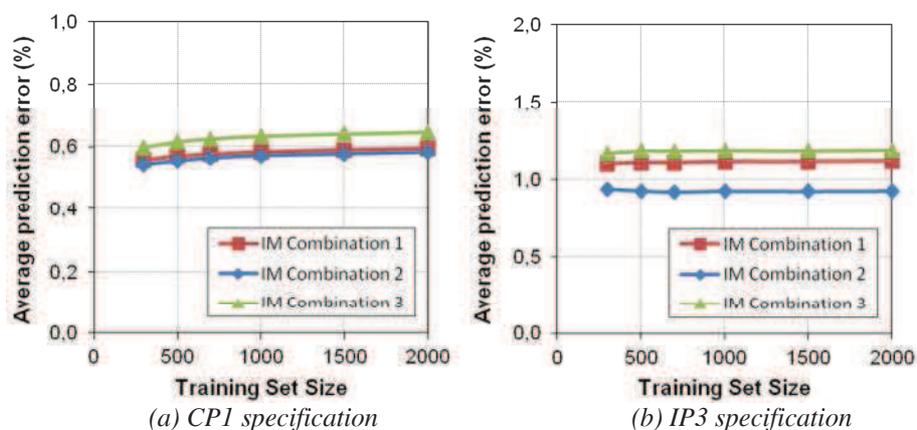


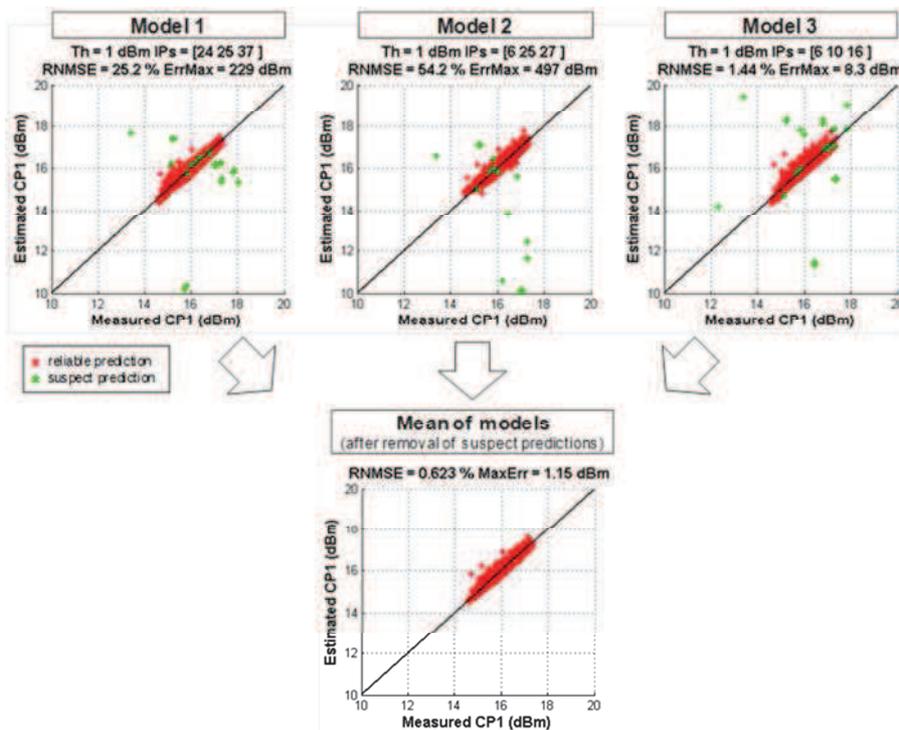
Fig. 36 : Average prediction error of selected redundant models vs. TSS.

Experiments have been performed on the same case study in order to evaluate the influence of the training set size. More precisely, the redundant models have been built for different sizes of training set ranging from 100 to 2,000 devices randomly chosen in the initial training set of 5,000 devices. Note that we have verified that the model performance remains unchanged for the different sizes of the training set. As illustrated in **Fig. 36**, the 3 selected models (for each specification) lead to similar performances with a constant average

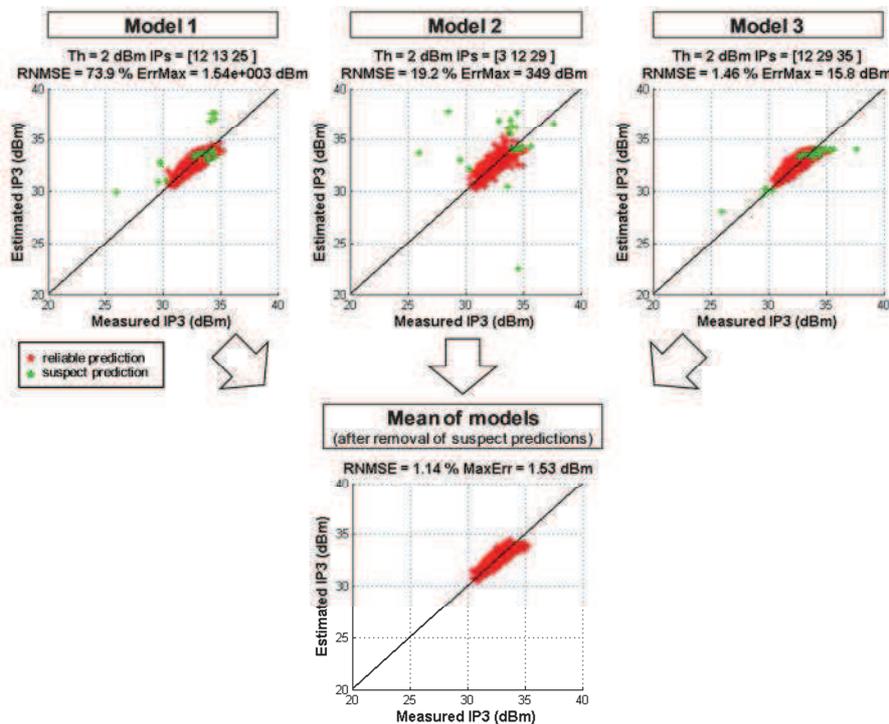
prediction error of about 0.6% for $CP1$ specification and about 1% for $IP3$ specification, whatever the TSS value.

Using these selected models, the efficiency of the proposed alternate test implementation has then been evaluated for different TSS . Detection threshold for prediction consistency checking has also been set to $1dB$ for $CP1$ specification and $2dB$ for $IP3$ specification.

As an illustration, **Fig. 37** gives correlation plots between predicted and actual values for $CP1$ and $IP3$ specifications respectively, associated to the 3 different models built on a set of 300 devices randomly chosen in the TS . For both specifications, it clearly appears that whatever the regression model, some circuits suffer from large prediction errors. However the efficiency of the procedure that checks prediction consistency is also clearly demonstrated since all circuits affected with large prediction error are correctly identified. As a result when suspect predictions are removed, a good agreement is observed between the value computed as the average of the 3 predictions and the actual value, for all circuits. On this example, 25 circuits are identified as circuits likely to be affected with large prediction errors regarding $CP1$ specification, and 27 regarding $IP3$ specification. After removal of these circuits, maximal prediction errors are limited to $1.15dB$ for $CP1$ specification and $1.53dB$ for $IP3$ specification, respectively. Note that some circuits are identified as suspect circuits for both specifications, so the number of circuits directed to the second tier of the test flow is actually only 38, which corresponds to 0.76% of the 5,000 processed circuits.



(a) CP1 specification ($\epsilon_{th}=1dB$)



(b) IP3 specification ($\epsilon_{th}=2dB$)

Fig. 37: Correlation plots for CP1 and IP3 predictions using 3 different models and after removal of suspect predictions (models built on 300 training devices)

More generally, the efficiency of the proposed alternate test implementation for different training set sizes is summarized in **Fig. 38** that compares prediction errors observed when using each regression model individually (excluding *OOB* predictions) to prediction errors observed after removal of suspect predictions and averaging of the 3 prediction results. It clearly appears that the proposed alternate test implementation offers better performance than the classical one, both in terms of average and maximal prediction errors. More in details, it can be observed that when using a single regression model and excluding *OOB* predictions, results depends on both the used model and the training set size. The average prediction error ranges between 1.12% and 1.99% for *CP1* specification and between 1.32% and 1.90% for *IP3* specification, whereas expected prediction accuracy from the training is of about 0.6% for *CP1* specification and about 1% for *IP3* specification.

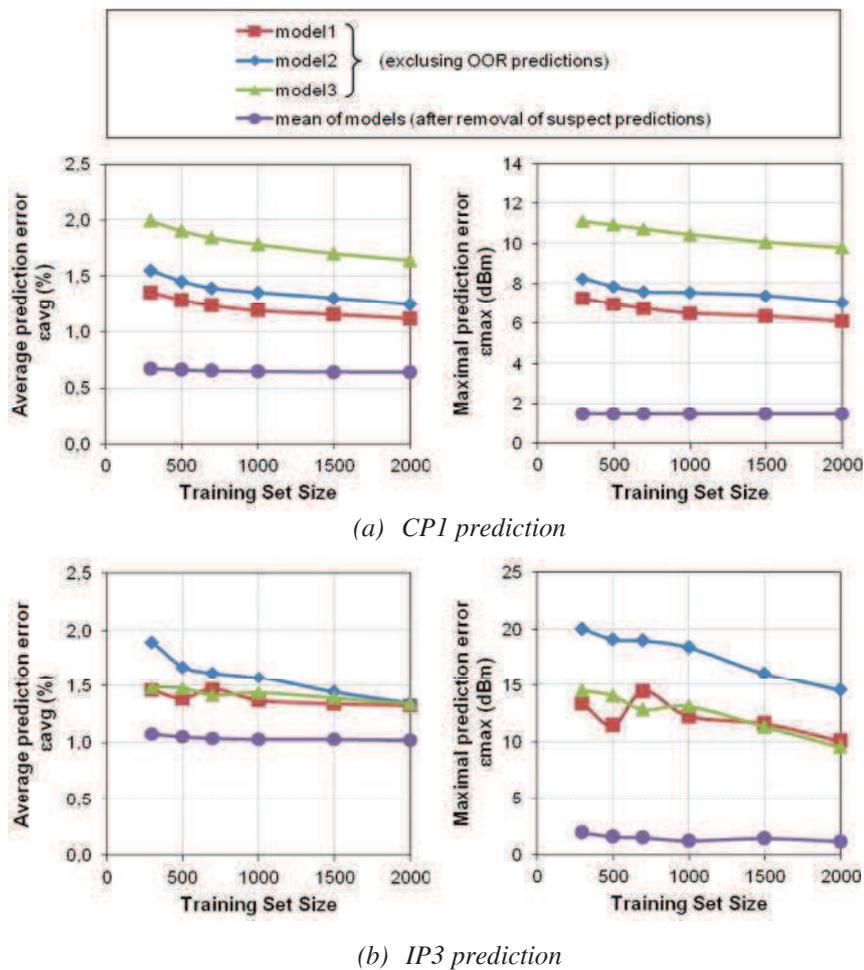
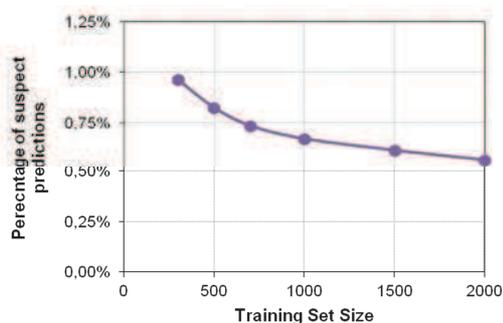


Fig. 38: Average and maximal prediction errors vs. *TSS*

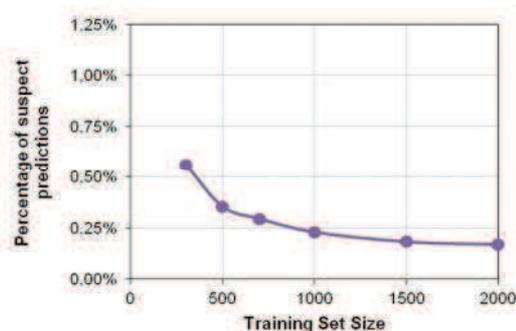
In the same way, even if OOR predictions are excluded, large prediction errors are still obtained for some circuits with a maximal prediction error that ranges between $6.1dB$ and $11.1dB$ for *CP1* specification and between $9.5dB$ and $20.1dB$ for *IP3* specification.

In contrast when the prediction consistency checking procedure is introduced in the test flow, results are independent of the training set size and good prediction accuracy is obtained for all circuits evaluated by the alternate test tier. The average prediction error is around 0.7% for *CP1* specification and 1.1% for *IP3* specification, and the maximal prediction error is limited to $1.9dB$ for *CP1* specification and $1.5dB$ for *IP3* specification, whatever the *TSS*. Of course, these good results are obtained at the price of some circuits directed to the second tier of the test flow where additional testing should be applied to characterize them. However it is worth noting that the number of these circuits is extremely small.

To illustrate this point, **Fig. 39** reports the percentage of identified suspect predictions regarding *CP1* and *IP3* specifications for different training set sizes. As expected, this percentage increases when the number of instances in the training set reduces, but in both cases it remains below 1% . Consequently, even with a small training set of only 300 devices, less than 1.5% of the circuits are directed to the second tier of the test flow.



(a) *CP1* prediction



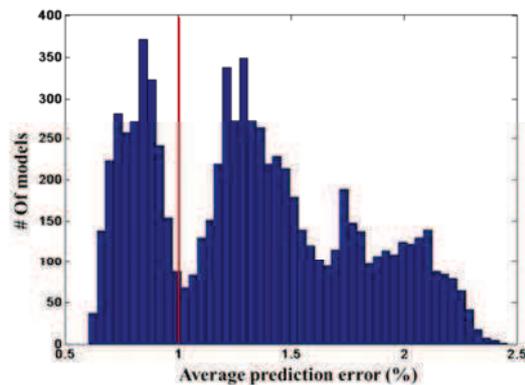
(b) *IP3* prediction

Fig. 39: Percentage of suspect predictions vs. *TSS*

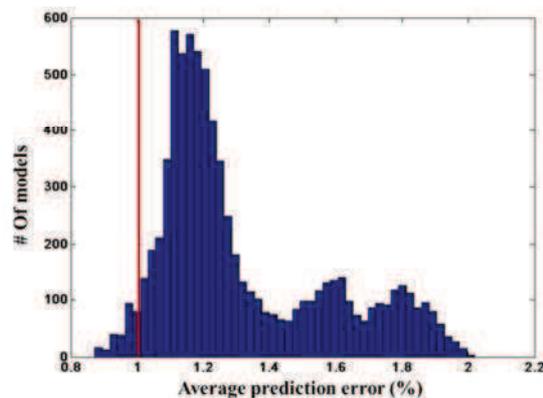
To summarize, these results show that on the contrary to the classical implementation where prediction accuracy degrades when reducing the training set size, the proposed approach permits to preserve prediction accuracy independently of the training set size, while only a very small number of devices are directed to the second tier of the test flow.

IV.5 Model redundancy to balance the lack of correlation between IMs and RF performance parameters

In this section, we investigate the efficiency of the proposed strategy in case of non-accurate redundant models. Indeed results presented in the previous sections rely on the use of 3 accurate models involving different combinations of indirect measurements for each specification. However depending on the product to be tested, it may happen that the number of potential *IMs* and/or the correlation between these *IMs* and the product specifications is limited. In this case, it may be difficult to build redundant models with a satisfying accuracy. The objective of this section is to analyze such a situation using the *PA* as a case study.



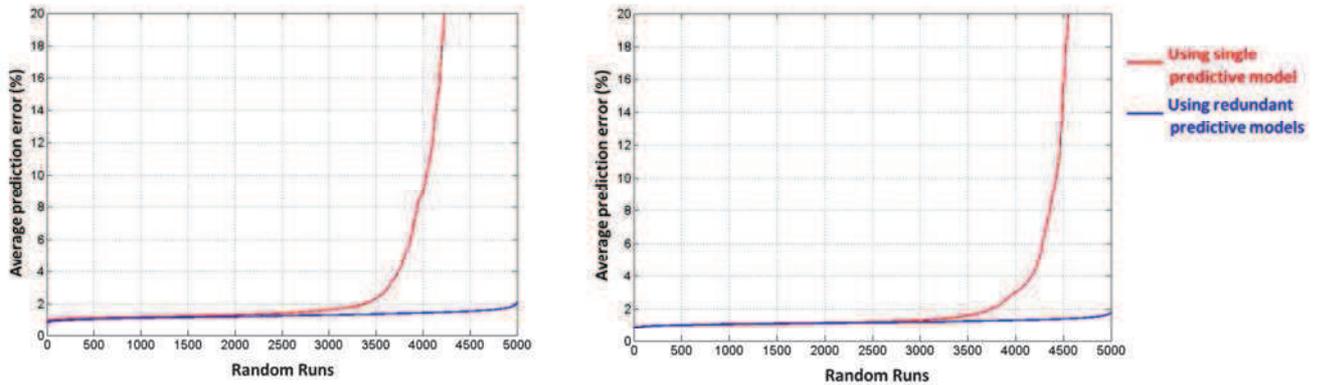
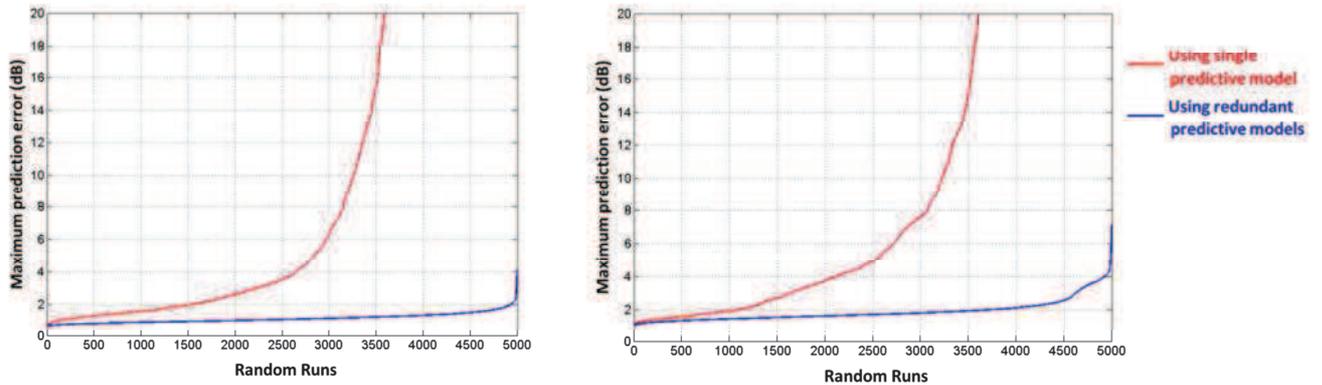
(a) *CPI specification*



(a) *IP3 specification*

Fig. 40: Distribution of average prediction error (computed on *TS*) for the exhaustive set of predictive models built using triplets of *IMs*

As previously mentioned we have a set of $37IMs$ for this test vehicle, which means that we can build $7,770$ different models based on triplets of IMs for each specification. Obviously not all these models exhibit a good accuracy, as illustrated in **Fig. 40** that shows the distribution of the average prediction error (computed on the training set of 5,000 devices) for all the $7,770$ models. For both specifications, there are some models with very good accuracy but the majority of the models have an average prediction error above 1% .

(a) *CPI specification*(b) *IP3 specification***Fig. 41:** Average prediction error observed on VS over the 5,000 runs(a) *CPI specification*(b) *IP3 specification***Fig. 42:** Maximum prediction error observed on VS over the 5,000 runs

For the purpose of this study, we intentionally choose to implement the redundant strategy using models with a degraded accuracy. More precisely, we choose to implement the strategy using 3 models randomly chosen among the set of models with an average prediction error (computed on TS) exceeding 1% for each specification, and repeating this experiment

5,000 times. For each run, the efficiency of the strategy is evaluated in terms of average and maximal prediction errors observed on the validation set of 5,000 devices, and compared to the conventional implementation of the indirect test strategy using a single predictive model for each specification (among the 3 redundant models, only the model with the lowest prediction error computed on *TS* is kept).

Results are summarized in **Fig. 41** and **Fig. 42** for *CP1* and *IP3* specifications respectively. Note that for readability reasons, values are sorted in ascending order for each curve. The benefit of the redundant strategy clearly appears from these graphs. Indeed good accuracy is observed over the 5,000 runs when using this strategy, with an average prediction error evaluated on *VS* that remains below 2% for both specifications whatever the random run. In contrast when using the conventional strategy with a single predictive model, the average prediction error exceeds 2% for about 35% of the random runs in case of *CP1* specification, and about 25% in case of *IP3* specification. The superiority of the redundant strategy is even stronger when looking at maximal prediction errors. In case of *CP1* specification, the maximum prediction error is below 2dB only for 30% of the random runs when using the conventional strategy, while the redundant strategy permits to reach this condition for 99% of the random runs. Similarly in case of *IP3* specification, the percentage of occurrences for which the maximum prediction error is below 2dB is around 20% with the conventional strategy, against 75% with the redundant strategy. Note that these good performances are achieved with only a small number of circuits removed from the indirect test tier, as illustrated by **Table 3** that reports the percentage of circuits identified as suspect predictions for both *CP1* and *IP3* specifications.

Table 3: Percentage of suspect predictions:
average and maximum values over the 5,000 runs

	Percentage of suspect predictions	
	Average	Maximum
<i>CP1</i>	0.97%	3.46%
<i>IP3</i>	0.40%	1.24%

More generally, these results demonstrate the efficiency of the redundant strategy to correctly identify model-based outliers, even if the predictive models have a limited accuracy. This is a very interesting feature because by relaxing the constraint on model accuracy, it widens the application field of the indirect test strategy to the case of products with a reduced number of exploitable *IMs* and/or *IMs* that present imperfect correlation with specifications. It

also gives great flexibility for test engineers to choose the set of used *IMs*, and therefore offers interesting perspective in terms of cost-optimized implementations.

IV.6 Different implementation schemes of redundancy

IV.6.1 Motivation

In the initial redundancy strategy, prediction confidence is established based on the use of 3 regression models that involve different combinations of indirect measurements, for each specification. As described previously, the *CPC* module qualifies the different predictions as confident and suspect predictions [44]. The efficiency of the proposed scheme relies on the assumption that it is very unlikely that 3 regression models built using different combinations of indirect measurements will erroneously predict the device performance with the same error. To widely validate the assumption, we have performed additional experiments on the PA case study. In these experiments, we have evaluated the performances achieved by the redundant strategy regarding *CP1* specification when using 3 regression models randomly chosen among a set of 5,000 acceptable models and repeating this experiment 100 times.

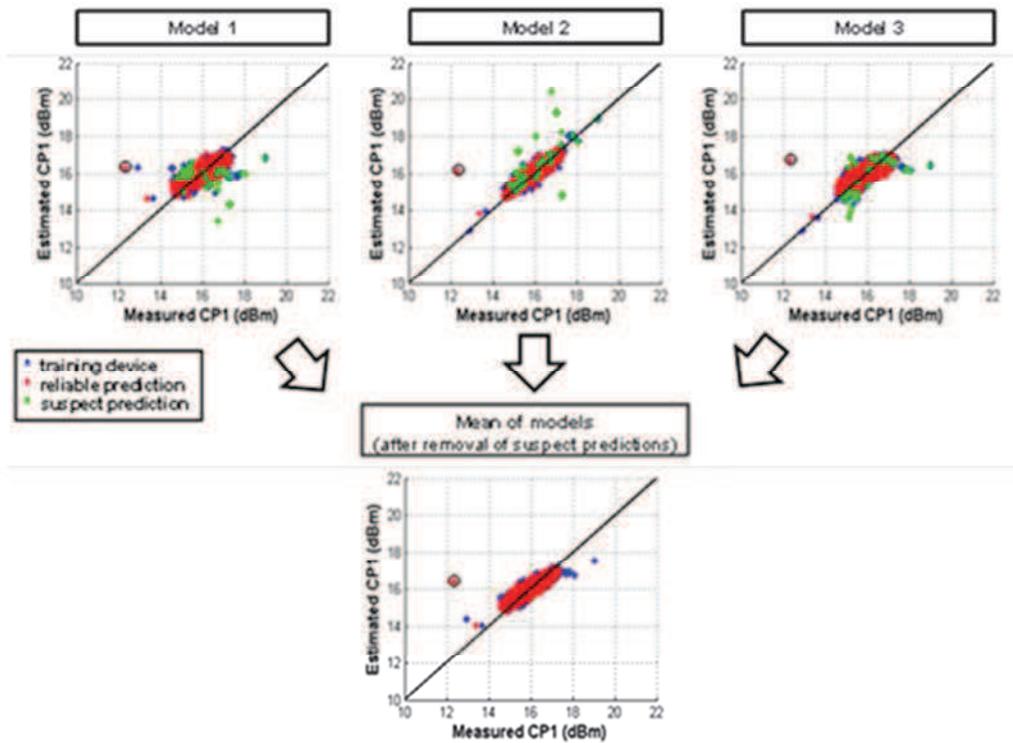


Fig. 43: Particular *CP1* prediction example illustrating the limitation of the proposed implementation.

Results show that there is only one case over the 100 experiments, for which the procedure fails to identify a relatively large error on one device prediction of the validation set. This example is illustrated **Fig. 43**. It can be clearly seen that the *CPC* procedure permits to identify all devices affected with relatively large prediction error, except one. For this particular device and these particular regression models, the 3 models actually yield to an erroneous performance prediction with a similar error (difference between each pair of predicted values is less than 1dBm). The prediction is therefore not detected as a suspect one and the device is not directed towards the appropriate test tier.

Our objective is to reinforce redundancy in order to avoid such situation. More precisely, the idea is to increase the number of redundant models used in the procedure that checks prediction consistency in order to diminish the probability that all models give the same erroneous prediction for one device.

IV.6.2 Principle of augmented redundancy CPC schemes

A first option to increase redundancy is to continue to exploit the different indirect measurements and to build, for each specification, not only 3 but a higher number of regression models based on different *IM* combinations. However it is clear that selecting a higher number of different *IM* combinations will inevitably imply a higher number of indirect measurements that have to be performed, which has a direct impact on the testing cost. This option is therefore not favored. Instead, our idea is to exploit another attribute of the data available for the construction of regression models related to training devices. Indeed, data used for the construction of a regression model clearly involves two aspects, i.e. indirect measurements and training devices. Any change in the set of indirect measurements or in the set of training devices results in a different model. Up to now, redundant models were built by changing the set of considered indirect measurements. The basic principle is to maintain this feature but also to exploit the other aspect, i.e. to build redundant models by changing the set of training devices. In particular, the idea is to split the training set in a number of partitions and to build regression models for the different partitions.

Our proposal for the generation of redundant regression models, for each specification S_j , is summarized in **Fig. 44**. On one hand 3 combinations of indirect measurements are selected from the set of available indirect measurements, and on the other hand the training

set is split in 3 distinct partitions. For each selected *IM* combination C_i , a regression model is then built considering the different partitions of the training set. A total number of 9 redundant models are therefore built, which exploit both different *IM* combinations and different partitions of the training set. These 9 models will be used during the testing phase in order to establish prediction confidence.

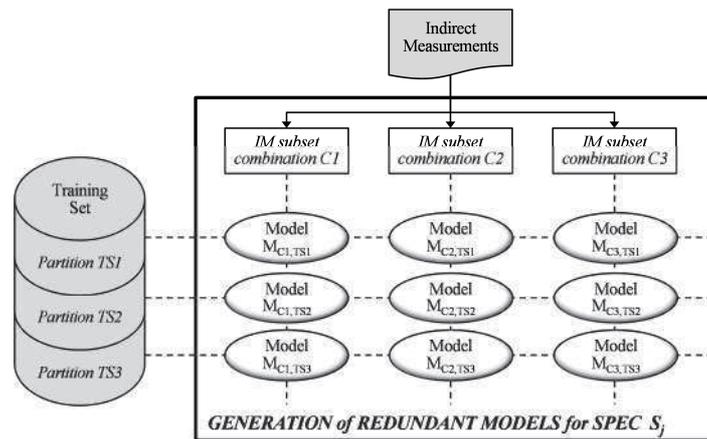
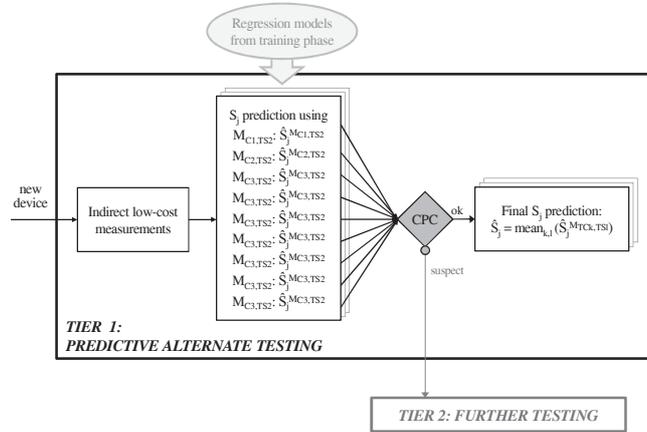
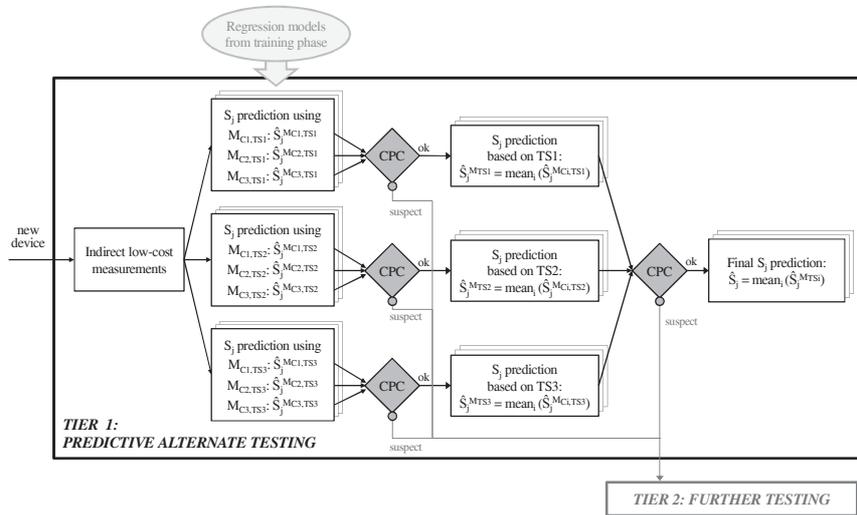


Fig. 44: Generation of redundant regression models

We have actually explored different implementations of the procedure that *CPC*. In the first version, consistency between the values predicted by the 9 models is checked in a single step, i.e. the difference between the predicted values is computed for each pair of models and prediction is considered reliable only if all these 36 differences are inferior to a given threshold ϵ_{thi} . In the case, the final predicted value is computed as the mean of the 9 predicted values. In the two other versions, prediction consistency is verified in two steps. In the first step, consistency between predicted values is checked considering models 3 by 3 according to the used training set partition in one version or the used *IM* combination in the other version; then 3 intermediate predicted values can be computed. In the second step, consistency between these 3 intermediate values is checked and the final predicted value is computed as the mean of these 3 values. These different implementations of the two-tier alternate test scheme with augmented model redundancy are summarized in **Fig. 45**.

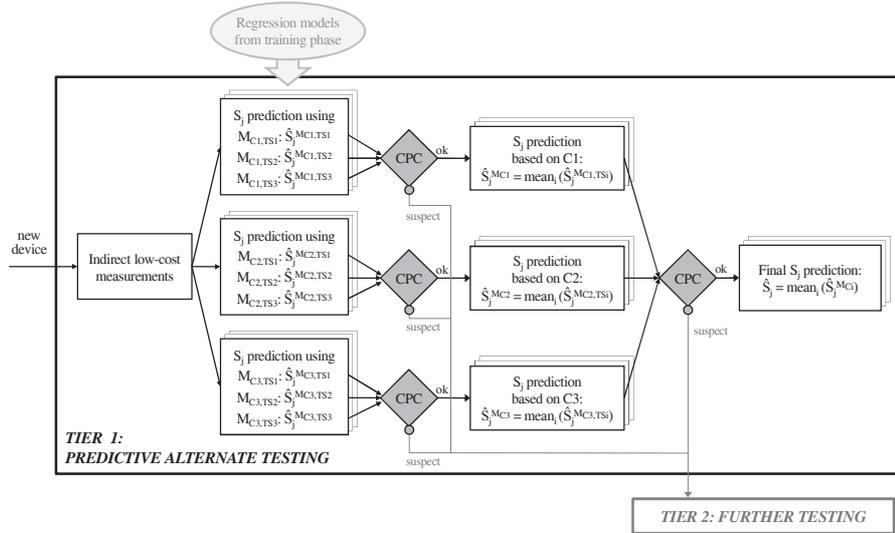


(a) Version A - one-step CPC procedure



(b) Version B - two-step CPC procedure

(Check prediction consistency between results issued from different IM combinations first)



(c) Version C - two-step CPC procedure

(check prediction consistency between results issued from different partitions of the training set first)

Fig. 45: Implementations of the two-tier indirect test scheme with augmented model redundancy

IV.6.3 Experimental Results

A number of experiments have been performed on our case study (*PA*) to validate these new implementations. First, we have considered the particular *CP1* prediction example discussed in previous sub-section, for which the initial implementation fails to identify a relatively large prediction error on one device of the validation set. Results are summarized in **Table 4** that reports both average and maximal prediction errors observed for the different implementations. Note that implementations with augmented model redundancy involve the partitioning of the training set in 3 disjoint subsets. Although the training set comprises a rather large number of devices, results are quite sensitive to the repartition of training devices in the different subsets. Consequently, we create several random splits of the training devices in 3 partitions and we report minimum, mean and maximum values observed over all runs (100 runs in this experiment). Finally for the sake of comparison, we also report results obtained when using a reduced-cost implementation that exploits model redundancy based only on different training set partitions (a single *IM* combination is considered and only 3 redundant models are built using the 3 different partitions).

Table 4: Average and maximal prediction errors observed using the different implementations for the particular *CP1* prediction example of previous sub-section

		Initial Implem.	New Implementations			Cost-reduced Implem.
			Version A	Version B	Version C	
ϵ_{avg}	<i>min</i>	1.28 %	1.17 %	1.17 %	1.17 %	1.65 %
	<i>mean</i>		1.18 %	1.18 %	1.18 %	1.71 %
	<i>max</i>		1.19 %	1.19 %	1.20 %	1.76 %
ϵ_{max}	<i>min</i>	4.08 dB	0.77 dB	0.77 dB	0.78 dB	2.03 dB
	<i>mean</i>		0.81 dB	0.81 dB	0.82 dB	3.35 dB
	<i>max</i>		0.96 dB	0.96 dB	0.96 dB	5.09 dB

From the results of **Table 4**, it can be observed that the 3 versions of the new implementation of augmented model redundancy have equivalent performances. The efficiency of these new implementations is clearly demonstrated, since they permit to reduce average and maximal prediction errors compared to the initial implementation. The improvement is particularly significant regarding the maximal prediction error that reduces from *4.08dB* with the initial implementation down to less than *1dB*. Also it can be observed that the cost-reduced implementation that exploits model redundancy based only on different training set partitions offers degraded performances compared to other implementations.

Table 5: Comparison of the different implementations for *CP1* prediction

		Initial Implem.	New Implementations			Cost-reduced Implem.
			Version A	Version B	Version C	
ϵ_{avg}	<i>min</i>	0.69 %	0.66 %	0.66 %	0.66 %	0.64 %
	<i>mean</i>	0.97 %	0.96 %	0.96 %	0.97 %	1.11 %
	<i>max</i>	1.28 %	1.25 %	1.25 %	1.26 %	5.03 %
	σ	0.16 %	0.15 %	0.15 %	0.15 %	0.31 %
ϵ_{max}	<i>min</i>	0.70 dB	0.52 dB	0.53 dB	0.53 dB	0.66 dB
	<i>mean</i>	1.06 dB	0.91 dB	0.93 dB	0.95 dB	1.57 dB
	<i>max</i>	4.08 dB	2.16 dB	2.17 dB	2.16 dB	38.49 dB
	σ	0.42 dB	0.28 dB	0.29 dB	0.29 dB	0.78 dB

Table 6: Comparison of the different implementations for IP3 prediction

		Initial Implem.	New Implementations			Cost-reduced Implem.
			Version A	Version B	Version C	
ϵ_{avg}	<i>min</i>	0.87 %	0.85 %	0.85 %	0.85 %	0.91 %
	<i>mean</i>	1.06 %	1.06 %	1.06 %	1.06 %	1.17 %
	<i>max</i>	1.31 %	1.41 %	1.43 %	1.44 %	1.58 %
	σ	0.07 %	0.07 %	0.07 %	0.07 %	0.09 %
ϵ_{max}	<i>min</i>	1.01 dB	0.97 dB	0.97 dB	0.97 dB	1.17 dB
	<i>mean</i>	1.77 dB	1.51 dB	1.55 dB	1.59 dB	2.93 dB
	<i>max</i>	7.35 dB	4.47 dB	6.28 dB	6.50 dB	14.66 dB
	σ	0.96 dB	0.50 dB	0.53 dB	0.56 dB	1.74 dB

Results on the previous *CP1* prediction example are obtained considering 3 particular IM combinations. To further corroborate these results, we have conducted a large campaign of experiments varying the IM combinations used to build the redundant models. More precisely for each specification, we perform 50 random selections of 3 different IM combinations among all the combinations corresponding to models with a satisfying accuracy. For each 3 selected IM combinations, redundant models are then built considering 100 random splits of the training devices in 3 partitions. Implementations with augmented model redundancy are therefore evaluated over 5,000 different cases of generated redundant models. Results are summarized in

Table 5 and

Table 6 for *CP1* and *IP3* predictions respectively.

These results confirm the observations made on the particular *CP1* prediction example, and in particular the superiority of the new implementations with augmented model redundancy. Indeed although impact on the average prediction error is not significant, there is a substantial improvement regarding the maximal prediction error, with a reduction not only of the mean and maximum values observed over the different cases of generated redundant models, but also of the standard deviation. In particular, this standard deviation reduces from 0.42dB with the initial implementation down to 0.28dB for *CP1* specification, and from 0.96dB down to 0.5dBm for *IP3* specification. This constitutes an important improvement of the robustness of the technique. Regarding comparison between the performances offered by

the 3 versions of the new implementation, they are equivalent for *CP1* specification, but version “A” leads to slightly better results for *IP3* specification; this implementation may therefore be preferred. Regarding the cost-reduced implementation that exploit model redundancy based only on different training set partitions, this solution should not be retained as it presents degraded performances compared to other implementations.

Finally we have compared the different implementations regarding the percentage of circuits for which suspect predictions are identified. Results are summarized in **Fig. 46**. As expected, implementations with augmented model redundancy lead to a higher number of suspect predictions compared to implementations with model redundancy based only on either different *IM* combinations or different training set partitions, both for *CP1* and *IP3* specifications. However the percentage remains extremely low, less than 1.5%, which means that the new implementations permits to improve prediction results while maintaining a low test cost overhead.

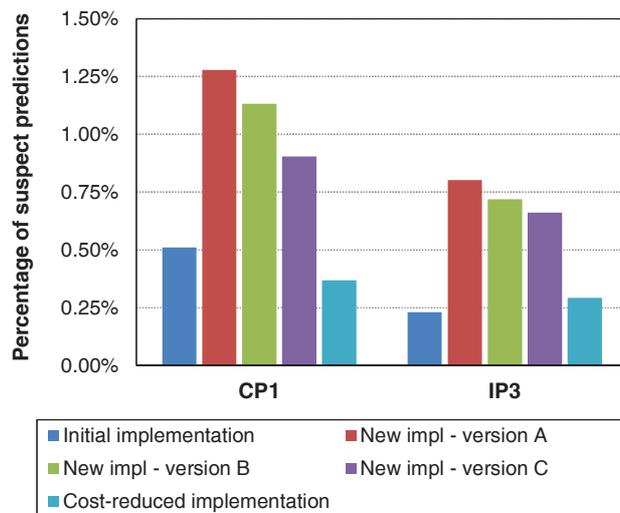


Fig. 46: Comparison of different implementations in terms of percentage of suspect predictions

IV.7 Conclusion

The main obstacle against the wide development of indirect test is the lack of confidence in the predicted *RF* performance values. As shown along this chapter, although circuit specifications are accurately predicted for most of tested devices, there are some devices for which prediction is affected by large errors; experiments show that these anomalies appear regardless of the used IM combination and the size of the training set, i.e. even though well correlated IMs and large TSS are used for model construction. Nevertheless, the percentage of these circuits increases when using models with low accuracy or reduced set of training devices.

To cope with these issues, we have proposed in this chapter a novel implementation of the prediction-oriented indirect test strategy. The idea is to introduce in the test procedure a kind of “safety” mechanism in order to prevent incorrect predictions which may cause test escape or yield loss. For this, we have exploited model redundancy and we have implemented a simple procedure that permits to identify devices for which prediction is prone to error. These devices are then removed from the indirect test tier and directed to a second tier where further testing may apply. In this way, we expect that most of the devices are evaluated through the low-cost indirect test tier and only a small fraction of devices are evaluated through a more expensive test procedure. As a result, the overall test cost is reduced compared to standard specification testing while accuracy is preserved.

This redundancy-based strategy has proven its efficiency along many experiments. Moreover it improves prediction accuracy and robustness in different contexts, such as building of models with training set of reduced size or low correlation IMs. It therefore permits to relax some important constraints faced by the test engineers for the efficient implementation of indirect test. Finally, we have proposed new implementations of the *CPC* procedure to improve its efficiency. These new implementations are based on redundant models that involve not only different combinations of indirect measurements as initially proposed, but also different partitions of the training set. Combining these two aspects permits the generation of a higher number of redundant models while maintaining the same cost, i.e. the same number of indirect measurements to be performed. All implementations redirect only a very small fraction of the devices to second tier, therefore incurring very low test cost overhead.

Conclusion

With the development of complex IC manufacturing processes and advanced IC integration schemes, it becomes increasingly challenging to produce operational chips. The quality assurance of the devices is becoming a very important criterion over the phases of the manufacturing process. Testing is of course an essential function to ensure product quality. In the context of analog and RF IC, the conventional practice is based on measuring the circuit performances and comparing them to specifications given by the datasheet. This specification-based test strategy, in spite of providing a golden test quality, has prohibitive cost. Two main factors contribute to this prohibitive cost: first the capital investment on the required *RF* test resources which could reach over than 60% of the total cost of an *ATE*; second the long test time needed to perform all *RF* measurements. In this context, testing of systems including *RF* parts has become a bottleneck to obtain low-cost products with high quality with respect the datasheet specifications and with reasonable time-to-market.

Some cost-reduced *RF IC* testing techniques have emerged in last decades. A quick overview on these techniques is given in the first chapter. One of the promising techniques is the indirect test strategy. This strategy can be applied at different levels in the manufacturing process; i.e. it can be adopted at wafer-level as well as package-level.

In this work, we have investigated the implementation of the indirect test strategy in the context of prediction-oriented test. This technique dramatically decreases the cost of the test solution, in particular when *DC* measurements are used to predict the circuit *RF* performances. In this case, there is no need for expensive *RF* equipment to perform the test. Moreover, the test time is considerably reduced. In addition because it gives access to the value of the device *RF* parameters, prediction-oriented indirect test also allows the industrials to perform additional post-silicon quality assurance operations, such as quality binning or process monitoring.

In spite of the clear advantages offered by this technique, the industrials have not yet entire confidence to widely adopt such a test strategy. The lack of confidence in the predicted value of *RF* parameters comes from different sources. As seen in the second and the third chapters, learning conditions and used *IMs* for the construction of a predictive model have a significant impact on the model accuracy. It also impacts the robustness of the model

regarding its ability to perform accurate prediction for new devices, and incorrect predictions observed for some devices might generate additional test escape or yield loss.

This thesis intends to improve confidence in the prediction-oriented indirect test strategy. Model accuracy and robustness are the key features to achieve trustable predicted values of the RF parameter. The model accuracy mainly depends on how much the used IMs express the RF parameter variation. Wise selection of the IMs is therefore required to build accurate models. In the literature, there are several methods for pertinent data selection. In chapter II, some of them have been investigated and compared to a reference solution based on an “exhaustive” search of the pertinent IMs subset. Due to evident computational issues; the search is restricted to subsets constituted only of 3 IMs . In spite of the limited space search, we have been able to build models with better accuracy than those built from IMs selected by the existing techniques. Moreover, thanks to the large number of prediction models built for each RF parameter, optimization of the cost of the test solution can be performed. So, clever IM subset selection can contribute to accurate model and more cost efficient test solutions. To this aim, we have proposed a method for pertinent IM subset selection based on three steps: (i) dimensionality reduction of the IM space based on PCA , (ii) iterative construction of the search space based on the selection of a limited number IM subsets at each iteration, and (iii) selection of an optimized IM subset through coverage problem resolution. The proposed method has been validated on the LNA test vehicle using simulation data; results show that predictive models with good accuracy can be built for the 12 RF performances using only 18 DC measurements selected among the 152 available IMs . Note that the selection of a cost-effective solution is a difficult task when dealing with a large number of RF parameters to be predicted and when IMs have different individual costs (i.e. the assumption that the test solution cost is directly proportional to the total number of used IMs is no longer valid). So, further work on this area could be the development of an optimization algorithm able to handle these aspects in the selection of a cost-effective solution.

In addition to the appropriate IMs selection for good model accuracy (i.e. low average prediction error), another challenge is to ensure high model robustness (i.e. low maximal prediction error). Unfortunately, good accuracy is achieved for most of the devices when using the conventional implementation of the prediction-oriented test strategy, but some devices exhibit large prediction errors. Even if the number of these devices is very small, this constitutes a real obstacle for wide development of the indirect test strategy. In this context, a novel implementation of the prediction-oriented indirect test strategy has been presented in

chapter IV. The key idea is to exploit model redundancy in order to establish prediction confidence for each tested device, which can then be used to define a robust test procedure. The proposed scheme is an adaptive test strategy based on two different tiers. All devices for which good prediction confidence is established are directed to the first tier, which corresponds to the indirect test procedure. In case of low prediction confidence, devices are removed from the indirect test procedure and directed to the second tier. The devices may be submitted to additional tests or simply discarded depending on the industrial strategy adopted for each product. Experiments presented in chapter IV using measurement test data have proven the efficiency of the proposed implementation to improve prediction accuracy and robustness against outliers. Moreover it also permits to get rid of the model performance dependency with the training data, and therefore obtain good efficiency even in case predictive models are built with training set of reduced size or low-correlation IMs. Note that the proportion of devices redirected to the second tier is very low (around 1%), which means that the large majority of devices are tested with the low-cost indirect test tier.

Another issue of the indirect test strategy, which is not addressed in this work, is the predictive model validity with respect to potential manufacturing process shift. Indeed it is likely that predictive models should be relearned in case of process shift for more efficient implementation. An interesting perspective is to investigate whether the number of circuits redirected to the second tier could constitute a good indicator for model relearning.

References

- [1] International Technology Roadmap for Semiconductors, "Radio Frequency and Analog/Mixed-Signal Technologies for Communications," 2011.
- [2] International Technology Roadmap for Semiconductors, "Test and Test Equipment," 2011.
- [3] International Technology Roadmap for Semiconductors, "Assembly and Packaging," 2011.
- [4] P. 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*, Vols. 21, Issue:3, pp. 349-361, Mar 2002.
- [5] S. Ellouz, P. Gamand, C. Kelma, B. Vandewiele and B. Allard, "Combining Internal Probing with Artificial Neural Networks for Optimal RFIC Testing," *IEEE International Test Conference (ITC)*, pp. 1-9, Oct. 2006.
- [6] S. Xiaoqin, V. Kerzérho and H. Kerkhoff, "Predicting dynamic specifications of ADCs with a low-quality digital input signal," *IEEE European Test Symposium (ETS)*, pp. 158-163, May 2010.
- [7] IBM, "Lowering Test Costs for RF ICs," IBM-Technical Note, Dec. 2002.
- [8] J. Ferrario, R. Wolf and S. Moss, "Architecting Millisecond Test Solutions for Wireless phone RF ICs," *IEEE International Test Conference (ITC)*, pp. 1151-1158, Oct. 2002.
- [9] J. Dabrowski and R. Ramzan, "Built-in Loopback Test for IC RF Transceivers," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 6, pp. 933-946, Jun. 2010.
- [10] M. Negreiros, L. Carro and A. Susin, "An Improved RF Loopback for Test Time Reduction," *Negreiros, M.; Carro, L.; Susin, A.A.; , "An Improved RF Loopback for Test Time ReduAutomation and Test in Europe Conference and Exhibition (DATE)*, pp. 1-6, Mar. 2006.
- [11] J. Dabrowski, "Lookback BiST for RF front-ends in digital transceivers," *International Symposium on System-on-Chip*, pp. 143-146, Nov. 2003.
- [12] J.-S. Yoon and W. Eisenstadt, "Embedded loopback test for RF ICs," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 5, pp. 1715-1720, Oct. 2005.
- [13] M. Onabajo, J. Silva-Martinez, F. Fernandez and E. Sanchez-Sinencio, "An On-Chip Loopback Block for RF Transceiver Built-In Test," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 56, no. 6, pp. 444-448, Jun. 2009.
- [14] H.-G. Stratigopoulos and Y. Makris, "Error Moderation in Low-Cost Machine-Learning-Based Analog/RF Testing," *IEEE Trans on Computer-Aided Design of Integrated Circuits and Systems*,

vol. 27, no. 2, pp. 339-351, Feb. 2008.

- [15] H.-G. Stratigopoulos, S. Mir, E. Acar and S. Ozev, "Defect Filter for Alternate RF Test," *IEEE European Test Symposium (ETS)*, pp. 101-106, May 2009.
- [16] H.-G. Stratigopoulos, P. Drineas, M. Slamani and Y. Makris, "Non-RF to RF Test Correlation Using Learning Machines: A Case Study," *IEEE VLSI Test Symposium (VTS)*, pp. 9-14, May 2007.
- [17] H.-G. Stratigopoulos and Y. Makris, "Nonlinear decision boundaries for testing analog circuits," *IEEE Trans on Computer-Aided Design of ICs and Systems*, vol. 24, no. 11, pp. 1760-1773, Nov. 2005.
- [18] H. Stratigopoulos, P. Drineas, M. Slamani and Y. Makris, "RF Specification Test Compaction Using Learning Machines," *IEEE Trans on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 6, pp. 998-1002, June 2010.
- [19] P. Variyam, S. Cherubal and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE trans on CAD of integrated circuits and systems*, vol. 21, no. 3, pp. 349-361, March 2002.
- [20] V. Stopjakova and et al., "Defect detection in analog and mixed circuits by neural networks using wavelet analysis," *IEEE transactions on reliability*, vol. 45, no. 3, pp. 441-448, September 2005.
- [21] R. Voorakaranam, S. S. Akbay, S. Bhattacharya, S. Cherubal and A. Chatterjee, "Signature Testing of Analog and RF Circuits: Algorithms and Methodology," *IEEE Trans on Circuits and Systems*, Vols. 54, Issue: 5, pp. 1018-1031, May 2007.
- [22] S. R. a. S. R. Nassif, "Accurate spatial estimation and decomposition techniques for variability characterization," *IEEE Trans. Semiconductor Manufacturing*, vol. 23, no. 3, pp. 345-357, Aug. 2010.
- [23] T. Liu and A. IBM, "A general framework for spatial correlation modeling in VLSI design," *IEEE Design Automation Conference (DAC)*, pp. 817-822, Jun. 2007.
- [24] H. Ke, H. Ke, C. J. John M. and M. Yiorgos, "On Combining Alternate Test with Spatial Correlation Modeling in Analog/RF ICs," *IEEE European Test Symposium (ETS)*, pp. 64-69, May 2013.
- [25] P. Variyam and A. Chatterjee, "Enhancing test effectiveness for analog circuits using synthesized measurements," *IEEE VLSI Test Symposium*, pp. 132-137, Apr 1998.
- [26] D. Han and A. Chatterjee, "Robust Built-In Test of RF ICs Using Envelope Detectors," *IEEE Asian Test Symposium (ATS)*, pp. 2-7, Dec 2005.
- [27] H. Friiz, "Noise Figures of Radio Receivers," *IEEE Proceedings of the IRE*, Vols. 33, Issue: 2, pp. 125-127, Feb. 1945.

- [28] Agilent, *Fundamentals of RF and Microwave Noise Figure Measurements*.
- [29] B. S. Keith and K. Joe, Production testing of RF and system-on-a-chip devices for wireless communications, ARTECH HOUSE, 2004.
- [30] J. Gints, "Gints Jekabsons' webpage," 2008. [Online]. Available: <http://www.cs.rtu.lv/jekabsons/regression.html>. [Accessed 2011].
- [31] F. Jerome H., "Multivariate Adaptive Regression Splines," *The annals of statistics*, vol. 19, no. 1, pp. 1-141, 1991.
- [32] J. QUINLAN, Induction of Decision Trees - Machine Learning 1, Boston: Kluwer Academic Publishers, 1986, pp. 81-106.
- [33] M. M.D., B. R.J. and C. W.J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *TECHNOMETRICS*, vol. 21, no. 2, pp. 239-245, May 1979.
- [34] R. Iman, D. J.M. and D. Zeigler, Latin hypercube sampling (program user's guide), Sandia Laboratories, 1980.
- [35] S. Bhattacharya and A. Chatterjee, "A DFT Approach for Testing Embedded Systems Using DC Sensors," *Design & Test of Computers, IEEE*, vol. 23, no. 6, pp. 464-475, June 2006.
- [36] R.Voorakaranam, S. Akbay, S. Bhattacharya, S. Cherubal and A. Chatterjee, "Signature testing of analog and RF circuits: algorithms and methodology," *IEEE trans on circuits and systems*, vol. 54, no. 5, pp. 1018-1031 , May 2007.
- [37] P. Langley, "Selection of relevant features in machine learning," *Proceedings of the AAAI Fall Symposium on Relevance*, 1994.
- [38] K. K. and R. L. A., "A practical approach to feature selection," *Machine Learning: Proceedings of the Ninth International Conference*, 1992.
- [39] K. I., "Estimating attributes: Analysis and extensions of relief," *Proceedings of the European Conference on Machine Learning*, 1994.
- [40] K. I., "On biases in estimating multi-valued attributes.," *IJCAI*, pp. 1034-1040, 1995.
- [41] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information : criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol. 27, no. 8, pp. 1226-1238, Aug. 2005.
- [42] H. e. a. Ayari, "Smart selection of indirect parameters for DC-based alternate RF IC testing," *IEEE, VLSI Test Symposium (VTS)*, vol. Apr., pp. 19-24, Apr. 2012.

- [43] H. Ayari, F. Azais, S. Bernard, M. Comte, V. Kerzerho and M. Renovell, "On the Use of Redundancy to Reduce Prediction Error in Alternate Analog/RF Test," *IEEE, Mixed-Signals, Sensors and Systems Test Workshop (IMS3TW)*, pp. 34-39, 2012.
- [44] H. Ayari, F. Azais, S. Bernard, M. Comte, V. Kerzerho and M. Renovell, "Making predictive analog/RF alternate test strategy independent of training set size," *IEEE International Test Conference (ITC)*, pp. 1-9, Nov. 20012.
- [45] H.-G. Stratigopoulos and Y. Makris, "Nonlinear decision boundaries for testing analog circuits," *IEEE Trans on Computer-Aided Design of ICs and Systems*, vol. 24, no. 11, pp. 1760-1773, Nov. 2005.
- [46] J. Kasten, C. S. Corp. and B. Kaminska, "An introduction to RF Testing: Device, method and system," *IEEE VLSI Test Symposium (VTS)*, pp. 462 - 468, Apr 1998.
- [47] S. Abdennadher, C. Oregon State Univ. and S. Shaikh, "Practices in mixed-signal and RF IC testing," *IEEE, Design & Test of Computers*, vol. 24, no. 4, pp. 332 - 339, Aug. 2007.
- [48] P. Variyam, G. I. o. T. A. G. U. Sch. of Electr. & Comput. Eng. and A. Chatterjee, "Enhancing test effectiveness for analog circuits using synthesized measurements," *IEEE, VLSI Test Symposium*, pp. 132 - 137, Apr. 1998.
- [49] P. Variyam and A. Chatterjee, "Specification driven test generation for analog circuits," *IEEE Trans., Computer-Aided Desig*, vol. 19, no. 10, p. 1189-1201, Oct. 2000.
- [50] H.-G. Stratigopoulos, S. Mir and Y. Makris, "Enrichment of limited training sets in machine-learning-based analog/RF test," *IEEE Design, Automation & Test in Europe (DATE)*, pp. 1668-1673, Apr. 2009.
- [51] K.-T. C. W. Z. X. L. a. K. B. H.-M. Chang, "Test cost reduction through performance prediction using virtual probe," *IEEE International Test Conference (ITC)*, pp. 1-9, Sep. 2011.
- [52] S. Bhattacharya and A. Chatterjee, "A DFT Approach for Testing Embedded Systems Using DC Sensors," *IEEE Design & Test of Computers*, vol. 23, no. 6, pp. 464-475, Jun. 2006.
- [53] International Technology Roadmap for Semiconductors, "Executive Summary," 2011.
- [54] H.-G. Stratigopoulos, S. Mir, E. Acar and S. Ozev, "Defect Filter for Alternate RF Test," *IEEE European Test Symposium*, pp. 101-106, May 2009.

Related publications

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “Making predictive analog/RF alternate test strategy independent of training set size” IEEE International Test Conference (ITC), 2012

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “Smart selection of indirect parameters for DC-based alternate RF IC testing” IEEE VLSI Test Symposium (VTS), 2012

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “On the Use of Redundancy to Reduce Prediction Error in Alternate Analog/RF Test”, IEEE International Mixed-Signals, Sensors and Systems Test Workshop (IMS3TW), 2012 (*Best Paper Award*)

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “Implementing model redundancy in predictive alternate test to improve test confidence” IEEE European Test Conference, 2013 (*Poster*)

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “Enhancing Confidence in Indirect Analog/RF Testing against the Lack of Correlation between Regular Parameters and Indirect Measurements” Microelectronics Journal 2014.

Ayari, H; Azais, F. ; Bernard, S. ; Comte, M. ; Kerzerho, V. ; Renovell, M. “New implementations of predictive alternate analog/RF test with augmented model redundancy “ Design, Automation & Test in Europe (DATE) March 2014

List of Figures

Fig. 1: Specification-based test strategy	2
Fig. 2: Test cost versus test time.	8
Fig. 3: ATE cost increase with additional added features.....	8
Fig. 4: Wafer-level RF IC test (RF prob).....	10
Fig. 5: Principle of integrated test solution	11
Fig. 6: Principle of the loopback testing	12
Fig. 7 Underlying idea of indirect testing.	13
Fig. 8: Classification oriented Indirect Testing.....	14
Fig. 9: Prediction oriented Indirect Testing	14
Fig. 10: Wafer level test cost reduction based on spatial correlation.....	15
Fig. 11 Indirect Test: classical implementation	19
Fig. 12: Block diagram of the tuner	21
Fig. 13: Block diagram of the PA.....	22
Fig. 14: SNR degradation of a signal passing through a semiconductor device.....	24
Fig. 15: Definition of the 1 -dB Compression Point	25
Fig. 16: Definition of the 3 rd order intercept point.....	25
Fig. 17: In-band Intermodulation Measurement	26
Fig. 18: Single Neuron	29
Fig. 19: The regression tree.....	30
Fig. 20: Statistical properties of Training and Validation subsets regarding CP1 specification.....	31
Fig. 21: Comparison of the correlation plots obtained with 4 different regression algorithms.....	33
Fig. 22: Example of aberrant prediction for one particular device	34
Fig. 23: MSE variation over models built with different IM combinations for different regression algorithms	35
Fig. 24: Illustration of prediction error for different sizes of TS and VS	37
Fig. 25: Illustration of prediction error for different regression models (with same training and validation subsets)	39
Fig. 26: $CP1$ prediction for two different TSS	40
Fig. 27 Influence of Training Set Size (TSS) on prediction accuracy for $CP1$ specification.....	41
Fig. 28: Iterative search of IM subsets for specification prediction.....	50
Fig. 29: Comparison of six different feature selection methods based on achieved model accuracy.....	52

Fig. 30: Comparison of the Best* solution computed in the space of 7-IM subsets and the optimum solution in the reduced space of 3-IM subsets.....	54
Fig. 31: Accuracy the solutions given by Best*, Optim* and REF selection strategies	57
Fig. 32: Cost of the solutions given by Best*, Optim* and REF selection strategies.....	58
Fig. 33: Flow diagram of the proposed two-tier indirect test scheme using model redundancy.....	64
Fig. 34: Correlation plots for CP1 and IP3 predictions using 3 different models and after removal of suspect predictions (models built on 5,000 training devices).....	66
Fig. 35: PA performance prediction with model redundancy: impact of the threshold value $\epsilon_{CP1/IP3}$ used during identification of suspect predictions	68
Fig. 36 : Average prediction error of selected redundant models vs. TSS.....	69
Fig. 37: Correlation plots for CP1 and IP3 predictions using 3 different models and after removal of suspect predictions (models built on 300 training devices).....	71
Fig. 38: Average and maximal prediction errors vs. TSS	72
Fig. 39: Percentage of suspect predictions vs. TSS	73
Fig. 40: Distribution of average prediction error (computed on <i>TS</i>) for the exhaustive set of predictive models built using triplets of IMs.....	74
Fig. 41: Average prediction error observed on <i>VS</i> over the 5,000 runs	75
Fig. 42: Maximum prediction error observed on <i>VS</i> over the 5,000 runs	75
Fig. 43: Particular CP1 prediction example illustrating the limitation of the proposed implementation....	77
Fig. 44: Generation of redundant regression models	79
Fig. 45: Implementations of the two-tier indirect test scheme with augmented model redundancy.....	81
Fig. 46: Comparison of different implementations in terms of percentage of suspect predictions.....	84

List of Tables

Table 1: Average accuracy over the RF performances for the six feature selection methods and combined strategy according to specification to be predicted.....	53
Table 2: Average and maximal prediction errors for CP1 and IP3 specifications	65
Table 3: Percentage of suspect predictions: average and maximum values over the 5,000 runs	76
Table 4: Average and maximal prediction errors observed using the different implementations for the particular CP1 prediction example of previous sub-section.....	82
Table 5: Comparison of the different implementations for CP1 prediction	82
Table 6: Comparison of the different implementations for IP3 prediction.....	83

List of acronyms

acc to	According to
ATE	Automatic Test Equipment
BIST	Built in Self Test
CFS	Correlation-based Feature Selector
CP1	1-dB Compression Point
CPC	Check Prediction Consistency
dB	Decibel
DC	Direct Current
DCT	Discreet Cosine Transform
DFT	Design for Test
DUT	Device Under Test
ENR	Excess Noise Ratio
GP	Gaussian Process
IC	Integrated Circuit
IF	Intermediate Frequency
IIP₃	Input third Order Intercept Point
IM	Indirect Measurement
IP₃	Third Order Intercept Point
KGD	Known Good Die
LNA	Low Noise Amplifier
LHS	Latin Hypercube Sampling
MARS	Multivariate Adaptive Regression Splines
MiM	Metal insulator Metal
MLR	Multiple Linear Regression
mRMR	minimum Redundancy Maximum Relevance
MSE	Mean Squared Error
NF	Noise Figure
ANN	Artificial Neural Networks
OIP₃	Output third Order Intercept Point
OOR	Out Of Range
PA	Power Amplifier
PC	Principle Component
PCA	Principle Component Analysis
PPM	Part Per Million
PWL	Piece Wise Linear
RF	Radio Frequency
RMSNE	Root Mean Squared Normalized Error
SiP	System in Package
SoC	System on Chip
SNR	Signal to Noise Ratio
TS	Training Set

TSS	Training Set Size
VP	Virtual Probe
VS	Validation Set
VSS	Validation Set Size

