SIGNATURE DRIVEN LOW COST TEST, DIAGNOSIS AND TUNING OF WIRELESS SYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

Shyam Kumar Devarakond

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May, 2013

SIGNATURE DRIVEN LOW COST TEST, DIAGNOSIS AND TUNING OF WIRELESS SYSTEMS

Approved by:

Dr. Abhijit Chatterjee, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Xiaoli Ma
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Linda Milor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Suresh K. Sitaraman
School of Mechanical Engineering
Georgia Institute of Technology

Dr. David Keezer
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: March 12, 2013

To my loving family…

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Professor Abhijit Chatterjee for his constant support and advice throughout the course of my PhD. I would like to thank him for providing me an opportunity to conduct research under his guidance. His energy and motivation has always inspired me to do my best. Over the years, our numerous discussions have had a significant influence in my professional and personal outlook. He is a remarkable person who has made this entire journey highly interesting.

I would like to take this opportunity to thank my committee members, Prof. David Keezer, Prof. Linda Milor, Prof. Xiaoli Ma, Prof. Suresh Sitaraman for taking time to serve on my committee and offering valuable suggestions in regards to my research. I would like to thank Semiconductor Research Corporation (SRC), Gigascale Systems Research Corporation (GSRC), National Science Foundation (NSF) and National Semiconductor for their support during various stages of my research at Georgia Tech.

I would like to thank Nash Khouzam, Dr. Zhenhai Fu, Dr. Soumendu Bhattacharya and John M. Carulli Jr. for their guidance and help during my internships at National Semiconductor and Texas Instruments. The opportunities they provided helped me understand the many practical challenges relating to IC industry. Their timely guidance has provided direction to my research.

I would like to thank all my colleagues in my research group whose support and suggestions have been invaluable during the course of my stay here at Georgia Institute of Technology. I would like to thank Rajarajan Senguttuvan, Vishwanath Natarajan,

Shreyas Sen, Hyun Choi, Mudassar Nisar, Sehun Kook, Jayaram Natarajan, Joshua Wells, Aritra Banerjee, Debashis Banerjee, Debesh Bhatta, Sabyasachi Deyati, Nicholas Tzou, Sen-Wen Hsiao, Thomas Moon, Barry Muldrey, Xian Wang, and Suvadeep Banerjee. I am lucky that my path crossed with all these unique individuals.

I would like to thank my friends Jayaram Natarajan, Vijay Sukumaran, Sourabh Khire, Kaushik Ramachandran, Debashis Banerjee, Rishiraj Bheda, Gokul Kumar, Rohan Verma, Srikaanth Niranjan, Divyanshu Agarwal, Hemant Sane, Bravishma Narayan, Krishna Bharath, Ashok Kumar, Kiran Iyer. Thanks to these people, I had a wonderful time during my time here in Atlanta.

I would like to thank Dr. Sriram Rallabhandi, Dr. Tushar Kumar for their guidance and help during my initial stages of PhD and my stay here in Atlanta. I would like to thank my family members Anuradha Ikkurti and Vijay Yellai for their help during my initial phase of life in US.

Finally yet importantly, I am ever grateful to my parents Sarath Kumar and Vijayasri Sarath for their unconditional love, constant support, and wishes. Their numerous personal sacrifices for my ambitions have made this possible. I am lucky to have such parents who have helped me through all the high and low points of my life. I would also like to thank my brother Shravan Kumar for all his love and moral support. I would like to thank my grandfather Venkateswarlu Ikkurty for his support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RF	Radio Frequency
SISO	Single Input Single Output
MIMO	Multiple Input Multiple Output
DUT	Device Under Test
CUT	Circuit Under Test
E-test	Electrical-test
MB-OFDM	Multi Band Orthogonal Frequency Division Multiplexing
DAC	Digital to Analog Converter
ADC	Analog to Digital Converter
UWB	Ultra Wide Band
EVM	Error Vector Magnitude
BER	Bit Error Rate
CCDF	Complementary Cumulative Distribution Function
SM	Spatial Multiplexing
SD	Spatial Diversity
IFFT	Inverse Fast Fourier Transform
PA	Power Amplifier

ITRS	International Technology Roadmap for Semiconductors
DCT	Discrete Cosine Transform
PDF	Probability Distribution Function
RMS	Root Mean Square
VCO	Voltage Controlled Oscillator
PLL	Phase Locked Loop
SOC	System On Chip
Tx	Transmitter
LUT	Look Up Table

SUMMARY

As the demands for cost efficient, low power wireless communication solutions increases, analog and radio frequency (RF) systems have been implemented in scaled nanometer nodes. At these nodes, multi-dimensional process variations affect the fidelity of these systems resulting in highly unstable yields, in-field wear-out, and signal-integrity problems. Besides random defects, high parametric variations in the fabrication phase have a significant impact on quality and yield of these systems. As technology nodes scale into the extreme nanometer regime, the "guarantee by design" approaches lead to highly pessimistic designs that require significant area and/or power. As a result, the process variations in these systems have mandated the need for increased efforts in post-manufacturing test and diagnosis phases of the devices.

As the trend for integration of complex SoCs continues, the lack of access to internal nodes and cost considerations has increased the testing focus towards system-level test techniques. Traditionally, the measurement of the system specifications requires the use of high cost instrumentation and long test times. These system-level test specifications need to be obtained in an intelligent manner to reduce the overall manufacturing costs of these devices. At the diagnosis-level, traditional process variation management involving the use of test/kerf structures to measure individual process parameter is insufficient in terms of diagnosis resolution and turnaround time.

Finally, it is necessary to develop fast and efficient techniques for performing post-manufacture tuning in the devices for improving the yield and reliability of these systems. Current state of the art techniques for yield improvement through post-manufacture tuning involves performing trimming or switching of various resistors and current sources to compensate for an individual specification. Further, these trimming techniques are, in general, performed sequentially leading to high tune or trim times in the production environment. To offset the parametric performance deviations in these

analog/RF circuits and systems, intelligent post-manufacture compensation framework that can be used to tune concurrently multiple specifications of a system in a time-efficient manner is required. In addition, such techniques need to be able to compensate for process variation with optimal power performance to ensure reliable operation.

In this thesis, the above-mentioned bottlenecks in test, diagnosis, and self-tuning/self-healing are addressed. A concurrent test and diagnosis methodology is developed to obtain the specifications of the DUT as well as the critical circuit process parameters measurements on a per-die basis using specialized diagnostic signatures. The die-level simultaneous test and diagnosis approach aids in identifying process variations and providing feedback for correcting it in a quick and efficient manner. An on-line performance monitoring methodology that uses real-time signal to obtain the performance degradation characteristics of the single input single output (SISO) RF system is investigated. A rapid low cost system-level test technique has been presented for advanced wireless systems such as the multiple input multiple output orthogonal frequency division modulation (MIMO-OFDM) RF systems.

A framework for performing low cost, signature driven post-manufacture tuning of RF systems is developed. Depending on the system computational resources and tuning knobs, different approaches have been detailed for SISO/MIMO-OFDM systems. Considering reliability and power budget limitations of the system, an adaptive power performance tuning approach has been developed that attempts to maximize the device performance considering the system power constraint.

The strategies developed in this work facilitate rapid diagnosis of process variations as well as tune for these variations leading to yield improvement. The solutions presented in this thesis have a significant impact in reducing the manufacturing cost and the time to market of these systems. The outline of the thesis is presented in Figure 1.

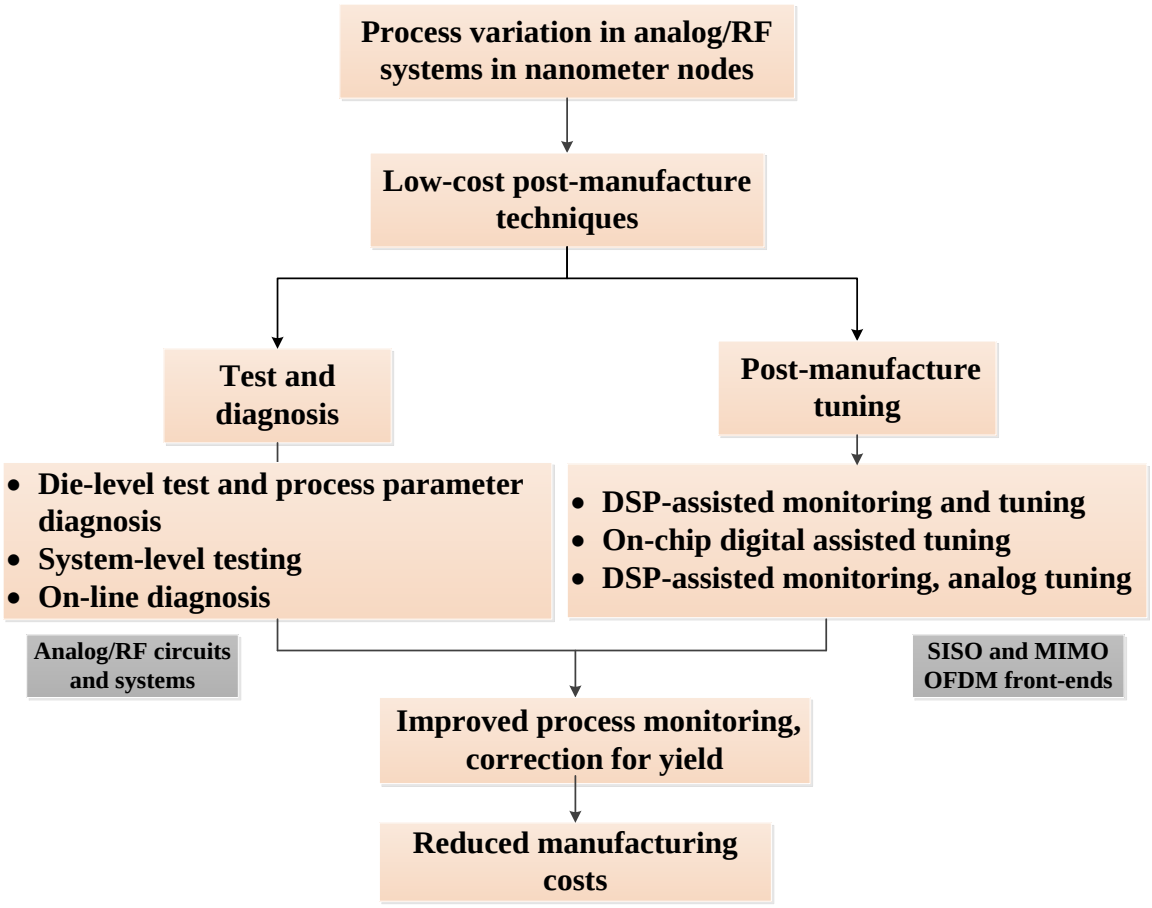


Figure 1: Dissertation outline and scope.

CHAPTER 1. INTRODUCTION

The increasing demands for higher data rates and low power wireless solutions have required the implementation of complex RF systems in nanometer nodes. Increased push towards system on chip (SoC) solutions has made the implementation of integrated wideband multi-standard wireless systems feasible. Commercially available state-of-the-art radio products integrate conventional communication standards along with application specific wireless solutions such as FM radio (100 MHz), RFID (13 MHz), digital TV (800-1600 MHz), GPS (1.5 GHz), Bluetooth (2.4 MHz), 802.11 (2.4 and 5 GHz), WiMax (2.5-3.5 GHz) etc. [1]. The recently adopted IEEE 802.16e and IEEE802.11n standards implement multiple input multiple output (MIMO) front-end chains with multiple antennas in a quest for increased data rates. To reduce the cost of these systems, the integration of multiple chains on a single chip is essential. The ITRS roadmap for RF and mixed-signal technologies developed for wireless applications emphasizes on the integration of various systems and process technology nodes in a single platform as shown in Figure 2. A critical aspect for co-existence of these multi-standard systems is the high fidelity of each of these front-end modules. The high fidelity ensures that there exists minimal interference between the different chains.

However, the resulting circuits implemented in these scaled nodes have been increasingly susceptible to manufacturing process variations. Further, low-power design techniques have made the goal of obtaining high yields at these nanometer nodes a greater challenge. At nodes such as 60 nm and below, process variations in the manufacturing of silicon systems by integrating analog/RF with digital pose a significant challenge to designers [3]. These challenges include problems in signal integrity, verification, and process variations among other problems. As increased efforts are been made to implement more analog and digital IP on the same chip, the optimization of a

technology node for digital performance has led to significant yield challenges for analog components.

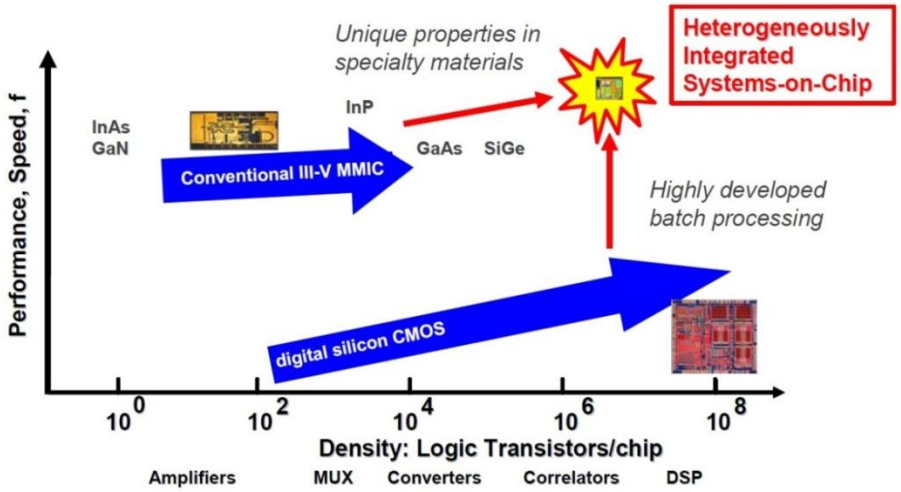


Figure 2: Future trends in wireless systems (Source: ITRS [2]).

Performance deviations (both inter-die and intra-die variations) due to parametric variations are of significant importance in wideband and multi-standard wireless devices over its narrowband counterparts. This challenge is due to the need to maintain the performance over a wider range of frequencies (broadband signals) as well as produce minimal interference. Further, as the number of components or chains in multi-standard wireless devices increase, the possibility of potential yield loss increase rapidly as well. Due to input-output pin constraints, it is highly unlikely that most of these systems have access to internal nodes. Hence, testing these complex SoCs consisting of complex circuitry that have limited access to internal nodes in production environment in an intelligent and low cost manner is critical for commercial success. The low yield returns in these systems further mandate the use of post-manufacturing calibration/tuning methodologies to improve performance metrics.

1.1. Process Variation and Yield loss

As analog and RF systems are implemented in nanometer nodes, the yield of these systems is affected by multi-dimensional process variation. Increased variance in process parameters as the technology nodes scale has been reported extensively in prior literature. In [4], the authors show that due to process variations, the worst-case specification variations in a VCO module is close to 43 %. The authors provide a technique to account for these process variations in the design flow using optimization techniques. In [5], it has been reported that the intra-die variation in threshold voltage has doubled as technology scaled from 130 nm to 45 nm. The substantial effect of drain induced barrier lowering (DIBL) on threshold voltage of CMOS transistors at 65 nm node is shown in [6]. The increased process variations cause the corresponding specifications of the devices to vary beyond the acceptable limits thereby leading to lower yields. To compensate for increased process variations, new design approaches have been formulated. Design centering approaches [7] that try to compensate for the effect of process variations on a given design by using genetic algorithm-based yield optimization on neutral network process models has been investigated in the past. Another approach is designing circuits based on process corners. In this approach, the circuit is designed such that the circuit's specifications are within limits for extreme process parameter variations [8]. However, such a technique leads to over design of the circuit, usually at the cost of higher power consumption.

Current industrial practices for performing process variation simulation at circuit-level involve the use of Pelgrom models [9]. Pelgrom models have been used traditionally, to analyze the effect of mismatches and process variations in a group of transistors. While these are commonly used in the design product development kits (PDK), they are not accurate enough for chips of larger dimensions. Due to the increased process variations, leading to parametric as well as catastrophic variations in the DUT, there exists a need for

development of low cost, time efficient techniques for test and diagnosis of analog/RF circuits. These manufacturing variations, along with other signal integrity impairments such as coupling (arising due to higher levels of integration and packaging), lead to degradation of the overall Quality of Service (QoS) metric in wireless systems. To improve yield in the presence of these high multi-dimensional process variations, it is essential to devise methods of performing post manufacturing tuning at different levels depending on the availability of test and tuning resources.

To summarize, as complex analog and RF systems are implemented in scaled nanometer nodes, for achieving low manufacturing costs and improved yields, the following aspects need to be addressed:

- Fast and low-cost techniques for testing and diagnosis of complex SoCs are required. These include techniques that are suitable for production and in field environment.
- Rapid post-manufacturing techniques need to be developed to increase the yield of the process-skewed instances in the production environment considering power and reliability.

1.2. Conventional Specification Testing of Analog/RF Circuits and Systems

Conventional approach for production testing follows a sequential approach to test for each individual specification of the device. As modern analog/RF system DUTs consists of a slew of specifications in different modes, during the characterization phase of the product, bench test is performed to determine each specification along with its guard bands. During the production-testing phase, a set of critical specifications is identified based on various criteria as well as initial yield characteristics [10]. The DUT is then tested in a sequential manner for these critical specifications using certain test-ordering techniques to minimize the test cost of the DUT. Each test configuration involves a

particular hardware setup, application of the test stimulus, settling time of the DUT response, response capture, and finally analysis. Further, tests of different nature such as DC and AC tests are applied to the DUT using different signal sources and relays are used to switch between the different paths on the load board. With increasing levels of integration, the production test cost of complex wireless systems can amount to as much as 45% of total manufacturing test cost [11]. For example, a wireless MIMO transceiver is characterized by hundreds of specifications in different set of modes as shown in [12]. Some of the critical specifications for these chips include gain, non-linearity or distortion parameters such as input referred third order intercept point (IIP3), I/Q amplitude and phase imbalances, noise figure, DC offsets, transmitter input impedances, transmitter non-linearities, and VCO phase noise.

1.3. Alternative Methodologies for Testing of Analog/RF Circuits

To reduce the test cost and test time of standard specification-based testing approach, a number of techniques that involve concepts such as test ordering, compaction, test generation, designs for test have been developed. The authors in [13][14] provide a good overview of techniques for testing analog, mixed-signal and RF circuits. Defect-oriented testing is a technique that focuses on identifying faults in the system [15]. In this technique, fault models or dictionaries are developed through correlation of data. These fault models are used to avoid tests with high cost or test time. In quiescent current testing (Iddq), the current from various DUTs is monitored to provide information of fault coverage [16]. Structural testing of circuits is another low cost methodology for performing testing of analog/RF circuits [17]. In structural test, as opposed to functional test, the test emphasis is on the circuit implementation rather than the circuit specification. Oscillation-based test technique is discussed in [18]. In this technique, the oscillation frequency of the DUT under certain condition is measured to determine its functional specification. In VDD ramp test, the current signature over time is monitored

as the VDD of the DUT is stepped up, based on which the pass/fail decision is obtained [19]. Alternate test techniques or supervised learning techniques have been used in the past to replace specification tests [20][21][22][23]. In this technique, regression models are developed to predict multiple specifications of the device using the test measurements obtained from the DUT. The DUT is excited by an optimized stimulus such that its output response shows good correlation to specifications in the presence of process variations. Besides building regression models, another low cost test methodology for multiple specifications determination is the use of iterative model-solving techniques for mixed-signal/RF systems. In this methodology, the system is represented using a behavioral model whose parameters can be iteratively changed such that its characteristics match that of the DUT. In [24], random OFDM frames are used as test inputs, to compute a number of specifications by analysis of demodulated data. The use of optimized test stimulus for diagnosis of RF specifications using nonlinear model solving was demonstrated in [25]. In [26], the authors implemented different learning techniques to determine go/no-go tests for RF devices.

Another approach for low cost testing of wireless and wireline systems is the loopback testing. In this technique, the output of the wireless transmitter is fed back to the receiver, with or without frequency translation using an external mixer. The received signal is then processed in the tester or digital baseband (in built in test (BIT) methodology) to determine if the device is good or defective [27][28]. To perform loopback testing without external components, an internal programmable switch can be used between the transmitter output and receiver output. However, the isolation provided by the switch and the non-linearity effects of the switch are critical parameters that need to be taken into account.

BIT methodology is a low cost technique that is particularly useful for complex SOC where access to all the internal nodes might not be feasible. To perform BIT testing of

high frequency devices, various sensors such as the peak detectors, RMS detectors and envelope detectors have been used to obtain the low frequency or DC signal corresponding to high frequency outputs [29][30][31]. To implement BIT methodology, area overhead and signal integrity effects due to the sensor loading are some of the important aspects that need to be considered.

1.4. Diagnosis for Analog/RF Circuits and Systems

In the context of integrated circuit manufacturing, diagnosis is the process of determining the causes of variation/failure in the DUT caused by manufacturing variations. As technology nodes continue to scale, die-to-die and wafer-to-wafer has variations have become significant resulting in inconsistent device yield. Ideally, the circuit should be designed to ensure that the specifications of the device, under process variations, lie within the acceptable region of the device. The acceptable region of a device is a complex multi-dimensional hyper surface that defines a region where all the specifications are within its lower and upper bounds. However, as process technology scales, ensuring compliance with acceptable region limits consistently is a challenge. This challenge is due to the high variability and time-varying effects of process parameters and their interaction effects occurring in production environment. As a result, continuous monitoring of process deviations and correction of these deviations are required for shorter time to market and yield sustainability of the device. This process of monitoring, determine the cause of a parametric shift or failure and providing feedback is referred to as diagnosis in this thesis. As process shifts are tracked, the variation in these shifts is fed back to the fab, to prevent any yield loss.

The causes of failure/variation can be attributed at various levels of manufacturing. These include manufacturing variations such as etch variation, photo resist variations, chemical composition variations, photolithographic variations. These variations are reflected in variations in the Spice-level process parameters such as Tox, Vth0, Lln,

mismatches in capacitances, transconductance of transistors, and other layout parameters etc.

Depending on the nature of the variations, there exist two different types of yields namely functional and parametric yield. Functional yield of a device characterizes the defects occurring due to the random nature of process variations. Parametric yield characterizes the failure of devices due to systematic shifts in process parameter values that cause the system specifications to spread beyond acceptable limits. In deep nanometer nodes, parametric variations and the associated yield problems are more prevalent and challenging as opposed to functional yield problems, which can be pruned out with minimal testing efforts at early stages of manufacturing testing. Diagnosis efforts in this work focus solely on developing an efficient methodology for identifying parametric variations with high resolution without adding significant overhead to existing manufacturing cycle, and providing feedback at the earliest to correct for these process deviations.

Current state-of-art diagnosis efforts rely on electrical test (e-test) measurements and their correlation to parametric variations in specifications. Using the above information, parametric yield loss is characterized and yield management is performed. On each wafer, the e-test parameters are measured at a limited number of test sites. To account for any deterministic e-test parameter variation that would result in yield loss, a significant amount of data needs to be collected for a long timeframe. Faster feedback can be achieved by increasing the number of test structures per wafer. However, the drawback of such a technique would be longer test time and increased silicon area.

In the past, parametric test measurements [32] for zero-yield wafers were used to predict process parameters. Data-mining [33] techniques have been used for diagnosis of process integration errors. In [34], process variation testing using sensing circuitry and frequency domain analysis, to obtain information pertaining to specific regions of a

wafer, is proposed. In [35], the authors use Bayesian theory aided by the construction of fault dictionaries to detect large parametric variations. To increase diagnostic resolution they perform online computations as well. Defect filtering and regression modeling were used to diagnose faults in a LNA circuit in [36]. Certain parameters of the circuits such as R, L, and C are obtained by analyzing the polynomial coefficients of the behavioral models of the analog circuit in [37]. A methodology for diagnosing critical process and circuit parameter values from diagnostic performance measurements (generated from specific measurements at critical nodes of the circuits) using reverse solution of forward regression models (mapping process/circuit parameters to the test measurements) was developed in [38]. In [39], the nearest neighbor residual (NNR) technique is implemented on Iddq tests to help identify outliers.

In digital chips such as microprocessors, a common trend in the recent years is to implement ring oscillators, delay lines for speed measurements, and memory matrices to identify certain MOSFET process parameters such as transistor threshold voltages, channel length, gate capacitance, drain to source resistance etc. [40][41]. These on-die measurements can be used to compensate for local process variations using adaptive bias of substrate etc. However, such techniques are not common for analog chips because of the high area overhead involved in implementing the on-die parametric measurements.

A technique for performing concurrent testing and diagnosis of analog/RF circuits in production environment on a per chip basis is developed in this work. This allows both, the specifications and critical spice-level device parameters, of the analog/RF circuits to be predicted accurately from the DUT response with lower test-time and test-hardware cost compared to standard testing techniques and allows for rapid yield debug and process correction.

1.5. System-Level Specification Testing

Recent years have witnessed a surge for high data rate portable wireless communication devices while low power wireless front-ends have become ubiquitous in day-to-day life. Currently, complete system-on-chip (SoC) implementations of WLAN modules are available in the wireless market. To meet the demand for higher data rates, MIMO systems have rapidly emerged as leading technology. Within a relatively short time, MIMO has been adopted into modern communication standards such as WLAN, WiMax and LTE. To maintain healthy profit margin, with higher levels of integration, it is necessary to develop efficient system-level testing solutions, thereby eliminating the test time and test cost associated with performing individual specification tests for each RF sub-module. While circuit-level specifications characterize the analog/RF modules of the individual circuit, system-level specifications such as error vector magnitude (EVM), bit error rate (BER), transmit spectral mask, receiver sensitivity, complementary cumulative distribution (CCDF), and transmit center frequency leakage can be used to quantify the end-to-end imperfections that exist in the wireless systems. These quantify the effect of all the imperfections that are present in analog/RF front end and the mixed signal circuits (such as ADC and DAC clock jitter, offsets, distortion, quantization noise etc.) as well as the interaction efforts that exist between these impairments.

Testing advanced RF modules, especially for system-level metrics such as EVM, BER, and CCDF etc. is challenging due to the capital and computational costs involved, thereby limiting parallel multi-site testing architectures. In the past, specialized patterns of test stimuli have been used to develop regression-based prediction of system specifications of single input single output (SISO) transceiver in [42]. This technique requires an explicit training phase on hardware measurements and hence uses standard testing procedures for calibration.

A number of techniques have been developed for determining the system-level EVM of SISO systems. An order of EVM computation test time is saved using technique discussed in [43] by sharing data while testing different modulators. A built-in test approach that uses the baseband processor for EVM computation is discussed in [44]. In [45], the authors develop methods for enhancing EVM test using refined EVM computation in conjunction with measurement of path input-output impedances. In [46], the authors discuss a technique for performing EVM testing by rotation of constellation points in a WLAN OFDM SISO system. In [47], the authors propose the decomposition of transceiver performance into static and dynamic non-idealities for calculating the EVM metric of RF SISO systems using regression functions. In [48], the authors propose design of simple digital design for test (DFT) circuits in the baseband of the RF transmitter to determine the overall transceiver EVM. It was shown that EVM could be determined from knowledge of RF system non-linearities and the noise spectrum by expressing EVM as an analytical function of these parameters [49]. All the above techniques have been proposed to determine system-level EVM specification for SISO systems. In [50], the authors provide a low cost methodology for performing BER testing of pulsed transceivers. In [52], a technique is developed for reducing the test time to determine the EVM and CCDF specifications for MB-OFDM UWB systems. The technique uses specialized digital streams obtained through test generation for exciting the DUT rather than the conventional pseudo-random digital inputs used to test the above specifications. These digital streams excite the imperfections of the system in an optimal manner as compared to the standard digital streams thereby reducing the time taken to determine these specifications.

In the case of advanced wireless transmitter modules such as MIMO RF SOCs, the authors in [53] present a low cost technique of measuring MIMO specifications using a combination of signal sources and switches that define a set of RF measurements. The

Gain, IIP3 and the phase offset of various RF chains are measured. Multiple RF chains are tested concurrently by applying tones at specific frequencies to different RF chains in such a way that the fidelity of each chain can be determined by observing spectral content across non-overlapping frequency bands of the observed response. However, the authors determine specifications such as cross coupling by testing the RF chains sequentially. The scalability of the methodology with RF chains is not possible without switching matrices and multiple sources. Further, the authors do not tie the lower-level specifications to higher system-level test parameters like EVM, transmit spectral mask. While commercial systems are available for performing system-level testing of different MIMO system specifications [54], these incur high cost in terms of the test equipment and test time.

In this work, a methodology for performing parallel system-level testing of MIMO-OFDM RF transmitter modules using optimized bandwidth-partitioned stimulus applied from the embedded DSP module of the RF system and a simple combination of sensors on the load board. A comprehensive set of behavioral specifications of multiple RF modules chains are computed simultaneously from the observed DUT response using a single data acquisition and are used to compute the system-level specifications. As multiple chains of the MIMO front-end system are tested concurrently, the presented technique enables lower test cost and test time.

1.6. Prior Work in Tuning of Analog/RF Circuits

As analog/RF circuits approach the nanometer regime, process variations assume great significance in the overall yield. The emphasis on design for manufacturability (DFM) and design for yield (DFY) in the case of analog/RF circuits as technology is scaled is highlighted in [55]. The two essential components of yield improvement through post-manufacturing tuning or calibration are as follows:

- Existence of a control unit that runs algorithms to estimate or monitor the imperfections in the DUT due to process variations. This can be the system DSP or the tester in production environment.
- Analog or digital tuning "knobs" for correcting the imperfections at various points of the system. Compensation of the non-idealities of a RF system can be performed both in the DSP domain using digital linearization techniques such as pre-distortion (digital knob) and in the analog domain using techniques such as feedback, tuning of varactors, inductors, bias current.

The tuning and control strategy that is adopted depends on numerous factors such as area, power, and complexity of the chip. Based on the above two factors, various tuning techniques that have been proposed in prior literature can be broadly classified as one of the following categories:

- Digital baseband (also known as the digital signal processor (DSP)) monitoring and tuning
- On-chip monitoring and analog tuning
- Digital baseband monitoring and analog and digital co-tuning

The overview of the tuning techniques is shown in Figure 3.

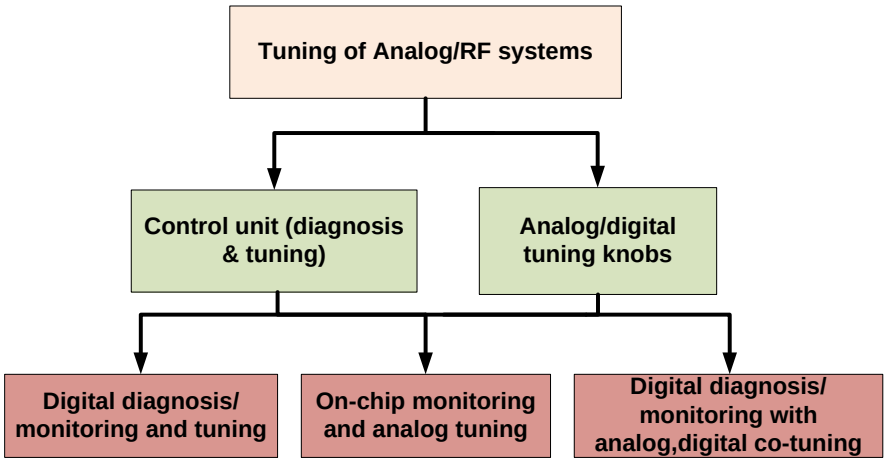


Figure 3: Overview of the tuning concept.

1.6.1. Digital Baseband or DSP-Based Monitoring and Tuning

This methodology involves estimating the imperfections in the front ends using measurements obtained from sensors/auxiliary components present in the front ends and correcting the imperfections in the baseband. The baseband is either before the transmitter front end or after the receiver front end. The imperfections that are corrected include DC-offsets, I/Q mismatches, and non-linearities. The advantage of using digital baseband for monitoring and correction is the high level of controllability in terms of resolution and precision that can be obtained. Unlike analog tuning knobs, no special hooks need to be incorporated into the front-end circuits for post-manufacture tuning.

Further, these techniques can be characterized as direct learning or non-adaptive and indirect learning or adaptive methods based on the implementation technique. In adaptive compensation, continuous online monitoring and calibration of the system is performed by using techniques like least squares (LS), least mean squares (LMS) or recursive least squares (RLS). In [56], a digital compensation technique is described for correction of I/Q impairments up to 10% amplitude mismatch and 10° of phase mismatch. Adaptive techniques for I/Q mismatch compensation, DC-offset cancellation is described in [57] [58]. Non-iterative techniques based on one-time characterization for quadrature compensation is presented in [59]. For non-linearity correction, digital predistortion is a widely used linearization technique. The popularity of the above technique can be attributed to its flexible adaptive nature as well as the extent of implementation accuracy at lower cost. Baseband predistortion creates inverse characteristics of the transfer function of the RF power amplifier, thereby increasing the effective 1dB compression point of the power amplifier. Adaptive predistortion uses methods such as the recursive least squares (RLS) [60] [61] algorithm to tune the predistorter coefficients during field operation. Standard techniques used for adaptive predistortion perform tuning of the predistortion coefficients using the receiver down conversion chain [60]. The power

amplifier output is downconverted, demodulated and fed back through the internal receiver chain. The signal is then processed digitally to improve the overall end-to-end linearity. The digital correction and tuning technique, however, is affected by receiver LNA and down conversion mixer non-linearities, I/Q demodulator mismatches, and ADC non-linearities arising in the down conversion procedure. Further, the drawback of such an adaptive compensation technique is the time the system takes to converge to its optimum performance. In non-adaptive compensation, a one-time calibration of the system is done during production test and the results are stored in a look up table (LUT) and used in real time operation [59]. The drawback of this methodology is that it does not take care of ageing related performance drifts. Digital compensation techniques cannot correct large impairments due to the limitations in the dynamic range of the data-converters in the system, amplification of DC-offset or saturation of front-end analog modules. However, these techniques are sufficient for correction of amplitude and phase impairments arising due to I/Q mismatches and DC-offset errors.

1.6.2. On-Chip Monitoring and Analog Tuning

In this methodology, the impairments are estimated by circuitry on the chip and are from the estimated performance criteria the device circuit parameters are tuned. These techniques can be analog or digital in nature. In analog tuning or compensation, the calibration/tuning is performed in the analog domain by changing the circuit parameters such as bias, supply voltage, passive components such as capacitors, inductors, and resistors. Generally, this methodology involves some form of circuit-level feedback. A completely analog scheme for tuning I/Q mismatches and the resulting image rejection ratio (IRR) performance in a two-stage down conversion receiver is presented in [62]. In this technique, an auxiliary chain performs a two-stage down conversion by either using In-phase or Quadrature phase LO at any point of time. The final output of the main transceiver is mixed with the output from the auxiliary chain to obtain the mismatch

values in the DC component after low pass filtering. More than 15 dB of IRR performance improvement is achieved using this technique. Analog pre-distortion for non-linearity improvement is performed using matching network design with independent amplitude and phase control [63]. A self-calibration scheme for tuning the impedance mismatches of a LNA by using a variable inductor with taps is presented in [64]. The technique involves real-time current sensing whose magnitude varies with the input match of the LNA, which is then amplified and peak detected to obtain a dc signal corresponding to the input match. To determine the frequency offset in matching, two tones of different frequencies are applied one after the other and the difference in the corresponding DC voltage is used to calibrate the shift in S_{11} of the circuit. In this technique, the inductor has taps with switches connecting to these taps. The parasitics of the switches as well as those of the interconnects need to be characterized. The advantage of using on-chip monitoring and tuning is the fact that short calibration time (in the order of microseconds) is achieved as opposed to digital baseband calibration schemes (generally in the order of milliseconds). The bandwidth of the analog feedback system generally decides the convergence in this methodology. In the digital monitoring design paradigm, on-chip digital logic is used to monitor and aid the compensation of mixed-signal/RF performance due to process variations. By implementing such a technique, the design focus is shifted from the analog circuitry to the digital circuitry. There has been significant work in the area of digitally assisted tuning of analog circuits including PLLs, frequency synthesizers, and ADCs. In the past, there has been numerous digital techniques for calibration of PLL, which is predominantly a digital block [65][66][67]. A LMS based technique for calibration of pipelined ADC is presented in [68].

In this methodology, there is no requirement for interaction with the system baseband leading to possibly lesser number of pin counts. This technique reduces the load on the system baseband processor. However, as opposed to analog circuit, voltage or current

sampling for feedback is difficult to achieve in RF circuits without actually affecting the output response. Depending on the type of feedback circuitry, the PVT variations in the feedback circuitry itself can affect the calibration performance.

However, the above research examples focus primarily on testing and tuning of specific analog/RF/ADC specifications. There is no generic scalable methodology that can concurrently tune multiple design specifications using on-chip circuitry. In this work, a self-contained on-chip digital-logic-based tuning methodology is developed.

1.6.3. Digital Baseband Monitoring, with Analog and Digital Co-tuning

In this methodology, the monitoring and control unit is the DSP of the transceiver system and numerous knobs in the baseband as well as analog front end have been incorporated in the system to tune for front-end imperfections. Such a methodology has gained significant attention in the past decade. The bias voltages of various front-end circuits are controlled using DACs that can be configured in post manufacturing phase for tuning. In [69], the authors use current DACs to compensate for the LO feed through (LOFT). The output of the transmitter is envelope detected and the frequency spectrum of the output calculated in the DSP is used for impairment correction. In [70], the authors develop an adaptive technique to compensate for I/Q mismatches observed in typical RF front ends, by making use of a variable delay gain cell to feedback correction vectors to the system LO. In [71], a dual mode 802.11b/Bluetooth radio along with tunable bias for LNA, Mixer and LOFT cancellation. In [72], the authors propose a scheme for compensating static and time varying dc offset as well as TX LO leakage using the baseband processor. There exist a number of internal calibration paths, which are controlled through switches, for calibrating the front-end components during test mode. In [73], a technique for tuning of RF front-end modules by running a gradient algorithm that attempts vary the output response of the transmitter to the reference output is presented.

To develop a monitoring scheme that is generic for performing multi-dimensional tuning, it is necessary to observe the variation in specifications simultaneously. Alternate test methodology or supervised learning methodology is well suited for this purpose [20][21][22][23]. The authors in [74] develop a one-time tuning technique in which the values of the analog tuning knobs of a LNA are directly predicted using supervised learning techniques. The methodology leads to lesser calibration time. However, it is limited by the accuracy of the regression models and does not guarantee power optimality. In [75], the circuit is reconfigured to oscillate in test mode and the oscillation signature in conjunction with regression models are used to tune the amplifier for yield improvement. A critical aspect of the analog tuning is the availability of hooks in the front-end circuits. Various papers have been published in the past that discuss the various tunable circuits. In [72], current steering DACs are used for dc-offset cancellation. Conversion gain of each I and Q path and IIP2 correction technique in mixers have been proposed in [76]. The conversion gain can be used for tuning of I/Q amplitude mismatch. In [77], an orthogonal tunable LNA in which the gain and the non-linearity characteristics of the device can be controlled independently without affecting each other is presented. A technique for controlling the input match, gain, and resonant operating frequency of LNA is presented in [78]. Baseband analog filter tuning can be achieved using RC reconfigurable filters [79]. This scheme is relatively easily to implement as the signal is in the low-frequency domain where the parasitics have less impact. A self-healing 60 GHz PA amplifier implemented in 65 nm technology nodes that can self-tune its gain and 1dB compression point is presented in [80]. Three control knobs operating at different power-levels are used for tuning.

With respect to the above tuning techniques, multiple specifications are tuned sequentially one after the other leading to considerable tune times. Further, in each step of the tuning, standard specification testing is performed. In this work, a framework for

performing rapid low cost iterative tuning of multiple RF systems in a power conscious manner using the system DSP is presented. Finally, none of the prior techniques investigates the effect of the power increase due to self-tuning. In this work, an adaptive tuning methodology is developed that attempts to optimally trade-off performance metrics against power, i.e. a methodology that maximizes performance of the device for a given power constraint is developed.

CHAPTER 2. PRINCIPLES OF SUPERVISED LEARNING AND MODEL-SOLVING-BASED TEST TECHNIQUES

2.1. Supervised Learning or Alternate Testing Technique

As explained in chapter 1, conventional test strategies for mixed-signal, analog and RF circuits are based upon specification-based testing techniques where the DUT specifications are measured using one or more input signals with possibly different test setups for each specification. These specifications are compared against pre-defined specification bounds to make device pass/fail decisions. This requirement of multiple test stimulus and different test set-up for each specification results in increased costs in terms of test time and test infrastructure. Alternate testing or supervised learning methodology involves the use of a single test configuration and predicts multiple specifications of interest with a single highly optimized input stimulus, thereby, reducing the overall test time and test cost. The fundamentals of alternate test or supervised learning framework are briefly described in this chapter.

Consider variations in the process parameter space P as shown in Figure 4 that affect the specifications of the DUT. For low cost diagnosis, an alternate set of measurements is determined such that the test measurements under process variations are strongly correlated with variations in the test specification values of the DUT. This set of measurements defines the measurement space M. Any deviation in the observed measurements from the expected implies a corresponding deviation of the measured RF specifications of the DUT from the expected in the specification space S, due to perturbations in the process space P. Hence, if an optimized stimulus is used such that the corresponding output response of the DUT varies due to process variations resulting in specification variations, then a model can be developed relating the specification (S) and measurement (M) domains. This model can be stated as follows:

$$S = f(M). \qquad \text{Equation 1}$$

There exist a number of different techniques for developing the regression mapping function f relating the test measurements and the specifications. In alternate test technique, a nonlinear regression model is developed using the technique called Multivariate Adaptive Regression Splines (MARS) [81]. The devices are excited by an optimized stimulus and the regression model is developed using measurements obtained on a "training set" of devices and the corresponding specifications. In the production environment, this model is then used to predict the RF test specifications S of the DUT from the observed alternate test measurements M as shown in Figure 4. The MARS algorithm selects a set of basis functions (linear or higher order) using the input variables. It also selects the coefficients for the basis functions to develop the regression function. Based on the input data variations, the MARS algorithm uses the concept of recursive partitioning to develop the model. The MARS algorithm consists of two steps that are referred to as the forward step and the backward step. In the forward step, basis functions are added to the model. In the backward step, the basis functions that contribute minimally to the least square fit of the model are removed. The backward step is evaluated using the generalized cross-validation error (GCVE) criterion. The criterion balances between over-fitting of the data and the residual error. The failure coverage of the alternate testing technique depends on the choice of the input test stimulus.

Learning or training: The optimized test stimulus is applied to each of these DUTs and their resulting responses are sampled and stored. Simultaneously, the output specifications of these devices are measured using the conventional specification-based test set-ups. The sampled transient measurements are mapped onto the specifications of the device using non-linear regression functions. During production test, this non-linear regression function is used to predict the specifications of a DUT from its response to the same optimized test stimulus. The production test implementation of alternate test or

supervised learning technique is shown in Figure 5. The requirement of standard test setup during the training phase is a disadvantage of the supervised learning technique.

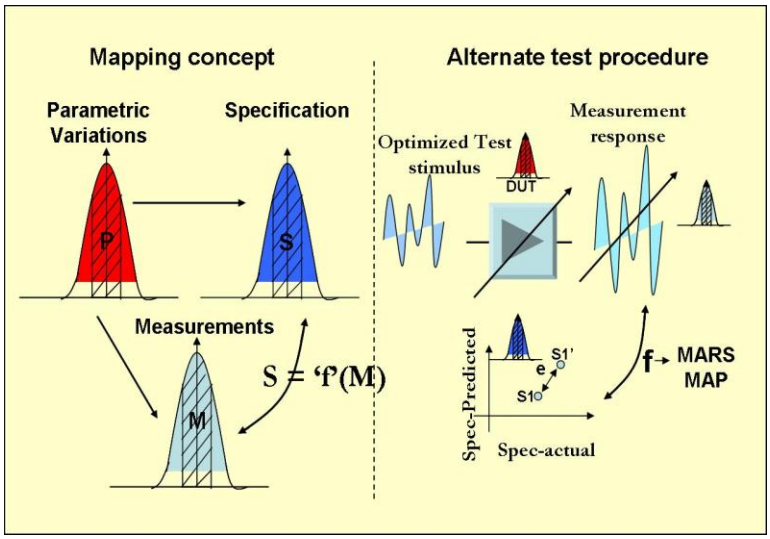


Figure 4: Alternate test or supervised learning technique.

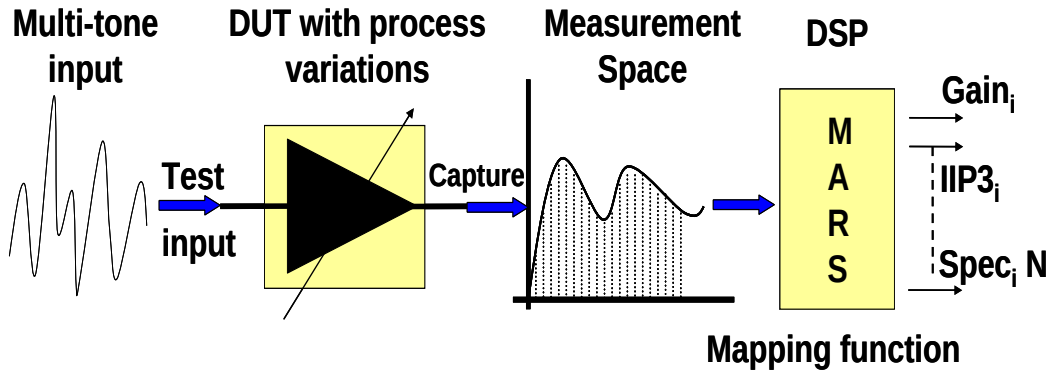


Figure 5: Implementation of supervised learning in production floor.

2.2. Model-Solving-Based Testing/Monitoring

Model-solving-based testing/monitoring, as the name suggests uses a parametric behavioral model of the system for testing purposes. In this technique, an optimized stimulus, obtained through pre-computed test generation scheme is used to excite the DUT as well as a behavioral model of the DUT simultaneously [24][25]. The output from both is compared by computing the least square error between their responses, which is

used to guide a non-linear solver in updating the behavioral model of the DUT. When the error in response is minimized, from the computed behavioral parameters, the system specifications can be computed using correlation techniques or through simulating the behavioral parameters in the software domain to compute the DUT specifications. Many different non-linear optimization techniques can be used for obtaining the behavioral parameters of the system. In this work, we use the trust-region-reflective algorithm for determining the behavioral parameters [82]. The bounds on the values of the behavioral model parameters of the DUT need to be determined during the characterization phase devices. Alternatively, the model can be solved directly without using any iterative techniques. The choice of the technique depends on the availability of computational resources, model complexity, and accuracy required in determining the model parameters.

In this technique, it is assumed that the behavioral model closely matches the behavior of the device. The concern with this technique is that behavioral parameters must accommodate uncertainties in measurement, which can be alleviated using a carefully calibrated and accurate measurement setup. The overview of the technique is shown in Figure 6. The advantage of this technique is that it avoids the need for a training phase and cost associated with it.

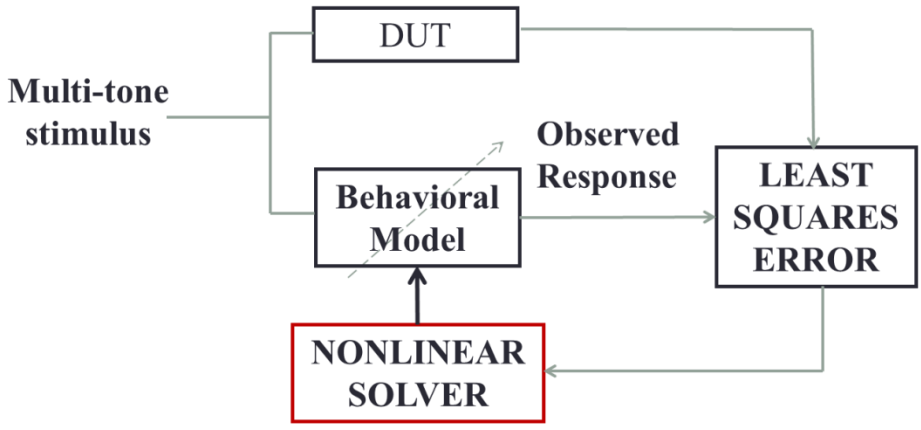


Figure 6: Model-solving-based testing.

2.3. Test Optimization

The test optimization procedure is a one-time offline optimization that needs to be performed to ensure that the input test stimulus is carefully optimized to result in an output response that is highly correlated with the specification values of interest. This forms the basis of supervised learning technique as well as model-solving-based test technique. This step can be either performed in simulation environment on the DUT circuit or on the initial set of characterization devices (provided there is a considerable process variation distribution among these devices). The optimization of the test stimulus depends on the type of DUT being tested and the performance parameters being evaluated. For a given DUT, there are several test generation algorithms available for optimizing the input test stimulus. An initial population of the DUT is selected such that it covers the entire process space with reasonable accuracy and the optimization is performed on this set. In the simulation environment, more than the expected process variations (which is in general $\pm 3\sigma$ where σ is the standard deviation of process parameter) are used to ensure that the process space of the DUT is well represented. The effectiveness of each test stimulus at each stage of test generation is evaluated by using the captured test response from all the devices and using the test technique (either supervised learning technique or model solving based technique) to compute their specifications. The prediction error in the specifications is used to drive the optimization. In this thesis, various versions of genetic optimization algorithm are used to determine the optimized stimulus to excite the DUT.

CHAPTER 3. TEST AND DIAGNOSIS OF RF CIRCUITS AND SYSTEMS

With increasingly shorter product cycles, yield entitlement is a critical parameter for commercial success of semiconductor devices. In the case of analog and RF circuits implemented in advanced nanometer nodes, attaining high yields is a challenge. Higher levels of circuit integration and implementation of circuits in scaled nanometer nodes has made the DUT test and diagnosis steps a crucial step in the manufacturing cycle.

In scaled technology nodes, to maintain good yield, it is necessary to monitor process variations continuously. Process variation management involves continuous process monitoring and feedback to offset process deviations due to environmental and thermal effects thereby improving chip yield. Current industry practice has two lines of defense that prevent misclassification of devices and thereby affect the yield of a device. These lines of defense are the fab parameter limits and test specification limits. The overall yield of a given device is determined by these limits.

At the fabrication level, a number of statistical control methods are used to monitor individual e-test parameters. The values of the e-test parameters are used to decide whether a wafer is good or defective. In general, a fab monitors a large gamut of e-test parameters based on the process technology used. These parameters are monitored through control charts where each e-test parameter is expected to be within an upper control limit (UCL) and a lower control limit (LCL). Current industry practice to determine yield/specification compliance of analog/RF devices at the test floor involves testing the DUT through standard specification testing techniques. To determine if the device has passed or defective, various specifications of a device are checked against their respective specification limits. In general, analog/RF circuits are characterized through a multitude of specifications. For each individual specification, the standard

testing technique involves a distinct test setup and utilizes expensive instrumentation that increases test time and test cost. Further, these testing techniques do not provide any insight into process parameter variations. While the above-mentioned test and fab checkpoints exist, each of these checkpoints operates independently.

3.1. Concurrent Test and Diagnosis of Analog/RF Circuits

The problem with the existing technique is that it does not take into account the sensitivity criterion that exists between a given e-test parameter and a test specification. Further, on each wafer, the e-test parameters are measured at a limited number of test sites. Thus, to account for any deterministic e-test parameter variation that would result in yield loss, a significant amount of data needs to be collected across a large interval of time. In production environment (time t = 0), it is essential for performing the above mentioned test and diagnosis steps in a cost effective manner and in a time efficient manner. In the present process control and monitoring techniques, a significant volume of data across multiple lots of wafers needs to be stored and analyzed before high confidence in parametric process diagnosis is possible. In general, multiple reasons for process-related deviations are generated and the feedback cycle can take several weeks. This problem becomes severe in the case of analog/RF technologies, which suffer significantly from parametric yield issues that force higher power consumption or lower yield design solutions. A further challenge is in the development of SoC and SiP packaging solutions that are in high demand across advanced product lines and the need to use known good die (KGD) in these packaging solutions. Hence, there exist a need for developing efficient and rapid test and process diagnosis techniques for analog/RF technologies.

In this work, an efficient methodology for die-level test-and-diagnosis for analog/RF circuits is developed. The key contribution of this work lies in the ability to both determine the DUT specifications as well as the underlying Spice-level model parameters

from the same DUT test response on a per-chip basis, thereby providing higher diagnostic resolution. A computationally efficient algorithm enables the test and diagnosis procedure for test stimulus generation. The algorithm simultaneously targets test sensitivity and Spice-parameter diagnosability. This allows cause-effect analysis to be performed that relates perturbations in the Spice-level model parameters to the DUT performance metrics (or specifications). Further, the cause-effect diagnosis is achieved at a test cost comparable to prior testing schemes that target only pass/ fail classification of tested devices [20][21][22]. The overview of the proposed technique is shown in Figure 7. Concurrent process diagnosis and testing enables of circuits enables rapid feedback providing insight into process variations occurring in volume manufacturing.

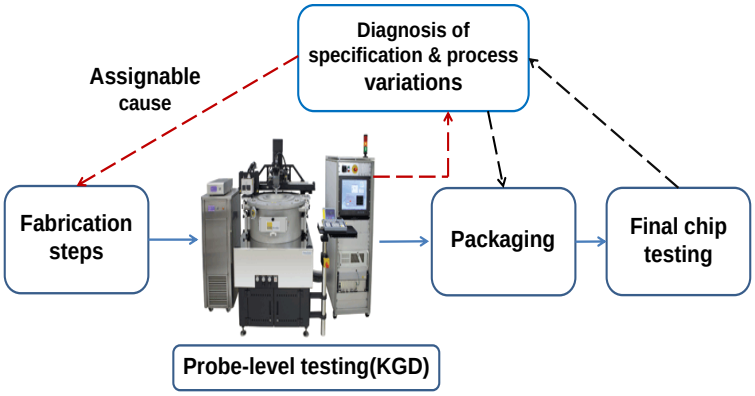


Figure 7: Proposed concept.

3.1.1. Proposed Methodology

In general, the specifications of analog or RF circuits are continuous functions of the circuit or device parameters. In the absence of a hard or catastrophic fault, significant information about process variations is contained in the DUT response. If the test-stimulus is designed (e.g., by varying its amplitude or frequency or both) to exhibit strong statistical correlation with its performance (test) specifications, then a forward regression mapping $x_1 : M \rightarrow S$(also denoted as x_{1ms} - such a notation will be used throughout the chapter) can be built by mapping the test-response measurements (M) to the device

specifications (S)(see Figure 8). This regression model is built in the circuit simulation environment by obtaining several devices through statistical sampling of an extended (to cover the range of expected process shifts) process space (P). Using these devices, a regression model ω_1 that maps measurements (M) of the devices to their specifications (S) is built. A correction regression function ω_2 is then determined using the measurements obtained on a set of hardware training devices (characterization wafer lots) such that x_{1ms} is the composition of ω_1 and ω_2, i.e., $x_{1ms} = \omega_2(\omega_1)$. This is done to compensate for calibration errors in hardware measurement. Initially the regression function ω_1 is used on the characterization wafers to predict their specifications. Then the actual specifications of the DUTs are measured using standard specification testing. Finally, a map ω_2 is built between the actual specifications of characterization wafers and the predicted specifications. From the same statistical sampling procedure used for building ω_1, a reverse regression model $x_2: M \rightarrow P$ is computed by mapping the measurement space M to the Spice-level model parameter space P (Figure 8). Since the actual Spice-level parameters of a tested IC are not known, it is not possible to apply any correction to the mapping x_{2mp} to account for modeling inaccuracies. In this case, it is assumed that the simulation models have been accurately calibrated through independent measurements on electrical test structures [83] [84]. At the minimum, in the case that any residual process modeling inaccuracies persist, the proposed diagnosis scheme will track perturbations in the process-parameters and specifications. Since measurements are made on a per-chip basis, the focus is to capture the systematic inter-chip variations as opposed to intra-chip random variability effects. It should be noted that a defect-filter is used to determine DUTs with catastrophic or large parametric defects. The defective devices that are determined by the defect-filter are not used to build the mappings x_{1ms} and x_{2mp}. The test methodologies developed in prior literature such as in [31][21][20] using Alternative test or Supervised learning are targeted exclusively for specification-based

testing and do not guarantee high correlation to process-parameters. Hence, these techniques, as opposed to the presented methodology, cannot reveal why certain specifications fail.

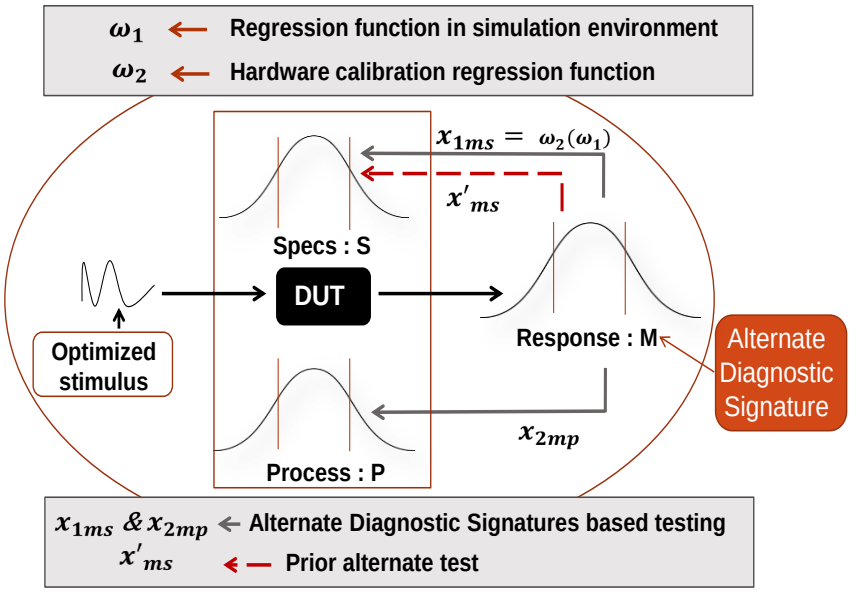


Figure 8: Overview of the methodology.

3.1.2. Test Generation

For effective diagnosis, the response of the DUT to the test stimulus used should have a strong correlation to the Spice-level parameters and specifications. In addition, for practical implementation, the test generation algorithm should be computationally efficient. The test generation process is a one-time process that is performed in software using the DUT circuit design. For obtaining the optimum stimulus, a Genetic algorithm is used. The algorithm minimizes a cost-function that is based on a differential sensitivity criterion. The sensitivity matrix criterion is used to avoid exhaustive statistical simulation of DUT during the test optimization phase. The proposed theory holds good for the case of small variations about the nominal values of process parameters.

In this work, the process domain of the DUT is represented by a vector $P_j = [p_1, p_2, p_3 \ldots, p_j]$, the output measurement domain of the DUT comprising of n response measurements is represented by a vector $M_n = [m_1, m_2, m_3 \ldots, m_n]$, and the specification domain of the DUT comprising of k specifications is represented by a vector$S_k = [s_1, s_2, \ldots s_k](k, j \ll n)$. A change δP_j in Spice-level parameters causes a change δM_n in measurement parameters and δS_k in specification parameters. Considering the first-order approximation, the Spice-level parameters and measurements can be related as

$$\delta M_n = A_{nj}\delta P_j, \qquad \text{Equation 2}$$

where A_{nj} is the sensitivity matrix defined below in Equation 3:

$$A_{nj} = S_p^m = \begin{bmatrix} \frac{|\delta m_1|}{\delta p_1} & \ldots\ldots\cdot & \frac{|\delta m_1|}{\delta p_j} \\ \vdots & \ddots & \vdots \\ \frac{|\delta m_n|}{\delta p_1} & \cdots & \frac{|\delta m_n|}{\delta p_j} \end{bmatrix}. \qquad \text{Equation 3}$$

The ability to diagnose the Spice-level parameters can be evaluated using singular value decomposition (SVD) of the sensitivity matrix. The SVD of the matrix A_{nj} is defined as

$$A_{nj} = U\Sigma V^T, \qquad \text{Equation 4}$$

where U, V are unitary matrices and Σ is a diagonal matrix whose entries are called the singular values of matrix A_{nj}. To determine the j process-parameters from n measurements uniquely, the effective rank of the sensitivity matrix should be equal to j. The effective rank of the sensitivity matrix is defined as the number of singular values (Σ) of the matrix that are above a predefined threshold. Hence, to determine the j parameters, a test optimization algorithm that increases the values of all the singular values (Σ) of the sensitivity matrix is used. The algorithm generates different singular

values (Σ) of A_{nj} by varying the amplitudes and frequencies of different tones in the test-stimulus. Since Genetic optimization minimizes the cost-function, the inverse of the sum of singular values is calculated. This cost-function is mathematically stated as follows:

$$costfunction\ (cf) = \frac{1}{(\sum_{i=1}^{j} \Sigma_i')}. \quad \text{Equation 5}$$

In the above stated cost-function, a thresholding function is implemented to modify the singular values of the sensitivity matrix. The thresholding function is used to ensure that all singular values have a significant contribution to the overall cost-function. The thresholding function is defined as follows:

$$if \quad \Sigma_i > \beta, then\ \Sigma_i' = 1;$$
$$else \quad \Sigma_i' = {\Sigma_i}/{\beta} \quad \text{Equation 6}$$

where β is a predetermined threshold, Σ_i and Σ_i' are the original and modified singular values, respectively.

Similar to Equation 2, the matrix relating δS_k with δP_j can be defined as

$$\delta S_k = C_{kj} \delta P_j, \quad \text{Equation 7}$$

and the matrix relating δS_k with δM_n can be defined as

$$\delta S_k = B_{kn} \delta M_n. \quad \text{Equation 8}$$

Using the matrices in Equation 2, Equation 7, and Equation 8 following equation is obtained:

$$C_{kj} = B_{kn} A_{nj}. \quad \text{Equation 9}$$

When the rank of the matrix A_{nj} is j (after performing test optimization), from Equation 9, matrix B_{kn} can be written as

$$B_{kn} = C_{kj}({A_{nj}}^T A_{nj})^{-1}{A_{nj}}^T. \quad \text{Equation 10}$$

Now in order to determine the k specifications uniquely from the measurements, the effective rank of the B_{kn} matrix should be equal to k. Once the effective rank condition (explained earlier) for the matrix B_{kn} is satisfied, the resultant stimulus is the optimum stimulus. In general, the sensitivity matrix may be rank-deficient due to the existence of linear dependences between the various manufacturing process-parameters. Such parameters are said to belong to ambiguity groups. A method presented in [54] is used to handle ambiguity groups. The flowchart presented in Figure 6 explains the overall proposed test generation algorithm. Using the optimized stimulus, the DUT is excited and the statistical regression models relating the measured responses to the specifications and the Spice-level parameters are developed. A tool called MARS (Multivariate Adaptive Regression Splines) [81] is used to develop the regression models. The complete methodology of our technique is shown in Figure 9.

3.1.3. Determination of Significant Device/Circuit Parameters

To diagnose the Spice-level model parameters, the parameters that have significant effect on the circuit performance need to be determined. In [85] , the authors have shown that the number of parameters that define the process space of the circuit can be reduced to a small number of parameters. In this work, a simple three-level factorial analysis is performed on the initial parameter space and the significant parameters above a threshold are obtained. In this work, the layout-area dependent process-parameters [85] are not considered.

3.1.4. Simulation Results - Diagnosis and Testing

3.1.4.1. Case Study I - Operational Amplifier

The DUT is a two-stage Op-Amp designed in AMI 0.5um technology node. The gain of the DUT is 45 dB and the f_{3dB} is 10 MHz. The circuit is implemented in Cadence Spectre environment. The circuit is interfaced with Matlab for test generation and diagnosis purposes. During test optimization, the amplitudes and frequencies of the test stimulus is optimized. The sensitivity matrix is computed by varying each of the process parameter, one at a time, about its nominal value ($\pm 1\sigma$) where σ is the percentage standard deviation of each process parameter. The optimized stimulus obtained is shown in Figure 11. The optimized stimulus has nine tones (from 100 KHz to five MHz) of varying amplitude.

During the training phase, the significant process parameters are varied using a random uniform distribution of $\pm 5\sigma$ (σ being the standard deviation of each process parameter) and 1000 such DUT instances were obtained. The defective parts (for e.g., those parts which have a catastrophic fault such as a short or open), were pruned out using specification-based defect filter. The DUT measurements are the output voltage sampled at 40 MHz and the transient response of the supply current sampled at 5 MHz. The responses are used to develop the models x_{1ms} and x_{2mp}. During evaluation phase, 150 instances of Gaussian distribution were considered. The relative error in the prediction of the process parameters as well as the specifications is shown in Figure 12. β is the current mirror ratio of the input stage of the Op-amp, Lln and Wln are the length and width reduction factors respectively, Vth0 is the zero-bias threshold voltages, Tox is the thickness of oxide, and Vsat is the saturation velocity at nominal temperature. The terms 2nd HD and 3rd HD are the second order and third order harmonic distortion components respectively. The n and p terms in parenthesis in the Figure 12 indicate the corresponding parameters for NMOS and PMOS transistors respectively. As can be seen from Figure 12, the proposed methodology accurately predicts the test and diagnosis parameters with a prediction error of less than 9% for all parameters.

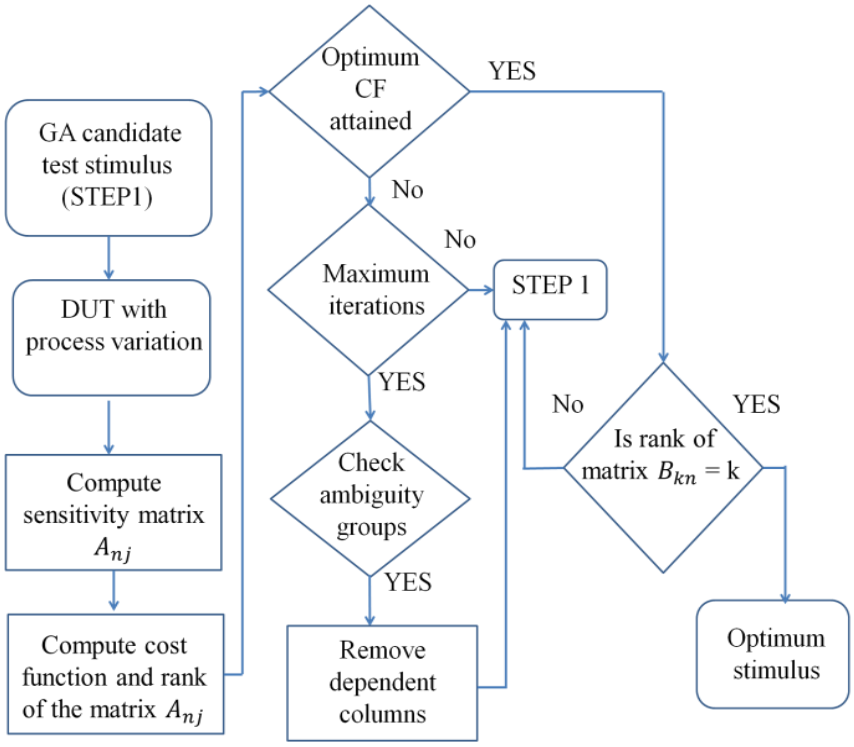


Figure 9: Flowchart of the test generation algorithm.

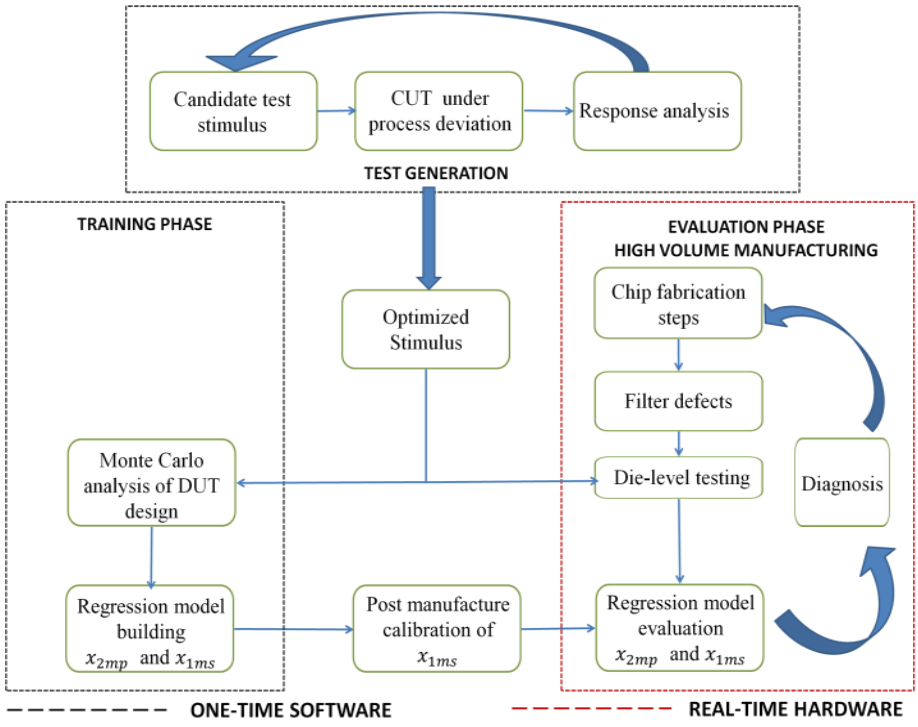


Figure 10: Complete methodology of proposed technique.

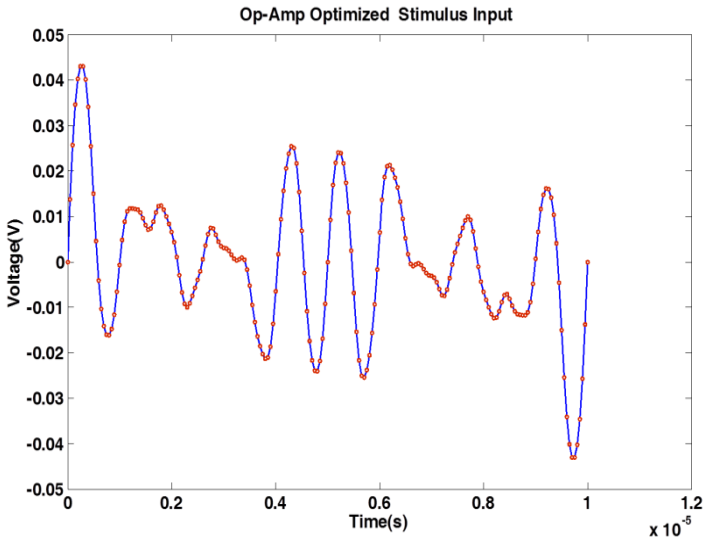


Figure 11: Optimized input stimulus.

Figure 12: Prediction error in specifications and process parameters.

3.1.4.2. Case Study II - Low Noise Amplifier

The case study considered is a 3 GHz radio frequency (RF) low noise amplifier (LNA). The three-stage LNA design has a nominal gain of 16 dB and IIP3 specification of -9.5 dBm. The DUT is implemented in 180 nm CMOS technology. The Spice-model used for this design is BSIM3v3. For simulating the LNA, a high frequency air coplanar

probe (ACP) model is used to account for the input and output probe parasitics. Initially 25 parameters of the BSIM3v3 model are considered and seven significant parameters using the methodology discussed in section 3.1.3 are obtained. The significant parameters are Lln(n), Vth0(n), Tox, Wln(n), Vsat(n), Cgdo(n), L1(input stage inductor).The test-stimulus optimization routine is performed using the technique described in Section 3.1.2. The frequencies and amplitudes of the input tones are optimized. The progression of the cost function with generations is shown in Figure 13.

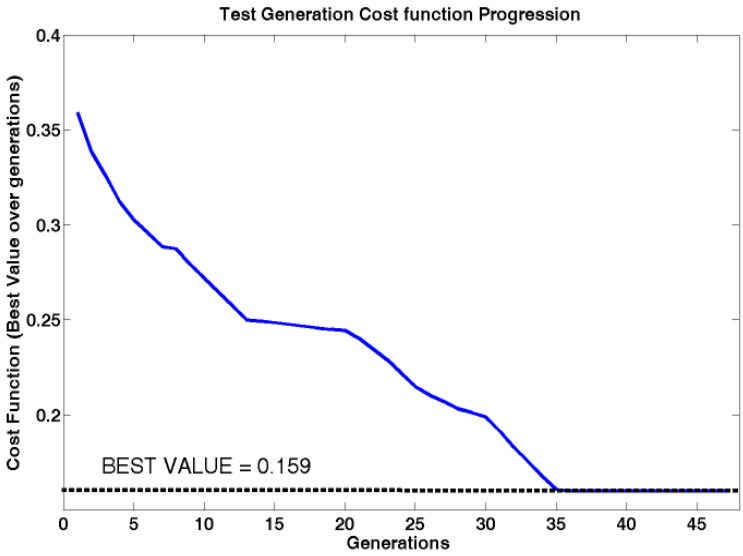


Figure 13: Test generation cost function progression.

The final stimulus consists of 18 tones spaced within a bandwidth of 150 MHz. The LNA multi-tone output response is down-converted to a low frequency response by a mixer (implemented in Matlab) and the transient response along with the transient supply current is used as the alternate diagnostic signature. For training the regression model, 2000 circuit instances are used. These instances are drawn from a randomly distributed uniform variation of all process-parameter perturbations ($\pm 5\sigma$, where σ is the percentage standard deviation). In the evaluation phase, 150 instances of the DUT are used. These instances are generated using a Gaussian distribution. Prior to training, a simple

specification based defect filtering approach is used to eliminate the inclusion of defective DUTs. The prediction plots for some of the device and circuit process-parameters are shown in Figure 14. The relative and root mean square (RMS) error (expressed in percentage) in the prediction of the specifications and process-parameters are shown in Figure 15.

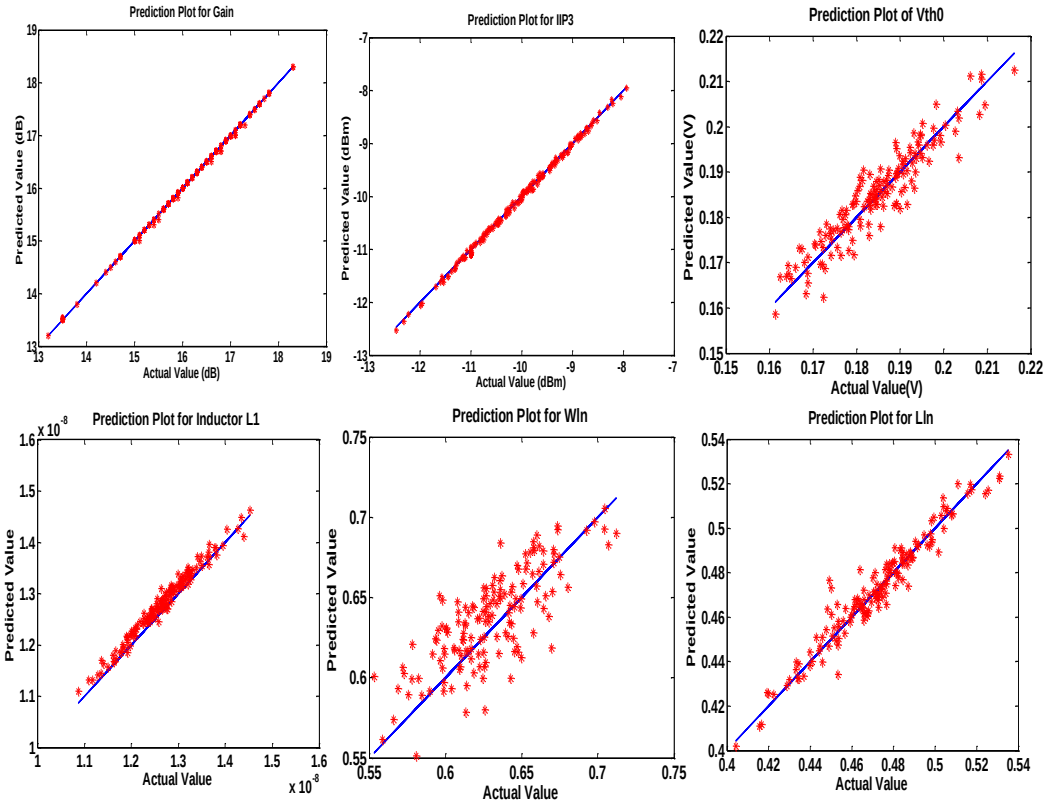


Figure 14: Prediction plots for specifications and process parameters.

The proposed technique is an efficient method to track process shifts occurring during volume manufacturing. For a suitably estimated regression models, the technique provides the ability to track any shift in the nominal value of the process-parameter as well as its distribution (over different wafers during production). To validate the above concept, a known (± 1 to 2σ) process shifts are injected into multiple process-parameters of the LNA. The original, injected, and tracked histograms are shown in Figure 16. Notice that the variations are induced both in the mean and in standard deviation of the

process-parameters. The results obtained from the histograms can be used for detecting multi-dimensional excursions in the fabrication facility during early stages of product manufacturing.

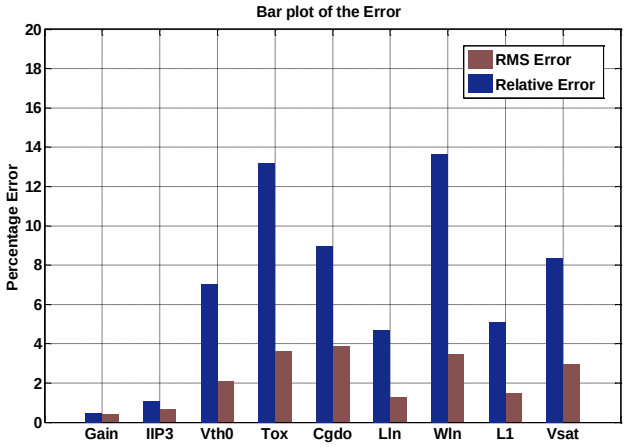


Figure 15: Error in prediction of the parameters.

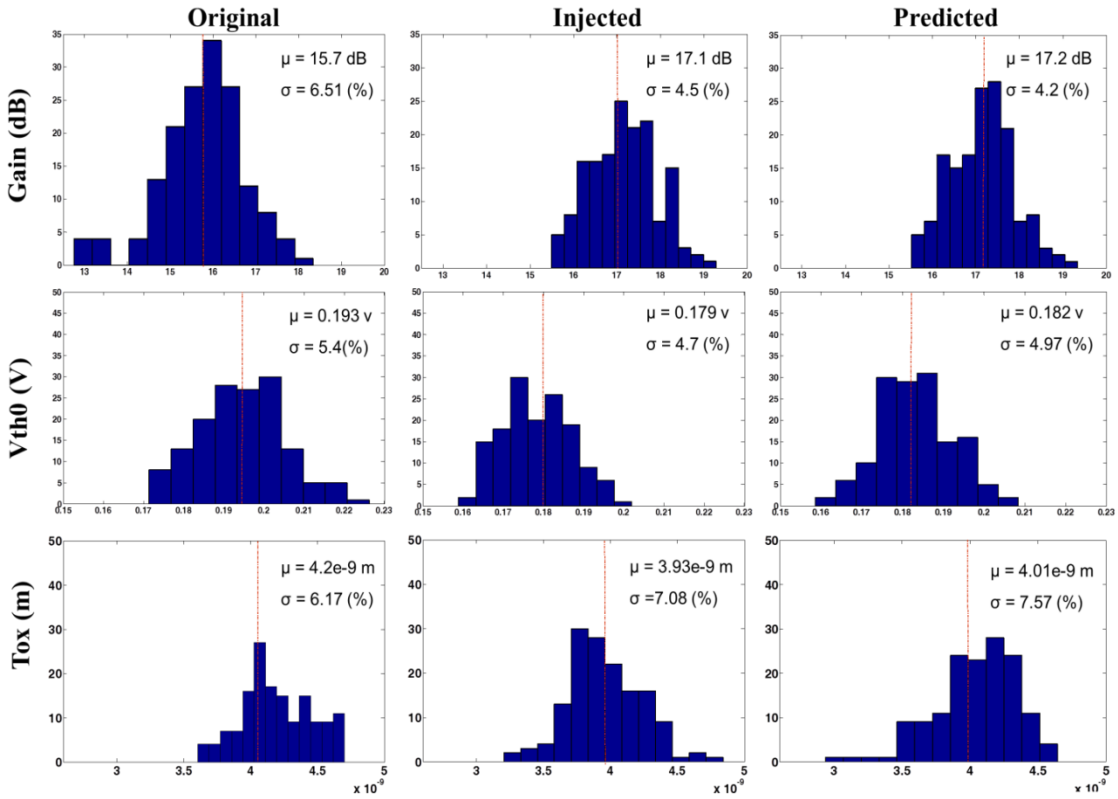


Figure 16: Specification and Spice-level parameter histograms.

3.1.5. Simulation Results - Cause-Effect Monitoring

The ability to diagnose process-parameters during wafer-level testing for a given circuit enables us to analyze the effects of the process variation on specifications of the circuit. This methodology is termed as Cause-Effect analysis. Using the process and specification domain information, a regression model $f_{ps}(p)$ relating both these distributions is developed. The presented analysis technique uses this regression model to determine the relative contribution of each of the process-parameter distribution to the DUT specification distribution. The analysis involves regrouping the function $f_{ps}(p)$ into multiple smaller functions. Each of these smaller functions consists of either individual process-parameter or multiple process-parameters. The variance contribution of each of the smaller function to the overall specification variance is determined. The regrouped function is shown as

$$f_{ps}(p) = \sum f_i(p_i) + \sum f_{i,ii}(p_i p_{ii}) + \sum f_{i,ii,iii}(p_i p_{ii} p_{iii}) .., \quad \text{Equation 11}$$

where p_i is one of the process-parameters in the vector $P_j = [p_1, p_2, p_3 \ldots ., p_j]$ and $f_{ps}(p)$ relates the process-parameters to one of the k specifications represented by $S_k = [s_1, s_2, \ldots s_k]$. The first term in Equation 11 accounts for all the functions corresponding to individual parameters p_i that determine the specification s_k. The second term in the equation indicates all the functions that consist of a combination of two different parameters and so on. In Equation 11, $i, ii, and\ iii$ vary from one to j. Considering that p'_{ix} is the diagnosed value of x^{th} instance of parameter p_i, the contribution of each parameter to the specification is calculated using metrics stated as follows:

$$m_{\acute{p}_i} = \frac{1}{N}\sum_{x=1}^{N} f_i\,(p'_{ix}), \quad \text{Equation 12}$$

$$Var(\acute{p}_i) = \frac{1}{N-1}\sum_{x=1}^{N}(f_i(p'_{ix}) - m_{\acute{p}_i})^2\ , \quad \text{Equation 13}$$

where N is the number of samples of the distribution, $m_{\acute{p}_i}$ and $Var(\acute{p}_i)$ is the mean and variance of each function $f_i(p_i)$, respectively. The mean and variance of each interaction function is computed in a similar manner. Equation 12 and Equation 13 represent the ANOVA decomposition technique for a regression model [81]. Such an analysis is performed for both the gain and IIP3 specifications of the original and process-varied distribution (used for diagnosis in Figure 16) and the Cause-Effect analysis results are presented in Figure 17. As evident from Figure 17, the gain variation (between the two distributions) is affected by the Tox, Lln, and Vth0 variations while the IIP3 variation is affected significantly by the length reduction factor (Lln) and Vth0. In addition, the variations in second order interaction-effects also contribute to the specification variation, but to a lesser degree of significance. The error portion in the graph is the percentage variation that cannot be explained by any process-parameter (error from the model). As evident from the graphs, most of the variations in the specifications can be explained by our diagnosis approach. The accuracy of the analysis is affected by the error in the prediction of the process-parameters. However, it does still provide significant information in regards to the process-variation trends that can be fed back to the fabrication facility for process monitoring.

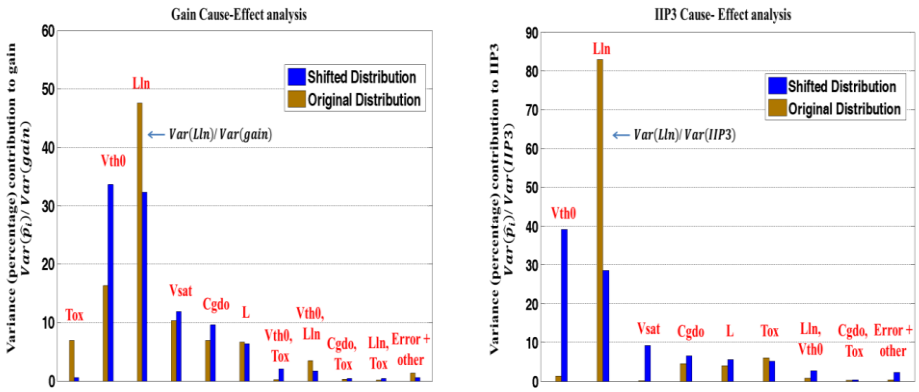


Figure 17: Cause-Effect analysis of gain and IIP3 specifications of LNA.

3.1.6. Hardware Validation of Proposed Methodology

Traditionally, the effect of process variations is studied using measurements of e-test parameters at different sites (also called as shots or reticles) of the wafer. In the absence of a catastrophic defect, the variation in the output response of a DUT depends on the variation of its process parameters. To validate the proposed work, instead of developing the regression functions in simulation environment, the e-test parameters of the device are used as the Spice-level parameter replacement. Hence, from the test measurement data of the DUT, one should be able to obtain information about the e-test parameters that influence the device. The above information regarding e-test parameters is obtained using a mathematical regression model that relates the output measurements of the device to each of its e-test parameters. The number of sensitive e-test parameters that can be determined depends on the number of independent measurements spanning the test measurement space. In production environment, a combination of probe-level tests (called Multiprobe) and final tests constitute the test measurement space. The regression model relating the e-test parameter and the test measurements can be developed using the initial characterization or design of experiment (DOE) wafers. In these wafers, many process parameters are varied beyond their normal expected variations in the production environment. Ideally, the e-test parameters that significantly influence the circuit specifications can be determined using experiments such as central composite design (CCD) in the simulation environment [86] (a variant of it is used in this work as explained earlier in Section 3.1.3). If the e-test parameters cannot be varied during the device characterization phase, the regression mappings need to be developed using the initial set of production wafers. Ideally, there are tens of e-test parameter measurements made on an individual wafer. Hence, to capture the spatial variation trends in the presence of noise as well as missing measurements would require many measurements over lots. This would lead to greater amount of time to implement the feedback

framework. Alternatively, there exists significant process variation information on each wafer in the shots where the measurements have not been made. This spatial process variation on each wafer can be obtained by interpolating the existing e-test parameter measurements to other locations on the wafer. Once a higher resolution wafer map of the process e-test parameter is obtained using a few wafers of the first lot, regression models relating the test data of the devices and the interpolated process e-test parameters can be developed (see Figure 18). Please note that spatial interpolation of an e-test parameter does not indicate its significance or its insignificance to the device. Any parameter that has spatial correlation can be interpolated. Hence, it is important to develop the regression model relating these parameters and the test data for each circuit. If the test measurement space is not large enough, then prediction of all the sensitive e-test parameters will not be feasible. However, the presented technique would still provide information about the variation in some of the critical parameters. Once the regression model is developed, using the test data from different locations of a wafer, the e-test parameters at the corresponding locations can be determined. As the test measurements made on every die are used to determine the e-test parameters, any drift occurring in the parameters can be quickly identified. At this point, feedback can be provided to the fab to tune the corresponding process parameter before any future yield loss occurs. To determine the sensitive e-test parameters, mathematical regression models are developed by mapping all the test measurements to each e-test parameter. In this step, a methodology called Multivariate Adaptive Regression Splines (MARS) is used for regression function development. The overview of hardware validation approach is shown in Figure 18.

3.1.6.1. E-test Parameters and Process control

As mentioned in introduction of the chapter, current methodology of process control monitoring involves measuring process e-test parameters at either 21 or 9 locations on

each individual wafer in a fab. These e-test parameters are measured using test structures placed in scribe lines (in between the dies) in specific sub-regions of a wafer called photo shots (also called as shots or reticles of the wafer). A hypothetical shot-map with the different shots/sites across the wafer where measurements are made is shown in Figure 19. The e-test parameters are measured using specialized test structures at a certain limited number of test shots on each wafer. Each shot is square/rectangular region composed of dies, and test structures that are placed in the scribe lines in between the dies. As each wafer provides only 9 or 21 measurements, a significant amount of data needs to be collected across large intervals of time (wafers and lots) to account for any deterministic process e-test parameter variation. Faster feedback can be obtained using increased number of test structures but at the cost of greater test time and silicon area. Each test structure measures a particular e-test parameter and the wafer is classified as a good wafer as long as the individual e-test parameter lies within certain limits. Often there exist hundreds of parameters, and the fab does not have any knowledge of the relative importance of the parameters and the limits with which the parameters need to be controlled for a given circuit.

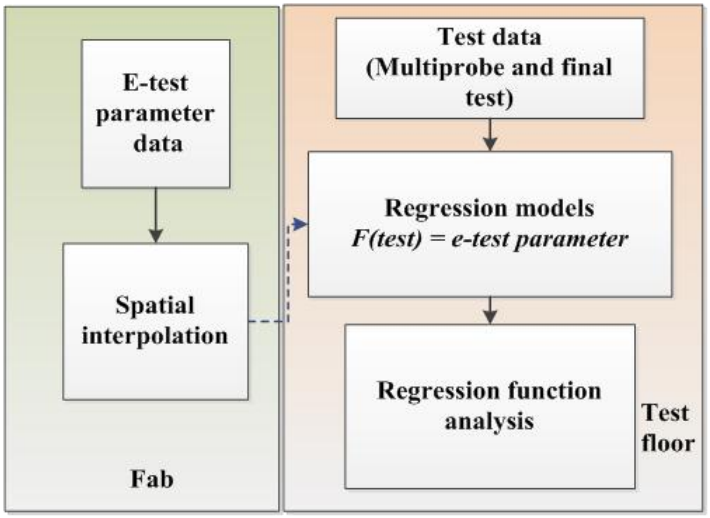


Figure 18: Hardware validation approach.

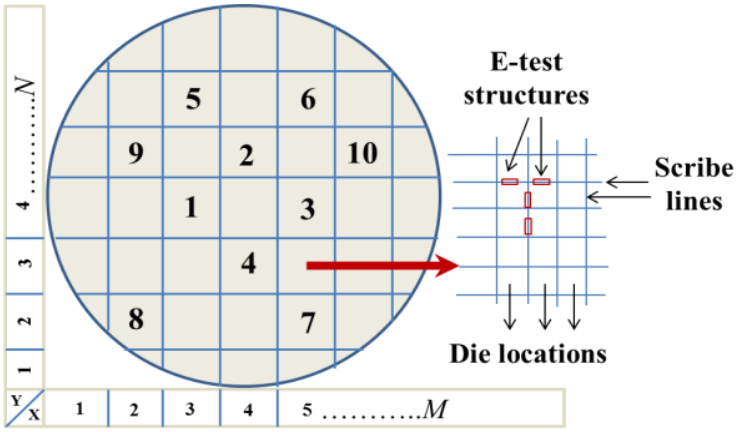


Figure 19: Hypothetical shot-map of the wafer.

3.1.6.2. Spatial Interpolation

In this work, the Virtual Probe (VP) technique is used for performing spatial interpolation [87]. The technique applies the principle of compressed sensing to semiconductor e-test parameters for achieving wafer-level spatial interpolation. If an e-test parameter has high amount of spatial correlation along the standard coordinate dimensions, then the e-test parameter in the spatial frequency domain (i.e., fourier or cosine transform of the spatial variation) is sparse with large number of coefficients being small or close to zero. The e-test parameter that needs to be spatially interpolated can be expressed as a two dimensional function in X-Y domain namely $f(x, y)$, where $x \in [1,2 \ldots . M]$, y $\in [1,2 \ldots . N]$ (see Figure 19) are the coordinates that define the variation of the e-test parameter along their respective directions. If the discrete cosine transform (DCT) of the two dimensional function can be stated as $F(p, q), where\ 1 \leq p \leq M, 1 \leq q \leq N$, then the corresponding inverse discrete cosine transform (IDCT) function can be stated as shown in .

$$f(x, y) = \sum_{p=1}^{M} \sum_{q=1}^{N} \alpha_p \beta_q F(p, q) \cos \frac{\pi(2x-1)p}{2M} \cos \frac{\pi(2y-1)q}{2N}, \quad \text{Equation 14}$$

$1 \leq x \leq M, 1 \leq y \leq N,$

In Equation 1, α_p, β_q are the scaling coefficients. Now considering that the e-test parameter is being measured in L locations (L being 21 or 9), i.e. the values of $f(x, y)$ at these locations are known. The goal is to use these measurements to find $f(x, y)$ at other locations using Equation 14. Further, it has been shown in that the DCT coefficients $F(p, q)$ are sparse due to the existence of high amount of spatial correlation across the wafer. Hence, by considering the joint probability distribution function (PDF) of all the sparse discrete cosine transform (DCT) coefficients and using maximum a-posteriori probability (MAP) estimation, it has been shown that the DCT coefficients can be obtained by formulating the optimization problem as a L1-norm regularization problem [86]. The above L1-norm regularization can be solved using the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm [88]. Once the estimate for all the DCT coefficients is obtained, the interpolated values of the parameter can be calculated using inverse discrete cosine transform (IDCT). The accuracy of interpolation obtained using the VP technique was found to depend on the extent of spatial correlation of each e-test parameter. The assumption of high amount of spatial correlation is true for most e-test parameters. Please note that spatial interpolation of an e-test parameter does not indicate its significance or its insignificance to the device. Any parameter that has spatial correlation can be interpolated.

3.1.6.3. Device Results

The device considered is a SOC implemented in a 180 nm process that is currently in production. It has both Analog and digital components that include DC-DC converters, amplifiers, ADCs and DACs among other modules. The simulation environment used for interpolation and regression analysis is Matlab. This section provides the experimental validation of the spatial interpolation technique implemented. A set of 18 wafers is

implemented with test structures at all the 70 shots of a wafer rather than the conventional 21 or 9 shots. Each e-test parameter is measured once in a shot. The choice of 21 (called category A parameters) or nine (called category B parameters) locations depends on the parameter significance (input from the designer & process engineer based on experience).

To use the VP interpolation technique, it is necessary that the e-test parameters exhibit a significant amount of spatial correlation across the wafer. The spatial correlation condition is verified by performing a DCT transformation using all the measured values of a parameter in a wafer. As an example, a DCT transformation of the N-Well sheet resistance parameter is performed using all its 70 measured values. The histogram of the DCT coefficients is shown in Figure 20. A large number of DCT coefficients of the N-Well sheet resistance parameter are close to zero indicating that there exists high spatial correlation across the wafer. Once the existence of high spatial correlation is verified, the conventional 21 and nine site measurements are chosen out of the 70 shot measurements and are used to interpolate to the remaining sites. The error metrics obtained for various e-test parameters by interpolating measurements made at 21 locations is provided in Table 1. Similar error metrics obtained by interpolating measurements made at nine locations are listed in Table 2. The relative error is calculated using the formula shown in Equation 15

$$Relative\ error = \left\{\left(\tilde{f}(x,y) - f(x,y)\right)/f(x,y)\right\} * 100, \qquad \text{Equation 15}$$

where $\tilde{f}(x,y)$ is the interpolated value and $f(x,y)$ is the measured value. The mean of the absolute value of relative error and the standard deviation of the relative error at various locations and wafers are tabulated.

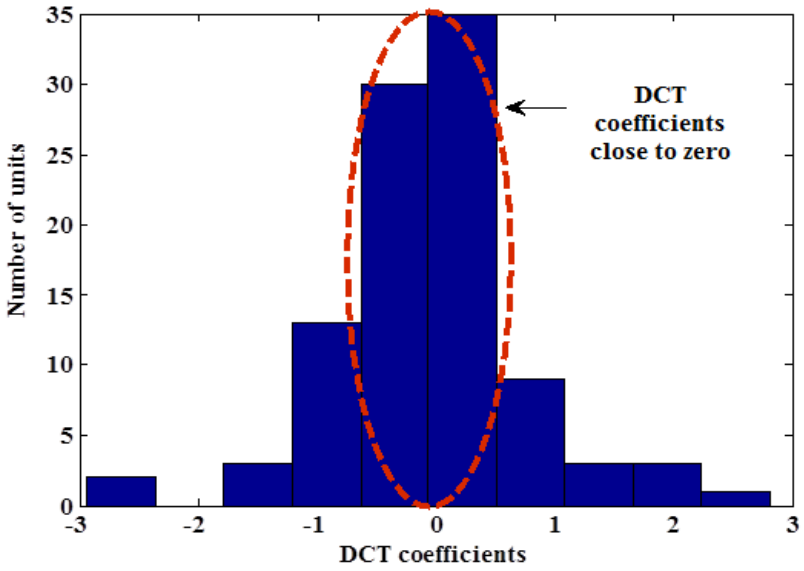


Figure 20: Histogram plot of the DCT coefficients.

Out of 126 e-test parameters that are measured for interpolation, 107 parameters are interpolated with mean relative error of less than five percent. The original and interpolated wafer maps of the N-well sheet resistance parameter are shown in Figure 21. E-test parameters having significant outlier measurements with non-normal distribution or parameter measurements with insufficient measurement resolution are the common causes for high interpolation error. While the basic method of VP has been implemented here, more advanced versions of this methodology that use multiple wafers of the same lot have been proposed [89]. The technique discussed in [89] could aid in providing interpolation results with greater resolution.

For the purpose of regression analysis, data from 108 wafers across four lots is used (close to 170,000 dies). On each of the 108 wafers, 97 e-test parameters are measured at 21 sites (Category A), and 168 e-test parameters are measured at nine sites (Category B). The measurements of each e-test parameter are screened to remove outliers. This outlier detection is an important step as these outliers can have a significant effect in the interpolation and the regression development steps. A non-parametric inter-quartile range

(IQR) technique is used to remove the outlier observations. The above outlier removal step helps in removing spurious data that occurs because of measurement error and various other related error sources.

To avoid defects, only dies that passed all the tests are considered in the experiment. Hence, to perform regression analysis, the probe-level test measurements of all the pass dies in a single shot are collected. The median of all the measurements from the collected die measurements is calculated. Out of the 347 probe-level test measurements, statistical distribution across devices is not observed for 50% of the measurements.

Table 1: Error metrics obtained by interpolating 21 measurements.

Parameter	**Mean absolute relative error (%)**	**Standard deviation in relative error (%)**
N-well sheet resistance	0.513	0.637
Unit capacitance 1	1.772	2.345
PMOS VT	0.505	0.512
PMOS 2 off current	5.126	3.871
PMOS drive current	0.505	0.532

Table 2: Error metrics obtained by interpolating nine measurements.

Parameter	Mean absolute relative error (%)	Standard deviation in relative error (%)
N-well sheet resistance	0.734	0.729
Unit capacitance 1	0.729	2.232
PMOS VT	0.709	0.600
PMOS 2 off current	12.94	5.198
PMOS drive current	1.064	0.886

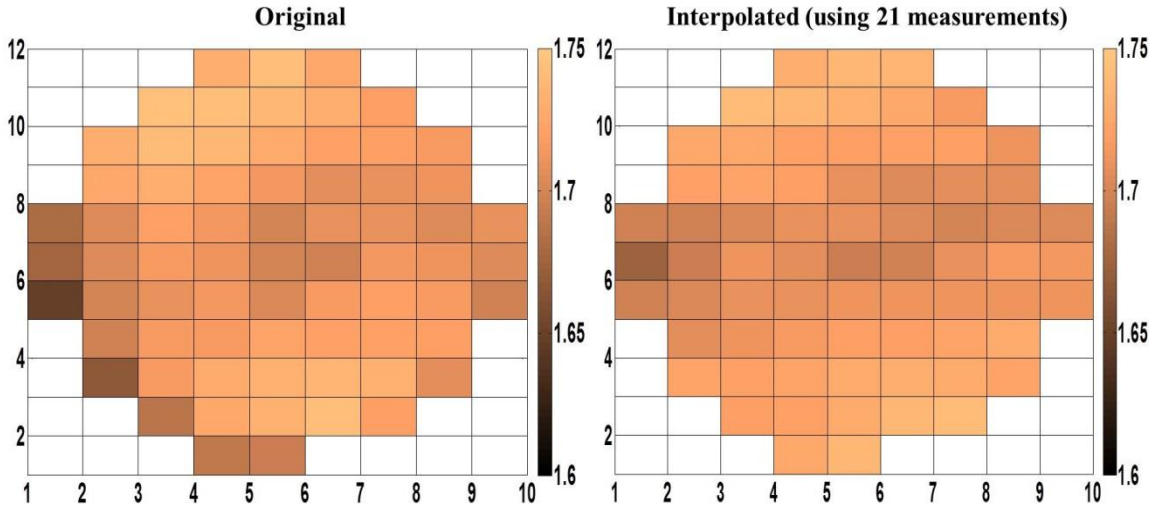


Figure 21: Original and interpolated e-test wafer maps for N-well sheet resistance.

For each of the e-test parameters, regression functions are developed using all the remaining test measurements that showed statistical distribution. In the training phase of

the regression function, 1400 observations are selected randomly from the wafers in Lot one and Lot two. For training the regression functions for the e-test parameters, 1400 observations were selected randomly from all the wafers in the first 2 lots as well as from 22 wafers of lot three. A combination of the actual and interpolated e-test parameters were used for training. For evaluation, 201 observations collected from the three wafers of lot three not included in the training and from all the wafers in Lot four were used. The scaled prediction plots for the various process e-test parameters are shown in Figure 22. As can be seen from the graphs, using this technique, accurate prediction of the process e-test parameters to which the circuit is sensitive is performed.

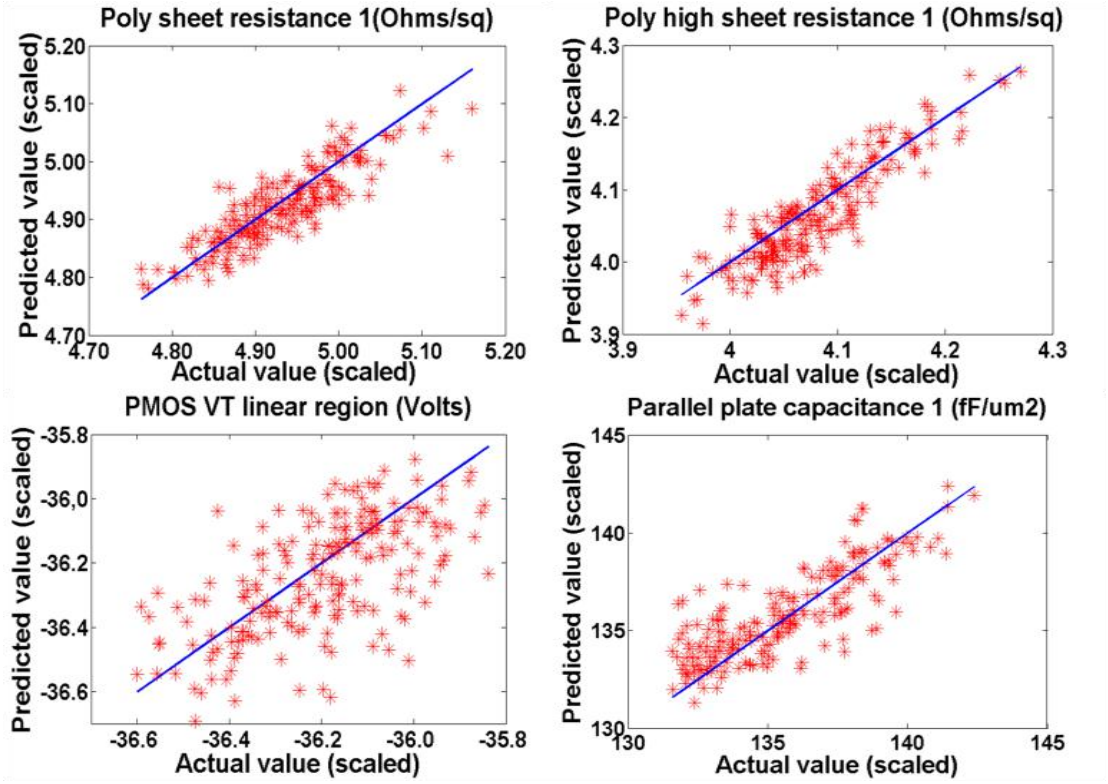


Figure 22: Scaled prediction plots for process e-test parameters.

The error values for some of the process e-test parameters belong to different categories that could be predicted are given in Table 3. Along with the mean of relative

error, the normalized root mean square $\left(NRMS = \frac{\sqrt{\frac{\sum(actual-predicted)^2}{n}}}{range(actual)} \right)$ is calculated. The goodness-of-fit metric (linear correlation) between the actual and the predicted values is also provided. 14 parameters in category A and 35 parameters in category B were predicted with their NRMS values ranging from nine to 20%. The minimum goodness-of-fit metric for these parameters was 0.683 (PMOS VT) which is the worst predicted parameter. The PMOS drive current prediction across different wafers/lots is shown in Figure 23. The mean relative error for wafer 23 is 0.54 % and the mean relative error for wafer 4 is 2.3%. The wafer predictions at all the 70 shots/sites (3x resolution increase) as well as for the all the pass devices in the wafer are shown in Figure 24.

Even though the regression function is developed at the shot-level, the measurements of every individual die can potentially be used to predict their respective e-test parameters with acceptable accuracy. As can be seen from Figure 24, there exists some white spots in the wafer map corresponding to the failed devices. Hence, a greater resolution of wafer map can be achieved using this technique. This technique also enables a tighter multi-dimensional control of process parameters (based on the values obtained from pass devices) as opposed to using uni-dimensional control charts. The accuracy of this prediction can be made better if an interpolation algorithm that is capable of interpolating to the resolution of a die is used and the regression model is developed using the test measurements of every die rather than the median of the dies in the shot. Advanced outlier detection mechanisms applied to the measurement domain can also improve the prediction accuracy.

If process monitoring is performed using test data of every die (close to 1300 values in a wafer as opposed to the traditional 21/9 e-test structure measurements), any process shifts can be quickly identified and information can be fed back to the Fab, before any yield loss occurs. The number of parameters that can be predicted can be increased further by using the post-packaging final test data for prediction.

An analysis of the regression functions was performed to determine the relative contribution of each test measurement. The regression function for each individual process e-test parameter is given by the equation below

$$p = f_{mp}(m) = \sum f_i(m_i) + \sum f_{i,ii}(m_i m_{ii}) + \cdots, \qquad \text{Equation 16}$$

where the function is regrouped to provide the relative contribution of each measurement towards the process e-test parameter prediction. In Equation 16, $f_i(m_i)$ is the contribution of the measurement to the process e-test parameter p. This analysis can be considered providing a coefficient for each test measurement. Such an analysis is performed for the process e-test parameters Poly sheet resistance 1 and Parallel plate capacitance 1 and shown in Figure 25 and Figure 26 respectively. As can be seen from Figure 25, T18, T19, T20 contribute significantly to the poly sheet resistance 1 parameter, these measurements are the I/O resistance and current measurements. This observation concurs with the design knowledge of the circuit. Further, as a methodology to cross validate the developed regression models, we perform a bivariate analysis between each predicted e-test parameter and test measurement. Spearman's rank correlation is used as a metric to obtain the correlation of a process e-test parameter to a test measurement. The rank correlation metric indicates the extent to which the relationship between two variables can be expressed by a monotonic function.

Table 3: Prediction of e-test parameters.

Parameter	***NRMSE (%)***	***Relative error (%)***	***Goodness-of-fit***
PMOS drive current	12.85	1.491	0.752
Parallel plate capacitance 1	13.92	0.841	0.818
Poly sheet resistance 1	10.64	0.658	0.895
Poly sheet resistance 2	9.860	1.325	0.817
VIA Resistance 1	12.17	1.689	0.894
NMOS drive current	14.68	0.980	0.801
Metal 6 resistance	13.55	6.282	0.854
Poly high sheet resistance 1	9.860	1.325	0.817
PMOS VT	19.72	0.344	0.683
Depletion MOS DVT	11.80	1.054	0.883

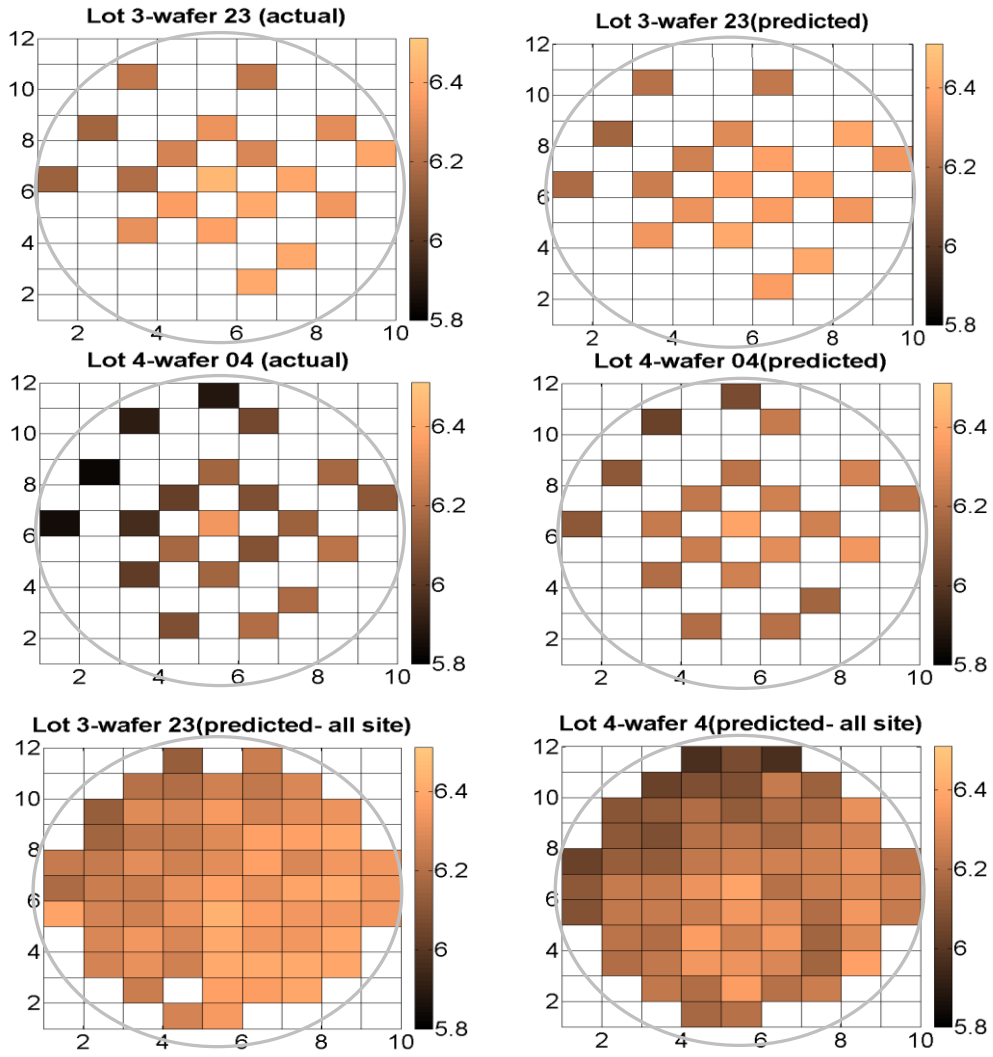


Figure 23: Scaled prediction plots of PMOS drive current for different wafers.

The rank correlation metric is mathematically stated as

$$\rho = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2 \sum_{i=1}^{N}(y_i-\bar{y})^2}}, \qquad \text{Equation 17}$$

where x_i, y_i are the ranked variables and $\bar{x}$, $\bar{y}$ are their respective means; N is the number of measurements. The rank correlation between the poly sheet resistance 1 parameter and the top test measurements is listed in Table 4. Thus, the test measurements that contribute to the prediction of the parameter have a high amount of correlation with the parameter, thereby validating the model as well.

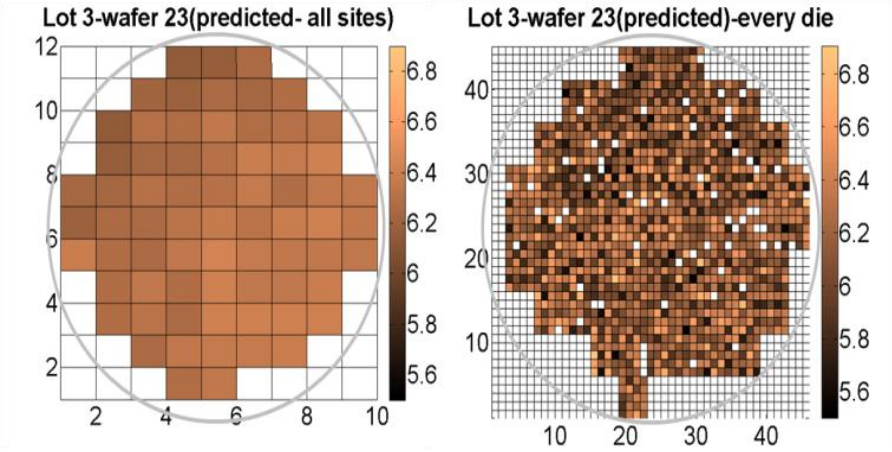


Figure 24: Prediction plots of PMOS drive current at site-level and die-level.

Table 4: Rank correlation between test measurements and poly sheet resistance.

Test Number	Spearman's rank correlation
T20	0.93
T19	-0.70
T18	0.96
T17	0.69
T16	0.95
T15	0.89

All the e-test parameters that were predicted had Spearman's rank correlation factor of at least 0.45 to one or more of the test measurement. For all the cases where the

correlations exceeded the threshold (0.45), p-value is estimated to ensure that valid correlation exists between the e-test parameters and the test measurements. E-test parameters that were not predicted did not show any substantial rank correlation to any test measurement. From Figure 26, it can be inferred that no single test measurements contribute significantly to the parallel plate capacitance 1. This makes sense from a design perspective as a number of test measurements get affected by capacitance and hence contribute to its prediction. The variation in the coefficients of the regression function over time can be guard banded to determine the extent of process variation as well as periodic recalibration of regression models.

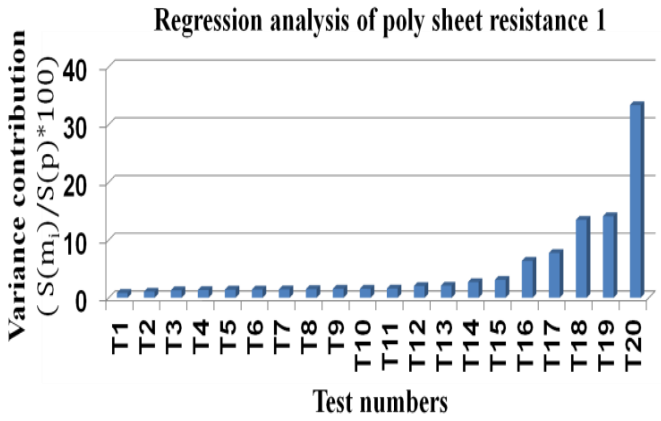


Figure 25: Analysis of regression model for poly sheet resistance 1.

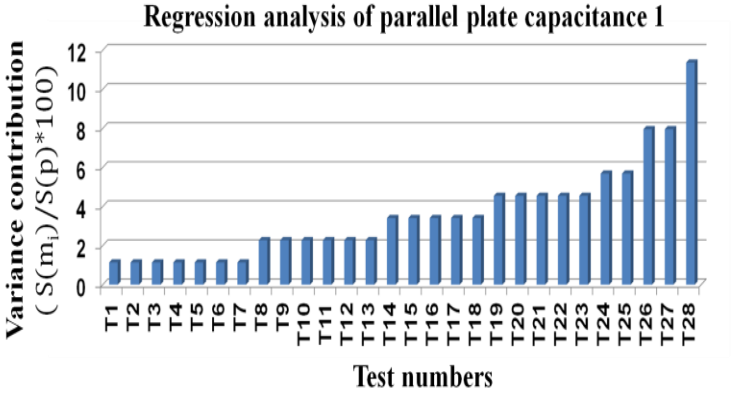


Figure 26: Analysis of regression model for poly sheet resistance 1.

3.1.6.4. Significant E-test Parameters (Dimensionality Reduction)

To implement an intelligent test based feedback control, the e-test parameters to be monitored in the process domain need to be selected intelligently. Currently, the parameters measured using the test structures account for many different variants of a particular parameter that are highly correlated with each other. Current technologies have as many as 500 process e-test parameters that are measured on very wafer. If a feedback control were to be implemented using all the process e-test parameters, such a technique would be highly complicated.

However, we can exploit the fact that many manufacturing steps are highly correlated affecting multiple groups of parameters (such as MOS parameters, resistance parameters etc.) in a similar manner. Hence there exists an opportunity to perform dimensional reduction and monitor at all times lesser number of parameters rather than the entire set of e-test parameters that are currently measured. To achieve the objective of dimensional reduction, Principal Component Analysis (PCA) is adapted in this work. Principal Component Analysis is a multivariate methodology that transforms observations in a given data set with high amount of correlation into a data set consisting of linearly uncorrelated variables that are termed as principal components.

If we considered a vector $X_1 = [X_{11}\, X_{12} \ldots X_{1p}]$ consisting of one set of measurements of p process e-test parameters and X_{np} is a matrix of n such measurements. Z_{np} is the data set in the new coordinate system and A_{pp} is the orthonormal transformation matrix called projection matrix. The transformation matrix A_{pp} consists of eigenvectors of the covariance matrix $\Sigma_{pp} = {X_{np}}^T X_{np}$. The eigenvalues of the covariance matrix is the variance explained by each of the principal component. The eigenvector with the highest eigenvalue forms the first principal component of the new data set Z_{np} and so on. The first principal component accounts for the largest

possible variance in the data set and the second principal component accounts for the next largest possible variance that is not accounted by first component.

In this work, the goal is to select a subset of the total process e-test parameters p that can be used to explain majority of the variation in the original larger data set. Hence, the objective is to reduce the dimensions of the original data set. There exist a number of techniques for performing dimensional or variable reduction using PCA. One such technique that was implemented is explained below:

Start

- Implement PCA on the original dataset X_{np} consisting of p process e-test parameters and n observations or measurements.
- Find all eigenvectors whose eigenvalues are lesser than a predetermined threshold
- Let the number of selected eigenvectors be k ($1 < k < p$)
- For i = 1 to k

{

  - Find the parameter j whose coefficient A_{ji} of the eigenvector i that has the maximum value
  - Remove the parameter from the original data
  - Reduce p = p -1

}

- Obtain the final list of e-test parameters that need to be monitored.

Stop

The idea behind the e-test parameter reduction principal is that the parameter that contributes maximum to the eigenvector that explains less amount of variation in the data can be eliminated without losing any variance information. If there is high amount of

correlation between a set of process e-test parameters, then the technique picks one parameter from the correlated set. It is important to note that this technique provides a reduced parameter set that explain majority of the overall process variation. However, not all these selected process e-test parameters variations cause variations in the circuit. Thus, the parameters that can be predicted through regression methodology should be essentially a subset of the above set determined to the PCA methodology. This way our predictions can be cross-validated.

The measured process e-test parameter data collected from the four lots (108 wafers) consisting of 2057 observations (for 97 parameters belonging to category A) and 645 observations (for 168 parameters belonging to category B) is used for the study. These observations are obtained after applying the outlier routines. This outlier screening procedure is important, as the PCA transformation is very sensitive to gross outliers. As the units of the different e-test parameters are different, the data matrix is normalized such that observations of each e-test parameter have a zero mean and unit variance. Further, the data set is rearranged into groups depending on the nature of measurements (e.g., MOS parameters, capacitance measurements, BJT measurements). There exist multiple methodologies for selection of threshold parameter. One technique involves the use of scree plot. The scree plot is the plot of the eigenvalues in descending order. As the eigenvalues indicate the variance accounted for by the principal components, the eigenvectors corresponding to low eigenvalues are selected for parameter reduction. In this work, the threshold was selected to be 0.7. Various studies performed in prior literature have shown that in principal components analysis using correlation matrices, variances values greater than one should be considered. The results of parameter reduction for different groups of process e-test parameters are shown in Table 5.

Table 5: Dimensionality reduction results for e-test parameters.

Process e-test parameter group	**Selected subset of parameters**
CMOS parameters (25 variables) (dimensional reduction weakest)	**NMOS 3 drive current**
	PMOS 3 VT (linear region)
	NMOS 2 off drive current
	NMOS 2 drive current
	PMOS 2 drive current
	NMOS drive current
	NMOS VT (linear region)
	PMOS off drive current
Resistor parameters (12 parameters)	**Poly sheet resistance 2**
	Active resistance
	Poly sheet resistance 3
	Sheet resistance Nwell

Via Chain and Contact parameters (24 variables)	Poly contact chain
	Contact resistance
	Kelvin contact Via resistance 2
	Via resistance 5
	Via resistance 1
	Kelvin contact Via resistance 1

Hence, the technique provides an efficient methodology for picking process e-test parameters for monitoring and implementing a feedback based on it. The parameters provided in Table 5 in bold are predicted through regression models. Other parameters that are predicted have a strong correlation to one of the above-predicted parameters. It is important to note that all parameters that have variation need not necessarily impact the device as these might not be used in the design or do not affect the device due to its design or layout methodology.

3.1.7. Key Contributions and Applications

The main contribution of this work is that it attempts to develop a means of relating the optimized output measurements of every device to the process/circuit parameters that affect it. Such a methodology has immense potential in performing fine-grained diagnosis at device level, which can be utilized to provide faster feedback to the fab. The key benefits or accomplishments of the proposed approach are as follows:

- Concurrent specification and Spice-level parameter prediction capability with much faster test time and lower test cost compared to standard testing.
- Cause-Effect analysis and rapid feedback providing insight into process variations occurring in volume manufacturing. The test and diagnosis implementation is simple, and can be run independently in any programming environment. It is obtained almost "free of cost".

The following benefits can help the various steps of manufacturing cycle in following ways:

- The learning of the relationship between test measurements and process e-test parameters can be helpful for performing diagnosis on customer return devices, i.e., Cause-Effect analysis in post-Si diagnosis.
- The methodology enables in defining new pass device-based variation limits for important process e-test parameters when the devices are transferred from the existing fab or other fabs. These limits are different from the traditional uni-variate control of process parameters as these limits would be based on the circuit design, and would enable identify those parameters that need to be controlled with tighter limits.
- Finally, the results of the presented analysis would help the design engineers and process-modeling engineers by providing a means to compare process models (Spice models) with variations occurring in production environment

In recent years, there have been an number of different spatial interpolation techniques that have been developed [87][89]. As mentioned earlier, spatial interpolation techniques eliminate the random component of process variation. This random component is reflected only in the device test measurement. Hence, by predicting the e-test parameter using measurement and comparing it with the spatial interpolated value may give an indication of extent of randomness in that parameter. This assumes that the noise in the

measurement and the residual error in the regression model are minimal (first order of approximation). However, this idea needs to be investigated further.

3.2. Online Testing or Performance Monitoring of Analog/RF Front Ends

Aggressive scaling of CMOS silicon real estate for higher circuit densities has led to greater variability in the silicon manufacturing process. The consequent reliability of process-skewed devices is a key concern for systems operating in the field where device monitoring and maintenance is difficult to perform. This work presents a new approach to real-time performance monitoring of analog portion of mixed-signal circuits and RF front ends using signal processing algorithms running on the baseband DSP that perform rapid RF front-end parameter estimation. High power RF modules suffer from temperature-induced memory effects, which lead to performance deviations. State-of-the art RF front-ends employed in the base stations are designed to operate at high power levels exceeding 50 Watts resulting in high rates of heat dissipation and thermal cycling. High power modules suffer from temperature induced memory effects that lead to performance deviations [90].

The majority of research in the on-line testing realm has targeted error detection in VLSI circuits. Prior research has mainly focused on developing Built-In Self-Test (BIST) techniques for detecting defects and performance degradation in analog/RF systems[91][92][93]. However, they are not suitable for online monitoring of wireless RF systems, particularly wireless base stations that must be operational round the clock without significant “down-time”. Such an on-line monitoring system is essential in the case of a system such as base station where it is imperative to have no down time as well as in continuously adaptive circuits, which would require constant monitoring of its own current state while adapting to the external environmental conditions.

In [94], the authors propose the use of built-in current sensors for on-line testing. In [95], an on-line testing methodology for testing identical circuits by output comparison is presented. In [96], the authors propose techniques for detecting gross-delay faults. The majority of research in the on-line testing realm has targeted error detection in VLSI circuits. In the recent past, there has been significant interest in developing online testing techniques for mixed signal circuits such as data converters and operational amplifiers [97][98][99][100][101]. In [97], Gao and Wang propose a reconfigurable ADC with online testing capability using a configurable switch network. In [98], Peralias et. al., propose a practical implementation of a DfT technique for a pipelined ADC. In [99], Kolarik et al., propose to develop self-exercising analog checkers for concurrent detection in analog and mixed signal circuits. In [100], the authors propose to develop a programmable window comparator with adaptive error threshold for analog online testing. In [101], an online BIST approach using window comparators and current-based checker circuits for mixed signal systems is presented. While techniques have been developed for mixed-signal circuits, online testing of RF circuits and systems is still in its early stages. Some of the listed techniques involve circuit-level feedback that is more challenging to implement in RF circuits due to stability issues arising from parasitic effects. In [102], Haralampos et al, discuss the design of an adaptive checker for concurrent error detection based on common mode signal analysis. An on-line test strategy is proposed for testing RF circuits in [103]. The technique uses simple voltage comparator to analyze the power spectral density functions of RF circuits and determine their characteristics. In [104], Natarajan et al., propose an online testing technique to assess the health of wireless RF transmitters. The spectral features of real-time signals at the baseband input and the corresponding spectral content at the output of the RF node are jointly processed by the DSP in the system to determine the performance of the transmitter without impeding normal service. The proposed technique requires extensive “supervised learning” during the production phase of the device. As opposed to earlier proposed techniques of utilizing

supervised learning algorithms [104], the current approach accomplishes real-time performance monitoring of RF front ends using signal processing algorithms running on the baseband DSP that perform rapid RF front-end parameter estimation without any prior learning.

3.2.1. Proposed Approach

The approach presented in this work uses a signal processing module in the baseband DSP to compute what the (transmitted) down-converted signal would look like if the RF transmitter were ideal (linear). The model that allows this computation is called the "parallel model". Using minimal hardware overhead in the form of an embedded power detector sensor and a mixer, the output signal is referenced against the processed input in the OFDM transmitter baseband unit. Using minimal computation power, the parameters of the system are estimated to acceptable accuracy limits. The log of the behavioral parameters versus time can be used to perform wear out and aging related prognostics for preventive maintenance. In this work, we do not discuss the causes of the performance degradation but model the degradation in the form of changes in specifications. There exist a number of physical phenomenons such a hot carrier injection (HCI), oxide soft breakdown (SBD) that cause performance degradation in RF circuits. These effects cause changes in the device parameters such as tranconductance, mobility, and threshold voltages shifts. These parameter changes result in deviations of circuit specifications such as gain, non-linearity, matching, and noise figure of RF circuits [105].

3.2.2. Architecture of Proposed Solution

The architecture of the proposed technique is shown in Figure 27. The shaded blocks are the online test circuitry. Due to the high peak to average ratio (PAPR) of the OFDM system, the power amplifier is designed to exhibit high levels of linearity across a wide power range of the input signal. Hence, in field estimation of distortion levels is feasible

only with high power input signal levels. Consequently, the validation engine can be turned on only when the input signal power is above a predetermined (calibrated) threshold level. The RF output of the transmitter is down converted using an auxiliary mixer, which can be designed to have relatively high linearity in comparison to the front-end mixer. The auxiliary ADC present in the system is used to sample the down converted signal and the sampled time domain signal is referenced against the real-time input. A QR factorization-based method is used to determine the behavioral parameters of the system, and is used to determine the performance deviations of the transmitter. The delay unit is programmed based on the delay (phase delay) through the RF transmitter path and provides a reference for synchronization of the output signal with input. A switch present at the local oscillator output is used to periodically monitor the performance of the I and Q paths. The operation of the system is described in Figure 28. Since the online test circuitry is used only during certain periods of the test intervals and is switched off rest of the time, degradation effects on the test circuitry are less prominent compared to the main RF transmitter module.

3.2.3. Mathematical Framework

In this section, the basic theory of parameter estimation technique of the RF DUT using its real-time transient signal is developed. The conceptual block diagram for N-order model estimation is given in Figure 29. Assuming there exist three different signals as shown in the Figure 29. For a given Reference Test Stimulus (RTS), the Golden Response Signal (GRS), and the Distorted Response Signal (DRS) are captured. For simplicity, the RTS is shown to be a single tone sine wave. Golden Response Signal (GRS) is the signal output if the transmitter was to have ideal characteristics (assuming gain of one V/V), and Distorted Response Signal (DRS) is the actual signal at the output of the transmitter (shown in Figure 29).

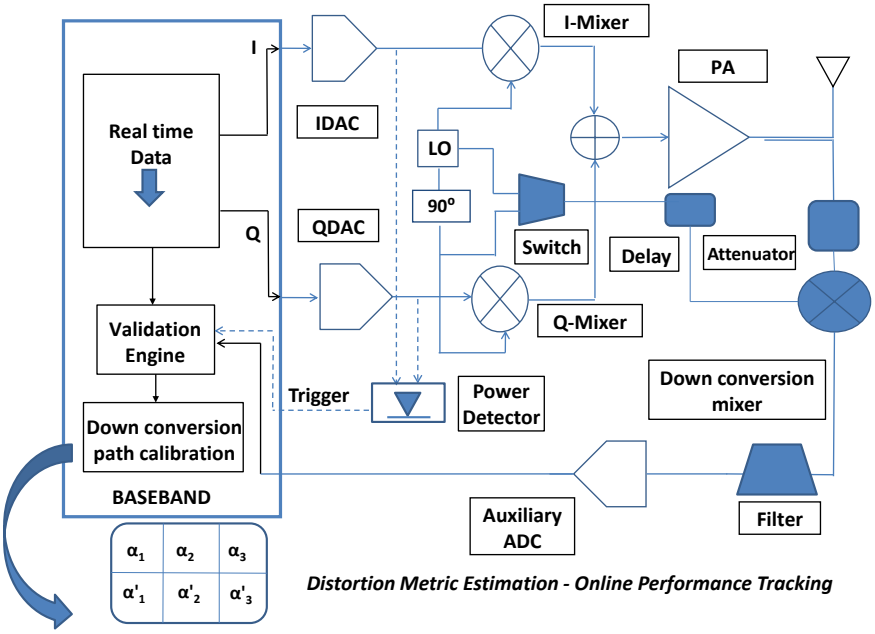


Figure 27: Block diagram of the proposed technique.

Key observation: The polynomial that maps the golden response signal (GRS) to the distorted response signal (DRS) captures the distortion characteristics of the nonlinear transfer function of the RF device under test (DUT). If the GRS and the DRS of a DUT are known, the polynomial that characterizes the distortion can then be computed by constructing the "Vandermonde" matrix v. Let us assume that the GRS is assigned as x and DRS is assigned as y. The objective is to determine the polynomial 'p' that maps x to y. Let the degree of the polynomial be defined to be 'n'. For a given x and a degree 'n' the Vandermonde matrix v can be constructed as follows:

$$v_{i,j} = x_i^{n-j}, \qquad \text{Equation 18}$$

where $v_{i,j}$ is an element of the Vandermonde matrix v with row index 'i' and column index 'j'. Once the Vandermonde matrix is computed, the polynomials are obtained by solving the following equation in the 'least squares' sense.

$$v.p \approx y. \qquad \text{Equation 19}$$

Equation 19 is solved by using standard QR factorization techniques to calculate p in a computationally efficient manner.

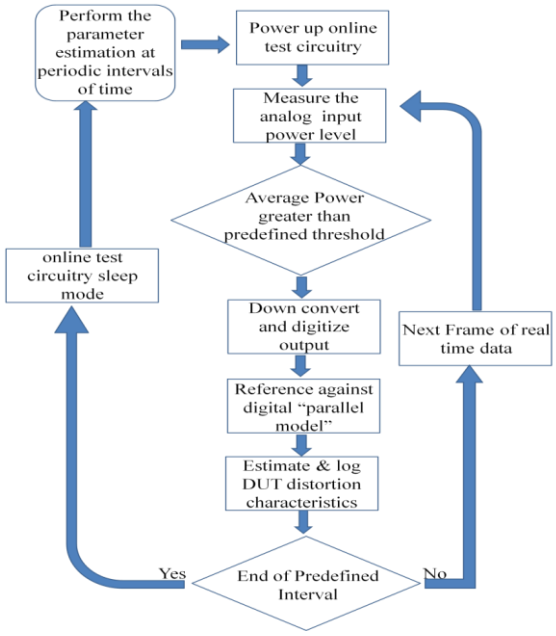


Figure 28: Flowchart of the proposed methodology.

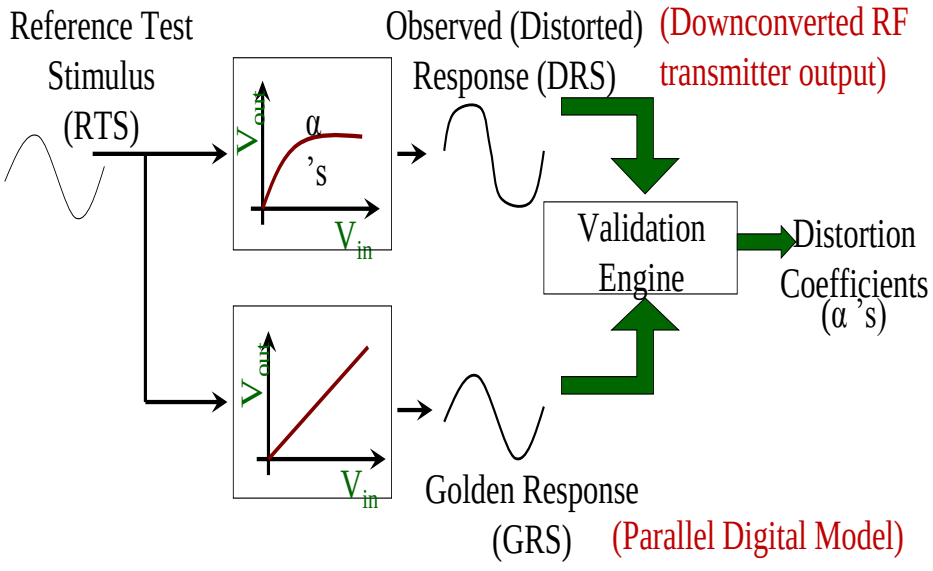
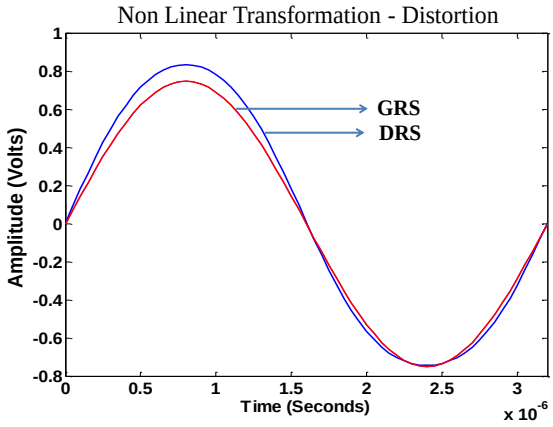


Figure 29: Conceptual diagram for proposed theory.

An example of the mathematical theory is provided in the Figure 30. The responses at the output of the RF devices (GRS and DRS) to a single tone excitation are shown only for explanation purposes. In practice, the developed theory of comparison-based estimation makes no assumption of the waveform of GRS and DRS. The above proposed concept is one of the many optimization algorithms that can be implemented to determine the distortion characteristics. Depending on the computation and performance demands of the required solution, many other linear and non-linear optimization algorithms can also be implemented to determine p in Equation 19.



Distortion Polynomial (p) = $f(x) = 0.009 + 1.1894x^1 + 0.0663x^2 - 0.2410x^3$

Figure 30: Sine wave example of the mathematical theory.

3.2.4. Application to Transmitter Framework

In quadrature modulator architecture as shown in Figure 27, the input to the power amplifier (PA) can be written as follows:

$$V_{PAin}(t) = I\big(cos(wt)\big) - Q\big(sin(wt)\big). \qquad \text{Equation 20}$$

Let us consider that the non-linearity of the PA is characterized by a function *f*. Hence, the PA output can be written as:

$$V_{PAout} = f\left(I\left(cos(wt)\right) - Q\left(sin(wt)\right)\right)$$ Equation 21

$$V_{PAout} = \alpha_1(V_{PAin}) + \alpha_2(V_{PAin})^2 + \alpha_3(V_{PAin})^3$$ Equation 22

The α_2 term produces frequency components that are either at baseband or twice the carrier frequency. The terms $\alpha_1$and α_3 contribute to components at carrier and higher frequencies respectively. The output after simplification is stated as follows:

$$V_{PAout} = \alpha_1\left(I\left(cos(wt)\right) - Q\left(sin(wt)\right)\right) + \alpha_3 I^3 \frac{\left(cos(3wt)+3cos(wt)\right)}{4} -$$
$$\alpha_3 Q^3 \frac{\left(sin(3wt)-3sin(wt)\right)}{4} - \alpha_3 I^2 Q \frac{\left(3sin(wt)\right)}{4} + \alpha_3 Q^2 I \frac{\left(3cos(wt)\right)}{4} +$$
$$3\alpha_3 Q^2 I \frac{\left(cos(3wt)\right)}{4} - 3\alpha_3 I^2 Q \frac{\left(sin(3wt)\right)}{4}.$$ Equation 23

Hence, the resultant signal around the carrier signal can be approximated to as shown in Equation 24.

$$V_{PAout} = \alpha_1\left(I\left(cos(wt)\right) - Q\left(sin(wt)\right)\right) + \alpha_3 I^1 \frac{\left(3cos(wt)\right)}{4}(I^2 +$$
$$Q^2) + \alpha_3 Q^1 \frac{\left(3sin(wt)\right)}{4}(Q^2 - I^2).$$ Equation 24

Now this high frequency signal is down converted by a built-in down conversion mixer running at carrier frequency. This down converted signal is filtered, sampled by an auxiliary ADC. The down converted signals have the distortion information and can be used to find the nonlinearities in the I and Q path as follows:

$$V_{PAout} * \left(cos(wt)\right) =$$
$$\alpha_1\left(I\left(cos^2(wt)\right) - Q\left(cos(wt)sin(wt)\right)\right) + \alpha_3 I^1 \frac{\left(3cos^2(wt)\right)}{4}(I^2 +$$
$$Q^2) + \alpha_3 Q^1 \frac{\left(3sin(wt)cos(wt)\right)}{4}(Q^2 - I^2).$$ Equation 25

The above stated signal is low pass filtered, and the resultant signal is digitized.

$$LPF\left(V_{PAout} * \left(cos(wt)\right)\right) = \alpha_1 \frac{I}{2} + \alpha_3 \frac{3I^1\left(I^2+Q^2\right)}{8}.$$

When $I \gg Q = \alpha_1 \frac{I}{2} + \alpha_3 \frac{3I^3}{8} = f(I).$ Equation 26

where LPF is the low pass filtering of the signal. Similarly, multiplying the output PA signal with quadrature LO signal results in a signal of the form shown in Equation 27.

$$LPF\left(V_{PAout} * \left(sin(wt)\right)\right) = -\alpha_1 \frac{Q}{2} - \alpha_3 \frac{3Q^1\left(Q^2-I^2\right)}{8}.$$

when $Q \gg I = -\left(\alpha_1 \frac{Q}{2} + \alpha_3 \frac{3Q^3}{8}\right) = -f(Q).$ Equation 27

Hence, LO signal to the down-conversion mixer is switched and the auxiliary ADC is used to obtain the response to find out the gain and non-linearities in both I and Q paths separately. This switching can be performed at periodic intervals.

3.2.5. OFDM Transmitter Baseband Modeling

In OFDM, orthogonal frequency encoded onto a large number of closely spaced orthogonal sub-carriers. OFDM is a spectrally efficient technique and has inherent robustness to inter-symbol interference and multi-path fading across a channel. Hence, the OFDM technique is the preferred choice for most of the commercial standards present today. The incoming data bits to an OFDM transmitter is split into parallel data streams according to the number of subcarriers present in the system. After digital modulation, the data streams are mapped onto the sub-carriers using an inverse fast fourier transform (IFFT) to result in a noise-like time domain OFDM signal. Guard time and cyclic prefix is encoded onto the OFDM signal just before transmission to result in an OFDM frame. The signals (real-time) such as the one shown in Figure 31 are used as the inputs for validation engine for parameter estimation technique and the subsequent performance monitoring.

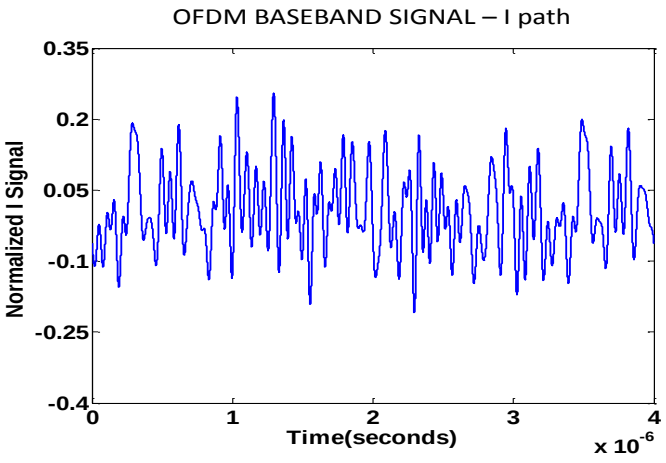


Figure 31: Baseband OFDM signal.

3.2.6. Modeling RF Front Ends

The incoming baseband signal is converted into an analog signal using digital to analog converters (DAC). The analog/RF front-end consists of mixers and amplifiers required for the up-conversion of the baseband signal into RF frequencies and subsequent amplification.

Mixers: In this work a mixer is modeled as an ideal multiplier C = x1(t).x2(t) for an input signal x1 and 'LO' signal x2. The ideal multiplier is followed by a non-linear transfer function block defined by polynomial distortion coefficients.

Amplifiers: In this work, amplifiers are modeled as a non-linear transfer block. A model so chosen is suitable for online monitoring applications. For a given input signal x(t), output y(t) can be represented as a 3rd order polynomial as shown in Equation 24. The coefficients are extracted from the Vin vs. Vout circuit level simulations using HP Advanced Design System (ADS) where α_1 is the gain and α_2, α_3 are 2nd and 3rd order nonlinearity relating to IIP2 and IIP3 respectively. Once the behavioral parameters are identified, these can be used in the simulation environment to determine the gain and distortion characteristics of the circuit.

3.2.7. Simulation Setup and Results

The OFDM system consisting of 200 frames of data is used in the simulation setup. Each frame consists of the 128-point IFFT data and the guard band bits. Noise performance of the system is considered through the addition of additive white Gaussian noise in the simulation setup. The mixer and power amplifier are modeled as discussed in the previous sections. The simulation results presented are discussed with reference to the Figure 32. The input power signal to the transmitter (in the I-path shown in Figure 27) is monitored continuously over the data frames. The initial set of data frames (up to 50) experience low power signal levels in the I-path. Performance deviations are induced during the OFDM data frames through deviations power amplifier characteristics and the ability of the system to track the performance deviations is presented. It is to be observed that gain tracking is realized during all the frames, whereas IP1dB has a greater dependence on the input power level in I channel. This is visible in the first 50 frames where the tracking is better when the input power amplitude is higher. At Frame 50, the IP1dB is increased from the nominal value. The increased linearity requires the power to be higher than that of normal power levels for the distortion characteristics to be excited. Hence, during this period the distortion estimation trends follow the input power level. For Frames 80 to 200, the IP1dB is reduced from -11.73 dBm to -19.0 dBm. As a result of reduction in IP1dB, lower power levels excite the non-linearity characteristics of the front end and IP1dB performance deviation is tracked. The relative error in tracking the system specifications when power levels exceed threshold levels are provided in Table 6.

As an experiment, performance deviation in the down conversion mixer is considered and the simulation was repeated by injecting up to 10% mismatch in the down conversion mixer and its corresponding digital model. The relative error in the specification tracking is provided in Table 7. It should be noted that even under performance variation in the mixer the tracking trend is maintained. To explain the logistics of performance tracking

the normalized power spectrums of Frame 65 and 120 are shown along with the corresponding input signals. From Figure 33 and Figure 34, it is clear that there exists spectral leakage in the data frame 120 caused by the amplifier distortion. This distortion is captured in the down converted time domain signal thereby enabling the performance deviation monitoring. The distortion performance evaluation can be used to tune the system for increased reliability as well as better performance. The tuning approach involved is beyond the scope of this current work and hence is not discussed here.

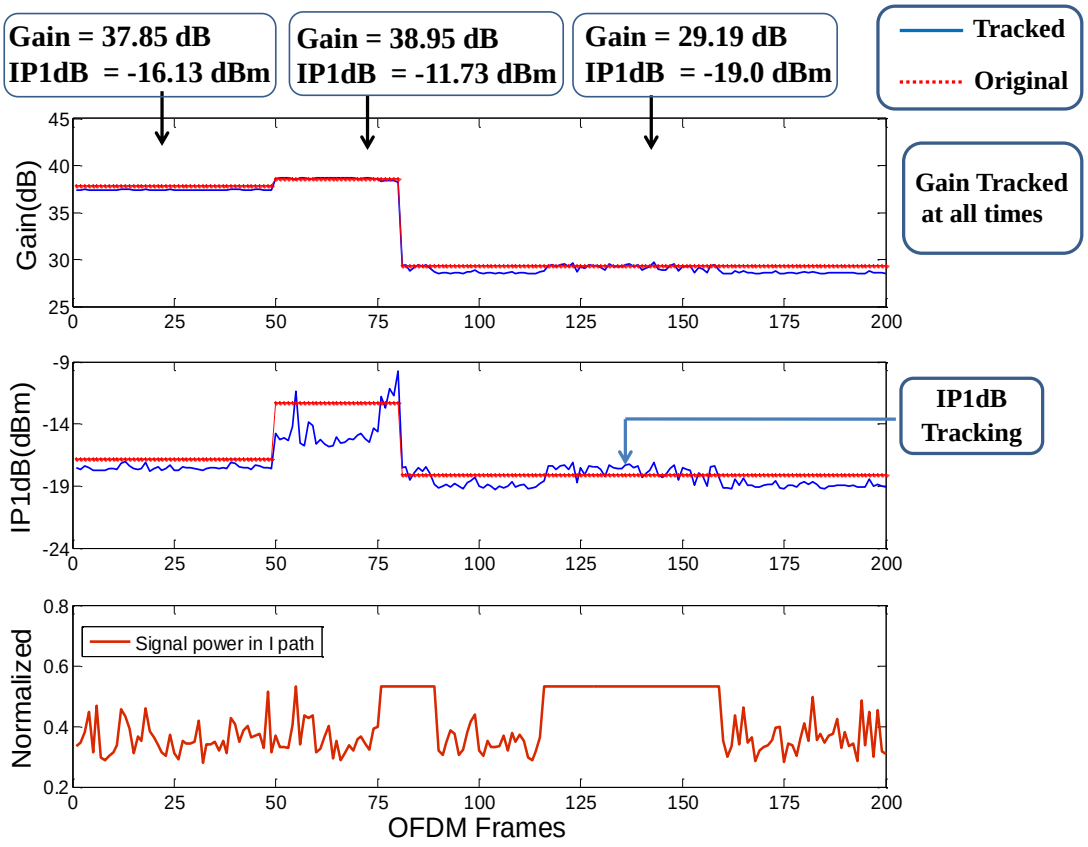


Figure 32: Performance monitoring of in-field RF transmitter.

3.2.8. Key Contributions

This work introduces a novel real time parameter estimation technique that provides for accurate performance monitoring of in-field RF/analog modules. The technique uses a "parallel digital model" of the RF transmitter path to capture the distortion performance

variations. The proposed architectural overhead is minimal in terms of computation complexity and hardware. The presented technique does not involve any supervised learning techniques and has strong implications in providing prognostics of in-field DUTs using the log of the in-field system performance deviations. Alternative algorithms can be used to solve the model parameters. However, the choice of the algorithm depends on the implementation and model complexity.

Table 6: Relative error in specification monitoring of the system.

Specification	**Gain(dB)**	**IP1dB(dBm)**
Relative error (%)	3.414	7.54

Table 7: Specification monitoring in presence of feedback performance deviation.

Specification	**Gain(dB)**	**IP1dB(dBm)**
Relative error (%)	6.414	10.91

In recent times, a non-linear constrained optimization technique has been used in conjunction with optimized pilot symbols to determine the transmitter performance deviation in real-time [106]. While such a technique might provide performance tracking with higher accuracy, it requires explicit test generation scheme on pilot symbols and increases the DSP workload in terms of the solver complexity. Thus, there exists a trade-off between the accuracy of performance tracking, the characterization steps, and the optimization complexity.

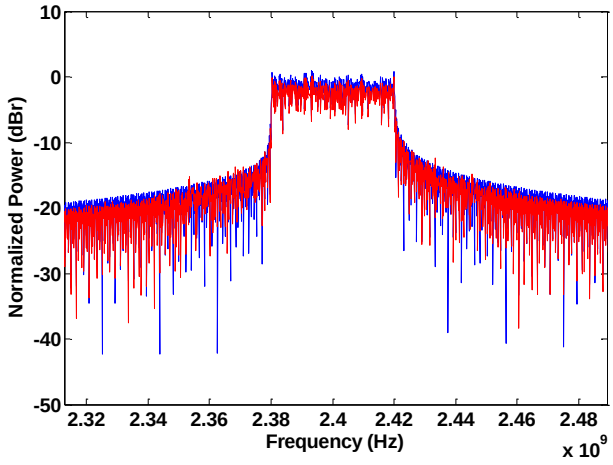


Figure 33: Spectral plot of frame 65-minimal distortion.

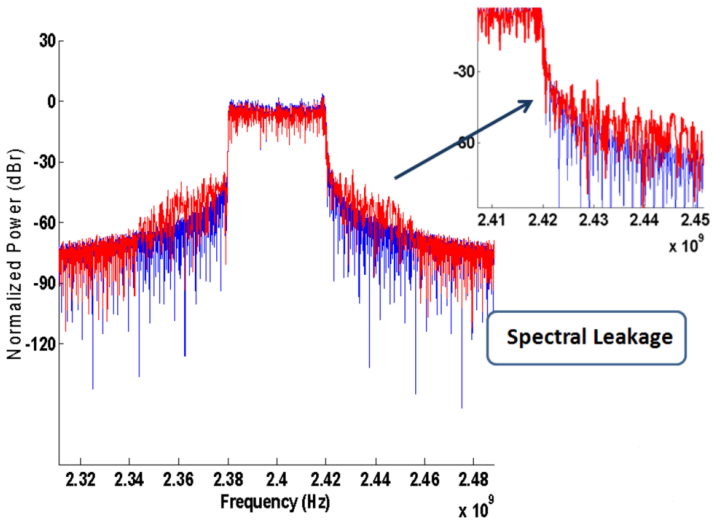


Figure 34: Spectral plot of frame 120-spectral leakage causing distortion.

CHAPTER 4. DIGITALLY ASSISTED TUNING TECHNIQUES FOR ANALOG AND RF SYSTEMS

Higher levels of CMOS radio integration and technology scaling trends pose a substantial challenge for producing highly efficient and linear RF circuit designs in the face of large process variations. Low cost manufacturing techniques for scaled technology nodes cause increased process variations, the impact of which, present themselves as imbalances and distortion in millimeter wave devices. Design methodologies to combat these process variations have led to over-designing the circuit in terms of power and area. Hence, in this work, intelligent post-manufacture tuning methods for reducing the impact of process variations on the performance of such devices are investigated to enhance yield. To perform the multi-dimensional tuning of these analog/RF devices in a time-efficient manner, at every step the state (in terms of specifications) of the devices needs to be evaluated. In this work, intelligent methods to determine and correct the circuit specifications by performing an iterative test-tune-test methodology is described.

In the past, digital compensation techniques that address specific impairments such as I/Q mismatch, skew, IIP3, etc. have been developed. Further, to perform tuning using digital compensation, most tuning techniques rely on signal processing operations performed by the digital baseband (or the digital signal processor) in the case of a BIST scenario or the tester in the production environment. In analog compensation, the calibration is performed in the analog domain by using local feedback to modify circuit characteristics such as bias voltages or currents or using passive variables circuit elements such as varactors, inductors with taps and resistor banks. These methods focused on providing tuning solutions for particular specifications of individual circuits by either using localized feedback/feed-forward to tune for the specification. In the cases

where different parameters are tuned, these are performed in a sequential manner leading to high test/tune times.

4.1 Overview of Self-Tuning/Self-Healing Methodology

The aim of this research revolves around developing a framework for designing self-healing RF systems (SISO and MIMO) using intelligent post-manufacture built-in diagnosis and tuning algorithms that will adapt autonomously to manufacturing process variations with the least (negative) impact on power consumption (power-conscious self-healing). Various frameworks that perform self-healing based on the system-level resources available are explored. The basic ideology is shown in Figure 35. A control unit (which may be the DSP of the system or a simple on-chip dedicated digital circuitry) uses sensors present in the system to diagnose the health. These sensors can be at the system output as well as at intermittent nodes. The sensors enable low frequency capture of design for test (DFT) responses to perform tuning with a low-cost infrastructure, amenable for production floor deployment (time T=0) as well as in field self-healing/self-tuning (built-in tuning). Further, the control unit tunes for the RF impairments of the system according to a control algorithm. The RF front-end modules that have process-induced imperfections are designed with built-in tuning knobs that provide the capability to tune the performance. As the performance of the device is tuned, the power consumption of the device varies. Finally, an adaptive self-healing methodology that attempts to maximize the performance of wireless devices across channels for a given system power constraint is presented. The methodology trade-offs the performance of the device under certain conditions for a given system-level power constraint. Such a technique results in devices that are tuned for various power-performance combinations and results in binning of wireless devices according to system-level performance constraints.

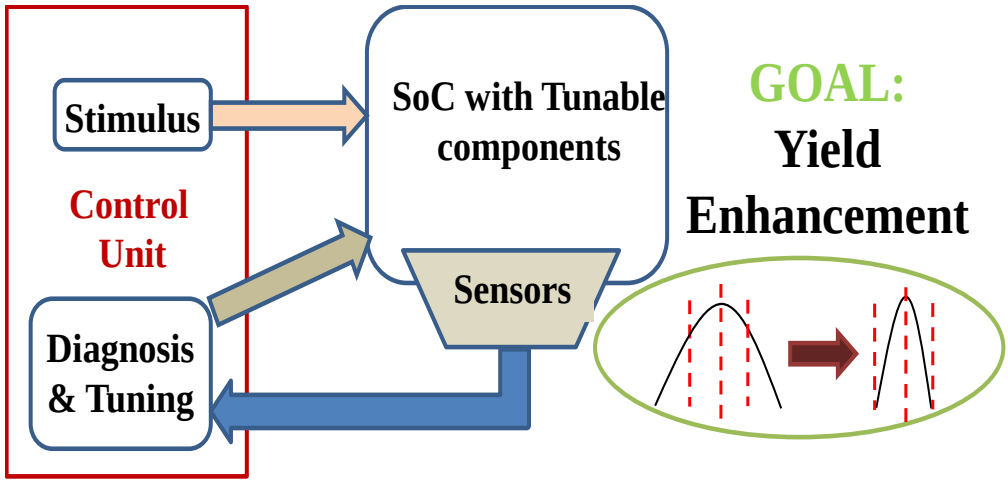


Figure 35: System-level self-healing conceptual diagram.

The key components of the self-healing methodology are as follows:

- Optimal stimulus generation such that the output response of the device shows high sensitivity simultaneously to the specifications of the DUT under process variation space as well as the tuning knob space.
- Analog or digital signature-driven tuning algorithm that tunes the tuning knobs of the system to perform efficient test-tune-test methodology thereby enabling an overall yield improvement.

4.2. Principles of the Proposed Tuning Approach

4.2.1. Optimal Stimulus Generation

In alternate testing or supervised learning (Section 2.1), the test stimulus to a mixed-signal/RF DUT is designed in such a way that under process variability effects, the test response (observed in the time or frequency domain) exhibits strong statistical correlation with its test specification values. Using the observed response, the test specification can be predicted from the former using nonlinear regression functions that map the obtained response to the corresponding DUT specification values.

Considering there exist j process parameters represented by $P_j = [p_1, p_2, p_3 \ldots., p_j]$, n DUT response measurements represented by $M_n = [m_1, m_2, m_3 \ldots., m_n]$, k specifications represented by $S_k = [s_1, s_2, \ldots s_k]$, and a set of tuning knob values $T_l = [t_1, t_2, t_3 \ldots., t_l]$, an optimum test stimulus is developed such that a strong statistical correlation between the observed test response and a specified set of DUT specifications is exhibited under large process variations across the parameters P_j for a range of tuning knob values T_l. In the above, strong statistical correlation across simultaneous multi-parameter perturbations in the vector $[P_j, T_l]$ is implied. For simulation purposes, the parameters P_j are assumed to be Gaussian with larger than normal (calibrated) values of σ, where σ is the vector of standard deviations of each parameter in each element of vector P_j. The parameters of T_l are assumed to be uniformly distributed across a predetermined (calibrated) range of values. The T_l parameters are used for performing tuning. The methodology is illustrated in Figure 36. The optimum stimulus generation for performing tuning is developed by considering each tuning knob setting of each process instance as an individual DUT for genetic algorithm-based test generation.

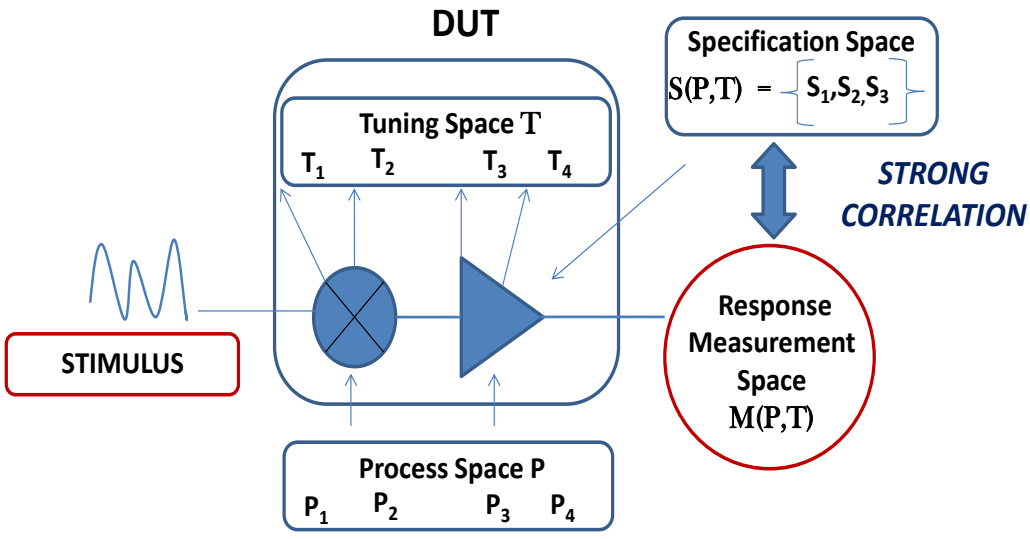


Figure 36: Optimum test stimulus generation concept.

4.2.2. Tuning Methodologies Overview

As mentioned in Chapter 1, there exist different tuning approaches that have been used in the past. These approaches include using the system DSP, on-chip logic as the control unit for performing tuning. Different tuning methodologies have used different knobs for compensation of impairments. These include digital baseband filters, current DACs, resistor switches, digital gain, and phase correction. As mentioned in Section 1.6, prior tuning methodologies can be broadly classified as one of the following categories:

- Digital baseband (also known as the digital signal processor (DSP)) based monitoring and tuning
- On-chip monitoring and analog tuning
- Digital baseband monitoring with analog and digital co-tuning

In this work, techniques have been developed in these categories that aid in performing faster and low-cost tuning of multiple specifications of device.

Typically, the main drawback of these digital compensation techniques is the long convergence time for the baseband adaptive filter coefficients. Further, they use the traditional receiver chain in loopback mode to perform the compensation. In analog on-chip tuning techniques, the bias voltages/currents of different circuits are varied to bring the circuit-level specification back within its specified limits. These circuit-level feedback techniques might be difficult to implement for RF circuits, due to stability issues. Finally, the DSP or the tester in the production environment can be used as a control engine to perform the iterative test-tune-test steps. In this technique, in the BIST scenario, the DSP chip of the system controls the performance of the front-end chip. The DSP uses both digital and analog knobs to correct the imperfections. However, the techniques discussed in the past do not perform power-conscious tuning and often correct

the imperfections in a sequential manner. In this work, a framework is developed to perform concurrent tuning of multiple parameters in a low-cost iterative manner.

The tuning techniques developed in this thesis, contribute in the following manner:

- In the realm of DSP based monitoring and tuning, a methodology for reducing the compensation time without using receiver chain in loopback mode is proposed.
- In this work, a low-cost DSP-based methodology that simultaneously tunes multiple specifications with minimal overhead in power consumption is proposed. Further, in the case of devices that consume more than an acceptable power consumption, an adaptive tuning technique that maximizes the performance of the device for a given power constraint is presented. An optimal trade-off methodology that trade-offs performance under certain channel conditions for a wireless system against power consumption is explored.
- Alternative to the DSP-based tuning, an architecture that uses simplified on-chip digital logic to perform tuning of process-induced circuit imperfections is developed in this work. This technique aids in reducing the workload on the system DSP and is amenable for performing BIST at wafer-level before front-end integration with system baseband. This work provides a complete on-chip self-tuning solution for tuning multiple specifications of a RF transmitter concurrently using digital logic.

4.3. Digital Baseband (DSP) Based Monitoring and Tuning

4.3.1. Low-Cost Digital Correction Scheme for PA Imperfections

RF power amplifier linearity is a critical metric that determines the quality of the transmitted signal, and the out-of-band power determines if the stringent FCC spectral mask regulations are met. However, the compressive nature of RF power amplifiers (RF circuits in general) causes them to operate in a non-linear manner at high power levels. This nonlinear operation results in intermodulation distortion products and spectral regrowth resulting in poor transmitted signal quality. To ensure linear PA operation under signals with high peak-to-average ratio as in OFDM, the power amplifier is backed off by several dB from saturation. This however, causes the power amplifier to operate at low levels of efficiency resulting in higher than necessary power consumption.

In general, AM-AM and AM-PM distortion effects are relatively uniform across small bandwidths of operation (20 MHz) for "good" narrowband RF power amplifier designs. However, wideband devices (devices designed to operate at more than one carrier frequency, such as a WiMax PA that operates from 2.3 to 2.7 GHz) suffer different distortion effects at frequencies spaced far apart from each other in the frequency domain (100 MHz) [107]. Hence, there has been increased emphasis on linearization techniques, both adaptive and non-adaptive [60][61].

Standard techniques used for adaptive predistortion perform tuning of the predistortion coefficients utilizing the receiver chain [60]. The power amplifier output is downconverted, demodulated and fed back through the internal receiver chain. The signal is then processed digitally to improve the overall end-end linearity. This technique, however, is affected by receiver LNA and down conversion mixer non-linearities, I-Q demodulator amplitude and phase mismatches and ADC non-linearities arising in the down conversion procedure [108]. As transistor sizes continue to decrease and the impact of process variations on the RF components becomes significant, large perturbations in the receiver noise and nonlinearity specifications are caused across different devices.

These variations in the receiver chain affect the optimality of the RF power amplifier's nonlinearity compensation using predistortion.

4.3.1.1. Proposed Methodology

Our proposed method provides for testing and adaptive compensation of a wideband RF power amplifier using a built in envelope detector sensor on the tester load board that consumes little area. A specially crafted multitone test stimulus generated from an external test system (test board) allows concurrent testing and compensation of the distortion characteristics of the RF power amplifier at each of the upconverted frequencies contained in the multi-tone stimulus. The captured response from the envelope detector connected to the RF power amplifier output is mapped to the AM/AM and AM/PM characteristics of the RF power amplifier using non-linear regression mapping functions built from calibration experiments. The estimated behavioral parameters are used to obtain the inverse predistortion transfer function (digital) for the RF power amplifier. The least mean squares algorithm is used to tune the predistortion coefficients using the envelope detector output to guide the search for the best tuning parameters using iterative test application. This allows faster convergence of the predistortion coefficient optimization process than current methods. In addition, as opposed to existing methods in which multiple tuning procedures need to be run at different carrier frequencies, a single testing and tuning procedure is used to perform concurrent testing and tuning at all the frequencies concerned, concurrently.

A multi-tone stimulus consisting of the RF frequencies at which diagnosis and linearization are to be performed is passed through the power amplifier and the output obtained is used as an input to the envelope detector. To develop the MARS mapping functions a set of PA processes instances are used and the envelope responses for each instance is obtained.

The behavioral parameters of the power amplifier instances at each of the different carrier frequencies (considered in the above step) are obtained using the conventional test technique and the behavioral-level specifications at each carrier frequency is extracted.

The envelope detector output for each of the selected instances is sampled and processed at the baseband to develop a mapping function between behavioral parameters(calculated in the second step) and the obtained response (in the first step) and the AM/AM and AM/PM distortion effects of these instances of the power amplifier at each desired frequency is estimated.

This test can be applied from an external tester to the power amplifier during production testing and response of the power amplifier can be captured and processed to determine the non- linearities.

The predicted behavioral parameters are stored in a LUT indexed according to the carrier frequencies. The inverse transfer function coefficients are then computed and stored in LUT of the baseband system indexed according to frequency of operation. To obtain higher accuracy in predistortion, LMS algorithm is then used to fine-tune the predistorter filter coefficients. The overall proposed approach is shown in Figure 37 and Figure 38 respectively.

4.3.1.2. Behavioral Modeling of Power Amplifier

Behavioral models of power amplifiers provide an efficient method of observing the power amplifier characteristics without intricate and complex transistor-level simulation. Behavioral modeling of power amplifiers has been extensively studied and the models can be classified as (a) memoryless non-linear systems, (b) quasi-memoryless and (c) non-linear systems with memory [109][110]. Volterra series and neural network based methods are used in modeling these systems. A simple memoryless modeling technique is used for modeling AM/AM and AM/PM characteristics. Since the system considered

operates with narrow bandwidth (20 MHz) where the frequency dependent effects are not considerable, the use of the following model is justified. Hence, in this model the outputs depend only on the instantaneous inputs. The model is described using the following two Saleh equations [60] [109]:

$$A(\gamma(t)) = \frac{\alpha_1 \gamma(t)}{1+\alpha_2 \gamma^2(t)}, \quad \text{Equation 28}$$

$$\emptyset(\gamma(t)) = \frac{\beta_1 \gamma^2(t)}{1+\beta_2 \gamma^2(t)}, \quad \text{Equation 29}$$

where $A(\gamma(t))$ is the amplitude distortion and $\phi(\gamma(t))$ is the phase distortion in the input signal represented in the form shown below in Equation 30,

$$x(t) = \gamma(t) e^{j\omega(t)}, \quad \text{Equation 30}$$

resulting in a output signal shown below in Equation 31,

$$y(t) = A(\gamma(t)) e^{j\left(\omega(t)+\emptyset(\gamma(t))\right)}. \quad \text{Equation 31}$$

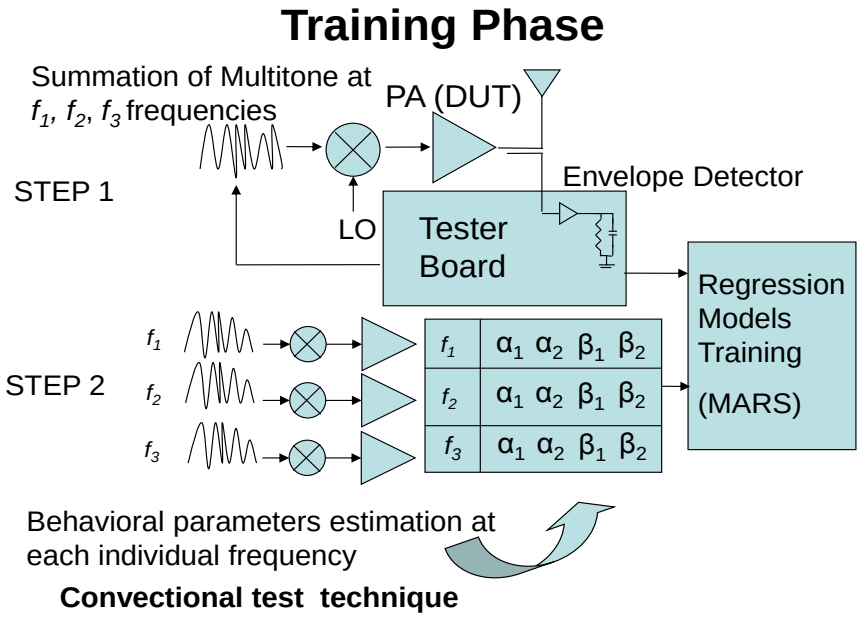


Figure 37: Training phase of the proposed methodology in production.

The power amplifier is designed in 0.18u CMOS technology in Agilent Advanced Design System (ADS). It consists of two stages namely the driver and the power stage. Input, inter-stage, and output matching is provided. It has a gain of 22 dB and a P1dB of 24 dBm at a frequency of 2.4 GHz. The parameters of the power amplifier are estimated in standard least-squares approach using the theory and equations explained in [111]. The input and output voltages are normalized at their respective values in saturation. The AM/AM and AM/PM characteristics at these frequencies are shown in Figure 39. If the amplitude and the phase distortions in the predistorter are denoted by $\Delta(t)$and $\vartheta(t)$ respectively, then the desired response of an ideally linearized amplifier is given by the following equations.

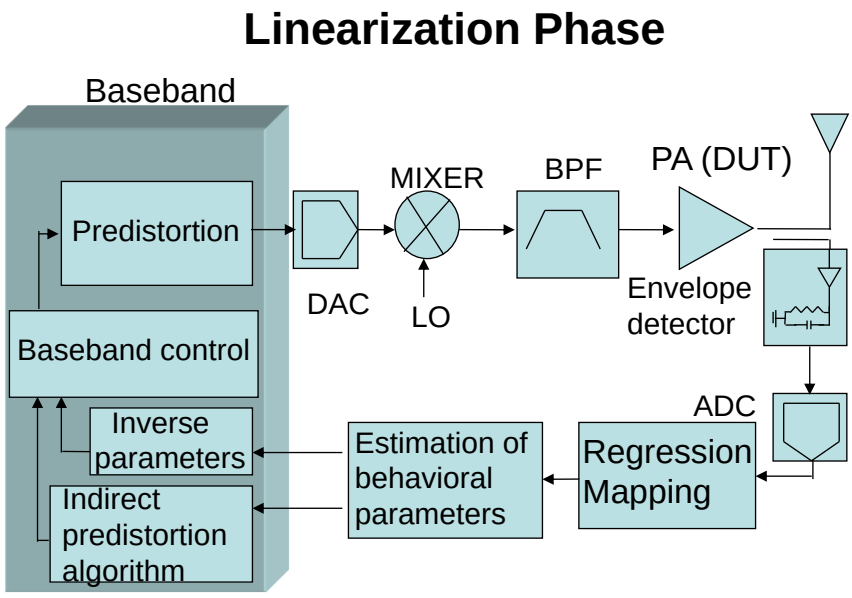


Figure 38: DSP-assisted linearization of PA.

$$A\left(\Delta\left(\gamma(t)\right)\right) = K\gamma(t), \qquad \text{Equation 32}$$

$$\emptyset\left(\gamma(t)\right) + \vartheta\left(\gamma(t)\right) = \mu, \qquad \text{Equation 33}$$

where K is the ideal gain of the amplifier and μ is the reference phase. Indirect learning predistortion is a standard approach implemented for linearization of power amplifier in

current transmitter systems [61]. However, the technique requires considerable iterations because it does not have prior knowledge of the power amplifier characteristics and starts with random initial coefficients. Our proposed technique uses a modified structure using LMS algorithm, which minimizes the error between the pre-distorted signal and the reference signal. This is used in addition to the analytically computed inverse transfer function coefficients obtained using Equation 32 and Equation 33. Using the low cost envelope detector to capture the response of the PA, the responses and the behavioral parameters are correlated using a non-linear mapping technique known as Multivariate Adaptive Regression Splines (MARS).

4.3.1.3. Simulation Results

A set of 70 instances were used in the training phase for building the regression model. The above instances are developed through Monte Carlo simulations of the power amplifier circuit. The output of the amplifier is passed through an envelope detector modeled in Matlab. A set of 30 instances are used in the prediction of the behavioral parameters simultaneously at the three carrier frequencies. This method provides for a decrease (3x) in the testing time of wideband power amplifiers compared to conventional testing methodologies. The estimated relative error in the prediction of the behavioral parameters of the power amplifier at three different carrier frequencies is shown in Table 8. In the proposed scheme, the initial predistorter filter coefficients are obtained using the alternate diagnostic testing in addition to behavioral parameters of the power amplifier. The LMS algorithm is further implemented to fine tune the coefficients of the predistorter filter. A comparison between the conventional and proposed adaptive predistortion techniques is shown in Figure 40. It can be observed based on the error progression in our scheme that faster convergence (an error of 5e-3 is obtained in 40 cycles) of the predistorter coefficients is obtained compared to the conventional indirect adaptation

scheme where the error progression is gradual (an error of 5e-3 is obtained in 160 cycles). EVM is an efficient metric used for indicating the linearity in the system.

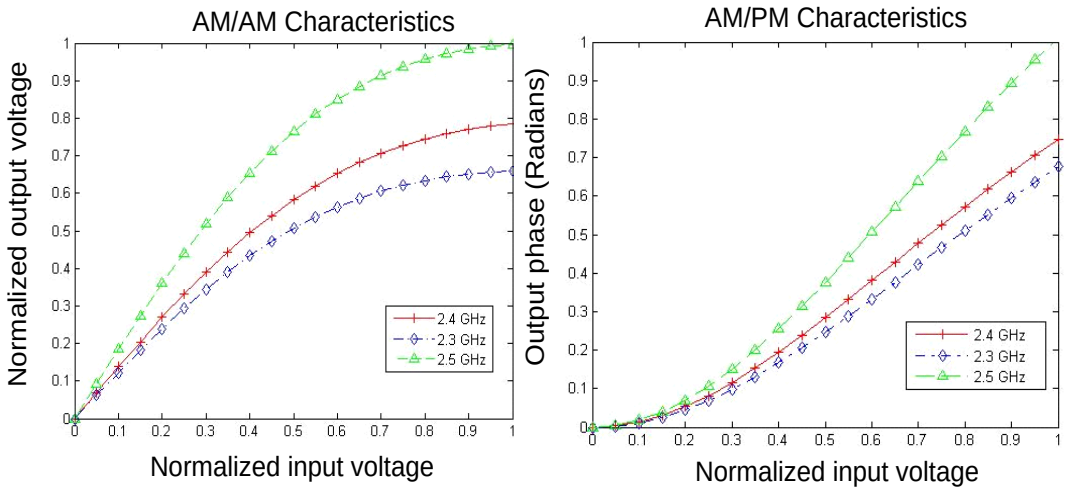


Figure 39: AM/AM and AM/PM characteristics at different carrier frequencies.

Table 8: Relative error in AM/AM and AM/PM characteristics.

Frequency	**Relative Error (%)**			
	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$
2.3 GHz	4.421	2.309	4.546	5.811
2.4 GHz	2.814	3.991	4.987	6.151
2.5 GHz	2.267	3.778	4.360	5.886

Hence, the EVM specification of system with and without predistortion are computed for a QPSK system. It can be observed that the predistortion technique improves the EVM of the system. A best-case improvement of 9.5% is obtained at 2.3 GHz. It can be

noticed that the predistorted constellation points has a lesser skew compared to the constellation points of the system without predistortion indicating the decrease of amplitude and phase distortions. The EVM plots are shown in Figure 41.

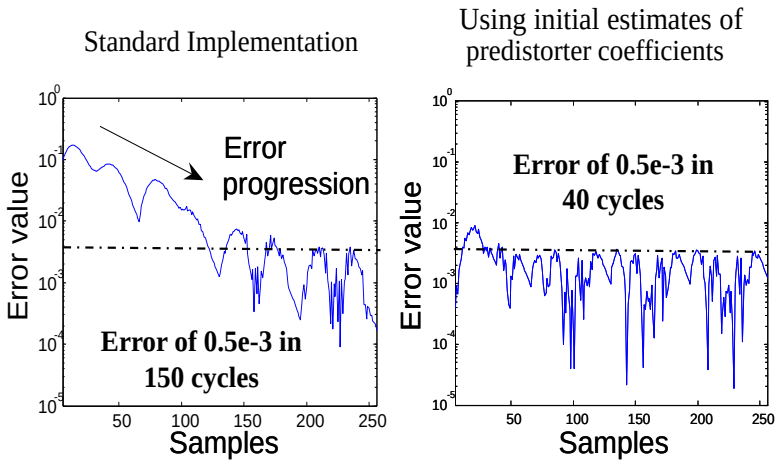


Figure 40: Error progression comparison.

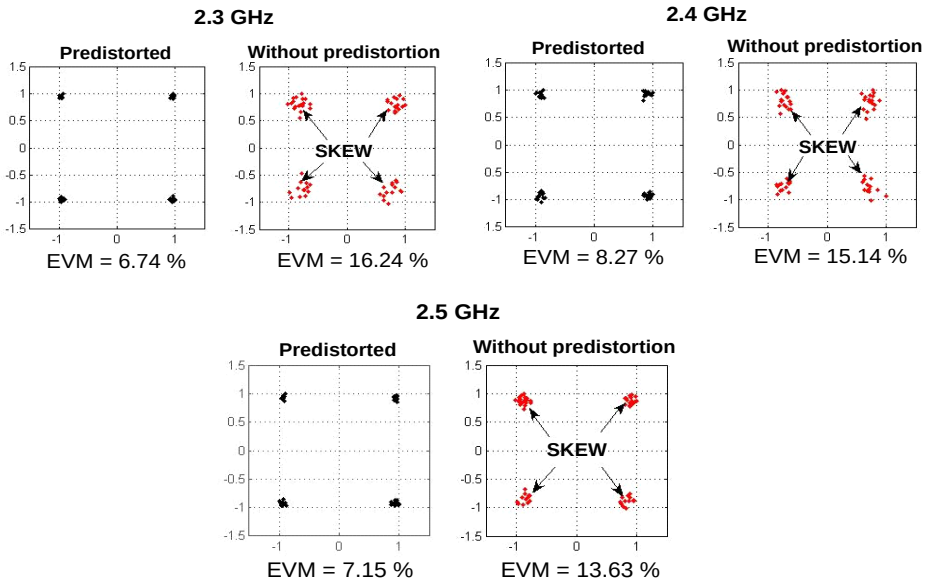


Figure 41: Constellation points of the transmitter.

4.3.2. BIST-Assisted Digital Compensation for MB-UWB Transmitters

High speed data transfer (up to 480 Mbps) over short-range distances (up to ~10 m), has made UWB a viable technology for wireless personal area network (WPAN) environments. The ability of UWB standard to coexist with other standards makes it an attractive "overlay" technology in the near future. There exist a number of applications such as Bluetooth, radar, and biomedical imaging etc., which plan to adapt UWB working standards. The FCC has legalized the frequency spectrum between 3.1 GHz - 10.6 GHz for commercial use by UWB devices with an allowed power spectral density of -41.25 dBm/MHz. Currently, multi band orthogonal frequency division multiplexing (MB-OFDM) and direct sequence code division multiple access (DS-CDMA) have been the two proposed approaches for UWB operation. The MB-OFDM system inherits the advantages of OFDM systems such as spectral efficiency, resistance to narrow band interference (NBI) and multi-path robustness. Further, the system has the ability to capture multi-path energy and the capability to turn on and off certain specific frequencies dynamically with a resolution of 4 MHz [112],[113]. The ECMA 368 standard defines the physical and the medium access layers of the MB-OFDM UWB system. The frequency spectrum of 3.1-10.6 GHz is divided into 14 bands consisting of 6 Band groups as shown in Figure 42. The bandwidth of each band is 528 MHz. UWB systems have a flexible data rate ranging from 53.3 Mbps to 480 Mbps. Operation in Band Group #1 is stated to be the mandatory mode. Techniques such as frequency and time domain spreading are provided to ensure the use of these systems under a variety of channel conditions. The wide operating range is exploited by using Band Groups #1 and #2 for longer-range applications and Band Groups #3 and #4 for shorter-range applications [114]. In [115], MB-OFDM UWB is presented as a cognitive approach for preventing interference with pre-existing technology standards. The transmitter architecture of MB-OFDM UWB systems is shown in Figure 43. Time-frequency coding with time-frequency interleaving determines the manner of frequency hopping from one

band to the other. A 128-point IFFT is employed using 100 data carriers, 10 guard tones, 12 pilot tones and 6 NULL tones.

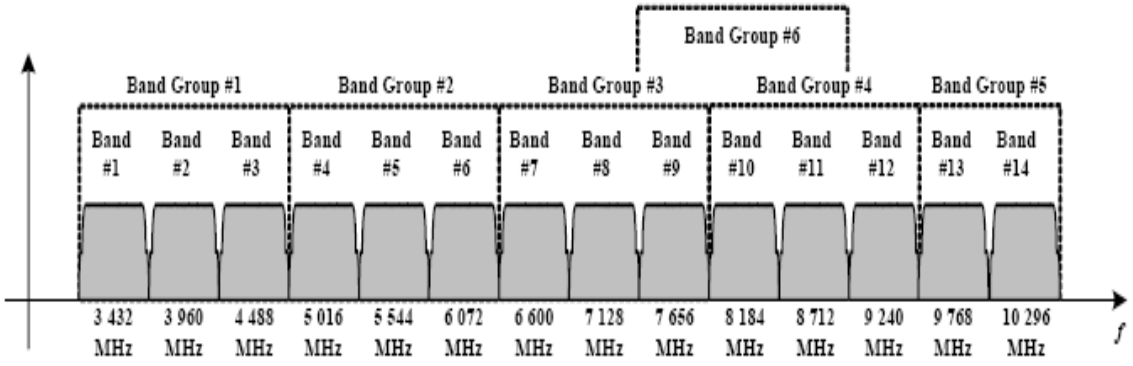


Figure 42: Band groups in MB-OFDM UWB scheme.

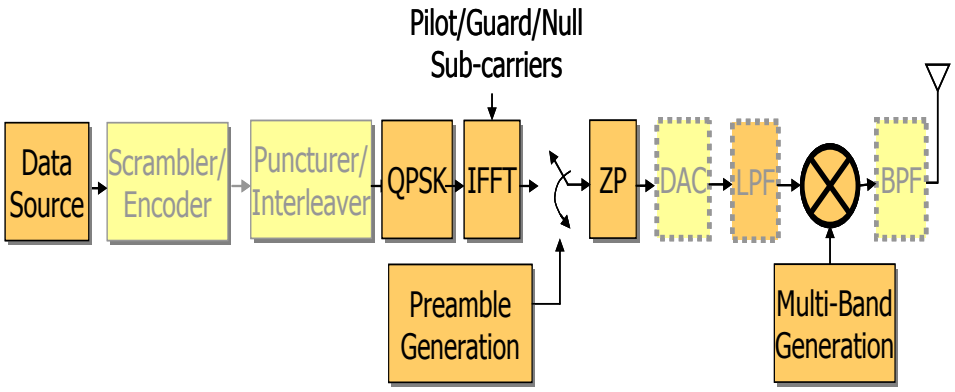


Figure 43: Traditional MB-OFDM transmitter

The present RF technology capabilities and existing standards provide opportunities for the utilization of band group #1 (3.168 - 4.752 GHz) for commercial applications. Alereon, Intel and Texas Instruments have made significant progress towards the development of RF and digital platforms required for implementation of UWB systems [116]. The authors in [113] provide a detailed methodology for future expansion into other band groups as well as development in design and industrial implementation capabilities. The standard UWB transmitter architecture does not necessitate the use of a power amplifier, though few architectures using power amplifiers have been discussed in

the literature [117]. Several such solutions capable of operating in multiple bands or the entire licensed UWB frequency range [118], [119], and [120] have been proposed in the recent past. To ensure operational reliability, wideband systems entail a "relatively flat performance characteristic" in the frequency range of interest. The above stated aspect becomes even more significant under severe process variation in the nanometer regime. Hence, there is a need to develop a low-cost wideband compensation methodology to maintain system performance and improve manufacturing yield. Compensation can be performed through analog circuit-level tuning techniques or through digital assisted techniques. In a MB-OFDM UWB system where the LO continuously hops across frequency bands, the implementation of "on the fly" analog compensation techniques will prove to be challenging. This is because the circuit switches in a period that is smaller than the settling time of circuit that is tuned. Further, digitally assisted compensation techniques provide for a relatively low cost solution with greater flexibility of implementation. In our proposed approach, a frequency dependent digital compensation block in the baseband provides the inverse of the RF front-end system characteristics, thereby increasing the linearity of the cascaded system (end-to-end wireless chain linearity increases).

4.3.2.1. Motivation

A number of present day commercial implementations of RF components are narrowband and hence the variation in their characteristics over the operating frequency range is not of primary importance. However, the implementation of MB-OFDM UWB front end necessitates that the devices work over several GHz of frequencies. Most of the current implementations of MB-OFDM are based on Band group #1 due to the limitations determined by hardware complexity. Companies like Alereon have started providing complete RF front-end solutions covering the complete UWB band of 3.1 to 10.6 GHz. Present day wideband design capabilities provide for the in-band variations

(528 MHz) to be within the tolerance limits. Nevertheless, due to power consumption constraints there would always be significant inter-band variation (from Band 1 to Band 14). In MB-OFDM, the Local Oscillator (LO) sweeps between any two specified frequency bands within a short interval of time ranging in nanoseconds. Therefore, the baseband signal experiences different transfer characteristics if the inter-band variation is significant. This variation in-turn increases the effective non-linearity of the system. This end-to-end nonlinearity can be reduced if the inter-band variation could be reduced by intelligent frequency dependent compensation.

In this work, a compensation scheme for carrier frequency dependent non-linearity in wideband devices is developed. A low cost multitone-driven BIST technique for estimating frequency dependent diagnostic behavioral parameters of the system is developed. A multi-way compensation method for compensating distortion characteristics individually in each band of interest (528 MHz wide) is presented. Further, a novel unified compensation method for compensating for frequency dependent non-linearities using just one compensation function for all the bands is developed thereby reducing the hardware requirement of the compensation block significantly. The frequency dependent compensation methodology is shown to work in an effective manner for non-idealities induced due to process variations as well.

4.3.2.2. Compensation Technique

The overall proposed approach for compensation is shown in Figure 44. The compensation for frequency dependent variation for nominal devices as well as process-skewed instances is performed in the digital baseband domain preceding the digital to analog converter (DAC). The compensation block comprises of essentially one or more polynomial transfer functions chosen intelligently taking into consideration the tradeoff occurring between the maximum end-to-end linearity and complexity of implementation aspects. The use of inverse characteristics for the compensation works on similar lines to

that of predistortion. However, predistortion is generally a technique used in the compensation of non-linearities in narrowband systems for high power amplifiers (PA) which exhibit AM-AM and AM-PM effects. The output power of UWB system being relatively low, system implementations do not necessitate the use of a PA before signal transmission. The RF front-end (i.e., up-conversion mixer in transmitter or the LNA and down-conversion mixer in receiver) in UWB system exhibits frequency dependent non-linearity due to wide frequency range of operation. The above aspect has not been addressed in earlier literature and hence an effective, low cost methodology for transfer function estimation and compensation scheme required to reduce the effects of variations across different operating frequencies in systems such as UWB is presented.

In conventional estimation and compensation schemes, the RF transfer function characteristics are estimated using a loopback technique. This is affected by the non-linearities existing in the receiver chain such as I/Q mismatch, LNA and down conversion mixer non-idealities. In the proposed approach, the estimation of transfer function is performed using an envelope detector at the output of the transmitter. The use of the above technique allows for an estimation technique, which is independent of receiver non-linearities. A well-crafted multitone stimulus is passed through the DUT (UWB mixer) and the mixer is characterized at its center frequency for each band. The response of the envelope detector which is a low frequency signal capturing the mixer non-idealities is sampled and processed at the baseband to develop the mapping functions used to characterize the DUT. The mapping functions are developed using a regression tool called MARS during the calibration phase [81].

Two approaches for compensation have been presented in this paper. One involving the use of multiple digital compensation functions and the other is using a single digital compensation function obtained using an iterative test application and optimization algorithm. The tradeoffs involved in either of the approaches have been presented.

4.3.2.3. System Description

The schematic of the wide-band mixer designed in 0.18u technology is shown in Figure 45. The proof of concept design is based on a double balanced Gilbert cell with inductor degeneration for improved linearity over the operating frequency range of 3 to 8 GHz. A balun is used for single to double-ended conversion. The mixer provides a gain of 15 dB, NF of 8 dB at 3.5 GHz. It consumes 6.6mW power at a nominal supply voltage of 1.8V and bias of 0.8V. The above operational frequency range(between Band Group #1 to Band Group #3) is considered for the evaluation of proposed technique, though due to conflicting standards, operation in Band Group #2 is not being considered for immediate future UWB implementations. The mixer specifications do not vary significantly within a 528 MHz frequency band(less than 0.5 dB) as shown in Figure 46, whereas over the complete frequency range it exhibits significant variation as shown in Figure 47. As the variation of the mixer characteristics over the maximum bandwidth concerned in MB-OFDM (528 MHz) is significantly small (less than 0.5dB variation), it is sufficient to characterize the mixer at its center frequency of operation.

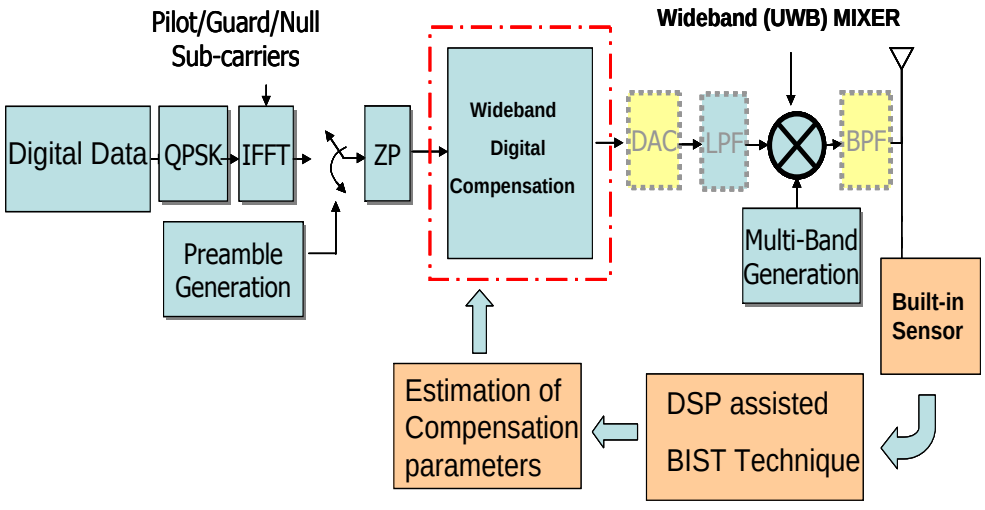


Figure 44: MB-OFDM system with digital compensation.

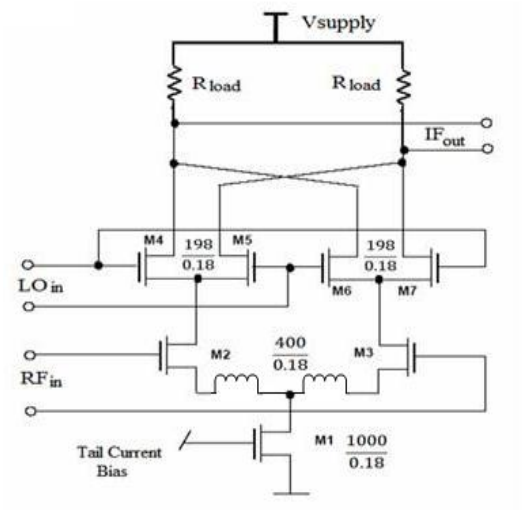


Figure 45: Schematic of wide-band mixer.

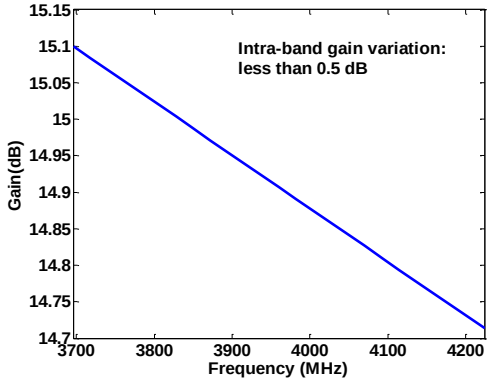


Figure 46: Variation of mixer gain over a single band of frequency (528 MHz).

The variation of the gain and third order intercept (IIP3) of the UWB mixer with frequency is shown in Figure 47 for several process instances generated by circuit-level Monte Carlo simulation of the designed mixer in Agilent Advanced Design Systems (ADS). With increasing frequency the gain rolls off as the capacitors (Cgs) in design start to get shorted whereas the output IP3 (OIP3, a linearity specification) does not exhibit any significant dependence on frequency and hence the Input IP3 (IIP3) =OIP3-gain keeps progressively increasing with frequency within the working range of the design. The work presented in [118] shows similar results for UWB devices. This change in gain

and TOI with frequency at different operating bands is compensated for using digital compensation techniques at the baseband.

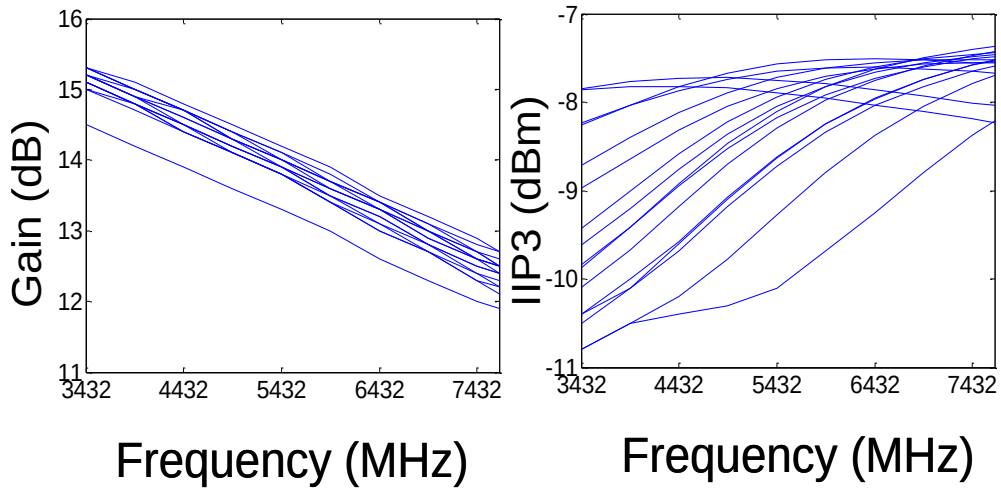


Figure 47: Mixer specifications variations over an operating frequency range.

4.3.2.4. BIST for Non-Linearity Parameter Estimation

The system is set into the test mode during the testing phase. The LO is configured in a manner to switch to all the frequencies in Band Groups #1,#2 and #3 sequentially. The BIST technique in our approach is performed using the alternative diagnostic testing [121], [122] where in contrast to conventional techniques a single signature is used to test many parameters of the DUT simultaneously using regression mapping techniques known as MARS. In this technique, a stimulus consisting of multi-tones are generated in the baseband. The multi-tone stimulus is passed through the mixer for a duration of time in which the LO sweeps through all the frequencies of interest and response at the mixer output is captured using an envelope detector. The collected signatures for different bands are then sampled and processed at the baseband. The overall block diagram representation of the BIST procedure is shown in Figure 48.

To build the regression mapping, process variations are induced into the mixer by varying the behavioral parameters of the mixer and the signatures of every instance is then sequentially obtained for each operating band. The above stimulus can be generated from the baseband during the post manufacturing tuning or during the time the system is idle in the field. The use of the envelope detector provides for a simple low cost sensing technique. This provides a method of testing for the transfer characteristics of the mixer at all its desired operating frequencies using a single multi-tone test, thereby providing a low cost efficient test solution for UWB devices.

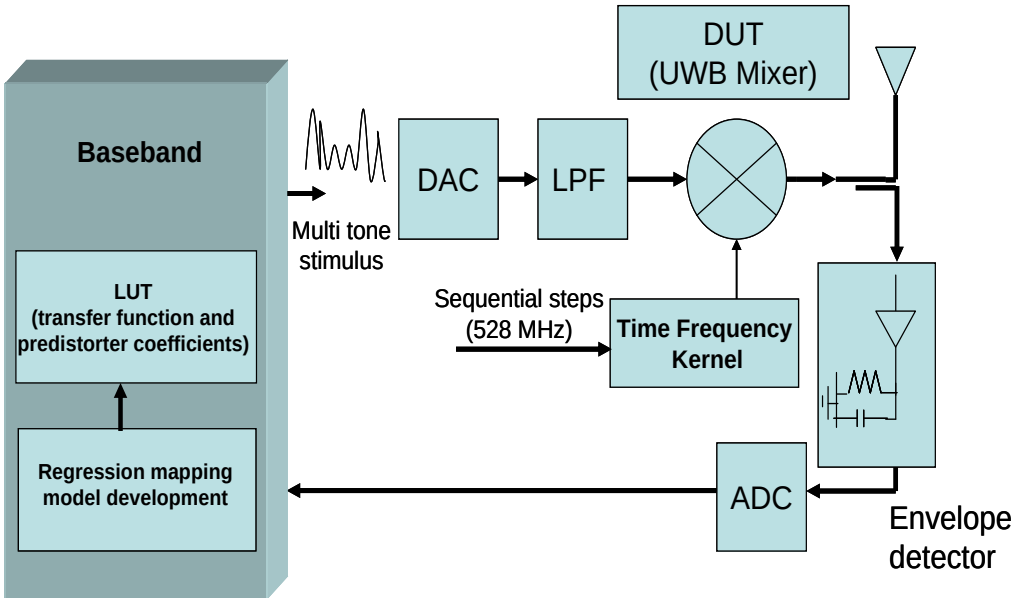


Figure 48: Block diagram of BIST for non-linearity parameter estimation.

The UWB mixer at each of its center frequency is modeled by a third order polynomial equation as shown in Equation 34.

$$y = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t), \qquad \text{Equation 34}$$

where α_0 is the DC offset, α_1 is linear gain and α_2, α_3 are the coefficients which give rise to non-idealities. The above model is implemented in Matlab in a MB-OFDM UWB system. The model characterized at its center frequency is used based on the explanation

provided in subsection A. The system model includes those blocks that are shaded (dark) in Figure 44. The model implemented performs frequency hopping in Band Groups #1, #2 and #3 ranging from 3.168 GHz to 7.920 GHz. Hence, nine different polynomial equations are used to characterize the UWB mixer. The mixer diagnostic parameters are estimated using the concept of alternative diagnostic testing.

4.3.2.5. Simulation Results - Multi-Way Compensation Technique

An ensemble of 100 devices were considered in which 70 were used to training for the development of the mapping functions and the remaining 30 were used to evaluate it. Figure 49 shows the scatter plots obtained different frequencies of operation. The x-axis lists the actual values and the y-axis shows the predicted values. The closeness of the scatter plots to the line with slope equal to one shows the accuracy in estimation of diagnostic parameters. In this section, the two proposed compensation techniques are explained and their tradeoffs are briefly outlined.

The obtained coefficients are stored in a LUT table at the baseband and the inverse of these coefficients are computed individually for each operating frequency. The compensation polynomial functions, are realized as FIR filters at the baseband. The signal processing control unit responsible for the control of the PLL during the field operation of the system based on the time frequency codes (TFC), controls the application of the compensation function. The block diagram representation of the above scheme is as shown in Figure 50. Taking into consideration the constraints of complexity and power consumption at the DSP unit, the above compensation can be applied using the two approaches:

- The inverse transfer function coefficients can be programmed into the LUT table present at the baseband unit and can be read from it to load the filter for every frequency hop made.

- The compensation coefficients are realized in the form of individual filters for each operating frequency band and the control unit coordinates a switching mechanism between the various filter blocks.

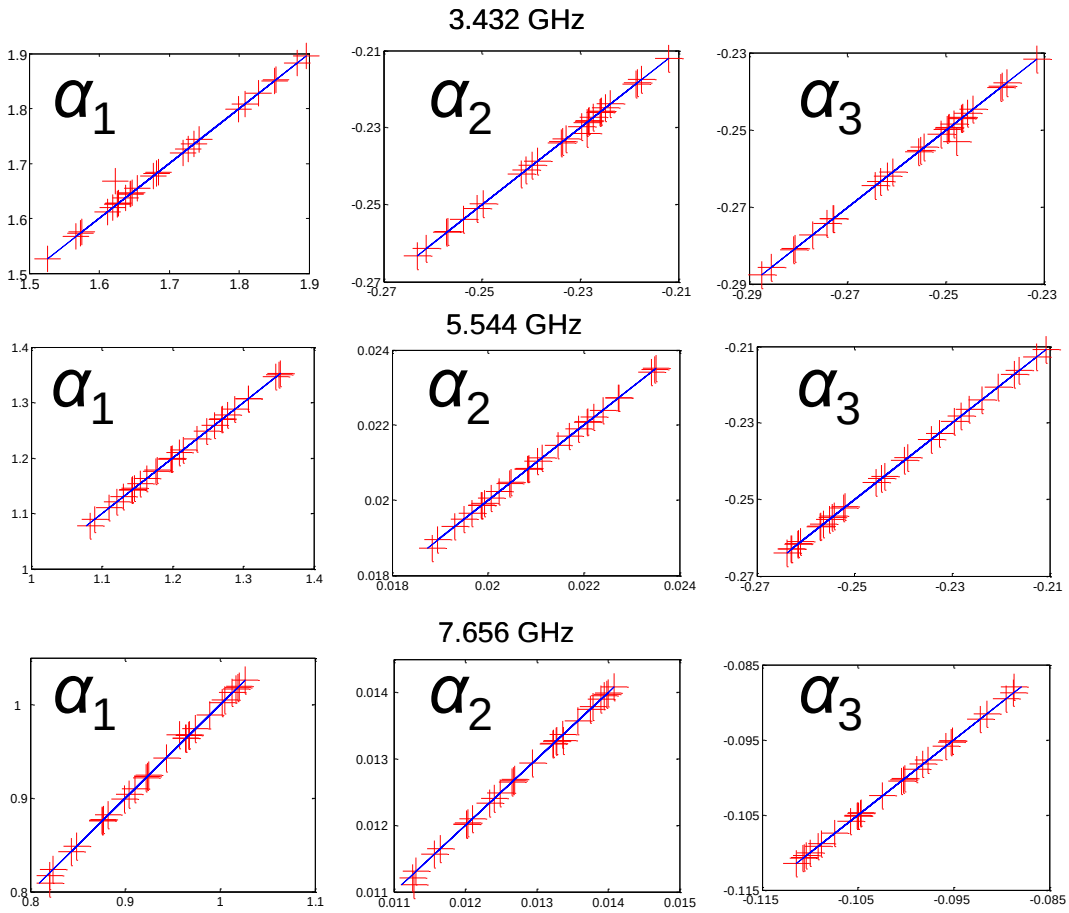


Figure 49: Behavioral parameter prediction for the three bands.

The EVM calculated with and without the above-mentioned compensation technique by downconverting the signal is shown Figure 51. The EVM calculated is the intrinsic EVM of the device and no external non-linearity contribution of any form was considered. The above technique has the drawback of intense hardware complexity and increased power consumption involved in implementation and operation in MB-OFDM UWB systems where high switching speeds at the digital and RF circuitry are a current major concern.

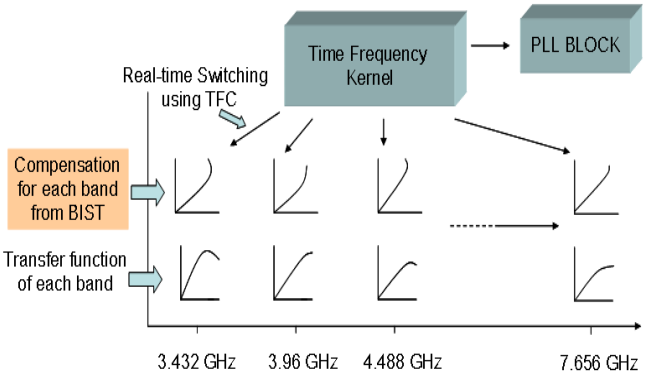


Figure 50: Multi-way compensation technique.

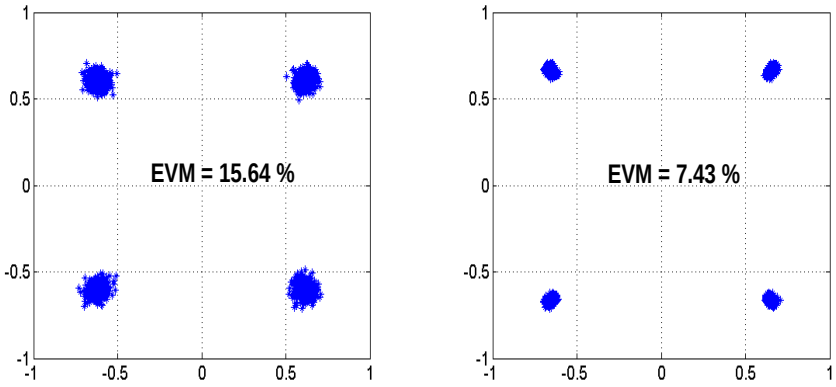


Figure 51: EVM without and with compensation (multi-way).

4.3.2.6. Simulation Results - Unified Compensation Technique

Considering the above drawback of complexity involved in implementing high speed switching digital circuits, this approach presents an intelligent compensation solution that provides a single filter based compensation technique. The initial compensation coefficients are chosen by taking the inverse of the transfer function in the mid-range of the operating frequency range. The LMS algorithm is used with iterative test application to drive the search for the optimum compensation coefficients. The LMS algorithm is optimized by considering the average of the error of a particular input sample in all the operating frequency bands. The compensation coefficients are updated after a time when

the input signal is affected by all frequency dependent transfer characteristics in the operating frequency range of the UWB mixer. The coefficients are updated according to

$$Cn, k+1 = C\,n\,, k + \eta * E\_avg * xn \qquad \text{Equation 35}$$

where $C\,n\,, k$ is the nth order coefficient, k is the iteration count of the baseband cycles. E_avg is the average of the instantaneous error samples(difference between the desired and actual response) for a particular sample collected over the desired operating frequency range and xn is the baseband test input samples. This methodology provides for a specialized search process that is capable of obtaining the tuning parameters that is best suited for all the frequencies of operation. The error progression during compensation of an instance is shown below in Figure 52, and it can be seen that convergence is achieved for 40 cycles. This technique necessitates the use of the control block for computation of LMS only during post manufacturing tuning and consumes lesser power (in real-time operation) than the former technique that consumes power throughout the operational life of the device.

The EVM calculated for a system (assuming ideal down-conversion of the transmitter output) without compensation, using multi-way and unified compensation respectively is shown in Figure 53. Both the techniques show EVM improvement (7.4 % in the case of Multi-way compensation method and 8.6 % in the case of Unified compensation method) compared to the uncompensated system (15.6%). The multi-way provides incremental improvement at the cost of higher hardware and power compared to the unified compensation technique.

4.3.2.7. Effectiveness of Digital Compensation Under Process Variation

With transistor size scaling, the detrimental effect of process variation on wideband devices is greater than their narrowband counterparts as the variation of the characteristics of the devices over different frequency bands could exhibit different

amount of deviation from its nominal. A compensation method, which can compensate for this process induced variations on top of frequency dependent variations would ensure the linear operation for wideband devices even under severe process variation. The technique developed in this research uses built in self-test (BIST) to estimate the nonlinearity parameters in post manufacturing tuning phase, and hence it captures the effect of process on each device while calculating the compensation function for multi-way or unified technique. This makes the compensation technique even more effective under process-induced variations.

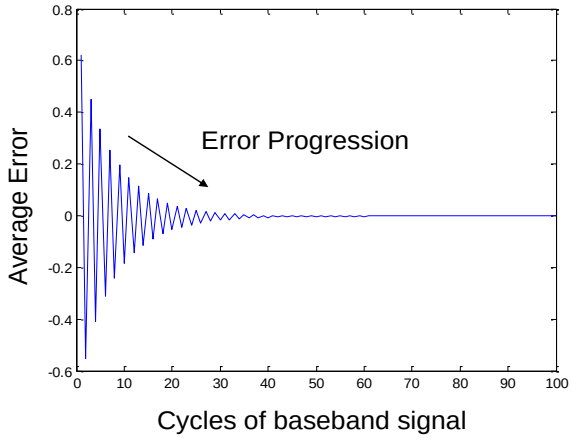


Figure 52: Error progression for optimum compensation coefficients.

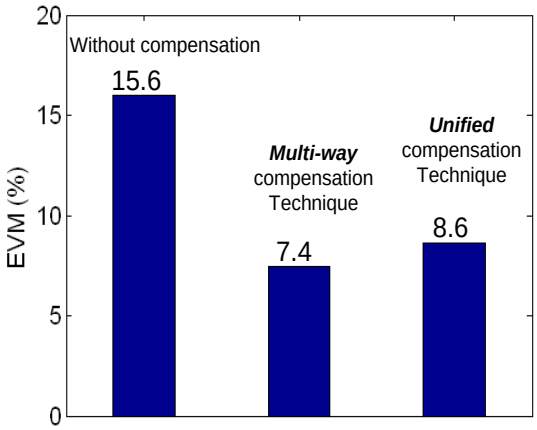


Figure 53: EVM variation due to proposed technique.

Several instances were generated by Monte Carlo simulations and the EVM for uncompensated (+) and compensated (*) case using multi way method is shown in Figure 54. It can be seen for all the cases there is significant EVM improvement, the maximum and average being 12.8% and 10.1% respectively: both of which are greater than nominal improvement of 8.2%. This shows that the proposed BIST assisted digital compensation technique is extremely effective under both process and frequency variations.

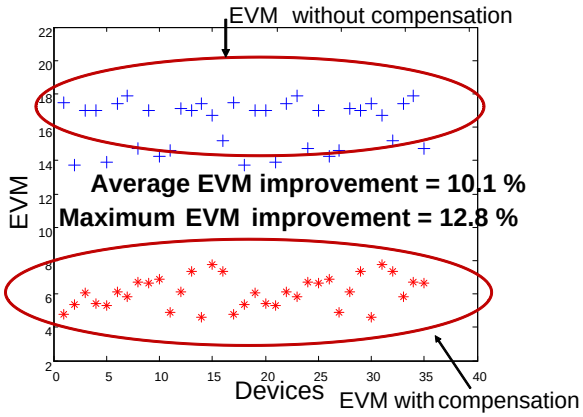


Figure 54: EVM of the devices with and without compensation.

4.3.3. Key Discussions

An efficient methodology for BIST assisted testing, estimation and compensation of wideband (MB OFDM UWB) and OFDM SISO devices has been investigated. The regression-based digital compensation schemes have been presented. In the case of OFDM system, the time taken for digital compensation is reduced using these regression techniques. In the case of UWB systems, it should be noted that analog tuning of the modules is not feasible due to the frequency hopping of the LO. As a result, digital compensation is the only method of compensation. Two different methodologies that trade-off accuracy with computational resources is presented. These techniques are not affected by the receiver chain impairments and do not suffer from long convergence times as is the case in traditional digital compensation schemes.

4.4. On-Chip Digital-Assisted RF System Tuning Methodology

The DSP-assisted system tuning techniques rely on signal processing operations performed by the system baseband processor. However, most existing RF designs incorporate the RF front end and the baseband processor (from different manufacturers) on different chips for signal integrity reasons. In such a scenario, the use of the baseband DSP for test and tuning of the integrated RF front end is difficult and can be achieved only after package integration. In the DSP-assisted technique, there needs to be a high amount of interaction between the baseband DSP and RF front end. As a result, such a technique would require a complex calibration block to interface between the baseband and front end. This requirement would lead to increased load in the baseband unit. Finally, this solution would affect the cost and time to market of the wireless units as yield improvement would only be feasible only after putting all the units together on a single package. One approach that alleviates the above problem is that of on-chip digital logic assisted tuning of mixed-signal/RF systems. In this design paradigm, on-chip digital logic is used to compensate for loss of mixed-signal/RF performance due to process variations. As the digital circuitry is relatively more robust to process variations in comparison to analog circuitry and is efficient to implement, the sensing and tuning circuitry is implemented using digital logic. The overview of the proposed technique is shown in Figure 55.

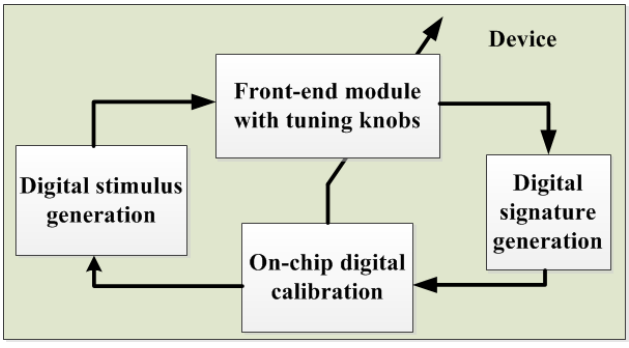


Figure 55: Conceptual diagram of the proposed methodology.

There has been significant work done in the area of digitally assisted tuning of analog circuits including PLLs, frequency synthesizers, and digital radio. However, the research above focuses primarily on testing and tuning of *individual* analog/RF/ADC specifications. To resolve this issue, this work proposes a generic tuning algorithm that can be used to test and tune multiple design specifications concurrently within a single comprehensive testing and tuning framework without using an external processor. In this work, the digital signature obtained from the device in response to an optimized stimulus is used to assess the extent of process variations in the circuit and compensate for the variations. With increased drive towards complex system on chip (SOC), system in package (SiP) and 3D integration technologies, the demand for known good die (KGD) has increased. The presented technique facilitates die-level calibration of the RF system thereby increasing the yield before integration.

The tasks of tuning the RF front ends are performed by dedicated circuitry enabled by digital signatures called "Hamming Distance Proportional" (HDP) signatures with the unique property that the hamming distance of the observed signature from the ideal is directly proportional to "how bad" the analog/RF circuit test response is under process variability effects relative to the nominal design. Due to this unique property, a hardware synthesizable sign-sign LMS-based algorithm can be used to tune the circuit for performance improvement. The objective of the optimization algorithm is to minimize the signature hamming distance through the tuning process.

In this section, the term DUT is used to refer to the device that is being tuned. In this methodology we do not perform any explicit testing to determine the specifications of the system. Under the theoretical framework of Figure 36, because the DUT response to the optimized stimulus has strong statistical correlation with the DUT specifications under the stated ranges of tuning knob values, we argue that:

- If the observed DUT response is different from the ideal for nominal values of the tuning knob values T, then one or more of the DUT's specs needs to be tuned.
- If the response of the RF DUT after tuning is identical to the expected "golden" response of the DUT to the applied test stimulus, then theoretically, the specifications of the DUT after tuning are identical to the nominal specifications of the DUT.

As a result, the measurement can be directly used for tuning the knobs of the process-skewed device, without actually predicting the specification. The signature obtained from the process-skewed device can be compared to its ideal signature obtained from a reference or golden device and the difference between them can be reduced to tune the specifications of the process-skewed instance back to its nominal values. An ideology similar to the one stated here is used in [73]. However, the algorithms running in system DSP are used to perform calibrations operations. This would require the integration of the RF module and the system DSP.

In this work, the analog response or signature obtained from the process-skewed device is processed on-chip to obtain a digital signature. This digital signature is compared to a reference digital signature to obtain a difference or error metric that is used to tune the device using a hardware driven algorithm. The principles for obtaining the digital and the difference signatures are discussed in the following sections.

4.4.1. DUT Performance Evaluation

The tasks of analysis and tuning the RF front end are performed by dedicated circuitry enabled by digital signatures, which we term as "Hamming Distance Proportional" (HDP) signatures. For DUT performance evaluation, the Hamming Distance Proportional (HDP) signature is developed from the low-frequency signature obtained at the device output. The proposed technique is showcased for a transmitter. However, the proposed

methodology can be used for a receiver as well. The architecture discussed is shown in Figure 56.

Consider the device response to the optimized stimulus is periodic with period N, where N is the number of S/H clocks (the S/H clock period is normalized to a value of "1") and the response is captured at the output of the envelope detector using a 1-bit comparator. The other input to the 1-bit comparator is a reference waveform generator that produces a linear ramp waveform with period M. The system can be designed such that across all process variations and tuning knobs settings of the device, the dynamic range of the reference waveform signal is greater than the device output response. One period of the periodic device response waveform at the sample and hold output is defined to consist of the samples $Y(t) = [Y(0), Y(1), Y(2), \dots, Y(N-1)]$ and one period of the reference ramp consists of the samples $R(t) = [R(0), R(1), R(2) \dots, R(M-1)]$ and $M \neq N$. Thus, $Y(t)$ and $R(t)$ are periodic waveforms offset in frequency. For $t > (M-1), R(t) = R(t\ mod\ M)$ and for $t > (N-1), Y(t) = Y(t\ mod\ N)$). A total number of $(M * N)$ comparisons at the comparator output are acquired. The sampling approach is similar to Vernier sampling used for measuring fractional distances. However, the manner in which the test response is interpreted is completely different.

Let X be defined as a $M \times N$ matrix with row indices going from 0 to $M-1$ and column indices going from 0 to $N-1$ (left to right). At time t, $R(t\ mod\ M)$ is compared against $Y(t\ mod\ N)$. $Y(t)$ is connected to the positive input of the comparator. At any clock comparison at the comparator, if $Y(t) > R(t)$, then the output of the comparator is high (1), else the output is low (0). The following observations can be obtained using the comparator digital output $D(t)$.

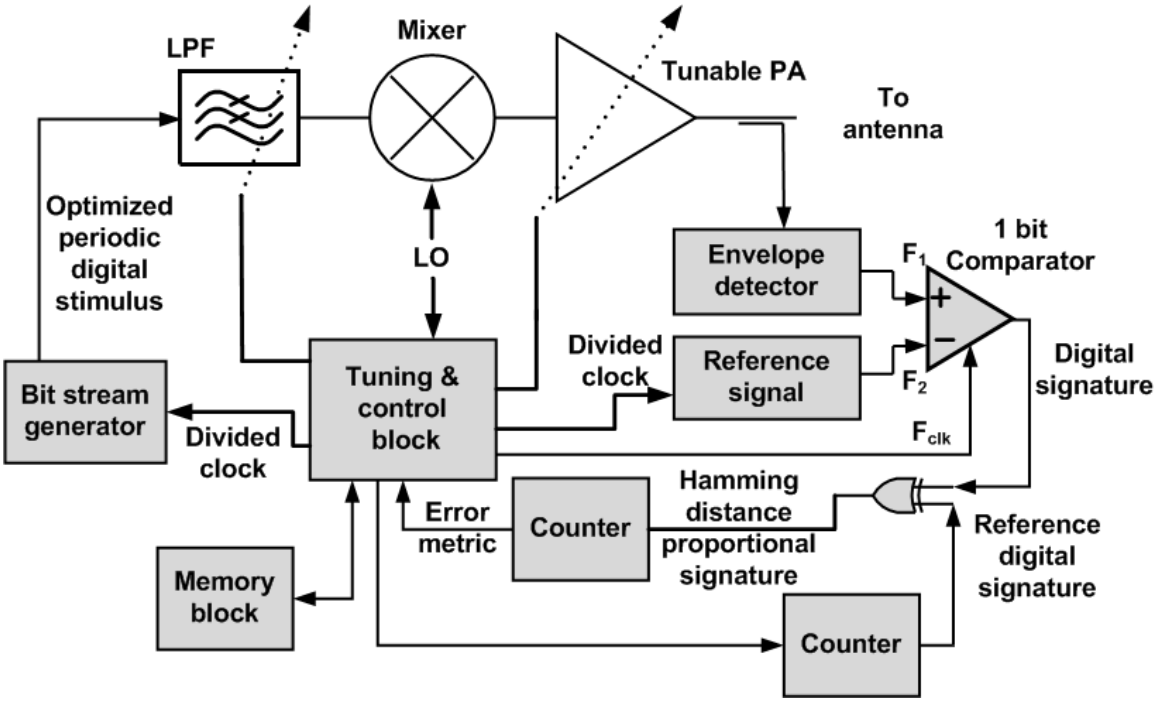


Figure 56: On-chip digitally assisted DUT architecture.

Observation 1: If each comparison of $Y(t)$ against every point of $R(t)$ obtained over $(M * N)$ comparisons are regrouped and arranged as a column of the matrix X. Then the total number of 1s in $X[I, J]$ is directly proportional to the area under the curve of Y(t).

Explanation: Consider a sample point $Y(0)$. The sample point is compared against $R(0)$, at t=0. This comparison gives the value of $X[M-1,0]$. The next comparison of $Y(0)$ is obtained at $t = N$ by comparing $Y(0) = Y(N \, mod \, N)$ against another point on $R(t)$. Subsequently, at $t = 2N$, $Y(0) = Y(2N \, mod \, N)$ is compared against a third different point on $R(t)$. Proceeding likewise, it is seen that $Y(0)$ is compared against $R(t)$ for all $0 \leq t \leq M-1$. All comparisons correspond to a single column in the matrix X. In a similar fashion, every point in $Y(t)$ is compared with every element in R. Collecting all the comparisons for each element of $Y(t)$ to form a vector and rearranging as column vector with all zeroes followed by ones and transforming to a column vector. The column vectors (total of N) so obtained for each element of $Y(t)$ can be

concatenated to form the matrix X.The number of 1s in each column represents the incremental area under the curve for the value of t, $0 \leq t \leq N-1$, corresponding to that column. The sum of the ones in each column is proportional to the amplitude at that time point and integrating over all the values of t (columns of X) gives the stated result.

This is analogous to comparison of every amplitude point of $Y(t)$ to M different levels (equivalent to a $log_2 M$ bit converter). However, in this case, the comparisons happen over a time cycles of the periodic output response waveform from the device and the reference waveform.

Example: A theoretical example that explains the concepts of Observation 1 is shown in Figure 57. At each of the time steps $t = 0$ through $t = 11$ (corresponding to N = 4 and M = 3 in observation 1), the result of the comparison between $Y(t)$ and $R(t)$ is shown in $D(t)$. In this case, for ease of illustration, it is assumed that both $R(t)$ and $Y(t)$ are piecewise ramp signals. The matrix X exhibits a ramp in its lower right corner with a high level of quantization representing the signal $Y(t)$. As can be seen the first column on the matrix has a single one, the second column has two and the third column has three.

Observation 2: If $Y(t)$ corresponds to the process-skewed response of the DUT with a time period of N, and $Y_{nom}(t)$ is the response of the nominal or reference DUT, then the absolute difference in the time-domain response between the two waveforms is given as shown below in Equation 36.

$$\sum_{t=0}^{(N)-1} |Y(t) - Y_{nom}(t)| \propto \sum_{t=0}^{(M*N)-1} XOR\big(D(t), D_{nom}(t)\big). \qquad \text{Equation 36}$$

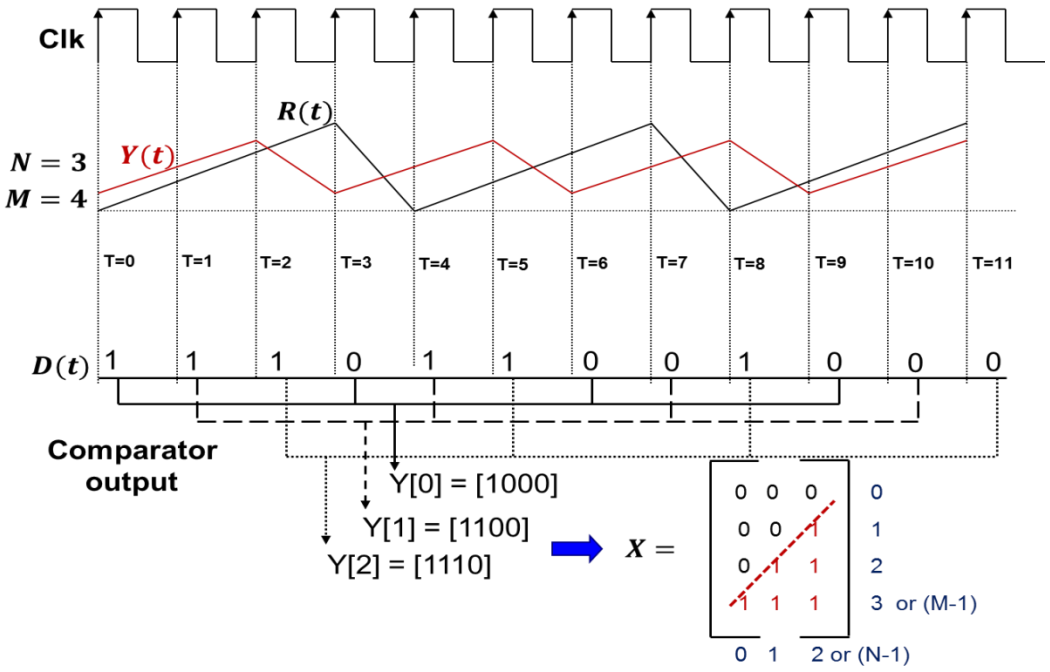


Figure 57: HDP signature example.

Equation 36 is the area between the curves $Y(t)$ and $Y_{nom}(t)$. In prior work [73], it has been shown that the area metric has a strong correlation to specification variations. This area is directly proportional to the number of ones obtained by taking the XOR of the digital response obtained at the output of the comparator for the process-skewed device $\big(D(t)\big)$ and the expected "golden" bitstream at the output of the comparator $\big(D_{nom}(t)\big)$ corresponding to the response $Y_{nom}(t)$.

The digital stream produced at the output of the XOR gate is the Hamming distance between the observed digital signature and the "golden" digital signature and is called the Hamming Distance Proportional (HDP) signature. The output is called so as the total number of 1s produced at the output of the XOR gate is the Hamming distance between the observed digital signature and the golden digital signature. By observation 2, the larger the value of difference between the two time-domain responses, the larger is the Hamming distance and vice versa. This Hamming distance value can be determined by a counter that is reset to zero at the start of the test and counts by one every time it sees a

"1" at the output of the XOR gate. The value of the counter after counting all the ones represents the hamming distance, and is called as an error metric or error count. This counter value, which determines the difference between $Y(t)$ and the golden or reference response $Y_{nom}(t)$, must be minimized in order to ensure that all the DUT specifications are tuned to their nominal values. Hence, the test is repetitively applied through different tuning steps, and the tuning algorithm is designed to minimize the counter value.

A key outcome of the above is that, on a per clock cycle basis, comparisons between $R(t)$ and $Y(t)$ (resulting in $D(t)$) and "golden" response correspondingly arranged $\left(D_{nom}(t)\right)$ are compared at the XOR gate input and the number of ones at the XOR gate output counted in the order of comparisons is directly proportional to the area of the curve between $Y(t)$ and $Y_{nom}(t)$. Due to Observation 2, the matrix $X[I,J]$ need not be constructed by the hardware. It is shown in Observation 1 only for the sake of explanation. Just counting the number of ones produced by the XOR gate output in the order that the comparisons between $R(t)$ and $Y(t)$ are performed gives the stated result and simplifies the hardware implementation.

While the explanation in the example is provided using ramp signal as the device response, the theory explained here holds good for any waveform with any number of multitones.

As shown in Observation 1, the frequency of $Y(t)$ is $f_Y = \frac{f_S}{N}$, and $R(t), f_R = \frac{f_S}{M}$. The N samples of the process-skewed device response waveform need to be compared against all the M different levels of the reference waveform. In order for such a condition to hold, the frequencies of $Y(t)$ and $R(t)$ should have a certain relationship given by the following Equation 37.

$$\frac{f_Y}{f_R} = \frac{M}{N} \; is \; irreducible \qquad \text{Equation 37}$$

Hence, according to Equation 37, the values M, N are co-prime. For any other random values of M, N the each sample of N would repeatedly be compared against a small subset of M rather than all its values over the $(M * N)$ clock comparisons cycles. Under this condition, the number of ones in a column would not be proportional to the amplitude and the integrated value to the area under the curve. Note that the example where N = 4 and M = 3 is a one case of Equation 37 as any two consecutive numbers are co-prime. The choice of M, N in our case can be based on implementation flexibility.

4.4.2. Tuning Architecture

The power amplifier and mixer circuits were designed in Advanced Design System (ADS) environment in 0.18um CMOS technology. The mixer is a Gilbert cell-based differential design. The PA is a two-stage CMOS design and is shown in Figure 58.

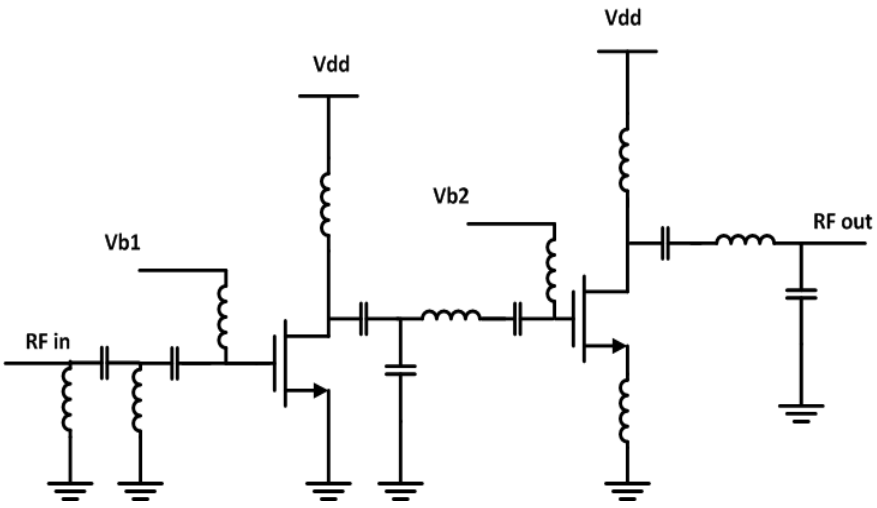


Figure 58: Two-stage PA design with tunable elements.

The tuning knobs used to tune the DUT are the bias knobs in the two stages of PA. The knobs can be implemented using CMOS switches. An envelope detector is connected to the output of the PA. The envelope detector captures the time-domain envelope of the

high frequency PA output signal, and hence contains the information of both the non-linearity and gain of the PA. The envelope detector output is fed to the comparator's positive input. The comparator output is at either logic 1 (if the device output voltage at any time is more than the voltage of the ramp signal at that time) or logic 0. The CUT test response waveform is transformed to a digital bit-stream that can be clocked at its output.

The negative input of the comparator is fed by the reference ramp signal. Other reference waveforms that can be used include simple RC output. There has been a lot of literature in the past that discusses generation of ramp signals on-chip for performing BIST for components such ADCs where the frequencies of the ramp signal range from a few KHz to MHz [123][124]. In [123], calibration schemes for correction of the slope of the ramp signal have been presented as well. In this scheme, the on-chip ramp signal precision is accurate enough to test 15-16 bit ADCs. The basic operating principle of the ramp signal generation is shown in Figure 59 . The current source drives the capacitor according to the switch signal, which can be derived from the system clock using a counter serving as a divider. The value of M can be chosen depending on the ease of implementation of the counter.

The XOR gate compares the output obtained from the comparator with the golden signal that is extracted from the nominal instance and stored in the on chip memory. However, as can be seen from Figure 57, the digital stream from t = 6 to 11 seen at the output of the comparator can be obtained by flipping the ones and zeroes of the digital stream obtained from t = 5 to 0. Hence, the entire digital stream does not need to be determined and comparisons only up to $(M * N)/2$ time points are sufficient. Further, instead of storing the actual digital stream, the number of continuous ones or zeros present in the golden digital bit stream, can be stored as a word in the memory. This technique reduces the memory requirement required to store the golden digital stream. By using a counter that counts to the value of word during which a bit (either one or zero) is

repeated in consecutive clock cycles, the bit pattern can be reproduced. Since the comparator operates at a much higher speed than the input signals, the ones and zeros in the digital stream do not alter in a frequent manner. The memory consists of words, which store the golden digitized pattern in the following format:

Number of continuous Ones	One
Number of continuous zeroes	Zero
Number of continuous Ones	One

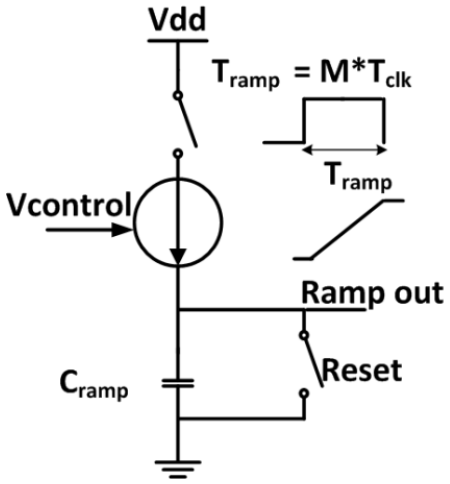


Figure 59: On-chip ramp-signal generator [123].

To achieve the divided clock frequency, the digital bit stream, and the error metric, various counters need to be implemented. A simple control unit facilitates the test stimulus initialization, golden digital bit stream generation, the ramp signal generation, as well as processing the error to perform the tuning of the PA circuit knobs.

4.4.3. Tuning Methodology

In this work, an efficient algorithm that is completely digitally synthesizable is used. The algorithm is based on the principle of Sign-Sign Least Mean Square (SS-LMS) algorithm. The SS-LMS is a variant of the LMS class of algorithms that are commonly used for adaptive filtering based digital compensation of analog/RF impairments. In this work, the principle of SS-LMS algorithm is used to tune analog knob settings of the tunable device. In our algorithm, the error metric obtained at the output of the XOR gate directs the search for the tuning knob settings of the PA. Let $e(n)$ be defined as the instantaneous error between the golden envelope or corresponding digital stream $D_{nom}(t)$ and the circuit response or corresponding digital bit stream $\big(D(t)\big)$ for a knob setting '$k_1(n)$'. The knob k_1 is then updated to its new value as shown in the equation below in Equation 38

$$k_1(n+1) = k_1(n) - \delta . ei(n) . ki(n), \qquad \text{Equation 38}$$

where the sign functions $ei(n)$ and $ki(n)$ are then defined as follows:

$$If\ e(n) - e(n-1) > 0, \qquad \text{Equation 39}$$

then $ei(n)$ = 1;

else $ei(n)$ = -1;

In the above equation, $e(n-1)$ is the error due to the $n-1$ iteration and $e(n)$ is the error in the n iteration. Similarly $ki(n)$ can be defined with respect to $k_1(n)$. The approach tunes the knobs one at a time. The sequence of the knobs can be repeated for a pre-determined number of repetitions. The selection of the order of the knob can be obtained in the characterization phase by examining the variation of the specifications surfaces with the knobs. Due to the simplicity in the implementation of the algorithm, there is no adaptive step size selection. As a result, the optimum knob solution can converge to a local optimum value. To take care of this problem, multiple starting

locations for the tuning knobs are used. The flowchart explaining the tuning algorithm iteration for each knob setting is shown in Figure 60**.** The basic idea of the algorithm is to change the direction of search in a knob setting when the error increases.

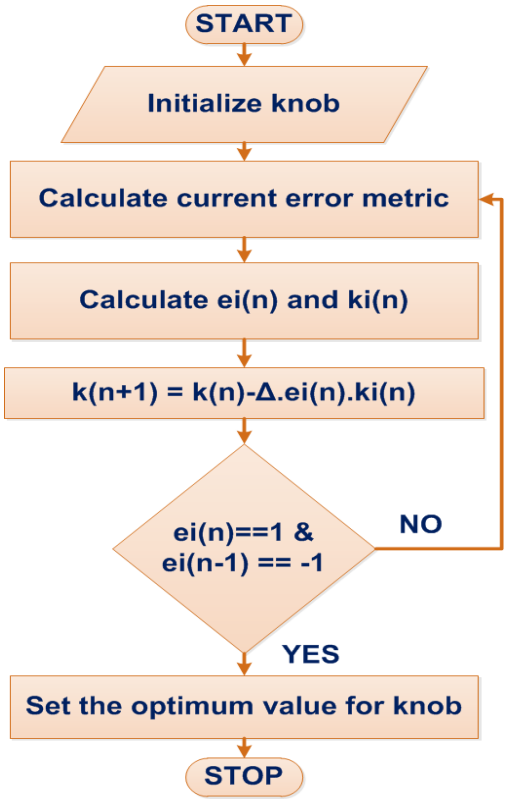


Figure 60**:** Sign-Sign LMS-based algorithm for tuning process-skewed instances**.**

4.4.4. Cost Function Formulation

As stated in Observation 2, the error metric or error count that is counted at the output of the XOR gate is directly proportional to the area difference (outcome of Observation 2), and can be used as a cost function for tuning the RF circuit "knobs" to tune the specifications of the circuit back to its nominal value. In this work, the difference in the gain and non-linearity metrics between the process-skewed device and the nominal device contribute to the total computed error count between the reference digital stream and the process-skewed digital stream. However, since multiple metrics (Gain and OIP3

in this case) are to be tuned, it is required that the cost function is sensitive to every individual specification. In general, it is observed that the contribution from variation in gain specification takes dominance over the contribution from the OIP3 specification. This can be intuitively explained as gain variation is present at all power levels and variation due to non-linearities becomes significant at relatively higher power levels when the non-linearity of the device is excited. To increase the sensitivity of the error to non-linearity of the DUT multiple error count measurements are made at different power levels. This can be explained as follows:

For ease of explanation, let us consider that the transmitter is modeled as a third order polynomial as shown below:

$$y = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t). \quad \text{Equation 40}$$

When $x(t)$ amplitude is low, the amplitude of higher order terms (corresponding to α_2, α_3) are small and hence ignored. Error $E1$ (calculated as in Observation 2) can be approximated as follows:

$$E1 = \sum_{t=0}^{(N)-1} |Y(t) - Y_{nom}(t)| = |(\alpha_1 - \alpha_{1nom})x(t)|. \quad \text{Equation 41}$$

$E1$ is the error metric obtained from the counter output that counts the number of ones at the XOR gate when the input amplitude is low. Now if amplitude of $x(t)$ is relatively higher (K times) corresponding to a higher input power level when the non-linearity characteristics of the device are excited, then the Error $E2$ is given as:

$$E2 = \sum_{t=0}^{(N)-1} |Y(t) - Y_{nom}(t)| = |(\alpha_1 - \alpha_{1nom})kx(t) + (\alpha_2 - \alpha_{2nom})k^2x^2(t) + (\alpha_3 - \alpha_{3nom})k^3x^3(t)|. \quad \text{Equation 42}$$

If the higher order terms are insignificant, then the above error E_2 can be stated as KE_1. However, at the higher input power level, the effect higher order terms are significant leading to distortion characteristics. Hence, the difference between the

Equation 41 and Equation 42 corresponds to the non-linearity in the response and is stated as

$$NLM = E_2 - KE_1. \quad \text{Equation 43}$$

The cost function (also called as the error count) is formulated as shown in Equation 44,

$$costfunction(cf) = W1 * k * |E1| + W2 * |NLM|, \quad \text{Equation 44}$$

where $W1$ and $W2$ are the weights of the two error terms. This cost formulation provides a better metric for obtaining the effect of both gain and non-linearity terms on the final error cost function. To obtain the cost function, the device is excited using a stimulus of various amplitudes. This can be done by programming the gain of the filter or by changing the setting of the variable amplifier gain, which usually precedes the mixer of a transmitter. From Observation 2, it is known that E_2, E_1 are error metrics that can be obtained from the counter value at the output of the XOR gate for different input power levels. Further, the golden digital signature obtained for the nominal device at both the power levels used for process-skewed device needs to be stored in the memory.

4.4.5. Optimized Stimulus Generation

In this work, as explained in previous section, an error count that indicates the extent of variation of the specification from its nominal value is developed. In order for this metric to be effective, it is essential that an input stimulus be chosen such that it increases the sensitivity of error metric calculated for different process skewed instances and their corresponding specification variations. To obtain such a stimulus, a genetic algorithm driven stimulus optimization algorithm that excites the effects of process variation across tuning knob settings is proposed. In the proposed stimulus generation method, a digital bit pattern is optimized by a binary elitism-based genetic algorithm (GA). The bit stream generator shown in Figure 56 produces the optimized digital stream. This generator can

be a simple state machine or an on-chip register controlled by the tuning and control unit. Each chromosome in the population of the GA represents a bit pattern and by crossover and mutation operation, new chromosomes are created in every generation. The bit pattern is first band pass filtered to get a test signal within the pass band of the device and response of the device is captured by an envelope detector. Similarly, response of the nominal device is also captured to get the golden signature. For every chromosome (bit pattern), fitness value is calculated by applying that bit pattern to a set of devices from different process corners across all tuning knob settings and then evaluating the test generation cost function. GA converges after several generations and finds out the test stimulus that gives the optimum value of the cost function. The stimulus generation cost function $(SGCF)$ is designed such that it maximizes the error count or cost function for different process instances across different tuning knobs. Hence, using Equation 44, the stimulus generation cost function can be stated as follows:

$$SGCF = maximize(\sum_{i=1}^{Z} cf). \qquad \text{Equation 45}$$

In the above equation, Z corresponds to the total number of process instances under different tuning knob conditions. In the above formulation, a limiting function is used to prevent the contribution of any particular process instance under a knob setting from dominating the overall cost function.

4.4.6. Simulation Results

The proposed method is validated on a behavioral model of a wireless RF transceiver in MATLAB. Monte-Carlo simulation was performed to generate different process-skewed instances of the PA and mixer. The process variations that were injected were the change in threshold voltages of the transistors, the length reduction factors of the transistors, oxide thickness, gate-source capacitance, channel mobility of the NMOS transistors. The input power characteristics vs. output power characteristics for different

instances were extracted, and the voltage transfer characteristics of the different instances were used to perform system-level simulations. The PA is modeled as a 5^{th} order polynomial and the mixer is modeled as a 3^{rd} order polynomial. The nominal specifications of the transmitter are shown below in Table 9. The test generation routine discussed in Section 4.4.5 is performed using a set of 50 process-skewed instances whose gain and IIP3 specifications are varied from the nominal device. The cost function progression over generations is shown in Figure 61. The optimized digital pattern obtained using test generation is shown in Figure 62. The digital bit pattern is 128 bit that is clocked using the system clock. In this simulation framework, the system clock is considered to be $f = \frac{1}{T} = 100e6$ Hzs. Such an on-chip clock frequency is quiet prevalent in today's analog/RF modules and can be generated from the system LO. The optimized digital pattern is filtered to obtain a multi-tone stimulus of period 128 T (N=128) and is shown in Figure 62. The reference ramp signal of time period of 311 T *(M=311).* This value of M is equivalent to a 8-bit ADC.

Table 9: Nominal specifications of the device.

Specification	Nominal	Bound
Tx gain	24 dB	3 dB ($\pm$1.5 dB)
Tx IIP3	-3.5 dBm	> -4.5 dBm

The envelope response obtained at the output of the PA is shown along with the ramp signal in Figure 63. For the ramp signal simulation, an offset error of 1% (mean) is modeled in different instances along with additive white Gaussian noise. The actual envelope output as well as the reconstructed signal (using Observation 1) for three process skewed different DUT instances are provided in Figure 64. The envelopes shown

here correspond to the higher input power level and the reconstructed envelopes have been scaled for comparison.

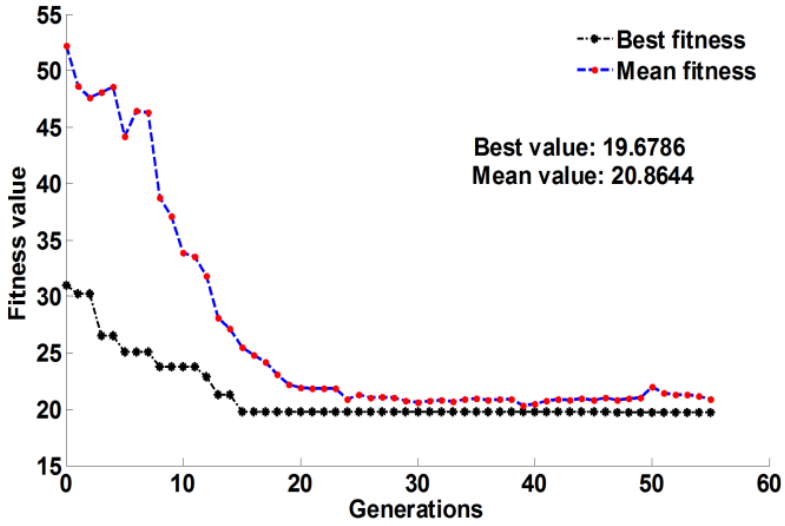


Figure 61: Test generation progression.

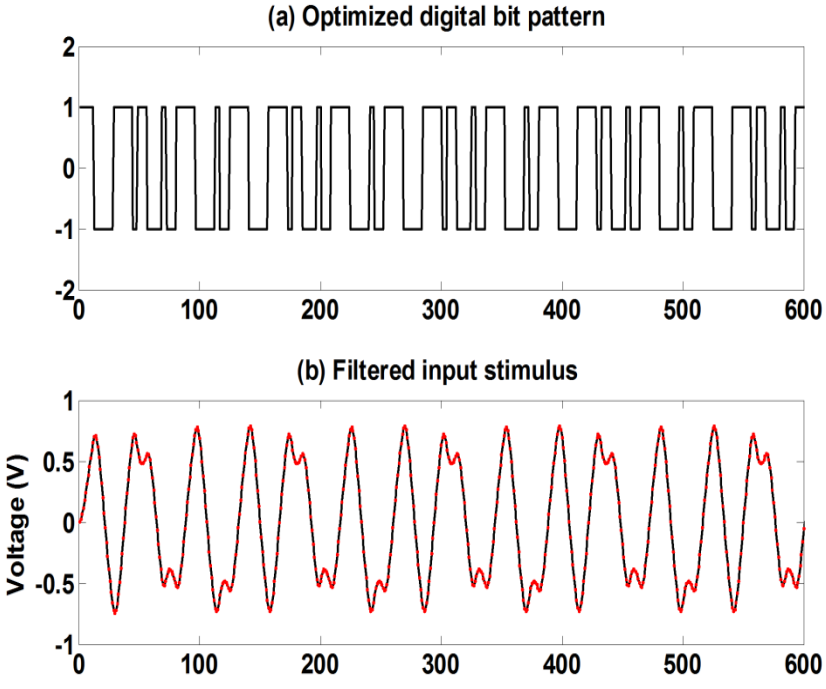


Figure 62: Optimized digital bit pattern and input stimulus.

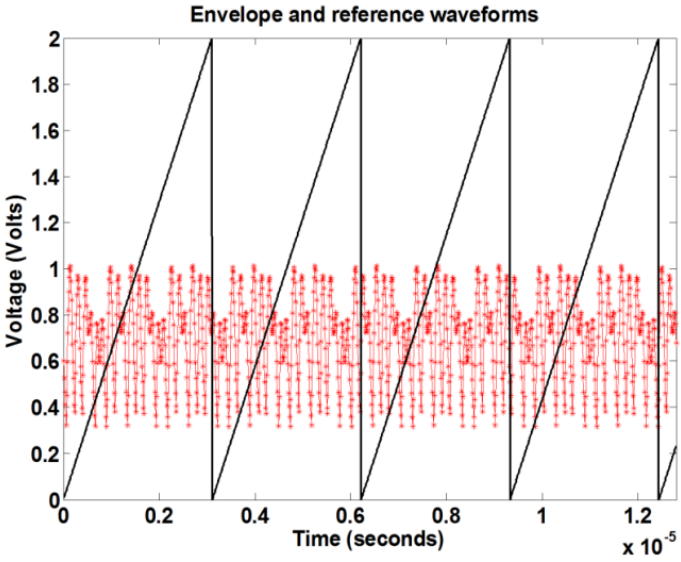


Figure 63: Comparator input signals.

The total error count, the specifications of the different process instances and the normalized root mean squared (NRMS) error also called as the reconstruction error between the reconstructed envelope (scaled) and the original envelope responses are provided in Table 10.

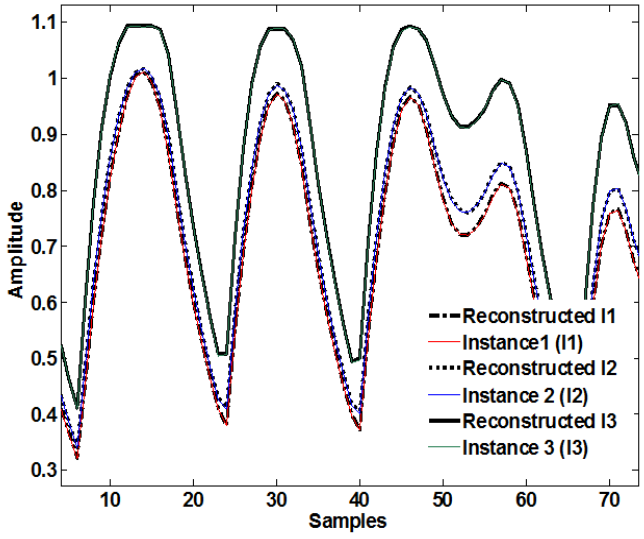


Figure 64: Validation of Observation 1.

Table 10: Error and reconstruction error for different instances.

Instance	Gain (dB)	IIP3 (dBm)	Error count	Reconstruction error (%)
1	24.05	-3.5	0	0.4
2	24.5	-3.9	274	0.39
3	27.3	-5.5	2795	0.31

This reconstruction error arises because of the quantization effect among the ramp signal levels. The quantization effect can be reduced by increasing the comparator operation speed. This would result in higher memory requirements along with more precise sample and hold circuit.

The cost function variation for a few process-skewed instances across tuning knobs is shown in Figure 65. As can be seen from the graphs, the error count reduces when both the gain approaches the nominal value of 24.0 dB and the IIP3 value is beyond its nominal bound of -4 dBm. Further, note that the error is high when only one of the specifications is within the pass bounds. It should be noted that complete decoupling of the non-linearity and gain effects is difficult to obtain by using a single error count. However, the use of two input power-levels and stimulus optimization aid in increasing the relative contribution of the two specification variations to the total error count or cost function.

For validating the tuning methodology, 275 instances were used for performing the yield simulations. The yield histograms before and after tuning are shown in Figure 66.

The initial yield of the system was 70.6%, and the final yield was determined to be 86.5% and a yield improvement of 15.9% is obtained by the proposed tuning methodology. The algorithm takes 15 iterations on an average to converge in the optimum value.

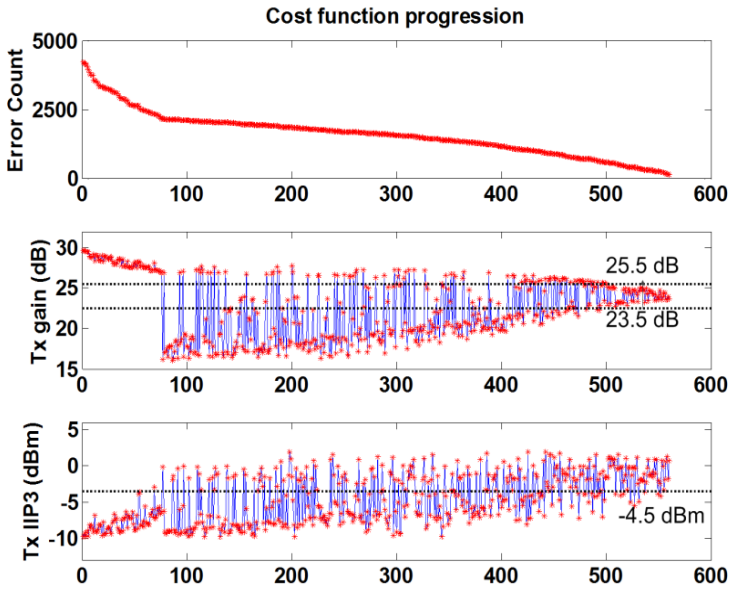


Figure 65: Cost function variation across tuning knobs and process instances.

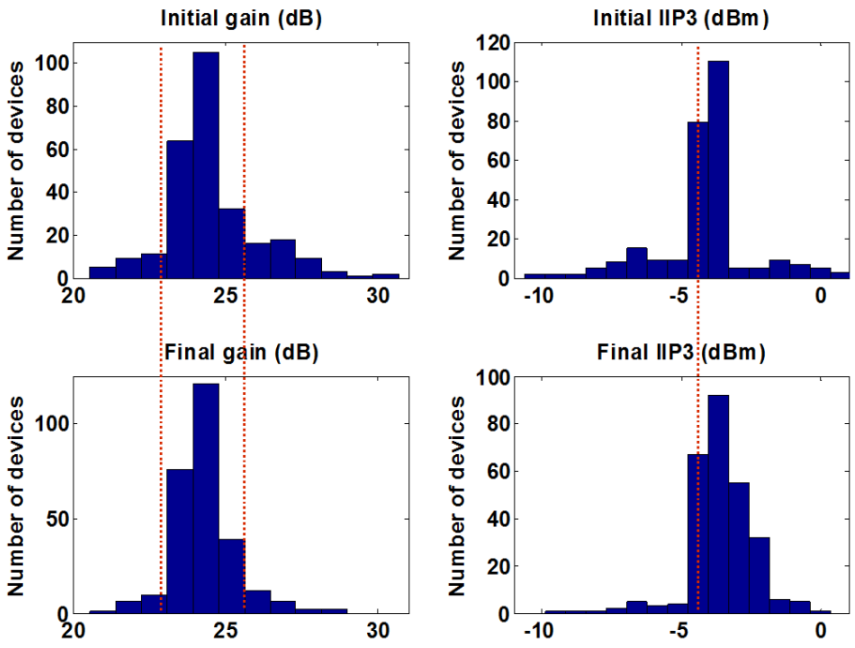


Figure 66: Yield histograms.

4.4.7. Hardware Validation

For the purpose of experimental validation of the proposed concept, an industrial RF module consisting of PA/LNA module in production with Texas Instruments is used as a platform for demonstrating the proposed concept. The entire tuning methodology is demonstrated using off the shelf components and is shown in Figure 67. The system consists of a mixer MAX2039 that is used to upconvert the baseband signal generated using Agilent 33220A function generator. The mixer is powered using a Keithley 2400. The signal is upconverted at a LO frequency of 2.2 GHz generated using the HP E8648D RF signal generator. The upconverted signal is fed to the PA that has tunable control knobs. The PA chip is fed through a RF socket on the TI tester board as shown in Figure 67. The output of the PA is downconverted and the low-frequency response is extracted using a custom-made envelope detector. The envelope detector is designed such that it has a cut-off frequency of 10 MHz. To prevent the comparator from loading the envelope detector, a buffer is used. The clocked comparator used is Hittite HMC874LC3C with an internal sample and hold. The clock signal used is 10 MHz and is provided by the Agilent 81133A pulse pattern generator. The reference signal used is a ramp signal generated using AFG320. The digital bit stream at the output of the clocked comparator is fed to the digitizer in the NI PXI 1073E DAQ chassis. The digitizer is interfaced through NI LabVIEW to the Matlab simulation environment in a PC. The tuning algorithm is implemented in Matlab as well. NI 488.2 GPIB controller is interfaced with Matlab for controlling the tuning knobs. The nominal specifications of the transmitter system are shown in Table 11.

For the transmitter setup, eight fabricated instances are used for process emulation. The tuning knobs used for compensating for process variation are the power amplifier supply (Vcc:*2.4 to 3.9 V*) and the digitally control ($Vcontrol$: 2.4 to 3.9). These knobs control the gain and OIP3 specifications of the device. During characterization, an

Agilent ESA4407B spectrum analyzer was used in a standard two-tone test measurement setup for determining the specifications of the transmitter. The variation of the specifications for a particular instance with the tuning knobs used is shown in Figure 68. As shown in Figure 67, the trigger signal generated from the DAC initiates the input stimulus to the mixer using the AFG320 as well as the reference ramp signal using the Agilent 33220A function generator. The reference clock that was used was the 10 MHz signal generated on the DAQ. Considering N = 100, a 100 KHz tone was used to create a two-tone stimulus at the input of the PA. The ramp signal was generated with a M value of 311, resulting in a frequency of 32.154 KHz. The number of comparison points collected at the output of the comparator is 31100. The amplitude of the ramp signal was selected to be 1.38 Vpp and symmetry of 70 %.The amplitude is selected such that it has a dynamic range greater than that of envelope detected PA response over tuning knob ranges and across all process instances.

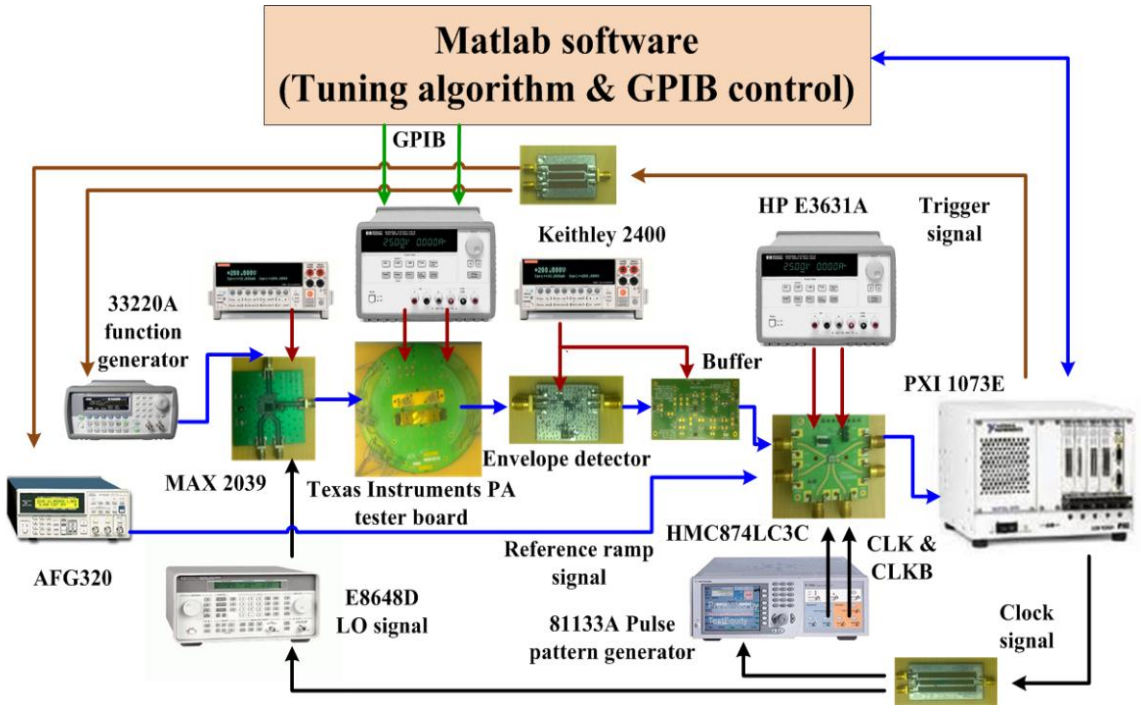


Figure 67: Hardware validation of the proposed concept.

Table 11: Nominal hardware specifications.

Specification	Nominal	Limits
Gain (dB)	8	±1
IIP3 (dBm)	14.5	> 13.5
Power (W)	0.58	

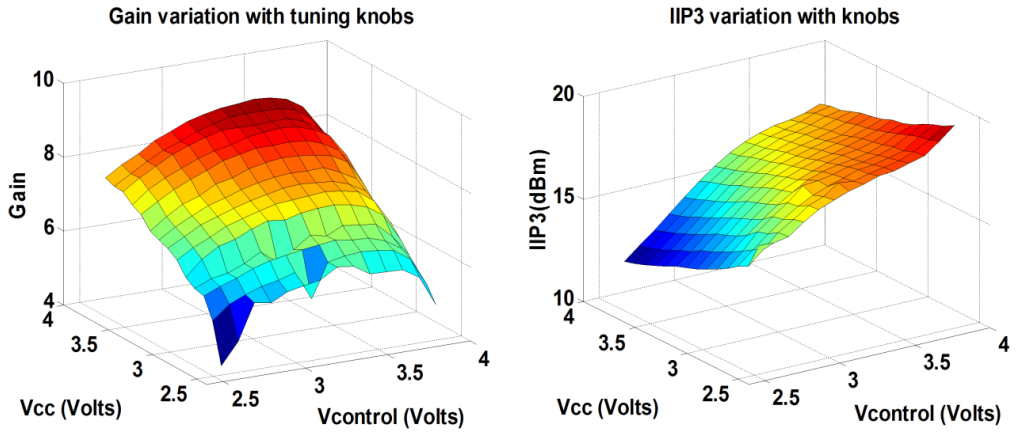


Figure 68: Specification value variation with tuning knobs.

The main clock signal is generated in the DAQ chassis and is fed as a reference clock to both the pulse pattern generator (clock signal) and the RF signal generator (LO signal). The digital bit pattern obtained at the output of the comparator is collected for 31100 clock cycles and is transferred to the PC. In the software domain, the comparison with the reference digital bit stream is performed using the XOR operation between the reference waveform and the process-skewed waveform, and the error count is calculated. the tuning knob selection based on the error count is applied using the GPIB control. Two power

levels of -5 dBm and -1 dBm are used at every tuning iteration to generate the cost function as discussed in Section 4.4.4 . The initial and the final measured values of the specification that are obtained before and after tuning are shown in Table 12. The average increase in the power consumption due to tuning in the tuned devices is 4.0 % and the average number of iterations performed to obtain the final tuned specifications is 18. Two different starting points were used in performing the tuning and the during 1st and 2nd iteration Vcc and Vcontrol were used as the first knob to be tuned respectively.

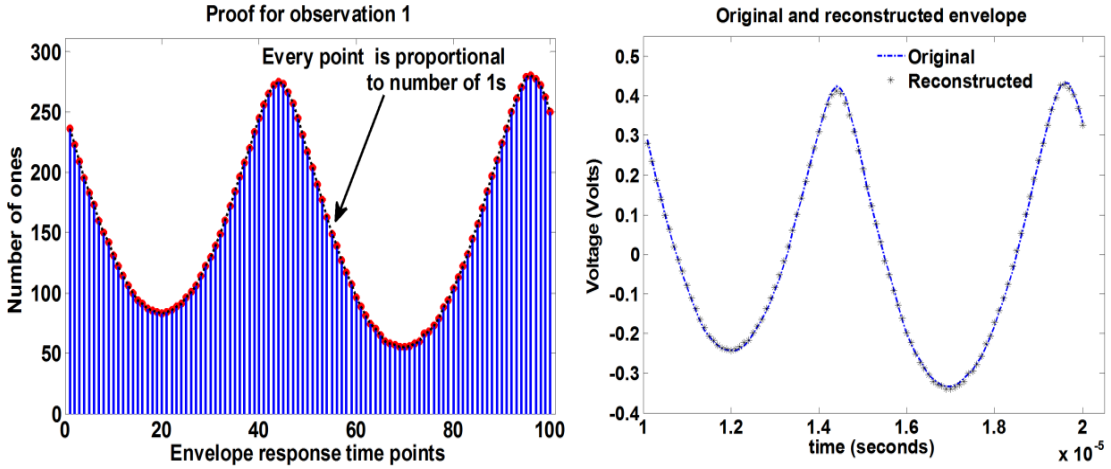


Figure 69: Reconstructed and original envelope signals.

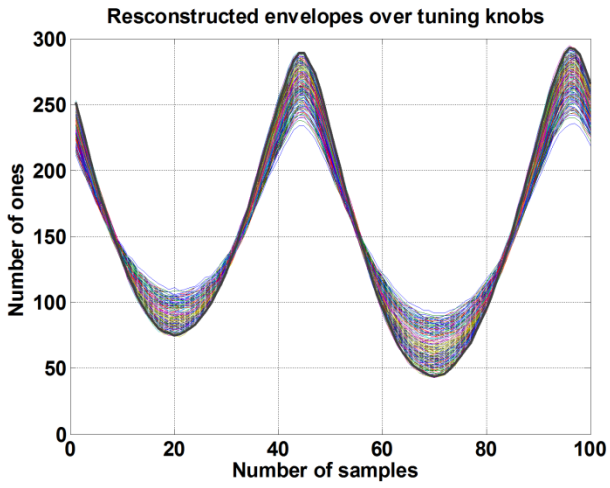


Figure 70: Reconstructed envelopes across tuning knobs.

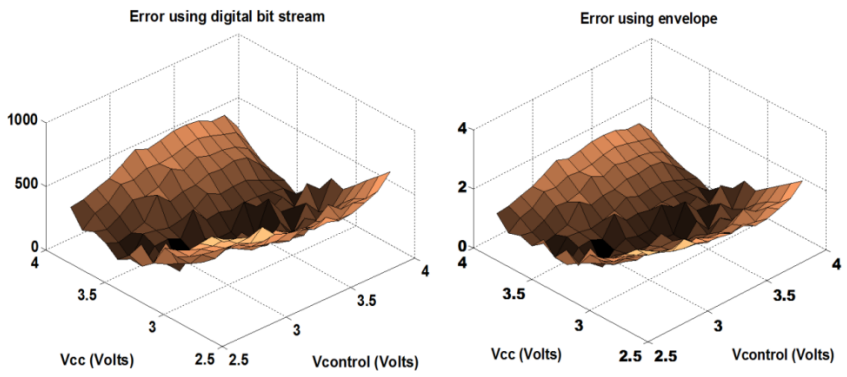


Figure 71: Error from envelope and XOR output (proof of Observation 2).

4.4.8. Key Discussions

- Test Time Overhead: The test time is a strong function of the comparator speed. Conservatively, on-chip clocked comparators of up to 0.512 GHz are feasible at 0.18um CMOS technology nodes. In general, higher speeds would require a more precise sample and hold circuitry. Hence, considering our hardware setup acquisition size of 31100 samples, the total time taken to obtain the digital stream twice would be around 0.06 ms (considering only half of the bit stream needs to be acquired for each iteration due to the symmetry in the digital bit stream). The total number of tuning iterations required per device on an average is 18. As a result, the total time for performing the tuning is in the order of couple of milliseconds (ignoring circuit-tuning time, which would be in the order of microseconds).
- Area Overhead: The area overhead is dominated by the reference signal generator due to the capacitor required for generating the ramp signal. The area for other components including the memory block (which can be implemented as a one-time programmable memory) is insignificant. The area of an on-chip ramp signal generator with calibration circuitry in 0.18um CMOS is 0.18 sq. mm [123]. A typical RF transmitter with on-chip power detector (similar to

envelope detector) and its control logic unit implemented in 0.18um CMOS is 4.8 sq. mm [125]. Hence, the additional overhead due to the reference is around 4%.

Table 12: Hardware validation of the technique.

Instance	**Initial measured value of specifications**			**Final measured value of specifications**			**Iterations**
	Gain	**IIP3**	**Power**	**Gain**	**IIP3**	**Power**	
1	10.284	17.75	0.580	9.03	20.54	0.589	19
2	5.666	10.28	0.551	8.2	15.88	0.61	9
3	8.972	11.7815	0.587	8.91	15.12	0.598	11
4	9.58	15.9635	0.572	8.63	15.57	0.602	9
5	6.26	12.993	0.545	8.1	15.8	0.581	11
6	3.092	8.9715	0.54	8.29	15.92	0.63	13
7	9.258	14.412	0.597	8.15	13.06	0.565	19
8	9.844	15.6405	0.585	8.54	13.77	0.544	17
9	7.206	9.3925	0.579	8.8	16.16	0.618	18

4.5. DSP-Assisted System Architecture with Analog and Digital Co-tuning

The methodology described in Section 4.4 involves using on-chip digital logic to perform tuning of RF circuits or systems. In this section, the system processor is used to accomplish the tuning objective. The essential components involved in DSP-assisted system architecture are as follows:

- DSP-based BIST for computing multiple performance metrics of RF circuits
- DSP-based complex control law for tuning a circuit performance parameters

The advantages of this tuning methodology is the fact that system-level tuning can be performed by using measurements (both analog and DC) from different circuits to simultaneously monitor or predict different system specifications. Further, complex calibration algorithms can be implemented in the system's baseband processor and can be run to determine the optimal knob settings that guarantee system performance criteria with minimal impact on the power of the DUT. As opposed to the prior technique, in the DSP-based techniques, explicit specifications determination can be performed at multiple frequencies at every step of the test-tune-test methodology.

4.5.1. DSP-Based BIST for Multiple Performance Metrics Estimation

It is critical to develop low cost and test time efficient BIST technique that can evaluate multiple performance metrics to ensure an overall improvement in performance of a RF device. The primary focus of past research has been to develop compensation techniques in which conventional (time-consuming) system-level tests for specific performance metrics are run iteratively, each time a tuning knob is turned. While such techniques are useful to develop compensation schemes for a specific RF performance metric, they may not fare well in a real device that suffers from several such non-idealities. Supervised learning techniques (Section 2.1), on the other hand, have the capability to estimate multiple performance metrics from a single response capture for an

optimized test stimulus. The supervised learning is performed using an optimized test stimulus that is obtained through the technique discussed in Section 4.2.1. For creation of a mapping function, each circuit knob setting for the system is considered as a separate instance of the device. This is done to facilitate the development of a non-linear regression model (running on the DSP) that can estimate multiple performance metrics across all process variations and across the entire tuning range for all possible circuit knob settings.

4.5.1.1. Control Law for Tuning a Circuit Knob

For a given RF device, we assume that multiple tuning knobs are available for modulating device performance (e.g. tuning can be performed on one or more of the passive circuit component values (capacitances, inductances and resistances) or active device biasing parameters (voltages, currents of a circuit module). In general, the number of tuning knob combinations can be quite large especially if fine-tuning is desired. Exhaustive enumeration of all possible such combinations to find the best one that restores the specifications of the device to its original values with minimal impact on power consumption can be very time consuming and expensive. Hence, a diagnostic BIST driven optimal control law that determines the best way to tune multiple knobs in an iterative manner in the shortest possible time is desirable. While the method of [123] develops such an approach, it does not allow multiple RF specs to be evaluated and therefore does not allow RF specs to be specifically traded off against each other through the tuning process. The proposed methodology is shown in Figure 72. In this work, a multi-dimensional gradient search algorithm with adaptive step size is used as an optimization engine to determine the knob settings corresponding to the "best" compensation that can be performed for each process instance. Let $K = [K1, K2, \dots, KN]$ denote the vector describing N control knobs present in the system. Depending on the approach, a cost function f is formulated to tune for multiple

specifications in the system. A central-difference-based first derivative approximation of the cost function f is used to obtain the gradient value along the steepest direction for each tuning knob as shown in Equation 46. The gradient vector for all the knobs present in the system (as shown in Equation 47) is then used to obtain the steepest descent direction S_k as shown in Equation 48. Once the direction is computed, the optimum step size (h_{opt}) for each iteration is computed using a 'golden ratio' based line-search technique. Golden ratio based adaptive step size is used to ensure faster convergence. The control knob vector K is then updated accordingly.

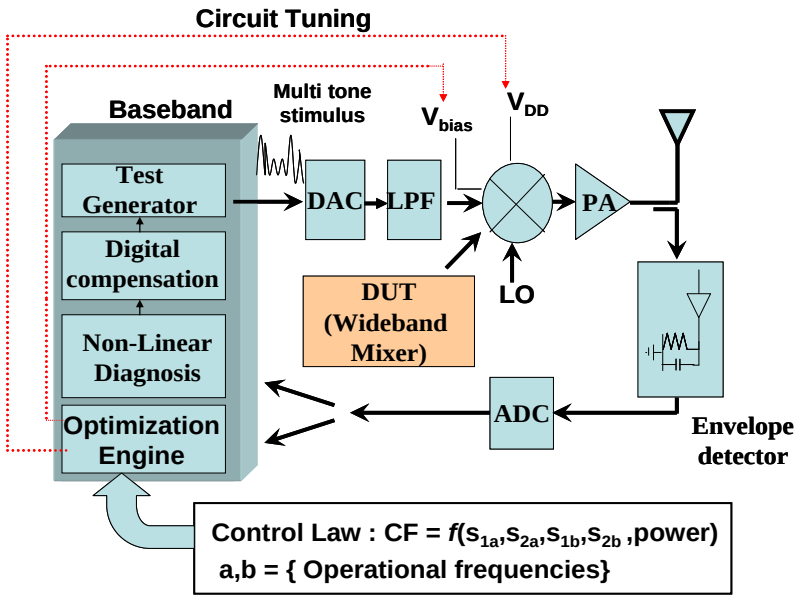


Figure 72: DSP-based tuning approach.

The overview of the tuning algorithm is shown in Figure 73. During run-time, the optimization engine optimizes the control knob voltages, iterating from a random staring point to obtain the best possible knob settings. The set of equations describing the mathematical formulation of the optimization technique is shown below,

$$\frac{\Delta f}{\Delta K_i} = \frac{f(K_i+\delta)-f(K_i-\delta)}{2\delta}, \qquad \text{Equation 46}$$

$$\frac{\Delta f}{\Delta K} = \left[\frac{\Delta f}{\Delta K_1}, \frac{\Delta f}{\Delta K_2}, \frac{\Delta f}{\Delta K_3} \ldots \frac{\Delta f}{\Delta K_n}\right], \quad \text{Equation 47}$$

$$S_k = -\frac{\frac{\Delta f}{\Delta K}}{\left|\frac{\Delta f}{\Delta K}\right|}, \quad \text{Equation 48}$$

$$h_{opt} = K + h * S_k, \quad \text{Equation 49}$$

$$K_{new} = K + h_{opt} * S_k. \quad \text{Equation 50}$$

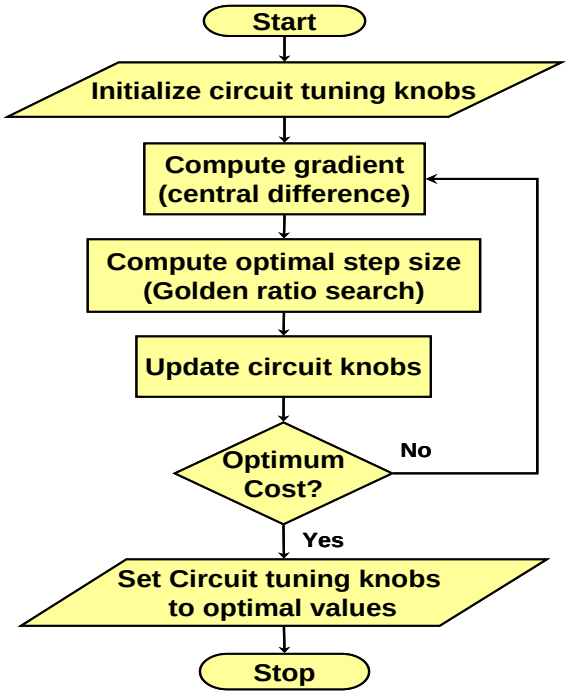


Figure 73: Gradient search with adaptive step size.

4.5.1.2. Cost Function Formulation

The key idea of the proposed post-manufacturing tuning technique is to offset any process variability to increase the overall yield of the system with minimum impact in power consumption. Hence, to facilitate the idea, the cost function is formulated as a weighted function of multiple specifications across the operating range of frequencies. The power consumption for a given knob setting is considered as an additional specification in the cost function formulation. In Equation 51, $S_{i,freq}^{j}$ denotes the 'ith'

specification for the 'jth' control knobs setting for a frequency band of operation defined by '$freq$'; $Snom_{i,freq}$ denotes the nominal specification value; $Pnom$ denotes the nominal power consumption value; $W_{i,freq}$ denotes the weight assignments for individual performance metrics and W_P denotes the weight assignment for power consumption value.

$$f(j) = \sum_{freq=1}^{L} \sum_{i=1}^{K} W_{i,freq} * \left(1 - \frac{S_{i,freq}^{j}}{Snom_{i,freq}}\right)^2 + W_P * \left(1 - \frac{P^j}{Pnom}\right)^2. \qquad \text{Equation 51}$$

By assigning appropriate weights the proposed cost function can be manipulated to achieve various trade-offs between multiple performance metrics across frequency bands of operation for overall yield improvement with minimal impact in the power consumption.

4.5.1.3. Simulation Framework

The schematic of the wide-band mixer designed in 0.18u technology for verification purposes is shown in Figure 74. The proof of concept design is based on a double balanced Gilbert cell with inductor degeneration for improved linearity over the operating frequency range of 3 to 7 GHz. A balun is used for single-to double-ended conversion. The power consumption of the device at a nominal supply voltage of 1.8V and bias of 0.8V is 6.6mW. For demonstration of the proposed methodology, two frequencies, namely, 3.4 GHz and 5.0 GHz where chosen. Process variations were induced in the above circuit with the aid of Monte Carlo simulations in Agilent Advanced Design Systems (ADS). In this work, Noise Figure (NF) is taken into consideration while determining the tuning ranges of the analog knobs based on circuit heuristics to ensure a guard banded noise performance as shown in Figure 75.

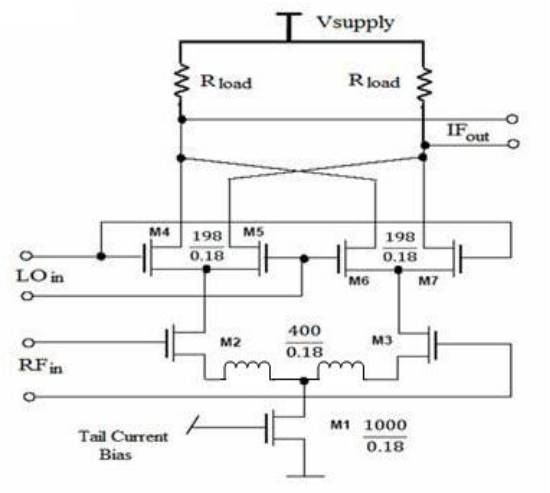


Figure 74: Schematic of the wideband mixer used as a case study.

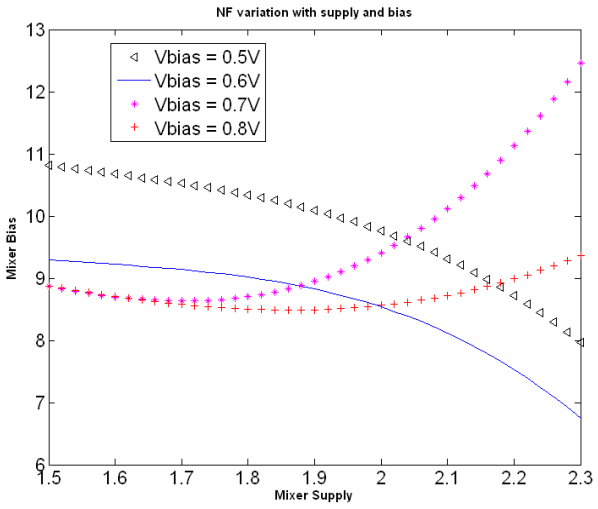


Figure 75: NF vs. circuit knobs.

The mixer supply and the mixer bias values are chosen as the circuit knobs for this piece of work. Based on the acceptable NF margins the tuning ranges were limited from 1.5 to 2.3V for the mixer supply and 0.5 to 0.8V for the mixer bias values. The NF plot across the supply and bias knob values at 3.4 GHz for a single process instance is shown in Figure 75. For the tuning ranges, the circuit knobs define a surface for all possible power values and individual specification values. The individual surfaces of one such metric (Gain) at 3.4 GHz and power are shown in Figure 76. By appropriately weighting

each of the surfaces, an optimal cost function surface can be obtained. The mixer of the RF subsystem is modeled as a non-linear transfer function followed by an ideal multiplier. The frequency mixing operation is realized by the multiplication operation as shown below as

$$y(t) = C.x_1(t).x_2(t), \qquad \text{Equation 52}$$

where C represents conversion gain of the mixer, $x_1(t)$ is the IF signal and $x_2(t)$ is the LO signal. Circuit-level simulations were performed to obtain the transfer curves for each process instance for all possible knob settings.

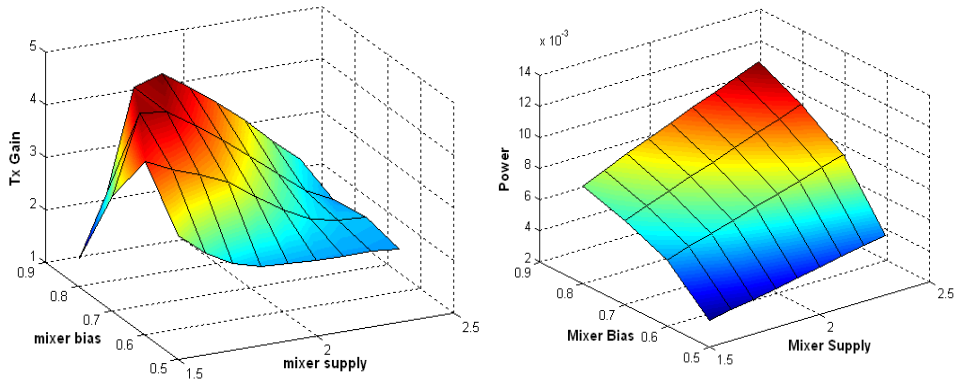


Figure 76: Gain and power surfaces for a process instance.

These transfer curves are then modeled behaviorally using polynomial coefficients up to a third order degree in Matlab. The behavioral modeling of the transfer function is expressed mathematically in Equation 53 as follows:

$$A\big(x(t)\big) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t). \qquad \text{Equation 53}$$

The tuning is performed using the optimization algorithm and regression analysis is implemented in Matlab. A diagnostic BIST procedure is developed to estimate the multiple performance metrics of the wideband mixer module across the frequency bands of interest. A set of 20 process instances across all possible knob values are used as 'training instances' to develop the non-linear regression model. The developed model is

evaluated for an ‘evaluation set’ of 10 process instances across all possible knob settings. While multiple performance metrics are evaluated, the scatter plots for one such performance metric across two frequency bands of interest (3.4 and 5.0 GHz) is shown in Figure 77. The closeness of the predicted values to the 45-degree line demonstrates a good accuracy in estimation of the performance metrics.

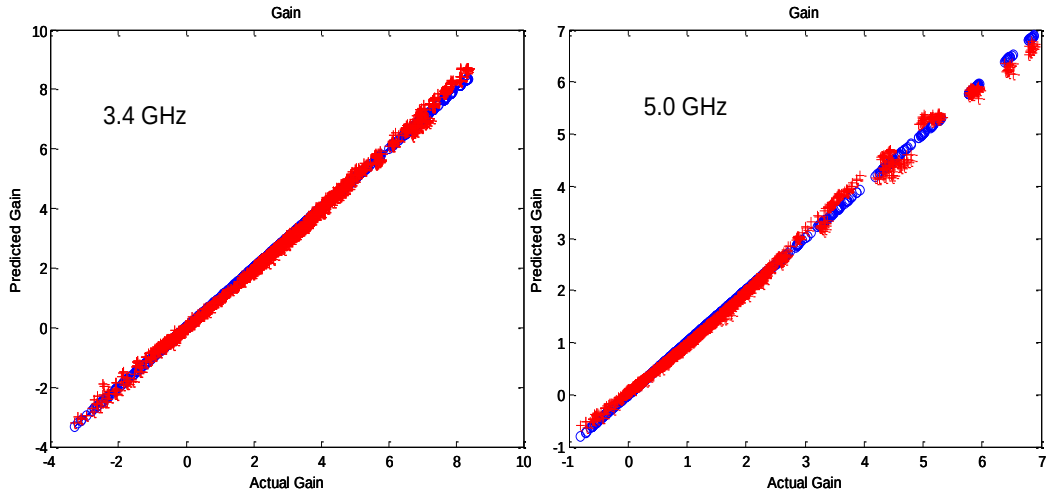


Figure 77: Gain performance metric estimation.

In this work, gain and IIP3 values are estimated by the proposed BIST technique across multiple frequencies, however other RF impairments can be estimated in a similar manner as well. In this case, study the performance of the tuning technique for various weight assignments are studied to analyze the trade-offs between multiple specs and power consumption for a boundary process instance. The nominal specs for gain, IIP3 etc. are shown along with their test bounds in Table 13. In Table 14, the performance specifications of the boundary process before performing the post-manufacture tuning technique are shown. It should be noted that these experiments are analyzed only to demonstrate the cost function formulation. During run-time, the surfaces corresponding to specs at multiple frequencies and power are weighted simultaneously for yield improvement. In experiment I, only one performance metric (gain) at 3.4 GHz is

weighted and all others are assigned to zero. The corresponding cost function is as shown in Figure 78. In experiment II, the gain specification at 3.4 GHz is weighted along with the power consumption values. In experiment III, the gain specification at 3.4 GHz and 5.0 GHz and the IIP3 specification at 3.4 GHz and 5.0 GHz along with the power consumption values are weighted. The objective of the experiment is to show the performance enhancements between various weighting assignments by analyzing the trade-offs between multiple specifications along with power considerations.

Table 13: Nominal test specifications limits.

	GAIN(dB) 3.4 GHz	**GAIN(dB) 5.0 GHz**	**IIP3(dBm) 3.4 GHz**	**IIP3(dBm) 5.0 GHz**	**Power (mW)**
Nominal	4.124	3.11	22.74	20.04	6.6
Lower bound	3.6	2.8	18.21	16.3	N/A
Upper bound	4.5	3.4	26.35	23.5	N/A

The tuning results for all the three experiments are shown in Table 15. It can be observed that experiment I results in the best possible scenario for gain enhancement at 3.4 GHz and experiment III results in an improvement in the overall performance of the system by bringing in all the specs from outside to inside the test bounds with a minimal impact in the power consumption. Further, in the proposed post-manufacturing tuning technique, digital pre-distortion is performed to improve the performance of the system.

Table 14: Performance metrics of the boundary process.

	Gain(dB) 3.4 GHz	**Gain(dB) 5.0 GHz**	**IIP3(dBm) 3.4 GHz**	**IIP3(dBm)**	**Power (mW)**
Before tuning	3.5	1.474	23.88	22.30	7.6
After tuning	3.99	2.83	23.49	21.44	9.5

In this technique, diagnostic BIST routines are used to compute the inverse correction polynomials to compensate for the static non-linearities present in the system. To demonstrate the effectiveness of the digital pre-distortion technique, the EVM constellation plots before and after tuning for QPSK modulation at 3.4 GHz and are shown in Figure 79. Such an analog, digital co-tuning technique cannot be performed in the previous on-chip digital logic assisted technique.

Table 15: Tuning trade-offs for a boundary process.

Experiment	**Gain(dB)**	**Gain(dB)**	**IIP3(dBm)**	**IIP3(dBm)**	**Power (mW)**
I	4.0203	2.013	28.92	24.74	11.2
II	4.355	2.242	31.33	25.72	10.6
III	3.99	2.83	23.49	21.44	9.5

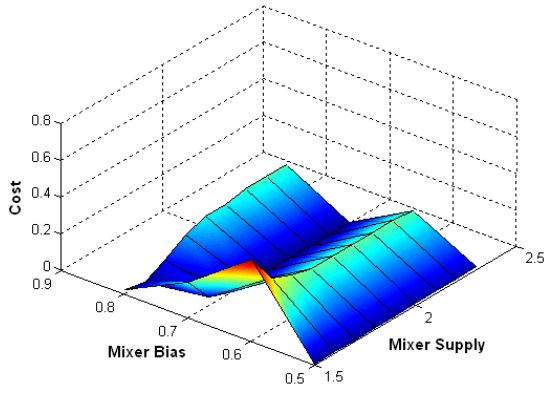


Figure 78: Cost function surface for gain at 3.4 GHz with power.

Up to 6.5% improvement was observed at 3.4 GHz and up to 6% improvement was observed at the 5.0 GHz.

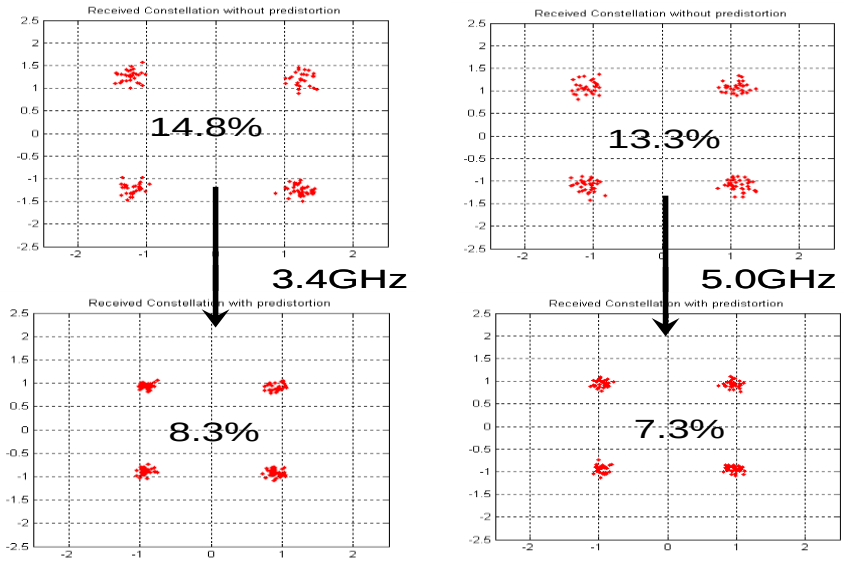


Figure 79: EVM improvements at 3.4 GHz and 5.0 GHz.

4.5.1.4. Yield Analysis

In this section, results pertaining to yield analysis of a process lot before and after the proposed post-manufacture tuning technique are presented. The results are shown in Figure 80. While the plots are shown for gain, the yield analysis is performed by taking

all other specifications such as IIP3 and NF into consideration across a wide range of operating frequency. The test bounds were fixed according to the values enumerated in Table 13. Based on the given bounds, the yield was observed to be 60.68% before tuning 93.19% after tuning. Results show significant improvement in yield by 32.51% from use of the post-manufacture tuning technique.

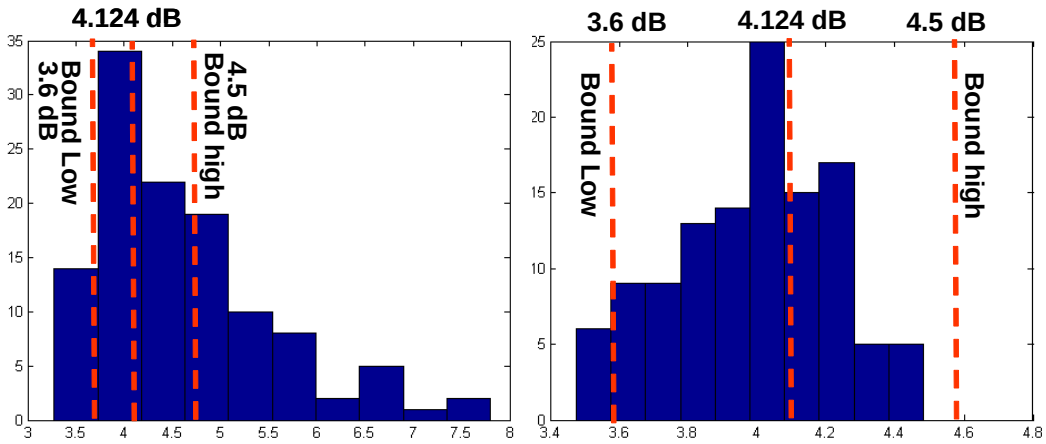


Figure 80: Yield enhancement.

4.5.1.5. Hardware Validation

In this section, the results of the proposed methodology are validated using a 2.4 GHz RF transmitter chain. The transmitter chain was constructed using mixer and PA ADL5320, which is a bias adjusting circuit and a varactor is used in the matching circuit. The variation in the varactor of the driver amplifier is used to create process variation. The driver amplifier is followed by a PA, which in turn is followed by an envelope detector. The output of the envelope detector is captured by the data acquisition (DAQ) system and the tuning is performed in the software environment in Matlab. The overall hardware setup is shown in Figure 81. The nominal as well as the final tuned values of the specification of the RF transmitter setup is shown in Table 16.

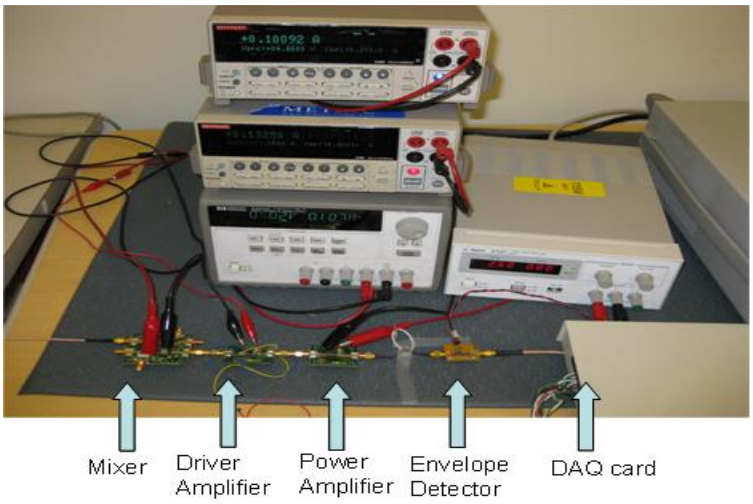


Figure 81: Hardware setup of the proposed approach.

Table 16: Hardware validation of proposed methodology.

	Gain (dB)	**IIP3(dBm)**	**Power (W)**
Nominal	3.5	10.5	1.2
Before tuning	6	8	1.26
With power consciousness	3.965	10.2290	1.402
Without power consciousness	4.45	10.04	1.652

In the first case, the weight was provided to the power term and in the latter case, tuning was performed without considering the power term. As can be seen from the

results, the power consumption of the final tuned term varies significantly in both the cases.

Further work in this direction for SISO RF systems have been performed using different tuning algorithms as well as larger number of tuning knobs in [126][127].

4.5.2. System-Level Test and Adaptive Tuning of MIMO RF Systems

As the demands for higher data rates and increased transmission range of wireless devices are ever increasing, recent standards have adapted the use of MIMO-OFDM technology that incorporates multiple RF front-end chains operating simultaneously. Current 802.11n standard allows up to eight transceiver chains to be incorporated into a single chip. With increased functionality of these wireless devices, meeting the power, performance, and area constraints has becoming a challenge. The above factors have mandated the implementation of the RF front ends in scaled nanometer nodes such as 22 nm and lower. At these nodes, the development of low-power solutions in an area-efficient manner is making the task of attaining a high yield an increasingly difficult task [55]. With the profusion of MIMO SoCs, testing each sub module of the RF chain might be infeasible (due to lack of access to internal nodes) and expensive. Hence, it is cost and time efficient to test these for systems for system-level test metrics. However, testing of these systems especially for system-level metrics is expensive in terms of both test instrumentation. The advantage in performing system-level tests is that it captures multiple impairments along with interaction effects and enables end-to-end system performance calibration. As a result, there has been increased focus towards the development of cost and time-effective system-level post-manufacturing tuning techniques that can be implemented in the production environment. Such a post-manufacturing tuning technique would involve performing iterative testing and tuning of the device.

A simple extension of a SISO testing techniques to multiple chains of a MIMO system would be feasible only if there exists independent instruments for each channel or if there were to be tested serially. While the former solution is cost hungry, the latter is expensive in terms of test time. Further, testing system-level specifications such as EVM, transmit spectral mask, adjacent channel power ratio (ACPR), etc. involve sending and acquiring real-time signals over an extended period of time and requires complex test instruments for modulation and demodulation [54]. For each iteration of post-manufacture tuning, performing these measurements using the standard setups can be time consuming. Hence, there exists a need for developing a solution for obtaining a low cost setup for performing the iterative test-tune-test methodology.

In general, yield improvements obtained by performing the above-mentioned post-manufacturing tuning methodology usually comes at the cost of increased power consumption. Besides power-budget related issues, power consumption beyond an acceptable threshold in a device can lead to reliability related failures. Hence, those devices that meet the nominal specifications requirement at the cost of more than acceptable power need to be discarded (i.e., yield loss). Alternatively, these devices can be tuned to new performance metrics at the cost of reduced power consumption, and the tuned devices can be binned according to their final tuned performance metrics. In general, such a technique will be essential for analog or RF in scaled nanometer nodes where high process variation can lead to low yields, and post-manufacturing techniques cannot guarantee a nominal performance solution within power-budget and reliability concerns. Hence, a methodology for adaptive tuning that trades off the power of devices violating power budget or reliability constraints against their performance metrics in an optimal manner is presented in this work.

4.5.2.1. Overview of MIMO-OFDM Systems

One of the primary reasons for the commercial success of MIMO systems is that they provide the flexibility to operate in both spatial multiplexing (SM) and spatial diversity (SD) modes [129][130]. SM mode is used for good channels (i.e., channels with high signal to noise ratio (SNR)) where the MIMO systems transmit multiple parallel independent data streams over multiple antennas thereby multiplying the data rate over the same frequency bandwidth. In SD mode, by utilizing space-time block coding (STBC) techniques (such as the Alamouti algorithm for 2 x 2 systems); the effective data rate is reduced to that of a SISO system [129]. However, using multiple antennas the receiver effectively combines the signal power across multiple paths leading to an increase in signal strength (diversity gain). As a result, the SNR of the signal becomes higher leading to greater range of transmission of the signal and reliability of transmission.

To enable MIMO wireless operation requires incorporation of multiple transmitters and receiver chains on-chip. Traditionally, MIMO-OFDM RF systems operate in spatial multiplexing or spatial diversity mode. They are tested in either of these modes. In spatial diversity mode, the data is transmitted in a specific sequence to maximize its robustness under multipath fading channels. On the receiver side, upon channel estimation, the transmitted symbols are reconstructed as shown in Figure 82 and the system EVM is calculated.

In SM mode, as independent data streams are sent, the error vector magnitude (EVM) of each chain is critical, and in SD mode, overall system EVM due to data coming from both the chains is critical. The system switches in these different modes depending on the channel conditions. For good channels, the system attempts to maximize data transmission and hence operates in SM mode and for bad channels, the system attempts to increase reliability or transmission SNR by operating in SD mode [129]. We term a channel as a good channel when it allows the higher data rate operation (for e.g., channel

1 allows for 64 QAM SM mode allows for higher data transfer rate than 16 QAM in SM mode which in turn allows higher data rate than QPSK SD mode). The typical variation of EVM with transmission range or channels for a 2 x 2 MIMO system arranged according to the highest data rate is shown in Figure 83 [130]. Each channel is a Rayleigh fading channel with path loss and noise level associated with it. In SM mode, the EVM refers to EVM of each transmitter and system operates using the maximum of the EVM values of the two chains.

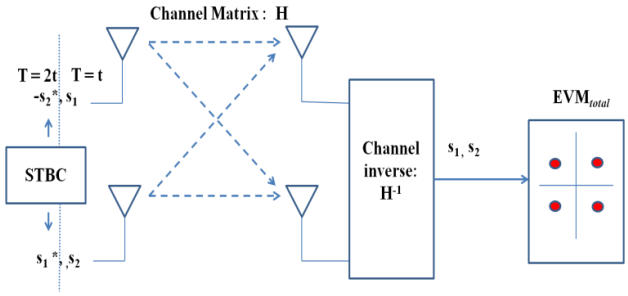


Figure 82: 2 x 2 MIMO-OFDM system in spatial diversity mode.

The reason for the system EVM to reduce when the system switches modes (from 16 QAM SM to 64QAM SD) is due to the reliability gained by switching from transmitting two independent streams (SM mode) to the redundancy of the same stream in both chains (SD mode) as shown in Figure 83.

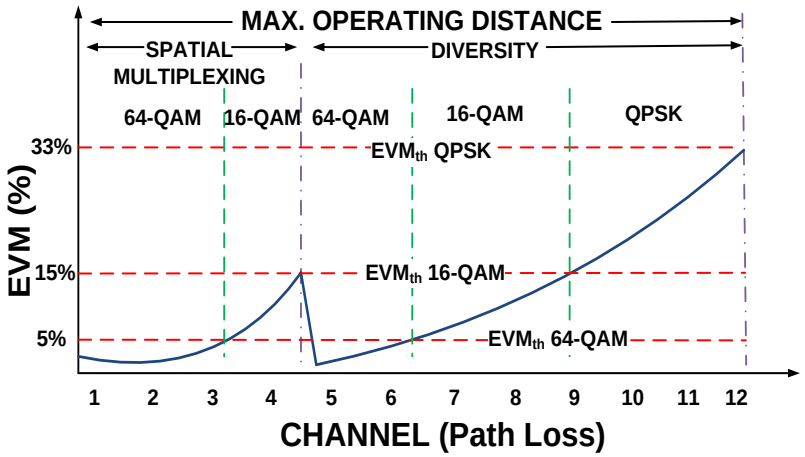


Figure 83: EVM variation in nominal transmitter device.

4.5.2.2. Test and Tune Architecture and Methodology

The testing technique developed here involves concurrent testing of the multiple chains of the RF transmitter. The testing technique developed here involves partitioning the input bandwidth of the system into regions/bands that are equal to the number of chains present in the MIMO system. The optimized multi-tone stimulus in each region (corresponding to each chain) is generated on-chip using the IFFT block present in the MIMO-OFDM baseband system as shown in Figure 84. In the test mode configuration, a set of pre-computed (complex) coefficients are loaded to the IFFT inputs to generate the multi-tone stimulus for exciting the RF front end. The values of these coefficients are determined using a genetic optimization algorithm that is run one-time offline on a computation engine. The output of the DUT responses are combined using a combiner and down converted using an envelope detector, which is then digitized and processed in the tester. Single channel acquisition is performed by combining the output of the DUT as shown in the Figure 84 thereby minimizing the need for another sensor and ADC. The output response is captured over a period of time N slots and the time averaged signal is used to obtain the static parameters of the system. The time-averaged signal is used to correlate with the static system-level specifications of the MIMO RF front ends. The variance in the signal captured over periods is used to determine the noise in the signal, which is used to correlate to the dynamic component of system-level specifications. Once the behavioral parameters of the MIMO transmitter are determined, the regressing mappings relating the static behavioral and noise parameters to the system-level specifications are developed (explained in Section 4.5.2.5). It is important to note that such a mapping relating system model parameters to the system-level specifications is developed via simulation models and there does not exist any "training" phase on actual production devices. If the specifications are beyond the normal expected values, then an iterative test-tune-test approach is developed using a "power-conscious" cost metric to

drive the tuning procedure (explained in Section 4.5.2.6). It is assumed that the MIMO RF front-end modules have "tuning knobs" that can be used to trade-off power for performance under large process variations. Finally, depending on the final device power consumption metric, an adaptive tuning strategy is investigated that trades-off the power consumption of the device with its performance (explained in Section 4.5.2.7). The steps involved in the proposed methodology are shown in Figure 85.

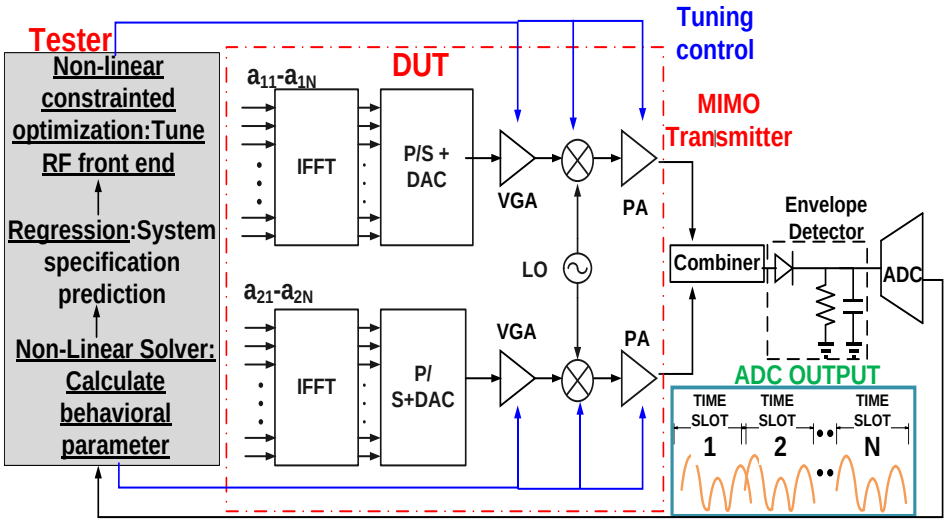


Figure 84: Test architecture and methodology.

4.5.2.3. Optimized Stimulus Generation for MIMO RF Transmitter

An analytical approach to determine the initial search space for test tones in such a way that there exists minimal interference between the responses of each chain is developed. Considering a M x M system that performs a N-point IFFT, the frequency ranges for the test tones lie between 0 and $(N-1)f$. Considering any two tones f_1 and $f_2 = f_1 + \Delta$, the third order intermodulation products exist at frequencies $f_1 - \Delta$ and $f_2 + \Delta$. Hence, two tones separated by a maximum distance Δ can have intermodulation terms that exist at a distance of Δ from them. Considering a 64 (N) point IFFT, if a sub-band of Δ tones from 0 and from $63f$ (i.e., $63f - \Delta$) are used for testing the two chains of a 2x2 RF MIMO front end (M = 2), the minimum separation between them should be

2Δ such that there is minimal interference in their intermodulation products (3rd order) when the outputs of the signals are combined. Now considering the total baseband bandwidth at the IFFT input ranges between 0 and$(N-1)f$. Hence we have $4\Delta = Nf$ and we obtain the $\Delta = Nf/4$. In general, if they are M chains, then to minimize the interference between channels, then the total bandwidth can be generalized to as shown in Equation 54.

$$(M-1) * 3\Delta + \Delta = Nf \qquad \text{Equation 54}$$

In this work, a 2 x 2 transmitter is considered. Hence, N = 64 and M = two; hence, tones considered in the first sub-band lie between 0 and $15f$ and the tones in the second sub-band lie between $48f$ and $63f$. If N = 1024, and M = two, $\Delta = 512f$. While, there does not exist any formal proof that the selection of these tones improve test accuracy, the selection of tones in this manner minimize the interference effects of one channel on other. Each sub-band Δ is used to test a transmitter chain. For initial tone spacing calculation, it is assumed that effect of higher order distortion does not affect significantly in the other sub-bands. However, similar equations with different Δ values can be obtained if the higher order distortion terms are significant. The tone selection scheme is shown in Figure 86. Considering these frequency bands as an initial point for genetic optimization, the amplitudes and phases of the frequency tones in each of the bands are optimized to estimate the behavioral model parameters of the chains of MIMO RF front end. During test generation, instances generated using Monte Carlo analysis of the circuits over different process corners as DUT instances are considered. In order for the test stimulus to be optimal for testing and tuning, the test generation needs to be performed across tuning knobs as well. In our test generation algorithm, for each process instance, at each knob setting, the model-solving-based test technique is used (see Section 2.2). A non-linear solver attempts to reduce the difference in the observed response (obtained from the circuit) and the response from the behavioral model by updating the

model parameters (see Figure 87). The overall cost function for test generation that needs to be minimized is the relative error in determination of the static behavioral parameters obtained using the non-linear solver for all the instances across tuning knobs. This cost function determines the choice of test stimuli amplitudes and phases to be used in each sub-band.

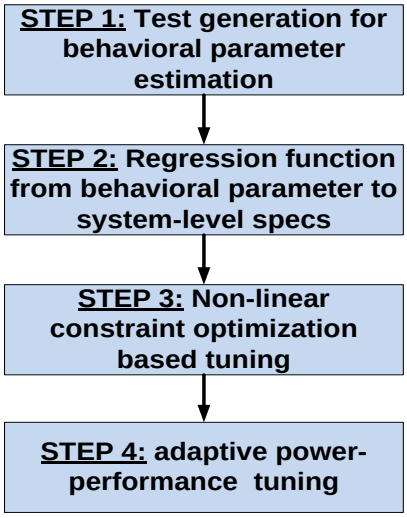


Figure 85: Proposed steps in the technique.

A bandwidth partitioning technique is also proposed in [53]. However, the authors determine specifications such as cross coupling by testing the RF chains sequentially. The methodology does not scale easily across multiple RF chains without the use of switching matrices and multiple signal sources. Further, the authors do not tie the lower-level RF specifications to higher system-level test parameters such as EVM and transmit spectral mask. In this work, the test generation/optimization approach presented in this research allows testing of a comprehensive (larger) set of specifications, scalability across diverse test specifications and most importantly, test stimulus design automation. The system-level specifications include I/Q gain or phase mismatch, DC-offset, nonlinearities, EVM, transmit mask offset, transmit center-frequency leakage, and cross-coupling effects between RF chains. For showcasing our concept, 2 x 2 transmitter is considered but the

methodology is scalable to even M x M chains (M being the number of RF transceiver paths), without the need for many signal sources or increased test time. For brevity, we do not discuss the details of genetic optimization here. The reader is recommended to [131] and discussion in Section 2.3 for more details on genetic optimization. In the case of N being large enough to prevent bandwidth partitioning, in such a case, along with bandwidth partitioning, time sequencing can be performed. For example, in the case of 16 x 16 transmitter with N = 64, then 4 transmitters can be tested at a time with 16 frequency tones. This technique would still lead to significant test savings.

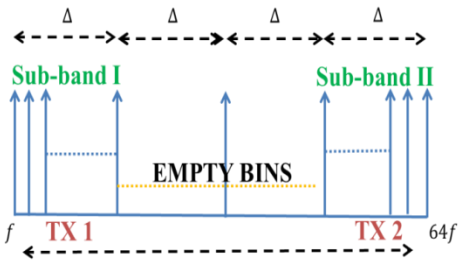


Figure 86: Selection of test tones scheme.

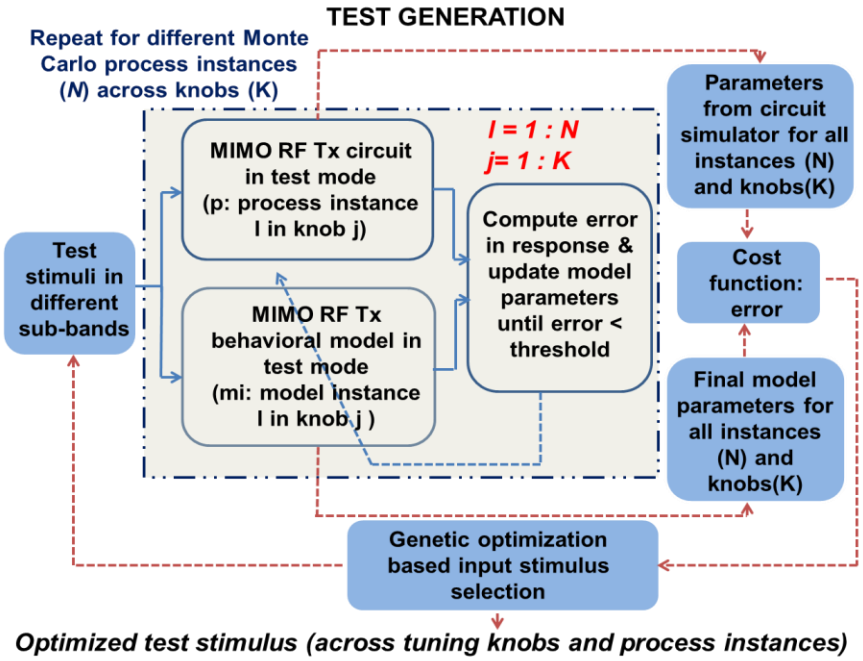


Figure 87: Flowchart describing test generation algorithm.

4.5.2.4. MIMO RF Transmitter System Impairment Modeling

In this work, a 2.4 GHz MIMO OFDM WLAN 2 x 2 transmitter platform is modeled. The modeling of the system includes impairments such as I/Q amplitude and phase mismatch, relative phase mismatch between MIMO transmitter chains, non-linearities of the front end, cross-coupling effects between the transmitter chains, DC offset in the chains, and phase noise in the LO signal. The modeled effects are shown in Figure 88. The multi-tone signal of each chain at the output of mixer is modeled as shown in Equation 55.

$$x(t) = real\left(\left((\beta + I) + jQ(1+\delta)\right) * L(t)\right) \quad \text{Equation 55}$$

$$L(t) = cos\left(wt + \theta + \emptyset(t)\right) + jsin\left(wt + \emptyset(t)\right) + \gamma_1 y_1(t) + j\gamma_1 y_1(t) + \gamma_2 y_2(t) + j\gamma_2 y_2(t) \quad \text{Equation 56}$$

$$Y(t) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x(t)^2 + \alpha_3 x(t)^3 \quad \text{Equation 57}$$

where $L(t)$ is the LO signal, β, δ and θ are the DC-offset, I/Q amplitude and phase imbalances respectively. Y_1 and Y_2 in Equation 56 are the signals at the output of the two transmitter chains. γ_1 and γ_2 are the cross-coupling coefficients. In MIMO systems, the signal at the output of each of the PA couples with other PA outputs through linear and to the LO signal through non-linear coupling. This effect is explained in [132], where it is shown that linear cross coupling is corrected to great extent through channel estimation and, hence does not have significant impact on the transceiver performance. However, non-linear coupling effects cannot be corrected for and have to be accounted for, while layout of the chip. Y_1 and Y_2 signals couple with both the I and Q LO signal paths. Hence, the signal components are represented with both real and imaginary components. $\emptyset(t)$ is the phase noise of the LO signal .The signal $x(t)$ undergoes distortion due to the mixer and power amplifier non-linearities that are each modeled as 3rd order polynomials shown in Equation 56. It is assumed the amount of coupling in both the I and Q LO signals path

is equal for each chain. A relative phase shift between the two transmitter chains arising due to asymmetry in the paths is modeled at PA output as a time delay in signal. A model such as the one presented above was used in the test generation phase.

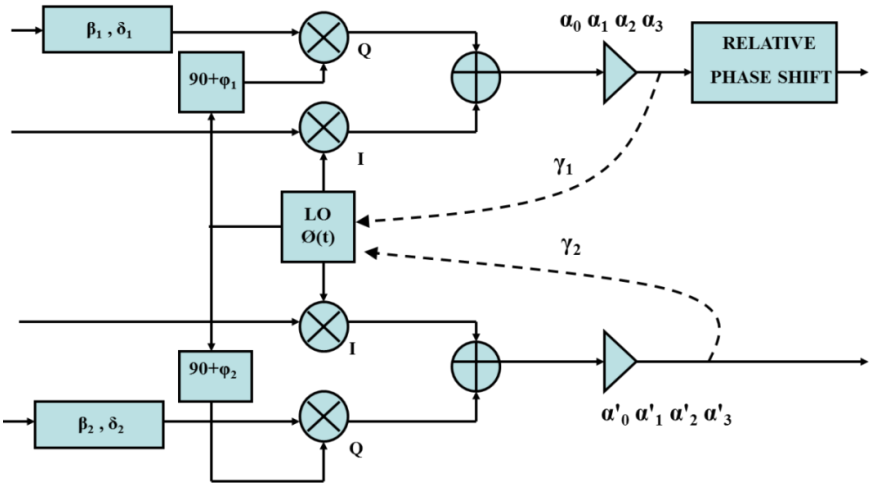


Figure 88: Model of the MIMO-RF front end impairments.

4.5.2.5. System-Level Specification Testing

The system-level parameters that are determined for the transmitter (Tx) include DC-offset in each chain, IIP3 (each Tx), I/Q mismatch (each Tx), average output power (each Tx), spectral mask measurements (each Tx), RF chain center frequency leakage, transmitter constellation error (TCE) or the EVM in SM and SD mode. These specifications can be obtained simultaneously for different chains of the transmitter without having to test them sequentially using a high-speed switching matrix (to switch between chains) as is the case in standard tests or prior literature. These system-level specifications can be classified as static and dynamic specifications. The static system specifications are specifications such as gain, IIP3, I-Q mismatch, cross coupling, spectral mask, and average output power. These are determined using a time windowed averaged output response of the captured signal. The static systems specifications are estimated by determining the values of the behavioral model parameters of the system

from the observed test results. This is performed using the model-solving-based test technique (explained in Section 2.2). We assume that the behavioral models used for analysis, accurately capture all the prominent non-idealities of the DUT. The DUT test stimulus is computed in such a way as to enable accurate determination of the DUT model parameter values from the observed DUT response. This happens when the response of the DUT to the test stimulus shows strong statistical correlation with variations in the relevant DUT model parameter values (explained in Section 4.5.2.3). Once the above test stimulus condition is satisfied, a nonlinear optimizer iteratively perturbs the behavioral model parameters in simulation until the response of the simulated model matches the observed response of the DUT.

Dynamic or noise measurements are, in general, averaged across time or frequency to yield aggregate noise measures. To this effect, the variance of the captured time-domain signal estimated at the output of the digitizer (function of dynamic non-idealities) is used in conjunction with the RF chain static specifications to build a model (regression based) in simulation environment for predicting system level EVM contribution due to noise parameters(EVM_{rand}) is determined. In this work, we utilize a tool called MARS for regression analysis. Note that the proposed use of MARS does not require the use of a set of training devices for building the regression mapping as in prior supervised learning methods [40][21][20] and can be designed directly from simulation data.

Some of the system-level specifications that are determined in this work are discussed below:

Tx Gain, IIP3, I/Q mismatch, Cross-coupling, DC-offset: Ideally, these parameters can be obtained directly by determining the behavioral parameters of the system. Limitations in the model accuracy might give rise to mismatches between the specifications obtained through behavioral parameter estimation and the actual system specifications. These limitations can be overcome by developing a standard regression between the behavioral

parameters and specifications in the simulation environment. On each device, the behavioral parameters are determined initially and then specifications can estimated in the background using a regression model.

TCE or EVM: It is known [47],[49] that system-level EVM can be computed as a function of the static and dynamic non-idealities of the RF system. This decomposition is performed in order to make the EVM computation without increasing the test time or utilizing complex test instrumentation. This technique of decomposition of EVM_{device} helps in EVM prediction more accurately using regression. The static non-idealities cause movement of the "center of gravity" of each constellation cloud while the dynamic non-idealities determine the "spread" of the constellation cloud around this "center of gravity" in the constellation map. To accomplish the computation of EVM, the EVM of the system can be modeled ([47], [49]) as shown below in Equation 58.

$$EVM_{total}{}^2 = EVM_{static}{}^2 + EVM_{rand}{}^2 \qquad \text{Equation 58}$$

Such an equation is true for a system where the noise and static parameters are statistically independent of each other. EVM_{static} represents the contribution to EVM_{total} due to all the static parameters (phase noise and thermal noise being zero), and EVM_{rand}, the contribution only due to phase noise and thermal noise (all static parameters being zero).

The EVM_{static} is determined by simulating the system using behavioral parameters representing static non-idealities of the system and developing a regression mapping between the static behavioral parameters and EVM_{static} [50]. It has been shown in the past that there exists a high amount of correlation between them and can be used to predicted using regression mappings developed one time in the simulation environment. For determining the EVM_{rand}, it is required that all the static parameters are zero. However, the actual measurement from the optimized multi-tone output obtained over the

time (N time windows) consists of noise as well as the static impairments. To determine this noise parameter, we determine the absolute error between time averaged multi-tone signal (averaged over N time windows, see Figure 84) and multi-tone signal in each time window. For simplicity, if we consider that the output is a single tone of period T, this can be mathematically represented as shown below in Equation 59.

$$Err_{jt} = \frac{1}{N}\sum_{i=1}^{N} x_i(sin(wt + \emptyset_i(t)) - x_j\, sin(wt + \emptyset_j(t)) \quad for\ j = 1,2 \ldots N \ \&\ t = 1,2, \ldots \ldots T. \qquad \text{Equation 59}$$

Considering the standard deviation over all time points, the error can be stated as shown in Equation 60.

$$std_{err} = std\left(abs\left(Err_{jt}\right)\right) \qquad \text{Equation 60}$$

The above stated noise or variance metric, along with the static behavioral parameters of each instance, is used to build a regression model to determine the EVM_{rand} during training phase and upon obtaining a suitable model, the value of EVM_{rand} can be predicted and the EVM_{total} is calculated according to Equation 58. While these EVM components cannot be isolated in the hardware, once the static and noise behavioral parameters are calculated, the regression mappings between the static behavioral parameters and the static component of EVM and the combination of static and noise parameters with the EVM_{rand} can be developed. The flowchart that outlines the procedure for determining the EVM is given in Figure 89.

Transmit spectral mask: The spectral mask of the transmitter determines the compression of the system and the resulting spectral regrowth that affects the signal fidelity in adjacent channels. A typical transmit mask is shown in Figure 90. The difference between the power spectral density (PSD) values (for e.g., $P_1 - P$) at different frequency offsets (10MHz and 20 MHz) provides the values of the spectral mask in dBr.

Once the system behavioral parameters are determined, the regression function is developed in software domain mapping the static behavioral parameters to the power levels P, P_1, P_2, P_3, P_4 and the difference in PSD levels is calculated.

Transmit central frequency leakage: Typical OFDM systems transmit null carriers at DC in order to prevent a DC-offset from saturating the receiver. However, DC offsets and non-linearities arising in the transmitter cause power to be leaked into carrier frequency of the transmit signal. The power leakage in the device is determined using a correlation function relating the leakage power and the static behavioral parameters developed in the simulation environment.

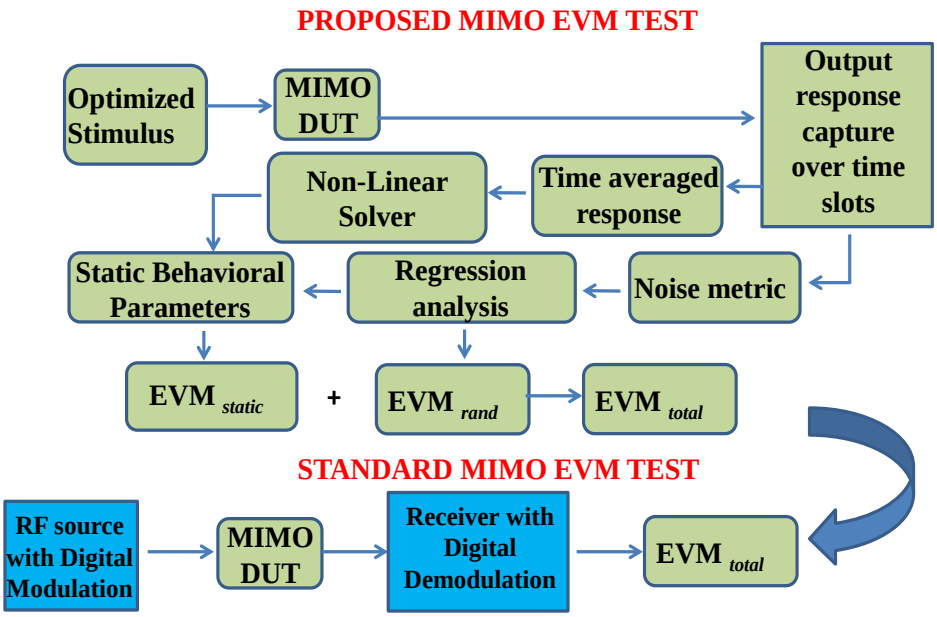


Figure 89: Overview of the proposed EVM testing methodology.

The following algorithm explains the methodology for determining the system-level specifications.

1. *Excite model and DUT with optimized stimulus.*
2. *Capture the DUT output response repeatedly over a predetermined period. Calculate time-averaged signal and variance in the time-domain signal*

3. *Use the difference between time-averaged output response and model output response to update the static behavioral parameters of the system*

4. *Correlate static parameters along with variance (noise metric) to determine system-level specifications such as* EVM_{rand}

5. *Correlate static parameters along to determine system-level specifications such as* EVM_{static} *, transmit average power etc.*

The correlation is step 4, 5 was performed through regression analysis using multivariate adaptive regression splines (MARS) [81]. The overview of the work is shown in Figure 91.

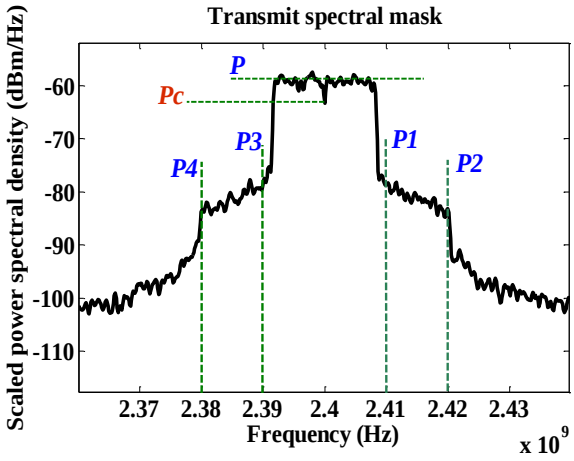


Figure 90: Typical transmit spectral mask.

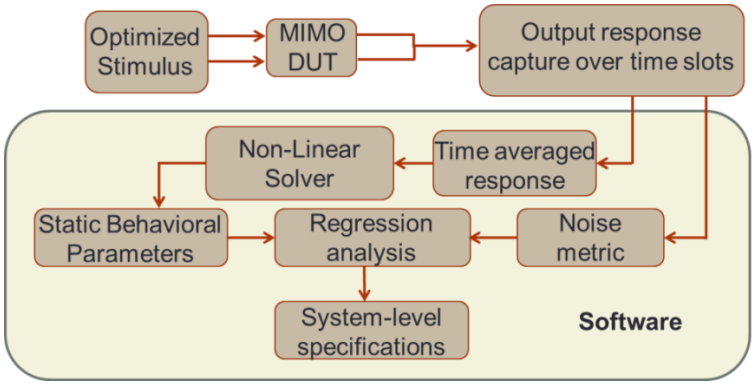


Figure 91: Overview of specification testing.

4.5.2.6. Tuning Methodology

During each iteration of the tuning methodology, the behavioral parameters of the system under the tuning knob conditions are determined and using a regression function developed across tuning knobs, the system-level specifications are predicted. In this work, a system-level constraint driven tuning solution is proposed. The goal is to find a set of tuning knob values such that a set of system-level bounds are met while consuming minimum power. Thus, for process-skewed instances, if $\{T_i\}$ be the set of all tuning knobs, then our aim is to find an optimum element of this set $\{T_{i0}\}$ such that:

$$min\big(Power(\{T_i\})\big) = \{T_{i0}\} \qquad subject\ to \qquad \text{Equation 61}$$

$$EVM_{i,SM,64QAM}(\{T_{i0}\}) < EVM_{th64QAM} \quad each\ transmitter\ i \qquad \text{Equation 62}$$

$$T_1 < Tx_{i\ o/p}(\{T_{i0}\}) < T_2 \quad for\ each\ transmitter\ i \qquad \text{Equation 63}$$

$$Tr\ Mask_{i,64QAM}(\{T_{i0}\}) < Tr \quad for\ each\ transmitter\ i \qquad \text{Equation 64}$$

where $Tr\ Mask_{i,64QAM}$, $Tx_{i\ o/p}$, $EVM_{i,SM,64QAM}$ are the transmit spectral mask at 20 MHz offset, transmitter output power and EVM in 64QAM SM mode for each transmitter respectively. T_1 and T_2 are the upper and lower bounds on the transmit output power levels respectively. While EVM ensures that the in-channel spectral distortion is controlled, the spectrum mask ensures that the adjacent channel distortion is under control. Ideally the transmit spectral mask has multiple measurements at different offsets that need to be satisfied. In this work, we consider one point while performing the tuning.

A standard interior-point algorithm with inequality constraints is used for performing optimization. The cost function to be minimized is the system power and the system specifications are formulated as inequality constraints [133]. To avoid the problem of local minima, a multi-start algorithm that selects multiple start points is used.

4.5.2.7. Adaptive Tuning Framework

As the process-skewed instances are tuned, it has been observed from prior results [73][74] that a number of tuned instances exist such that their final tuned power consumption is much higher than the nominal power consumption of the device. In a post-manufacture adaptation framework, limitation in the degree of performance tuning of the DUT comes from two factors namely power budget and reliability constraints. In a battery powered handheld devices, power consumption of the devices has to remain within a certain limit in order to achieve acceptable battery life. Hence, performance of the DUT cannot be traded-off arbitrarily with power consumption. During device production, there exists an upper limit to the acceptable power consumption of the device. Secondly, higher currents and voltages in a circuit module result in higher electrical stress on the components that reduce its lifetime. In the past, a number of techniques for dynamically varying the transmitter power consumption during its real-time operation to save power have been proposed. These techniques depend on the operating conditions encountered by the device during in field operation. So considering the above factors, an adaptive tuning methodology for production environment is proposed for devices that have a final tuned power consumption value beyond the acceptable limit.

In general, all wireless systems are built with a certain signal margin (called fade margin) that define their capability of operation under different channel conditions. Greater is the margin; better is the chance of operating in worst-case channels. The system-link budget that decides the system design can be provided as follows:

$$Received\ sensivity\ (dBm) = Tx\ Output\ power(dBm) + Gain(dB) - Loss(dB) - link\ margin(dB)\ , \quad \text{Equation 65}$$

where $Tx\ Output\ power$ is the output power of the transmitter, the $Gain$ term accounts for the transmit and receive antenna gain, and the $Loss$ term accounts for the free-path attenuation. The term $fade\ margin$ or $link\ margin$ is the margin that accounts for

different types of fading that occur in a multipath environment. The extent of $fade\ margin$ is related to the probability of coverage in the presence of fading channel. The $fade\ margin$ incorporated in a given system depends on the application of the system [134][135]. Typical $fade\ margin$ numbers are anywhere between 25 to 40 dB. The system is designed such that with the fade margin a very high probability (close to 99 % in some applications) of coverage at the cell boundary exists.

Hence, for those devices having beyond acceptable tuned power consumption, in this work, a tuning solution that attempts to reduce the number of channels of operation under SD mode (or alternatively it can be considered as reducing the probability of coverage), while trying to reduce power consumption of the device is proposed. In this way, a methodology that attempts to maintain its operation in good channel conditions while trying to trade-off performance of the system under bad channel conditions against power consumption is presented. Here, it is assumed that the probability of occurrence of a bad channel with time is lower than that of good channel [136]. While very good channels also do not occur with high probability, they still allow transmission in the highest possible data rate whenever they occur, and optimizing for such a channel is justifiable. Hence, the adaptive methodology attempts to maximize the performance of the device across its different operating conditions for a given system power constraint.

In the SD mode of operation, the combined EVM of the signal is depends on the combined performance of both the transmitters (especially the output powers of the two transmitters), which affect the power consumption of the MIMO system. Hence, if devices are tuned to worse performance specifications in SD mode, the power consumption decreases. For those devices having beyond acceptable tuned power consumption, a tuning solution that attempts to reduce the number of channels of operation under SD mode while trying to reduce the dc power consumption of the device

is proposed. These devices are binned according to their performance constraints. The optimization problem involves finding a new set of tuning knobs $\{T_{i0}{}'\}$ such that

$$min\big(Power(\{T_i\})\big) = \{T_{i0}{}'\} \qquad subject\ to \tag{Equation 66}$$

$$EVM_{i,SM,64QAM}(\{T_{i0}{}'\}) < EVM_{th64QAM} \quad each\ transmitter\ i \tag{Equation 67}$$

$$EVM_{\ 1} < EVM_{SD,QPSK}(\{T_{i0}{}'\}) < EVM_{\ 2} \quad for\ system \tag{Equation 68}$$

$$Tr\ Mask_{i,64QAM}(\{T_{i0}{}'\}) < Tr \quad each\ transmitter\ i \tag{Equation 69}$$

where $EVM_{\ 1}$ and $EVM_{\ 2}$ are the bounds for the EVM in SD mode for a bad channel. In this iteration, instead of bounds on individual transmitter outputs, a bound is placed on the system EVM in the SD mode of operation under bad channel conditions. The values of $EVM_{\ 1}$ and $EVM_{\ 2}$ used in 2nd iteration (2nd bin) are higher than the values of the nominal device (i.e., devices in 1st bin). Similarly, the values used for 3rd iteration (3rd bin) are higher than the 2nd iteration and so on. Thus, by increasing the bounds of EVM in SD mode, this methodology enables us to trade-off performance metrics against power consumption. General RF circuit knowledge dictates that this is a common trend; however, the presented algorithm attempts to find the performance combination for which the power performance is optimal. Thus, the presented technique attempts to:

- Maximize the number of channels of operation of the device for a given power constraint, i.e., reduces coverage for bad channels while attempting to keep performance in good channels.
- Find the knob combination with least power consumption that satisfies the new system constraints.

The bounds decided for the various iterations can be obtained from a system characterization study. The number of bins depends on the acceptable levels of loss of performance in the case of bad channels, which in-turn will be application dependent.

4.5.2.8. System Results

In this section, the results obtained on a 2.4 GHz 2 x 2 RF MIMO transmitter implemented in Matlab are presented. The phase noise of the system is added according to details provided in [132]. In this work, it is assumed that the phase noise in the multiple chains of the transmitter is completely correlated. The optimized multi-tone stimulus is captured for a time period equivalent to the time taken to sending 100 OFDM symbols each having 64 subcarriers. The combined output response of the transmitter chains is envelope detected, and is digitized using a 16 bit, 80 MHz digitizer. The amplitudes and phases of the frequency tones (selected by the methodology discussed in Section 4.5.2.3) are optimized using the genetic algorithm. The fitness value progression over generations is shown in Figure 92. The process variations for I/Q phase and amplitude, the relative phase shift in each RF chain and the cross-coupling parameters were generated using 3σ (where σ is the standard deviation, σ = 6-7 %variance) Gaussian variations from their nominal values. For the distortion characteristics, process instances were generated using the amplifier and mixer circuit under Monte Carlo analysis. From the power sweep of various instances, the instances were modeled in Matlab as explained in Section 4.5.2.5.

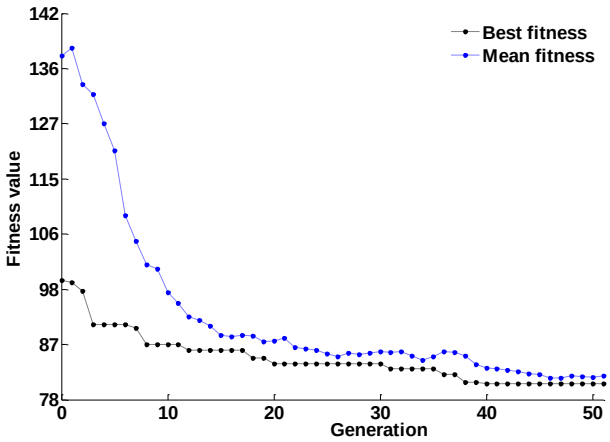


Figure 92: Cost function progression over generations.

The optimized test stimulus is used to determine the static behavioral parameters for all the process instances generated using Monte Carlo simulations. The optimized stimulus used in each of the I and Q channels of the two chains are shown in Figure 93. The envelope responses for a couple of process-skewed instances are shown in Figure 94.

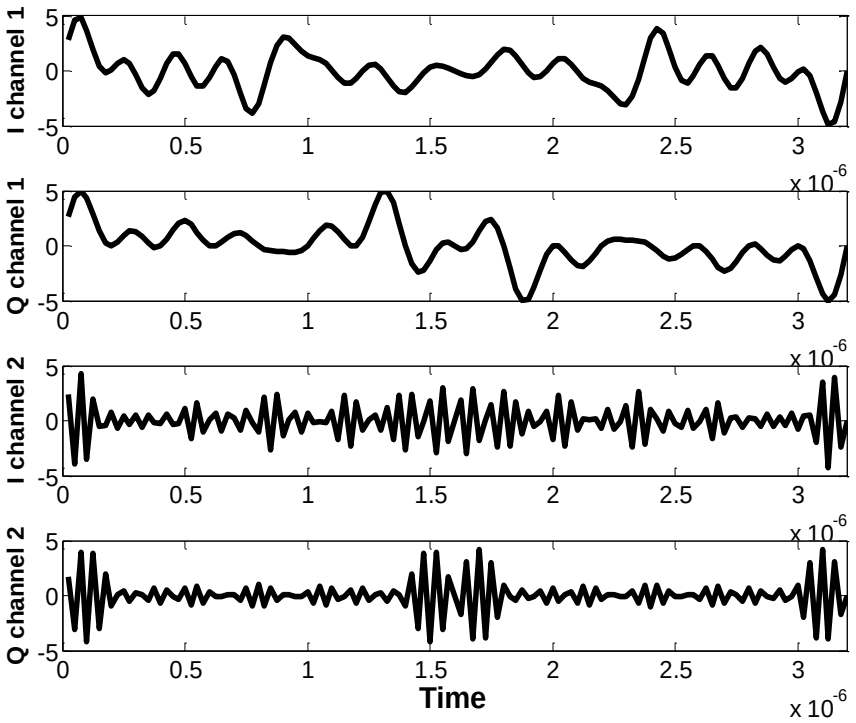


Figure 93: Optimized input stimulus.

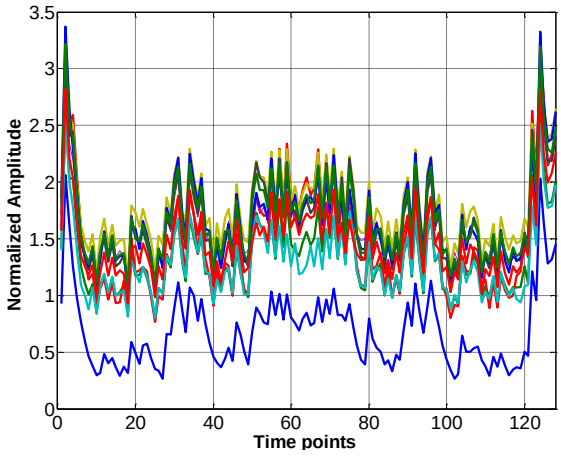


Figure 94: Envelope response of the combined signal.

The prediction plots for some of the behavioral parameters are shown in Figure 95. The X-axis of plots is the actual values and Y-axis is predicted values. 370 instances of system are tested in nominal system settings. The behavioral parameters are estimated for these instances. The worst-case mean absolute percentage error obtained for all the static parameters was less than 3%. The nominal test condition for EVM is in SM mode for a good channel. In this condition, the contribution of EVM is completely from the device and negligible from the channel characteristics. The decomposition of the EVM_{total} into EVM_{static} and EVM_{rand} is validated for randomly chosen 200 instances and is shown in Figure 96.

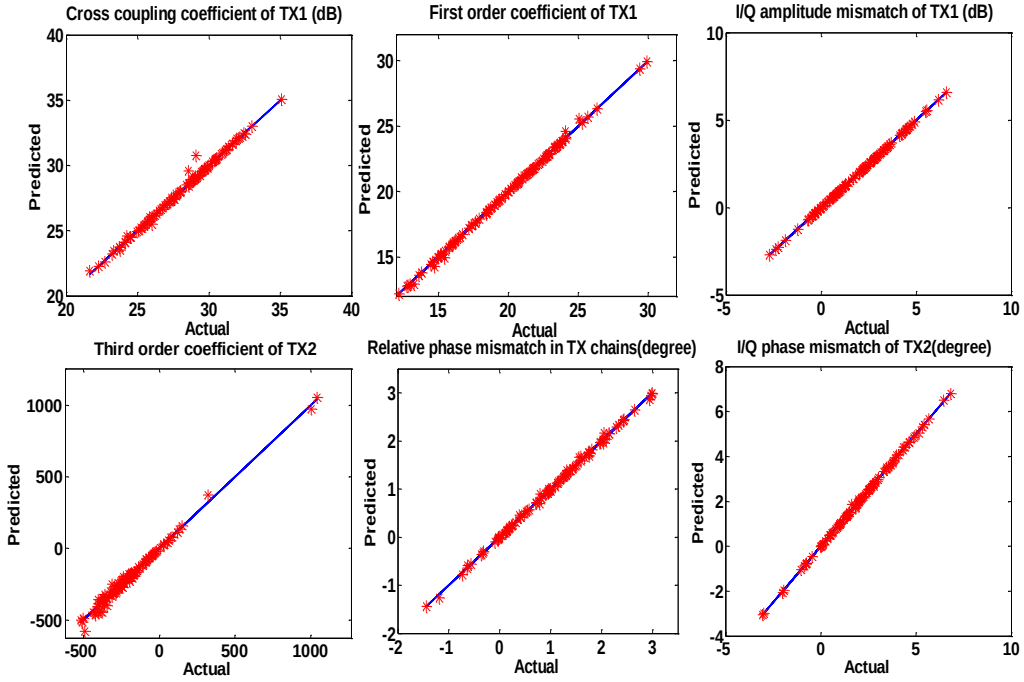


Figure 95: Prediction plots of static parameters.

Using Equation 59 and Equation 60, the std_{err} for all the 370 process instances were calculated from captured the time domain response. For the purpose of estimation of transmitter EVM_{total}, 220 instances are used to build the regressing maps and 150 instances are used to evaluate it. The prediction plot for EVM_{rand} in SM mode for 150 random instances at nominal system settings is shown in Figure 97.

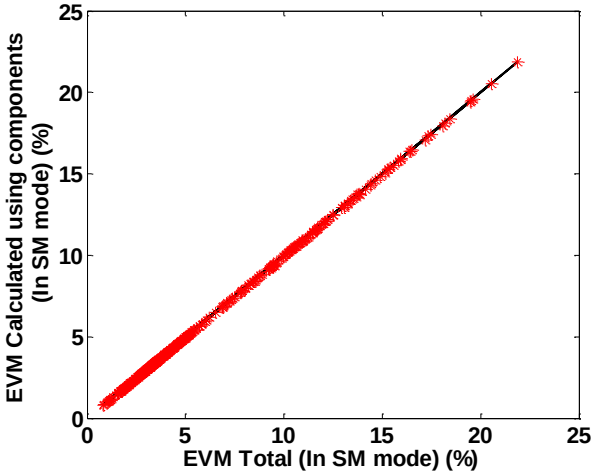


Figure 96: EVM calculated using decomposition versus actual EVM.

Using the predicted EVM_{rand} and EVM_{static}, the EVM_{total} is calculated using Equation 58. The variation of the relative error in prediction of EVM_{rand} with varying resolution of the digitizer is shown in Table 17. The normalized root mean squared error is a metric indicative of the residual variance in the error. From the obtained static behavioral parameters of the system, the system-level specifications are calculated. The relative error in prediction of system-level specifications under nominal test conditions of the transmitter chains is calculated according to Equation 70 and is tabulated in Table 18.

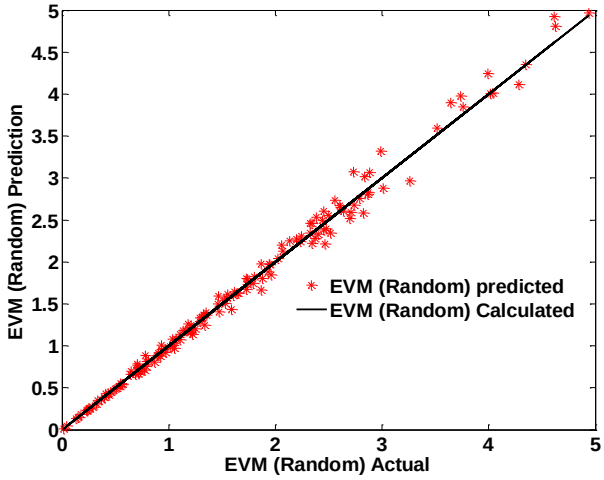


Figure 97: EVM_{rand} prediction plot.

$$Relative\ error\ = mean\{abs((actual - predicted)/actual)\} * 100. \quad \text{Equation 70}$$

The tuning knobs used are the two bias knobs of the two-stage PA, the bias and supply knobs of the gilbert cell mixer and the VGA gain settings (see Figure 84). A total of 9224 knobs combinations are used. 306 instances are obtained after removing extremely large parameter deviation or defective instances (greater than four σ where σ is % variance in specifications). Instances that failed the EVM_{total} in SM mode due to high values of EVM_{rand} in the 64QAM mode were discarded. Before performing analog tuning, digital correction for I/Q amplitude and phase mismatch was performed. Out of the 306 instances, 122 instances are initially within bounds of the system. The nominal power consumption of the MIMO system is 0.816 Watt. The tuning algorithm is used to tune the devices to nominal specifications (as explained in Section 4.5.2.6) with bounds given in Table 19.

Table 17: Error variation with ADC word size.

Digitizer	**Normalized root mean squared error in EVM_{rand} (%)**	**Worst Case prediction of EVM_{total} (%)**
16 bit 80 MHz	3.74	0.62
14 bit 80 MHz	4.87	0.97
12 bit 80 MHz	6.12	1.5

Out of the 204 devices that failed to meet the specification limits initially, 126 devices are tuned to within nominal specification bounds. At each step of tuning, the specifications were predicted. The prediction plot for EVM in SM mode across randomly selected tuning knobs and process instances is shown in Figure 98. Of the 126 devices, 90 devices are tuned with increase in power consumption. The initial and final tuned system specifications are shown in Figure 99. Considering the maximum tolerable power consumption overhead of 25% more than nominal, an upper bound of 1W on power consumption is selected. 50 out of 90 instances are tuned with a power consumption of more than 1W. These devices along with the other untuned devices are tuned in an adaptive manner.

For the nominal MIMO transmitter system, the entire transmission range is divided into 14 channels (channel 1 being the best channel (64QAM SM operation mode) and channel 14 being the worst-case channel (high EVM in SD mode under QPSK modulation). The EVM in the SD mode for the nominal device at channel 11 is between 17% and 22%. The 50 instances that consumed more than acceptable power are then tuned (2^{nd} iteration or 2^{nd} bin) in an attempt to trade-off performance and power according to the formulation given in the adaptive tuning Section 4.5.2.7 with the EVM ranges for channel 11 in the SD mode being between 24% (EVM_1) and 30% (EVM_2), $EVM_{th64QAM}$ = 3% and Tr = -26dBr. Thus, the goal is to maximize the chances for the devices to operate in good channel conditions while losing coverage during bad channels. The same bound that was used in first iteration on the EVM value in SM mode is used for second iteration. The prediction of the EVM in SD mode for channel 11 across knobs and process instances is shown in Figure 98. For predicting the EVM in SD mode, the static behavioral parameters along with the noise metric was used to build a map.

30 of the 50 initial instances could be tuned for this condition. Additionally, 29 instances that could not be tuned in the first tuning step are tuned for the given

constraints. The final tuned output power and power consumption of the 59 devices before and after tuning is shown in Figure 100. For a particular device tuned in this iteration, there exist 417 knob combinations that satisfy all of the EVM constraints with power values ranging from 0.659W to 0.92W and the optimization converged at its lowest value.

Table 18: Error in the prediction of various MIMO transmitter specifications.

Specifications	**Relative error (%)**
Transmit (Tx) average power (dBm)	1.63
Tx spectral mask (20,10 MHz offset) (dBr)	2.03,1.37
EVM (SM 64QAM mode) (%)	4.42
I/Q amplitude mismatch (Volts), phase mismatch (degree)	1.18, 2.03
Tx center frequency leakage (dB)	2.41
DC-offset(Volts)	1.02
Tx IIP3(dBm)	2.92
Tx cross coupling	2.51

Table 19: MIMO transmitter nominal specifications and pass bounds.

Tx O/P power $Tx_{i\,o/p}$	$EVM_{th64QAM}$	**TX spectral Mask at (20 MHz offset)** Tr
22.0 dBm (±1.5 dBm)	< 3 %	< -26 dBr

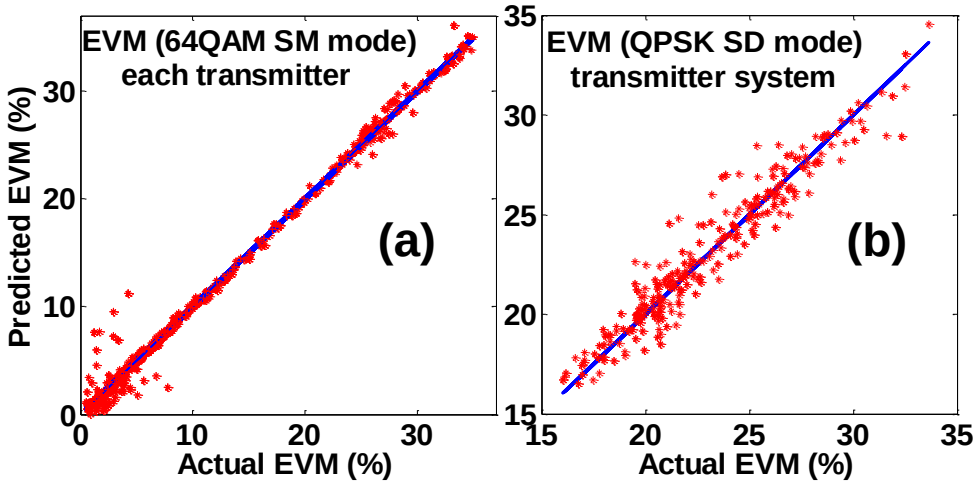


Figure 98: Prediction plots of the EVM under different channel conditions.

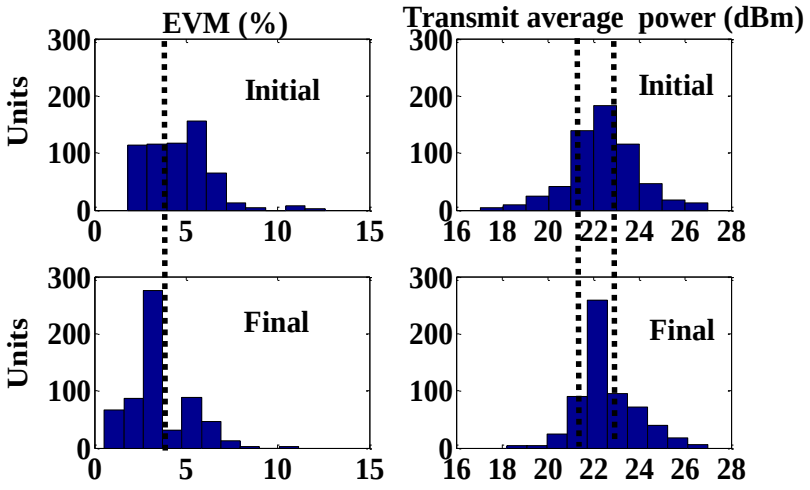


Figure 99: Devices tuned for nominal specs (1st iteration or bin).

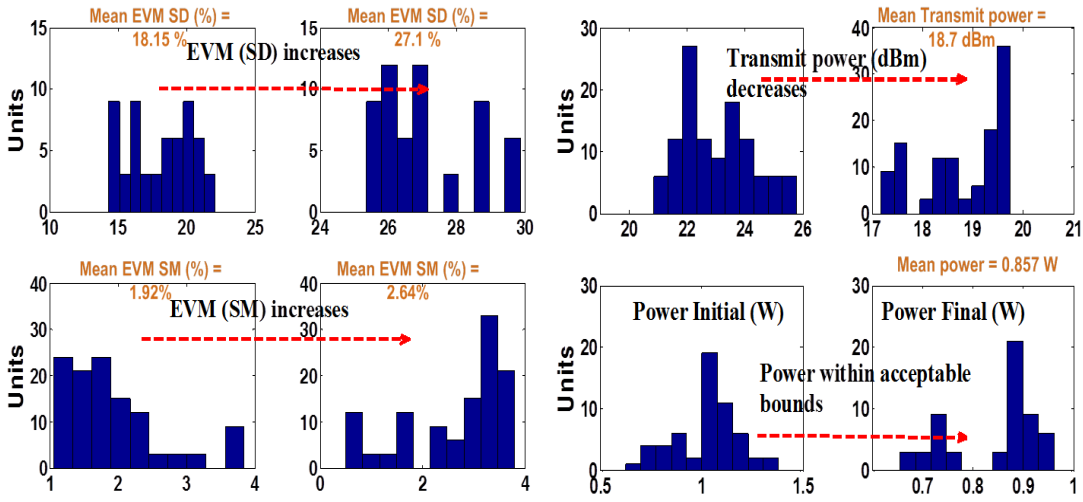


Figure 100: Histogram plots of devices tuned for 2nd tuning.

The system-level performance variation for devices in different bins or iterations is shown in Figure 101. The first subplot shows the system operation for a nominal device. The second subplot shows the EVM variation with channels for a process-skewed untuned instance. The third subplot shows the same for the process-skewed instance after tuning it to the nominal system specifications (1st iteration or bin). After tuning the device, the performance of the instance comes closer to the nominal. However, the final tuned power of the device is more than acceptable system power constraint of 1W. Hence, the device is tuned for 2nd bin specifications. After performing the second tuning, the performance over channels for good channels is better than that of process-skewed device and closer to the tuned device after 1st iteration but the performance is much worse for bad channels (i.e., EVM of 29%). Finally, the power value obtained at this iteration (0.937W) is the lowest among all possible knob combinations that satisfy the system specifications bounds for 2nd bin.

The remaining 20 instances are then tuned (3rd iteration or bin) for lower performance criteria of EVM in SD mode of channel 9 being between 24% and 30% and EVM in SM mode of 4%, and Tr = -26dBr. Out of the 20 instances, 15 instances are tuned.

Additionally 12 instances of the original untuned set are tuned for these performance criteria. The average performance and power metrics for the 27 tuned instances in the 3rd iteration or bin is shown in Table 20.

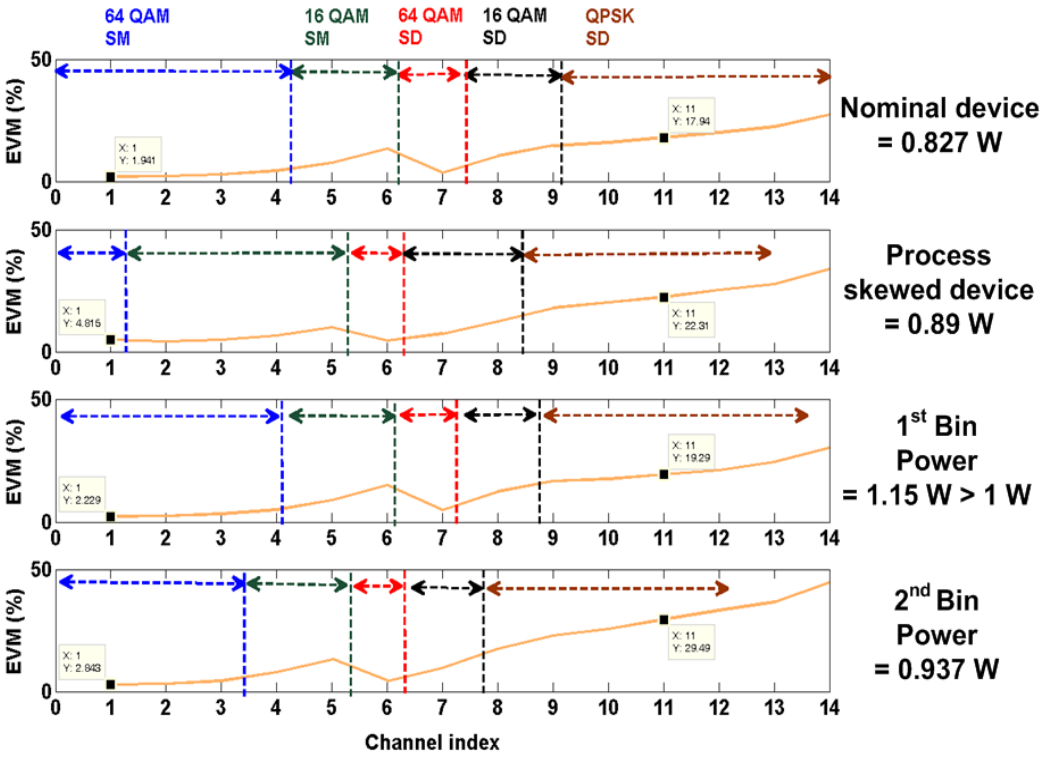


Figure 101: System performance variation for devices in different bins.

Hence for this system, the initial yield before tuning of the system is 122/390 (31%) and after performing self-tuning, 126 devices passed the specifications, out of which 50 instances are beyond acceptable power limits resulting in 198/390 (50.5%) tuned devices for nominal specs (1st bin). Additionally, 86 devices are tuned by using the power-performance trade-off concept. The yield numbers are calculated considering the devices that were initially removed as extreme outliers or defects. The overall result of the adaptive tuning methodology performed across the channels is shown in Figure 102. The Y-axis is the percentage of chips classified as pass and the X-axis is the number of channels across which the devices operate. Different such curves can be obtained depending on the system power constraints and system-level design constraints.

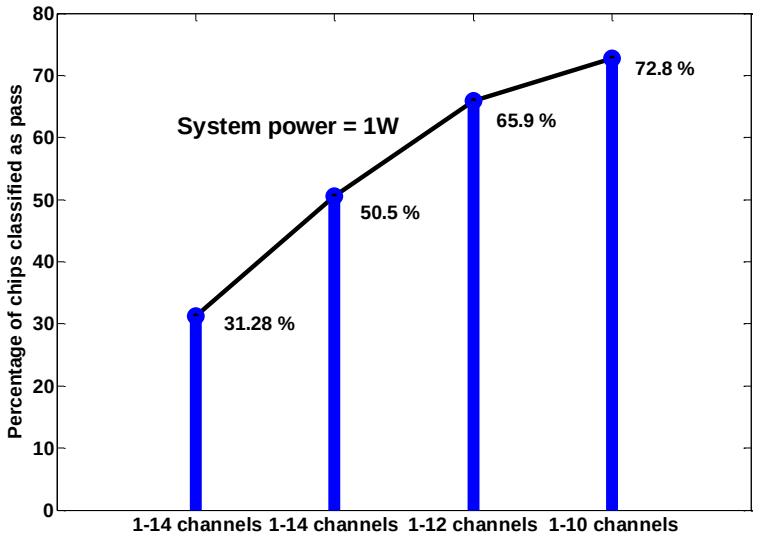


Figure 102: Adaptive tuning results.

While in this work, an algorithmic framework has been developed that attempts to maintain operation capability in good channels while losing coverage in the bad channels, other possible tuning mechanisms that place different constraints on different channels are also feasible. For e.g., one such technique would be to trade-off performance against power consumption by losing the capability to operate in high data rate modes (good channels) while attempting to maintain coverage in bad channels. The advantage of this work is that it provides a framework that can be used to adaptively tune these devices and bin them according to different system design and power constraints.

Table 20: Average performance metric of devices tuned in 3rd iteration.

Transmit power (dBm)	**EVM % (SM)**	**EVM % (SD) (channel 9)**	**Power (W)**
15.5	2.9	26.9	0.725

4.5.2.9. Hardware Validation

A MIMO 2 x 2 transceiver is implemented using external off-the shelf components. The baseband processing of MIMO system including channel emulation is implemented in Matlab. In the test mode architecture, only one downcoversion chain is used as shown in Figure 103.

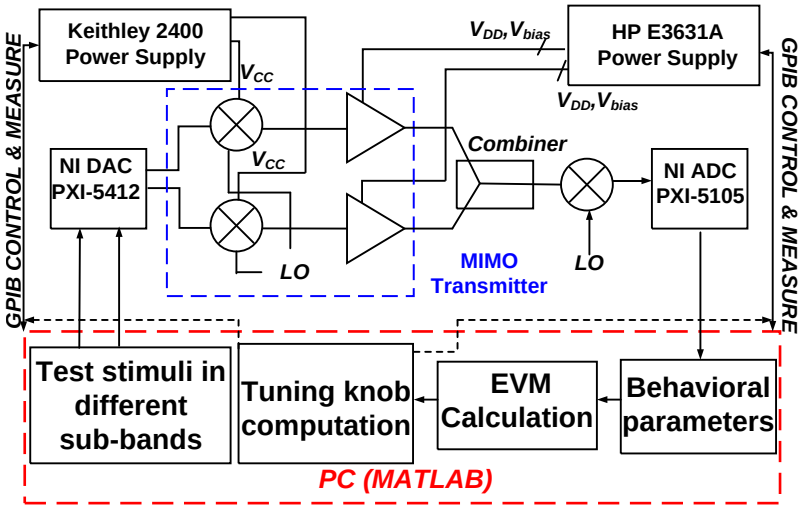


Figure 103: Hardware setup for MIMO RF transmitter.

The MIMO RF transmitter system consists of an upconversion mixer followed by an amplifier in each transmitter chain. For test purposes, the output of the two chains are "added" using a combiner and the combined signal is attenuated and down converted using a down-conversion mixer (MAX 2039). A NI-DAQ system is used for test stimulus delivery and response acquisition. It consists of two arbitrary waveform generators (PXI 5412) and one digitizer channel (PXI 5105), which are used to perform the data conversion operations. HP power supplies are used for providing power and are used to tune the bias and supply voltage of the amplifier. For the purpose of creation of process instances, the bias of up conversion mixers (ADL5801) is varied using a Keithley supply (one supply controls both mixers) and a varying amount of noise is added to the signal path. Each chain of the system was tested using bandwidth partitioned stimulus. In each

chain, five equally spaced tones are used (chain I - 120-125 KHz and chain II - 170-175 KHz). The inputs of the each chain experience different non-linearity characteristics that affect the signal swings and extent of compression in their respective output responses. The output responses of each chain are shown in Figure 104. The two responses are captured individually to show the difference in signatures obtained from each RF chain. However, in the test mode, the responses are combined. From this combined test response signature, the input-output linearity characteristics of the transmitter chains determined (see Figure 105). As can be seen from the graphs, the non-linearity characteristics have been captured accurately.

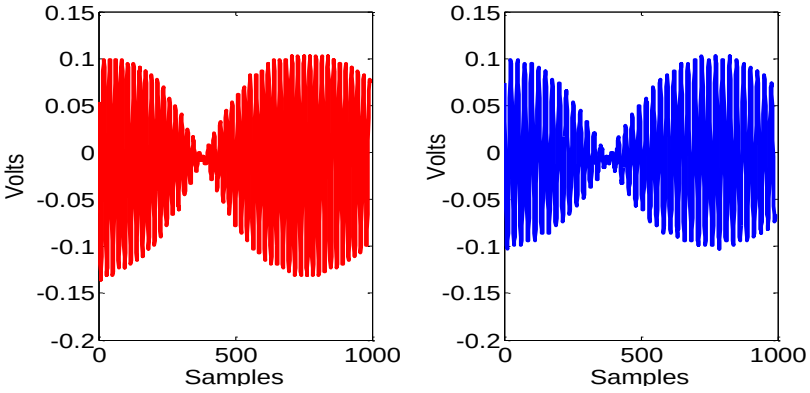


Figure 104: Time domain output waveforms from each chain.

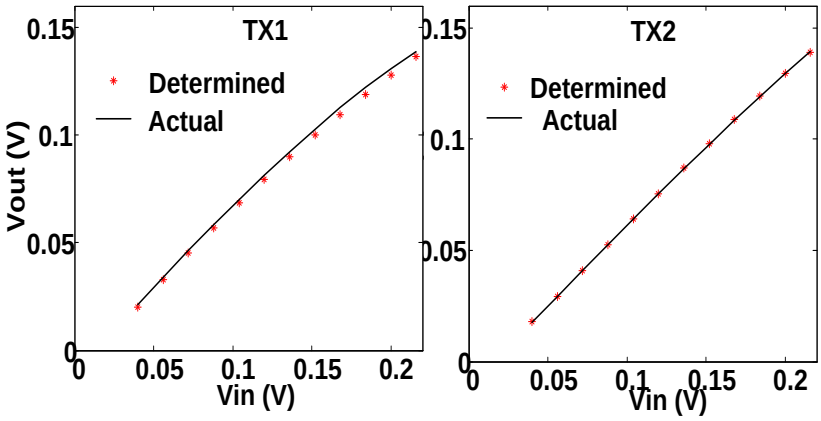


Figure 105: Non-linearity characteristics of the transmitter chains.

A regression mapping is then developed between EVM and the parameters. To develop the regression, the EVM (in SM mode of each transmitter) is calculated by implementing the channel in baseband (Matlab). Then the mapping between the behavioral parameters and EVM_{SM} is determined. The nominal specifications of the each chain is EVM_{SM} = 4.6% (bound: < 5.6%), Tx power = 1 dBm (±1 dBm) and Power = 1.6W. The average of the transmitter's initial and final performance metrics for three instances along with total initial and final dc power values are shown in Table 21.

Table 21: Tuning of performance metrics for MIMO-RF transmitter instances.

Instance	**Initial specifications**		**Final specifications**		**Power (mw)**	
	Tx O/P power	**EVM**	**Tx O/P power**	**EVM**	**Initial**	**Final**
1	-0.8	3.2	1.13	5.0	1.44	1.70
2	-0.3	7.4	0.9	4.1	2.2	2.64
3	2.6	4.7	1.5	5.1	2.31	1.82

CHAPTER 5. CONCLUSION AND FUTURE WORK

The objective of the research is to develop strategies for performing low cost test and diagnosis for analog/RF systems, and to enable post manufacturing self-tuning for yield improvement and reliability. Methodologies that attempt to detect and correct process variability and its effects through intelligent test, diagnosis, and compensation techniques have been presented. Solutions amenable to both production environment and in field operation are presented. The presented work targets OFDM-based single input single output (SISO) and multiple input multiple output (MIMO) systems. The benefits of the presented techniques are to reduce manufacturing costs.

In Chapter 3, a technique for performing concurrent testing and diagnosis of analog/RF circuits on a per-chip basis in production environment in presented. The method relies on the use of alternate diagnostic tests under which the DUT response (alternate diagnostic signature) exhibits strong simultaneous correlation with its specifications as well as critical Spice-level device parameters. This allows both the specifications and critical Spice-level device parameters of the analog/RF circuits to be predicted accurately from the DUT response with lower test time and test cost compared to standard testing techniques. The ability to predict the critical parameters of the circuit on a per-die basis aids in establishing a faster feedback methodology to correct for process variation shifts. In the latter part of the chapter, a technique to perform online diagnosis for OFDM transmitter systems using DSP-based time domain real-time signal monitoring technique is presented.

In Chapter 4, various methodologies for performing post-manufacturing tuning of OFDM-based SISO and MIMO systems for yield improvement is presented. The methodologies discussed in this work use intelligent post-manufacture built-in test and tuning algorithms that are aided by simple on-chip support infrastructure.

- In the realm of digital compensation, a low cost digital compensation technique is presented for both OFDM SISO systems and MB UWB OFDM systems where quicker digital compensation of RF impairments is achieved by using regression functions to predict the initial compensation values. The techniques do not use the receiver to estimate the imperfections in the transmitter.
- An effective self-contained methodology for tuning of RF modules using on-chip digital logic is developed. In this methodology, digital signatures obtained from the device are used to monitor the performance deviations and correct for it. This technique helps in performing DSP free die-level self-tuning or self-healing.
- An alternate approach that uses a gradient descent algorithm running on a DSP of the SISO OFDM system to tune multiple specifications in a power conscious manner is investigated.
- Finally, a methodology for performing efficient parallel system-level testing of MIMO-OFDM RF transmitter modules is presented. An adaptive power performance tuning technique for these modules that attempts to maximize system performance for given power constraint is also developed.

In future, the developed diagnosis technique needs to be extended to enable determination of the critical process parameters at an intra-die resolution. Such a methodology with intra-die resolution would be essential for performing diagnosis and providing feedback to the fab especially in scaled nanometer nodes below 45 nm where significant mismatch effects and random effects are observed in analog/RF circuits,. Further, a methodology that can leverage the diagnosis data to perform adaptive testing of these devices in the production floor can be investigated. An extension of the framework developed in this thesis is to leverage the process parameter information in the tuning

framework to perform a more design-specific tuning. While the circuits used in the tuning techniques incorporate some preliminary circuit tuning knobs, an in-depth study of the various possible tuning knobs for different analog/RF modules is required. With greater trends towards dynamic adaptation as well as process compensation, intelligent selection of tuning knobs in circuits is essential. One such effort is made in [77], where an orthogonal tunable LNA is presented. In this work, the gain and the non-linearity characteristics of the LNA can be controlled independently. The presented system-level test and adaptive tuning techniques can be extended to MIMO-OFDM RF receiver systems. Various other frameworks of performing adaptive tuning by trading off performance against power under different operating conditions need to be investigated. Such an adaptive tuning technique will be essential in nanometer node SoCs where the extreme process variations in the devices will make the goal of tuning the performance specifications of devices to one set of nominal performance metrics highly challenging.

REFERENCES

[1]. Luiz M. Franca-Neto, R. Eline, B. Balvinder, “Fully Integrated CMOS Radios from RF to Millimeter Wave Frequencies, “ Intel Technology Journal, Vol. 8, Issue 3, ISSN 1535-864X, Aug. 2004, pp. 241-258.

[2]. http://www.itrs.net/Links/2009ITRS/2009Chapters_2009Tables/2009_Wireless.pdf

[3]. “Analog-RF IP Integration challenges SoC Designers”, May 2006 issue,Chip Design Magazine.

[4]. Ghai, D.; Mohanty, S.P.; Kougianos, E.;, "Design of Parasitic and Process-Variation Aware Nano-CMOS RF Circuits: A VCO Case Study," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.17, no.9, pp.1339-1342, Sept. 2009.

[5]. C. Chiang, J. Kawa, “Design for Manufacturability and Yield for Nano-scale CMOS” (Springer, Dordrecht, 2007), pp. 14–15.

[6]. W. Zhao, Y. Cao, F. Liu, K. Agarwal, D. Acharyya, S. Nassif, K. Nowka, “Rigorous extraction of process variations for 65 nm CMOS design”, in Proceedings of European Solid-State Device Research Conference (ESSDERC), Sept. 2007, pp. 89–92.

[7]. Pratap, R.J.; Sen, P.; Davis, C.E.; Mukhophdhyay, R.; May, G.S.; Laskar, J.; , "Neurogenetic design centering," Semiconductor Manufacturing, IEEE Transactions on , vol.19, no.2, pp. 173- 182, May 2006 doi: 10.1109/TSM.2006.873517.

[8]. Georges G.E. Gielen;, "Design methodologies and tools for circuit design in CMOS nanometer technologies," Solid-State Circuits Conference, 2006. ESSCIRC 2006. Proceedings of the 32nd European , vol., no., pp.21-32, Sept. 2006.

[9]. Pelgrom, M.J.M.; Duinmaijer, A.C.J.; Welbers, A.P.G.; , "Matching properties of MOS transistors," Solid-State Circuits, IEEE Journal of , vol.24, no.5, pp. 1433-1439, Oct 1989.

[10]. J. B. Brockman and S. W. Director, “Predictive subset testing: optimization IC parametric performance testing for quality, cost, and yield,” IEEE Trans. Semi. Manufacturing, vol. 2, no. 3, pp. 104-113, Aug. 1989.

[11]. Roberts, G.W.; Dufort, B.; , "Making complex mixed-signal telecommunication integrated circuits testable," Communications Magazine, IEEE , vol.37, no.6, pp.90-96, June 1999.

[12]. MAX2839: 2.3GHz to 2.7GHz MIMO Wireless Broadband RF Transceiver. Available at http://datasheets.maxim-ic.com/en/ds/MAX2839.pdf

[13]. Milor, L.S.; , "A tutorial introduction to research on analog and mixed-signal circuit testing," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on* , vol.45, no.10, pp.1389-1407, Oct 1998.

[14]. Kwang-Ting (Tim) Cheng and Hsiu-Ming (Sherman) Chang, "Recent Advances in Analog, Mixed-Signal, and RF Testing," in IPSJ Transactions on System LSI Design Methodology (TSLDM), vol. 3, pp. 19-46, February 2010.

[15]. S. Ozev, A. Orailoglu and H. Haggag, “Automated test development and test time reduction for RF subsystems,” IEEE Intl. Symp. on Circuits and Sys., 2002, pp. 581-584.

[16]. Faust, M.G.; , "ATE features for Iddq testing," Test Symposium, 1998. ATS '98. Proceedings. Seventh Asian , vol., no., pp.153-157, 2-4 Dec 1998.

[17]. Devarayanadurg, G.; Soma, M.; Goteti, P.; Huynh, S.D.; , "Test set selection for structural faults in analog IC's," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.18, no.7, pp.1026-1039, Jul 1999.

[18]. Arabi, K.; Kaminska, B.; , "Oscillation built-in self test (OBIST) scheme for functional and structural testing of analog and mixed-signal integrated

circuits," Test Conference, 1997. Proceedings., International , vol., no., pp.786-795, 1-6 Nov 1997.

[19]. Silva, E.; Pineda de Gyvez, J.; Gronthoud, G.; , "Functional vs. multi-VDD testing of RF circuits," Test Conference, 2005. Proceedings. ITC 2005. IEEE International, vol., no., pp.9 pp.-420, 8-8 Nov. 2005.

[20]. Variyam, P. and Chatterjee, A., "Specification Driven Test Generation for Analog Circuits," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 19, No. 10, October 2000, pp. 1189-1201.

[21]. Voorakaranam, R., Akbay, S.S., Bhattacharya, S., Cherubal, S. and Chatterjee, A, "Signature Testing of Analog and RF Circuits: Algorithms and Methodology," IEEE Transactions on Circuits and Systems, Vol 54, Issue 5, May 2007, pp. 1018-1031.

[22]. Bhattacharya, S. and Chatterjee, A., "Wafer-Probe and Assembled-Package Test Co-Optimization to Minimize Overall Test Cost," ACM Transactions on Design Automation of Electronic Systems, Vol. 10, No. 2, pp. 302-329, 2003.

[23]. Halder, A., Bhattacharya S. and Chatterjee, A., "System-Level Specification Testing of Wireless Transceivers," IEEE Transactions on VLSI, Vol 16, Issue 3, March 2008, pp. 263-276.

[24]. E. S. Erdogan and S. Ozev, "Detailed Characterization of Transceiver Parameters Through Loop-Back-Based BiST", IEEE Transactions on VLSI Systems, Vol. 18, No. 6, 2010, pp. 901 - 911

[25]. Banerjee, A.; et, al., "Optimized Multitone Test Stimulus Driven Diagnosis of RF Transceivers Using Model Parameter Estimation," VLSI Design pp.274-279, 2-7 Jan. 2011.

[26]. Stratigopoulos, H.-G.D.; Drineas, P.; Slamani, M.; Makris, Y.; , "Non-RF to RF Test Correlation Using Learning Machines: A Case Study," VLSI Test Symposium, 2007. 25th IEEE , vol., no., pp.9-14, 6-10 May 2007.

[27]. Dabrowski, J.J.; Ramzan, R.M.;, "Built-in Loopback Test for IC RF Transceivers," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.18, no.6, pp.933-946, June 2010.

[28]. Srinivasan, G.; Chatterjee, A.; Taenzler, F.;, "Alternate loop-back diagnostic tests for wafer-level diagnosis of modern wireless transceivers using spectral signatures," VLSI Test Symposium, 2006. Proceedings. 24th IEEE , vol., no., pp.6 pp.-227, April 30 2006-May 4 2006.

[29]. Qi Wang; Soma, M.;, "RF front-end system gain and linearity built-in test," VLSI Test Symposium, 2006. Proceedings. 24th IEEE , vol., no., pp.6 pp.-233, April 30 2006-May 4 2006.

[30]. Qizhang Yin; Eisenstadt, W.R.; Fox, R.M.; Tao Zhang; , "A translinear RMS detector for embedded test of RF ICs," Instrumentation and Measurement, IEEE Transactions on , vol.54, no.5, pp. 1708- 1714, Oct. 2005.

[31]. Bhattacharya, S.; Chatterjee, A.; , "Use of embedded sensors for built-in-test RF circuits," Test Conference, 2004. Proceedings. ITC 2004. International , vol., no., pp. 801- 809, 26-28 Oct. 2004.

[32]. Andrew Hamilton, Richard Schofield, "An Expert System for Process Diagnosis," Proc. IEEE Intl. Conf. On Microelectronics Test Structures, vol. 2, no. 1,1989, pp. 55-57.

[33]. F. Di Palma, G. DeNicolao, G. Miraglia, O.M. Donzelli, “Process Diagnosis via Electrical-Wafer-Sorting Maps Classification,” IEEE International Conference on Data Mining, 2005, 4 pp.

[34]. Mehrdad Nourani; Arun Radhakrishnan; , "Modeling and Testing Process Variation in Nanometer CMOS," Test Conference, 2006. ITC '06. IEEE International , vol., no., pp.1-10, Oct. 2006.

[35]. Fang Liu; Nikolov, P.K.; Ozev, S.;, "Parametric fault diagnosis for analog circuits using a Bayesian framework," VLSI Test Symposium, 2006. Proceedings. 24th IEEE , vol., no., pp. 6 pp., 30 April-4 May 2006.

[36]. Ke Huang; Stratigopoulos, H.-G.; Mir, S.; , "Fault diagnosis of analog circuits based on machine learning," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2010 , vol., no., pp.1761-1766, 8-12 March 2010.

[37]. Sindia, S.; Singh, V.; Agrawal, V.D.; , "Parametric Fault Diagnosis of Nonlinear Analog Circuits Using Polynomial Coefficients," VLSI Design, 2010. VLSID '10. 23rd International Conference on , vol., no., pp.288-293, 3-7 Jan. 2010.

[38]. Cherubal, S.; Chatterjee, A.; , "Parametric fault diagnosis for analog systems using functional mapping," Design, Automation and Test in Europe Conference and Exhibition 1999. Proceedings , vol., no., pp.195-200, 9-12 March 1999.

[39]. Daasch, W.R.; McNames, J.; Madge, R.; Cota, K.; , "Neighborhood selection for IDDQ outlier screening at wafer sort," Design & Test of Computers, IEEE , vol.19, no.5, pp. 74- 81, Sep-Oct 2002.

[40]. Miyazaki, M.; Ono, G.; Ishibashi, K.; , "A 1.2-GIPS/W microprocessor using speed-adaptive threshold-voltage CMOS with forward bias," *Solid-State Circuits, IEEE Journal of* , vol.37, no.2, pp.210-217, Feb 2002.

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

[42]. Halder, A.; Bhattacharya, S.; Chatterjee, A.; , "System-Level Specification Testing Of Wireless Transceivers," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , pp.263-276.

[43]. M. Helfenstein, E. Baykal, K. Muller, and A. Lampe, "Error vector magnitude (EVM) measurements for GSM/EDGE applications revised under production

conditions," IEEE International Symposium on Circuits and Systems, Kobe, Japan, 2005, vol. 5, pp. 5003-5006.

[44]. J. Dabrowski and J. G. Bayon, "Mixed loopback BiST for RF digital transceivers," Proceedings 19th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, Cannes, France, 2004, pp. 220-228.

[45]. E. Acar, S. Ozev, K. B. Redmond, "Enhanced Error Vector Magnitude (EVM) Measurements for Testing WLAN Transceivers," ICCAD 2006, pp. 210-216.

[46]. Senguttuvan et.al..;, "Efficient EVM Testing of Wireless OFDM Transceivers Using Null Carriers," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.17, no.6, pp.803-814.

[47]. Natarajan, V.; Choi, H.; Lee, D.; Senguttuvan, R.; Chatterjee, A.; , "EVM Testing of Wireless OFDM Transceivers Using Intelligent Back-End Digital Signal Processing Algorithms," Test Conference, 2008. ITC 2008. IEEE International, vol., no., pp.1-10, 28-30.

[48]. Yilmaz, E.; Nassery, A.; Ozev, S.; Acar, E.; , "Built-in EVM measurement for OFDM transceivers using all-digital DFT," Test Conference, 2009. ITC 2009. International , vol., no., pp.1-10, 1-6.

[49]. Nassery, A.; Ozev, S.; Verhelst, M.; Slamani, M.;, "Extraction of EVM from Transmitter System Parameters," European Test Symposium (ETS), 2011 16th IEEE , vol., no., pp.75-80.

[50]. Vishwanath Natarajan, Hyun Woo Choi, Aritra Banerjee, Shreyas Sen, Abhijit Chatterjee, Ganesh Srinivasan, Friedrich Taenzler,Soumendu Bhattacharya: Low Cost EVM Testing of Wireless RF SoC Front-Ends Using Multitones. IEEE Trans. on CAD of Integrated Circuits and Systems 31(7): 1088-1101.

[51]. Senguttuvan, R.; Bhattacharya, S.; Chatterjee, A.; , "Test Method for Measuring Bit Error Rate of Pulsed Transceivers in Presence of Narrowband

Interferers," Microwave Theory and Techniques, IEEE Transactions on , vol.55, no.9, pp.1942-1950, Sept. 2007.

[52]. Bhattacharya, S.; Senguttuvan, R.; Chatterjee, A.;, "Production test enhancement techniques for MB-OFDM ultra-wide band (UWB) devices: EVM and CCDF," Test Conference, 2005. Proceedings. ITC 2005. IEEE International , vol., no., pp.10 pp.-245, 8-8 Nov. 2005.

[53]. Acar, E.; Ozev, S.;, "Low Cost MIMO Testing for RF Integrated Circuits," IEEE TVLSI, vol.18, no.9, pp.1348-1356, Sept. 2010.

[54]. SCHAUB K.;, "MIMO challenges existing ATE".

[55]. Buhler, M., et al. "DATE 2006 Special Session: DFM/DFY Design for Manufacturability and Yield - influence of process variations in digital, analog and mixed-signal circuit design," Proceedings of Design, Automation and Test in Europe, 2006. Volume 1, 6-10 March 2006 Page(s):1 – 6.

[56]. Valkama, M.; Salminen, K.; Renfors, M.; , "Digital I/Q imbalance compensation in low-IF receivers: principles and practice," Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on , vol.2, no., pp. 1179- 1182 vol.2, 2002.

[57]. J.K. Cavers, M.W. Liao, "Adaptive Compensation for Imbalance and Offset Losses in Direct Conversion Transceivers", IEEE Transactions on Vehicular Technology, Volume 42, Issue 4, November, 1993, pp. 581-588.

[58]. C.H. Park, J.H. Paik, Y-H You, M-C Ju, J-W Cho, " Techniques for Channel Estimation , DC-offset Compensation, and Link Quality Control in Bluetooth System", IEEE Transactions on Consumer Electronics, Volume 46, Issue 3, August 2000, pp. 682-689.

[59]. G-J Van Rooyen, J.G. Lourens, "Non-Iterative Compensation for Software-Defined Radio Quadrature Front-end Inaccuracies", Proceedings of Wireless

Communications, Networking and Mobile Computing, Volume 1, September 2005, pp. 598-601.

[60]. W. J. Kim, K. J. Cho, S. P. Stapleton, and J. H. Kim, "Baseband derived RF digital predistortion," Electronics Letters, vol. 42, pp. 468-70, 2006.

[61]. Hsin-Hung Chen; Chih-Hung Lin; Po-Chiun Huang; Jiunn-Tsair Chen; , "Joint Polynomial and Look-Up-Table Predistortion Power Amplifier Linearization," *Circuits and Systems II: Express Briefs, IEEE Transactions on* , vol.53, no.8, pp.612-616, Aug. 2006.

[62]. Montemayor, R.; Razavi, B.; , "A self-calibrating 900-MHz CMOS image-reject receiver," Solid-State Circuits Conference, 2000. ESSCIRC '00. Proceedings of the 26rd European , vol., no., pp.320-323, 19-21 Sept. 2000.

[63]. Jaehyok Yi; Youngoo Yang; Myungkyu Park; Wonwoo Kang; Bumman Kim; , "Analog predistortion linearizer for high-power RF amplifiers," Microwave Theory and Techniques, IEEE Transactions on , vol.48, no.12, pp.2709-2713, Dec 2000.

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

[65]. M. H. Perrot, T. L Tewkbury III, and C. G. Sodini, “A 27-mW CMOS fraction-N synthesizer using digital compensation for 2.5-Mb/s GFSK modulation,” IEEE J. Solid-State Circuits (JSSC), vol. 32, no. 12, pp. 2048- 2060, Dec. 1997.

[66]. C.-H. Wang, et. Al., “A direct digital frequency modulation PLL with all digital on-line self calibration for quad-band GSM/GPRS transmitter,” in Proc. of 2009 Symposia on VLSI Technology and Circuits (VLSI2009), June 2009.

[67]. Ping-Ying Wang; Zhan, J.-H.C.; Hsiang-Hui Chang; Chang, H.-M.S.; , "A Digital Intensive Fractional-N PLL and All-Digital Self-Calibration Schemes," Solid-State Circuits, IEEE Journal of , vol.44, no.8, pp.2182-2192, Aug. 2009

[68]. Hsiu-Ming Chang; Kuan-Yu Lin, Chin-Hsuan Chen and Kwang-Ting Cheng, “A Built-In Self-Calibration Scheme for Pipelined ADCs,” Proceedings, International Symposium on Quality Electronic Design, March 2009, pp. 266-271.

[69]. Lee, C.P.; Behzad, A.; Ojo, D.; Kappes, M.; Au, S.; Meng-An Pan; Carter, K.; Tian, S.; , "A Highly Linear Direct-Conversion Transmit Mixer Transconductance Stage with Local Oscillation Feedthrough and I/Q Imbalance Cancellation Scheme," Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International, vol., no., pp.1450-1459, 6-9 Feb. 2006.

[70]. A. Mostafa, I. Elmala' and H. K. Sherif, “A Self-Calibration Technique for Mismatches in Image-Reject Receivers”, IEEE 2002 CICC.

[71]. Darabi, H.; Chiu, J.; Khorram, S.; Kim, H.; Zhimin Zhou; Lin, E.; Shan Jiang; Evans, K.; Chien, E.; Ibrahim, B.; Geronaga, E.; Tran, L.; Rofougaran, R.; , "A dual mode 802.11b/Bluetooth radio in 0.35μm CMOS," *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International* , vol., no., pp. 86- 479 vol.1, 2003.

[72]. Vassiliou, I.; Vavelidis, K.; Georgantas, T.; Plevridis, S.; Haralabidis, N.; Kamoulakos, G.; Kapnistis, C.; Kavadias, S.; Kokolakis, Y.; Merakos, P.; Rudell, J.C.; Yamanaka, A.; Bouras, S.; Bouras, I.; , "A single-chip digitally calibrated 5.15-5.825-GHz 0.18-μm CMOS transceiver for 802.11a wireless LAN," *Solid-State Circuits, IEEE Journal of* , vol.38, no.12, pp. 2221- 2231, Dec. 2003.

[73]. Natarajan, V., Senguttuvan, R., Sen, S., Chatterjee, A., “ACT: Adaptive Calibration Test for Performance Enhancement and Increased Testability of Wireless RF Front-Ends,” VLSI Test Symposium, 2008. VTS 2008. 26th IEEE April 27 2008-May 1 2008 Page(s):215 - 220.

[74]. Donghoon Han; Byung Sung Kim; Chatterjee, A.; , "DSP-Driven Self-Tuning of RF Circuits for Process-Induced Performance Variability," Very Large Scale

Integration (VLSI) Systems, IEEE Transactions on , vol.18, no.2, pp.305-314, Feb. 2010.

[75]. Goyal, A.; Swaminathan, M.; Chatterjee, A.;, "A novel self-healing methodology for RF Amplifier circuits based on oscillation principles," *Design, Automation & Test in Europe Conference & Exhibition, 2009. DATE '09.* , vol., no., pp.1656-1661, 20-24 April 2009.

[76]. Kivekas, K.; Parssinen, A.; Ryynanen, J.; Jussila, J.; , "Calibration techniques of active BiCMOS mixers," Solid-State Circuits, IEEE Journal of , vol.37, no.6, pp.766-769, Jun 2002.

[77]. Sen, S.; Banerjee, D.; Verhelst, M.; Chatterjee, A.;, "A Power-Scalable Channel-Adaptive Wireless Receiver Based on Built-In Orthogonally Tunable LNA," *Circuits and Systems I: Regular Papers, IEEE Transactions on* , vol.59, no.5, pp.946-957, May 2012.

[78]. Che-Hong Liao; Huey-Ru Chuang; , "A 5.7-GHz 0.18-μm CMOS gain-controlled differential LNA with current reuse for WLAN receiver," *Microwave and Wireless Components Letters, IEEE* , vol.13, no.12, pp.526-528, Dec. 2003.

[79]. Amir-Aslanzadeh, H.; Pankratz, E.J.; Sanchez-Sinencio, E.; , "A 1-V +31 dBm IIP3, Reconfigurable, Continuously Tunable, Power-Adjustable Active-RC LPF," *Solid-State Circuits, IEEE Journal of* , vol.44, no.2, pp.495-508, Feb. 2009.

[80]. Liu, J.Y.-C.; Tang, A.; Ning-Yi Wang; Gu, Q.J.; Berenguer, R.; Hsieh-Hung Hsieh; Po-Yi Wu; Chewnpu Jou; Chang, M.-C.F.; , "A V-band self-healing power amplifier with adaptive feedback bias control in 65 nm CMOS," *Radio Frequency Integrated Circuits Symposium (RFIC), 2011 IEEE* , vol., no., pp.1-4, 5-7 June 2011.

[81]. J.H Friedman,"Multivariate adaptive regression splines," The Annals of Statistics, vol 19,no.1.pp.1-141,1991.

[82]. Byrd, R.H., J. C. Gilbert, and J. Nocedal, "A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming," *Mathematical Programming,* Vol 89, No. 1, pp. 149–185, 2000.

[83]. Achilles G.Kokkas, “Comprehsive semiconductor test structure”, Patent US Patent No. 4672314.

[84]. Lukaszek, W.; Grambow, K.G.; Yarbrough, W.J.; , "Test chip based approach to automated diagnosis of CMOS yield problems," *Semiconductor Manufacturing, IEEE Transactions on* , vol.3, no.1, pp.18-27, Feb 1990.

[85]. Chieh-Yuan Chao; Milor, L.S.; , "Performance modeling using additive regression splines," Semiconductor Manufacturing, IEEE Transactions on , vol.8, no.3, pp.239-251, Aug 1995.

[86]. Ramakrishnan, H.; Shedabale, S.; Russell, G.; Yakovlev, A.; , "Analysing the effect of process variation to reduce parametric yield loss," Integrated Circuit Design and Technology and Tutorial, 2008. ICICDT 2008. IEEE International Conference on , June 2008.

[87]. Wangyang Zhang; Xin Li; Liu, F.; Acar, E.; Rutenbar, R.A.; Blanton, R.D.; , "Virtual Probe: A Statistical Framework for Low-Cost Silicon Characterization of Nanoscale Integrated Circuits," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.30, no.12, pp.1814-1827, Dec. 2011.

[88]. R.Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society. Series B (Methodological) , Vol. 58, No. 1 (1996), pp. 267-288.

[89]. Wangyang Zhang; Xin Li; Acar, E.; Liu, F.; Rutenbar, R.; , "Multi-Wafer Virtual Probe: Minimum-cost variation characterization by exploring wafer-to-wafer correlation," Computer-Aided Design (ICCAD), 2010 IEEE/ACM International Conference on , vol., no., pp.47-54, 7-11 Nov. 2010.

[90]. Y.C. Chou, D. Leung, Q. Kan, M. Biedenbender, D. Eng, R. Lai, T. Block, A. Oki, “Reliability model for Predicting Long Term DC/RF Performance in GaAs PHEMTS”, in IEEE Compound Semiconductor Integrated Circuit Symposium, Nov 2005,pp.4.

[91]. B. R. Veillette and G. W. Roberts, “A built-in self-test strategy for wireless communication systems,” International Test Conference, 1995, pp. 930-939.

[92]. D. Han, S. Bhattacharya and A. Chatterjee, "Low-Cost Parametric Test and Diagnosis of RF Systems Using Multi-Tone Response Envelope Detection," IET Proceedings on Computers & Digital Techniques, Vol. 1, Issue 3, May 2007, pp. 170-179.

[93]. T. Zhang, W. R. Eisenstadt, R. M. Fox, Q. Yin, “Bipolar Microwave RMS Power Detectors,” IEEE Journal of Solid State Circuits, Vol. 41, No. 9, September 2006, pp. 2188-2192.

[94]. Jien-Chung Lo, “Online current testing,” IEEE Design & Test of Computers, Volume 15, Issue 4, Oct.-Dec. 1998, pp.49 – 56.

[95]. I. Pomeranz, S.M. Reddy, “Concurrent online testing of identical circuits using non-identical input vectors,” IEEE Transaction on Dependable and Secure Computing, Volume 2, Issue 3, July-Sept. 2005, pp.190 – 200.

[96]. M. Favalli, C. Metra, “Online testing approach for very deep-submicron ICs” IEEE Design & Test of Computers Volume 19, Issue 2, March-April 2002, pp.16 – 23.

[97]. Yueran Gao, Haibo Wang, "A Reconfigurable ADC Circuit with Online-Testing Capability and Enhanced Fault Tolerance", 24th IEEE symposium on Defect and Fault Tolerance in VLSI systems,2009, Page(s): 202-210.

[98]. E.Peralias, A.Rueda, J.A.Prieto and J.L.Huertas, "DFT & On-lin Test of High Performance Data Converters: Practical Case", International Test Conference, 1998, pp.534-540

[99]. V.Kolarik, M.Lubaszewski and B.Courtois, "Designing Self-Exercising Analogue Checkers", 12th VLSI Test Symposium, 1994, pp. 252-257.

[100]. A. Laknaur, R. Xiao and H. Wang, "A Programmable Window Comparator for Analog Online Testing", 25th VLSI Test Symposium, pp.119-124.

[101]. J.Velasco-Medina, I. Rayane and M. Nicolaidis, "On-line BIST for Testing Analog Circuits", International Conference on Computer Design, 1999, pp. 330-332.

[102]. Stratigopoulos,H-G.D.;Makris.Y; "An Adaptive Checker for the Fully Differential Analog Code," IEEE Journal of Solid-State Circuits, Vol.41,No.6, pp.1421-1429,June 2006.

[103]. Negreiros, M.; Carro, L.; Susin, A.A.; , "Low cost on-line testing of RF circuits," On-Line Testing Symposium, 2004. IOLTS 2004. Proceedings. 10th IEEE International , vol., no., pp. 73- 78, 12-14 July 2004.

[104]. Natarajan, V.; Srinivasan, G.; Chatterjee, A., "On-line error detection in wireless RF transmitters using real-time streaming data," International On-Line Testing Symposium, Lake of Como, Italy, July 10-12, 2006, pp.159-164.

[105]. Qiang Li; Jinlong Zhang; Wei Li; Yuan, J.S.; Yuan Chen; Oates, A.S.; , "RF circuit performance degradation due to soft breakdown and hot-carrier effect in deep-submicrometer CMOS technology," Microwave Theory and Techniques, IEEE Transactions on , vol.49, no.9, pp.1546-1551, Sep 2001.

[106]. Deyati, S.; Banerjee, A.; Chatterjee, A.; , "Pilot symbol driven monitoring of electrical degradation in RF transmitter systems using model anomaly diagnosis," On-Line Testing Symposium (IOLTS), 2012 IEEE 18th International , vol., no., pp.142-145, 27-29 June 2012.

[107]. J. Liu, H. Arslan, et al., "Impact of Carrier Frequency Dependent Power Amplifier Behavior on 802.11a WLAN System," European Conference onWireless Technology.Armsterdam, pp. 289–292, 2004.

[108]. Marsalek, R.; Prokopec, J.; , "Digital Predistorion with Transmitter Imperfection Simulation," Radioelektronika, 2007. 17th International Conference , vol., no., pp.1-4, 24-25 April 2007.

[109]. M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," IEEE Transactions on Microwave Theory and Techniques, vol. 54, pp. 348-59, 2006.

[110]. Z. Anding, J. Dooley, and T. J. Brazil, "Simplified Volterra series based behavioral modeling of RF power amplifiers using deviation-reduction," in 2006 IEEE MTT-S International Microwave Symposium Digest, San Francisco, CA, USA, 2006, p. 4 pp.

[111]. Adel A.M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers',IEEE Trans.Comm., Vol.29,pp 1715-1720, November , 1981.

[112]. A.Batra,J.Balakrishnan, et al., TI Proposal 03/141r3.802.15.3a, IEEE standards, May 2003.

[113]. A.Batra,J.Balakrishnan, et al., "Multi-band OFDM physical Layer Proposal for IEEE 802.15 Task Group 3a,"IEEE P802.15-04/493rl-TG3a,September 2004.

[114]. http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-368.pdf

[115]. http://www.ecmainternational.org/activities/Communications/tg20_UWB_Background.pdf

[116]. A.Batra,J.Balakrishnan, "Multi-band OFDM: A Cognitive Radio for UWB", Proc. IEEE Int. Circuits and Systems Symp.,vol.5, 4 pp.

[117]. http://www.alereon.com/products/chipsets/

[118]. A.Ismail,A.Abidi, " A 3.1 to 8.2 GHz direct conversion receiver for MB-OFDM UWB communications," 2005 ISSCC Dig. Tech Papers, Feb 2005.

[119]. Lee, F.S.; Chandrakasan, A.P. , “A BiCMOS Ultra-Wideband 3.1-10.6 GHz Front-End”, Proc. Custom Integrated Circuits Conference,2005,pp.153-156.

[120]. Park_Yunseo, “Direct Conversion RF Front-End Implementation for Ultra-Wide and GSM/WCDMA Dual-Band Applications in Silicon-Based Technologies”, Doctor of Philosophy Thesis.

[121]. Bhattacharya, S., Halder, A., Srinivasan, G. and Chatterjee, A., “Alternate Testing of RF Transceivers Using Optimized Test Stimulus for Accurate Prediction of Systems Specifications,” Journal of Electronic Testing: Theory and Applications, Vol. 21, No. 3, pp. 323-339. 2005.

[122]. D. Han and A. Chatterjee, “Robust built-in test of RF ICs using envelope detectors,” Asian Test Symposium, pp. 2-7, 2005.

[123]. Provost, B.; Sanchez-Sinencio, E.; , "On-chip ramp generators for mixed-signal BIST and ADC self-test," Solid-State Circuits, IEEE Journal of , vol.38, no.2, pp. 263- 273, Feb 2003.

[124]. Jiun-Lang Huang; Chee-Kian Ong; Kwang-Ting Cheng; , "A BIST scheme for on-chip ADC and DAC testing," Design, Automation and Test in Europe Conference and Exhibition 2000. Proceedings , vol., no., pp.216-220, 2000.

[125]. Po-Chih Wang; Chia-Jun Chang; Wei-Ming Chiu; Pei-Ju Chiu; Chun-Cheng Wang; Chao-Hua Lu; Kai-Te Chen; Ming-Chong Huang; Yi-Ming Chang; Shih-Min Lin; Ka-Un Chan; Ying-His Lin; Lee, Chao-Cheng, "A 2.4GHz Fully Integrated Transmitter Front End with +26.5-dBm On-Chip CMOS Power Amplifier," *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE* , vol., no., pp.263,266, 3-5 June 2007.

[126]. Natarajan, V.; Devarakond, S.K.; Sen, S.; Chatterjee, A., "BIST Driven Power Conscious Post-Manufacture Tuning of Wireless Transceiver Systems Using Hardware-Iterated Gradient Search," *Asian Test Symposium, 2009. ATS '09.* , vol., no., pp.243,248, 23-26 Nov. 2009.

[127]. Natarajan, V.; Sen, S.; Devarakond, S.K.; Chatterjee, A., "A holistic approach to accurate tuning of RF systems for large and small multiparameter perturbations," *VLSI Test Symposium (VTS), 2010 28th* , vol., no., pp.331,336, 19-22 April 2010.

[128]. Practical Manufacturing Testing of 802.11OFDM Devices, litepoint.

[129]. Heath, R.W.; Paulraj, A.J.; , "Switching between diversity and multiplexing in MIMO systems," Communications, IEEE Transactions on , vol.53, no.6, pp. 962-968, June2005.

[130]. Chan-Byoung Chae; Forenza, A.; Heath, R.W.; McKay, M.R.; Collings, I.B., "Adaptive MIMO transmission techniques for broadband wireless communication systems [Topics in Wireless Communications],"Communications Magazine, IEEE , vol.48,May 2010.

[131]. David E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Wesley Publishing Company, 1989.

[132]. Palaskas, Y.; , "A 5GHz 108Mb/s 2x2 MIMO Transceiver with Fully Integrated +16dBm PAs in 90nm CMOS," Solid-State Circuits Conference, 2006. ISSCC 2006. IEEE International , pp.1420-1429.

[133]. Bonnans, J. Frédéric; Gilbert, J. Charles; Lemaréchal, Claude; Sagastizábal, Claudia A. (2006). *Numerical optimization: Theoretical and practical aspects*. Universitext (Second revised ed. of translation of 1997 French ed.). Berlin: Springer-Verlag.

[134]. Garuda, C.; Ismail, M.;, "A multi-standard OFDM-MIMO transceiver for WLAN applications," *Circuits and Systems, 2005. 48th Midwest Symposium on* , vol., no., pp.1613-1616 Vol. 2, 7-10 Aug. 2005.

[135]. http://www.tranzeo.com/allowed/Tranzeo_Link_Budget_Whitepaper.pdf

[136]. Broyde, Y.; Messer, H., "A cellular sector-to-users path loss distribution model," *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on* , vol., no., pp.321,324, Aug. 31 2009-Sept. 3 2009.