Solutions pour l'auto-adaptation des systèmes sans fil

Martin Andraud

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Nano Electronique et Nano Technologies**

Arrêté ministériel : 7 Août 2006

Présentée par

Martin ANDRAUD

Thèse co-dirigée par **Emmanuel Simeu et Haralampos Stratigopoulos**

préparée au sein du **laboratoire des Technologies de l'Informatique et de la Microélectronique pour l'Architecture des systèmes intégrés** et de **l'Ecole Doctorale Electronique Electrotechnique Automatique et Traitement du Signal**

Solutions pour l'auto-adaptation des systèmes de communication sans fil

Thèse soutenue publiquement le **14 Juin 2016**,
devant le jury composé de :

Lars Hedrich
Professor, Goethe-Universitat Frankfurt am Main, Germany, Rapporteur
Josep Altet
Associate Professor, Universitat Politecnica de Catalunya Barcelona, Spain, Rapporteur
Ian O'Connor
Professeur, Ecole Centrale de Lyon, Examinateur
Yann Deval
Professeur, IMS Bordeaux, Examinateur
Emmanuel Simeu
Maître de conférences, TIMA Grenoble, Co-Directeur de thèse
Haralampos Stratigopoulos
Chercheur CNRS, LIP6 Paris, Co-Directeur de thèse
Marian Verhelst
Associate Professor, Katholieke Universiteit Leuven, Belgium, Invitée
Serge Bernard
Chercheur CNRS, LIRMM Montpellier, Invité

ABSTRACT

SOLUTIONS FOR THE SELF-ADAPTATION OF WIRELESS SYSTEMS

The current demand on ubiquitous connectivity imposes stringent requirements on the fabrication of Radio-Frequency (RF) circuits. Designs are consequently transferred to the most advanced CMOS technologies that were initially introduced to improve digital performance. In addition, as technology scales down, RF circuits are more and more susceptible to a lot of variations during their lifetime, as manufacturing process variability, temperature, environmental conditions, aging, etc. As a result, the usual worst-case circuit design is leading to sub-optimal conditions for the circuit, in terms of power and/or performances. In order to counteract these variations, to increase the performances and reduce the power consumption, adaptation strategies must be put in place. More importantly, the fabrication process introduces more and more performance variability, which can have a dramatic impact on the fabrication yield. That is why RF designs are not easily fabricated in the most advanced CMOS technologies, as 32nm or 22nm nodes for instance. In this context, the performances of RF circuits need to be calibrated after fabrication so as to take these variations into account and recover yield loss. This thesis work is presenting on a post-fabrication calibration technique for RF circuits. This technique is performed during production test with minimum extra cost, which is critical since the cost of test can be comparable to the cost of fabrication concerning RF circuits and cannot be further increased. Calibration is enabled by equipping the circuit with tuning knobs and sensors. Optimal tuning knob identification is achieved in one-shot based on a single test step that involves measuring the sensor outputs once. For this purpose, we rely on variation-aware sensors which provide measurements that remain invariant under tuning knob changes. As an auxiliary benefit, the variation-aware sensors are non-intrusive and totally transparent to the circuit. Our proposed methodology has first been demonstrated with simulation data on an RF power amplifier as a case study. Afterwards, a silicon demonstrator has then been fabricated in 65nm CMOS technology in order to fully demonstrate the methodology. The fabricated dataset of circuits is extracted from one typical and two corner wafers. This feature is very important since corner circuits are the worst design cases and therefore the most difficult to calibrate. In our case, corner circuits represent more than the two third of the overall dataset and the calibration can still be proven. In details, fabrication yield based on 3 sigma performance specifications is increased from 21% to 93%. This is a major performance of the technique, knowing that worst case circuits are very rare in industrial fabrication.

ACKNOWLEDGMENTS

First of all, I want to think about all the people without who this work would not have been possible. In particular, I sincerely thank my advisors, Emmanuel Simeu and Haralampos Stratigopoulos for their support, their very useful help and their advices that were always appropriate. Then, I thank Salvador Mir for hosting me in the RMS team, and for his support during these years. Generally, I want to acknowledge all people of TIMA Laboratory (administrative staff, IT service...) for the work that they accomplish every day.

These words are also for all the people that helped me during this Ph.D work (and I hope that I will not forget anyone). Thanks to Alexandre Giry and Alexandre Siligaris from cea-leti for his very helpful support and for the discussions we had about this chip that caused us a lot of work, thanks to Skandar Basrour for the test room facilities, thanks to CIME (Alexandre, Loic, Aurélien) for their help with testing platforms, thanks to the RF team of cea-leti, especially Alexandre Siligaris, for their help with the chip layout. This work would have also be significantly worse without all my colleagues from RMS team (Guillaume, Thanasis, Asma, Rshdee, Hani, Brice, Laurent, Diane, Miroslav, Louay...) and from the others (Alban, Simon, Franois, Quentin, Emilie, Adrian...) and also to the TIMA football team.

At the end, last but not least, I want also to have a word for my families and friends, that were here at all times. It is a chance for me to have them by my side.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

GENERAL INTRODUCTION

1.1 CONTEXT

Mobile phones, computers, and in general every connected objects, play a constantly increasing role in our everyday life. A chief reason of the pensiveness of this technology is the evolution of wireless Radio Frequency (RF) transceivers and communication protocols. Consumers always demand more from technology, and the industry push hardware limits further to continuously increase the capabilities of wireless communication.
This current demand for ubiquitous connectivity imposes stringent requirements on the functionality of RF transceivers. Designers are charged with providing RF transceivers that can exchange higher volumes of data at higher speeds while consuming less power. Towards this objective, RF designs are transferred to more advanced CMOS technologies that are originally introduced to improve digital performance. In fact, to reduce the form factor, power consumption, and fabrication cost, the trend nowadays is towards Systems-on-Chip (SoC) where RF transceivers are integrated together with the digital processor and memory onto the same die. However, developing SoC with many heterogeneous circuits is a very difficult task.

TECHNOLOGY SCALING AND ITS CONSEQUENCES On one hand, although digital performance has often largely benefited from technology scaling, RF designs at 65nm and below start becoming very susceptible to process variations to the point where there it can have a dramatic impact on fabrication yield. For this reason, RF transceivers are not easily fabricated in the most advanced 45 nm, 32 nm, and 22 nm technologies. In 2011, the International Technology Roadmap for Semiconductors (ITRS) identified manufacturability issues as main challenges in the design of future circuits [10]. Critical aspects highlighted in the report are among others, the device parameter variability and performance/power variability. Moreover, it is also explained that in the case of analog and RF devices, solving these issues is even more challenging than for digital devices.

Technology Node	250 nm	180 nm	130 nm	90 nm	65 nm	45 nm
Threshold voltage variability (%)	4.7	5.8	8.2	9.3	10.7	16

Table 1: From [1] List of threshold voltage variability for several technology nodes

Indeed, the variability of device parameters (like threshold voltage, resistivity or fluctuations in the physical dimensions) is constantly increasing as technology nodes decrease, which complexes even more the design of analog and RF circuits. To illustrate this issue, Table 1 shows the intra-die variability of the threshold voltage V_{TH} for a typical MOS transistor,

for several technology nodes. The variability is calculated by dividing the typical standard deviation of the threshold voltage σV_{TH} measured on different circuits, by its mean value μV_{TH}. It can be clearly seen that threashold voltage variability increases when the technology node is reduced, starting from 4.7% at 180nm to increase up to 16% for 45nm node. This increased variability will have an impact on the circuits' performances since block-level and system-level specifications are affected as well as individual transistors themselves [1].

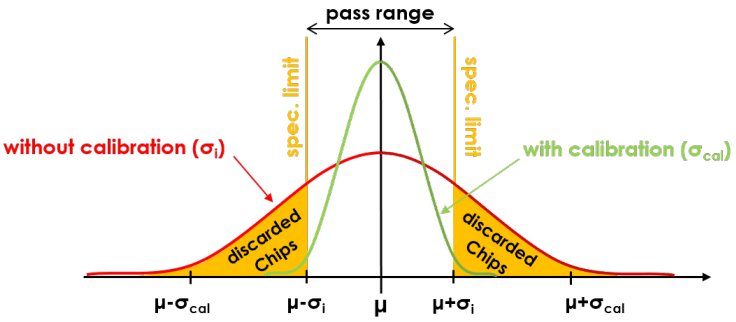

Figure 1: From [1]: Typical distribution of a circuit's performance with and without calibration

For instance, typical gaussian distributions of a performance is plotted in Fig.1, with the mean performance value denoted μ and the standard deviation denoted σ. Initial performance distribution, without any calibration mechanism, is plotted in red. It is first important to note that if variability is increased (i.e. σ becomes more important), the corresponding performance distribution will be widened. By fixing the specification limits for this performance, both the number of chips that are discarded, i.e. that do not satisfy the specifications (under the yellow area in Fig.1), as well as the corresponding fabrication yield, are fixed. As the demand for more performances on the circuit is leading to tighter specification limits, the yield is decreased at an expense of high cost for the manufacturer since all the chips under the yellow area have to be thrown away. Evidently, this is a show-stopper in the pursuit of ubiquitous connectivity and for the evolution of SoC. That is why, already for the 65 nm technology, elaborate ad-hoc calibration can be required to compensate for these large variations, and increase the fabrication yield. By using calibration techniques, the number of discarded chips can be largely reduced, with the same specification levels as illustrated in green color in Fig.1. But once again, this calibration may come with a cost in terms of area, power, etc that must be considered.

MODERATION OF THE POWER CONSUMPTION On the other hand, power consumption is a increasingly constraint in the design of electronic systems, especially for RF circuits that are responsible for a large fraction of the power consumption of wireless communication systems. As Illustrated in Fig.2 for a typical RF transmitter, the RF part of the circuit that transmits the power (which corresponds to the Power Amplifier (PA) and the transmitted power itself) represents almost 50% of the total transmitter power consumption. The remaining power is consumed by the front-end part for 20%, the Digital Signal Processor (DSP) together with Medium Access Control (MAC) protocol for 25% and Digital to Analog Conversion (DAC) for the remaining 6%.

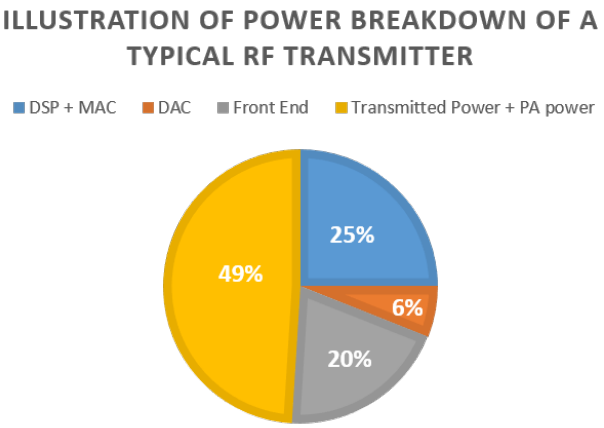

Figure 2: From [2]: Repartition of the power consumption for a typical transmitter used for WiFi communication (Intel Prism-II)

OTHER TYPES OF VARIATION In addition circuits' functionalities and performances can degrade over time in the field during their lifetime, due to temperature, aging, and environmental variations for instance. As an example of environmental variations, we can consider the communication in the RF channel that constantly changes, having an impact on the communication quality. Susceptibility to aforementioned variations is more or less important depending on the target device, but in mission-critical or safety-critical applications, some or all these variations can quickly become a major concern.

A NEED FOR SOLUTIONS Several solutions can then be envisaged, especially at system level. For example, process variations could be moderated thanks to ad-hoc calibration procedures, in order to recover yield loss and enable the fabrication of circuits in the most advanced technologies. DC power consumption can be brought under control by adapting the circuit to its workload so as to consume only the minimum required power to achieve a target task. The other types of variations can also be tackled by adapting the circuit with dedicated mechanisms.

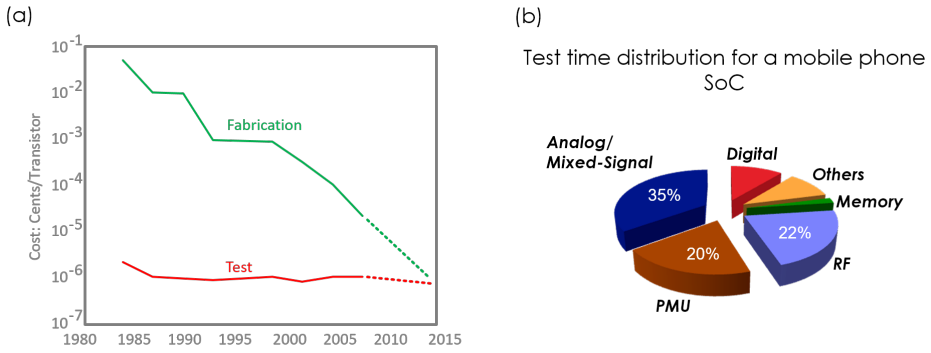

Figure 3: (a) From [3] fabrication cost versus test cost - (b) From [4]: test time of different parts of a SoC

TESTING THE RF PERFORMANCES In this context, the test of RF performances becomes even more important. Indeed, even if the design phase is still critical to ensure that a RF circuit will meet stringent specifications, efficient test is needed to ensure that the chips sent to the customers are functional. This test must (a) verify whether the circuit is functional or not with a high accuracy and a minimal error (ideally a few part-per-million), but also (b) must be performed as fast as possible using a low-cost of test equipment to minimize test time and cost. Regarding test cost, the semiconductor research association has published data that clearly show that, although fabrication cost is almost constantly decreasing from one technology generation to the next, the cost of test has remained almost constant since 1985 [3]. This is illustrated in Fig.3(a). Test cost reduction is therefore an intensive research field, especially for analog and RF circuits, that can represent more than 50% of the total test time of a mobile phone SoC, as shown Fig.3(b) [4].

Moreover, specifically considering self-adaptive circuits, the test might be performed on-chip by the circuit itself, which adds serious constraints. Those constraints can for instance be defined in terms of area overhead, power consumption overhead, and added computational cost so as to perform the entire adaptive procedure inside the chip. Consequently, adapting the RF performances cannot be done at an expense of prohibitive area, power, and/or computing overheads. A significant amount of research work has been carried out to develop many different calibration and/or adaptation strategies, but only limited or ad-hoc procedures have usually been considered. As a result, there is a need to develop more global adaptation strategies, able to simultaneously handle different types of variation, which is one of the main objective of this work.

1.2 OBJECTIVES AND CONTRIBUTIONS OF THIS WORK

The general subject of these thesis is to rethink about solutions to adapt the RF circuits, regarding the multitude of variations that can affect them. As highlighted in the previous paragraphs, process variations are a major concern since they affect all fabricated circuits, especially RF circuits. In addition, the circuits are more and more affected by these variations as technology scales down.

The first issue that we then decided to tackle is the compensation for process variations. In this work, compensation for process variations will be called *calibration*. Many challenges have to be overcome to obtain an efficient and low-cost calibration strategy, especially for RF circuits, where the test cost is a major concern. Among the concepts presented in the literature, *alternate test* has proven its feasibility, by performing low-cost measurements (that are DC or low-frequency) and by predicting the performances of an RF circuit based on machine-learning algorithms. Based on an existing alternate test paradigm, this work aims at extending the current alternate test methodologies to perform a post-manufacturing calibration in "one-shot", which means that there is a unique test/tune step to calibrate the RF circuit under test. Indeed, even if alternate test is a low-cost test technique, it is not always sufficient to enable cost-effective calibration algorithm. The one-shot feature is a key aspect of this work to further reduce the cost of calibration.

Moreover, the developed technique relies on non-intrusive variation-aware sensors that offer an image of the process variations encountered by the RF circuit. This key property offers

an advantage compared to other techniques since (a) sensors are not electrically connected to the circuit and are totally transparent to it, then the performances of the circuit are not degraded and (b) thanks to this non-intrusive property, the measurements and the tuning settings of the RF circuit becomes totally independent of each other, which enables a true one-shot calibration procedure.
As a first case study, an RF power amplifier has been designed to prove the concept with simulation data. Neural networks are trained so as to be able to predict four main performances of this power amplifier, depending on the process variations (measured by the non-intrusive sensors) and the tuning knobs used for calibration (the bias and power supply voltages). A silicon demonstrator has then been built to fully prove the concept, in 65nm CMOS technology provided by STmicroelectronics. Results show that the calibration methodology is able to calibrate even the chips that are fabricated in the most extreme process cases as it will be proven along this manuscript.

1.3 STRUCTURE OF THIS MANUSCRIPT

This manuscript is divided in five different chapters, the first being this general introduction.

In the second chapter, an introduction to the RF circuits and their adaptation principle will be given. The aim of this introduction is to have a general overview of how a basic RF transceiver is built, which are its main components and performances, and which associated metrics are used to evaluate them.

In the third chapter, the state-of-the-art will be presented, divided in three main parts. First of all, advances in terms of RF circuit testing, which is the basis for the development of adaptive circuits, will be reviewed. Then, previous calibration (i.e. compensation of process variations) techniques will be detailed the research domain and place our work in perspective with other previous works. Finally, adaptive RF circuits with dedicated algorithms will be the object of the last section, in order to understand the needs, requirements, and the strategies that must be developed so as to compensate for other types of variations.

In the fourth chapter, we will go into the details of our proposed one-shot calibration methodology. Basic principles of the methodology will first be explained, along with the key points of the general strategy. The calibration framework and the non-intrusive sensing strategy will be explained in details, and a proof-of-concept with simulation data will validate our initial objectives.

In the fifth chapter, the design of a complete RF CMOS silicon chip will first be detailed, in terms of layout, and test board design. Then, results will be presented in order to fully prove the calibration methodology.

Finally, a sixth chapter will concludes and draw the perspective of future works, especially in terms of self-adaptive RF circuits.

2

RF TRANSCEIVERS AND THEIR ADAPTATION PRINCIPLES

The design of RF transceivers is subject to many constraints, among others respecting wireless communication norms (center frequency, channel, safety, etc), while obtaining high performance, high reliability and low power consumption, which is especially for mobile communication devices. This chapter is devoted to give a general view of how RF transceiver design is trying to comply with all these constraints.

To this end, section 2.1 will first give an overview of a basic RF transceiver architectures and different data modulation schemes, focusing on the most recent techniques. In section 2.2, the main performances and figures of merit of RF circuits will be presented. An introduction to power amplifier design will be presented section 2.3, since power amplifiers will be the focus of this manuscript. Section 2.4 will provide details about the main parameters that have to be compensated and/or adapted, in the case of RF transceivers. Finally, section 2.5 will give a general overview of current adaptation techniques that have been put in place to adapt RF transceivers in view of different types of variation.

2.1 RF TRANSCEIVERS: BASIC CONCEPT AND OPERATION

2.1.1 *Basic transceiver topology*

A basic RF transceiver topology is shown Fig.4. It is composed of a transmitter and a receiver. On the transmitter's side, the baseband signal is first encoded, following a protocol specific to the modulation technique (some of them are detailed in section 2.1.2). After that, the signal is up-converted through a mixer. The reference signal of the mixer, at the operating frequency of the system, is usually provided by an on-chip Local Oscillator (LO). The signal is then amplified by a Power Amplifier (PA) so as to be transmitted through the antenna. On the receiver's side, the very weak RF signal received through the antenna needs first to be amplified by a Low-Noise Amplifier (LNA), and is then down-converted to baseband, through a mixer fed by the LO and demodulated following a specific coding protocol. A lot of different topologies can be derived from this basic scheme, guided by the constraints inherent to specific transceiver design, such as modulation coding complexity, cost, power consumption, noise immunity, etc. Some of these constraints are critical and are detailed in section 2.2.

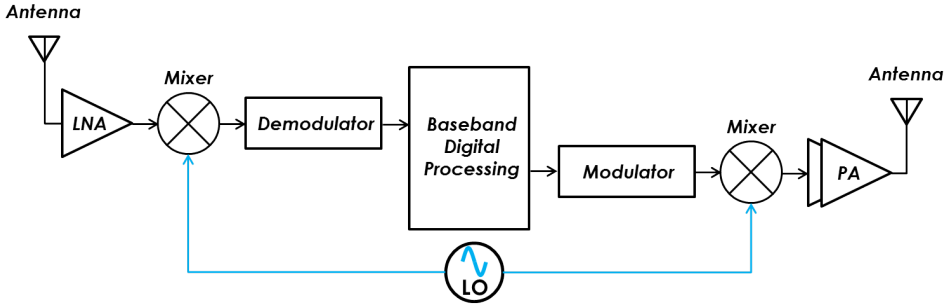

Figure 4: Basic topology of an RF transceiver

2.1.2 *Overview of coding and modulation schemes*

Each communication protocol is based on specific coding and modulation schemes for the RF signal. This section is dedicated to give an overview of the most used modulation schemes, focusing on digital modulation that is dominant in today's RF transceivers.

2.1.2.1 *Basic digital modulation schemes*

The general aim of modulation is to encode a baseband signal and transmit it through an RF channel by employing an RF carrier signal (a sinusoid for example) at the communicating frequency. Considering that an analog signal is created with three main parameters (frequency, amplitude, phase), the signal to be transmitted can be encoded into the amplitude, frequency or phase of the RF carrier. We refer to Amplitude Modulation (AM), Frequency Modulation (FM) or Phase Modulation (PM), respectively. Modern RF transceivers usually use digital signals to be modulated, instead of analog signals, and they will be the focus of this section. A digital signal ('0' or '1') can also be encoded into amplitude, frequency or phase of the carrier to obtain, by analogy with AM, FM and PM, Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), or Phase Shift Keying (PSK), as illustrated respectively in Fig.5 (a), (b) and (c).

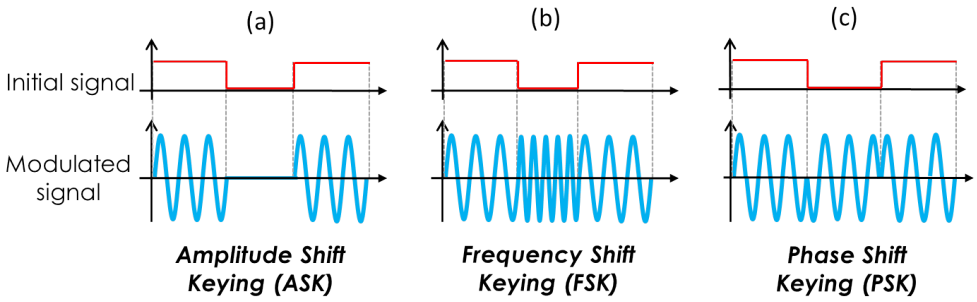

Figure 5: Illustration of basic digital modulation schemes: (a) amplitude shift keying, (b) frequency shift keying and (c) phase shift keying

In practice, the most used modulation schemes are FSK and PSK [11]. In order to obtain more reliable or more power-efficient modulation schemes, specific values of frequency or phase have to be chosen. In this objective, Binary FSK (BFSK) and Binary (BPSK) are the most suitable candidates.
Considering BPSK, two opposite phases of the RF carrier are chosen to modulate the digital baseband signal. Looking more precisely into the modulated signal, and considering that the carrier frequency has a sinusoidal form given by $x_{carrier} = A_c.cos(\omega_c.t)$, we can write the modulated BPSK signal as:

$$x_{BPSK} = A_c.cos(\omega_c.t + \phi) = (A_c.cos\phi).cos(\omega_c.t), \tag{1}$$

where A_c is the amplitude of the RF signal, ω_c is the carrier frequency and ϕ is the modulated phase, equal to 0°C or 180°C (π in radians).

Considering BFSK, two frequency signals are chosen, with equal amplitudes, to modulate the digital baseband signal. With the same RF carrier, BFSK modulated signal is given by:

$$x_{BFSK} = \alpha_1.cos(\omega_1.t) + \alpha_2.cos(\omega_2.t), \tag{2}$$

where [α_1 α_2] = [A_c 0] or [0 A_c] depending on the transmitted symbol.

A more general approach is to look at modulated signals in constellations. This visualization scheme is particularly appropriate for more advanced quadrature modulation schemes, that will be seen in the next section. In constellation diagrams, modulated waveforms are visualized in terms of the coefficients of their inner products. Examples are presented for BFSK and BPSK, respectively in Fig.6 (a) and (b).

Considering BFSK, the vector [α_1 α_2] can ideally take only two values, [A_c 0] or [0 A_c], as presented Fig.6(a). Considering BPSK, the phase of the RF carrier is modified, which consists in having two opposite cosine functions with amplitudes $\pm A_c$ in the constellation diagram Fig.6(b).
However, ideal constellation values are altered by noise, interferences, etc. and move from their initial location. A decision boundary is usually put to distinguish one symbol from the other. For example, the boundary applied to BFSK is illustrated Fig.6(c).

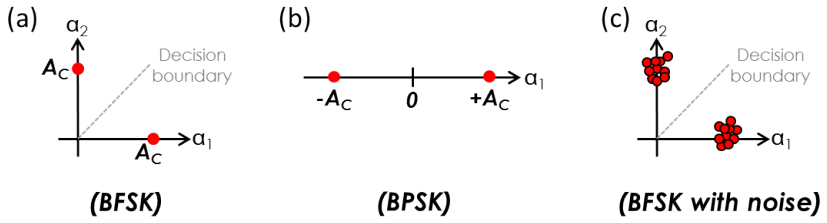

Figure 6: Illustrations of basic digital modulation schemes: (a) binary frequency shift keying BPSK and (b) binary phase shift keying BPSK

2.1.2.2 *Advanced modulation schemes*

Based on these basic schemes, more advanced modulation techniques have been developed [11]. The most obvious motivation is to increase the speed and throughput of the communication. For example, considering FSK and PSK modulations, more than two values of frequency or phase can be chosen. For example, if data is encoded with two bits per symbol, it gives four different possibilities. By extension, we then refer to MFSK and MPSK, M representing respectively the number of frequencies or phases used in the modulation. However, the performance of a modulation technique can be evaluated with three other criteria:

- Bit-Error Rate and noise immunity, which measures how reliable is the communication
- Spectral efficiency, which evaluates the spread of the signals into the RF spectrum
- Power efficiency, which evaluates the power consumption needed to enable modulation

All these criteria must be traded off so as to find the most suitable modulation for a given application.

INTRODUCTION TO QUADRATURE MODULATION It is possible to transmit more that one symbol simultaneously, by splitting the binary data stream into two different bit streams [11]. These two streams are generated before modulation and split in separated paths, as illustrated Fig.7(a).

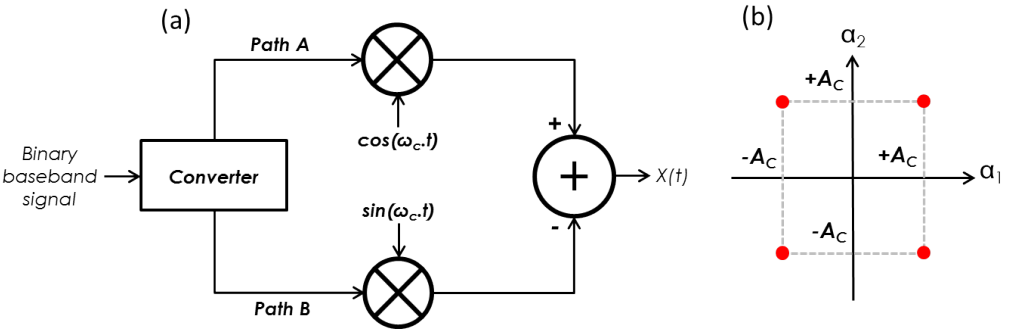

Figure 7: (a) schematic of a quadratic modulator, and (b) example of obtained constellation

This technique is called *quadrature modulation* and is extensively used in modern communication systems. The two generated bit streams (path A and B) are combined and modulated into a *single* carrier:

$$x(t) = \beta_m.A_c.cos(\omega_c.t) + \beta_{m+1}.A_c.sin(\omega_c.t), \tag{3}$$

where $cos(\omega_c.t)$ and $sin(\omega_c.t)$ are orthogonal functions, which is a necessary condition to carry the signals in a single carrier, and β_m and β_{m+1} are the digital rectangular pulses streams split before modulation. A corresponding constellation diagram is presented in Fig.7(b), where β_m and β_{m+1} correspond to pulses with ± 1 values, in which case the modulated signal can be written as:

$$x(t) = \alpha_1.cos(\omega_c.t) + \alpha_2.sin(\omega_c.t) \tag{4}$$

with α_1 et α_2 that can take values of $\pm A_c$.

QPSK AND $\frac{\pi}{4}$ QPSK For instance, PSK modulation can be extended to a quadrature modulator. This type of modulation is known as Quadratic PSK (QPSK) and can give a similar constellation to the one presented in Fig.7(b). Indeed, if we want to obtain exactly the same constellation of Fig.7(b), we can see the modulated signal as a unique cosine function, with four different phases $k\frac{\pi}{4}$, $k = 1, 3, 5, 7$. This leads to a modulated signal equal to $x(t) = \sqrt{2}A_c.cos(\omega_c.t + k\frac{\pi}{4})$, which is equivalent to the previous expression in Eq.4. The number of phases corresponds to the number of symbols that can be transmitted.
Based on this principle, the number of symbols can be doubled by using two quadrature modulators in parallel. In this case, there is one entire quadratic modulator in each path, with for example two QPSK signals with a $\frac{\pi}{4}$ phase shift in between paths. As seen Fig.8(a) for 8 symbols, all the points in the constellation have constant amplitude and different phases, so they all belong to a circle of a given amplitude A_m whose center is in the middle of the constellation diagram. This modulation is called $\frac{\pi}{4}$-QPSK.

QUADRATURE AMPLITUDE MODULATION For a more important number of symbols, instead of using a single modulation, we can use both amplitude and phase to carry information [11]. This is the idea of Quadratic Amplitude Modulation (QAM). Following an example with 16 symbols, the amplitude of the signal is modulated with four different values ($\pm A_1$ and $\pm A_2$), on top of the QPSK phase modulation ($\pm\phi_1$ and $\pm\phi_2$). This gives 16 possibilities for the modulated signal and the resulting constellation is presented in Fig.8(b), for equally spaced amplitudes and phases. In this type of modulation, points are referred in terms of "in-phase" (I) and "quadrature" (Q) components, which gives the (I,Q) plane, equivalent to the (α_1, α_2) plane for the previously presented modulation techniques. QAM can be extended to a lot more values, to obtain 32-QAM, 64-QAM, etc. This is one of the main modulation type used today in wireless communication protocols, such as the IEEE 801.11n standard (WiFi communication).

ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING Until now, the modulated signal is carried on a single frequency signal. To increase the throughput of communication protocols, further extension can be made by using several frequencies in the same RF channel to transmit data. This technique is called Orthogonal Frequency Division Multiplexing (OFDM)

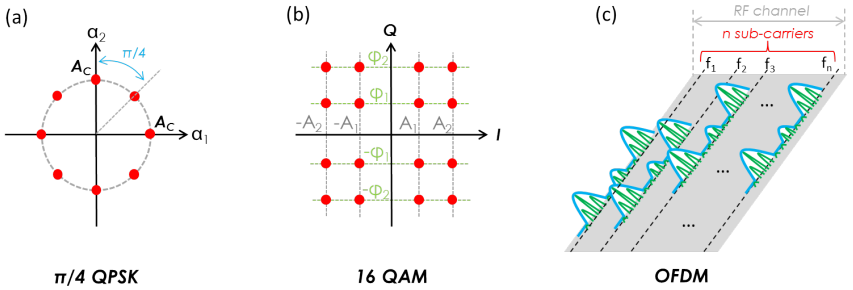

Figure 8: (a) constellation of a $\pi/4$ QPSK constellation, and (b) example of 16QAM constellation

[12, 13]. The basic idea is to encode data on multiple carrier frequencies, where each of the sub-carrier is modulated with a given protocol (like QAM for example), as presented in Fig.8(c). To eliminate interferences, each frequency has to be orthogonal to the other. Data can then be sent in parallel on the multiple sub-carriers. OFDM is used in today's 4G mobile phones, and for TV encoding.

2.2 MAIN PERFORMANCES AND METRICS OF RF TRANSCEIVERS

This section highlights the most important performances that have to be taken into account in the design of RF circuits. For each of the performances, we will detail some of the main influences that it has on the circuit and metrics used to evaluate them.

2.2.1 *Noise*

The term "noise" is defined as a set of perturbations superposed onto the desirable RF signal in the way that they affect the operation of the RF system. We can distinguish two types of noise, namely (a) *outside* noise, i.e. the noise introduced by the environment of the RF transceiver (channel noise, electromagnetic perturbations...), and (b) *intrinsic* noise, introduced into the signal by the RF components of the transceiver.

Different metrics are then available to quantify the noise:

- Signal to Noise Ratio (SNR): The SNR is defined as the power level of the desired signal over the power level of the background noise. This ratio is often expressed in decibels.

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{5}$$

- Noise Figure (NF): the NF of a system is (most commonly) defined as the SNR at its output over the SNR at its input. It measures the degradation of a signal when it passes through a defined system. This definition is generally used to evaluate the amount of noise added by a specific block in the RF chain.

$$NF = \frac{SNR_{OUT}}{SNR_{IN}} \tag{6}$$

Minimum SNR values are defined in wireless communication norms and have to be respected to ensure a reliable communication. On the other side, NF is a critical performance of some RF blocks, especially for LNAs and mixers.

SENSITIVITY Sensitivity of the receiver is defined as the minimum level of signal that can be detected at its input, denoted $P_{in,min}$, for a given SNR. Without loss of generality, the expression of the sensitivity depends of the noise floor derived from the bandwidth of the RF signal and the noise figure, and on the minimal tolerated SNR, denoted by $SNR_{out,min}$.

$$Sensitivity = P_{in,min} = noise\_floor + SNR_{out,min} \tag{7}$$

2.2.2 *Scattering parameters*

Scattering parameters (or S-parameters) are performances that describe the behavior of an electrical network. They are very useful for characterizing RF circuits, mostly because RF circuits are based on matched loads (i.e. impedance matching). The most used S-parameter representation for characterizing RF circuits is a two-port network, as illustrated in Fig.9.

Here, signals on each port p_x (x represented by x=1 or 2) are seen to have incident (a_x) and reflected (b_x) waves. The impedance at the input and output, Z_0, is called the

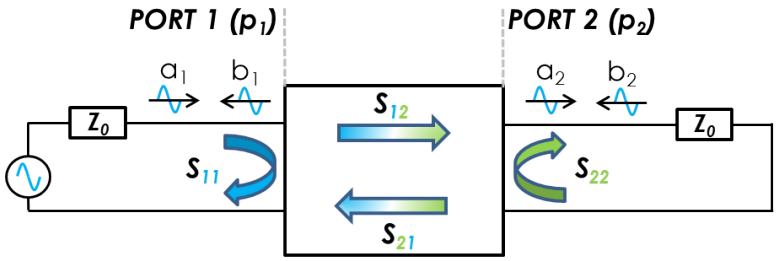

Figure 9: Illustration of S-parameters of an RF circuit

characteristic impedance of the system (usually 50Ω for RF circuits).
The S-parameters can be modeled in a matrix formula given by:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{8}$$

The S-parameters are then:

- S11: input port voltage reflection coefficient. It measures how well the input impedance of the circuit is matched to the characteristic impedance Z_0.
- S12: reverse voltage gain. It measures the voltage isolation between the input and the output ports of the circuit.
- S21: forward voltage gain. It measures the voltage gain between the input and output ports of the circuit.
- S22: output port voltage reflection coefficient. It measures how well the output impedance of the circuit is matched to the characteristic impedance Z_0.

S-parameters are the most used representation to evaluate isolation, gain and impedance matching of RF blocks, which are important performances of RF circuits.

2.2.3 *Transmission channel*

The communication of RF systems is regulated by strict norms, especially for wireless mobile devices. On one hand, the receiver must select only the signal on its communicating band while suppressing signals on the adjacent channels. This property is known as *selectivity*. On the other hand, the transmitter must send the RF signal into the allocated RF channel with enough power to enable a sufficient communication quality, while avoiding hanging over adjacent channels and disturb their communication. This is evaluated by (among others) the Adjacent Channel Power Ratio (ACPR), illustrated Fig.10 that is given by the ratio of the average power transmitted in the allocated transmission channel over the average power transmitted in the adjacent channels.

In terms of design, the PA is critical regarding ACPR, because it is the last block of the transmitter.

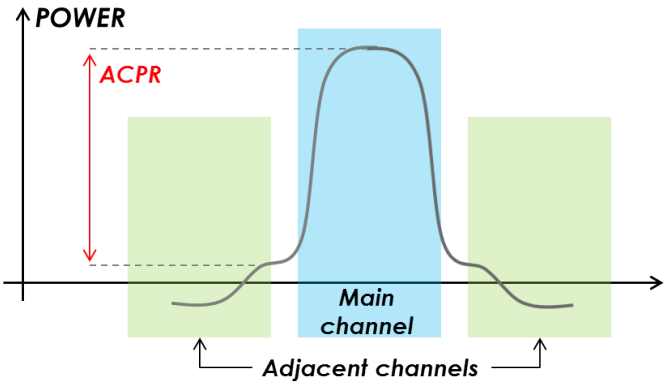

Figure 10: Illustration of Adjacent Channel Power Ratio ACPR

2.2.4 *Non-linear effects*

The RF system can also produce by itself undesirable frequency components, into is own communication channel. In this case, we call the system as "non-linear". The linearity is then an important criterion, especially for the RF blocks that process the signal (LNA, PA or mixer for instance). Basically, an RF block can produce interferers at any frequency, in-band or out-of-band with respect to the communication channel.

1-DB COMPRESSION POINT Harmonic distortion is defined as harmonics produced by a system where only a fundamental RF signal is fed at the input, as illustrated in Fig.11(a). The more the RF power is increased, the more harmonics are produced. In practice, the gain of an amplifier is constant without non-linearities and starts dropping for high RF amplitudes, as seen in Fig.11(b), which produces non-linearities, known as *compression*. To quantify it, we define that the non-linearity limit (the point where we consider that the system is not linear anymore) is the point where the gain has dropped from 1-dB from the gain at low power levels. This point is called 1-dB compression point and is a critical performanace for PAs for example.

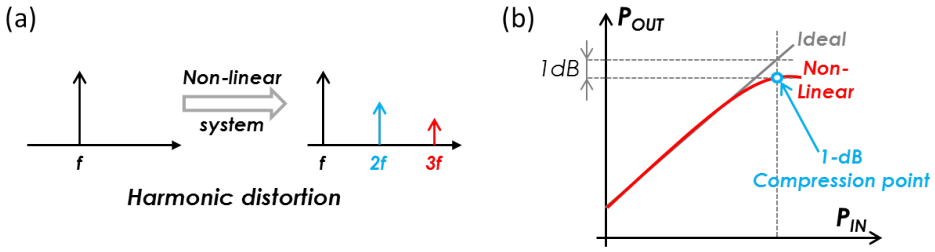

Figure 11: (a) illustration of harmonic distortion and (b) illustration of 1-dB compression point

THIRD-ORDER INTERCEPT POINT Intermodulation is defined as non-harmonic signals produced by a system, in addition to harmonic distortion, as presented in Fig.12(a). For instance, if two RF frequency tones f_1 and f_2 (close to the fundamental) are fed into an RF system, non-harmonic distortion terms can be present at its output. The more critical terms

are third order terms, i.e. at frequencies $(2f_1 - f_2)$ and $(2f_2 - f_1)$, because they are very close to the fundamental value and potentially into the RF channel bandwidth. This is illustrated in Fig.12(b).

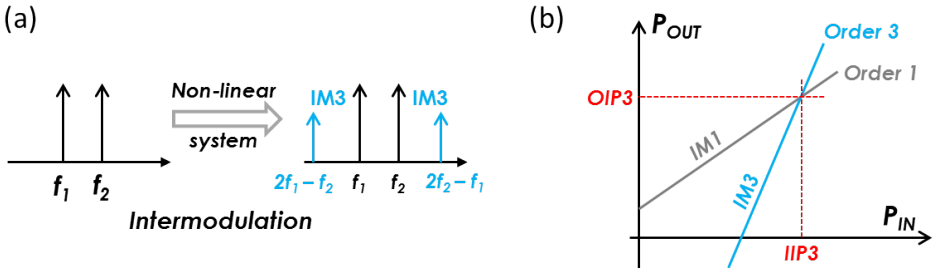

Figure 12: (a) illustration of intermodulation and (b) illustration of third-order intercept point

The point where third-order intermodulation products IM3 are intercepting first-order harmonics IM1 (i.e. at f_1 and f_2) is known as third-order interception point IP3 and is used as the main performance for measuring intermodulation.

2.2.5 *Power efficiency*

Specifically considering RF transceivers for wireless mobile communication, power consumption has become a critical issue, especially for extending battery life. Even if the power consumption is an obvious performance for each RF block, efficiency is more often considered. Efficiency E is given by the ratio of the RF power to the DC power:

$$E = \frac{P_{RF}}{P_{DC}} \tag{9}$$

Another performance is the Power Added Efficiency (PAE), which measures how well the DC power drawn from the supply is converted into "useful" RF power. It is given by:

$$PAE = \frac{(P_{RFout} - P_{RFin})}{P_{DC}} \tag{10}$$

In particular, power efficiency is critical for PA design since PA is usually the block that consumes the most power to amplify the RF signal to a high power level.

2.2.6 *Quality of communication*

The main metric to obtain a general view of the communication quality is the Bit Error Rate (BER), which is given by the ratio of erroneous bits received over the total number of bits transmitted. This metric has many advantages, since it is totally independent of coding and modulation schemes, and gives usable and intuitive results. However, it has a major drawback. Since the usual error rate over the communication is very small (a few erroneous bits per million of transmitted bits for example), a very large sequence of bits has to be transmitted and received to estimate accurate BER values, which requires a long time [2].
In order to reduce the estimation time, the Error Vector Magnitude (EVM) metric has been developed. As illustrated in Fig.13, by measuring the difference of magnitude between the

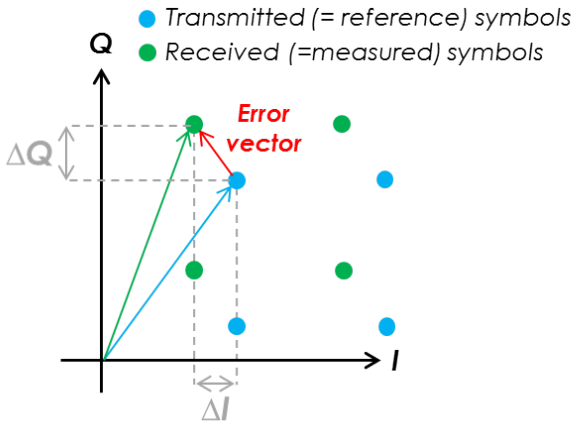

Figure 13: Illustration of Error Vector Magnitude in constellation

transmitted and received symbols in the constellation diagram (I,Q), it is possible to compute the EVM for each symbol.

Measuring EVM is much faster than measuring BER, but is not adapted to any modulation schemes and gives a more abstract value.

2.3 RF POWER AMPLIFIERS IN RF TRANSCEIVERS

This section focuses on PA design, since PA will be chosen as a case study for this work. In general, PA topologies can be broadly classified into two types, linear PAs and switched-mode PAs. We will focus here on linear PAs that are the most power-hungry PAs and their use is widespread in modern transceivers. An introduction to switched-mode PA will also be made at the end of the section.

2.3.1 *Introduction to linear PA design*

BASIC CLASS-A PA DESIGN The primary need of a PA is to amplify RF signals. In this perspective, first PA designs were dedicated to operate with MOS transistors seen as voltage-controlled current sources. An example of a schematic of a one-transistor linear PA is presented Fig.14(a). It consists of a transistor, a choke inductor that isolates the RF part from the DC supply, and a lossless matching network to the load resistance.
In particular, a MOS transistor is assumed to be "linear" between its cutoff and saturation regions [14]. Assuming that the saturation current is equal to I_{max}, and cutoff current is equal to zero, the transistor can be biased in the middle current point $\frac{I_{max}}{2}$, as seen Fig.14(b). In this case, the whole input voltage signal excursion is linearly converted into a drain output current. The conduction time of the PA is then 100%, but the term conduction angle is usually preferred, in this case 2π (conduction all the time). We refer here to class-A operation.
Since the whole excursion is used, class-A corresponds to a maximum linearity, but is also the most power-hungry mode of operation. Indeed, since bias point has to be set at $\frac{I_{max}}{2}$, it leads to a maximal current excursion at the output. In terms of efficiency, class-A operation has a maximal efficiency of 50%, in theory. Half of DC the power is then wasted, instead of being converted into RF.

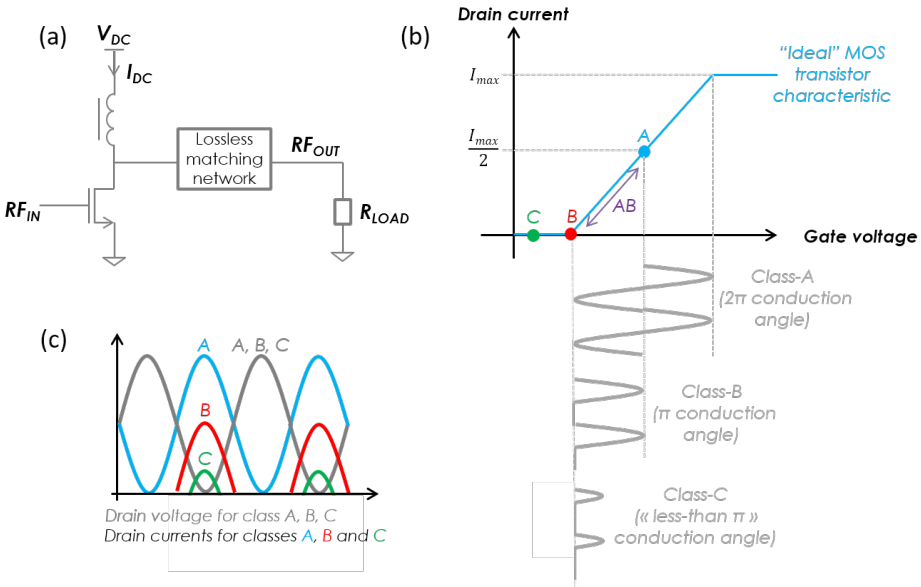

Figure 14: (a) Schematic of one-transistor PA, (b) illustration of different linear classes operation and (c) illustration of drain voltages and currents for linear PA classes

OTHER LINEAR PA In order to reduce power consumption and increase efficiency, the bias current of the transistor can be lowered [14]. Consequently, as seen Fig.14(b) and (c), conduction angle is reduced with bias current. We obtain the other linear classes of operation for a PA, namely class-B for a π conduction angle (which means that exactly half of the input excursion is converted into drain current), class-AB for a conduction angle between 2π and π, and class-C for conduction angle lower than π. It has to be noted, as illustrated Fig.14(c), that for each of these operation classes, drain current is modified but drain voltage stays constant (i.e. a full sinusoid).
In terms of power efficiency, class-B attains theoretically 78.5%, and class-C up to 100% for a conduction angle of 0 (even if this configuration is not used in practice).

2.3.2 *Figures of merit in linear PA design*

2.3.2.1 *Linearity versus Efficiency: power back-off*

A major trade off in PA design is the balance between linearity and efficiency. On one hand, with class-A operation, linearity is maximized, while efficiency stays relatively low [14]. In addition, the linear assumption made previously is very optimistic in practice, since the characteristic of MOS transistors is not exactly linear. The solution to more linearity is then to operate at *power back-off*, which means that the output current excursion (i.e. P_{RF}) is lowered, to reduce non-linearities. However, this configuration degrades the efficiency, since maximum RF power is decreased while DC power remains constant.
On the other hand, when lowering the conduction angle, the output drain current is not an entire sinusoid. Usually, a filter (RLC tank for instance) is placed at the output of the PA to eliminate undesirable harmonics and keep only the fundamental sinusoid. But linearity can be affected, especially for class-C operation. Practically, linearity stays constant for class-A and class-B operations (with a much lower power consumption for class-B) and drops for class-C operation.

2.3.2.2 *Efficiency versus gain: input drive requirements*

Another factor to take into account is the gain of the amplifier [14]. This gain is related to the input drive requirement, which corresponds to the input level required to fully drive the PA. In practice, class-B operation will require a much higher drive level than class-A, that can be up to 6 dB, corresponding to a power gain reduction of the amplifier of the same amount. This difference can be important for RF PA design, where intrinsic transistor gain is lower than that of low-frequency amplifiers. However, this power gain reduction is related to efficiency improvements (since DC power consumption is improved in class-B comparing to class-A). It then appears that there is a trade off between gain and efficiency in PA design, in addition to the efficiency versus linearity trade off.

2.3.2.3 *Multistage PA design*

The different trade offs can be more easily tackled using multistage PA designs. The advantage of this technique is to decompose the necessary requirements (linearity, efficiency, gain) on different cascaded PA stages. Usually, the final PA stage, called power stage, is preceded by a driver stage, which will bring more gain, since it is used at lower power levels (i.e. with

relaxing linearity requirements). On the contrary, the power stage can then be designed for maximum linearity (which will lead to lower gain).

2.3.3 *High-efficiency PA design*

The basic principle of high-efficiency PAs is to operate the MOS transistor as an on/off switch instead of a linear operation. These types of PAs are called Switching-Mode Power Amplifiers (SMPA), and can be also categorized in several classes. Their main advantage is that they can achieve a theoretical 100% efficiency.

CLASS-E AMPLIFIERS As an example of SMPA, Fig.15(a) shows a simplified schematic of a Class-E PA [15]. It is composed of a MOS transistor used as a switch, a choke inductor L that connects the drain of the transistor to the supply voltage, a bypass capacitor C in parallel of the drain-to-source channel of the transistor (that can be realized with the parasitic drain-to-source capacitance of the MOS transistor), a L_s-C_s filter tuned to let pass only the fundamental frequency of the system (and filter out all the other harmonics), and a load resistance R_L. The main currents and voltage waveforms are depicted Fig.15(b).

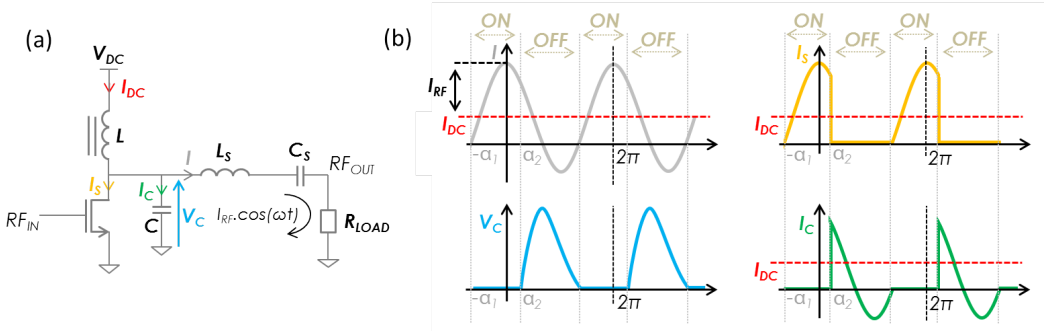

Figure 15: (a) basic topology of class-E amplifiers - (b) main current and voltage waveforms of class-E PA

The basic principle is that the MOS transistor is switched on for a fraction of the RF period and is switched off for the rest of the time [14]. Assuming that there is a "flywheel" sinusoidal current $I_{RF}cos(\omega.t)$ flowing into the L_S-C_S-R_L branch, in addition to the DC current I_{DC} provided by the supply, the total current in this branch is $I = I_{DC} + I_{RF}cos(\omega.t)$ as seen in gray in Fig.15(b). In the figure, the current into the MOS transistor I_S, the current into C denoted by I_C, and the voltage across C denoted by V_C, are respectively represented in yellow, green and blue. When the transistor is on, e.g. it is seen as a closed switch (as between $-\alpha_1$ and α_2), the current I is flowing into it and I_C and V_C are equal to zero. When the transistor is turned off, e.g. it is seen as an opened switch (as between α_2 and $2\pi - \alpha_1$), the current I is now flowing into C, and the combination L-C produces also a voltage across C. In the meantime, there is no current flowing into the MOS transistor. As it can be seen, there is no point where the current I_s and V_C coexists at non-zero values. This results in a theoretical conversion efficiency of 100%. Moreover, the switching transistor can be turned on when the voltage across it is zero, at times α_1 and $(2\pi - \alpha_1)$ for instance, which eliminates the losses due to switching. However, the PA works only around a specific operating frequency, tuned by L and C. This can cause several issues when the capacitor C is realized with the parasitic capacitance of the MOS transistor. Additional filters are also usually needed to eliminate

undesirable harmonics.

More architectures have been implemented to improve the performances of SMPAs. However, linear topologies are still more often used in practical transmitter designs because they bring more linearity, which is a primary requirement of PA design for modern wireless communication systems.

2.4 WHICH PARAMETERS TO ADAPT/COMPENSATE FOR ?

The aim of this thesis work is to develop methodologies that adapt a RF circuit against different types of variations. In order to understand how to compensate for these variations, this section is dedicated to give an introduction to the main variations that affect an RF circuit's behavior, explain their origin and discuss how current design methodology take them into considerations.

2.4.1 *Process variations*

The manufacturing process of integrated circuits is subject to many variations. Manufacturing variations result from non-uniform conditions during the different steps of the fabrication process, i.e. deposition of the layers and/or diffusion of the dopants. Process parameters include dopant concentration densities, oxide thickness, diffusion depths, etc. The variation of these process results in variation in the electrical characteristics of the component (MOS transistors, resistors, capacitors, etc), such as threshold voltage, mobility or sheet resistance. In the end, these variations impact the global performances of a circuit. Moreover, as technology is scaling down, these variations are more and more significant, as explained in chapter 1.
For a given circuit, process variations can be categorized into two main types, namely *inter-die* and *intra-die* variations, as seen in Fig.16. Inter-die variations are impacting in the same way identical components in a same die. They can be classified into different categories: the process variations between lots of wafers (lot-to-lot), between wafers (wafer-to-wafer), or between the dies of the same wafer (die-to-die). Intra-die variations affect differently identical components in the same die. Both types of variations impact a fabricated die and have to be considered simultaneously.

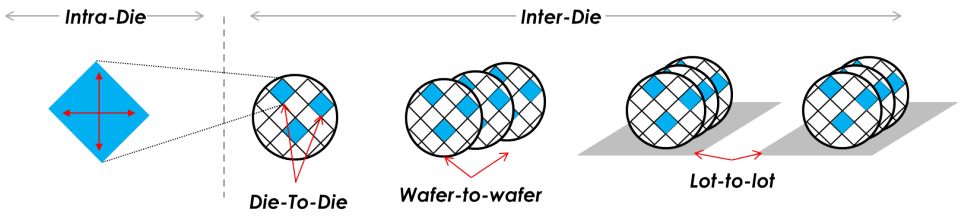

Figure 16: Overview of process variations for a given integrated circuit

The usual way to minimize the impact of process variation is to use *worst-case* design, i.e. designing the circuit such as it works even with extreme process variations.

2.4.2 *Environmental variations*

During its operation, a circuit is subject to many environmental variations, that can result in suboptimal performances and/or power consumption. In traditional worst-case design, the circuit is designed to function under the worst environmental conditions, but this can have a dramatic impact on its power consumption.

TEMPERATURE Any semiconductor circuit is sensitive to temperature variations. Indeed, temperature may have a negative impact on speed, power and reliability of CMOS circuits. In particular, it mostly affects the threshold voltage V_{TH}, the mobility μ and the velocity saturation v_{sat}. In traditional worst-case design, temperature variations are taken into account, but worst-case temperature condition is almost never encountered by the circuit, which leads to non-optimal power consumption and/or performances.

CHANNEL CONDITIONS Focusing on RF circuits, the quality of the communication through the channel varies from time to time, depending on many factors, such as the place of operation (city, countryside...), the distance from the antenna or from the other device, the noise into the channel, etc. All these factors affect the RF circuit and usually lead to non-optimal power and/or performances. On one hand, if the channel is degrading, the quality of communication is getting worse, and the circuit may not be able to provide the required quality of service. On the other hand, if the channel has a very good quality, the communication can be "too good" with respect to the minimum requirements. In other terms, the performance constraints on the RF transceiver can be relaxed so as to lower its power consumption.
As it will be seen in section 2.5, some techniques have been implemented and included in norms of several wireless communication protocols so as to obtain the required power consumption and performances depending on channel conditions.

2.4.3 *Aging*

As the circuit is used and stressed, its performances are degrading over time. This aging can cause failures in the circuit or can also be compensated by more power consumption to satisfy the requirements throughout the lifetime of the circuit.
Concerning MOS transistors for analog and RF circuits, three major transistor aging effects have been observed [16]:

- Bias Temperature Instability (BTI): during operation, the characteristics of a MOS transistor are affected by the changes in its bias and temperature conditions. This effect, known as BTI, degrades the behavior of the transistor. It mostly affects the threshold voltage [16, 17].
- Hot Carrier Injection (HCI): when the electric field into the MOS transistor is high some carriers may reach very strong energy, sufficient for them to be injected into normally forbidden regions (like crossing the gate oxide for example) [17]. There, they are trapped and cause defects that can be accumulated over time and degrade transistor's performances. This effect affects different parameters of MOS transistor including the threshold voltage, the mobility and the output conductance.
- Time-Dependant Dielectric Breakdown (TDDB): The electric field into the MOS transistor can also cause the gate oxide to break down. Either completely due to a very high field, or more slowly due to a low level field applied to the transistor for a long time. This slow breakdown is the most dominant effect for sub-180nm CMOS technologies [16], and causes a progressive degradation of the MOS transistor characteristics.

Moreover, with the scaling of CMOS technologies, these aging effects are more and more important.

2.5 ADAPTIVE MODERN RF TRANSCEIVERS

Many different solutions have been proposed to adapt the performances and reduce the power consumption of RF transceivers. A first approach is to design specific circuit architectures or models, as opposed to more generic adaptation strategies, based on dedicated digital algorithms. First, modern "mostly-analog" transceivers able to handle modern communication protocols will be presented. Then, solutions will be divided in several categories: channel-adaptive transceivers, digitally-assisted RF transceivers, and power/performance adaptation for receivers and transmitters (through PA).

2.5.1 *Example of a modern analog RF transceiver architecture*

Recalling the basic RF transceiver architecture seen in section 2.1 , tremendous improvements have first been made on the design of "mostly-analog" RF transceivers. A conventional analog RF transceiver design that can handle complex quadrature modulation schemes is presented in Fig.17. As it can be seen, its architecture is far more complex than the basic topology presented in section 2.1.

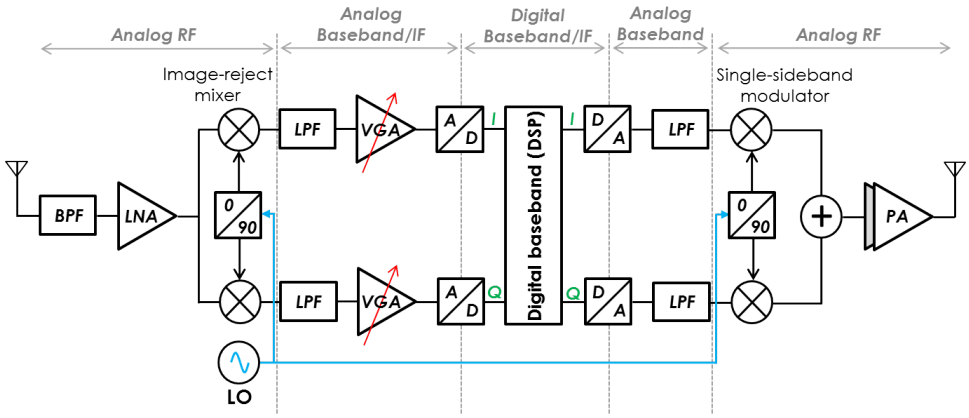

Figure 17: From [5]: example of a modern RF transceiver

On the transmitter's side, baseband data is first converted into I and Q digital signals (if the transmitter uses quadrature modulation for instance). Those signals are converted in analog through a digital-to-analog converter (DAC or D/A), undesirable harmonics are filtered by a Low-Pass Filter (LPF) then they are up-converted into RF thanks to a single-sideband modulator and are finally transmitted to the antenna through the PA.
On the receiver's side, out-of band blockers of the received signal are first filtered by a Band Pass Filter (BPF). A LNA amplifies the signal, which is then down-converted to baseband or to an Intermediate Frequency (IF). This conversion is done through a image-reject mixer that operates with the in-phase and out-of-phase complex components of the modulation. After passing trough a LPF and amplified with a Variable-Gain Amplifier (VGA) the signal is finally converted in digital through an analog-to-digital converter (ADC or A/D), to be processed in the DSP.

2.5.2 *Channel-adaptive transceivers : Adaptive Modulation Coding*

A first approach of adaptive transceiver design is to perform a global adaptation at a system (transceiver) level. Towards this goal, modern RF transceivers, especially those used for mobile applications (smartphones for instance), are incorporating an adaptive modulation of the coding scheme, so as to adapt the communication to different channel conditions, called Adaptive Modulation Coding (AMC) [18, 19]. More specifically, the underlying idea is to modify the type of coding scheme and /or the modulation coding rate used by the system depending on channel conditions, as illustrated in Fig.18 for IEEE.802.11n protocol [6]. Coding rate is an indication of how much of the data stream is actually being used to transmit usable data and is expressed as a fraction of the most efficient rate (for instance 5/6 corresponds to 83.3% of the data stream being used).

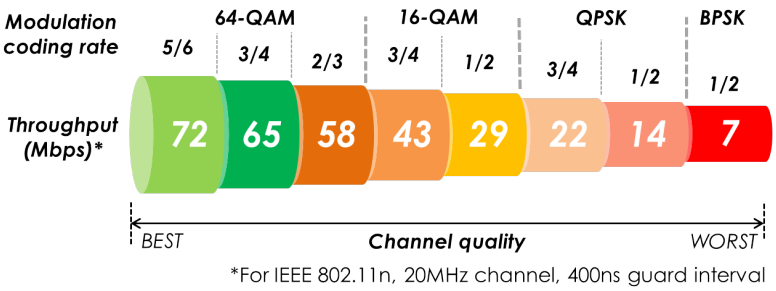

Figure 18: from [6]: principle of AMC

Indeed, in the different communication protocols, minimal SNR and receiver sensivity are defined, as shown in table 2 for the IEEE 802.11a protocol [20].

Data Rate [Mbps]	Modulation type	Minimum Sensitivity [dBm]	Required SNR [dB]
6	BPSK	-82	4.2
18	QPSK	-77	8.8
36	16-QAM	-70	14.8
54	64-QAM	-65	20.1

Table 2: Examples of specifications for IEEE 802.11n protocol

As it can be seen, switching modulation type, depending on data rate for instance, may enable to relax the constraints on performances (sensitivity and SNR for example) and consequently reduce the power consumption.

2.5.3 *Digitally-assisted RF transceivers*

As discussed in the general introduction, the aggressive scaling of CMOS technology is not suitable for obtaining high yield and reliable low-power analog RF transceivers. In contrast, digital circuits have always largely benefited from this scaling, where at each technology advancement, better performances, highers speed, less power consumption and area can be achieved for a relatively constant fabrication cost [5]. Based on this assessment, the idea

behind digitally-assisted RF transceivers is to replace most of the analog RF blocks in the transceiver with *digital RF* equivalents, with the underlying objective of totally replacing analog RF blocks. In other terms, the aim is to "help" analog RF blocks to be less affected by technology scaling, by providing them digital assistance. Towards this objective, extensive research work has been carried out, to develop digitally-assisted industrial RF transceivers, able to satisfy simultaneously the more and more stringent requirements of modern communication protocols (high performance, high reliability, low power, etc) [21].
It has to be noted that an extensive review of digitally-assisted blocks is not in the scope of this introduction, but an overview will be presented to understand the general "philosophy" of digital RF designs, with the help of Fig.19.

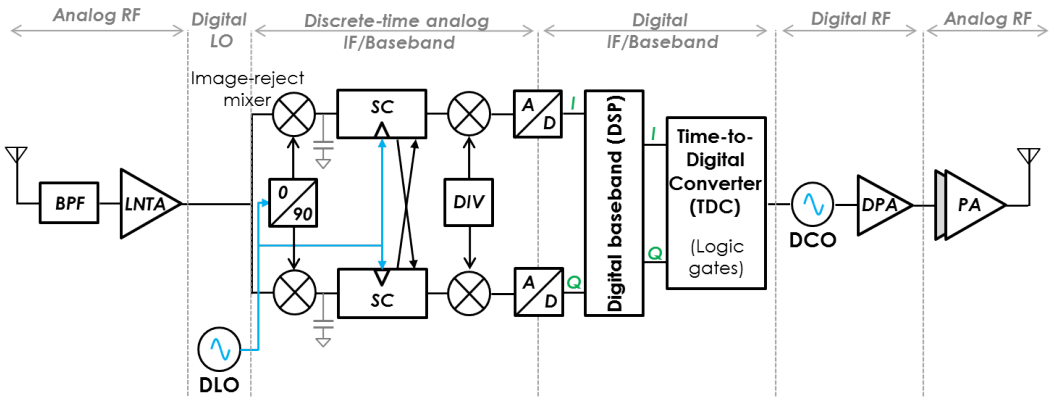

Figure 19: from [5]: example of mostly-digital RF transceiver

TRANSMITTERS Several transmitters have been converted to the digital domain [5]. For example, usual analog PLL is replaced by an All-Digital PLL (AD-PLL), composed by a Time-To-Digital Converter (TDC) realized with logic gates, and a Digitally-Controlled Oscillator (DCO). Polar modulation can be used with this topology, with a phase modulation through the AD-PLL and an amplitude modulation that drives a Digital PA (DPA). A more detailed overview of polar transmitters is given in section 2.5.5.

RECEIVERS Considering the receiver, "mostly-analog" receivers can also been replaced by direct conversion receivers for example. The basic idea is to convert the incoming RF signal into digital, or discrete time, as soon as possible in the receiving chain. Generally, it consists of replacing usual tuned RF elements in RF filters by high-frequency sampling. In other words, frequency translation is done by sampling, instead of filtering. Ideally, sampling has to be realized after the LNA, which has to be tunable (LNTA). However, the main bottleneck in this type of design is the ADC since at a sampling frequency close to the RF, it requires very high dynamic range [22]. This type of ADC is not available today with a suitable power consumption, this technique is then impractical. For this reason, the data conversion is usually performed at an Intermediate Frequency (IF), but the benefit of using a direct conversion is lost. If suitable ADC architectures are available in the next future, which could be possible since sampling frequency will improve with technology scaling, direct conversion architecture will achieve better performances and less susceptibility to process variations. Alternatively, RF-sampling receivers have also been developed. The frequency translation is also performed by sampling and analog-to-digital conversion, right after the LNA. Sampling

is here realized with Switched-Capacitor (SC) passive filters for example, that are digitally reconfigurable, well-controlled and suitable for low supply voltage operation (in modern CMOS technologies). Requirements on the ADC can then be relaxed by using the oversampled RF signal at the output of the SC filter.

However, in any case, the use of mostly-digital RF receivers does not eliminate the constraints linked to the receiver design, such as gain and noise balance, baseband filtering, digital processing, power consumption of each block, etc.

2.5.4 *Lowering receivers power consumption*

Adaptation can also be performed at a lower level. Regarding the receiver, there is no dominant RF block in terms of power consumption, as it is the case for the transmitter whrere the PA comsumes almost all the power. Adapting the power of RF receivers will then require to act on several blocks simultaneously. Some of the basic adaptation techniques are detailed in the following paragraphs [7].

SENSING OPERATING CONDITIONS Adjusting the performances of the receiver requires to sense its operating state which depends on the received signal power, with on-chip detectors for example. This measurement can then be used to calculate the required Signal To Noise Interferers Ratio (SNIR), that captures both the linearity (IP3) and noise figure (F), which can be considered as the two main performances of the receiver. This SNIR estimation, as illustrated in Fig.20(a), can be plotted into a noise-linearity plane. Acceptable SNIR corresponds to satisfying a minimum IP3 requirement ($IP3_{lim}$), and a maximum noise requirement F (F_{lim}). These values define an acceptable operating region shown in gray in Fig.20(a).

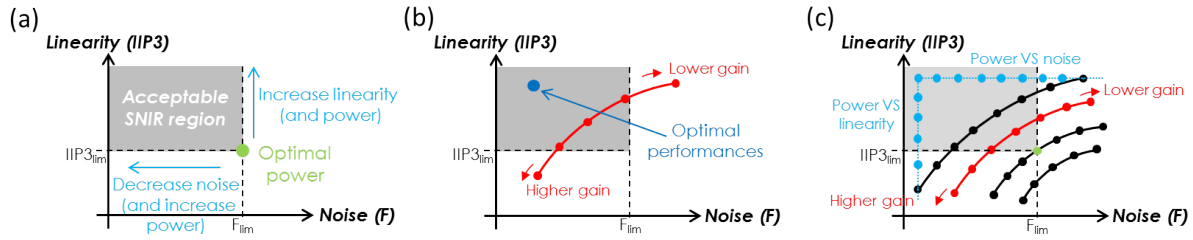

Figure 20: inspired from [7]: (a) illustration of acceptable SNIR and power control for receivers, (b) illustration of Automatic Gain Control, (c) illustration of combined gain and power control

POWER ADAPTATION On one hand, increasing linearity or decreasing noise is done at an expense of power consumption. If power reduction is required, a price has to be paid in terms of, at least, one of these two performances. The optimal operation point, represented in green, is then the point $[IP3_{lim}; F_{lim}]$ since it is the point where operating conditions are satisfied ($SNIR = SNIR_{min}$) and power consumption is minimized.

AUTOMATIC GAIN CONTROL On the other hand, tuning linearity and noise can be done by automatically adjusting the gain of the receiver to extend its dynamic range. This is done with Automatic Gain Control (ACG) by adding a variable-gain amplifier after the first low-pass

filter of the receiver. This technique enables to trade off noise and linearity, but both cannot be optimal at the same time, as illustrated in Fig.20(b). By using ACG, power consumption can be relaxed since both high linearity and low noise are not achievable simultaneously, compared to usual receiver's designs where both maximum linearity and minimum noise are targeted, represented by a blue point in Fig.20(b). However, all states of ACG operation have the same power consumption.

PERFORMANCE AND POWER ADAPTATION By combining ACG and power adaptation, the receiver can adapt its performances together with is power consumption, as in illustrated Fig.20(c). Power is then traded off either with linearity or with noise. We can obtain significant power savings with gain and performance control.

DISCUSSION All these techniques can be applied to reduce power consumption and find a suitable trade off between receiver's performances. However, they are usually working on pre-defined paths [7], without configuring independently the different performances and the power consumption. These issues can be solved by using more advanced adaptation techniques, with dedicated adaptation algorithms, as it will be discussed in the next chapter.

2.5.5 *Lowering transmitters power consumption through digital PA assistance*

In the transmitter, the PA is usually the most power-hungry block, and is also critical for maintaining the transmitter's RF performances. For theses reasons, PA is usually the main concern for power and performance adaptation. The basic paradigm of PA design is to simultaneously satisfy the different constraints, namely transmitting a signal with high power (1W or more) and maintaining linearity, while having the best energy efficiency and saving battery life. To this end, active power saving techniques have been developed for modern PA designs [6], as illustrated in Fig.21.
Traditional static PA designs, as seen in Fig.21(a), use a fixed supply voltage, provided by a Power Management Unit (PMU), which results in wasting a lot of power [6]. Indeed, since the PA is amplifying variable envelop signals, it is most of the time forced to work at power back-off, where linearity is guaranteed but efficiency is not optimal. Since static power consumption is fixed and suitable for maximum power levels, a significant amount of power can be wasted if the RF signal to be transmitted is not at maximum power. Adapting PAs is aiming at achieving two goals: (a) increasing the RF performances of the PA to be able to relax design constraints, and (b) reducing power consumption by finding on-line the most suitable power supply voltages to reduce the amount of wasted DC power. It can be noted that both goals can be achieved at the same time. Some techniques have been developed and summarized in Fig.21(b).

In order to increase the RF performances, the digital domain can come to the rescue. Several techniques, like Digital Pre-Distortion (DPD) [14], Crest Factor Reduction (CFR), or digital companding, can be used to shape the incoming RF input signal of the PA and bring more linearity and/or efficiency [6], as presented in Fig.21(b). In the meantime, Power Control (PC) can be achieved by communicating power information through the base station. For example, in Global System for Mobile communication (GSM) protocol, the Base Station (BS) controls the power of the mobile station by sending it required power information through the RF link [19].

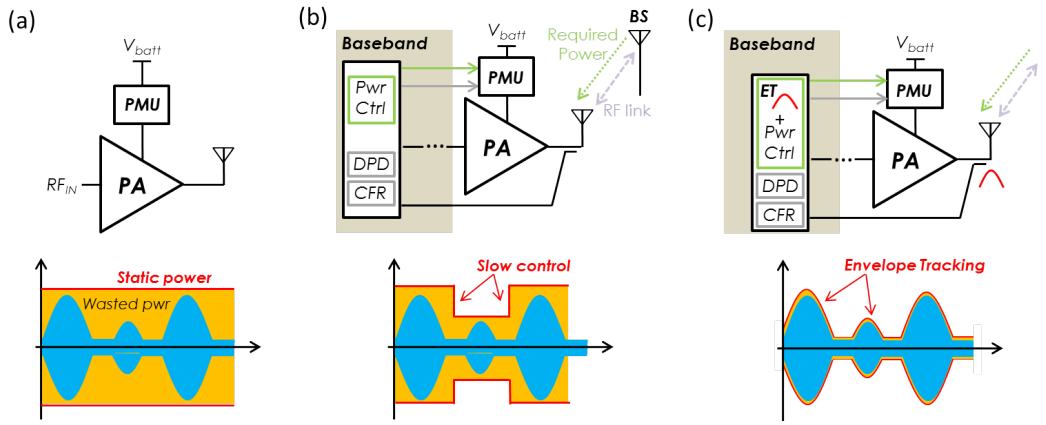

Figure 21: From [6], several PA architectures: (a) static PA, (b) slow power control and digital assistance, and (c) power control with envelope tracking

Further accurate power adaptation can be achieved by adapting power supply according to the envelope of the RF signal. This technique is known as Envelope Tracking (ET). Since a maximum efficiency is obtained when the difference between the envelope and the power supply is minimized, the aim of ET is that the power supply voltage follows as close as possible the envelope of the RF signal to transmit, as depicted Fig.21(c). Envelope detectors, or couplers can be used for this tracking and a control loop assigns the most suitable power voltage as fast as possible regarding the instantaneous tracked envelope.

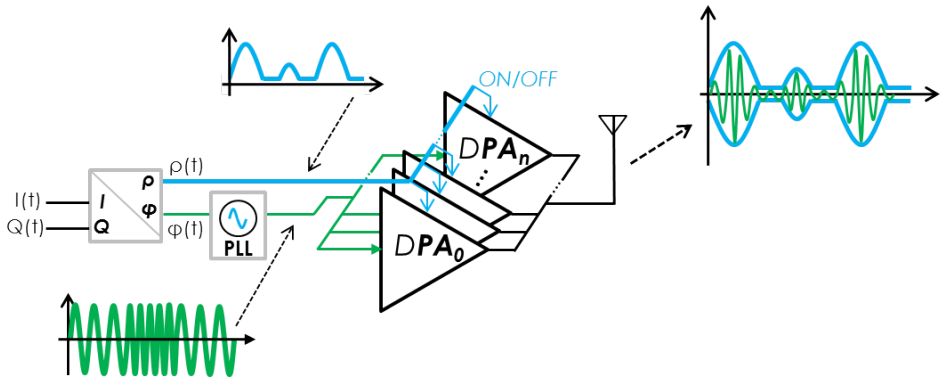

Figure 22: Principle of operation and usual topology of polar PA

A more drastic power consumption adaptation can also be done by using polar transmitters. The basic principle is to convert the usual cartesian signal (I,Q) of the quadrature modulator into polar coordinates (ρ, ϕ). The phase signal of the $\phi(t)$ path is modulated into the RF carrier through phase-modulation. The amplitude in the $\rho(t)$ path is then multiplied with the modulated phase to obtain the final modulated signal. This multiplication can be done by a multiplier before the final PA, or directly at the PA stage to control its supply. For instance, with the polar transmitter topology presented in Fig.22 [23], the amplitude information of ρ is used to turn on and off several "digital" PAs in parallel, to modulate the output amplitude of the PA. Each individual PA can be realized with a SMPA topology, to obtain highly energy-efficient modulated signals at the output of the PA, without complex power

control.

2.6 CONCLUSION

The design of RF transceivers has moved to sophisticated architectures able to communicate in the most stringent wireless protocols. A lot of constraints have to be simultaneously satisfied (noise immunity, linearity, high gain, the worst environmental conditions, etc) while lowering power consumption at maximum. Moreover, the RF transceiver is submitted to increasing process variations as we move to the most advanced technology nodes, that mostly affect its analog parts. As we have seen, many adaptation strategies exist and significant improvements have been made by developing adaptive modulation, digital RF circuits, adaptive LNAs, mixers and PAs, etc.
However, even with these developments the trade off between the various performances and power consumption of RF transceivers is still suboptimal. For example, as seen previously, the most drastic approach to increase PA efficiency and reduce its power consumption is to use Switched-Mode PAs (SMPA). However, the use of this type of PA is not feasible for all RF modulator types and linear PAs are still required in many cases. In the design of RF transmitters, the PA that is used has to take into account consideration the type of modulator. Indeed, the type of modulation has a main influence on the PA efficiency and by extension on the overall power consumption of the system [15].

The development of global adaptive RF circuits is then necessary to achieve simultaneously robustness against variations, high performances and low power. For this purpose, implementing a dedicated adaptation algorithm is a path that has been recently investigated. But developing a global strategy requires that the RF circuit has many capabilities, among others be able to test by itself its performances and tune them. The next chapter will go into details about these abilities, by providing a state-of-the-art of test, calibration and adaptation of RF systems.

3

STATE-OF-THE-ART

Developing a global adaptation strategy consists of giving new capabilities into RF circuits. The circuit should be able to self-test its performances and to be aware of its current operating conditions. Then, it needs to be equipped with different mechanisms that compensate for the performance or power losses due to variations. The aim of this chapter is then to understand how to test RF performances of a circuit, and how to compensate these variations in the circuit itself.

First of all, the circuit has to be able to test itself its performances, at low cost and with maximal efficiency. Towards this objective, advances in the test of RF circuits will be detailed in section 3.1. Then, once performances are estimated, the circuit must be able to tune them. This tuning will be divided in two parts: calibration, or process-variation compensation, that will be reviewed in section 3.2, and adaptation to the other types of variations (environment, aging...), that will be detailed in section 3.3. This state-of-the-art will guide us to choose our strategy, for developing a self-calibrated RF circuit and extending this methodology to further adaptation.

3.1 TESTING RF CIRCUITS

To be able to calibrate or adapt an RF circuit, we first need to know the target performances. However, testing the RF performances of a circuit is not straightforward. Indeed, RF signals have to be generated as test stimuli for the Device Under Test (DUT), and RF data has to be collected from the DUT to the tester with high-speed links in order to be analyzed. The test environment must have under control noise and/or interferences and has to be carefully designed. Moreover, all tests have to be done as quick as possible, with a minimal cost added to the fabrication, since test does not produce any added value to the circuit, excepting the fact that we ensure its functionality. To have a general overview of the Cost of Test (CoT), the following expression can be used [24]:

$$CoT = \frac{Fixedcost + Recurringcost}{Yield * Throughput}, \tag{11}$$

where *Fixedcost* are the tester, probes and installation costs, *Recurringcost* is referring to the maintenance cost of the tester (up to 10% of the fixed costs), *Yield* is the number of fault-free dies over the total number of tested dies, and *Throughput* is given by the number of fault-free dies over the total test time of all dies. If test cost is high with respect to fabrication costs, it can lead to the paradoxical situation that it is less costly to put the circuit into the final product and replace it after discovering a defect, than screening out and discarding a defective die at an expense of more time-consuming production tests. Test cost is then a critical issue.
As it will be seen in this section, the usual test techniques used at the post-fabrication stage suffer from critical disadvantages and are not suitable candidates to enable cost-effective calibration and/or adaptation of RF circuits. This section is then dedicated to present alternative strategies, aiming at extracting the RF performances of a circuit without requiring techniques that are too costly.

3.1.1 *Usual RF test with Automatic Test Equipment*

Industrial test is usually performed by extracting the performances of an RF circuit with an Automatic Test Equipment (ATE). ATEs are computer-controlled test stations, able to simultaneously test several integrated circuits as fast as possible during high-volume test phases where millions of devices are fabricated. Production test is usually divided into two parts:

- On-wafer test: all the dies of the same wafer are first tested individually with specific probes. This phase is typically used to detect grossly defective dies before packaging.

- On-package test: the dies that pass the on-wafer test are then packaged, and subjected to more tests to confrim whether they meet or not the specifications.

Considering on-wafer test, the standard procedure for a SoC, illustrated in Fig.23, is as follows [25]. The global performance test consists of sequential tests, each of them employing on a different configuration. In the beginning, the first test configuration is loaded into the tester. After some settling time, the measurements are taken by the ATE and stored, during the test time t_1. Then, another configuration is loaded into the ATE during the switching time

t_{12}, and the process is repeated sequentially, for the N different configurations, until all performances are obtained. This test methodology is known as *functional testing*, or *specification-based testing*. Indeed, each performance is compared to pre-defined specifications and, if the circuit does not lie within the specification range (upper or lower-bounded depending on the performance considered), then the circuit fails the test and considered as "defective". If specifications are satisfied, the circuit passes the test and is considered as "fault-free".

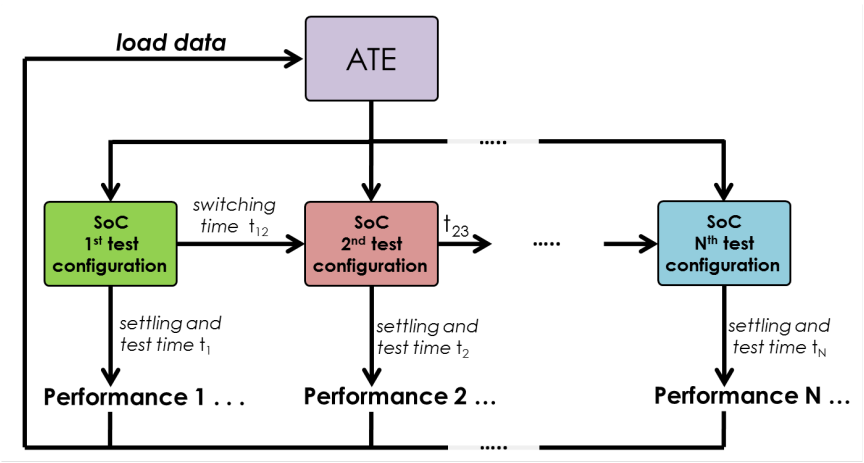

Figure 23: Usual test procedure with Automatic Test Equipment

DISCUSSION Taking test cost factors into account, production test with ATE is very expensive. Indeed, it requires to individually test each die of each wafer a first time, and packaged dies a second time, which is time-consuming. Knowing that an RF ATE can cost up to 2 billion dollars, the test cost of each die can be in the order of a few cents per second of test, which can quickly become critical for complex SoCs that require many tests [26]. In practice, RF test with ATE is too expensive and test cost must be reduced.

REDUCING ATE TEST COST? Based on the fact that some performances of the circuit are correlated together (since they are mostly affected by identical process variations), first works aiming at reducing the cost of ATE-based test were focusing on two main aspects:

- Reducing the number of functional tests, by eliminating redundant tests for example, so as to reduce the total test time
- Defining the best order for the different tests in the functional test sequence, so as to reduce the test sequence at maximum

Even if the overall test time can be reduced with these techniques, the ATE-based functional test is still very costly so not suitable for a low-cost testing strategy.

3.1.2 *Built-In Test techniques*

To alleviate this issue, costly RF links between DUT and ATE have to be replaced. This technique is known as Built-In Test (BIT) and has been developed for several digital, mixed-signal, and also RF circuits. The basic idea of BIT is to transfer a maximum of the tester functionalities (test stimuli generator, measurements...) onto the chip so as to use a low-cost

ATE, or ideally perform the entire test procedure on-chip, in which case we refer to Built-In Self-Test (BIST). A first benefit of this approach is that, since high-speed links and multiple connections are suppressed, the test procedure is much faster and with less interferences.
But, by replacing the standard test with BIT, the RF performances are not measured explicitly, since measurements are not performed at RF level on-chip. Instead, they are replaced by low-cost measurements that only *reflect* the actual state of the RF circuit. Test engineers have then to set test margins on BIT measurements so as to assess if the circuit is defective or not. However, this approach presents two main challenges:

- How to develop efficient on-chip sensors to obtain accurate low-cost measurements?
- How to detect defective dies solely based on low-cost measurements?

In order to answer those questions, some BIT sensors and alternative BIT frameworks will be detailed in the following subsections.

3.1.2.1 *Types of BIST measurements*

DC SENSORS (VOLTAGES - CURRENTS) A first approach to extract simple measurements is to measure DC voltages and/or currents directly into the DUT. Concerning DC voltages [27], all analog/RF circuits are designed to work under a given static operating point. If there is a defect into the analog/RF chip, this defect will also certainly affect its DC operating point. By setting proper test limits, it is then possible to detect defects solely by measuring DC voltages at critical DC points.
Concerning current sensing [28, 29], the strategy usually consists of dynamically measuring the DC current provided by the power supply. This technique has initially been developed for digital circuits, but can be adapted for analog and RF circuits as well. Indeed, if an RF circuit is defective, its dynamic current consumption will be modified and can then be detected by a current sensor inserted on the power line. For example, we can cite a Built-In Current Sensor (BICS) presented in [30].

ENVELOPE DETECTORS Another approach to obtain measurements on an RF signal at a lower frequency is to monitor its envelope. Indeed, the envelope is directly proportional to the amplitude of the RF signal. Based on this envelope monitoring, several detectors have been proposed [31, 32, 33, 34], to obtain different DC or low-frequency measurements. For example, one basic idea presented in [32] and illustrated in Fig.24 is to obtain a DC value proportional to the absolute or the RMS value of the RF amplitude. This DC sensor output is obtained via several amplifying, rectifying and low-pass filtering stages.

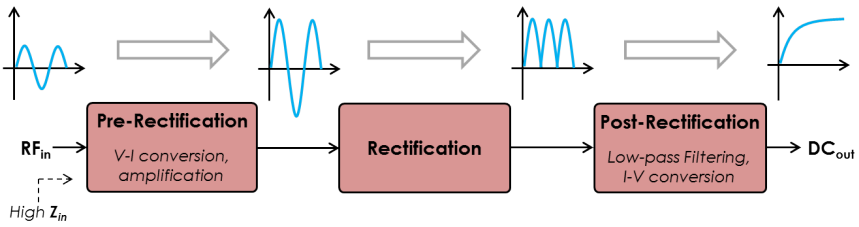

Figure 24: Illustration of amplitude detection based on envelope monitoring

THERMAL MEASUREMENTS When an analog/RF circuit is functioning, it dissipates power at both DC and RF levels, which causes a heating process around its different components. The underlying idea of thermal measurements is then as follows: if we can monitor temperature variations caused by this heating, we may be able to deduce the circuit's power consumption, and so estimate some of its RF performances. Towards this objective, thermal sensors have been proposed as a signature for monitoring circuit performances [35], as illustrated in Fig.25. Indeed, the temperature of a device (MOS transistor for instance) is directly related to its dissipated power. It can then be proven that by monitoring the neighboring temperature of the device, we can obtain an image of its power consumption and, thereby, an image of its performances non-intrusively. We can set proper test margins to discriminate defective from fault-free dies.
The non-intrusive property is a main advantage, especially for RF circuits, since usual "intrusive" sensors (like envelope detectors or DC probes) tap into the signal path of RF circuits and degrade their performances.

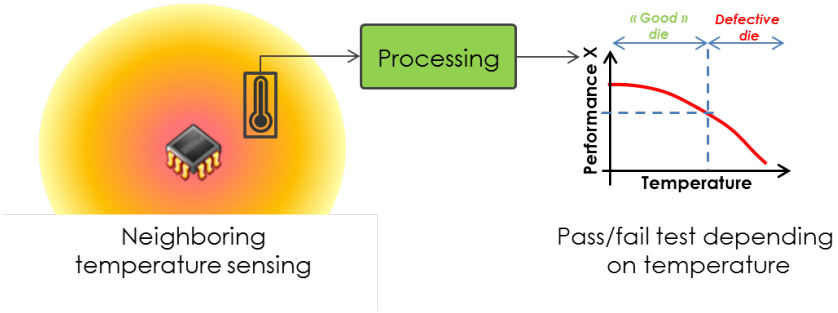

Figure 25: Illustration of non-intrusive thermal sensing as BIT measurements

The principle of thermal testing has also been demonstrated on RF devices, like PAs [36] and LNAs [37], targeting defect-oriented testing and monitoring of RF power levels, gain, linearity or even time domain characteristics.

3.1.2.2 *Case study of specific BIT strategy: loopback test*

Loopback test is a methodology that can be applied for testing a complete RF transceiver [38, 39]. The basic principle is to connect on-chip the transmitter and the receiver in a loop configuration, as shown in Fig.26.

For testing purposes, the output signal of the PA is transferred to the input of the LNA, after being attenuated to be in the sensitivity range of the LNA. By comparing the transmitted and received signals, it is possible to identify defects in the overall transceiver chain. The main advantage is that no external RF stimuli is required to test the whole transceiver. However, some defects can be masked (for example a gain loss in the PA can be compensated by a higher LNA gain in the loop configuration) and, the transceiver must be carefully considered with the loopback connection, making the design more complex. The masking issue can be avoided by adding other BIT sensors, for instance RMS envelope detectors, in addition to loopback.

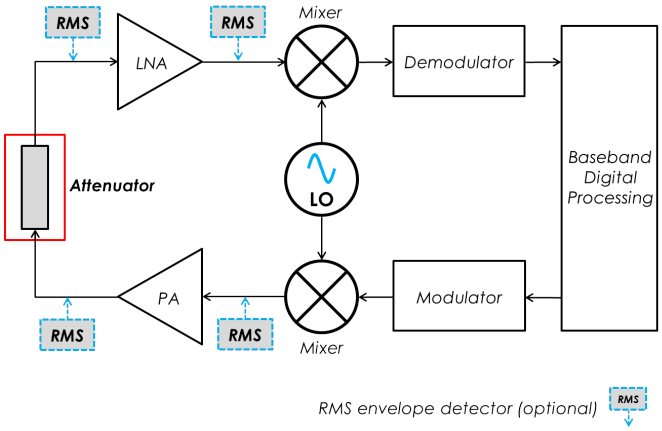

Figure 26: Illustration of loopback test

3.1.3 *machine-learning applied to the test of RF circuits*

3.1.3.1 *Introduction*

The initial objective of test is to detect defective from fault-free dies, without scarifying test quality, which is not straightforward. Using BIT, the test decision (pass of fail) has to be made solely based on *low-cost measurements*. In a nutshell, test engineers have to decide if the tested RF circuit has the necessary quality without explicitly measuring its RF performances [40]. In this perspective, test is based on the fact that process variations will affect both the DUT performances and its associated BIT sensor measurements, as illustrated Fig.27.

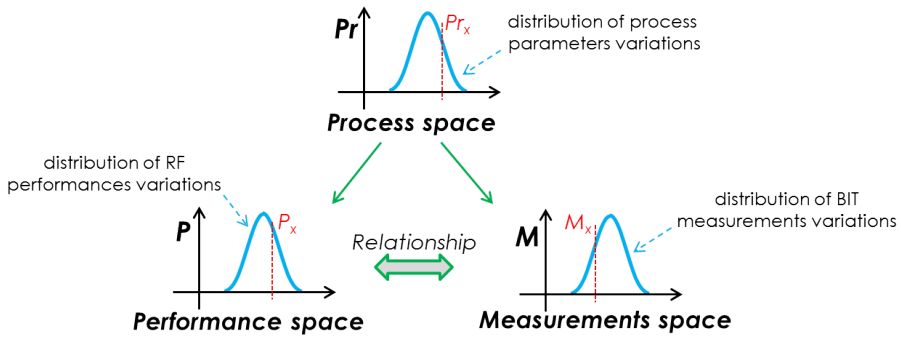

Figure 27: Illustration of the influence of process parameters on both RF performances and low-frequency BIT measurements

Process variations that affect the DUT are included into a given process variation space Pr. They will affect both the performances of the circuit and the low-cost measurements in two other variation spaces, respectively P and M. In particular, a given process variation Pr_x will be translated into a given DUT performance variation p_x and into a given measurement variation m_x. In order to make test decisions, i.e. to infer if the circuit is defective or not regarding performance specifications, machine learning can come to the rescue. Machine-learning will help to learn what is the direct relationship between P and M without explicitly measuring Pr.

3.1.3.2 *Machine learning strategies to test RF circuits*

The general aim of machine-learning is to use specific algorithms so as to estimate an unknown mapping between the inputs and outputs of a system from known input data [41]. Let us consider a system with two input variables M_1 and M_2 and one output variable OUT, as shown in Fig.28(a). In the case of circuit testing, M_1 and M_2 are two low-cost measurements extracted from the DUT. Moreover, OUT is a categorical variable that takes values of "pass" of "fail" identifying if the circuit is fault-free (it satisfies all specifications) or faulty (it violates at least one specification). Input and output variables are supposedly related together with a given mapping function $OUT = f(M_1, M_2)$ that has to be learned using a machine-learning algorithm. Towards this objective, different methods have been developed [40, 42, 43, 44, 45].
Generally, two main strategies can be used, namely classification-oriented test and prediction-oriented test. In both cases, supervised learning will be used. Supervised learning consists of teaching the algorithm what it will have to learn, by training it with appropriate data so as to achieve its learning objective, as illustrated Fig.28(c). For instance, the training data consists in a set of circuits, chosen to be the most representative of process variations. Both test strategies will be detailed in the following subsection.

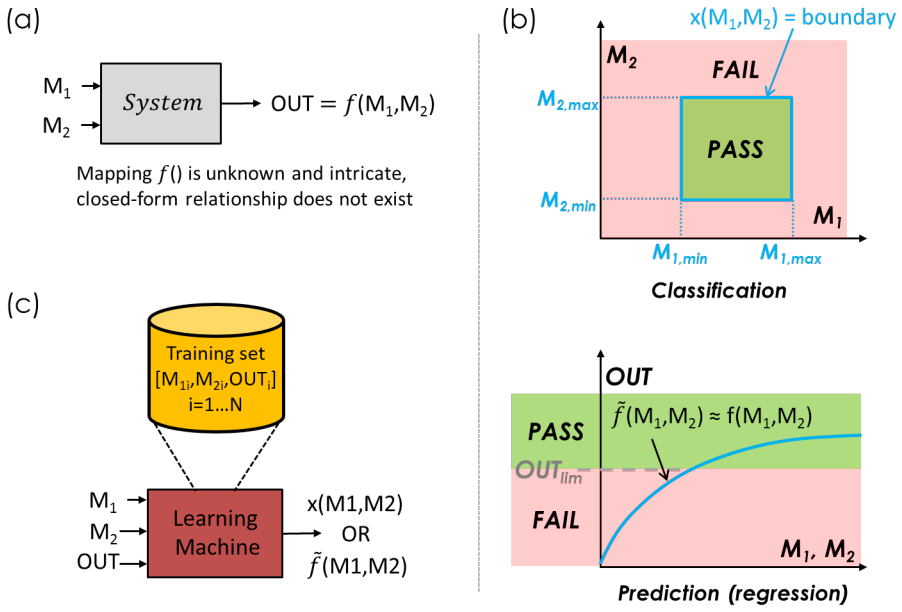

Figure 28: (a) example of a system with two input variables M_1 and M_2 and one output variable OUT, (b) Illustration of test results with classification and prediction algorithms, and (c) Illustration of training for a learning-machine

CLASSIFICATION-ORIENTED TEST In classification-oriented test, values of the output variable are decomposed in several classes, and the aim of the machine learning is to classify the incoming inputs values into one of the different output classes. In this strategy, f is not estimated explicitly, instead the algorithm use a decision boundary to classify incoming circuits with the less probability of misclassification [41]. More particularly, OUT is divided into two classes, "pass" or "fail" as explained previously. As presented in Fig.28(b), a decision is made depending on the measured input values M_1 and M_2. If both M_1 and M_2 are

into an acceptable range ($M_1 \in [M_{1,min}; M_{1,max}] \cup M_2 \in [M_{2,min}; M_{2,max}]$), then the circuit is classified as "pass". Unless it is classify as "fail".
Classification-oriented test flow is divided in two phases, namely training and industrial test (or production test), as illustrated in Fig.29. Firstly, the training phase aims at evaluating the relationship between the low-cost measurements and the performances of the DUT, in order to define the test boundaries. This is achieved by using a training sample of dies, and measuring both the RF performances and the associated low-cost measurements, compared to the circuit's specifications. In the industrial test phase, pass/fail decision is made by evaluating if the DUT footprint lies within or outside of the boundary.
The classifying process can be achieved with several algorithms, as the nearest neighbors, neural networks, decision trees... Then, in the industrial test phase, only low-cost measurements are measured on each new device, and pass/fail test is performed by evaluating if the device lie within the boundary specified in the training phase.

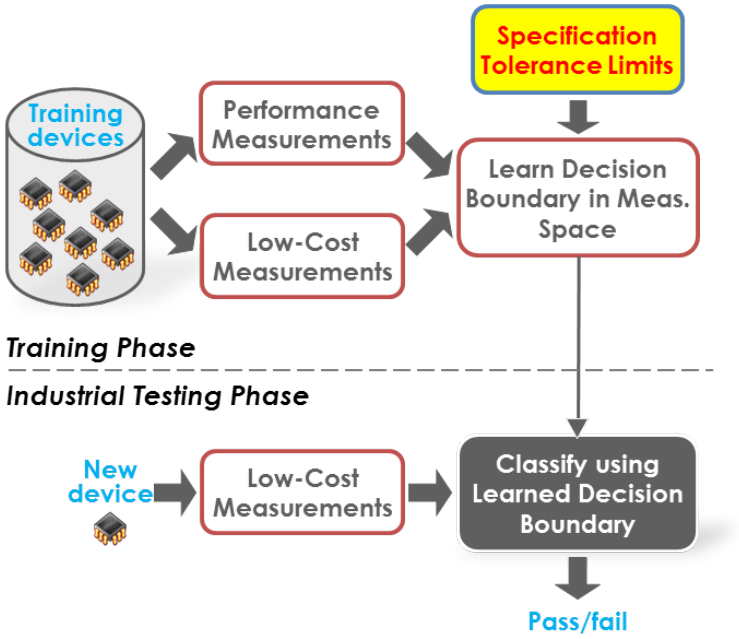

Figure 29: Illustration of classification-oriented test

In the classification-oriented strategy, defective and fault-free dies are determined without any information about their actual RF performances. Even if this reduces the test cost, we have no information about how far the performances are from the initial specifications [40]. Instead, test engineers are usually working with abstract quantities, representing for example linear combinations of low-cost measurements.
For these reasons, prediction-oriented strategy, more commonly known as *alternate test*, has been implemented.

PREDICTION-ORIENTED TEST: ALTERNATE TEST Considering alternate test, the goal of machine-learning is to explicitly estimate the relationship between M_1, M_2 and OUT, which means predicting the values of OUT for a given M_1 and M_2. This is done, as shown in Fig.28(d), using a regression model to map the inputs to the outputs and estimate f. In this case, the estimated function $\tilde{f}$ represents the relationship between the low-cost measurements

M_1 M_2 and the RF performances OUT. The regression function $\tilde{f}$ has to approximate f as accurately as possible to have a smallest possible prediction error [41].
Alternate test has initially been proposed in [40] and its test flow, illustrated Fig.30, contains also training and industrial test phases.

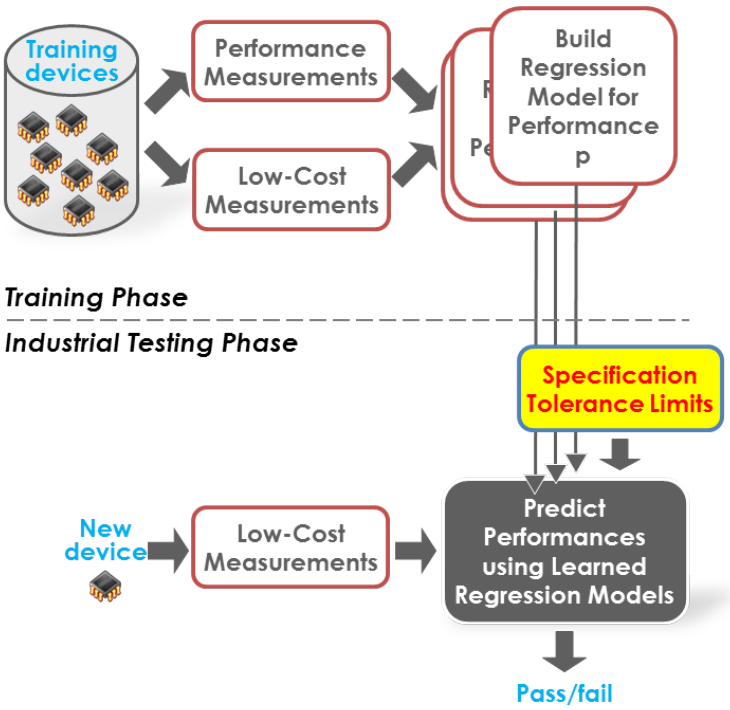

Figure 30: Illustration of prediction-oriented test strategy, or alternate test

During training, the regression model is build to map the RF performances to the low-cost measurements. In this case, an individual model is build for each performance (a number p of performances will lead to p different models). These regression models can be realized with several algorithms, among others Multivariate Adaptive Regression Splines (MARS), neural networks, polynomial regression, etc. Then, during the industrial test phase, only low-cost measurements are measured, each performance is predicted and pass/fail decision is finally made.

DEFECT FILTER FOR ALTERNATE TEST During the training of the machine-learning algorithm, some precautions have to be taken. Indeed, some of the circuit might have an different behavior, due to physical defects in the circuit itself introduced during the manufacturing process. Those circuits are known as *outliers*. Outliers must be removed previously to the training phase, because they are inconsistent with the statistical nature of the data that has to be analyzed by the machine-learning algorithm. They could then misguide the algorithm and degrade the performance of the obtained regression function. This issue can be prevented by using a defect filter, as the one presented in [46]. By using this defect filter, samples that exhibit a different behavior, i.e. that belongs to another statistical distribution, will be filtered out previously to training.

3.1.3.3 *Examples of alternate test applied to RF circuits*

Alternate test has largely been applied to RF circuits, with a lot of case studies (LNA, PA, mixers...). First of all, the most appropriate low-cost measurements have to be selected. This is equivalent to say that for performing the regression, we have to select the measurements that exhibit the highest correlation with the performances that have to be monitored. In this objective, several works have proven that alternate test can be used to predict many RF performances, using different sensors. For example, RMS envelope detectors can be placed at several locations of the RF receiver to predict global system performances, and individual performances of each RF block [47]. The implemented algorithm includes a selection of the best measurements so as to optimize prediction results. Results shows that the main performances of an LNA and a mixer can be accurately predicted. In [48], DC sensors, known as *function-mapped* sensors, are presented, specifically dedicated to alternate test. They have a simple architecture, and can be more easily integrated into the alternate test framework. Those sensors are used to accurately predict some RF performances of an LNA. Alternate test can be also successfully used for monitoring mixers, as in [49]. In particular, an envelope detector is integrated into the alternate test framework and successfully predict its IP3 and gain.
Feature selection process is also a major issue that has to be taken into account. This process consists of selecting the features (i.e. available low-cost measurements) that exhibit the highest correlation with the performances that have to be monitored. In [50], a methodology is proposed to select the test features that obtain the best correlation with targeted performances.

NON-INTRUSIVE ALTERNATE TEST Usual BIT and alternate test techniques suffer from a major drawback: they are (more or less) intrusive for the DUT. Indeed, inserting BIT sensors is tapping into the RF path of the circuit, which can degrade its performances. In order to avoid this degradation, non-intrusive test methodologies, based on alternate test, have been proposed. They consist of using sensors that have any contact with the DUT to perform low-cost measurements. Two types of sensors have been presented, namely thermal sensors [36, 37, 51] or *non-intrusive variation-aware sensors* [52]. Thermal testing has been demonstrated on LNAs and PAs for example. Non-intrusive alternate test has also been demonstrated on a 2.4GHz LNA, and will be seen in more details in the next chapters, since it will be used in this thesis work.

3.2 CALIBRATING RF CIRCUITS AFTER MANUFACTURING

Once RF performances are extracted or estimated, the next step towards adaptive RF circuits is to calibrate them. In this manuscript, calibration is referred to the tuning of RF performances after the fabrication of the circuit so as to compensate for process variations. As seen previously, process variations are more and more important for modern CMOS technologies, and have to be considered to enable fabrication of SoCs with embedded RF transceivers in the most advanced CMOS technologies. As in the previous section, the starting point chosen here is calibration based on ATE, which will nevertheless lead to a prohibitively solution. Several alternatives have been proposed in the literature to reduce the cost and are described in the following subsections.

3.2.1 *Modifying the performances with tuning knobs*

Calibration will only be performed if we are able to tune the performances of the DUT. This is achieved by inserting *tuning knobs* so as to render the design flexible. But the selection process of tuning knobs has to follow several rules.
First of all, only slight modifications can be tolerated in the topology of an RF circuit in order to enable test and calibration while maintaining an expected level of performances. Considering this aforementioned issue, tuning knob are chosen to have a minor impact on the RF circuit architecture. They typically consist of currents and voltages and can act on the different performances of the circuit. However, considering RF circuits, the different performances are not independent from each other. For instance, as seen in the previous chapter, gain, efficiency and linearity have to be traded-off in PA design. In addition, the dependencies between performances are very complex and usually non-linear. Acting independently on the performances is something very challenging to accomplish in RF design. If this is not achievable, the relationship between tuning knobs and performances will have to be taken into account, with several assumptions in some cases. This will be detailed in the following subsections.

3.2.2 *Straightforward calibration with ATE*

The most straightforward solution to perform post-manufacture RF circuit calibration is to use an iterative test/tune loop procedure, based on an ATE, as illustrated in Fig.31. In the test phase, the RF performances of the DUT are first extracted through the ATE. If these performances do not satisfy the specifications, then tuning knobs are adjusted in the tuning phase so as to modify the RF performances of the DUT. The test/tune procedure is then iteratively repeated until an optimal tuning knob configuration is found and all performances lie within the specifications.

DISCUSSION This procedure has the main disadvantage that test is repeated a each iteration of the calibration procedure, even so testing the RF functions with ATE only once is already responsible for a large fraction of the overall production cost, as discussed in section 3.1.1. This solution is therefore very costly. Cost-effective post-manufacture calibration methodologies can then be broadly categorized into two types:

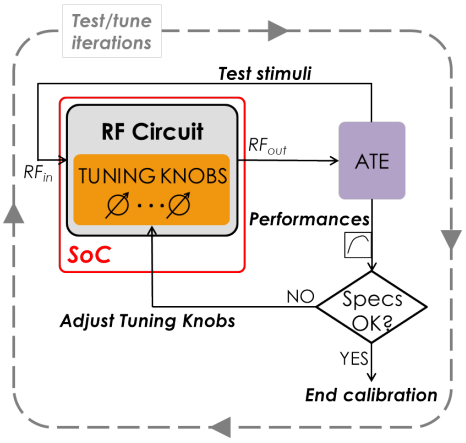

Figure 31: Iterative calibration setup with an Automatic Test Equipment

1. Performing calibration during production test on the ATE, using low-cost test techniques
2. Integrating the complete calibration mechanism on-chip so as to let the circuit autonomously perform its own calibration

3.2.3 *Post-manufacturing calibration at production test stage*

In the production test stage of a circuit, test cost can be reduced by the use of low-cost test strategies, like BIT and alternate test, as seen in section 3.1. A general view of this strategy is illustrated Fig.32.

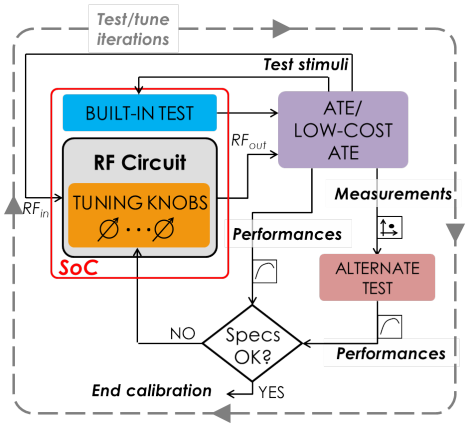

Figure 32: Illustration of cost-effective calibration methodologies

The basic approach is to use the principle of ATE-based calibration presented in the previous section. Indeed, an iterative test/tune scheme can be set up, but the difference here is that each test is performed at a lower cost with BIT and/or alternate test, which greatly reduces the overall calibration cost. A low-cost ATE can be used, since it does not have to measure any RF signal. However, even if cost per iteration is reduced, the total number of iterations has still to be minimized. In this objective, several works have been presented [53, 54, 55, 56, 57, 58].

SIGRATURE-DRIVEN TUNING For example, [55] proposed a methodology known as *signature-driven tuning*, in order to calibrate a complete RF transceiver. It is based on a comparison between an observed DUT envelope measured with an on-chip detector and a "golden" envelope, that corresponds in a nominal design. In this work, alternate test is used previously to calibration to optimize the test response of the system to a defined stimuli. This means that alternate test is only used to obtain an envelope response that correlates the most with the actual monitored performances so as to be able to accurately track the process variations. Then, the actual calibration procedure is an iterative optimization. The envelope response of the DUT is first measured for a first tuning knob setting and a cost function is computed. This cost function measures how far the DUT response is from the golden response. Then calculations are performed to optimize the search direction of the algorithm and to obtain a new tuning knob setting. The overall procedure is repeated until the optimal tuning knob setting is achieved (i.e. difference between golden and measured envelopes is minimal). Since the search direction is optimized at each iteration, the total number of iterations of the calibration procedure is minimized.
Power consumption can also be taken into account in the optimization algorithm [56], in which case it becomes a constrained algorithm. Indeed, on top of the calibration procedure, the algorithm includes features to find the best combination of tuning knobs that (a) satisfy the performance requirements and also (b) has a minor impact on power consumption.
An extension of *signature-driven tuning* has been presented in [57] and is relying on behavioral models. In this method, low-cost measurements are performed with an envelope detector only for a few tuning knob settings (3 or 5 depending on the prediction method used afterwards) and are varied such as the test response of the DUT and the behavioral model match. Then, this model is used to predict the RF performances of a transmitter. The problem is decomposed by individually predicting the performances of the mixer and the PA instead of the whole transmitter which results in less complexity. In the meantime, power consumption is taken into account, measured with dedicated current sensors, so as to choose the tuning setting that minimizes power consumption. Best tuning knob setting can then be predicted using an optimization algorithm that accounts for both power and performances.

CALIBRATION WITH BAYESIAN MODEL FUSION The modeling between performances and tuning knobs for efficient calibration is also very challenging. Towards this objective, alternate test can also be coupled with another algorithm, called Bayesian Model Fusion (BMF) [58], in order to improve the modeling. The basic principle of BMF is that data collected at the pre-silicon (schematic and post-layout simulations for example) and post-silicon (pre-production and production fabricated dies for example) stages are correlated together, even if they are not exactly identical [59]. Based on this observation, it is possible to improve post-silicon training by using pre-silicon results into the model. In other words, simulation results are included in the prediction algorithm, that only needs to be "updated" with a few data from silicon measurements. More prediction accuracy can be achieved, which greatly helps the calibration process.

DISCUSSION Even if the number of iterations is reduced and the modeling approach is improved, the process is still iterative. Performing several tests will increase the total test time, even if the cost of each test is lower than that of standard specification-based test ran on ATE. In the case of [57], it can be noted that testing and tuning phases are not iteratively performed. Instead, several tests are performed in a unique test phase, in the beginning of

the calibration procedure, for different tuning knob settings. Data is then used afterwards to find the best tuning knob setting, but test is not repeated. However, since several tests are required for each CUT, the procedure is still time consuming and costly, considering that the cost of the test is largely greater than the cost of tuning. For these reasons, improvements are still necessary.

3.2.3.1 *One-shot calibration*

Further cost reduction can be achieved by relying on a pre-trained model to infer the optimal tuning knob values directly from the result of a single test step in one-shot without needing to enter a test/tune loop, as shown in Fig. 33 [60, 61, 62].

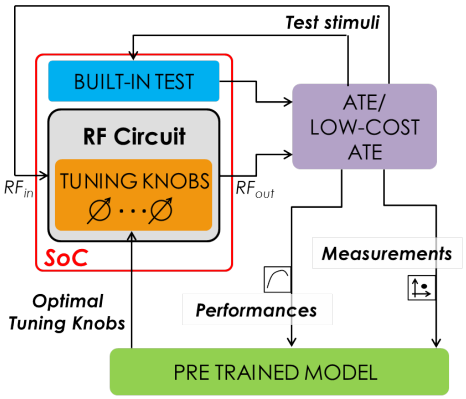

Figure 33: Illustration of one-shot calibration methodology

The basic principle of these methodology is to perform calibration in a unique test/tune iteration. Calibration time can be greatly reduced since test is performed only once. However, this methodology requires to infer, by performing only one test, the best combination of tuning knob that calibrates the performances. In this objective, several works have been proposed.

A first strategy is to simplify the learning process used to predict the performances. In [60], it is proposed to implement a one-shot calibration procedure for calibrating a LNA. The principle is to enable a feedback loop into the LNA under test forcing it to oscillate. Based on the obtained oscillation frequency, two tuning knobs (a varactor at the supply node of the LNA and a control of the LNA bias current) are set so as to calibrate two LNA performances, gain and 1-dB compression point. The particularity of the method is that tuning knobs are chosen to present a linear relationship with the performances. In this way, we can rely on a linear pre-trained model which maps the measured performances to the appropriate change that needs to be applied to the tuning knob values.

In [61], it is proposed to calibrate a 1.9-GHz LNA, with the information extracted from a envelope detector. The tuning knobs used for calibration are the two bias voltages of the LNA. The basic calibration procedure is as follows: the performances of circuit instances in a representative training set are first measured and, in addition, each circuit instance is calibrated manually, in order to identify appropriate tuning knob values. Using this data, a regression model is trained to infer the appropriate tuning knob values directly from the measured performances. Thereafter, this pre-trained regression model can be readily used for calibration, by measuring only once the output of the envelope detector.

PERFORMANCE POST-MANUFACTURE CALIBRATION OF RF CIRCUITS In [62], the developed methodology proposes to use an extended version of alternate test presented in section 3.1.3.2 by including tuning settings into the regression models. The framework is illustrated Fig.34.

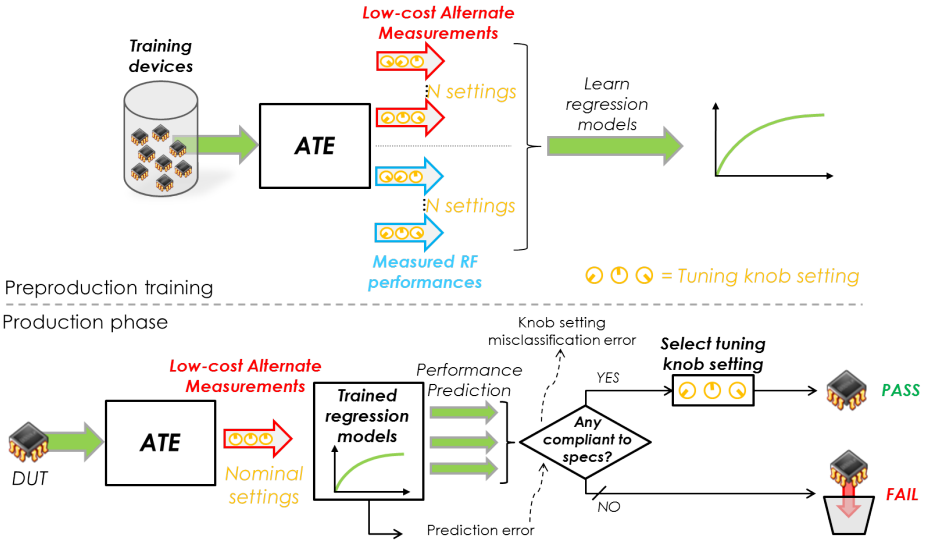

Figure 34: Illustration of on-chip calibration strategy

In more details, the calibration employs a pre-trained regression model that predicts directly the performances based on built-in low-cost alternate test. However, in contrast with previous work, training phase is modified by including both the low-cost alternate measurements and different tuning knobs settings, so as to be able to predict the RF performances of the circuit across both process and tuning variation spaces. Here, the measurements are aimed at offering an *image* of the process variations. Tuning knobs are not specifically designed to act independently on the performances, but their interactions with performances are directly taken into account into the model. Moreover, it is assumed that process variations and tuning knobs orthogonally act on the performances. Thus, a composite regression model can be built without including interaction terms of measurements and tuning knobs. Consequently, once training is performed each DUT is tested a first time with the nominal tuning knob setting to evaluate if its performances satisfy the specifications. If not, it has to be calibrated, and the optimal tuning knob values can be found by running an optimization using as an underlying function the regression model. The measurements are obtained once for the nominal tuning knob setting and they are fixed during the optimization.

3.2.4 *On-chip calibration strategies*

The complete calibration mechanism can also be integrated on-chip, as illustrated in Fig.35.

In this case, built-in test measurements that reflect the state of the performances are digitized and forwarded to the digital signal processor (DSP) where an algorithm finds a best guess of tuning knob values. This approach may still remain iterative, but the test/calibration loop is much faster since it is performed entirely on-chip without relying on external ATE.

First applications that were developed are more *ad-hoc*, such as controlling the input match of an LNA, alone [63, 64] or together with its gain [65]. Those methodologies usually in-

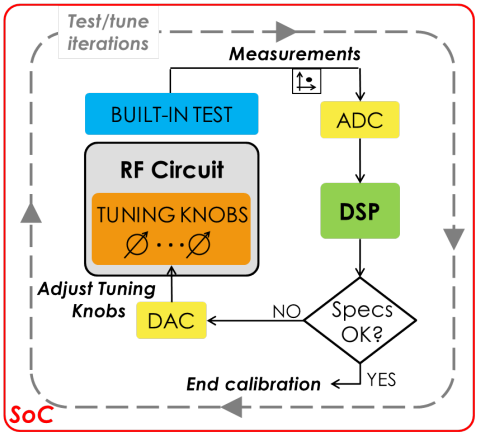

Figure 35: Illustration of on-chip calibration strategy

volve tunable matching components (inductors, varactors) and iterative test/tune approaches. Based on the same architecture, more global strategies have been demonstrated [66] for different RF circuits. Some of the techniques described previously in section 3.2.3 can also be implemented on-chip, as the technique in [56] for example.

SELF-CALIBRATING PA (HEALICS PROGRAM) Under the framework of Self-HEALing mixed-signal Integrated Circuits (HEALICs) program, a complete self-calibrating SoC has been developed [66]. The SoC contains "self-healed" ADC, PLL, microwave/millimiter-wave transceiver and PA. Here, we will focus on the self-healed PA presented in [8]. The global architecture of the PA is illustrated in Fig.36. The PA is designed with a specific 28-GHz two-stage 2-1 power combining class-AB topology.

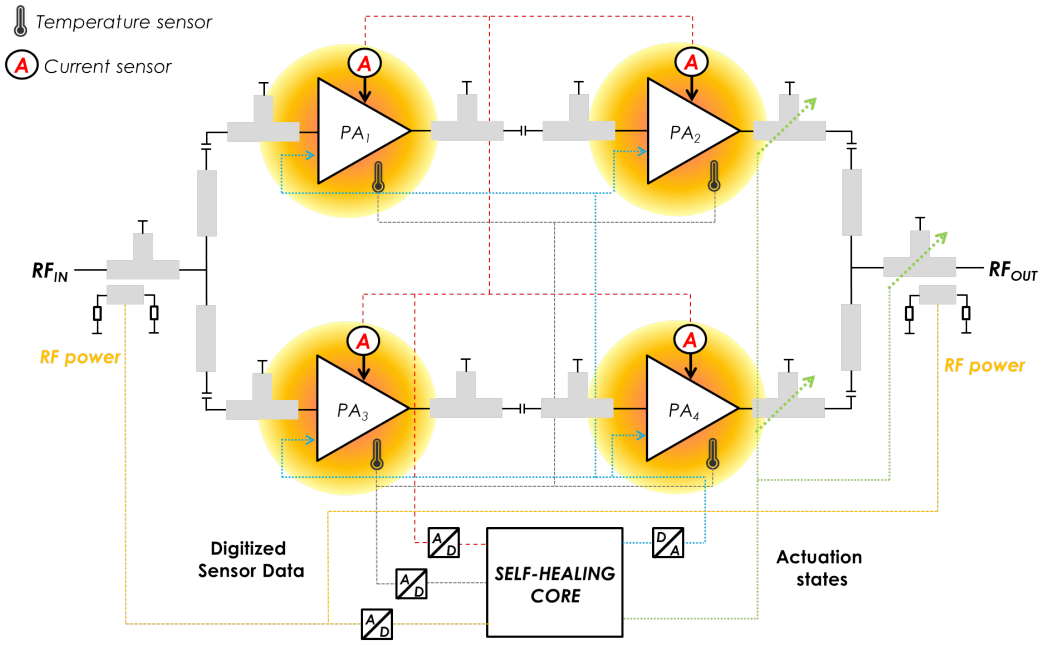

Figure 36: Illustration of self-healing PA proposed in [8]

The global aim is to tune different performances of the PA. On one side, several sensors are used to extract the different RF performances on-chip. Firstly, two 50-ohm coupled transmission lines sense the RF power at input and output of the PA and are used to measure load mismatches. Then, a dynamic current sensor, designed together with a power supply regulator is used to measure DC power consumption. Finally a thermal sensor, designed with interposed diodes in between the different fingers of the PA MOS transistors can track the power dissipated by the PA. On the other side, two types of actuators are used, namely DC control of bias currents of the PA stages to tune RF performances and tunable transmission line stubs at the output of the PA to modify its output impedance.
To control the tuning strategy, a digital algorithm has been set up, with two different algorithms to (a) maximize output power and/or (b) minimize DC power consumption for a fixed output power. The algorithms are implemented in a custom digital block on-chip and consist of global state machines that find the optimal tuning solution among all possible states. Finally, data conversions are performed on-chip with specific low-power converters. This strategy is used to increase the output power of the PA and/or minimize DC power consumption.

3.2.4.1 *Summary - classification of calibration methodologies*

As we have seen in the previous subsections, many different calibration strategies can be developed. We propose to classify these methodologies with several criteria (the list is non-exhaustive), as summarized in Table 3. First of all, calibration can be performed iteratively,

CALIBRATION	Type of strategies		
What is the type of calibration?	Iterative	”Minimally-iterative”	One-shot
How the calibration is performed?	Off-line	”On idle times”	On-line
Where does the calibration takes place?	Off-chip	”partially on-chip”	On-chip
Is the calibration power conscious ?	YES		NO

Table 3: proposed classification of calibration methodologies

or in one-shot, which is a first classification criterion. An intermediate category considers methodologies that minimize the number of iterations, which are called ”minimally iterative”. Then, this calibration can be performed on-line (i.e. during the operation of the circuit), in idle time or off-line (at post-manufacturing testing for example). The calibration can take place on-chip (totally or partially) or off-chip. Finally, the calibration can be power conscious or not, depending if calibration of RF performances is done by limiting the power consumption of the circuit.

3.3 ADAPTING RF CIRCUITS WITH DIGITAL ALGORITHMS

As seen at the end of Chapter 2, RF designs require more and more adaptation strategies to be able to satisfy all requirements. Based on the different techniques mentioned, it is possible to integrate complete adaptation algorithms that are able to monitor a large set of RF performances. With the evolution of SoC that include more and more powerful DSP, the introduction of these type of techniques is easier. They can include calibration (i.e. compensate for process variations) but are not limited to it. They are able to compensate also for environmental variations, aging, or to reduce power consumption. Some of the proposed techniques are reviewed in the next subsections.

3.3.1 *Adaptation against aging and/or temperature*

The aging of a circuit is translated into a slow degradation of the performances over time. This issue is usually not taken into account specifically, but can be handled with some of the calibration methodologies previously detailed. In particular, if the calibration is entirely performed on-chip, like in [56, 66], it can be used during the lifetime of the circuit, to periodically sense the performances and correct the system accordingly. As an example, the sustainability of 60-GHz millimeter wave (mm-wave) transceivers is a major concern, tackled in [67] with a self-healing loop that can monitor the performances throughout the lifetime of the circuit.
In particular, the PA is also a very challenging block at mm-wave and adaptive solutions have been proposed in [68] by implementing a two-level adaptation. The first level is a "local" adaptation loop designed in hardware with simple circuits. The aim is to perform an amplitude and phase compensation to improve EVM without the use of an adaptation algorithm. Amplitude is controlled with a simplified version of an envelope detector that dynamically adjusts the biasing of the PA to maximize linearity in real-time. Phase compensation is performed by controlling the gate-to-source capacitance of the MOS gain transistor of the PA, thus limiting phase distortion. The second level is a more global adaptation with a self-healing loop to adapt the PA gain and power consumption, against process, aging and temperature.

3.3.2 *Adaptation against channel conditions*

Channel conditions are critical for the proper working of RF circuits. As seen in Chapter 2, worst channel conditions lead to degraded communication and best channel conditions lead to wasted power. In order to adapt the RF transceivers to changing channel conditions, more elaborated strategies than Adaptive Channel Modulation (ACM introduced in chapter 2) have been presented, based on a dedicated digital algorithm. They concern the PA of a transceiver [2], a complete transceiver [69] or a complete receiver [69, 70]. The main objective of these different works is to be able to reduce power consumption while maintaining an expected level of performance that depends on channel quality.
The PA of the transmitter is the block of an RF transceiver that consumes the most power. Moreover, it is responsible of the final signal transmission on which depends a non negligible part of the communication quality. Targeting OFDM applications, a channel-adaptive PA has been presented in [2]. The adaptation is carried out in two ways. Firstly, Peak-to-Average Ratio (PAR) of digital OFDM signal is dynamically reduced with *companding* (a

signal-processing technique) and PA is adaptively re-biased. Since envelope amplitude is reduced, the PA can operate with less power back-off, linearity requirements are relaxed, and power can be reduced. Secondly, the quality of the channel is estimated through an EVM measurement. Indeed, PAR reduction induces more noise, which increases EVM and degrades the communication quality. Feedback adaptation of the PA is then required to ensure (a) less power consumption (depending on companding level) and (b) sufficient communication quality by keeping EVM below a given threshold. The overall power consumption of the system can be reduced under favorable channel conditions (up to 5.5x compared to non-adaptive transceivers).
Considering receivers, [70] has proposed an adaptive framework based on an orthogonally tunable LNA. Orthogonal tuning aims at obtaining an independent control of several RF performances or groups of RF performances, in order to enlarge the performance space that can be searched by the tuning algorithm and find a "more optimal" operating point. For instance, in this work, linearity can be tuned independently of the group (gain and noise). EVM is measured to evaluate channel quality and the LNA is tuned accordingly so as to operate below a given EVM threshold that guarantees the lowest power consumption while maintaining the performance level expected for each channel condition.

PROCESS-TOLERANT CHANNEL-ADAPTIVE RF RECEIVER In [6], a process, channel and performance adaptive RF receiver is proposed. Several strategies have been combined to obtain a more global adaptation of the receiver, where the LNA and the mixer are simultaneously tuned. The process is illustrated in Fig.37.

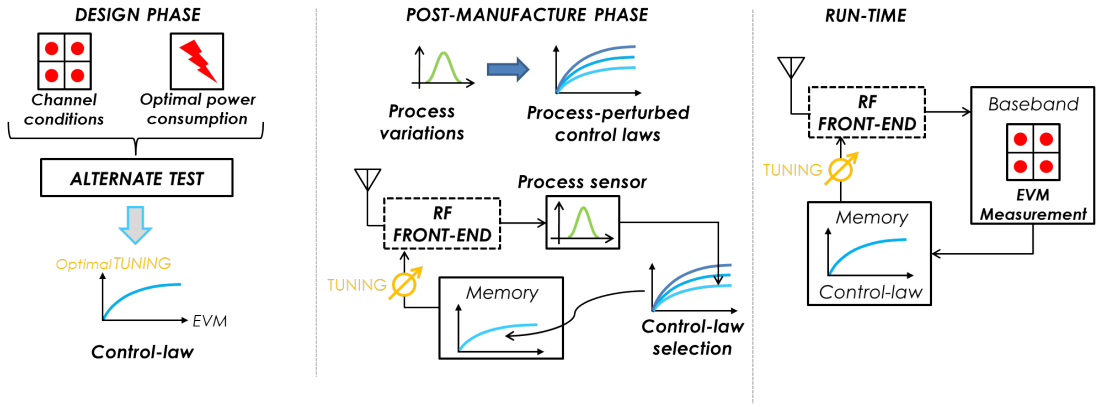

Figure 37: Illustration of adaptation strategy proposed in [6]

Channel adaptation consists of evaluating the best operating power state regarding EVM measurement. A control law (locus) is built at the design stage to map the optimal tuning knob settings (here the bias voltage of the LNA and the power supply of the mixer) for several channel conditions. Optimal setting is defined as the tuning state that maintains EVM below a given threshold and minimizes power consumption. This law is stored (as a look-up table) into the chip and used at run time into a feedback loop to optimize power consumption based on the EVM measurement. However, because the circuit is affected by process variations, the previously built locus is not process-independent.
A specific process adaptation procedure is thus performed with alternate loopback test. Since the receiver is usually coupled with a transmitter, it is possible to use this transmitter to tune the receiver. However, the transmitter needs also to be tested to ensure that it is not defec-

tive. This test is performed with an alternate test framework, by estimating the performances through an envelope detector. The transmitter can be calibrated if performances are not satisfied, with dedicated tuning knobs (the power supply of the PA and pre-distortion parameters). Then, the transmitter is used to test and calibrate the receiver. At the design phase, the locus discussed for channel adaptation is computed for different process-perturbed devices. We obtain then a set of locus functions, measured with loopback test for different process variations. Calibration consists of preforming loopback for the considered device and identifying the process-perturbed locus that is better adapted to the circuit. Only this locus will be programmed into the chip and used for channel adaptation.

Recently, in [69], adaptation has been extended to a complete Multiple-Input Multiple-Output (MIMO) transceiver. The adaptation principle is similar to the previously mentionned works, with EVM monitoring and evaluation of the best tuning setting over encountered channel conditions. Optimal tuning settings for every channel condition are stored into a look-up table and used in real-time. In order to achieve more power savings, several modes are defined, namely Data Priority (DP) for which the highest throughput is targeted for given channel conditions and Energy-Priority (EP) that targets the lowest energy-per-bit. This operation is performed on top of the EVM monitoring.

3.3.3 *Specific adaptation to reduce power consumption*

Power adaptation algorithms have been proposed for RF receivers, as in [7]. As seen in the previous chapter, power can be traded-off with RF performances (for instance noise and linearity for a receiver). However, classical adaptation schemes cannot achieve a two-dimensional parameter adaptation, which means having a flexible configuration of noise and linearity to adapt power consumption. In [7], an algorithm constantly monitors the Signal-to-Noise Interfers Ratio (SNIR) with a control loop based on the following phases: (a) measuring the SNIR, (b) estimating the interferers power and (c) calculating the required values of linearity and noise. SNIR is measured in the digital demodulator at baseband. If SNIR is greater than required, tuning knobs are tuned so as to decrease power consumption. If SNIR is lower than required, power consumption is increased. The receiver contains an adaptable LNA, mixer, a power-scalable operational amplifier and an adaptive analog baseband.

ADAPTIVE LOGICAL CONTROL FOR EFFICIENT ENERGY CONSUMPTION Another approach is to rely on models to build an adaptive control loop. For instance, in [9], alternate test is coupled with power prediction through a behavioral model inserted in a control loop, as illustrated in Fig.38.

At the design stage, a set of training samples is selected and two models are computed. Firstly, a behavioral model of the circuit with parameters θ is built from transient simulations. Secondly, a regression model is built to map the set of parameters θ to the performances P of the DUT, using alternate test. At run time, the control procedure is as follows. Firstly, power supply VDD is set at its maximal value. Then, a low-cost envelope detector is measured, behavioral parameters are computed $\tilde{\theta} \approx \theta$, performances $\tilde{P} \approx P$ are estimated from the model and are then compared to pre-defined specifications. Specifications are defined for several power mode operations (stringent specifications for maximum power mode, relaxed

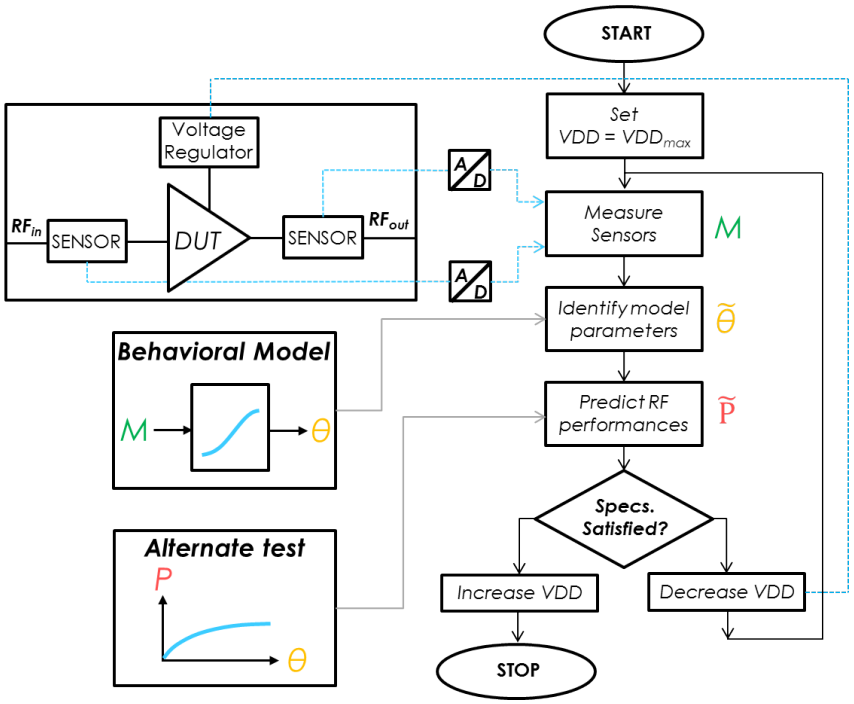

Figure 38: Illustration of power consumption control proposed in [9]

specifications for minimal power mode...). Knowing in which mode the circuit has to operate, the control algorithm iteratively evaluates if specifications are satisfied. If yes, then power supply voltage can be decreased (along with power consumption), and the procedure is repeated. If not, then power supply voltage is increased and the procedure stops. This adaptation procedure enables the control of the power consumption of an LNA for example, with pre-defined power modes, and can be done every time that a new power mode is required.

3.3.4 *Summary of proposed adaptation techniques*

In order to summarize this section, the adaptation methodologies can be classified depending on the type of variation they can compensate for, and the methodolgy that is being employed, as presented in Table 4.

3.4 OVERALL DISCUSSION

From our point of view, a global adaptation strategy should include target a post-manufacturing calibration, in addition to other parameters that affect the RF circuit (channel conditions, temperature, aging...). Since the impact of process variations are more and more important for modern CMOS technologies and affect their fabrication yield, we decided to first focus on low-cost and efficient post-manufacturing calibration. We have previously defined, in section 3.2, a classification of calibration methodologies presented in the literature. The general calibration strategy we targeted is highlighted in blue Table 5.
As we have seen, developing cost-effective calibration requires first to have efficient BIT sensors. However, for all aforementioned calibration techniques, the sensing of RF performances

ADAPTATION		Type of strategy	
Level of adaptation		By hardware	With dedicated algorithm
How is the adaptation performed?		On idle times	On-line
Does it integrate calibration?		YES	NO
Types of variations that are taken into account?	Temperature	YES	NO
	Aging	YES	NO
	Channel	YES	NO
Is it power conscious?		YES	NO

Table 4: Proposed classification of adaptation strategies

is always performed intrusively that is, with sensors that have an electrical contact with the RF path of the DUT. Even minimally intrusive, these sensing strategies may perturb the behavior of the DUT and degrade its performances. In this work, we employ non-intrusive sensors, enabling us to monitor the performances of RF circuits without any electrical contact with them. To this end, we added a new criterion to the classification of calibration methodologies in Table 5.

CALIBRATION	Type of strategies		
Type of	Intrusive		**Non-intrusive**
What is the type of calibration?	Iterative	”Minimally-iterative”	**One-shot**
Where does the calibration takes place?	Off-chip	**”partially on-chip”**	On-chip
How the calibration is performed?	**Off-line**	”On idle times”	On-line
Is the calibration power conscious ?	**YES**		NO

Table 5: Characteristics of our targeted calibration methodology

Since performances need to be predicted to be able to tune the circuit, classification-oriented test is not suitable. We have then chosen to use the alternate test so as to be able to predict RF performances depending on process variations and tuning settings. However, in current alternate test-based calibration techniques during post-production test, several limitations can be denoted. For instance, in [60] it is very challenging to accomplish a linear relationship between tuning knobs and RF performances by design in practice and, in addition, it is difficult to identify tuning knobs that act orthogonally on performances. In [61], the regression model does not include explicit information about process variations which may affect its overall accuracy and generalization capability. In [62], the error of the regression model can be large if the measurements vary nonlinearly across the tuning knob settings. Moreover, calibration based on alternate test-based frameworks can be iterative or one-shot, the latter significantly reducing the total cost. We then oriented our research towards one-shot calibration.

On-chip calibration approach is then very appealing, especially if we extend calibration to adaptation strategies in the field to account for aging and environmental effects. However, adding the feedback calibration loop on-chip unavoidably increases the area overhead and the overall complexity of the design and has to be taken into account. For these reasons, our first attempts towards a calibration methodology will not be performed totally on-chip and will be off-line.

Based on this state-of-the-art, the first aim of this work is to develop a calibration methodology, with the following constraints:

- **Use alternate test to enable calibration at low-cost.**
- **Develop a non-intrusive strategy so as to not degrade the performances of the DUT and leave the DUT intact.**
- **Perform calibration in "one-shot", that is in a single test/tune step to reduce test time and cost, while considering interaction between tuning knobs, performances and process variations.**
- **Consider power consumption as a direct variable in the calibration strategy to obtain a power-aware calibration**
- **Define a calibration strategy that can be further integrated into a more general adaptation framework and can extended to be "on-chip".**

Next chapter will go into further details about the methodology developed in this work.

4

POST-MANUFACTURING ONE-SHOT CALIBRATION OF RF CIRCUITS BASED ON NON-INTRUSIVE SENSORS

The different objectives that we defined at the end of chapter 3 for our calibration strategy will lead to specific choices, in terms of architecture and basic components used for calibration. The general objective of this chapter is to specify these choices and to introduce our own calibration procedure for RF circuits, which will be validated with simulation results as a first proof of concept.

Our choices with respect to the the state of the art will be summarized in section 4.1. Then, the basic principles and achievements of our methodology will be summarized in section 4.2. Section 4.3 will explain in details our proposed calibration framework. Then non-intrusive sensors which are key for the calibration methodology will be reviewed in Section 4.4. The design of the case study used for the first proof of concept, based on a PA, will be detailed in Section 4.5. Finally, Section 4.6 will explain in details the first simulating results in order to validate the calibration methodology.

4.1 INTRODUCTION

Existing calibration techniques can be classified with several criteria. As summarized in Table 6, we have chosen what we consider as the main features of calibration methodologies, namely (a) the number of iterations to obtain the optimal tuning knob setting, (b) the possibility of on-chip integration together with performing calibration on-line (e.g. adaptation), (c) the intrusiveness of the performance sensing strategy and (d) whether the methodology is power conscious or not. These features are displayed for many existing calibration techniques in the literature. Our proposed approach is one-shot and is simultaneously based on totally non-intrusive sensing of the RF performances. Moreover, our approach is power conscious but for the time being, will be performed off-line and off-chip.

Ref.	**Number of Iterations?**	**On-chip& On-line?**	**Intrusive?**	**Power conscious?**	**Comments**
[55]	A FEW	NO	YES	NO	The number of iterations is minimized
[56]	A FEW	YES	YES	YES	The number of iterations is minimized
[57]	ONE (*)	YES	YES	YES	One test for several tuning knob settings
[58]	MANY	YES	YES	NO	Use of Bayesian model fusion
[60]	ONE	YES	YES	NO	Linear tuning VS performances relationship
[61]	ONE	YES	YES	YES	Manual calibration for training
[62]	ONE	NO	YES	YES	interaction terms not taken into account
[63]	MANY	YES	YES	NO	exhaustive tuning knob search
[64]	MANY	YES	YES	NO	exhaustive tuning knob search
[65]	MANY	YES	YES	NO	exhaustive tuning knob search
[8]	A FEW	YES	YES	YES	Algorithms implemented in a custom block
This work	**ONE**	**NO**	**NO**	**YES**	**Use of non-intrusive variation-aware sensors**

Table 6: Summary of existing calibration techniques and their main characteristics

The constraints defined at the end of Chapter 3 are listed again below:

1. Develop a non-intrusive post-manufacture calibration strategy
2. Perform calibration in "one-shot", i.e. in a single test/tune step to reduce test time
3. Consider power consumption as a direct variable in the calibration strategy

4. Use our calibration strategy to be further integrated into a more general adaptation framework

The next section will summarize the solutions that we proposed to address these different points. As we shall see in this chapter, points 1 and 2 are closely related together. By using a new type of non-intrusive variation-aware sensors and an appropriate calibration algorithm, we will be able to achieve a true one-shot calibration strategy, which has not been developed before in the literature. Indeed, all previous work use intrusive sensors for testing the performances, and/or use some assumptions/simplifications in the prediction algorithm. Point 3 is achieved by considering power consumption as a performance to calibrate, in the same way than another RF performance. Finally, with point 4, we should take into account further improvements of the proposed methodology to be used on-chip and on-line.

4.2 BASIC PRINCIPLES AND OVERVIEW

4.2.1 *Introduction of proposed low-cost alternate calibration*

The development of low-cost post-manufacturing calibration methods has been closely related to the progress made in alternate test methodologies. We have chosen to improve upon an existing calibration paradigm for RF circuits, initially presented in [62], illustrated in Fig.39. Globally, calibration is performed during production test with minimum cost increase and aims to:

- Correct out-of-specification performances to recover yield loss
- Obtain a desired trade-off between performances and/or power consumption

Both goals are achieved while maintaining power consumption within a specified budget.

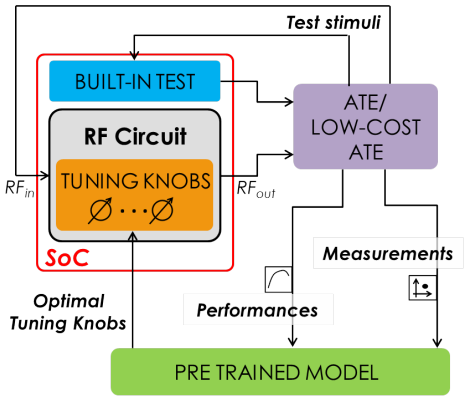

Figure 39: Illustration of one-shot calibration methodology

Calibration is enabled by judiciously inserting *tuning knobs* into the circuit. The tuning knobs add degrees of freedom in the design and can act on all the performances irrespectively. The calibration algorithm employs a *pre-trained regression model* that expresses the relationship amongst circuit performances, measurements that offer an "image" of process variations, and tuning knob values. This algorithm is based on the alternate test paradigm detailed in Chapter 3.
As we will see along this chapter, the key idea of our proposed methodology is that measurements are obtained on *non-intrusive sensors* that are not electrically connected to the circuit and, thus, are totally transparent to it. This non-intrusive property allows us to perform the calibration efficiently in *one-shot* using a single test step and optimizing tuning knob values quickly in software in the background, which greatly reduces the cost. In particular, thanks to the non-intrusive property, the sensor measurements stay invariant under changes of the tuning knobs. This means that we need to obtain the sensor measurements only once, plug their values into the regression model, and then optimize the resultant relationship between performances and tuning knobs without needing to repeat the sensor measurements for every tuning knob setting that is visited during the course of the optimization.

MATHEMATICAL FORMULATION OF THE PROBLEM The proposed calibration framework can be formulated mathematically. The RF performances of the circuit P are related to the process parameters P_r and the tuning knobs TK with a given function f:

$$P = f(P_r, TK). \tag{12}$$

Since it is not possible to measure directly the process parameters, we rely on a set of alternate measurements M. These measurements are aimed at sensing process variations, and they are then related to the process parameters with a function g:

$$M \approx g(P_r), \tag{13}$$

where the approximation accounts for the fact that the measurements may not reflect all process parameters. Substituting (13) into (12) gives

$$\begin{aligned} P &\approx f(g^{-1}(M), TK) \\ &\approx z(M, TK). \end{aligned} \tag{14}$$

The function z links the performances of the circuit with process variations (extracted with alternate measurements M), given tuning knob setting TK. However, z has an unknown and intricate closed form, thus we will use the alternate test paradigm and train a regression model to approximate it. More details about the training of the regression model are given in section 4.3.2.

4.2.2 *Enabling one-shot calibration*

The term "one-shot" refers to the fact that calibration is performed by testing the performances only once. Considering that we have obtained the regression function z as presented in section 4.2.1, the calibration will consist of two different phases:

1. Estimate initial RF performances of the circuit considering a nominal combination of tuning knobs.
2. If one or more performances fail the specifications, identify an appropriate tuning knob setting such that each performance meet the specifications

The necessary condition for obtaining point 2 in one-shot is that measurements M stay invariant under the changes of tuning knob setting TK. If this condition is satisfied, M and TK are completely independent from each other. We can then seek for the best value of TK, guided by the regression function z, knowing that M is constant and will not be affected.

To achieve this, we propose to use non-intrusive sensors to perform the measurements M. Indeed, as seen in Fig.40(a), if we use a typical alternate sensor, such as an amplitude detector for example, this sensor taps at the RF output of the DUT, denoted RF_{OUT}. Then, during the optimization of tuning knob settings, if we modify the value of TK, we modify the output of the DUT and thereby the output of the sensor as well. In this case, measurements of the sensor have to be updated for each visited tuning knob setting, which eliminates the possibility of "one-shot" calibration (the test is repeated for each visited TK), unless if some approximation is made. In contrary, by using a non-intrusive variation-aware sensor, as illustrated in Fig.40(b), there is no electrical contact between the DUT and the sensor. Then,

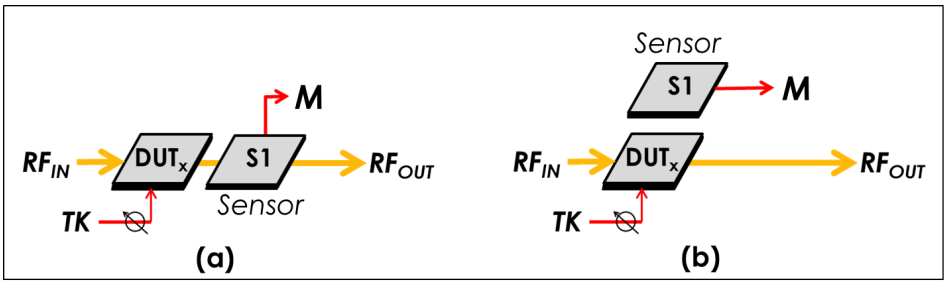

Figure 40: Illustration of different sensing methods to obtain measurements M: (a) usual "intrusive" sensor and (b) proposed "non-intrusive" variation-aware sensor.

if the tuning knobs of the circuit are modified, RF_{OUT} is affected but no modification is made at the sensor's output. This enables to visit several tuning knob settings so as to find the best combination and apply it to the circuit without measuring the sensor output multiple times. In other words, the value of M does not have to be updated in the model and stays constant for all TK.
Based on this sensing strategy, we can then set up a one-shot calibration procedure as will be detailed in section 4.3.

4.2.3 *Considering power consumption*

Some of the previously proposed calibration methodologies are explicitly considering power consumption. Most times [56, 57, 61, 62], power consumption is seen as an additional constraint for optimizing the tuning setting of the circuit. For example, if several tuning knobs satisfy all the specifications, the chosen tuning setting is the one for which power consumption is minimized. In [57], power is controlled by current sensing at the PA stage and used as a constraint within the optimizer.
Unlike previous works, we decided to integrate power consumption in the list of performances (i.e. as gain, linearity, etc). DC power consumption is then also predicted with alternate test and its optimization becomes an additional objective in the calibration algorithm, along with the objective of optimizing the trade-off of RF performances.

4.2.4 *Integrating a more general adaptation framework*

In order to integrate the proposed calibration methodology in a more general adaptation framework, it should be extended such that it can be performed on-chip and on-line. Although this is not the scope of this thesis, we investigated these goals and we envision solutions.

ADAPTING THE PERFORMANCES ON-CHIP The adaptation framework requires to perform the optimization of the performances on-chip. For this purpose, several solutions can be proposed. A first approach is to use Look-Up Tables (LUT) on-chip. LUT are in most ties used to store optimal tuning knob settings for every combination of on-chip measurements. The values contained into the LUT are coming from off-line training and usually fixed during the lifetime of the circuit. The adaptation then consists in measuring the sensors' outputs, searching for a corresponding optimal tuning setting in the LUT and tune the circuit accord-

ingly. Look-up tables are widespread in previous work as explained in Chapter 3. The main advantage is that calculations are avoided on-chip, which reduces the cost of the adaptation circuit. However, this solution does not offer a lot of flexibility since the results are stored only once in the look-up table.
In order to offer more flexibility, all calculations can be performed on-chip. For instance, a DSP can be used to embed the complete adaptation algorithm. If a DSP is already available in the system, it can be readily used to perform the calibration, and during lifetime to adapt the performances. Moreover, it offer a high flexibility since the algorithm could be run in real-time when it is necessary. However this solution can be very costly if the DSP has to be added on purpose.

BUILDING VARIABLE TUNING KNOBS In order to perform the adaptation of the performances on-chip, tuning knobs have to be integrated as well. In this purpose, variable currents and/or voltages must be designed. The problem is different depending on the target current or voltage.
Concerning bias voltages, that do not require high currents, specific DAC can be used directly as in [8, 63] for example. However, power supply voltage must drive high currents, especially if the tuning concerns a PA. Building a variable power supply voltage is then related to built a variable voltage regulator, either by adapting existing power management unit if it is available on the SoC, or by designing a specific voltage regulator.Concerning currents, a variable current generator can also been built, for example digitally-controlled by a digital core.

4.3 CALIBRATION ALGORITHM

In this section, we will describe in more details the different components of the calibration algorithm, namely the regression tool to derive the function z, the regression training, the error estimation and the optimization algorithm.

4.3.1 *Brief introduction to neural networks*

As shown in equation 14, we have to find the mapping between RF performances, tuning settings and alternate measurements. As explained previously, this relationship is usually non-linear and has an intricate form. The use of machine learning in this context is then necessary to learn this relationship. Several machine learning algorithms can be used to perform the mapping, including Multivariate Adaptive Regression Splines (MARS), or neural networks, etc. In this work, we have decided to use neural networks which is one of the most used tool in machine learning, for estimating non-linear functions. Indeed, it has been proven in the past that neural networks with one hidden layer are able to model any continuous non-linear functions accurately. This section will provide a basic introduction to neural networks.

A general view of a neural network topology is presented in Fig.41. A neural network is composed of a number of neurons connected together with several connexions named synapses [41, 71]. In this work, we will use *feedforward* neural networks. In this particular network topology, neurons are organized in three different types of layers: the unique input layer, one or more hidden layers (that process the data), and a unique output layer.
Each of these layers is usually composed of several neurons. For instance, in Fig.41, the input layer is composed of n neurons $X = (x_1, x_2...x_n)$, the hidden layer of m neurons $Z = (z_1, z_2...z_m)$, and the output layer of c neurons $Y = (y_1, y_2...y_c)$. Each neuron of the network receives connexions from all neurons of the previous layer, and is connected to every neuron of the following layer. For instance, neuron z_1 receives connexions from all input neurons (x_1 to x_n) and is then connected to all output neurons (y_1 to y_c). All connections are weighted, i.e. a weight $w_{i,j}$, between neuron i and neuron j, can be defined. Those weighted connexions are called *synapses* and can be seen as multipliers which multiply the "value" of the connexion and the weight. During the learning process, the values of the weights will be adjusted.

A general view of a neuron, is presented in Fig.42. It is composed of three basic functions, namely the propagation, the activation and the output functions.
The first step for the neuron is to transfer all information from its connexions with other neurons (seen as a vector) into a scalar network input. This step is known as propagation. For example, the neuron z_1 has a vector X of connexions with neurons (x_1 to x_n), but only one scalar input in_{z1} is available at the neuron input. Using for instance a weighted sum as a propagation function (which is a very popular function), we sum each output of previous neurons $out_i, i \in X$,with the associated weights of the connexions to neuron z_1 to obtain in_{z1}.

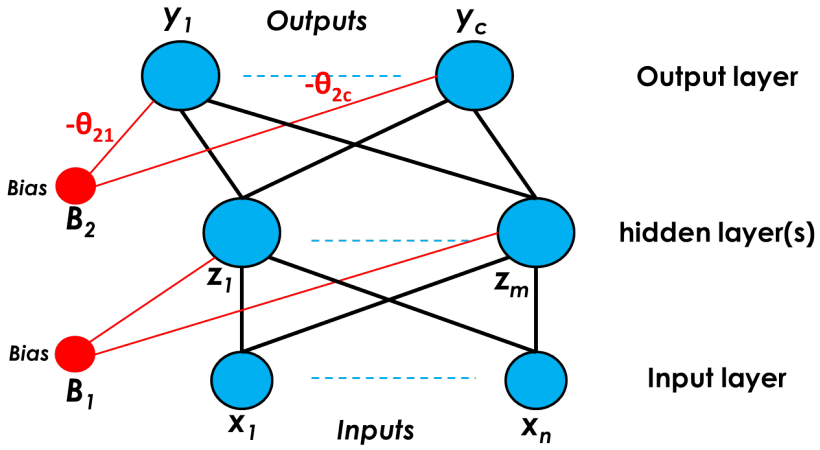

Figure 41: General topology of a feedforward neural network

The weighted sum is a combination of a multiplication (performed by the synapse) and an addition. We can write:

$$in_{z1} = \sum_{i \in X} (out_i . w_{i,z1}). \tag{15}$$

When the network input is processed, the input goes to the activation step. Basically, the neuron can be active or not and its activity is defined by the activation of the neuron $a_j(t)$ (t representing the time), which depends in general on the activation of the previous time step $a_j(t-1)$, the input of the neuron in_j, a threshold value θ_j given for each neuron and an activation function $fun()$. We can write:

$$a_j(t) = fun(a_j(t-1), in_j, \theta_j) \tag{16}$$

The activation function can take different forms, the most popular being the hyperbolic tangent that is comprised between -1 and 1. In the learning procedure, the threshold functions θ_j can be modified so as to adjust the activation to the function that has to be learned. Finally, the output of the neuron must be updated. This is the role of the output function that calculates the output values of the j^{th} neuron out_j depending on the activation a_j. Usually, this function can be considered as unity, i.e. the output of the network is equal to the activation function. We can then write:

$$out_j(t) = a_j(t) \tag{17}$$

During the learning process of the network, threshold values of each neuron and connexion weights between neurons must be updated to enable the network to well estimate the desired function. In order to facilitate learning, the different threshold values can be seen as a weighted connected links that are constantly activated and are known as biases neurons. As an example, bias neurons are represented Fig.41. For instance B_2 in red connects the neurons (y_1 to y_c), with respective weights ($-\theta_{21}$ to $-\theta_{2c}$).

In the context of our work, neural networks are seen as a tool for predicting the performances and will be implemented with the neural network toolbox available in the MATLAB software.

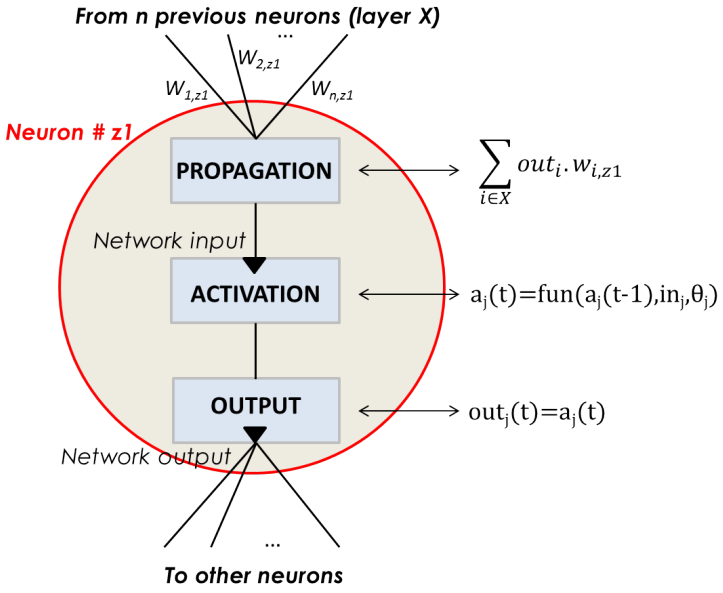

Figure 42: general data processing flow of a neuron

4.3.2 *Training and validation of the model*

The prediction model is built according to Fig.43. In the context of production testing, we have to employ a set of circuits issued from different pre-production wafers. This set constitutes a training set of N circuit instances. Then, on each circuit instance we obtain:

1. The RF performances that have to be monitored, by usual test with an ATE for example.
2. The measurements obtained with BIST sensors.

In order to take tuning into account, both are extracted for K different combinations of tuning knobs. We obtain $N * K$ different samples, that will be used to train our validate the prediction model.

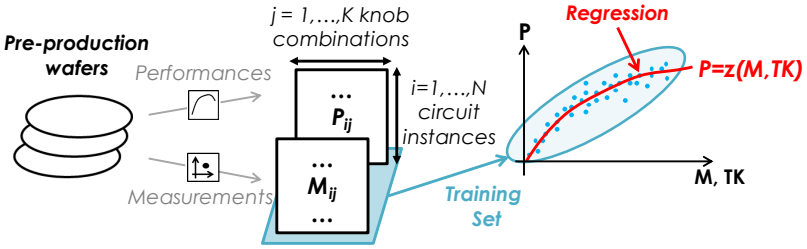

Figure 43: Training phase to derive the regression model that maps low-cost measurements to performances given the tuning knob setting.

Formally, let P_{ij} and M_{ij} denote the performances and measurements for the i-th circuit instance, $i = 1, \cdots, N$, and the j-th tuning knob combination denoted by TK_j, $j = 1, \cdots, K$. The pairs $[(M_{ij}, TK_j), P_{ij}]$ are used to fit the regression model. To guarantee a mapping that is accurate across the feasible space of process variations and tuning knobs, we use N circuit instances that are representative of the fabrication process and cover all process corners and K combinations of tuning knobs that span uniformly the feasible tuning knobs range.

Alternatively, simulation data can be used if the pre-production wafers are not available. To prove the concept, for instance, we can simulate the different circuit instances by performing Monte-Carlo simulation, in order to span the process space and obtain process-perturbed circuits. In order to uniformly span the process space, random sampling is not used. Instead, we use Latin Hypercude Sampling (LHS), that divides the overall process space in several bins, and collect uniformly samples for every bin. In this way, if we take for example one instance per bin, we are sure that all process space is represented (if the number of bins is sufficiently high).

4.3.2.1 *learning phase*

The learning phase is divided in two phases, training and validation. On the $N * K$ samples that we generated (by simulation or with data extracted from pre-production wafers), we select 80% for being used as training samples, and the remaining 20% to be used for validation purposes. Then, supervised learning is used, with a dedicated machine-learning algorithm. In our case, we have chosen neural networks, that will be implemented thanks to the neural network toolbox in the MATLAB software.

TRAINING OF THE NETWORK During the training phase, samples are iteratively presented to the neural network, with corresponding RF performances to be predicted. As explained in the framework of alternate test in chapter 3, one model is build for predicting each performance. For instance if we have 4 performances to predict, four separated neural networks are trained, with corresponding alternate measurements. Indeed, for each trained performance, we try to minimize the number of sensors that are used to obtain an easier performance prediction, with the lowest error.
The network training methodology used is called *backpropagation*. Essentially, it consists, for each input sample, of adjusting the weights of the different neurons of the network, so that the network output corresponds the most to the real value of the performance (target). After each sample, error between predicted and targeted performance values are calculated, and weights are updated for the next sample. The order of training samples is chosen randomly. The training continues while the calculated error decreases. If this error does not decrease for a given number of iterations, training is stopped to avoid *overfitting* and the model with the least error is kept for validation. Since training depends on several parameters (number of neurons, number of layers in the network, selection of training samples...), several models of neural networks should be trained to select the one that obtain the best performances.
In addition, it should be noted that the choice of the training set is crucial for prediction. Indeed, it has to be the most representative of the variety of process variations experienced by the circuit. In statistical terms, the training set has to be limited and non-representative. In our case, training sample selection is even more important, since training should be simultaneously representative of process variations and tuning knob settings. Statistical tools can be used to assess that the training set is representative enough.

VALIDATION OF THE NETWORK Once the network is trained, the remaining set of samples is used as a validation set, to compute the prediction performances of the neural network, and ensure that prediction is accurate enough. For each sample, performances are predicted with the network built at the training stage, network's outputs are compared with targeted

values and prediction results are computed.
The most used metric to obtain prediction error is to compute the RMS error e_{RMS}, given by:

$$e_{RMS} = \frac{100}{P_{i,nom}} \cdot \sqrt{\frac{\sum_{j=1}^{n} (P_{i,j} - \hat{P}_{i,j})^2}{N}} \tag{18}$$

Where e_{RMS} is expressed in %, $P_{i,nom}$ is the nominal value of performance i, $P_{i,j}$ is the real (targeted) value of the performance i, $\hat{P}_{i,j}$ is the estimated value of performance i with the neural network, and N is the number of samples in the validation set. Typical RMS error values are usually in the order of a few percent for obtaining good prediction results.

4.3.3 *Calibration algorithm*

After successful training and validation, the regression model can be readily used to calibrate a circuit. Provided that the measurements stay invariant under changes of the tuning knob values, the calibration can be performed in one-shot, as shown in Fig.44. First, we obtain the measurements and we predict the performances for the nominal tuning knob setting using the pre-trained regression model. In case the performances are unsatisfactory, we run an optimization using as an underlying function the pre-trained regression model. In the course of optimization, tuning knob values are varied in a directed search towards performance values that meet our calibration objective. If measurements stay invariant under changes of the tuning knob values, that is, if $M_{ij} = M_{ih}$, $j \neq h$, then they need to be obtained only once for any tuning knob setting and, subsequently, the values are plugged into the regression model and remain fixed during the course of the optimization. Otherwise, if this condition was not satisfied, then we would have needed to repeat the measurements in each iteration of the optimization algorithm for every visited tuning knob setting.

4.3.4 *Optimization algorithm*

Formally, let the circuit have d performances $P_1, \cdots, P_d$ with upper or lower specifications $s_1, \cdots, s_d$, respectively. For each performance we learn a regression model $P_i = z_i(M, TK)$ as explained above. The optimization algorithm can be formulated as

$$\min_{TK} \gamma \text{ such that: } \begin{cases} P' - \gamma * w \leq s' \\ TK_{\min} \leq TK \leq TK_{max} \end{cases}, \tag{19}$$

where $P' = [\alpha_1 * P_1, \cdots, \alpha_d * P_d]^T$, $s' = [\alpha_1 * s_1, \cdots, \alpha_d * s_d]^T$,

$$\alpha_i = \begin{cases} -1 & : \quad P_i \text{ has a lower specification } s_i \\ 1 & : \quad P_i \text{ has an upper specification } s_i \end{cases}, \tag{20}$$

and $w = [w_1, \cdots, w_d]^T$ is a vector with optional weighting factors for the performances.

The proposed calibration approach consists of obtaining a set of low-cost measurements only once and running an optimization algorithm quickly in software in the background using the ATE. Thus, it incurs overall a low cost that is a small fraction of the standard test cost. A final standard test may be performed after calibration is completed to measure the performances and confirm whether calibration has succeeded, as is the case for all the approaches described in Chapter 3, where the performances are calibrated without being

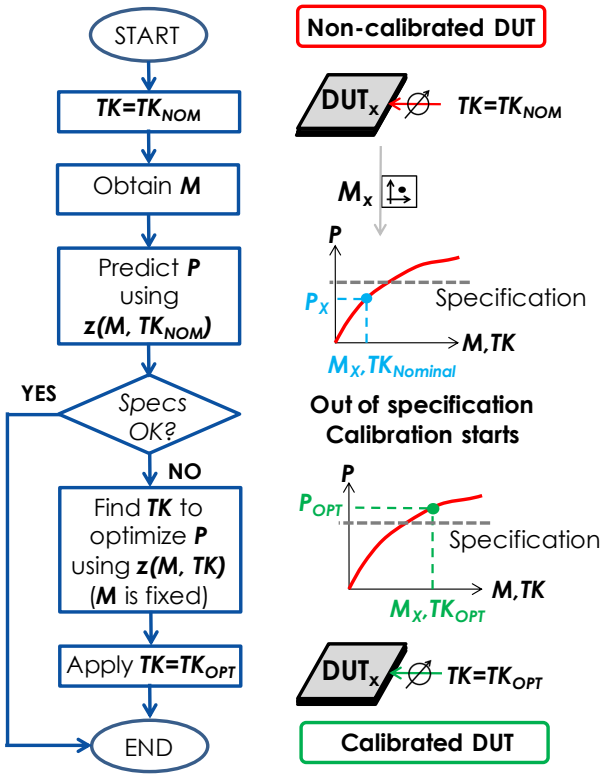

Figure 44: Proposed calibration procedure.

measured explicitly. In this way, we circumvent the risk of labeling a failing circuit as calibrated and functional. However, it should be noticed that the pre-trained regression model not only guides the calibration appropriately, but also predicts the performances for the optimal tuning knob setting at which the calibration converges, as is the case in [62]. Thus, if the regression model predictions are deemed accurate, the final standard test may be summarily eliminated. In this scenario, calibration and testing are performed together in one-shot at low-cost.

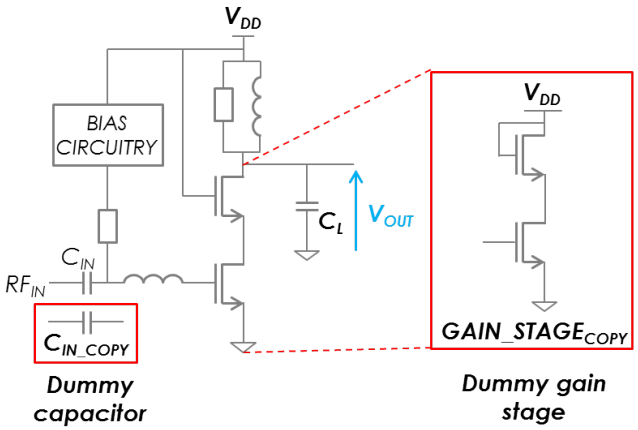

Figure 45: Examples of non-intrusive variation-aware sensors extracted from a LNA topology

4.4 NON-INTRUSIVE VARIATION-AWARE SENSORS

As explained already in section 4.2, the necessary condition to perform calibration in one-shot is to employ measurements that remain invariant under changes of the tuning knobs. In this objective, we propose to use non-intrusive variation-aware sensors. This type of non-intrusive built-in sensors has been demonstrated for RF circuits in [52] and for data converters in [72]. It is inspired by the process control monitors typically placed in the scribe lines of a wafer to monitor variability and identify off-target process parameters [73, 74]. A detailed review of non-intrusive sensors will be given in this section.

4.4.1 *Principle*

The basic idea is to extract an image of process variations experienced by a given RF circuit by using sensors that are not electrically connected to it. The principle is to rely on sensors that consist of dummy analog stages (i.e. bias stages, gain stages, current mirrors, etc.) and dummy single components (i.e. transistors, capacitors, etc.) that can be identified in the topology of the RF circuit. This sensor definition can be readily generalized for virtually any circuit. The dummy structures are "copied" from the circuit's layout and "pasted" in close physical proximity to the corresponding structures of the circuit on the layout, such that two identical structures are perfectly matched. For example, a RF LNA cascode topology is shown Fig.45.

Based on this schematic, one possible dummy stage that can be extracted is a copy of the gain stage, with the two transistors in a cascode configuration. Similarly, one possible dummy component that can be used is a copy of the input capacitance.
The underlying idea is that the dummy structures "witness" the same inter-die and correlated intra-die variations as the structures in the topology of the circuit. Thus, it is expected that the measurements on the dummy structures will be correlated to the performances of the circuit since they are both subject to the same inter-die and correlated intra-die variations. For example, the measurements on the dummy gain stage are expected to correlate with gain and noise figure performances since these two performances depend heavily on the transistor M1. Similarly, the dummy capacitor is expected to correlate to the S11 performance since C_{IN} defines the input matching. The only factor that affects the correlation is the uncorrelated

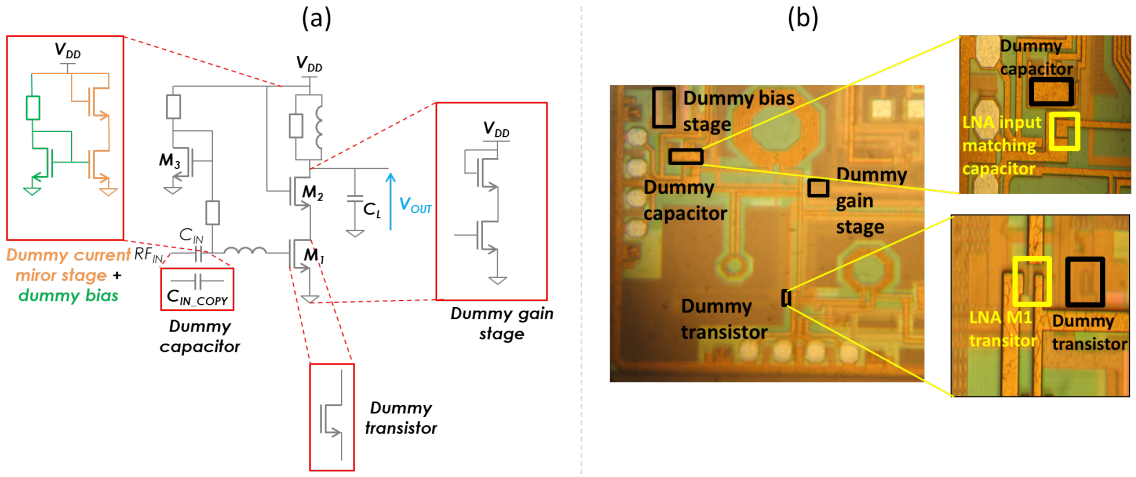

Figure 46: (a) schematic of the LNA and its associated sensors (b) die microphotograph to evaluate sensor integration

intra-die process variations (e.g. mismatch) that can be minimized by a careful layout.
The non-intrusive property of the sensors and the fact that they are readily extracted from the topology of the circuit is a clear advantage since designers need only to deal with inserting the tuning knobs.

4.4.2 *Existing examples*

As an illustration of how non-intrusive sensors can be used, we revisit an example extracted from [52]. In this work, four major RF LNA performances are targeted to be predicted by non-intrusive sensors using the alternate test paradigm. Those performances are gain, noise figure, and linearity measured by both 1-dB compression point and IIP3. In order to extract an image of process variations, five non-intrusive sensors have been built, as presented in Fig.46(a): one dummy gain stage, one dummy bias stage, one dummy current mirror, one dummy transistor and one dummy capacitor. In terms of layout, all sensors are placed very close to the structure they monitor, as it can be seen from Fig.46(b) that shows a photograph of the fabricated die.

As illustrated in Fig.47, the alternate test framework can be successfully used to predict the targeted performances with a relatively low error.

First of all, sensors' outputs and performances exhibit in some cases very high correlation. For instance, as seen in Fig.47(a), the DC current of a dummy current mirror (composed by the DC equivalent model of the LNA and a bias stage) is highly correlated with the gain of the LNA. Indeed, the gain of the LNA (G_{LNA}) can be roughly estimated by $G_{LNA} = g_m.Z_{OUT}$, where g_m is the gate transconductance of the MOS transistor (M_1 here) and Z_{OUT} the output impedance of the LNA. g_m is highly dependent of the current that flows into the MOS transistor, since it is given by $g_m = \frac{\partial I_{DS}}{\partial V_{GS}}$ where I_{DS} is the drain-to-source current and V_{GS} is the gate-to-source voltage. Following a similar argument, we obtain a good correlation betwwen the LNA IIP3 and the output of the dummy bias stage.
The different errors obtained, for the four main LNA performances are displayed in Fig.47(b). The machine learning algorithm used to predict the performances is a feedforward neural

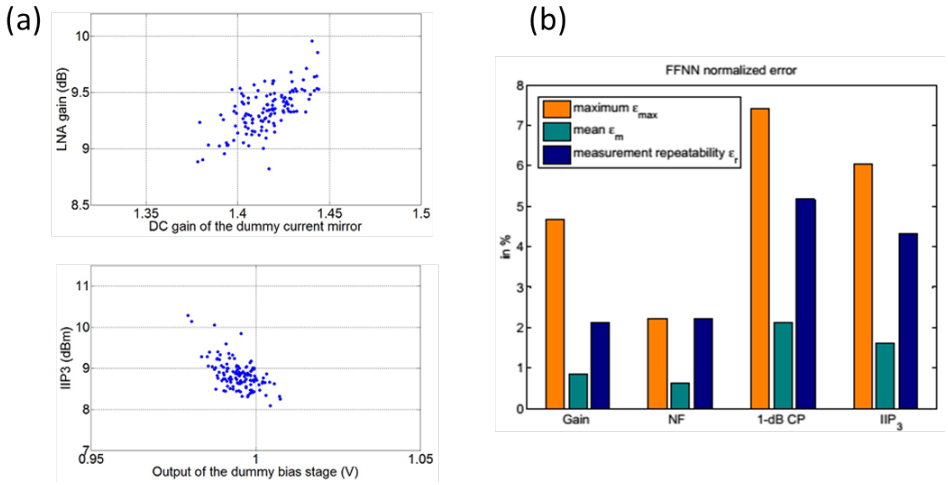

Figure 47: Non-intrusive sensors extracted from the topology of each PA stage.

network (FFNN), using all available sensors' outputs as network inputs. As it can be seen, the mean prediction error is very low for each performance (the worst prediction is done for 1-dB compression point with approximatively 2% of mean error). Maximum prediction error is kept relatively low (maximum around 7% for 1-dB compression point) and is comparable to the measurement repeatability error.

4.5 DESIGN OF A FIRST CASE STUDY

As a proof-of-concept, we have decided to apply the calibration methodology to a Power Amplifier (PA). Indeed, as mentioned in chapter 2 the PA is a critical block in terms of global transceiver performances, but also in terms of power consumption. Since our methodology is taking into account power consumption directly, we have decided to focus on the PA that can consume more than 40% of the total power sunk by the RF front-end of a typical RF WiFi transmitter [2]. Reducing PA power consumption, while maintaining an expected level of performance could significantly reduce the total RF transceiver power consumption.
As explained in more details in Chapter 2, linear PAs are still today the most used PA topologies since saturated-mode PAs that consumes less power, are not compatible with all modulation types and transceivers. We therefore decided to focus on the linear PA design.
Our PA will be designed in the CMOS065 65nm technology provided by STMicroelectronics.

4.5.1 *PA topology*

The first step of the design is to choose a topology for the PA. It can be noted that the context of our work, the design of the PA is nor targeted any particular protocol and is used a case study for prooving the concept of calibration. The chosen topology should respect several criteria:

1. Use a widespread topology.
2. Keep the topology simple, so as to avoid at maximum design issues and focus on the calibration methodology itself.
3. Enable the tuning of 4 main PA performances: gain, linearity, efficiency and DC power consumption.

All the aforementioned reasons have led us to use a two-stage PA. This enables us to have more tuning knobs and more flexible tuning of the different performances. The topology of each stage is chosen to be similar for reasons of design simplicity. Each stage was chosen to be in a self-biased cascode topology, initially presented in [75]. This topology has been used for several standards at 2.4GHz, like bluetooth for example. A simplified schematic of the PA is shown Fig.48(a) [75].

The particularity of self-biased cascode is that the output signal of the PA stage at the drain of the cascode transistor $M2$ is fed back to the gate of transistor $M2$ with a network composed by a resistor R_B and a capacitor C_B that connects the gate to the ground. This network is equivalent to a low-pass filter. By choosing the right cut-off frequency for this filter R_B-C_B, both transistors undergo an identical maximum drain-gate voltage allowing larger output voltage swing. This, compared to conventional cascode topologies, leads to a non-optimal gain but increases output power [75].
The complete PA, shown in Fig. 48(b), is composed of two similar self-biased cascode stages, namely a driver stage designed for maximum gain and a power stage designed for maximum output power. To optimize power transfer, two matching networks at the input (IM) and at the output (OM) have also been designed. It can be noted that, for simplicity reasons, this first case study does not contain interstage matching network for simplicity reasons.

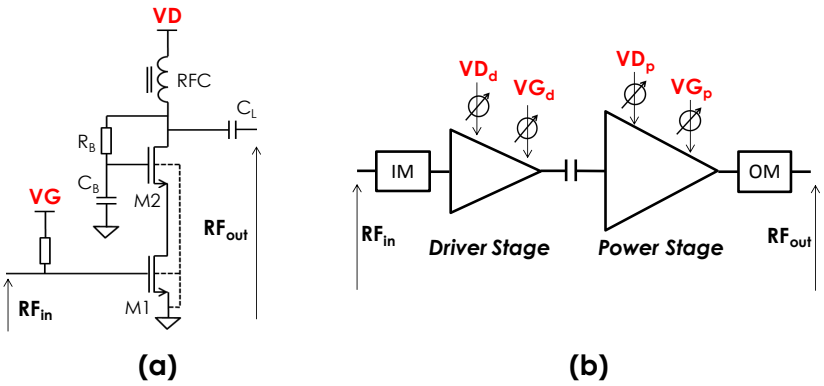

Figure 48: (a)Self-biased cascode topology chosen for each PA stage; (b) Two-stage PA with tuning knobs.

CLASS AB OPERATION The next step of the study is to define the operation class. For the reasons given in chapter 3, in practice a good compromise between the different requirements is achieved by operating at class AB, since:

- Linearity is ideally the same as Class A operation
- Power consumption is lowered compared to class A operation
- Efficiency and gain are balanced between class A and class B operation

For example, considering that operating in class-B will result in a 75% efficiency for a gain of 6dB and class-A will result in a 50% efficiency for a gain of 12dB, a middle-point class-AB operation will be a compromise between both, with a 65% efficiency (15% more that class-B), for a gain of 9dB (3dB more than class-A) [14].

4.5.2 *Choice of tuning knobs*

The choice of tuning knobs is also guided by several constraints:

1. They must have a significant influence on PA performances, so as to be able to efficiently tune the RF performances
2. They must not modify the PA topology and be convenient to designers

The aforementioned reasons guided us to chose tuning knobs that are available at the design level and keep the initial PA topology intact. These are the power supply voltages and gate bias voltages of each stage, as shown in Fig.48(b), that is, no alterations in the design have been made. On one hand, modifying the gate bias voltage of the PA basically translates to controlling the current into the MOS channel. On this RF current depend the RF power (used to calculate efficiency), the gain of the PA, and the linearity (the three are in fact traded-off). On the other hand, power supply voltage is mainly used to control the DC power consumption. These tuning knobs will be used for both stages in order to have a better control on all the performances that have to be monitored.

Table 7: Power amplifier nominal performances

RF performance	Nominal value
P_{DC}	46.6 mW
$OCP1$	9.1 dB
$Gain$	34.2 dB
PAE	31.9%

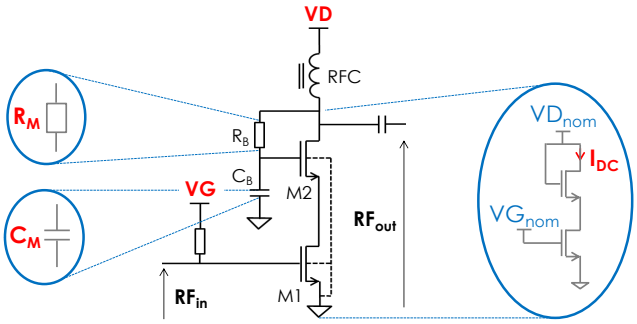

Figure 49: Non-intrusive sensors extracted from the topology of each PA stage.

4.5.3 *Nominal performance*

The three main performances of the PA are considered for calibration: linearity measured by the output referred 1-dB compression point ($OCP1$) which also corresponds to the maximum output power, efficiency measured by the maximum power added efficiency (PAE), and voltage gain ($Gain$) which is equal to the power gain if matching conditions are satisfied. Since our goal is to find an optimal performance trade-off without increasing power consumption, DC power consumption (P_{DC}) is considered as a fourth performance. The nominal design values in a typical simulation of the four performances are listed in Table 7.

4.5.4 *Choice of non-intrusive sensors*

The principle of non-intrusive variation-aware sensors is to monitor process variations in critical components of the RF circuit's architecture. Considering the chosen PA topology, we have several components to monitor (two MOS transistors, one capacitor, one resistor, and the components of the matching networks). In our case, we decided to focus on the PA itself, without considering matching components and RF choke inductor, since those components will be integrated off-chip onto a printed-circuit board in our test chip (see Chapter 5). Then, we have to consider:

- The two MOS transistors, especially the gain transistor (M_1), on which depend all the RF performances.
- The capacitor and the resistor of the self-biased cascode topology, since some RF performances, such as gain and linearity, directly depend on their values.

In this case study, we have then decided to use three non-intrusive sensors per stage, namely (a) a dummy single resistor, (b) a dummy single capacitor, and (c) a dummy cascode gain stage, all extracted from the topology of the PA stages, as shown in Fig.49. The associated alternate low-cost measurements are the values of the two dummy resistors, the values

of the two dummy capacitors, and the DC currents drawn by the two dummy cascode gain stages.

4.6 SIMULATION RESULTS

4.6.1 *Introduction*

As a first proof-of-concept we evaluate the methodology by simulation. The PA has been designed into the Cadence framework in the CMOS 65nm technology provided by ST microelectronics. Simulations are run in Cadence in the ADE XL environment using SpectreRF. All simulations consider both inter-die process variations and mismatch according to the distributions defined in the process design kit. Then, we discuss the regression model training to derive the mappings between the non-intrusive measurements and the performances. Finally, several calibration scenarios will be presented, to assess the effectiveness of the methodology.

4.6.2 *Regression model*

The first step towards calibration is the prediction of the performances based on non-intrusive sensor measurements. For this purpose, we generated training data though Monte Carlo simulation. The training data includes $K = 305$ different combinations of tuning knobs and for each tuning knob combination we perform a Monte Carlo simulation with $N = 100$ runs. The total number of training samples is then $N * K = 30500$. To cover a maximum of the process parameter space, we used Latin Hypercube Sampling (LHS) as the underlying sampling technique in the Monte Carlo analysis. We randomly selected 80% of the samples to train the regression model and the rest 20% of the samples were used for validation purposes, making sure that this ratio is preserved for every tuning knob setting. The regression model was built with a feed-forward neural network (FFNN) implemented in Matlab using the function `feedforwardnet` and the Levenberg-Marquardt back-propagation training algorithm.

In order to optimize prediction and avoid overfitting, we selected an optimal number of sensors for each performance prediction. Specifically, we then trained different models adding sensors one by one until the prediction error was sufficiently low. For example, on one hand, DC power consumption can be very accurately predicted by using only dummy gain stages (from which we extract the DC current I_{DC}) since the DC power consumption is directly related to the DC current drawn by the PA, which is monitored by the dummy sensor. The DC power consumption is the performance that is the most accurately predicted as it will be seen afterwards. If we add other sensors, they are not bringing any more information to the neural network and the model may start overfitting, increasing the error. On the other hand, if we use only the two dummy gain stages for predicting the linearity, we do not have enough information since the linearity does not depend only of the DC current that flows in the amplifier. We have to add both dummy capacitors and resistors so as to achieve a low error, since these components directly affect the linearity.

Table 8 shows the error of the regression model on the independent validation set in terms of the correlation between the true and estimated performance values and the root-mean-square (RMS) error, both expressed in %. As can be seen, the generalization of the regression model on the validation set is excellent and the model can be safely used to guide the calibration procedure.

Table 8: Generalization ability of the regression model measured on the validation set.

RF performances	Correlation Coefficient	RMS Error
P_{DC}	99.99%	0.45%
$OCP1$	97.98%	1.9%
$Gain$	97.08%	3.1%
PAE	85.12%	5.7%

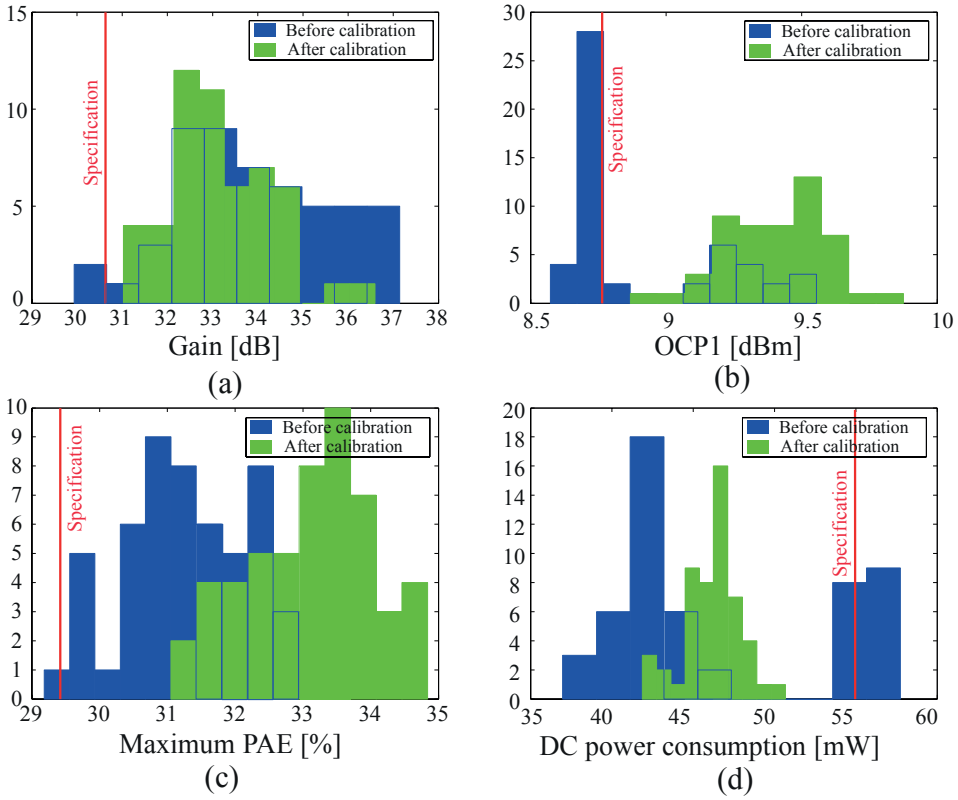

Figure 50: Circuit distributions before and after calibration for scenario 1.

4.6.3 *Scenario 1: "standard" calibration to recover yield loss*

In this experiment, we perform "standard" calibration on a set of failing circuits to recover yield loss. First of all, we performed an independent Monte Carlo simulation with 3600 runs. Then, from this set of instances, we identified in total 52 circuits that fail at least one performance using the standard test approach. Fig.50 illustrates the histograms of the performances for these 52 circuits before we attempt calibration, that is, for the nominal tuning knob setting. As can be observed, the majority of failing circuits violate the $OCP1$ and P_{DC} specifications listed in Table 7. After calibration is completed, the model predicts that all circuits are successfully calibrated. Final simulations with the resultant tuning knob values are performed to obtain the ground truth performance values and confirm the model prediction. The histograms of calibrated circuits are superimposed on the histograms of non-calibrated circuits in Fig. 50. As can be observed, the model prediction is confirmed and we have recovered 100% of the yield loss. The distributions of the calibrated circuits lie well within the specification limit and, in general, they are more centered showing less

RF performances	Goal attainment		
	Before calibration	After calibration	
		Model	True value
P_{DC}	92%	94%	92%
$OCP1$	35%	94%	94%
$Gain$	61%	98%	92%
PAE	58%	94%	96%

Table 9: Goals attainment before and after calibration.

dispersion. The fact that the model was not only adequate to achieve calibration, but also predicted correctly the performances of the calibrated circuits, shows that a final test is not really necessary.

4.6.4 *Scenario 2: "aggressive" calibration to obtain better performance trade-off*

In this experiment, we perform calibration on a set of functional circuits with the aim to obtain a better performance trade-off. We set an "aggressive" goal to improve $Gain$, PAE, and $OCP1$ such that they have values better than the nominal design values listed in Table 7, while tolerating a DC power consumption up to 51.26 mW, which corresponds to a 10% increase of the nominal design value, but nevertheless is still within the allowable specification of 54.9 mW. The aim of this experiment is to demonstrate that the proposed calibration method can be used to achieve stringent performance goals apart from yield loss recovering.

In this analysis, we employ 100 circuits extracted from an independent Monte-Carlo simulation. The percentage of circuits that attain the goal for each performance for the nominal tuning knob setting is shown in the first column of Table 9. Again, the "true" (e.g. simulated) value of performances is used to determine if circuits lie within the performance goals or not. As expected, around half of the circuits do not attain the goal for $Gain$, PAE, and $OCP1$ since for the these performances the goal is set at the nominal value which lies close to the mean value. Furthermore, nearly all circuit instances attain the goal for P_{DC} since the goal for P_{DC} is close to the specification. After calibration is completed, the model predictions of goals attainment are shown in the third column of Table 9. As it can be seen, according to the model, the calibration procedure succeeded to reach the goal for the vast majority of circuits. Final simulations with the resultant tuning knob values are performed to obtain the ground truth performance values and confirm the model prediction. The fourth column of Table 9 shows the ground truth goals attainment based on simulation. We observe that the percentage of circuits that attain the goal is very high and above 90% for all performances. For the $Gain$ the percentage increases from 35% to 94%, for the PAE the percentage increases from 61% to 92%, and for the $OCP1$ the percentage increases from 58% to 96%. Furthermore, these improvements are achieved without deteriorating P_{DC} for which the goal is still attained for 92% of the circuits.

By comparing the third and fourth columns of Table 9, we conclude that the model predictions are close to the ground trough percentages obtained by simulation, which implies that the model infers the performances of the calibrated circuits with sufficient accuracy. Table 10, which shows the RMS error of the model for the non-calibrated and calibrated circuits, confirms this result. As can be seen, the model achieves practically the same level of accuracy across the set of non-calibrated and calibrated circuits for each performance.

RF performances	RMS error	
	Before calibration	After calibration
P_{DC}	0.57%	0.94%
$OCP1$	1.52%	1.79%
$Gain$	2.81%	2.71%
PAE	5.13%	5.18%

Table 10: Estimation errors on the non-calibrated and calibrated circuits for scenario 2.

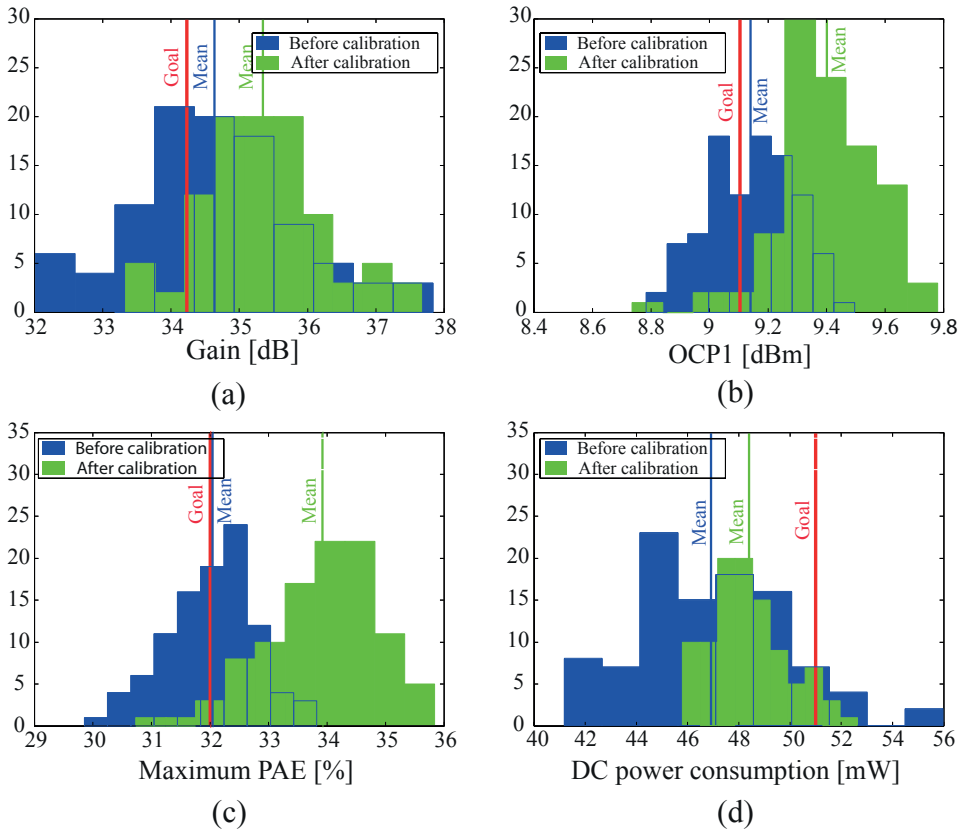

Figure 51: Circuit distributions before and after calibration for scenario 2.

A visual representation of the calibration efficiency is illustrated in Fig. 51, which shows for each performance the histograms of simulated circuits before and after calibration. As can be seen, for the three performances *Gain*, *PAE*, and $OCP1$ that have a lower specification, the sample means are moved clearly to the right. For the P_{DC} that has an upper specification the mean is also moved to the right, but the standard deviation is smaller and the largest part of the distribution, that is, 92% according to the third column in Table 10, lies on the left of the vertical line that corresponds to the goal.

4.7 DISCUSSION

In this chapter, we have demonstrated a true one-shot calibration methodology for RF circuits. this technique employs non-intrusive sensors, that are able to track the RF performances regarding process variations, and enable a one-shot tuning of the performances. Even if this methodology can be readily used for test and calibration purposes, several things can be pointed out. First of all, since alternate test is used in this framework, we rely on statistical data, that can give wrong results. Even if mis-prediction error is minimized by a proper training and validation of the model, the question than remains is: *"how far can we trust the prediction model?"*. A first answer that can be made is that, since the model has been sufficiently well trained, it is able to accurately enough predict the RF performances, and we can trust it without performing further investigations. This is the case for most of the one-shot alternate-based calibration frameworks presented earlier in the literature, since it is dedicated to replace standard test. We can "tolerate" than a small fraction of fabricated dies are mis-predicted, like a few percent, as for usual ATE test (this estimation could be made by calculating yield loss and test escapes for instance). However, our methodology have a fundamental difference with previous works, since non-intrusive sensors give only a *static value*, that is totally independent of the RF circuit. On one hand, this property enable the true one-shot calibration. But, on the other hand, we are working in a open-loop configuration and we do not have any indication about the actual DUT performances. In other words, several other scenarios should be taken into account:

1. There is a catastrophic defect into the CUT, that is not monitored by the non-intrusive sensors, so the CUT does not work even if we predict that it is working
2. There is a non-catastrophic defect in the CUT, that impacts the performances but cannot be taken into account in the model (for example a component that is defective) so the prediction is wrong
3. The prediction model mis-predict the RF performances and/or the best tuning knob setting for another reason, and calibration is not performed properly

Points 1 and 2 have already been investigated, and a solution has been proposed by monitoring CUT temperature with non-intrusive variation-aware sensors [52]. Indeed, a defect in the CUT will be translated to a shift in the current drawn by the CUT, and consequently to the power that it dissipates. This power shift can be detected through a non-intrusive temperature monitoring of a critical MOS transistor for instance. If there is a defect, it will impact the circuit's power consumption, and will be detected by the sensor. Considering point 3, there is currently no proposed low-cost solution (to the best of our knowledge) that does not have mis-prediction. Outliers can be detected with a defect filter, to minimize the possibility of mis-classification [REF]. However, to assess that calibration succeeded, we can only perform a final test with ATE. This is a costly solution, but added cost can be minimized considering that during this test phase, a maximum of circuits have been successfully calibrated by the previous calibration methodology and lie within the specifications, so yield loss is compensated.

4.8 CONCLUSION

In this chapter, we have presented a novel one-shot calibration methodology dedicated to RF circuits, based on alternate test. The key point of the approach is that alternate measurements are realized with non-intrusive sensors. This property enable to achieve a true "one-shot" calibration, where performances are teste only once to find the best tuning knob setting. The developed calibration procedure is able to correct out-of-specifications performances but also to find different trade offs between several performances. A proof of concept has been presented, by designing and calibrating a case study, in the form of a two-stage power amplifier, with the proposed method. Non-intrusive sensors have been built around the CUT, and the calibration methodology has been proven by simulation.

The next step of the study is now to further prove the concept on silicon. In this objective, next chapter will present a silicon chip that has been designed in 65nm technology provided by STmicroelectronics. An updated case study will be presented and tested to fully prove the concept of non-intrusive one-shot calibration.

5

EXPERIMENTAL RESULTS

A proof of concept of our calibration methodology though simulation has been presented in the last chapter. To fully prove the feasibility of this methodology, this chapter presents silicon results.
Towards this objective, we have fabricated a silicon chip that contains a Power Amplifier (PA) and its associated non-intrusive sensors. We present the design and the layout of the PA and its associated sensors in section 5.1. A Printed-Circuit Board (PCB) has been built to fully test the fabricated chip, as detailed in section 5.2. the measurement setup is detailed in section 5.3. Based on measured data, calibration results are presented and analyzed in section 5.4.

5.1 DESIGN OF A SECOND CASE STUDY

The design of the PA has been slightly modified with respect to the design presented in the previous chapter. The topology itself has not been changed, but several improvements have been made in order to facilitate the layout and to assure the functionality of the PA.
In general terms, the PA topology, with a driver and a power stage, both in a self-biased cascode configuration, has been kept identical. In addition, three matching networks have been implemented, respectively at the input (IM), at the output (OM) and at the inter-stage level (ISM). We have chosen not to integrate these matching networks on-chip, but to implement them with discrete components on a Printed-Circuit Board (PCB). This choice is motivated by several reasons:

- Matching networks may contain several passive components, in particular inductors, that can be area-hungry. Placing the matching networks on a PCB will save chip area and cost.
- A matching network at PCB level offers more flexibility, because we can modify the components in the case of sub-optimal matching.
- Considering our application, designing the matching networks on PCB will enable to eliminate variations in the matching components from chip to chip. In this case, all tested chips will be in exactly the same environment and only the process variations of the PA will be considered, which facilitates the demonstration of the calibration methodology.

It can be noted that, even if matching network variations will not be considered here, they could have been taken into account by adding sensors that monitor matching components. For instance, in [76], on-chip capacitors, resistors and inductors of an RF LNA are monitored by non-intrusive process sensors.

5.1.1 *PA design*

In the PA design methodology, we have chosen to implement the power stage as a number of driver stages in parallel. This design choice was made for simplifying the overall design of the PA, in order to have more guarantees on its functionality. Indeed, PA performances are not the primary concern for prooving the calibration methodology, but PA functionality is. This design choice offers several advantages:

- In terms of design, once the driver stage has been implemented, the power stage will result from a parallel combination of driver stages, which will facilitates device sizing.
- In terms of layout, attention is focused on the driver stage, since the power stage will is a combination of drivers stages.

MOS TRANSISTOR CHOICE AND DEVICE SIZING The sizing and design of the PA is not the primary focus of this manuscript. Consequently, design constraints will not be presented in details here.
The 65 nm CMOS065 RF technology of STmicroelectronics is used for this chip design. Several MOS transistors are available, namely low-power MOS transistors with a nominal power

supply of 1.2V, and two high-voltage transistors with a thicker gate oxide that can respectively work with nominal power supply of 1.8V or 2.5V. In our application, we have chosen high-voltage transistors mainly because a higher supply voltage increases the output power delivered by the PA.
In general terms, the width of the MOS transistors determines the maximum output power that the PA will be able to deliver. DC simulations have been carried out, with different transistor sizing, in order to determine the size that results in optimum output power. The targeted maximum power is 20 dBm (that corresponds to 100mW), which is less than in state-of-the-art PA designs (that can deliver an output power up to 30 dBm for 1W PAs), but sufficient to prove the calibration concept. Then, the driver stage has to drive the input of the PA power stage. A ratio of 1:4 has been chosen. The power stage will then be composed of 4 driver stages in parallel. In order to satisfy the requirements, we have chosen a width of $1600\mu m$ for the transistors in the power stage and a width of $400\mu m$ for the transistors in the driver stage.
The other components to size are the capacitors and resistors of the self-biased cascode structures. The strategy chosen to size these components is to obtain the maximum voltage variation on the output in order to maximize the compression point, without compromising too much the gain. Concerning the PA driver stage, resistors are chosen to be $1k\Omega$ and capacitors $250fF$. If the capacitor value is too much decreased, the drain-to-gate voltage of the common source transistor of the power stage could exceed 3V which is the absolute maximum voltage of the technology. Concerning the PA stage, resistors are chosen to be 250Ω and capacitors $1pF$ which correspond also to four components in parallel.

5.1.2 *Layout of the PA and its associated sensors*

The layout of the PA has been divided into several parts.

First of all, as explained before, the power stage is equivalent to four driver stages in parallel. Moreover, based on the non-intrusive sensors principle presented in the last chapter, all sensors have to be placed in the closest physical proximity to the structures they monitor in the PA. We have considered this constraint by layouting the PA core and its associated sensors simultaneously by using an elementary stage. The elementary stage is composed of the self-biased cascode structure of the driver stage and its associated non-intrusive sensors (one dummy gain stage, one dummy capacitor and one dummy resistor). Consequently, the power stage and its associated sensors will be composed of four elementary stages in parallel. By using this layout methodology, attention is focused on the elementary stage, that is afterwards duplicated in order to obtain the power stage.
The elementary stage has been integrated in the gap between the pads of the driver, as illustrated in Fig.52.The corresponding schematic, with the same highlighted colors, is given in Fig.53
In the design manual of 65nm technology, it is specified that two successive pads must not be closer than $20\mu m$. Since we want to minimize the PA area, this value has been used for placing the elementary stage. Each pad has a horizontal dimension of $60\mu m$, so the available area to integrate the elementary stage and its associated sensors is $60 * 20\mu m$. This choice of a layout between the pads has several advantages. First of all, it uses a space that is usually left empty, which optimizes chip area. Secondly, it minimizes the connexions between (a) the drain of the cascode transistor M2 and the output pad and (b) the source of the transistor M1

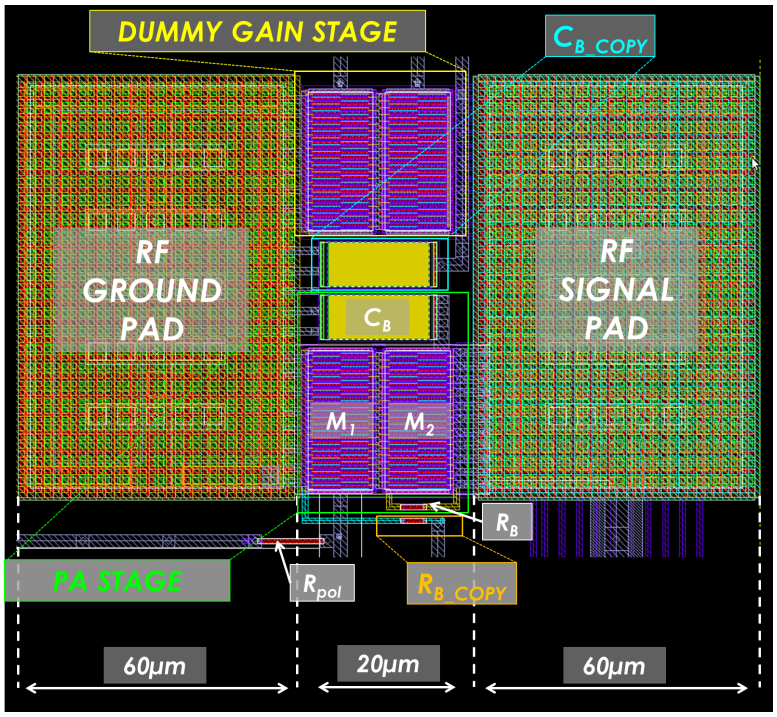

Figure 52: Layout of the elementary PA stage

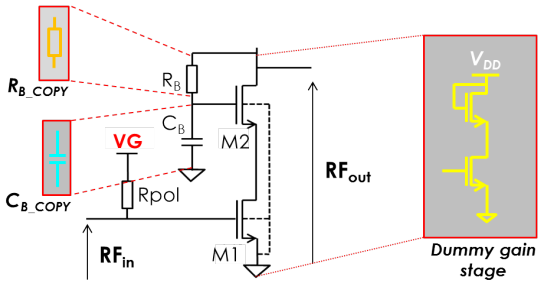

Figure 53: Schematic of the elementary PA stage

and the ground.

CHOOSING THE NUMBER FOR MOS TRANSISTOR FINGERS As seen previously, the elementary stage is composed of two transistors with a width $W = 400\mu m$. In order to fit those transistors to the gap between two pads, they have to be divided into several fingers. Moreover, the multiple finger technique is one of the most effective method to build a large size transistor since it results in lower gate resistance, lower RF noise, and higher frequency performance. However, reducing the transistor finger width or increasing the number of fingers can result in larger gate capacitance. We have then chosen to divide the MOS transistors into 40 fingers, which means that each of then is $10\mu m$ long to have a compromise between area and maximum current constraints. The total area of the MOS transistor is then approximately $10 * 25\mu m$.

LAYOUT OF THE ELEMENTARY STAGE The elementary stage, highlighted in green Fig.52, is composed of the two MOS cascode transistors M1 and M2, divided in several fingers. The self-biased cascode components C_B and R_B are respectively placed on top and at the bottom of these transistors. This compact structure is inserted into two pads, one RF signal pad connected to the output of the cascode structure and one ground pad. The RF input of the stage will be connected to another pad afterwards.
We have then chosen a nearly symmetrical layout. First , the copy of C_B (C_{Bcopy}) has been placed on top of C_B, as highlighted in light blue Fig.52. Then, the dummy gain stage, that is composed of the two MOS cascode transistors identical to those in the PA driver stage has been placed with the same orientation. The distance between the dummy gain stage and the PA driver stage is less than $15\mu m$. Finally, the resistor R_B, its copy R_{Bcopy} has been placed next to R_B so that both are matched, as highlighted in yellow in Fig.52. In terms of shielding, the capacitors and the transistors have been placed into guard rings, so as to be well isolated.

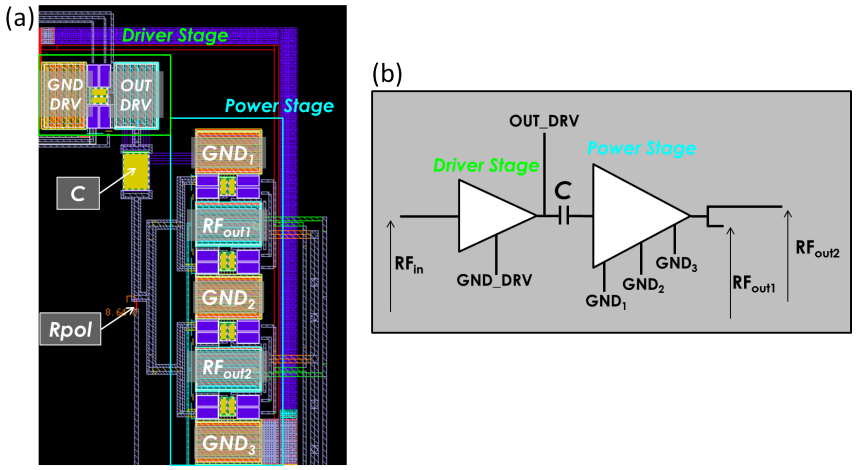

Figure 54: (a) layout and (b) associated schematic of the PA

LAYOUT OF THE POWER STAGE As explained previously, the power stage is equivalent to four elementary stages in parallel. In terms of layout, both the driver and power stages have been placed with a 90 degrees orientation difference in order to minimize the interferences between them. Specifically, the power stage has been designed in a symmetrical approach in between several pads. Since the power can be as high as 20 dBm, the output is supported by two pads instead of one, in order to have less current flowing into each pad. The configuration of the power stage layout is then to alternate between ground and RF output pads, with an elementary stage in between each pad. It requires a total of three ground pads and two output RF pads that will be thereafter combined on the PCB level. The arrangement is GND1-RF1-GND2-RF2-GND3, with a symmetry axis in the middle of pad GND2.

5.1.3 *Layout of the pad ring and the complete chip*

The RF and ground pads (7 pads in total) are considered together with the layout of the PA. The rest of the pads are analog pads, that provide access to the non-intrusive sensors (6 independent pads) and to the bias voltages of the PA (2 pads). Then two pads are put to have GND and VDD for the pad ring themselves, and a clamp is also put to prevent for

degradation due to high voltages. The pad ring has been designed to fit on one side of the chip, with 10 pads on the same side.
The PA and the embedded sensors have been integrated into a chip that contains also a 2.4 GHz LNA that is investigated in another project. The chip layout is presented Fig.55. The circuit is composed of two separated pad rings that are not connected together, one for the PA and one for the LNA. The PA is highlighted with the red contours. As it can be seen, since no matching network is present on-chip, the area of the PA is minimized, compared to the area of the neighboring LNA that has all inductors and matching networks integrated on-chip. The total chip area is $1200 * 800 \mu m^2$. A photograph of the fabricated chip is presented in Fig.56.

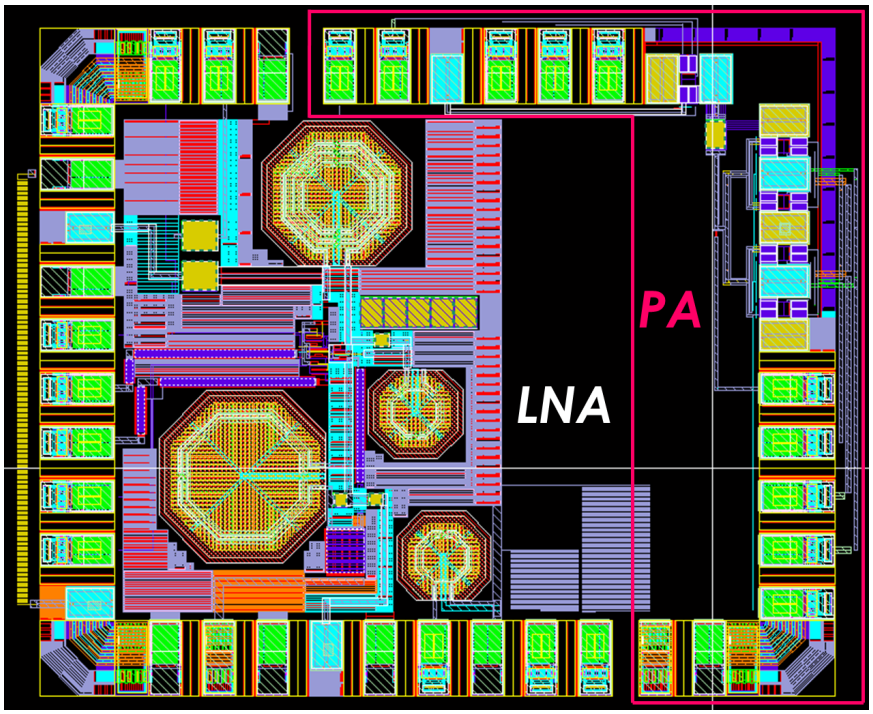

Figure 55: Layout of the complete chip with PA and LNA

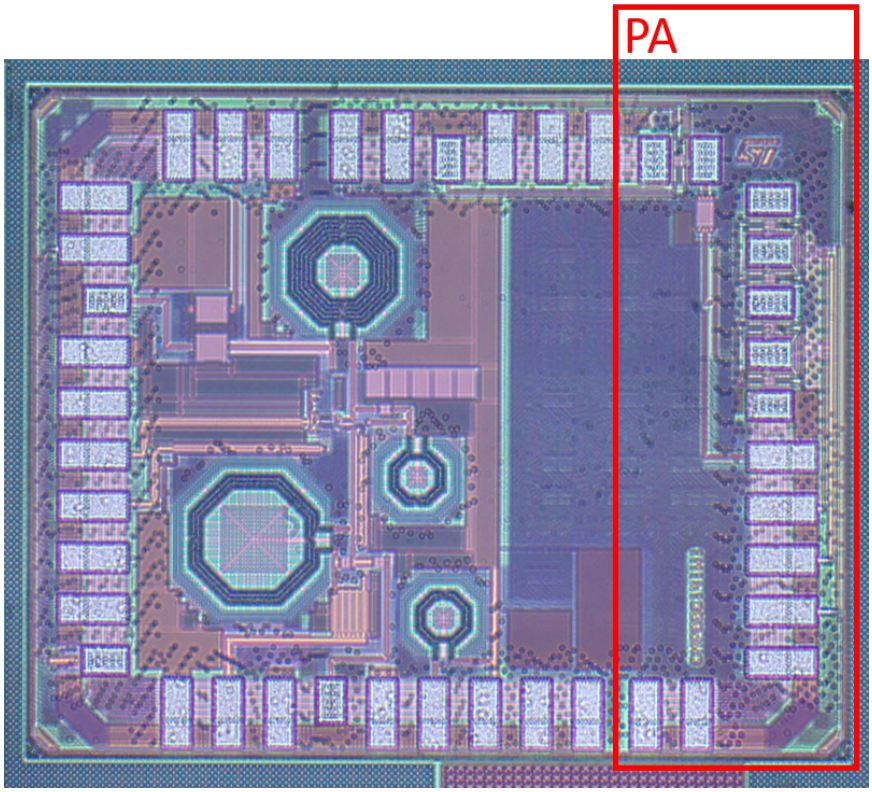

Figure 56: Die photograph of the fabricated chip

5.2 PCB DESIGN

The PCB is a crucial part of the case study since it contains the various matching networks that are necessary for the functionality of the PA.

5.2.1 *Overview*

A general view of the schematic of the PCB is presented in Fig.57, which highlights the PA section. The PCB contains the three matching networks, at the input of the PA (RF_IN), between the two stages (INTER_STAGE) and at the output (RF_OUT). It contains also the power lines, polarization lines, and the DC lines to extract the values of the non-intrusive sensors. The design of the PCB has several constraints, in order to obtain the desired RF circuit behavior. In the next subsections, we consider the design of the three matching networks.

5.2.2 *Impedance matching networks*

INPUT MATCHING NETWORK The input matching network aims to match the 50Ω characteristic impedance to the input impedance of the circuit (Z_{in}). This matching is evaluated by measuring the S11 parameter of the complete PA, that is minimized by choosing an appropriate impedance network. The general strategy to match the input impedance is to evaluate the S11 to extract the value of Z_{in} and then use the Smith chart to add appropriate elements

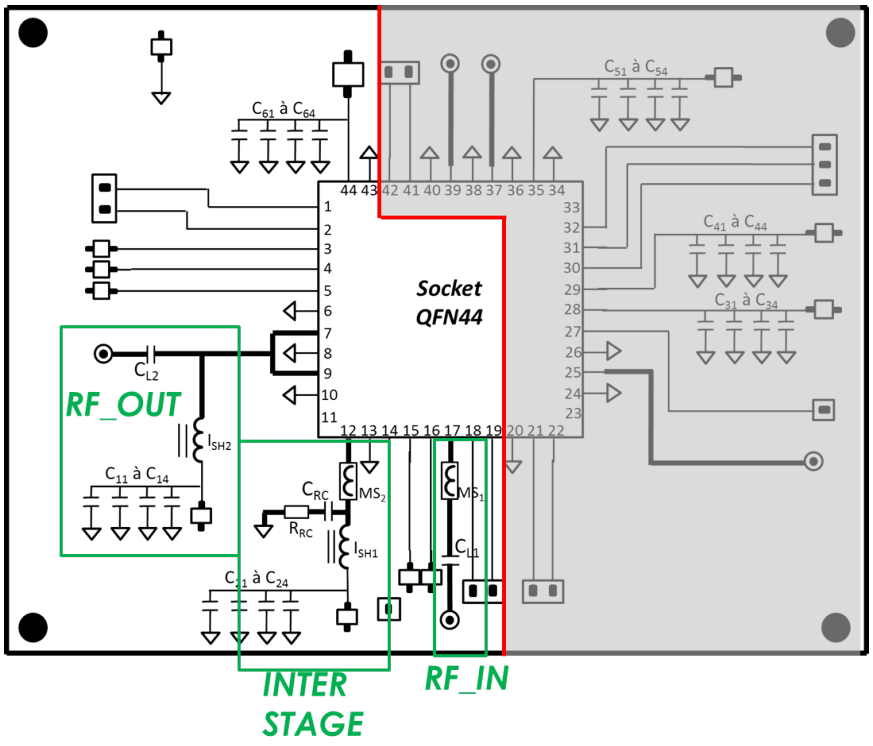

Figure 57: schematic of the test board

in the matching network.
In practical terms, the value of S11 must be always lower than -10dB for the considered frequency band to have a proper matching. On the contrary, if the value of S11 is above zero, this means that a negative impedance is presented at the input of the PA, which creates instabilities.

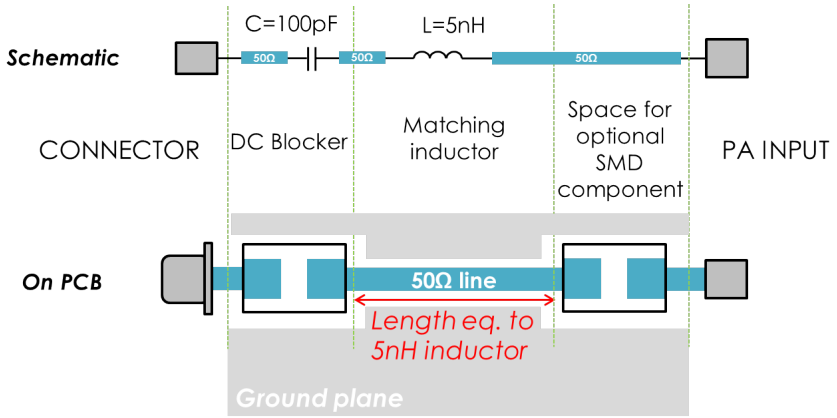

Figure 58: schematic and PCB design of the input matching network

The complete input matching network is presented in Fig.58. It consists of a DC blocker implemented by a series 100pF capacitor and a 50Ω line whose length is chosen to be equivalent to a 5nH inductor. This value corresponds to the value of the matching inductor between

50Ω and the input of the PA, so as to have a S11 below -10 dBm at 2.4GHz. Then, a very small capacitor (40 to 50 fF) is needed to achieve better matching, which may also been realized by the parasitic capacitor of the RF pad.

OUTPUT MATCHING NETWORK The output matching network has to match the optimal output impedance of the PA to the characteristic 50Ω impedance of the RF line. The standard matching strategy is known as conjugate matching. This method consists of matching the output impedance of the circuit (Z_{out}) with its conjugate, so as to maximize power transfer and achieve a maximal efficiency. Given that the circuit sees a 50Ω load at its output, we have to build a network to match 50Ω to the conjugate of Z_{out}. This condition is usually satisfied by extracting Z_{out} and/or by minimizing the $S22$ parameter at the center frequency of the circuit.
However, this strategy is not employed for PA design. Instead, we will use loadline matching. Loadline matching is most appropriate for PA design because its aim is to maximize the PA output power, which is a primary concern. In this case, it exists an optimal output impedance Z_{opt}, usually different from the output impedance Z_{out}, for which output power is maximized. The target of loadline matching is then to find the value of Z_{opt} by performing loadpull simulations. Loadpull consists of placing an impedance tuner at the output of the PA and running a series of simulations that evaluate the PA output power for different output loads. This type of simulation generates power contours that can be seen in a Smith chart which allows evaluating which impedance has to be presented at the output of the PA to have the best output power.
In our case study, the optimal output impedance is given by $Z_{opt} = 47 + j12$. This value is not very far from the characteristic impedance $Z_0 = 50 + j0$, so we decided to present directly a 50Ω line at the output of the PA, which leads to a sub-optimal load line matching but anyways the degradation is practically negligible, i.e. 0.3 to 0.5 dBm. Since we mainly target to have a functional PA, this output network will be easier to design and have less design constraints.

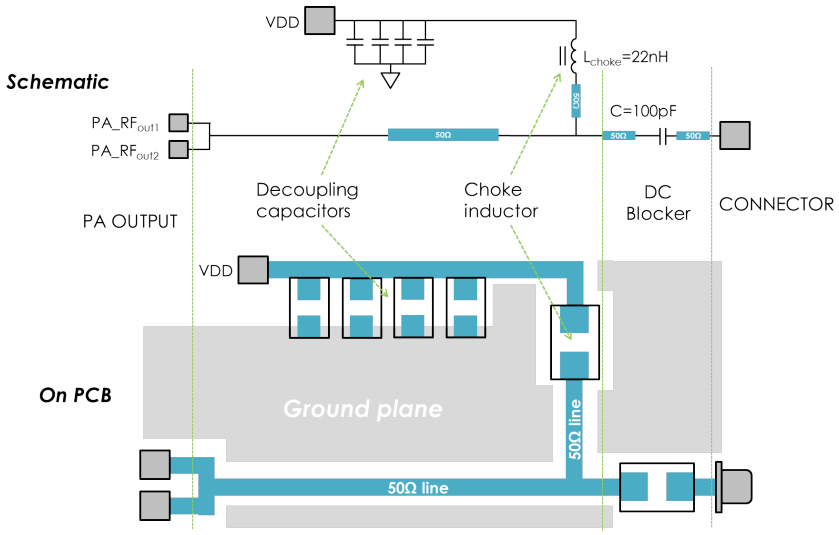

Figure 59: schematic and PCB design of the output matching network

An overview of the complete output network is presented in Fig.59. At the output of the PA, the two RF outputs are combined onto the PCB. Then, as explained earlier, we present

a 50Ω line directly at the output. The power supply is provided by another line, with also a 50Ω impedance, which is connected to the power supply via an RF choke inductor and decoupling capacitors. Finally, the link to the output connector is done via a DC blocker (100pF series capacitor) and another 50Ω line.

INTERSTAGE MATCHING NETWORK The most critical matching network is the one between the two PA stages. The aim of interstage matching is to match the optimal output impedance of the driver stage (Z_{outDRV}) to the input impedance of the power stage (Z_{inPWR}), and also to bring the DC power to the driver. The only component that has been put on-chip is the series capacitor between the driver and the power stage, whose value has been chosen to achieve matching with a transmission line that goes out of the driver into the PCB. The general interstage matching network that we have designed is presented in Fig.60.
In this context of impedance matching, the length of the transmission line that goes out of the driver stage is crucial. We have identified that this length has to be chosen to have a value of $\approx 6.5nH$ equivalent inductor, in order to have a proper matching. In addition, a series R-C network has to be added after this transmission line to achieve better stability. This network will also degrade the gain, but since PA functionality is the main concern, the R-C network should lead to a more stable PA.

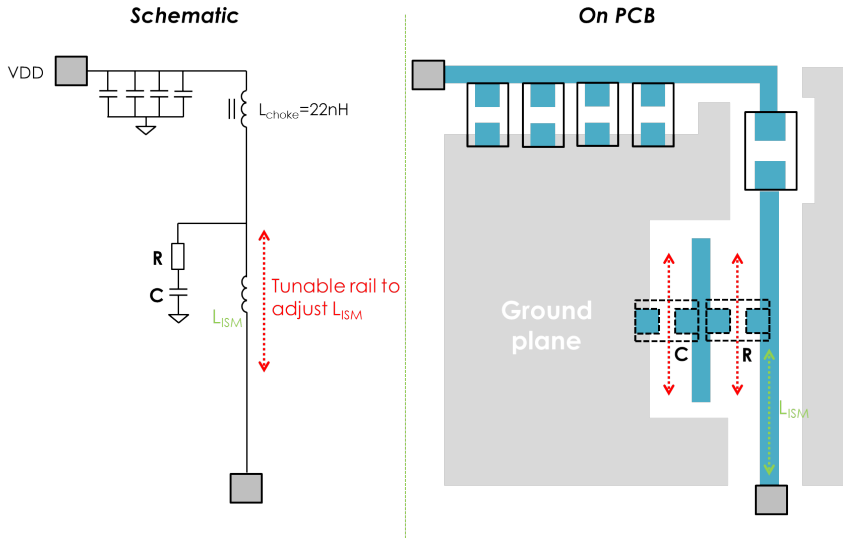

Figure 60: schematic and PCB design of the inter-stage matching network

In order to have more flexibility in the design, we have decided to build a tunable rail so as to place the R-C network onto the PCB. The idea is to be able to choose the length of the matching line by placing the R-C network at different locations on the PCB. In this way, we can adapt the matching to different parasitics, such as the parasitics of the bonding inductors or the RF pads. The value of the inductor can be tuned from 5nH to 8nH by moving the R-C network.
The last part of the interstage network is to place the RF choke inductor and the decoupling capacitors, and connect them to the power supply.

5.2.3 *fabrication of the PCB*

The PCB has two sides. At the bottom side, a specific RF layer has been implemented so as to design all the required RF lines. Photographs of the PCB are presented in Fig.61. The top side of the PCB contains a socket in order to test the different chips. It contains also all connectors and test points in order to perform all the tests and measurements. PCB varnish has been removed where further matching adaptation might be needed, which gives a direct access to the RF lines. Each power supply line is decoupled with four decoupling capacitors.

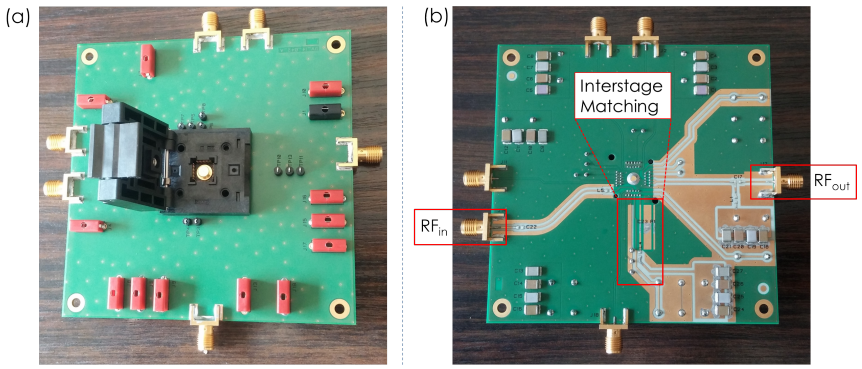

Figure 61: Photographs of the PCB (a) top side (b) bottom side

5.3 FABRICATED CHIPS AND MEASUREMENT SETUP

5.3.1 *fabricated set of chips*

The PA has been fabricated in the CMOS065 65 RF technology provided by ST microelectronics. In total, we fabricated 71 circuits. In order to have a set of circuits that is representative of the process variations and to prove the concept with out-of-specifications performances, the circuits have been extracted from three different wafers, in which transistor parameters have been varied. Our set contains chips from:

- A Typical (TYP) wafer, with typical transistors, resistors and capacitors (24 chips)
- A FAST wafer, with fast transistors, maximum poly resistance and typical capacitors (23 chips)
- A SLOW wafer, with slow transistors, maximum poly resistance and typical capacitors (25 chips)

Indeed, ST microelectronics can control three parameters in the fabrication process, namely the type of transistor, the poly resistance and the capacitance. Consequently, the transistor parameters in the FAST and SLOW wafers were selected after examining through simulation which two combinations among the eight possible combinations (fast or slow transistors, low or high ploy resistance and low or high capacitance) result in the most variations in the performances of the PA.

5.3.2 *Measurement environment*

Performing many measurements on approximately 71 different circuits under the same conditions is not a straightforward task, since we do not have automatic test stations at our disposal. The measurement procedure will consist of measuring the RF performances of all the circuits for a large number of tuning knob setting which requires a long time. In addition, we need to measure the outputs of the non-intrusive sensors just once for all the circuits since measurements are independent of the tuning knob settings. The measurement procedure must be automatized on a computer-aided platform so as to be able to perform all measurements in the same manner. The overall measurement setup is controlled by a station equipped with the LabVIEW software. This software controls all the instruments used in the measurement procedure and stores the measured data. The data that has to be extracted from each circuit are:

- The measurements of non-intrusive sensors: two resistors, two dummy gain stages, and 2 capacitors.
- The main RF performances of the PA for different tuning knob settings: GAIN, OCP1, PAE.
- The power consumption P_{DC} for the same tuning knob settings.

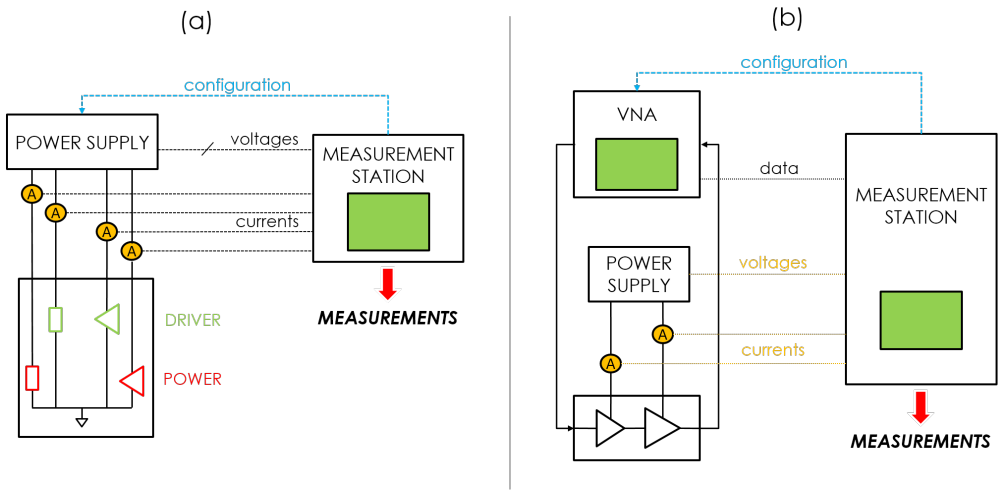

Figure 62: Schematic view of the measurement setups: (a) DC setup and (b) RF setup

NON-INTRUSIVE SENSOR MEASUREMENTS Non-intrusive sensor measurements is performed using the DC setup illustrated in Fig.62(a). Three different types of sensors must be measured (e.g. dummy resistor, dummy capacitor and dummy gain stage). The dummy resistor values can be extracted by applying a voltage across the resistor and measuring the current. Regarding the dummy gain stages, we apply the nominal power supply and bias voltages and we measure the current that flows through the transistor channel. Regarding capacitors, several methods can be considered. In our case study, the values of the capacitances are very small (around 1pf and 250fF). Even by using an accurate capacimeter, the parasitic capacitances of the different components around the capacitor (RF pad, connectors and most importantly the socket) will obscure the capacitance value and will make its extraction very challenging. Instead, we can measure the dummy capacitors included in the LNA that was designed in the framework of another project that was under way in parallel. However, as it will be seen in our results, the alternate test predictions, which define the quality of calibration, are still excellent if we exclude the dummy capacitor. For this reason, we removed the capacitors from the analysis.
For the dummy resistors and gain stages, the setup is as follows. Four power supplies bias the different sensors. We have chosen a voltage of 1V for the resistors and 2.5V for the dummy gain stages. Four current meters are placed in series with the power supplies, and their values are recorded by the LabVIEW program. The circuit are sequentially measured one after the other.

RF MEASUREMENTS The RF setup is illustrated in Fig.62(b). The three RF performances (e.g. GAIN, PAE and OCP1) are measured using a Vector Network Analyzer (VNA). Specifically, gain can be measured by fixing the power at the input and measuring the power at the output, OCP1 is measured with a power sweep, and the PAE is calculated by measuring the RF power at both the input and the output, performing the subtraction and dividing by the DC power consumption. DC power consumption is measured with current meters. The computer controls the VNA, in order to perform the different measurements. Power supplies, coupled with current monitors are used and their value is recorded by the computer. In this manner, all performances can be measured for one circuit, and the procedure is repeated for

every circuit in the training set.
The measurement procedure is as follows. For every circuit, we sequentially measure the performances for different tuning knob combinations. For every tuning knob combination, we measure one time each performance with a procedure automatized in LabVIEW. In details, the GAIN is first measures at low power (-35 dBm). Then, a power sweep is performed in order to measure OCP1 and the corresponding PAE and PDC values.
A picture of the measurement setup is illustrated in Fig.63. The different elements are also noted in the figure. We can observe the test board (1), the multimeters (2), the bias voltage supply (3), the VNA (4), the computer equipped with LabVIEW software (5) and the power supply (6).

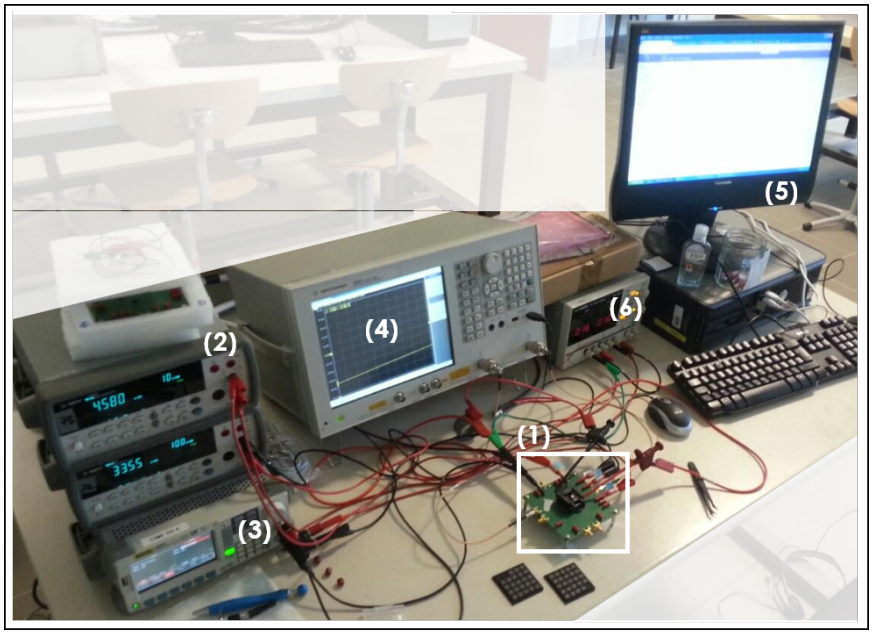

Figure 63: Pictures of the RF setup

5.3.3 *Nominal performances of the PA*

One PA from the typical wafer has been first measured and the performances are much lower than expected. This degradation is mainly due to matching imperfections and some design choices. First of all, the measured saturated output power is 12.4 dBm, instead of the expected value of 20 dBm. This represent a degradation of 5x (18mW instead of 100mW). Consequently OCP1 performance will be degraded as well. Also, the PAE is very degraded, around 2 to 3 percent instead of 25%.
The degradation of the performances is mainly due to matching imperfections. Indeed, in the design procedure, not all the environment around the PA have been modeled accurately enough, especially the RF socket. It appears that the parasitic inductance of this socket (in the order of 5 to 6 nH) has been under estimated in our model and completely degrades the matching networks previously implemented. Even by adding other components in this matching network, we were not able to increase the output power. We had assumed that the PA output was already close to the characteristic impedance of the line. Due to the socket, the PA output is actually far from the expected value and we have no way to correct this

degradation.
Nevertheless, we are still able to claim correct functionality of the PA, even with much lower performances than expected. As mentioned earlier, for the purpose of demonstrating the proposed calibration methodology, it is sufficient to have a functional PA.

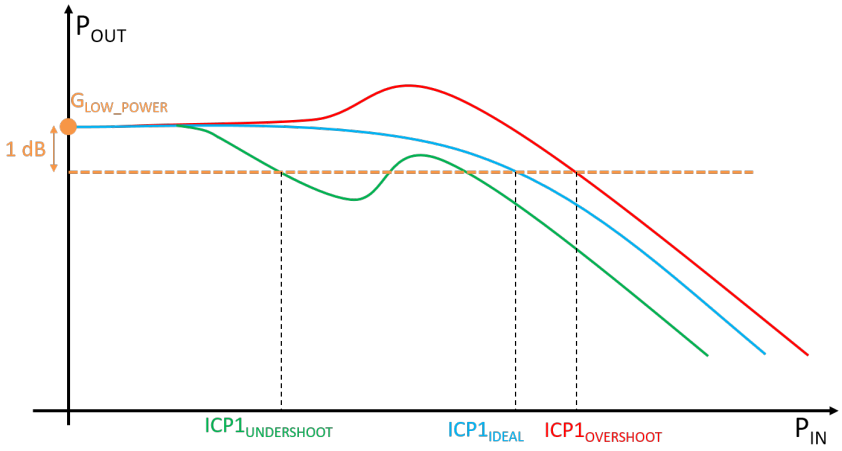

Figure 64: Illustration of different linearity behavior of the PA

5.4 RESULTS

5.4.1 *Performance measurements and dataset analysis*

By performing the first measurements on a few chips extracted from the three wafers, it appears that the PA is very susceptible to process variations, up to the point that there is no common range of tuning knob settings for which all the circuits work properly. here, by working properly, we mean that performances could be considered acceptable and that the circuit behaves as a PA. For instance, if the gain is negative or very low, the circuit cannot be considered as a PA. Consequently, the circuits extracted from each wafer have a different range of tuning knob settings for which they work properly. It has to be noted that this is not an issue for the calibration concept, since both process variations and tuning knob settings are considered in the prediction of the performances. However, this sets several challenges to the alternate test procedure and the calibration concept. Moreover, some of the fabricated circuits were not functional at all. The final dataset is then composed of 54 circuits, including 13 TYP, 25 FAST and 16 SLOW chips.

5.4.1.1 *Working range of the different circuits*

An accurate measurement of the RF performances is not feasible for every possible tuning knob setting. For instance, if the bias voltage of the amplifier is too low, the MOS transistors of the amplifier are turned off. Consequently, the measured gain is very low and some other performances have a non-relevant value. In our case study, the bias voltage values for which the MOS transistors are turned on is different for each wafer. This point is exacerbated by the fact that SLOW and FAST wafers correspond to the two most extreme cases. In a usual fabrication process, those devices are very rare, but in this study they constitute more that two thirds of our dataset. Consequently, variations are exacerbated and each wafer is working in a different biasing range.

Another issue comes with the definition of the 1dB compression point given in Chapter 2. 1dB compression point is defined as the point where the gain has dropped by 1dB compared to the gain at low power. It assumes that (a) the highest gain is obtained at low power and

Tuning Knob	SLOW		TYPICAL		FAST	
	Min value	Max value	Min value	Max value	Min value	Max value
VDp	2.2	2.5	2.2	2.8	2.2	2.8
VDd	2.2	2.8	2.0	2.2	2.0	2.2
VGp	0.76	0.77	0.685	0.695	0.62	0.63
VGd	0.65	0.71	0.57	0.61	0.53	0.55

Table 11: minimum and maximum tuning knob values for each wafer type

(b) that the gain is constant and then drops after a given power. However, the measured compression point is not always trustworthy, as seen in Fig 64, where the ideal compression case is represented in blue. For some biasing configurations (that are always avoided in a usual design procedure), the gain could have an overshoot before dropping, which is represented in red in Fig.64. In some other cases, the gain could drop a first time, then increases again and finally drops as expected. This undershoot is represented in green in Fig.64.

These linearity behaviors are due to many factors, such as the imbalance between the driver and the power stage for example. In these cases, the 1dB compression point cannot be measured correctly, but most importantly the PA does not work properly since it does not satisfy the linearity requirement. Concerning the undershoot case, the compression point is pushed backwards. In this case, the linearity performance is degraded and most likely the calibration algorithm will not choose the corresponding tuning knob values. However, concerning the overshoot case, the compression point is pushed forward, even if the behavior of the PA is not linear anymore when the gain starts to rise. Then, it is possible that this combination could be chosen by the optimization algorithm because 1 dB compression point is better, even though overall the circuit is no longer linear. The corresponding tuning knob settings must then not be used by the calibration algorithm.

For these reasons, we have decided to define a range in which all the circuits in a wafer type works properly, that is, the linearity measurements are trustworthy and gain is sufficient. The measured ranges are detailed in Table 11 and a graphical representation is presented in Fig65.

As it can be seen, only the power supply of the power stage VDp overlaps between the different wafers. For the other tuning knobs, there is a difference between each wafer type, most importantly concerning the bias voltage of the power stage VGp. On one hand, it can be seen that FAST transistors are active for lower bias voltages than TYP transistors, which in turn are active for lower bias voltages than SLOW transistors. On the other hand, FAST transistors are overdriven before TYP transistors are active and so on and so forth. This is explained by the fact that FAST and SLOW circuits are two very extreme cases. Moreover, whenever there is an unbalance between bias voltages, for instance when we increase VGp and decrease VGd too much, the behavior of the PA is not correct anymore, which gives a restricted range of possible bias voltages.
Moreover, for each wafer type, we will measure around 65 different tuning knob combinations in order to span the entire allowed tuning knob range.

In the following sections, we will take this range into account so as to ensure that every calibrated chip behaves in a correct way.

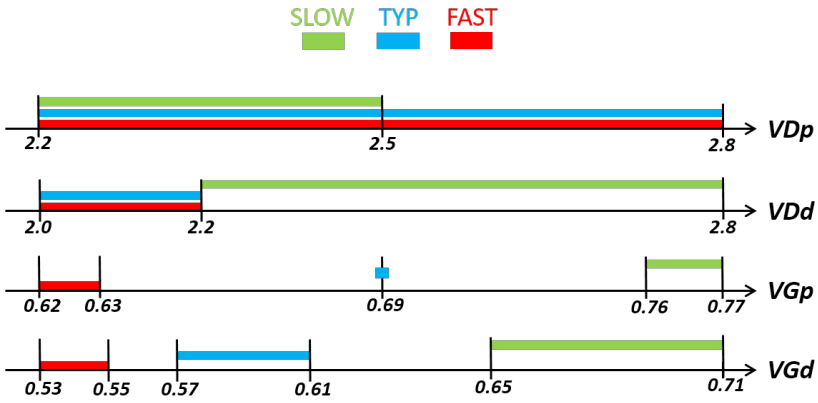

Figure 65: Illustration of the different tuning knob ranges for each wafer type

Table 12: Within-wafer variability and measurement repeatability error of performances.

	Gain (in dB)		OCP1 (in dBm)		P_{DC} (in mW)		PAE (in %)	
	Var	Rep	Var	Rep	Var	Rep	Var	Rep
FAST	1.38	0.062	1.51	0.088	8.92	0.149	1.23	0.028
SLOW	0.62	0.083	1.50	0.121	10.80	0.201	1.90	0.037
TYP	0.71	0.048	0.97	0.134	5.83	0.148	0.82	0.088

5.4.2 *Repeatability and variability analysis of the measurements*

Before proceeding with the analysis of the alternate concept and the one-shot calibration methodology, we need to ensure that the measurements that are taken are reliable enough. Specifically, two analysis will be performed:

- Repeatability: in this analysis, the same measurements are repeated 10 times for randomly chosen circuits extracted from each wafer type (one circuit per wafer type). The standard deviation is then calculated across these measurements for each wafer type and across all wafers, and then used to define the repeatability error of the measurement.

- Variability: in this analysis, measurements are taken from 5 randomly chosen circuits extracted from each wafer type (15 circuits in total). Standard deviation is calculated for each wafer type and across all wafers to define the variability of the measurement/performance.

First of all, since, there are no common tuning knobs ranges for which all chips from all wafers are functional. This observation allows us to claim that the across-wafer variability of performances is "infinite". Then, table 12 shows the within-wafer variability of performances versus the within-wafer measurement repeatability error of performances. The within-wafer measurement repeatability error of a performance is defined as the standard deviation of several measurements of the performance obtained on a randomly selected chip. The within-wafer variability of a performance is defined as the standard deviation of the performance values across all the chips of the wafer, where the performance values are obtained for the

Table 13: Within-wafer and across-wafer variability and measurement repeatability error of sensor measurements.

	IRD (in mA)		IRP (in mA)		ID (in mA)		IP (in mA)		CAP (in pF)	
	Var	Rep	Var	Rep	Var	Rep	Var	Rep	Var	Rep
FAST	0.025	$3.86 \cdot 10^{-4}$	0.013	$1.45 \cdot 10^{-4}$	0.186	0.015	0.54	0.072	0.053	$1.9 \cdot 10^{-3}$
SLOW	0.026	$3.7 \cdot 10^{-4}$	0.01	$1.22 \cdot 10^{-3}$	0.234	$8.2 \cdot 10^{-3}$	2.26	0.19	0.064	$2.4 \cdot 10^{-3}$
TYP	0.039	$1.06 \cdot 10^{-4}$	0.011	$3.16 \cdot 10^{-5}$	0.171	0.054	1.27	0.327	0.048	$2.3 \cdot 10^{-3}$
All	0.029		0.012		2.34		7.45		0.06	

tuning knobs setting that corresponds to the middle points of tuning knobs ranges for this wafer, as shown in Fig. 65. The performance values of a chip are obtained several times and they are averaged so as to remove the measurement repeatability error effect. It should be noted that this analysis is performed by obtaining performance values using the conventional test approaches. As it can be seen, within-wafer variability of performances is much larger than the within-wafer measurement repeatability error. In most cases, it is larger by more than one order of magnitude. Therefore, not only the across-wafer variability of performances is "infinite", but also the within-wafer variability is clearly noticeable.

Table 13 shows the within-wafer and across-wafer variability versus measurement repeatability error analysis for the sensor measurements. In this case, the across-wafer variability can be computed since the sensors are non-intrusive and independent of tuning knobs. As it can be seen, the within-wafer variability is at least one order of magnitude larger than the within-wafer measurement repeatability error. For ID and IP, the across-wafer variability is significantly larger than the within-wafer variability, whereas for IRD, IRP, and CAP the across-wafer and within-wafer variabilities are practically the same. This can be explained by recalling that the two corner wafers include FAST/SLOW transistors and high polylisicon resistance. Since ID and IP are obtained on dummy gain stages that include transistors, they show large across-wafer variability. In contrast, IRD, IRP, and CAP are obtained on single resistors and a capacitor and they show zero across-wafer variability.

In conclusion, our data set meets our original objectives set in Section 5.4.1 to have sufficient variability and chips failing due to excessive process variations.

5.4.3 *Proving alternate test concept*

The efficiency of the calibration concept depends on how accurately enough we can predict the RF performances through non-intrusive sensors based on alternate test paradigm. Therefore, we will learn the regression model based on the available data set and we examine in details its generalization accuracy based on several error metrics. This analysis is detailed in the following subsection.

5.4.3.1 *Evaluation procedure*

The regression model for each performance is learned in the MATLAB environment using a two-layer feed-forward neural network (FFNN) with sigmoid hidden neurons and a linear output neuron. This type of network can fit mapping problems arbitrarily well given an

Table 14: Pre-trained regression model efficiency according to several error metrics.

perfomances	ϵ_{RMS} (in units of performance)		ϵ_{max} (in units of performance)		ρ (in %)		FoM (dimensionless)	
	mean	std	mean	std	mean	std	mean	std
Gain (in dB)	0.36	0.0046	6.04	0.088	98.3	0.036	0.18	0.023
OCP1 (in dBm)	0.34	0.0088	2.88	0.14	97.43	0.14	0.23	0.0058
P_{DC} (in W)	$3.2 \cdot 10^{-3}$	$5.25 \cdot 10^{-3}$	$56 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	98.23	0.065	0.19	0.003
PAE (in %)	0.76	0.013	15.59	0.3	92.65	0.2	0.37	0.061

Table 15: Pre-trained regression model efficiency according to FPR error metric.

Gain			OCP1			P_{DC}			PAE		
ϵ_{th} (in dB)	FPR mean (in %)	FPR std (in %)	ϵ_{th} (in dBm)	FPR mean (in %)	FPR std (in %)	ϵ_{th} (in W)	FPR mean (in %)	FPR std (in %)	ϵ_{th} (in %)	FPR mean (in %)	FPR std (in %)
0.5	2.75	0.32	0.5	8.96	0.52	0.0025	19.41	1.26	1	4.2	0.62
1	1.64	0.05	0.75	3.78	0.26	0.005	4.64	0.35	2	1.05	0.1
1.5	1.35	0.046	1	2.01	0.17	0.0075	1.98	0.22	4	0.34	0.02
2	0.84	0.027	1.25	1.17	0.16	0.01	1.05	0.10	6	0.29	0.0089
3	0.33	0.015	1.5	0.72	0.096	0.02	0.32	0.015	10	0.17	$2.9 \cdot 10^{-17}$
4	0.14	0.029	2	0.34	0.03	0.03	0.28	0	12	0.11	0.0089
5	0.056	$7.3 \cdot 10^{-18}$	2.5	0.13	0.027	0.04	0.099	0.02	14	0.085	$1.5 \cdot 10^{-17}$
6	0.017	0.015	3	0.0028	0.0089	0.05	0.028	$3.7 \cdot 10^{-18}$	16	0.0028	0.0089
6.5	0	0	3.25	0	0	0.06	0	0	17	0	0

appropriate number of neurons in its hidden layer. The appropriate number of neurons can be selected using a trial and error approach. We observed that the error metrics improve initially as we keep adding hidden neurons and they saturate after a given number of hidden neurons. This number is around 15 in our experiments. The results that we report are therefore based on a network with 15 neurons in the hidden layer. Alternatively, we could have used any other advance regression tool, such as multi-adaptive regression splines (MARS), support vector machines (SVMs), etc.

Our data set includes a total of N=55 chips. For each chip we consider K=67 tuning knob combinations. We obtain the performance values $\mathbf{P}_{nk}$ for the n-th chip and the k-th tuning knob combination $\mathbf{TK}_k$, where $n = 1, \cdots, N$ and $k = 1, \cdots, K$. We also obtain the sensor measurements $\mathbf{M}_n$ for the n-th chip, where $\mathbf{M}_n = \mathbf{M}_{nk}$, $k = 1, \cdots, K$. In total, we have $N \cdot K = 55 \cdot 67 = 3685$ pairs (or "chip instances") $[(\mathbf{M}_n, \mathbf{TK}_k), \mathbf{P}_{nk}]$ for learning the regression functions.

To report trustworthy and unbiased error metrics, we perform 10 random shuffles of the chip instances and we use 5-fold cross validation. In particular, for a given random shuffle, we divide the chip instances into 5 folds, where in each fold we have $N_v = 3685/5 = 737$ chip instances. We repeat learning 5 times, each time using 4 folds in the training set and 1 fold in the validation set, such that each fold is left out once. The error metrics are computed on the validation set. We report the mean error metrics across the 5 learnings for a given random shuffle and across the 10 random shuffles.

Formally, let P_{ij} and $\hat{P}_{ij}$ denote respectively the true and predicted value of the i-th performance of the j-th chip instance, $i = 1, ...4$, $j = 1, ..., N_v$. Let also $\mathbf{X}_i = [P_{i1}, ..., P_{iN_v}]$ and $\mathbf{Y}_i = [\hat{P}_{i1}, ..., \hat{P}_{iN_v}]$. We use the following error metrics:

- Root mean square (RMS) error defined as

$$\epsilon_{\text{RMS}} = \sqrt{\frac{\sum_{j=1}^{N_v} \left(P_{ij} - \hat{P}_{ij}\right)^2}{N_v}}. \tag{21}$$

  It is expressed in units of each performance and indicates the average expected error.

- Maximum error defined as

$$\epsilon_{\text{max}} = \arg\max_j \left(\left|P_{ij} - \hat{P}_{ij}\right|\right). \tag{22}$$

  It is expressed in units of each performance and indicates the worst case maximum error.

- Pearson's correlation defined as

$$\rho = 100 \cdot \frac{\text{Cov}(\mathbf{X}_i, \mathbf{Y}_i)}{\sigma_{\mathbf{X}_i}\sigma_{\mathbf{Y}_i}}. \tag{23}$$

  It is expressed in % and indicates the linear correlation between the true and predicted performance values.

- Figure of Merit [77] defined as

$$\text{FoM} = \frac{\epsilon_{\text{RMS}}}{\sigma_{\mathbf{X}_i}} \tag{24}$$

  It is dimensionless and compares the average expected error with the variability of the performance.

- Failing prediction rate (FPR) [78] defined as

$$\text{FPR} = 100 \cdot \frac{1}{N_v} \sum_{j=1}^{N_v} I^i(j), \tag{25}$$

  where

$$I^i(j) = \begin{cases} 1 & : \quad \left|P_{ij} - \hat{P}_{ij}\right| > \epsilon_{th} \\ 0 & : \quad \text{otherwise} \end{cases}. \tag{26}$$

  It gives the percentage of chip instances for which the absolute prediction error is larger than a threshold ϵ_{th}.

In general, we consider that the regression model for a performance is accurate and that the learning has succeeded if (a) the average expected error is considerably smaller than the variability of the performance and (b) the absolute prediction error for every circuit is smaller or at least comparable to the measurement repeatability error when the performance is measured on ATE using the conventional test approach.

Criterion (a) is expressed directly by the error metric FoM. FoM is preferred from ϵ_{RMS} since it encompasses ϵ_{RMS} and it also accounts for the scenario where the variability of the performance is small. In this case, ϵ_{RMS} is expected to be small anyways and, thus, it does not necessarily reveal the quality of learning. FoM should be close to 0, but a value below 1 is considered to be largely satisfactory. Criterion (b) is expressed by the error metric FPR by setting ϵ_{th} equal to the measurement repeatability error of the performance. In this case, FPR

gives the expected percentage of chip instances which will not satisfy criterion (b). These chip instances may end up misguiding the calibration algorithm. FPR is preferred from $\epsilon_{\max}$ since $\epsilon_{\max}$ may be misleading in the presence of outliers in the data set. Besides, $\epsilon_{\max}$ is equal to the minimum value of ϵ_{th} that results in FPR=0. It should be noted that the error metric ρ is qualitative and should be taken with a pinch of salt. Ideally, it should be close to 100%, but a much lower value does not necessarily imply a prohibitively high absolute prediction error or that the regression model cannot guide efficiently the calibration. In short, FoM and FPR are the most important and comprehensive metrics for assessing the quality of learning. The rest of the metrics are given for the purpose of completeness since they have been extensively used in previous works.

The results are summarized in Tables 14 and 15. As it can be seen, for all error metrics, the standard deviation is at least one order of magnitude smaller than the mean value, which implies that the reported error metrics are robust and unbiased given the available data set. From Table 14 we observe that the largest FoM corresponds to PAE and is 0.37, that is, it is much lower than 1. The low FoMs indicate an excellent average quality of predictions for all the performances. ρ is higher than 92.65% indicating an excellent correlation between true and predicted values for all the performances. Table 15 shows the FPR for different ϵ_{th} values. The $\epsilon_{\max}$ in Table 14 should be carefully assessed by considering the FPR values in Table 15. For example, regarding the Gain, $\epsilon_{\max} = 6.04$ dB, yet we observe that only 2.75% of chip instances have a prediction higher than 0.5 dB. Regarding the PAE, $\epsilon_{\max} = 15.59\%$, yet only 1.05% of chip instances have a prediction error higher than 2%. It turns out that there are a few outlier chip instances in the data set that give rise to the maximum observed error. These outlier chip instances have tuning knob combinations at the edges of the tuning knobs range and/or have sensor measurement values at the edges of the sensor measurement distributions. In these outer input subspaces there are not enough observations to ensure a robust learning of the regression functions. As a result, the regression functions are extrapolated resulting in a prediction error. The maximum error would have been reduced if we had a larger data set that contains more such "exteme" circuits. Furthermore, such "exteme" circuits can be screened out by a defect filter [79] whose role is to identify suspect circuits that are foreign from the bulk of the statistical data and for which predictions entail a risk.

5.4.4 *One-shot calibration: methodology*

Once the alternate test concept has been proven, we can proceed to prove the calibration concept. Before going into the details of the methodology, we will detail the modification that has been made in the calibraton algorithm in order to accomodate the fact that we do not have common tuning knob ranges for every wafer. Then, we will detail the evaluation procedure and the results.

5.4.4.1 *Optimizer used for calibration*

As explained in the last chapter, optimization has been performed by using the function fgoalattain in the Matlab optimization toolbox. As explained previously, each wafer works in a different tuning knob range. In the optimization procedure, this range must be taken into account so that the optimizer finds a tuning knob combination within this range.
To reduce the measurement time, for each wafer type, we measure the performances of the PA only for the tuning knob settings that lie within the corresponding range. Outside the

corresponding ranges, the PA is not working properly thus we assign values to performances that are false (e.g. clearly outside the range), forcing in this way the optimizer to converge to tuning knob values within the range. Another reason for making this choice is that, since we measure the circuits with bench top standard equipment, all measurements must be carried out in one working day for one wafer, so as to have all circuits as much as possible in the same measurement conditions. With that said, not all measurements could have been carried out for In that case, all measurements could have been carried out for every tuning knob combination in one working day. This would have been possible if we had the opportunity to work on an ATE.

5.4.4.2 *Optimization procedure*

The procedure that we have chosen, detailed in Fig.66(a), is as follows. First the nominal tuning knob setting TK_{NOM} is chosen to be the medium range of the typical wafer measurements. Then, we predict the performances of the circuit under calibration. The prediction functions have only been trained at specific tuning knob range as explained above. If the current tuning knob setting does not lie into this range, alternate test cannot predict accurately the performance since the prediction will be somewhat random. However, we know that if the current tuning knob setting is not within the allowed range of values, the circuit does not work properly. That is why we have decided to give to the optimizer ”fake” non-acceptable performance values so as to avoid going out of the allowed range.
For instance, if the circuit to be calibrated is extracted from the wafer FAST or SLOW, the nominal tuning knob setting is initially out of the allowed range. To drive the optimization algorithm to the right direction, we decided to compute the Euclidian distance between the current tuning knob setting and the median of the allowed range for this circuit and to use it to define fake performance values. The larger this Euclidian distance, the largest the distance will be between the assigned fake performance values from the specifications. An example of this situation is depicted in Fig.66(b). Starting from a nominal tuning knob value TK_{NOM}, the algorithm will first quickly converge to the allowed tuning knob range and it will optimize the performances. This choice that we have made actually resembles reality since we expect that performances are degrading if the tuning knobs are going far from the allowed range. Of course, when the tuning knob settings lie into the allowed range, we give to the optimization algorithm the performances predicted by the alternate test, since the prediction algorithm has been trained at this range. An example is also given in in Fig.66(b) where TK converges to TK_{OPT} into the allowed range.
This procedure enables the optimization algorithm to converge for all circuits in the dataset. At the end of the calibration procedure, each tuning knob setting lies into the predefined range and all performances are predicted with the previously trained regression functions.

5.4.5 *Results*

In Section 5.4.3 we demonstrated that regression models can be learned with sufficient accuracy. In the second step, we employ these regression models for the purpose of calibration and we examine the efficiency of the proposed approach in correcting yield loss and in identifying a more desired trade-off between the performances.

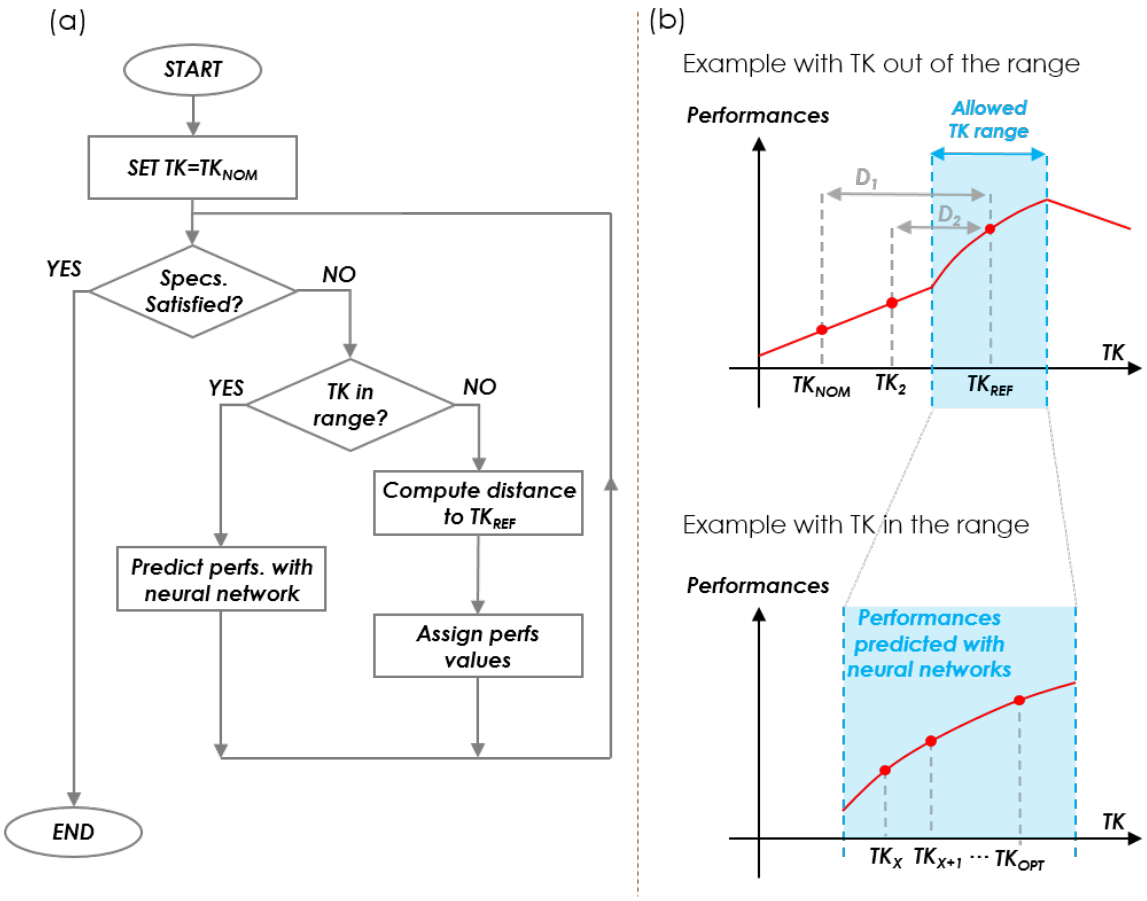

Figure 66: Histograms of chips from FAST wafer after calibration.

We selected a random chip from the TYP wafer and we performed an exhaustive search to identify the optimal tuning knob setting. We consider this setting to be the nominal tuning knob setting. The search resulted in $\mathbf{TK}_{\text{nom}} = [\text{VDp}_{\text{nom}}, \text{VDd}_{\text{nom}}, \text{VGp}_{\text{nom}}, \text{VGd}_{\text{nom}}] = [2.5\text{V}, 2.2\text{V}, 0.69\text{V}, 0.6\text{V}]$. The nominal performance values are calculated as the mean performance values measured on the chips from the TYP wafer at the nominal tuning knob setting. The specification of each performance is set at 3σ away from the nominal value. We consider that the test limits are set at the specifications, that is, the test limit for a performance equals the specification of the performance, such that a circuit fails the test if and only if one or more specifications are violated. Table 16 lists the resultant nominal values and the specifications for each performance.

The chips from the FAST and SLOW wafers do not function properly when applying the nominal tuning knob setting $\mathbf{TK}_{\text{nom}}$. In particular, the gain of the chips from the FAST wafer presents an overshoot before it starts dropping and, thereby, the linearity requirement is violated, resulting in an untrustworthy OCP1 measurement. The chips from the SLOW wafer have a very low gain. In short, the yield of the chips from the FAST and SLOW wafers is 0% and we employ the calibration algorithm aiming at boosting the yield of these two corner wafers.

We assume that available for training the regression functions are all chips from all wafers except the chip that is under calibration. Specifically, if the j-th chip is under calibration, we use the $N \cdot K = 54 \cdot 67 = 3618$ chip instances $[(\mathbf{M}_n, \mathbf{TK}_k), \mathbf{P}_{nk}]$, $n \neq j$, for learning the

Table 16: Nominal performance values and 3σ specifications.

Performance	Nominal value	3σ specification
Gain (in dB)	18.77	>15.6
OCP1 (in dBm)	11.4	>8.58
P_{DC} (in W)	0.106	<0.12
PAE (in %)	3.76	>0.91

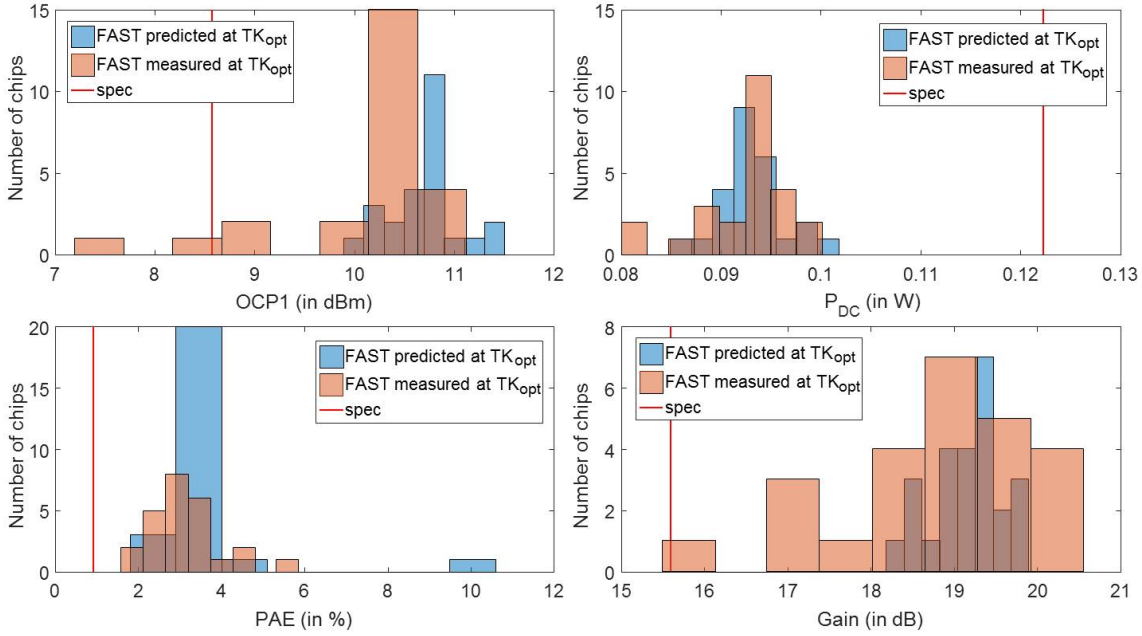

Figure 67: Histograms of chips from FAST wafer after calibration.

regression functions. Thereafter, we launch the calibration algorithm shown in Fig. 44. The algorithm converges to an optimal tuning knob setting $\mathbf{TK}_{\text{opt}}$ that offers the best possible trade-off between the performances. The regression functions offer a prediction of the performances at $\mathbf{TK}_{\text{opt}}$. We also measure the true performances at $\mathbf{TK}_{\text{opt}}$, in order to examine whether the calibration has succeeded.

Fig. 67 and 68 show the histograms of the predicted and measured performances after calibration for the chips from the FAST and SLOW wafers. Table 17 shows for all the performances the yield before and after calibration, as well as the mean and maximum error between the predicted and measured performance values after calibration. Regarding the chips from the FAST wafer, according to the histograms of the predicted values, the calibration has succeeded to recover 100% of the yield loss for all performances. According to the measured values, the situation is slightly different since 2 out of the 25 chips fail by a small margin the specification of OCP1. Thus, for OCP1 the true yield recovery is 96%. In general, the predictions point to the correct calibration decision for all except 2 chips. Regarding the chips from the SLOW wafer, the true yield recovery is 100% and the predictions point to the correct calibration decision for all chips. In short, the calibration resulted in drastic yield recovery and the regression models can be safely used to guide the calibration procedure.

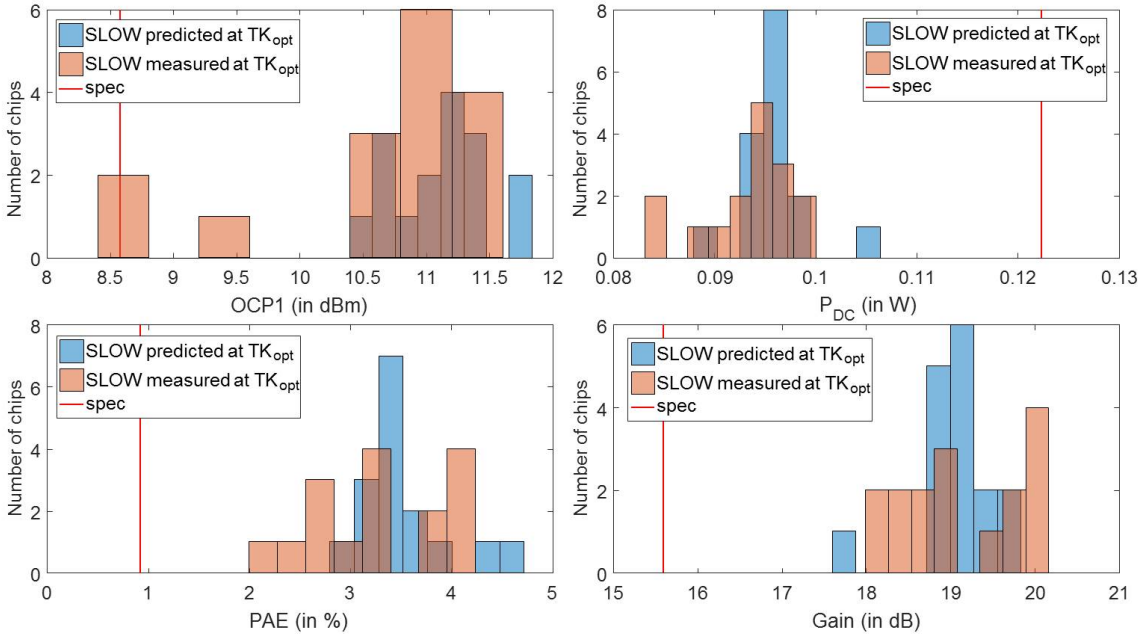

Figure 68: Histograms of chips from SLOW wafer after calibration.

Table 17: calibration results.

	FAST				SLOW				TYP			
	before calibration	after calibration			before calibration	after calibration			before calibration	after calibration		
	yield	yield	mean model error	max model error	yield	yield	mean model error	max model error	yield	yield	mean model error	max model error
Gain (in dB)	0	100%	1.11	3.026	0	100%	0.91	2.46	7.96%	100%	0.95	2.2
OCP1 (in dBm)	0	96%	1.11	3.46	0	100%	1.02	2.55	92.31%	100%	1.12	2.84
P_{DC} (in W)	0	100%	0.0047	0.0127	0	100%	0.0063	0.0127	100%	100%	0.0063	0.0143
PAE (in W)	0	100%	1.69	7.12	0	100%	0.88	2.22	100%	100%	0.85	1.71

The fact that the regression models predicted correctly the calibration result shows that a final standard test at $\mathbf{TK}_{\text{opt}}$ to confirm the calibration may not be even necessary. In this case, it suffices only to perform a fast defect-oriented test to screen out chips with gross defects.

As mentioned already, aside from calibrating non-functional chips, the proposed approach can also be used for achieving a better performance trade-off. For the purpose of demonstrating this experimentally, we consider the chips from the TYP wafer and we employ a starting tuning knob setting $\mathbf{TK}_0$ that is far from the nominal tuning know setting $\mathbf{TK}_{\text{nom}}$ as defined previously. In particular, we considered $\mathbf{TK}_0 = [\text{VDp}_0, \text{VDd}_0, \text{VGp}_0, \text{VGd}_0] = [2.2\text{V}, 2\text{V}, 0.67\text{V}, 0.56\text{V}]$ which is located at the edge of the tuning knobs range. Fig. 69 shows the histograms of the measured performances at $\mathbf{TK}_0$ before calibration and the predicted and measured performances after calibration. Table 17 shows for all the performances the yield before and after calibration, as well as the mean and maximum error between the predicted and measured performance values after calibration. We observe that at $\mathbf{TK}_0$ only 1 out of 13 chips satisfies the gain specification (i.e. yield is 7.96%) while 1 chip also fails the OCP1 specification (i.e. yield is 92.31%). The calibration algorithm was able to find appropriate tuning knob settings for these chips to fully recover yield loss. As it can be seen from Fig.

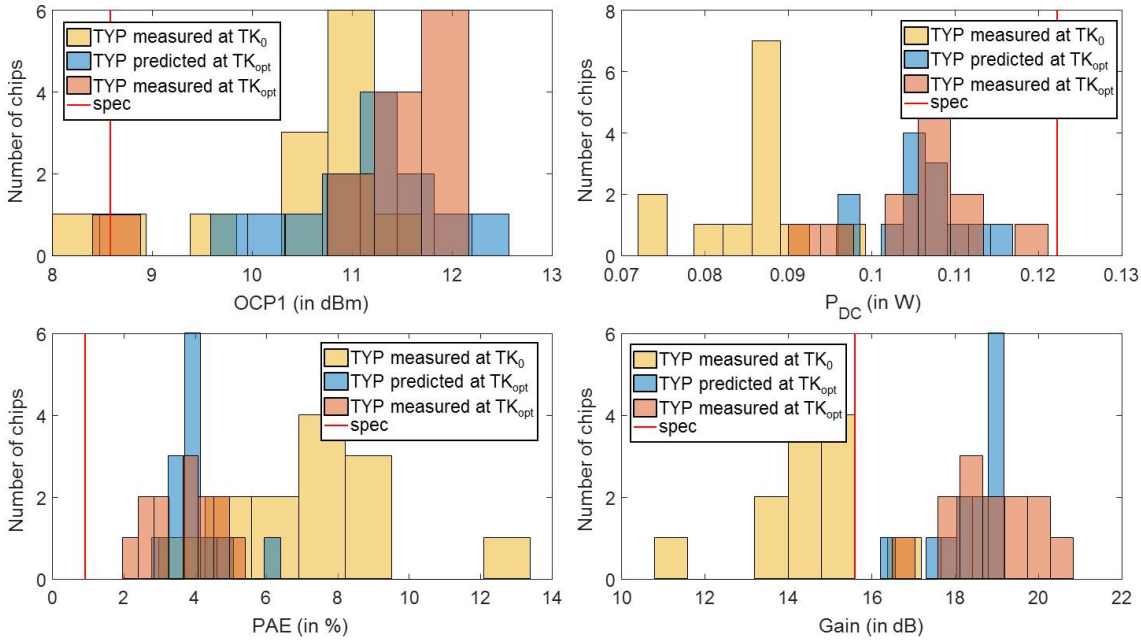

Figure 69: Histograms of chips from TYP wafer before and after calibration.

69, by looking at the shift of the histograms and the sample means, the Gain and OCP1 are improved at the expense of deteriorating P_{DC} and PAE, but P_{DC} and PAE are still within the allowable specification range.

It should be noted that the focus of this work is to provide a proof-of-concept of the calibration methodology. The PA may not have state-of-art performances, but it is functional and it largely suffices to provide such a proof-of-concept. The calibrated performances for the chips from the FAST and SLOW wafers are expected to be as good as the performances of the chips from the TYP wafer. The experiment demonstrates that the performances for the chips from the FAST and SLOW wafers are corrected to lie close to the performances of the chips from the TYP wafer. In other words, the fact that the calibrated performances are not state-of-the-art does not point to a limitation of the calibration methodology, but it is a result of not having a state-of-the-art nominal design. We have no reason to believe that the calibration methodology will not be effective for state-of-the-art PAs. Actually, the yield of state-of-the-art PAs is expected to be lower since such designs are aggressive and intuitively will suffer more from process variations, which makes the need for a calibration methodology even more pronounced.

5.5 CONCLUSION

In this chapter, silicon results are presented to fully prove the one-shot calibration concept. A dataset from fabricated typical, fast and slow circuits has been used. First of all, it can be demonstrated that performances can be accurately predicted by using non-intrusive alternate test. Based on these results, the calibration concept has also been proven. The optimization algorithm had first to be modified in order to handle the particularities of the fabricated chips. The vast majority of the circuits can be correctly calibrated with specifications set to to $\pm 3\sigma$. The yield of the fabricated dataset, which is close to 21% initially, grows up to 93% after calibration. Those results are very promising since more than the two third of the dataset is composed from worst case samples that are very rare in industrial fabrication. Moreover, it has also been proven that if the circuit has a suboptimal performance tradeoff, the algorithm is actually able to bring the circuit close to this optimal state.

6

CONCLUSIONS AND PERSPECTIVES

6.1 CONTRIBUTIONS

RF circuits are more and more susceptible to a large variety of variations since they are pushed forward to be fabricated in the most advanced CMOS technologies. Among those variations, those due to the fabrication process are of primary concern, since the fabrication yield of RF circuits is decreasing along with the technology nodes. That is why a calibration of RF circuits is required in most of the cases.
However, testing and calibrating a RF circuit is very costly, so there is a high need for low cost test strategies, in order to obtain affordable calibration methodologies. Consequently, many challenges have to be faced: lower the cost and time of test, obtain a fast calibration method, and do not disturb the RF circuit while it is under test.

In this work, a novel one-shot non-intrusive calibration methodology for RF circuits is demonstrated. The key achievements of this technique are (a) testing the performances of the RF circuit with low-cost alternate test, (b) use non-intrusive variation-aware sensors that do not disturb the RF circuit while it is functioning and (c) achieving a low-cost calibration in one-shot, that is, by testing the RF performances only once to obtain a suitable tuning knob setting that calibrates the performances. Moreover, power consumption is considered as well as a performance so that calibration have a minor impact on power consumption.

Our proposed methodology has first been demonstrated with simulation data. Afterwards, a case study has been fabricated with circuits extracted from typical and corner wafers. This feature is very important since corner circuits are the worst design cases and therefore the most difficult to calibrate. In our case, corner circuits represent more than the two third of the overall dataset and the calibration can still be proven. In details, fabrication yield based on $\pm 3\sigma$ performance specifications is increased from 21% to 93%. This is a major performance of the technique, knowing that worst case circuits are very rare in industrial fabrication. Moorover, we have also proven that if the initial tuning setting is far from the optimal setting, our algorithm is able to find the optimal tuning setting and consequently improve performance trade off.

To summarize, the main contributions of this work, compared to previous work in the literature, are:

- Proving the non-intrusive test concept with an RF power amplifier at 2.4 GHz
- Developping a novel post-manufacturing calibration methodology that is achieved by:

- Sensing the RF performances non-intrusivly, which means without any eletrical contact with the circuit under test
- Using sensors that are totally independent of the tuning knob setting. In this way, performances are tested once and optimal tuning knob is found in a single without repeating test.
- Using the non-intrusive alternate test methodology that enables to perform low-cost calibration
- Considering the power consumption as a performance in order to limit the impact of calibration on power consumption

6.2 PERSPECTIVES

Moreover, many different perspectives can be envisaged to follow up on this work. First of all, the results presented in this work could be refined by measuring the fabricated circuits with a more advanced measurement setup, that is not currently available in our laboratory. Indeed, a faster and more accurate setup will extend the measurement range and so the possibilities of our proposed calibration methodology.
Then, the calibration technique presented in this work is realized off-line. However, it could be possible to integrate the calibration algorithm on-chip in order to have a complete self-calibrated circuit. In this way, the non-intrusive calibration could be coupled with other adaptation techniques in order to build a general adaptation algorithm so as to take into account other types of variations (temperature, aging, environmental conditions, etc).
Also, since fabricated circuits for corner wafers are available, it could be possible to couple the proposed calibration methodology with Bayesian Model Fusion (BMF) in order to achieve better results. Indeed, with BMF, a few circuits are required to update the calibration algorithm and samples from corner wafer are the most interesting candidates. In order to implement this technique, a more accurate model of our case study has to be developed so as to take all parasitics into account, especially concerning the socket.

PUBLICATION LIST

PEER REVIEWED CONFERENCES

M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "One-shot calibration of RF circuits based on non-intrusive sensors," in *Proc. of 51th Design Automation Conference, San Francisco, USA, June 2014*, 2014, pp. 1–6

M. Andraud, A. Deluthault, M. Dieng, F. Azais, S. Bernard, P. Cauvet, M. Comte, T. Kervaon, V. Kerzerho, S. Mir, P.-H. Pugliesi-Conti, M. Renovell, F. Soulier, E. Simeu, and H.-G. Stratigopoulos, "Solutions for the self-adaptation of communicating systems in operation," in *Proc. of International On Line Testing Symposuim IOLTS*, 2014, pp. 234–239

INVITED TALKS IN CONFERENCES

A. Dimakos, M. Andraud, H.-G. Stratigopoulos, E. Simeu, and S. Mir, "Test and calibration of RF circuits using built-in non-intrusive sensros," in *Proc. of IEEE Computer Society International Symposium on VLSI*, 2015, p. 627

M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "Self-healing of RF circuits based on non-intrusive sensors," in *13th IEEE International New Circuits and Systems Conference, Invited Talk in Special Session, Grenoble, France*, 2015, pp. 1–6

COMMUNICATIONS IN INTERNATIONAL CONFERENCES

——, "Post-manufacturing one-shot calibration of RF circuits based on non-intrusive sensors," in *PhD Forum at Design Automation & Test in Europe, Best poster Candidate*, 2015, p. 1

BIBLIOGRAPHY

[1] M. Onabajo and J. Silva-Martinez, *Analog Circuit Design for Process Variation-Resilient Systems-on-a-chip*. Spinger Science & Business, 2012.

[2] S. Sen, R. Senguttuvan, and A. Chatterjee, "Environment-adaptive concurrent companding and bias control for efficient power-amplifier operation," *IEEE Transactions on Circuits and Systems*, vol. 58, no. 3, pp. 607–618, 2011.

[3] S. I. A. (SIA), "International technology roadmap for semiconductors (ITRS)," http://www.itrs.net/Common/2003ITRS/ExecSum2003.pdf, 2003 edition.

[4] F. Poehl, F. Demmerle, J. Alt, and J. Obermeir, "Production test challenges for highly integrated mobile phone SoCs," in *Proc. IEEE European Test Symposium*, 2010, pp. 17–22.

[5] R. B. Staszewski, "Digital RF and digitally-assisted RF (invited)," in *Proc. IEEE International Symposium on Radio-Frequency Integration Technology RFIT 2011*, 2011, pp. 9–16.

[6] S. Sen, V. Natarajan, D. S., and A. Chatterjee, "Process-variation tolerant channel-adaptive virtually zero-margin low-power wireless receiver systems," *IEEE Transactions of Computer-Aided Design of Circuits and Systems*, vol. 33, no. 12, pp. 1764–1777, 2014.

[7] M. Meghdadi and S. Bakhtiar, "Two-dimensional multi-parameter adaptation of noise, linearity and power consumption in wireless receivers," *IEEE Transactions on Circuits and Systems*, vol. 61, no. 8, pp. 2433–2443, 2014.

[8] S. Bowers, K. Sengupta, B. Parker, and A. Hajimiri, "Integrated self-healing for mm-wave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 3, pp. 352–363, 2013.

[9] R. Khereddine, L. Abdallah, E. Simeu, S. Mir, and F. Cenni, "Adaptive logical control of RF LNA performances for efficient energy consumption," in *Proc. of IEEE International Conference on VLSI and System-on-Chip*, 2010, pp. 161–166.

[10] S. I. A. (SIA), "International technology roadmap for semiconductors (ITRS)," http://www.itrs2.net/2011-itrs.html, 2011 edition.

[11] B. Razavi, *RF microelectronics (second edition)*. Pearson, 2011.

[12] I. Poole, "OFDM orthogonal frequency divison multiplexing tutorial," http://www.radio-electronics.com/info/rf-technology-design/ofdm/ofdm-basics-tutorial.php.

[13] M. Debbah, "OFDM tutorial (in French)," http://www.supelec.fr/d2ri/flexibleradio/cours/cours-ofdm-french.pdf.

[14] S. C. Cripps, *RF power amplifiers for wireless communication (second edition)*. Artech House, 2006.

[15] P. A. J. Nuyts, P. Reynaert, and W. Dehaene, *Continuous-Time Digital Front-Ends for Multistandard Wireless Transmission*. Springer International Publishing Switzerland, 2014.

[16] S. Mahoto and G. Gielen, "Impact of transistor aging on RF low noise amplifier performance of 28nm technology : Reliability assessment," in *Proc. IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS)*, 2013, pp. 413–416.

[17] R. Entner, "Modeling and simulation of negative bias temperature instability," http://www.iue.tuwien.ac.at/phd/entner/node21.html.

[18] A. Goldsmith, "Adaptive modulation and coding for fading channels," in *Proc. IEEE Inf. Theory Commun. Workshop*, 1999, pp. 24–26.

[19] D. Gauntlett, "Demystifying modulation and coding scheme (MCS) index values," http://www.digitalairwireless.com/wireless-blog/2012-09/demystifying-modulation-and-coding-scheme-index-values.html, 2012.

[20] F. Rivet, "Design of a radio frequency front-end receiver dedicated to software-radio for mobile terminals," Ph.D. dissertation, IMS laboratory, University of Bordeaux, France, 2009.

[21] S. Afridi, "Digitally assisted RF circuits," https://afriditech.files.wordpress.com/2013/12/digitally_assisted_rf_seminar_ssa.pdf, 2013.

[22] S. Andersson, "Multiband LNA design and rf-sampling front-ends for flexible wireless receivers," Ph.D. dissertation, Linkping University, Sweden, 2006.

[23] S. Zheng and H. C. Luong, "A CMOS WCDMA/WLAN digital polar transmitter with AM replica feedback linearization," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 7, pp. 607–618, 2013.

[24] K. B. Schaub and J. Kelly, *Production Testing of RF and System-on-a-Chip Devices for Wireless Communications*. Artech House, Inc., 2004.

[25] M. Burns and G. W. Roberts, *An Introduction to Mixed-Signal IC Test and Measurement*. Oxford University Press, 2001.

[26] V. Natarajan, "Self-healing RF SoCs: Low cost built-in and control driven simultaneous tuning of multiple performance metrics," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, USA, 2010.

[27] S. Ellouz, P. Gamand, C. Kelma, B. Vandewiele, and B. Allard, "Combining internal probing with artficial neural networks for optimal RFIC testing," in *IEEE International Test Conference*, 2006, pp. 4.3.1–4.3.9.

[28] I. Bathurone, J. L. Huertas, S. Sanchez-Solano, and A. M. D. Richardson, "Supply current monitoring for testing CMOS analog circuits," in *Proc. IX Conference on Design of Circuits and Integrated Systems DCIS, Sitges*, 1996, pp. 231–236.

[29] W. M. Lindermeir, T. J. Vogels, and H. E. Graeb, "Analog test design with i/sub DD/ measurements for the detection of parametric and catastrophic faults," in *IEEE Design Automation and Test in Europe*, 1998, pp. 822–7.

[30] Y. Kilic and M. Zwolinski, "Process variation independent built-in current sensor for analogue built-in self test," in *Proc. IEEE International Symposium on Circuits and Systems, ISCAS*, 2001.

[31] F. Jonsson and H. Olsson, "RF detector for on-chip amplitude measurement," *Electronic letters*, vol. 40, no. 20, pp. 1239–1240, 2004.

[32] A. Valdes-Garcia, R. Venkatasubramanian, R. Snrinivasan, J. Silva-Martinez, and E. Sanchez-Sinencio, "A CMOS RF RMS detector for built-in testing of wireless transceivers," in *Proc. VLSI test symposium*, 2005, pp. 249–254.

[33] H.-H. Hsieh and L.-H. Lu, "Integrated CMOS power sensors for RF BIST applications," in *IEEE VLSI Test Symposium*, 2006, pp. 234–238.

[34] A. Valdes-Garcia, R. Venkatasubramanian, J. Silva-Martinez, and E. Sanchez-Sinencio, "A broadband CMOS amplitude detector for on-chip RF measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1470–1477, 2008.

[35] J. Altet, A. Rubio, E. Schaub, S. Dilhaire, and W. Claeys, "Thermal coupling in integrated circuits: application to thermal testing," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, pp. 81–91, 2001.

[36] N. Detimple, J. L. Gonzalez, J. Altet, L. Luque, and E. Kerherv, "Design of a fully integrated CMOS self-testable RF power amplifier using a thermal sensor," in *Proc IEEE European Sold-State Circuits Conference ESSIRC*, 2012, pp. 398–401.

[37] E. Aldrete-Vidrio, D. Mateo, J. Altet, M. A. Salhi, S. Grauby, S. Dilhaire, M. Onabajo, and J. Silva-Martinez, "Strategies for built-in characterization testing and performance monitoring of analog RF circuits with temperature measurement," *Measurement Science and Technology*, vol. 21, no. 7, pp. 1–10, 2010.

[38] M. Jarwala, D. Le, and M. S. Heutmaker, "End-to-end test strategy for wireless systems," in *Proc. International Test Conference ITC*, 1995, pp. 398–401.

[39] A. Valdes-Garcia, J. Silva-Martinez, and E. Sanchez-Sinencio, "On-chip testing techniques for RF wireless transceivers," *IEEE Design & Test of Computers*, vol. 23, no. 4, pp. 268–277, 2006.

[40] P. N. Variyam, S. Cherubal, and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 3, pp. 349–361, 2002.

[41] V. Cherkassky and F. Mulier, *Learning from Data*. John Wiley & Sons, 1998.

[42] C. Y. Pan and K. T. Cheng, "Implicit functional testing for analog circuits," in *IEEE VLSI Test Symposium*, 1996, pp. 489–494.

[43] W. M. Lindermeir, H. E. Graeb, and K. J. Antreich, "Analog testing by characteristic observation inference," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 9, pp. 1353–1368, 1999.

[44] C. Y. Pan and K. T. Cheng, "Test generation for linear time-invariant analog circuits," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 46, no. 5, pp. 554–564, 1999.

[45] Z. R. Yang, M. Zwolinski, C. D. Chalk, and A. C. Williams, "Applying a robust heteroscedastic probabilistic neural network to analog fault detection and classification," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 1, pp. 142–151, 2000.

[46] H.-G. Stratigopoulos, S. Mir, E. Acar, and S. Ozev, "Defect filter for alternate RF test," in *IEEE European Test Symposium*, 2009, pp. 101–106.

[47] S. Bhattacharya and A. Chatterjee, "Use of embedded sensors for built-in-test of RF circuits," in *IEEE International Test Conference*, 2004, pp. 801–809.

[48] S. S. Akbay and A. Chatterjee, "Built-in test of RF components using mapped feature extraction sensors," in *IEEE VLSI Test Symposium*, 2005, pp. 243–248.

[49] C. Zhang, R. Gharpurey, and J. A. Abraham, "On-line calibration and power optimization of RF systems using a built-in detector," in *Proc. of IEEE VLSI Test Symposium*, 2007, pp. 398–401.

[50] M. J. Barragan and G. Leger, "A procedure for alternate test feature design and selection," *IEEE Design & Test of Computers*, vol. 32, no. 1, pp. 18–25, 2015.

[51] L. Abdallah, S. H., S. Mir, and J. Altet, "Defect-oriented non-intrusive RF test using on-chip temperature sensors," in *Proc. of IEEE VLSI test symposium*, 2013, pp. 1–6.

[52] L. Abdallah, H.-G. Stratigopoulos, S. Mir, and C. Kelma, "Experiences with non-intrusive sensors for RF built-in test," in *Proc. IEEE International Test Conference*, 2012, paper 17.1.

[53] A. Chatterjee, D. Han, V. Natarajan, S. Devarakond, S. Sen, A. Choi, R. Senguttuvan, S. Bhattacharya, A. Goyal, D. Lee, and M. Swaminathan, "Iterative built-in testing and tuning of mixed-signal/RF systems," in *Proc. of IEEE International Conference on Computer Design*, 2009, pp. 319–326.

[54] V. Natarajan, S. Devarakond, S. Sen, and A. Chatterjee, "BIST driven power conscious post-manufacture tuning of wireless transceiver systems using hardware-iterated gradient search," in *Proc. of IEEE Asian Test Symposium*, 2009, pp. 243–248.

[55] V. Natarajan, S. Sen, A. Banerjee, A. Chatterjee, G. Srinivasan, and F. Taenzler, "Analog signature-driven postmanufacture multidimensional tuning of RF systems," *IEEE Design & Test of Computers*, vol. 27, no. 6, pp. 6–17, 2010.

[56] V. Natarajan, A. Banerjee, S. Sen, S. Devarakond, and A. Chatterjee, "Yield recovery of RF transceiver systems using iterative tuning-driven power-conscious performance optimization," *IEEE Design & Test of Computers*, vol. 32, no. 1, pp. 61–69, 2015.

[57] D. Banerjee and A. Chatterjee, "Signature driven hierarchical post-manufacture tuning of RF systems for performance and power," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 2, pp. 342–355, 2014.

[58] S. Sun, F. Wang, S. Yaldiz, X. Li, L. Pileggi, A. Natarajan, M. Ferriss, J. Plouchart, B. Sadhu, B. Parker, A. Valdes-Garcia, M. Sanduleanu, J. Tierno, and D. Friedman, "Indirect performance sensing for on-chip self-healing of analog and RF circuits," *IEEE transactions on circuits and systems*, vol. 61, no. 8, pp. 61–69, 2014.

[59] F. Wang, W. Zhang, S. Sun, X. Li, and C. Gu, "Bayesian model fusion: large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data," in *Proc. of Design Automation Conference*, 2013, pp. 1–6.

[60] A. Goyal, M. Swaminathan, A. Chatterjee, D. Howard, and J. Cressler, "A new self-healing methodology for RF amplifier circuits based on oscillation principles," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 10, pp. 1835 –1848, 2012.

[61] D. Han, B. S. Kim, and A. Chatterjee, "DSP-Driven self-tuning of RF circuits for process-induced performance variability," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 2, pp. 301–314, 2010.

[62] N. Kupp, H. Huang, Y. Makris, and P. Drineas, "Improving analog and RF device yield through performance calibration," *IEEE Design & Test of Computers*, vol. 28, no. 3, pp. 64–75, 2011.

[63] T. Das, A. Gopalan, C. Washburn, and P. R. Mukund, "Self-calibration of input-match in RF front-end circuitry," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 52, no. 12, pp. 821–825, 2005.

[64] J. Wilson and M. Ismail, "Input match and load tank digital calibration of an inductively degenerated CMOS LNA," *Integration, the VLSI journal*, vol. 42, no. 1, pp. 3–9, 2009.

[65] K. Jayaraman, Q. Khan, B. Chi, W. Beattie, Z. Wang, and P. Chiang, "A self-healing 2.4ghz LNA with on-chip s11/s22 measurement/calibration for in-situ PVT compensation," in *Proc. of IEEE Radio Frequency Integrated Circuits Symposium*, 2010, pp. 311–314.

[66] C. Maxey, G. Creech, S. Raman, J. Rockway, K. Groves, T. Quach, L. Orlando, and A. Mattamana, "Mixed-signal SoCs with in situ self-healing circuitry," *IEEE Design & Test of Computers*, vol. 29, no. 6, pp. 27–39, 2012.

[67] A. e. a. Tang, "A low-overhead self-healing embedded system for ensuring high yield and long-term sustainability of 60ghz 4gb/s radio-on-a-chip," in *Proc. of IEEE International Solid-State Circuits Conference*, 2012, pp. 316–317.

[68] J. Y.-C. Liu, R. Berenguer, and M.-C. F. Chang, "Millimiter-wave self-healing power amplifier with adaptive amplitude and phase linearization in 65-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 5, pp. 1342–1352, 2012.

[69] D. Banerjee, S. Devarakond, X. Wang, S. Sen, and A. Chatterjee, "Real-time use-aware adaptive RF transceiver systems for energy efficiency under BER constraints," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2015.

[70] S. Sen, D. Banerjee, M. Verhelst, and A. Chatterjee, "A power-scalable channel-adaptive wireless receiver based on built-in orthogonally tunable LNA," *IEEE Transactions on Circuits and Systems*, vol. 59, no. 5, pp. 946–957, 2012.

[71] D. Kriesel, "A brief introduction to neural networks," http://www.dkriesel.com/en/science/neural_networks, 2005.

[72] A. Zjajo, M. J. Barragan Asian, and J. Pineda de Gyvez, "BIST method for die-level process parameter variation monitoring in analog/mixed-signal integrated circuits," in *Proc. of IEEE Design, Automation & Test in Europe Conference*, 2007, pp. 1–6.

[73] M. Bhushan, A. Gattiker, M. B. Ketchen, and K. K. Das, "Ring oscillators for CMOS process tuning and variability control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 1, pp. 10–18, 2006.

[74] L.-T. Pang and B. Nikolic, "Measurements and analysis of process variability in 90 nm CMOS," *IEEE Journal on Solid-State Circuits*, vol. 44, no. 5, pp. 1655–1663, 2009.

[75] T. Sowlati and D. Leenaerts, "A 2.4GHz 0.18-μm CMOS self-biased cascode power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1318–1324, 2003.

[76] A. Dimakos, H.-G. Stratigopoulos, A. Siligaris, S. Mir, and E. De Foucauld, "Parametric built-in test for 65nm RF LNA using non-intrusive variation-aware sensors," *Journal of Electronic Testing, Theory and Applications*, vol. 31, no. 4, pp. 281–394, 2015.

[77] M. J. Barragan, R. Fiorelli, G. Leger, A. Rueda, and J. L. Huertas, "Alternate test of LNAs though ensemble learning of on-chip digital enveloppe signatures," *Journal of Electronic Testing, Theory and Applications*, vol. 27, no. 3, pp. 277–288, 2011.

[78] S. Larguech, F. Azais, S. Bernard, M. Comte, V. Kerzérho, and M. Renovell, "Efficiency evaluation of analog/RF alternate test: Comparative study of indirect measurement selection strategies," *Microelectronics Journal*, vol. 46, no. 11, pp. 1091–1102, 2015.

[79] H.-G. Stratigopoulos and S. Mir, "Analog test metrics estimates with PPM accuracy," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, 2010, pp. 241–247.

[80] M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "One-shot calibration of RF circuits based on non-intrusive sensors," in *Proc. of 51th Design Automation Conference, San Francisco, USA, June 2014*, 2014, pp. 1–6.

[81] M. Andraud, A. Deluthault, M. Dieng, F. Azais, S. Bernard, P. Cauvet, M. Comte, T. Kervaon, V. Kerzerho, S. Mir, P.-H. Pugliesi-Conti, M. Renovell, F. Soulier, E. Simeu, and H.-G. Stratigopoulos, "Solutions for the self-adaptation of communicating systems in operation," in *Proc. of International On Line Testing Symposuim IOLTS*, 2014, pp. 234–239.

[82] A. Dimakos, M. Andraud, H.-G. Stratigopoulos, E. Simeu, and S. Mir, "Test and calibration of RF circuits using built-in non-intrusive sensros," in *Proc. of IEEE Computer Society International Symposium on VLSI*, 2015, p. 627.

[83] M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "Self-healing of RF circuits based on non-intrusive sensors," in *13th IEEE International New Circuits and Systems Conference, Invited Talk in Special Session, Grenoble, France*, 2015, pp. 1–6.

[84] ——, "Post-manufacturing one-shot calibration of RF circuits based on non-intrusive sensors," in *PhD Forum at Design Automation & Test in Europe, Best poster Candidate*, 2015, p. 1.