BUILT-IN SELF TEST AND CALIBRATION OF RF SYSTEMS FOR PARAMETRIC FAILURES

A Dissertation
Presented to
The Academic Faculty

by

Dong-Hoon Han

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, 2007

BUILT-IN SELF TEST AND CALIBRATION OF RF SYSTEMS FOR PARAMETRIC FAILURES

Approved by:

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

Dr. Madhavan Swaminathan
School of Electrical and Computer Engineering
Georgia Institute of Technology

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

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

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

Date Approved: [March 29, 2007]

To my wife, Jihee Youn,

and

To my sons,Seokhyun and Seungwoo Han

For their tremendous love, support, and belief in me.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Abhijit Chatterjee, for his precious support and advice guiding me through my whole Ph.D. career. Without his patience, guidance, and continued support I couldn't have made it this far. I would also like to thank my committee members, Dr. David Keezer, Dr. Madhavan Swaminathan, Dr. Linda Milor, and Dr. Suresh Sitaraman, for taking the time to review my work and providing valuable suggestions to improve my thesis.

I am greatly indebted to my colleagues at Testing and Reliability Engineering Lab, S. Bhattacharya, A. Halder, C. Seo, G. Srinivansan, S. Akbay, S. Goyal, H. Choi, R. Senguttuvan, M. Ashouei, V. Natarajan, S. Sen, and S. Kook, for their valuable suggestions on various aspects of my work and life.

I am thankful to all my parents and parents-in-law for their unlimited support, encouragement, and belief in me.

Most of all, I would like to thank my wife, Jihee Youn, and my sons, Seokhyun Han and Seungwoo Han, for their love, support and friendship. You make everything worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACPR	Adjacent Channel Power Ratio
ADC	Analog to Digital Converter
ATE	Automatic Test Equipment
BIST	Built-In Self-Test
BPF	Band-pass Filter
CMRR	Common Mode Rejection Ratio
COT	Cost of Test
DAC	Digital to Analog Converter
DAQ	Data Acquisition
DFT	Design for Testability
DIB	Device Interface Board
DSP	Digital Signal Processing
DUT	Device Under Test
EVM	Error Vector Magnitude
IC	Integrated Circuit
IIP3	Input third-order Intercept
IF	Intermediate Frequency
ITRS	International Technology Roadmap for Semiconductor
IMD	Inter-Modulation Distortion
LNA	Low Noise Amplifier
LPF	Low-Pass filter

MARS	Multivariate Adaptive Regression Splines
NF	Noise Figure
OIP3	Output third-order Intercept
PA	Power Amplifier
RF	Radio Frequency
SNR	Signal to Noise Ratio
SoC	System-on-Chip
TOI	Third Order Intercept

SUMMARY

Continuing advances in CMOS technology have resulted in hardware designs of ever increasing complexity. Systems can have billions of transistors that incorporate into a single die such mixed-signal systems as analog/RF/digital circuitry. In addition, the use of scaled CMOS technologies enables these to operate in multi-GHz frequencies. Such systems pose unprecedented challenges both in production testing and manufacturing yield. The need to reduce the costs of production tests and to improve parametric yields becomes even more crucial as processes move to geometries of less than 100 nanometers and process variations continually increase. In fact, the cost to test modern mixed-signal systems-on-chip (SoC) can be as high as 30 percent of their manufacturing cost, and yields for ICs with geometries below 100 nanometers may not exceed 50 or 60 percent.

This thesis proposes a multifaceted production test and post-silicon yield enhancement framework for RF systems. The three main components of the proposed framework are the design, production test, and post-test phase of the overall integrated circuit (IC) development cycle. First, a circuit-sizing method is presented for incorporating test considerations into algorithms for automatic circuit synthesis/device resizing. The sizing problem is solved by using a cost metric that can be incorporated at minimal computational cost into existing optimization tools for manufacturing yield enhancement. Along with the circuit-sizing method introduced in the design phase, a low-cost test and diagnosis method is presented for multi-parametric faults in wireless systems. This test and diagnosis method allows accurate prediction of the end-to-end specifications as well as for the specifications of all the embedded modules. The

procedure is based on application of optimized test stimulus and the use of a simple diode-based envelope detector to extract the transient test response envelope at RF signal nodes. This eliminates the need to make RF measurements using expensive standard testers. To further improve the parametric yield of RF circuits, a performance drift-aware adaptation scheme is proposed that automatically compensates for the loss of circuit performance in the presence of process variations. This work includes a diagnosis algorithm to identify faulty circuits within the system and a compensation process that adjusts tunable components to reduce the effects of performance variations. As a result, all the mentioned components contribute to producing a low-cost production test and to enhancing post-silicon parametric yield.

CHAPTER 1

INTRODUCTION

Wireless communications for both mobile and in-office (point-to-point communication) applications are undergoing a revolution because of the proliferation of different communication standards that span diverse communication bandwidths. In conjunction with these applications, advances in semiconductor process technology and in circuit design tools have increased the degree of circuitry integration and enabled most of the functionality to be placed on a single chip. These solutions thus permit a significant reduction in system-level manufacturing costs with less power consumption.

One of the significant challenges faced by the semiconductor manufacturer is the production test cost because of ever-increasing complexity of tests [1]-[3]. As shown in Figure 1, the test cost issue was predicted in the '97 SIA roadmap in which test capital per transistor was expected to exceed the silicon capital cost.

Although analog and RF circuitry is much smaller than digital circuitry in most modern systems-on-chip (SoCs), analog and RF circuits require the use of complicated test procedures, lengthening test times and increasing the cost of automatic test equipment (ATE). As a result, the test cost of modern mixed-signal SoCs can be as high as 30% of their manufacturing cost [1], [4] and is impacted significantly by the cost of testing the various embedded analog and RF circuit components.

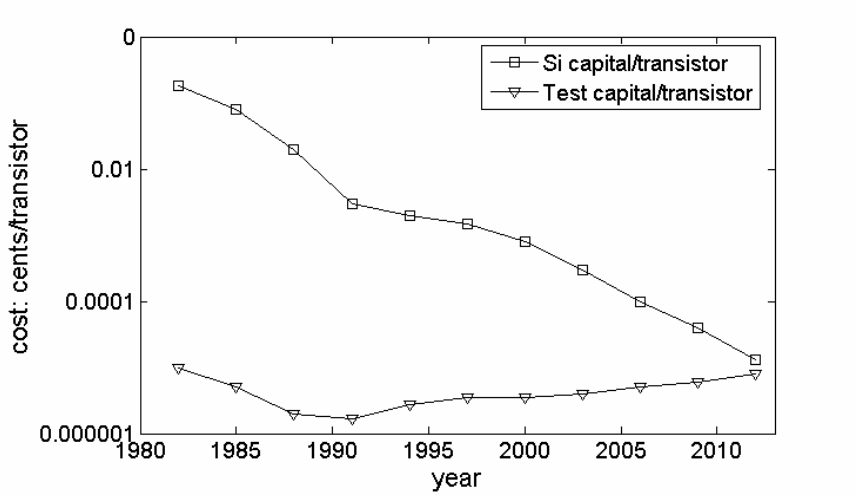


Figure 1. Silicon and test capital per transistor.

Along with the production test cost issue, technology scaling has been accompanied by increased device performance and power sensitivity to process variations [5]. Figure 2 clearly shows the current trends of 3σ process variations for three representative parameters in which process variations have be increased as the feature size of a transistor reduces.

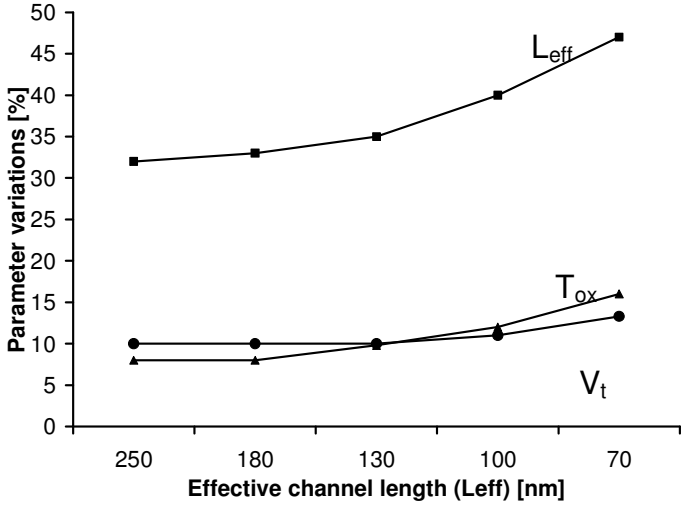


Figure 2. Technology parameter 3σ variations [5].

The result has been significantly lower yields of ICs. In addition, manufacturing yield can be degraded by an amount inversely proportional to the die area [6]. Unless some mechanism is invented to make CMOS devices robust to process variations, there is fear that CMOS technology scaling may grind to a halt.

1.1 Production Test of Analog and RF Circuits

The faults of analog and RF circuits are usually divided into two classes [7]-[9]. Parametric faults, which result from global fluctuations inherent in the manufacturing process, are usually modeled by small deviations in the circuit parameters. The second class, catastrophic faults, which occur because of local effects such as spot defects, are usually modeled by a topological change of the circuits. As devices grow more complex at the same time as their dimensions shrink, the performance of circuits is increasingly sensitive to deviations inherent in fabrication. Catastrophic faults are likely to result in abrupt degradation, whereas parametric faults typically lead to smooth variations of the measurements obtained from the device-under-test (DUT). Consequently, although parametric faults are becoming more and more important, their detection poses a much more difficult problem than the detection of catastrophic faults [9], [10].

A production test for parametric faults as well as for catastrophic faults in analog and RF circuits is typically specification oriented [10]-[13], which involves measuring all specifications of the DUT and then comparing them against the corresponding acceptance limits that determine pass or fail of DUTs.

The prohibitive cost of production testing of analog and RF circuitry results mainly for the following reasons:

- Each specification test requires a different test configuration and stimulus, resulting in long overall per-chip test time on expensive ATE. For example, the tests for the third-order input intercept and the noise figure require significantly different tests setups.
- Increased circuit complexity on a single die prevents testing access to internal nodes. The alternative of routing the internal nodes to external pins significantly increases the overall manufacturing cost. Furthermore, it is inapplicable to RF circuitry because it produces loading effects that in turn induce significant circuit performance drifts.
- Complicated specification measurements call for expensive ATE, which costs a few million dollars. In turn, the cost of a single second on a $3M RF tester comes to around a dime if the tester runs constantly and without interruption [14].
- Elaborate measurement setups and complicated load boards, such as high-performance sockets with precision pressure and electromagnetic isolation, result in increased test time [15], [16]. Moreover, these test setups call for complicated calibration processes so that calibration performance is maintained during the production test.
- Multi-site capability needs to be enhanced to increase test throughput on expensive ATE. However, because of the increased operating frequency of DUTs and the complicated device interface board (DIB) and tester load board placed between the DUT and the ATE, the number of test sites in a RF ATE cannot exceed eight.

These technical test issues of analog/RF circuits combine with increased manufacturing cost to elevate test-related expenditures to an ever larger percentage of the overall cost of chip production in the face of intense competition that prohibits price increases. The result is a squeeze on manufacturers' profits.

1.2 Prior Work in Production Test

Traditionally, to minimize test time incurred during specification testing, the statistics and joint probability density functions (PDFs) of various specifications have been used to test for only a subset of test specifications without compromising fault coverage. These test methods rely on two major principles:

- Use of test set compaction [10], [17], [18], which attempts to identify redundant tests and remove them from the test set with minimal impact on overall fault coverage
- Use of test scheduling or test ordering [10], [19]-[21] in which the average test time for a device across all the tests is minimized. In this approach, the failure-prone specifications are tested early in the device test plan and the test sequence is ordered in such a way as to minimize overall test cost.

Although these approaches have contributed to test cost reduction, they are mainly oriented toward single chip tests in which controllability and observability are fully provided. For complicated SoC tests having limited controllability and observability, alternatives to these approaches are essential to cut the ever-increasing cost of testing. In addition, the test cost reduction methods outlined above must be tailored to each device design and to the associated process statistics. The design procedure itself is typically

independent of the test procedure, with the latter being treated as a back-end process that is performed after design is completed. In recent years however, this thinking is beginning to change so that tests are considered early in the design cycle as a way to minimize time-to-market and to effectively manage test development costs for high-performance devices.

As IC designs become more complex, the problem of testing them becomes complex, too. This is compounded by the fact that observability and controllability only scale inversely compared with the functionality of ICs. The issue has been addressed by numerous design-for-test (DFT) and built-in self-test (BIST) techniques. In DFT approaches, test circuitry is deployed inside the device and ensures that the internal nodes are routed to dedicated test pins, or its signal is converted favorably to testing. For example, a scan design for digital circuitry has proliferated in industry. Furthermore, the BIST methodology provides a facility for on-chip testing. The main purpose of BIST is to reduce the complexity of testing and thereby decrease the cost and simultaneously lessen the need for reliance on expensive external test equipment. BIST reduces test costs in two ways: (1) reduced test-cycle duration, and (2) a simpler test/probe setup through reducing the number of I/O signals that must be driven/examined under test control. Both lead to a reduction in hourly charges for ATE service. Although DFT and BIST techniques for digital circuitry, such as logic and memory chips, have proliferated, their application to analog and RF circuits has lagged badly [22].

On the other hand, an alternate test methodology has been proposed in [23] as a low-cost test alternative to typical specification testing. Since the original introduction of the alternate test methodology, many researchers have successfully explored implementation and expanded application to analog, mixed-signal, and RF circuits [24]-

[29]. In contrast to specification-based testing, the alternate test methodology, without recourse to explicit measurements, employs just one test configuration and a single test stimulus to predict all the specifications of interest.

Process variations inherent in the manufacturing process affect the circuit specification as well as the measurement space, such as transient responses obtained from a test stimulus. Figure 3 illustrates the effect of the variation of one such parameter in **P** on the specification **S** and the corresponding variation of a particular measurement data in **M**. For any point p in the parameter space **P**, a mapping function onto the specification space **S**, f:**P**→**S** can be computed. Similarly, for the same point p in **P**, another nonlinear mapping function onto the measurement space in **M**, f:**P**→**M** can be computed. Therefore, for a region of acceptance in the circuit specification space **S**, there exists in the parameter space **P** a corresponding allowable region of variation of the parameters. This in turn defines a region of acceptance of the measurement data in the space **M**. A circuit can be declared faulty if the measurement data lies outside the acceptance region in **M**. Alternatively, by using nonlinear regression mapping tools, the mapping function f:**M**→**S** could be constructed for the circuit specification **S** from all measurements in the measurement space **M** [30].

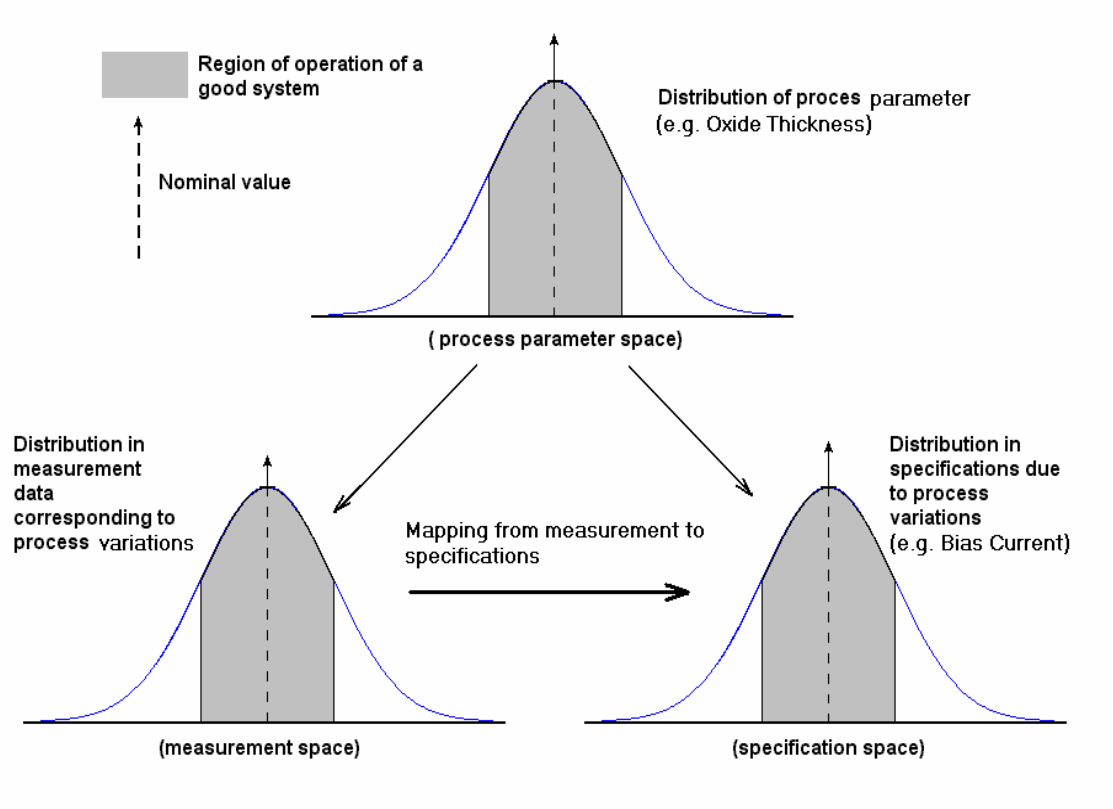


Figure 3. Relationship among process parameter, measurement, and specification spaces [28].

The overall procedure of the alternate test methodology is depicted in Figure 4. A specially crafted test stimulus is applied to the DUT in such a way that the resulting test response is strongly correlated with the specification values of interest. Hence, the statistical relation from the test response to the specifications of interest can be accomplished by nonlinear regression mapping [30] and the specification prediction can be achieved through implicit measurements.

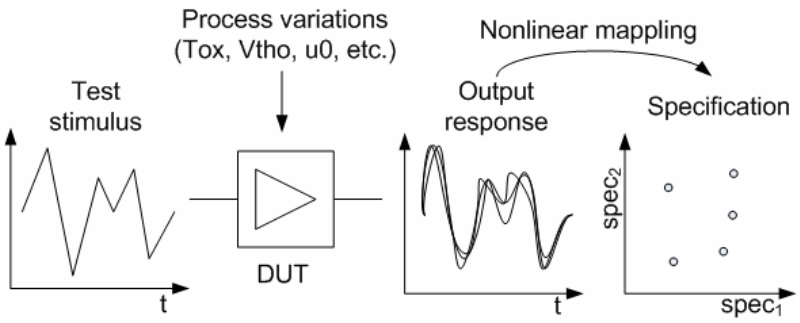


Figure 4. Alternate test methodology.

In this approach, a piecewise linear signal or multi-tone sinusoidal waveform as a test stimulus, in general, is used and can be optimized via special alternate test generation algorithms [23], [24], [27]. As a result, the alternate test methodology simplifies test setups and reduces the number of measurements and amount of switching between tests and accurately predicts specification values.

In this work, the alternate test methodology is extended to RF circuit testing and incorporated into the BIST technique, and thereby accelerates production testing on low-cost testers (or via reuse of embedded hardware resources) and simplifies the complexity of test setups needed for specification measurements.

1.3 Post-Silicon Yield Enhancement

The capability to take into account the natural random variability of the fabrication process is of strategic importance because yield relates directly to profitability. In the IC design phase, circuit designers construct a circuit topology for a set of predefined performance metrics and then search for a set of nominal parameters so that the circuit will meet its design performance specifications, such as those for speed and

power dissipation, under all operating conditions. Because of the random fluctuation in the fabrication process and random variations in the circuit operating conditions, one should select the parameters in such a way that the designed circuit has sufficient margins to meet the acceptability criteria. Statistical design approaches are usually employed to select the nominal parameters for enhancing manufacturability and thus maximizing manufacturing yields.

In the design phase, what designers want in designing an analog and RF circuit is (1) to find a way to compute how much variability is associated with any given choice of nominal design and (2) to find an efficient way to adjust the nominal design choices in the direction of reducing variability to acceptable levels. Monte Carlo analysis is generally used to handle the former goal. However, Monte Carlo analysis is computationally intensive and is quite unsuitable for achieving the second goal of reduced variability. This brought about the development of an efficient way to estimate variability at the system level [31], [32]. On the other hand, starting from nominal design, a local optimization process tries to push performance away from the specification boundaries so as to make the circuit more robust against technology parameter variations. This process is called design centering as post-design yield optimization [33]-[35]. Approaches toward analog synthesis or circuit sizing for manufacturability have been presented that combine nominal circuit optimization and variation analysis for enhanced manufacturing yield [36]-[38].

Classical electronic design described above, however, relies on analytical equations and numerical tables for choosing parameter values. With computer-aided design and appropriate software tools, optimal design can be more readily obtained. However, because of limitations on the exact realization of the analytical and simulation

output and on statistics regarding process variations, there should be a gap between real manufacturing and simulation results. Moreover, performance variations are anticipated to be increased significantly because of technology scaling in very deep-submicron regimes. To ensure efficient circuit manufacturability, self-calibration to compensate for performance variability needs to be done after fabrication.

In the past, such calibration has been difficult because of the inability to use on-chip mechanisms to measure the deviations of a RF circuit's performance metrics from what was expected. While it is possible to use complex external testers to test whether a circuit meets all its design specifications, the constraints of a production test cost budget prohibit on-chip measurement of all the design specifications of an RF circuit. Current RF BIST techniques are restricted to measuring only one or perhaps a few RF design specifications and are not yet accepted in the industry [22]. Further, self-calibration procedures require the capability to perform circuit diagnosis from the test results.

While there has been work in the past on diagnosis of performance loss in RF systems, prior research has focused on diagnosis and compensation for very specific performance loss mechanisms (such as gain loss, I-Q mismatch, DC offset errors [39]-[42]) using different testing methods for each specification. As a consequence, specific targeted compensation for a diagnosed non-ideal circuit parameter is performed with little emphasis on addressing the mitigation of the *combined effects of many circuit nonidealities* that are more likely to occur in practice and are harder to resolve with compensation mechanisms that address only a single circuit level parameter deviation. This drawback comes from lack of test and diagnosis methods for multi-performance deviations of RF circuits. Therefore, a single tuning mechanism is carefully designed to adjust only a single performance parameter with no or little impact on other performance

parameters. In addition, to simultaneously compensate for multi-performance variability, controllability of circuit performances needs to be enhanced using multiple tuning mechanisms. However, it leads to a higher complexity of optimization problems in searching for optimum trim of tuning mechanisms. Inevitably, this process involves iterative test and diagnosis of a circuit under the stringent constraints of manufacturing costs.

1.4 Research Contributions

The objective of the research in this dissertation is to develop a framework for low-cost production testing and post-silicon parametric yield enhancement. Each component of the framework is employed in the design, production test and post-test phase on the IC development cycle.

First, a circuit sizing method is proposed that aims to reduce device test time without adversely affecting its performance. The goal of this research is to *present a methodology for incorporating test considerations into automatic analog circuit synthesis/device resizing* algorithms. This is novel because existing analog and RF circuit design optimization tools for circuit synthesis and device sizing primarily focus on performance-driven circuit design [43]-[50]. Little research has been reported in the area of incorporating test considerations into automatic analog circuit synthesis and device sizing. The contributions of this component are

- A test cost reduction method called dynamic test elimination is introduced. Through analysis of the test redundancy among various specifications, one or more specification tests can be eliminated from the test suite.

- A cost metric for design optimization is formulated. It is shown that this test cost metric shows high correlation with the test complexity (test time) of a set of devices of the same type (test pilot) with different device sizes, all of which meet specified performance metrics. Hence, such a test cost can be easily incorporated into a circuit synthesis/sizing procedure that accommodates power, area and yield optimization as well. When employed in conjunction with yield enhancement methods, the cost metric can be extracted from the yield statistics and thereby no additional circuit simulations or evaluations are required.

To reduce production test costs, it is necessary to measure all the RF test specifications using a low-cost external test system. In this work, a low-cost test and diagnosis scheme for integrated RF systems is proposed. Though circuit designers have devoted to develop true RF detectors to measure power and RMS of RF internal signals, simple and small sensors are *designed into* the RF load board or the RF circuit itself to facilitate manufacturing test and diagnosis. The objectives of the work are as follows:

- To be able to test (i.e. generate pass/fail information) and diagnose an RF system under simultaneous multi-parameter perturbations (simultaneous performance variations in multiple modules) with accuracy similar to the results achievable for parametric failures with standard RF test instrumentation.
- To provide a low-cost test solution that will be performed using test response sensors on the load board or inside the die with only baseband processing support from an on-chip DSP. The resulting diagnostic data can be used for process debugging and rapid yield ramp-up. Even though a low-speed digitizer is used to process the test response signals, the predicted specifications are at speed test specification values.

- To study how test and diagnosis accuracy is increased by incorporating sensors at the outputs of the RF modules of a wireless system in addition to those sensors already available at its observable outputs. The goal is to study the additional test and diagnosis accuracy obtained in exchange for the extra effort expended in incorporating such sensors into the design of an RF front-end system. Note that we desire accurate diagnosis of each RF module even when its test input signal has changed because of parametric variations in other modules that feed its inputs (multi-parameter, multi-module parametric perturbations).

A self-calibration technique is proposed to develop global RF *specification-aware compensation methodology* that can trade off performance specifications against one another in a preferred way while performing *simultaneous multi-parameter compensation*. The key contributions of the proposed self-calibration approaches are as follows:

- The hardware cost of performing diagnosis and compensation should be very low relative to existing methods that can only perform limited parameter tuning. Further, the diagnosis and compensation techniques must be capable of functioning without any external RF tester support.
- The proposed fine-grained diagnosis and self-compensation technique will allow fine-grain tuning to be possible using completely on-chip hardware and software resources with a single shot. This will have the effect of reducing the overall cost to run a self-calibration procedure.
- Using the proposed techniques, it will be possible to maintain high manufacturing yield and robust long-life operation at high levels of reliability for scaled CMOS RF front-end designs.

In summary, the contributions of each of these components are highlighted in Figure 5.

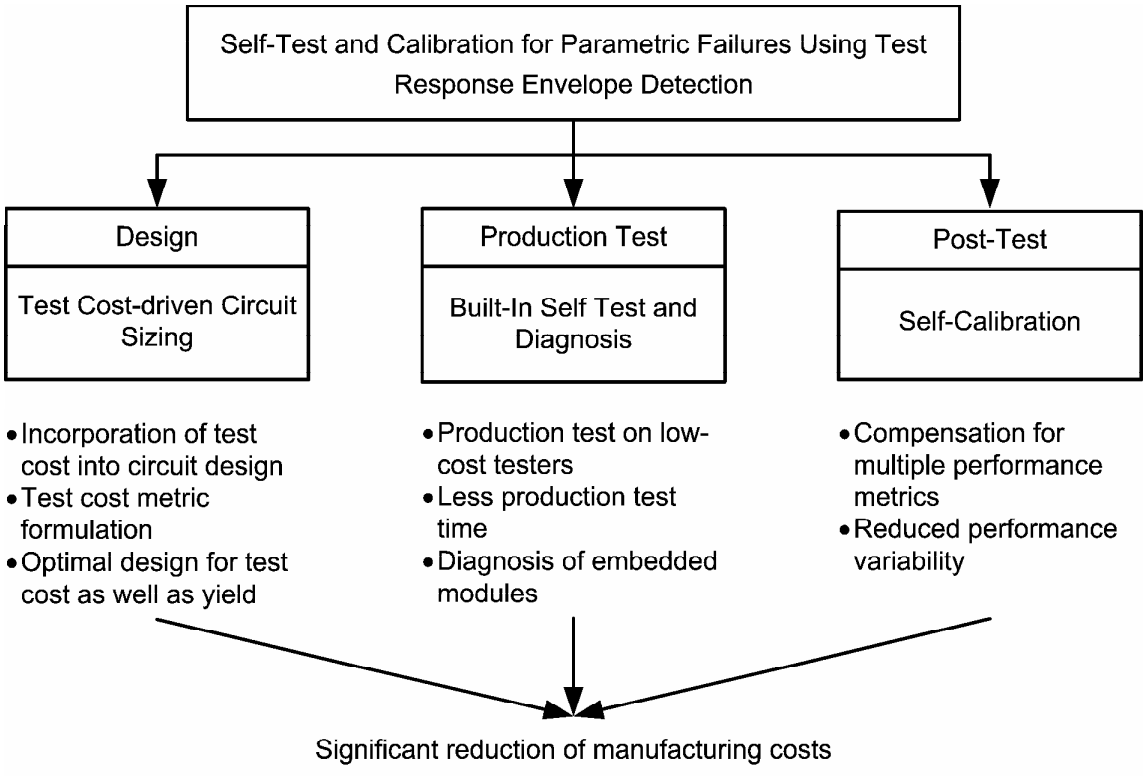


Figure 5. Highlighted contributions.

1.5 Overview of Contents

This dissertation is organized as follows. Chapter 2 presents a test cost metric and its application to circuit sizing. Chapter 3 shows the effectiveness of response surface modeling-based circuit sizing compared with a typical simulation-based method. In Chapter 4, a low-cost parametric test and diagnosis is described, and the validation of the method is provided with a case study of a 1.57GHz RF transceiver prototype. Chapter 5

introduces software-in-the-loop self-calibration for multi-performance variability, and implementation of a 1.9GHz LNA using TSMC 0.25um CMOS technology is also presented. Finally, Chapter 6 presents a summary of results and suggestions for future research.

CHAPTER 2

CIRCUIT SIZING TECHNIQUE INCORPORATING TEST COST METRICS

In this chapter, a circuit sizing methodology is discussed that *incorporates test considerations into automatic analog circuit synthesis/device resizing* algorithms. Currently, existing analog design optimization tools addressing the circuit synthesis and the device resizing problems primarily focus on performance-driven circuit design, i.e., automatic design methods to meet given target specifications, along with power/area minimization or yield maximization [43]-[50]. On the other hand, various production test techniques described in the previous chapter have been explored to reduce test cost, independent of circuit performance and yield. However, little research has been reported for *design optimization for test cost reduction* [43]. In principle, if a circuit can be designed to have lower test cost upfront without compromising (1) circuit performance, (2) yield, and (3) test coverage, the final manufactured circuit can be sold at a lower price per unit.

The rest of this chapter is organized as follows. Section 2 describes circuit sizing and optimization preliminaries. Section 3 presents the dynamic test elimination method and the proposed cost metric for evaluating test time is described in Section 4. In Section 5, circuit sizing for *simultaneous* test time reduction and yield enhancement is presented. Case studies for a two-stage CMOS op-amp and folded cascaded op-amp are presented

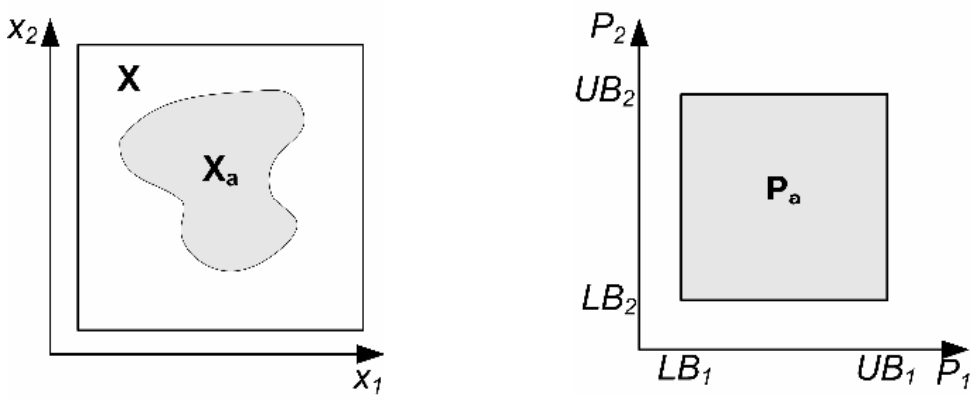
next. Simulation results showing the effectiveness of the proposed cost metric and circuit sizing data is presented in Section 6.

2.1 Preliminaries

In this section, first circuit sizing basics are presented. Then the core premise of the proposed approach is discussed.

2.1.1 Circuit Sizing and Test Cost Optimization

For a given circuit topology, circuit sizing is an optimization process that results in a design parameter set $\mathbf{X}_{opt}$ under which the circuit satisfies all the target specifications while maximizing manufacturing yield, minimizing power consumption and die area. In general, the design parameter set $\mathbf{X}$ includes transistor geometry dimensions, passive component values, and bias currents/voltages for transistor-level circuit sizing problems [43]. The concept of circuit sizing is depicted in Figure 6, where (a) shows the input space with two design variables and (b) shows the output space with the upper and lower bounds of two performance specifications p_1 and p_2. All input combinations $\{x_1, x_2\}$ in the subspace $\mathbf{X}_a$ of $\mathbf{X}$ in Figure 6 (a) map onto the rectangular space $\mathbf{P}_a$ of Figure 6 (b). Within the subspace $\mathbf{X}_a$, an optimum set of design parameters can be selected with consideration of other performance metrics such as power consumption, yield, and die area.



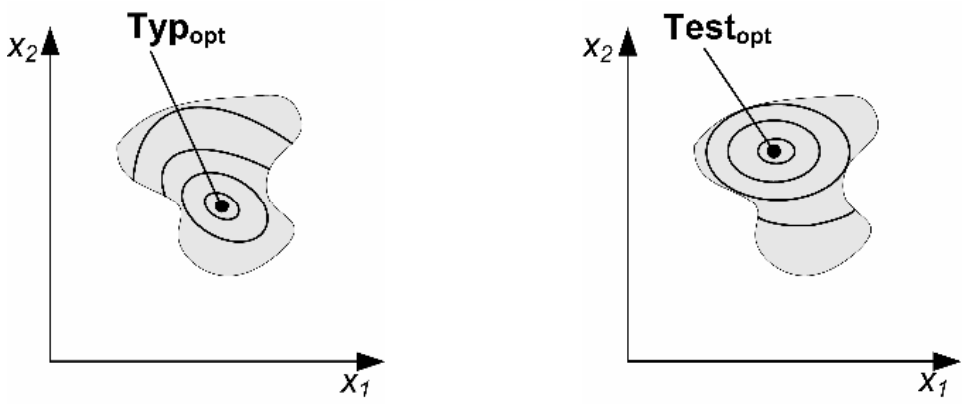
(a) **(b)**

Figure 6. Circuit sizing: (a) design variables and (b) circuit performance metrics.

Without loss of generality, this process can be converted into a single optimization problem and be stated as

find $x^* \in \mathbf{D}$ such that $\Psi(f(x^*)) \leq \Psi(f(x)), x \in \mathbf{D}$,

where $\mathbf{D} \subset \Re^n$ denotes the space of design variables, $f(\cdot)$ is a set of objective functions derived from performance specifications. The latter are computed from transistor-level SPICE simulation runs or closed form equations (if available). Typically, multiple objective functions $f(\cdot)$ can be converted into a single objective function in $\Psi(\cdot)$, such that the minimum or maximum value of the function corresponds to the optimum design for the given target specifications. The optimum design solution Typ_{opt} in X_a can be determined as shown in Figure 7. Each contour line in Figure 7 (a) (input variable/device sizing space) maps onto a contour line in Figure 7 (b) (performance and test cost space). The objective is to move across the contour lines of Figure 7 (a) via repeated cost metric evaluation such that the performance/test optimal point $Test_{opt}$ of Figure 7F (b) is reached.



(a) **(b)**

Figure 7. Contour plots of (a) typical circuit sizing and (b) test time.

2.1.2 Core Premise of Proposed Approach

From the literature on design centering for yield optimization [45], [47], [49], [50], it is clear that yield can be controlled through proper device resizing. As mentioned earlier, test cost and test time is also impacted by the same. Consider two circuit specifications P_1 and P_2 that are affected by a common set of device sizing parameters. If the common parameters dominate the statistics of the two circuit specifications then in all likelihood, these will be highly correlated as well. For example, open loop gain Av and common mode rejection ratio (CMRR) of a two-stage CMOS op-amp [6] are defined in terms of transconductance g_m and output resistance g_o of transistors , and given by Equations (1) and (2), below,

$$A_v = \left(\frac{g_{m2}}{g_{o2} + g_{o4}}\right)\left(\frac{g_{m7}}{g_{o6} + g_{o7}}\right) \quad \textbf{(1)}$$

$$CMRR = \frac{2g_{m2}g_{m4}}{(g_{o2} + g_{o4})g_{o5}} \quad \textbf{(2)}$$

where the index for g_m and g_0 denotes a particular transistor. Clearly the statistics of Av and CMRR will exhibit some degree of statistical correlation due to the presence of the common term $g_{m2}/(g_{02} + g_{o4})$ in Equation (1) and (2). The core premise of the proposed approach is based on the observation that the *statistical correlation between two or more specifications (such as Av and CMRR above), under specified manufacturing process variations can be further modulated (increased) by device sizing*. This, coupled with the fact that *increased specification correlation can be exploited to reduce test time and test cost* forms the core of our proposed synthesis-for-test approach. In the proposed device resizing methodology, test cost is directly impacted by device sizing.

2.2 Dynamic Test Elimination for Specification Testing

Dynamic test elimination exploits correlations between measured test specifications to reduce time. Measurements of a set of specifications (*estimator* specifications) are used to predict pass/fail values of other specifications (*estimated* specifications) with high confidence. If the values of the estimator specifications lie within predetermined ranges, then the estimated specifications can be predicted to be within a predetermined range as well. Hence, pass/fail analysis for the estimated specifications can be performed without explicitly measuring their values. If the values of the estimator specifications lie outside the predetermined ranges, then the values of the

estimated specifications need to be determined by explicit tests, which measure the respective specification values.

Dynamic test elimination relies on predictive subset testing [17], which has been used as a static test set compaction method for digital circuits. Predictive subset testing is based on the analysis of correlations among different specifications of the DUT. Suppose two test specifications of the DUT are highly correlated. Then one of the two specifications can be estimated from the other, assuming the other is measured accurately. Ideally, if two specifications are perfectly correlated, i.e., their linear correlation coefficient is 1 or -1, one can drop one of the two specifications from the test measurement procedure. However, in general, perfect correlation is implausible in real-life analog circuits. The resulting random error inherent in correlation driven estimation causes test loss called *overkill* [17]. Therefore, predictive subset testing is only applicable to circuits with strongly correlated specifications

To resolve the overkill, a test redundancy interval is determined for the *estimator* specification. If the estimator specification value is within this test redundancy interval, an explicit measurement of the estimated test specification need not be performed for pass/fail decision making. However, if the estimator specification value is *outside* the test redundancy interval, explicit measurements of the estimated test specification values must be made to generate a pass/fail decision for the DUT.

2.2.1 Test Redundancy Analysis

Test redundancy analysis is necessary for determining the test redundancy interval for the estimator specification. A large number of different *instances* of the DUT are logged during characterization testing or generated by varying the circuit process

parameters for given parameter statistics in simulation. For each instance, all the specifications of the DUT are measured and all possible pairs of specifications are selected for test redundancy analysis. For any selected specification pair, a polynomial regression model is generated to fit the scattered specification-pair data points (over different DUT instances) using least square error minimization. Such curves are shown in Figure 8 for the pair of specifications of an op-amp consisting of phase margin (vertical axis) and unity-gain bandwidth (horizontal axis) specifications.

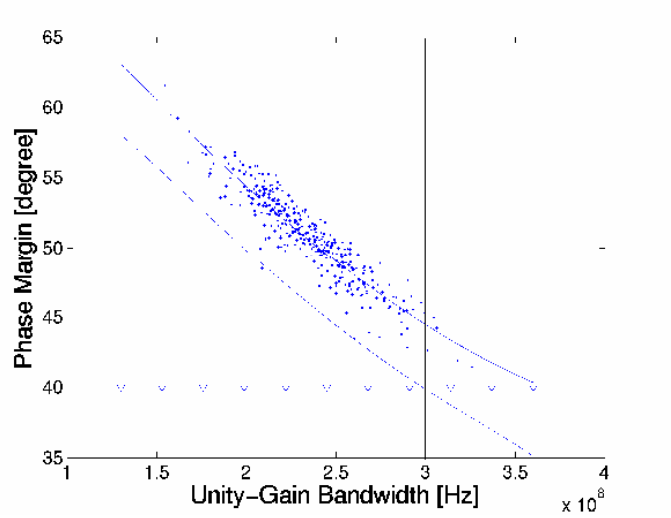


Figure 8. A pair of specification with polynomial fitting and its confidence bound.

Assuming that the errors for polynomial regression fitting are normally distributed with variance σ^2, a test redundancy confidence (TRC) value for the estimated specification value $\hat{p}$ based on the definition of confidence interval is defined. The true value of the estimated specification lies within the interval $[\hat{p}-n\sigma, \hat{p}+n\sigma]$ with a certain probability, where $\hat{p}$ denotes the estimated specification value, σ denotes the

standard deviation of the estimation error, and $\pm n\sigma$ denotes the TRC interval. For example, for n=3.3 the probability that the estimated specification lies within $[\hat{p} - n\sigma, \hat{p} + n\sigma]$ is 99.9%.

In the proposed approach, the test redundancy interval is defined as the region of the estimator specification corresponding to which one can make pass/fail decisions for the estimated specification. The test redundancy interval is determined by the degree of correlation between two specifications, distribution (i.e., mean and variance) of each specification and the test acceptance limit for each specification (two limits for two-sided specifications). In the example shown in Figure 8, the test redundancy computation procedure is shown for the phase margin and the unity-gain bandwidth frequency specifications of a two stage CMOS operational amplifier. Figure 8 shows the distribution of the specifications, a cubic fitting curve and its lower confidence bound corresponding to 99.9% TRC. Consider that the op-amp data sheet requires that a "good" circuit must have its phase margin equal to or above 40˚. Then, one can say from Figure 9 that the DUT will pass the phase margin (estimated specification) test if its unity-gain bandwidth (estimator specification) is measured to be below 300MHz. The test redundancy interval for the phase margin test can be equal to [-∞, 300MHz] over the unity-gain bandwidth specification, shown as the shaded area in Figure 9.

In contrast, a pair of uncorrelated specifications is shown in Figure 10, where the TRI is relatively wide. Hence, the unity-gain bandwidth specification always needs to be measured to determine pass/failure of the DUT. As can be inferred from two cases described above, TRI is mainly affected by the amount of correlation existing between two specifications, distribution shape of each specification, and the test specification bounds (i.e., acceptance limit for each specification).

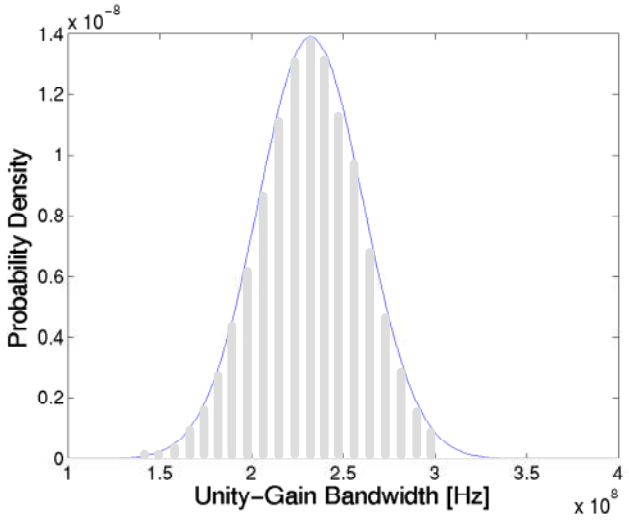


Figure 9. Test redundancy on the domain of estimator specification for the case shown in Figure 8.

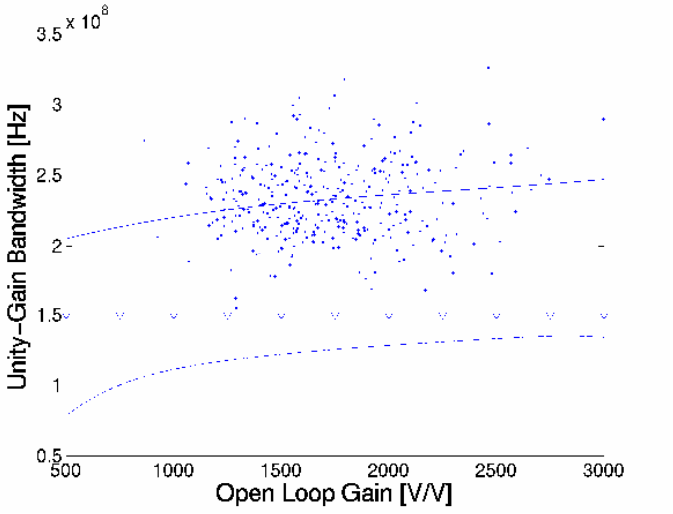


Figure 10. Uncorrelated pair of specifications.

2.2.2 Dynamic Test Elimination

The test redundancy interval helps in identifying the tests that are suitable for test elimination given that the measurement data for the estimator specifications are available from the applied test procedure. However, the actual test cost savings are not determined by the test redundancy interval alone. The test time for individual specification tests also affects the overall test time. Suppose a test set has n tests requiring test times of w_i, i=1,⋯,n, per test, then the average test time, t_{avg}, [10], in general, can be stated as

$$\begin{aligned} t_{avg} &= w_1 + w_2 Y_1 + w_3 Y_1 Y_2 + \cdots + w_n Y_1 Y_2 \cdots Y_{n-1} \\ &= w_1 + \sum_{i=2}^{n} w_i [\prod_{j}^{i-1} Y_j] \end{aligned} \quad \textbf{(3)}$$

where Y_j is the probability that the DUT passes the j-th test regardless of the previous test results. The first test in the test sequence needs to be performed. After that, each following tests are performed if the DUT pass all the previous tests. Otherwise, the DUT determined as a faulty instance can be dropped during the test sequence. For example, the average test time to perform the j-th test is $w_j Y_1 Y_2 \cdots Y_{j-1}$ where the term $Y_1 Y_2 \cdots Y_{j-1}$ implies the probability that the j-th test is performed. Combining the definition in Equation (3) with the test redundancy interval presented in this work, t_{avg} can be modified as

$$\begin{aligned} t_{avg} &= w_1 + w_2 Y_1 T_2 + w_3 Y_1 Y_2 T_3 + \cdots + w_n Y_1 Y_2 \cdots Y_{n-1} T_n \\ &= w_1 + \sum_{i=2}^{n} w_i [\prod_{j=1}^{i-1} Y_j] T_i \end{aligned} \quad \textbf{(4)}$$

where the proposed method introduces a probability factor T_i for test time reduction. The term T_i in Equation (4) represents the likelihood of performing real "measurements" for the *i*-th specification and can be described as

$$T_i = P(N_i \mid G_1 \cap \cdots \cap G_{i-1}) \quad \textbf{(5)}$$

where the term G_k denotes the instances that passes the *k*-th test in the test sequence and N_k is the instances for which the *k*-th test is not eliminated from the test set through test redundancy analysis based on the earlier specification measurements. Similarly, 1-T_i implies the probability that the *i*-th test is eliminated from the test set through test redundancy analysis based on the all the previous (1 … $(i\text{-}1)^{th}$) specification measurements. If the *i*-th specification is eliminated from any previous measurements, $P(N_i|G_1 \cap \ldots \cap G_{i-1})$ approaches 0. Also, if the *i*-th specification is uncorrelated with the 1st to (*i*-1)-the tests, $P(N_i|G_1 \cap \ldots \cap G_{i-1})$ approaches 1 and Equation (4) equals Equation (3). If TRI among specifications exists and then its value is below 1, the overall test time can be reduced by the factor of the probability T_i compared to the typical test scheduling in Equation (3).

2.3 Test Cost Metric

Any design sizing methods require one or more cost metrics related to target performances such as circuit specification performances, power/area, yield, etc., in terms of which the optimization algorithm attempts to search for an optimum. A typical structure of circuit sizing is mainly composed of optimization and cost evaluation procedure for the current instance. Depending on the type of the evaluator employed,

circuit sizing method can be divided into the simulation-based approach and equation-based approach. From the optimization perspective, global optimizer such as simulated annealing, genetic algorithm, etc. is in widespread as an optimizer to get a global optimum solution [43].

In the case of circuit sizing for test time reduction, the device sizing procedure is supposed to calculate and minimize average test time (t_{avg}) in Equation (4) in conjunction with typical cost metrics mentioned above. A possible circuit sizing for test time reduction as well as typical performance metrics can de depicted in Figure 11. As inputs, the target specification limits **P**, design variables **X** and their constraints, and a circuit topology can be included. To calculate the average test time for the current candidate, joint PDFs among the specifications of interest, first, need to be extracted via Monte-Carlo simulation, followed by dynamic test set compaction and test scheduling. This process is iterated until the current candidate meet all the specifications or the cost function defined converges to the minimum or maximum. The complexity of computing the average test time, for a set of n specification tests, is $O(n!)$, because $n!$ possible permutations of the test set exist. Also, for each permutation, the probability $P(N_i|G_1 \cap \ldots \cap G_{i-1})$ needs to be calculated. Furthermore, these computations are performed for each and every iteration in the device sizing algorithm, making calculation of t_{avg} infeasible for all practical purposes.

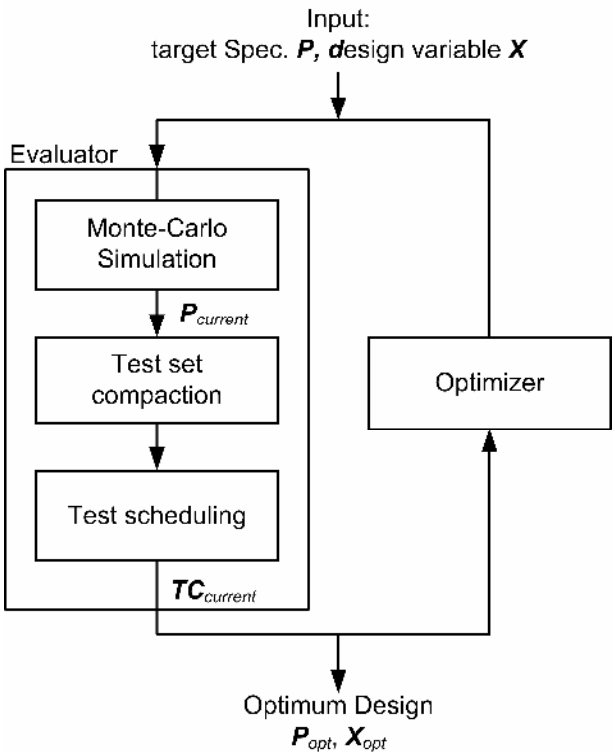


Figure 11. Exhaustive circuit sizing approach for test cost reduction.

To avoid the enormous computational cost expected in the approach shown in Figure 11, a heuristic cost metric is formulated to reduce computational cost of evaluating the average test time for the current candidate.

2.3.1 Generalized Test Redundancy

While a distribution for each specification is unknown without performing Monte-Carlo simulation, a variety of specification distributions can be well approximated by multivariate normal distribution. With this assumption, the linear non-homogeneous estimator is the best of all estimators when two specifications P_1 and P_2 are jointly Gaussian [51]. Then, the estimate $\hat{P}_2$ for the specification P_2 can be approximated by a

linear function of the observation P_1 with the constant values of correlation coefficient ρ_{12}, standard deviation σ_1, and mean η_1 of each specification, respectively. It can be given by

$$\hat{P}_2 = E(P_2 \mid P_1) = \frac{\rho_{12}\sigma_2}{\sigma_1}(P_1 - \eta_1) + \eta_2 \quad \textbf{(6)}$$

where $E(\cdot)$ denotes expectation [51]. Its mean-square-error (MSE) is

$$\varepsilon_{1,2}^2 = E[(P_2 - \hat{P}_2)^2] = \sigma_2{}^2(1 - \rho_{12}{}^2) \quad \textbf{(7)}$$

Then, test redundancy interval of the specification P_2 over the specification P_1 can be generalized in terms of a few variables mentioned above. Let LB_2 and UB_2 be the given lower and upper specification limit of the specification P_2. The test for the specification P_2 can be eliminated if the measurement of the specification P_1 is within the range of p_L and p_U shown in Figure 12, which correspond to the point $LB_2+n\varepsilon_{1,2}$ and $UB_2-n\varepsilon_{1,2}$ in the value of the estimator $\hat{P}_2$.

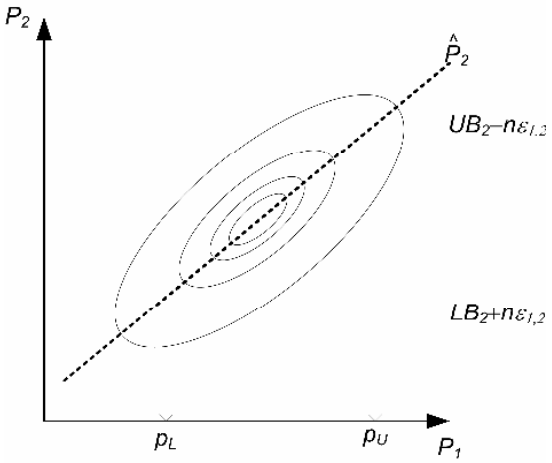


Figure 12. Contours of joint Gaussian PDF for two specifications.

The probability to be a redundant test of the specification P_2 over the observation P_1 can be generalized as

$$F_{1,2} = \int_{p_L}^{p_U} f(p_1)dp_1 \quad \textbf{(8)}$$

As a result, a test redundancy matrix **F** associated with all the specifications is derived from the definition in Equation (8) and given by

$$\mathbf{F} = [F_1, F_2, \cdots, F_n] = \begin{bmatrix} 0 & F_{12} & \cdots & F_{1n} \\ F_{21} & \ddots & & \vdots \\ \vdots & & & \\ F_{n1} & \cdots & & 0 \end{bmatrix} \quad \textbf{(9)}$$

where the element F_{ij} of **F** is the probability that the value of the *i*-th specification of a DUT instance is within the test redundancy interval corresponding to the *j*-th specification. Hence, F_{ij} represents the probability that pass or failure of the *j*-th specification can be determined from the knowledge of the *i*-th specification. On the other hand, the diagonal element F_{ii} of **F** is equal to zero.

2.3.2 Formulation of Test Cost Metric

The procedure to estimate the test time expressed in Equation (4) accommodates test scheduling as well as iterative calculations of the test redundancy matrix **F**. Though heuristic test scheduling methods [10], [19], [20], have been proposed and are applicable, the computational load can be significant as the number of test specifications increases.

With the test redundancy matrix **F** and the test time for each specification w_i, a test cost metric, *TC*, is defined as

$$TC = \sum_{i=1}^{n} w_i \left\| F_i \right\|_1 \quad \textbf{(10)}$$

where F_i=[F_{1i}, F_{2i} ,…,F_{ni}]. The test time w_i acts as a weighting factor in computing *TC*. *w*'s are estimated from the test-time profile for similar DUTs, and they remain constant during design optimization. Circuit sizing is driven such that *TC* is maximized and thereby larger test redundancy among the specifications exists with more weights on the specifications requiring longer test time. The simulation overhead for calculating *TC* is decided by the computation complexity of the covariance matrix among the specifications, described in Equation (7). In particular, for the problem of design centering or circuit sizing and synthesis for yield optimization [49], [50], statistical information regarding the specifications is already available as a part of the yield analysis. Therefore, computation of the covariance matrix and consequently computation of the *TC* introduces no additional simulation overhead for the current circuit candidate. It is obvious that the proposed cost metric can be directly imported into design centering or circuit sizing tools with no impact on circuit level simulation complexity.

2.4 Implementation of Test Cost-driven Circuit Sizing

In this section, a circuit sizing tool is presented, which automatically searches for the optimum design parameters with respect to test time as well as manufacturing yield of circuits. The overall conceptual diagram is depicted in Figure 13 that employs linear modeling to explore yield analysis, followed by evaluation of test cost metrics.

Note that there will be no limitation in applying the proposed test cost metric to other applications. For example, the cost metric can be applied to post-design optimization such as design centering for yield enhancement or incorporated with circuit synthesis tools.

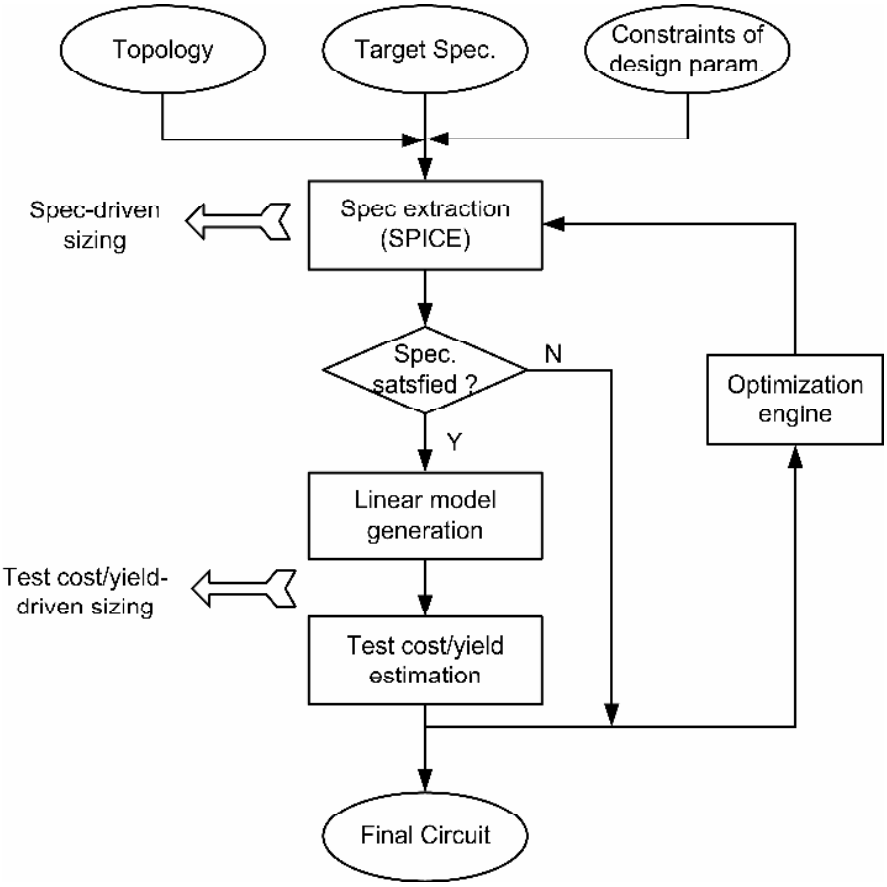


Figure 13. Circuit sizing procedure.

2.4.1 Statistical Analysis

Manufacturing yield of a circuit is the probability that the performance of a DUT, deviated from the nominal due to the process perturbation, meets the target performance metrics. The index C_{pk} present in [52] is used as a part of cost function in the optimization process, which is stated for one specification in terms of standard deviation σ, upper/lower specification limits, S^U, S^L, and nominal specification value $\bar{y}$ as

$$C_{pk} = \min\{\frac{S^U - \bar{y}}{6\sigma}, \frac{\bar{y} - S^L}{6\sigma}\} \quad \textbf{(11)}$$

C_{pk} is directly related to the yield of each specification assuming that the specification is normally distributed. For example, C_{pk}=0.00 implies 50% yield and C_{pk}=1.00 corresponds to 99.865% yield.

The required parameters for calculating C_{pk} are the standard deviation values of each specification. To extract the standard deviations, Monte Carlo simulation, in general, needs to be performed. However, considering its enormous computational cost, it is almost infeasible to use Monte Carlo simulation on circuit sizing tools. As an alternative to Monte-Carlo simulation, linear and quadratic statistical models in [53], are utilized to keep the computational cost low. Let **X** be a *p*-dimensional process and design parameters with mean $E[\mathbf{X}]$ and covariance matrix $D[\mathbf{X}]$. The input-output relationship in the linear model can be expressed as **Y=CX** where **C** is a constant matrix and **Y** denotes the performance specifications of interest. Specification distribution then can be characterized through $E[\mathbf{Y}] = \mathbf{C}E[\mathbf{X}]$ and $D[\mathbf{Y}]=\mathbf{C}D[\mathbf{X}]\mathbf{C}'$ for the linear model. Detail descriptions including quadratic statistical models can be found in [49], [53].

It is noted that yield optimization process requires the standard deviations of each specification. Also, the correlation coefficients as well as standard deviations are the required parameters for the proposed test cost metric in Equation (6) and (8), which can be calculated without additional computational cost using $E[\mathbf{Y}]$ and covariance matrix $D[\mathbf{Y}]$ (i.e., correlation coefficient $\rho_{ij} = D_{ij} / \sqrt{D_{ii} D_{jj}}$). Though the procedure in Figure 13 employs linear models to extract statistical information, there will be no limitation on use of any methods. It should be stressed that the test cost metric reuses information obtained in yield analysis. Therefore, test time and yield-driven circuit sizing can be performed with the computational cost just for yield optimization.

2.4.2 Circuit Sizing Steps

In general, manual circuit design first concentrates on nominal design, and then subsequent optimization of circuits for statistical fluctuations is followed. This process is to alleviate the high computational cost for yield analysis and also comes from no explicit manual design approaches without statistical analysis. However, circuit synthesis and sizing tools that imitate this process can often produce a bad starting point for gradient-based post yield optimization, so the yield improvement is prone to fail [47]. This stems from the fact that sizing tools typically drive a circuit design at an edge of performance space. It is also expected in test cost-driven circuit sizing problems.

To alleviate this problem, and to overcome redundant and computationally expensive statistical analysis for yield and test cost analysis, circuit sizing is divided into two phases; specification-driven sizing and test time/yield-driven sizing. If the current candidate circuit does not meet the target specifications of interest, this state, regarded as

specification-driven sizing, only concentrates on the nominal design without statistical analysis. Otherwise, the state corresponds to test cost/yield driven sizing. In this phase, statistical analysis is performed to extract indices for theses two factors. Thus, the cost function for circuit sizing can be defined as

$$\Psi(x) = \begin{cases} \sum_i a_i f_i(x), \text{if specification-driven sizing} \\ \lambda_1 Yield + \lambda_2 TC, \text{otherwise} \end{cases} \quad \textbf{(12)}$$

where a_i and λ_i are used as weighting factors. Here, the maximum of cost function corresponds to the optimum solution for the given circuit topology and design parameter constraints with respect to the test cost and yield.

2.5 Experimental Results

In this section, goodness of fit of the proposed test cost metric and its application to circuit sizing is described through two case studies, for which a two-stage CMOS op-amp [6] and a folded-cascode op-amp [54], shown in Figure 14 and Figure 15, were employed.

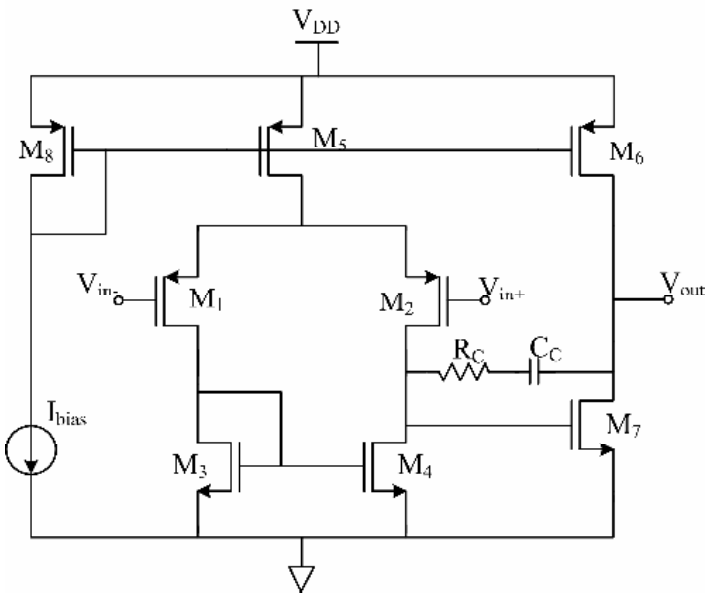


Figure 14. Two-stage CMOS op-amp.

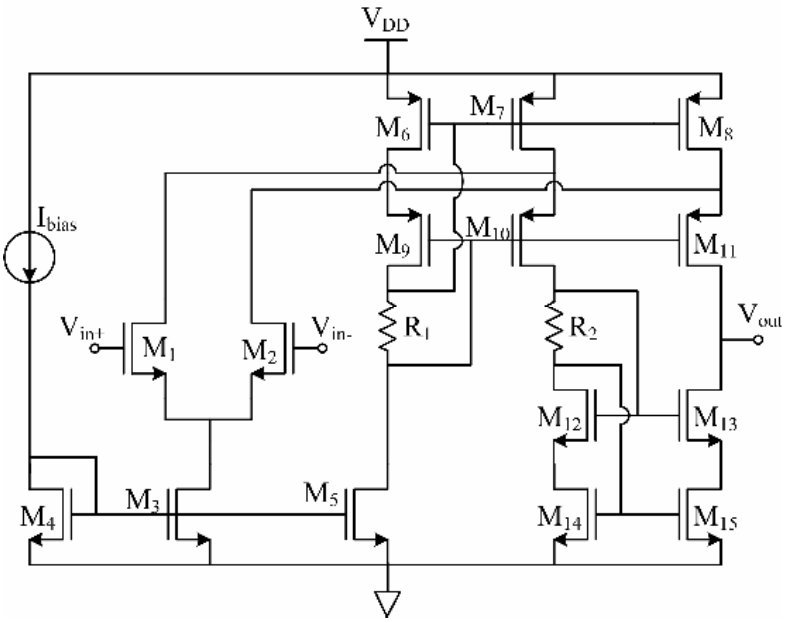


Figure 15. Folded-cascode op-amp.

During circuit sizing of these two circuits, a set of test specifications of interest were considered and are as follows:

- gain at DC (Av(0)),
- gain bandwidth (GB),
- phase-margin (PM),
- positive and negative slew-rate (SR+ and SR-),
- offset,
- equivalent input noise (EIN),
- positive and negative output voltage swing (OVS+ and OVS-),
- positive and negative power-supply rejection ratio (PSRR+ and PSRR-),
- common-mode rejection ratio (CMRR),
- quiescent power consumption (Pdiss).

To imitate process variations in reality, Monte-Carlos simulation was performed for the pre-defined statistics of the process variations. Due to the lack of statistical information of process parameter perturbations, perturbation parameters were assumed to be independent and normally distributed with zero mean and 5% standard deviation. The parameters perturbed are as followed:

- zero-bias threshold voltage of p-channel transistors
- zero-bias threshold voltage of n-channel transistors
- channel doping concentration
- low field mobility of p-channel transistors
- low field mobility of n-channel transistors
- drain and source diffusion sheet resistance of p-channel transistors

- drain and source diffusion sheet resistance of n-channel transistors
- length offset fitting parameter
- width offset fitting parameter
- resistors
- capacitors
- bias currents

2.5.1 Goodness of Fit of Test Cost Metric

To evaluate goodness of fit of the proposed test cost metric, the following experiments were performed for two different circuits with respect to various circuit performance statistics and test times for each specification:

1) a set of different op-amps for each type was designed to get various statistics including performance specifications and correlation degrees among specifications,
2) manufacturing yield values for the set of op amps were arbitrary adjusted above 1.5σ, corresponding to 93.32% manufacturing yield, representing the typical yield range for a high-yield manufacturing process,
3) 100 different sets of test time values, *w*'s, were generated, which were uniformly distributed, and normalized such that the sum of the test time for each set is unity,
4) each test time set was applied to the set of different op-amps to obtain the correlation degree between average test times and proposed test cost metric values. The average test time was extracted using the method (i.e., brute-force

search) shown in Figure 11, accompanying with Monte-Carlo simulations, test set compaction, and test scheduling.

The figure of merit is degree of correlation between averaged test times and test cost metric values. So higher the degree of correlation, higher the fitness of the proposed cost metric. Two examples for each type of op-amps are given in Figure 17 where each point in the plot corresponds to one op-amp instance at a particular set of test times. As can be seen, the test cost metric is inversely proportional to the normalized test time, and two case studies show correlation coefficient of -0.897 and -0.976, respectively. Figure 19 shows the histogram of the correlation degree of each op-amp with 100 different test time sets. The mean correlation values are -0.891 and -0.845, and the correlation values have standard deviation of 0.0459 and 0.0732, respectively. Based on the high correlation values obtained from the above experiments and the associated standard deviation degrees, it can be inferred that the proposed test cost metric is an effective measure for driving a circuit design for test time reduction for any arbitrary test time set. It can conclude that through exploiting the proposed test cost metric, one can determine in the design phase which circuit is better than the others from test cost point of view with less computational cost. In addition, the test redundancy can be estimated in the circuit design phase for evaluating the feasibility of test cost reduction.

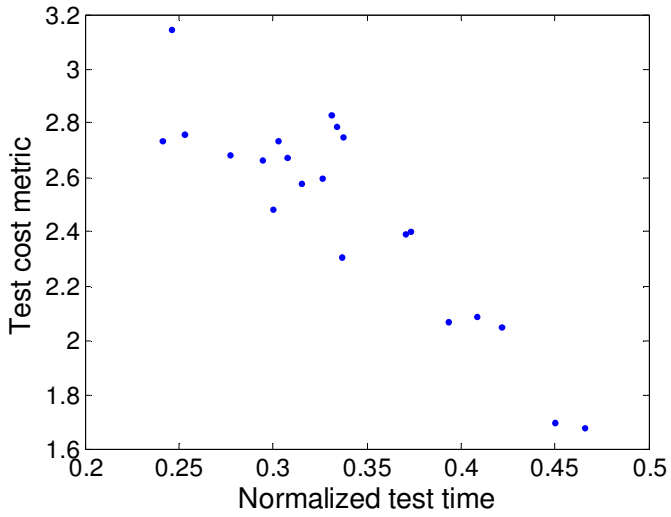


Figure 16. Test cost metric versus average test time of the two-stage op-amps.

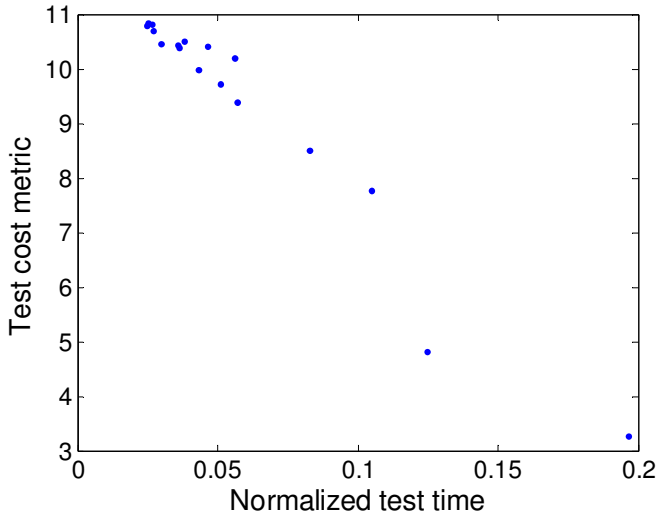


Figure 17. Test cost metric versus average test time of the folded-cascode op-amps.

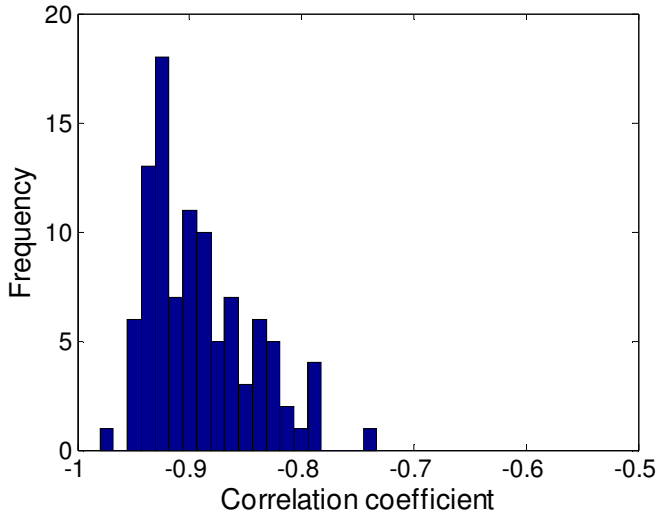


Figure 18. Histogram of correlation coefficients between average time and test cost metric for the two-stage op-amps.

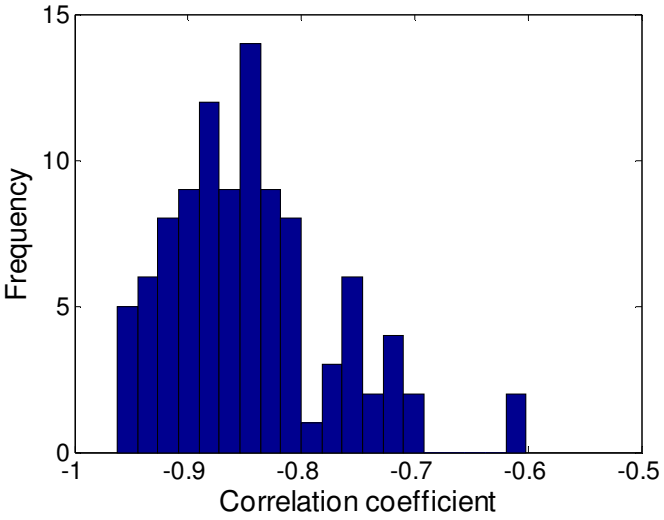


Figure 19. Histogram of correlation coefficients between average test time and test cost metric for the folded-cascode op-amps.

2.5.2 Circuit Sizing Application

To show the performance of circuit sizing in terms of test cost reduction and manufacturing yield, two op-amps shown in Figure 14 and Figure 15 were applied. The design parameters included bias currents, resistors, capacitors, and all the transistor widths while the transistor lengths were set to the minimum. The circuit sizing was processed in AMI CMOS 0.6μm technology and the simulations were done via Cadence SPECTRE with OCEAN scripts.

To compare the performances of the circuit sizing with respect to the average test time and yield, two cases were considered: 1) yield-driven circuit sizing, and 2) test cost and yield-driven sizing. The test cost and yield-driven circuit sizing was based on the method described in Section 2.4. For the yield-driven sizing, all the procedure was the same as that of the test cost/ yield-driven sizing except that the test cost metric was not evaluated and the optimization process was driven only by the estimated yield. The test times for all the specifications are assumed as listed in Table 1 for each op-amp. For example, the specification test for Av(0) requires 7 units, offset specification 2 units and so on.

Table 1 lists the test times and specification constraints as input parameters of circuit sizing and its results with all the specification values, yield, and average test time for two cases. All the specification values satisfy the target specification limits for both cases, where the final values including yield are similar to each other. However, as can be seen in Table 1, the final average test time for test cost and yield-driven sizing case shows better results (i.e., 14.48 units) with 45% reduction of test time and 0.57% of yield loss compared to the yield-driven circuit sizing. Table 2 lists the final values of the design parameters and design variable constraints employed in circuit sizing.

Table 3 and Table 4 show the circuit sizing constraints and corresponding results for the folded-cascode op-amp. All the specification values are similar to each other after circuit sizing. Similar to the results of the two-stage op-map shown in Table 1, the folded-cascode op-amp, incorporating test cost metric into circuit sizing, shows 36% test time reduction and 0.1% loss of yield compared to yield-driven circuit sizing case.

Table 1. Circuit sizing results for the two-stage op-amp.

Spec.	Test time	Spec. constraints	Yield-driven sizing	TC & yield-driven sizing
Av(0)	7	>67dB	69.4	69.8
GB	12	>18MHz	38.9	21.2
PM	14	>70°	74.9°	87.7°
Offset	2	<+/-4mV	-0.183	-0.313
SR+	11	>10V/µs	14.86	10.99
SR-	11	<-10V/µs	-15.79	-11.91
EIN	9	<14nV/$\sqrt{Hz}$	11.21	12.50
OVS+	2	>2.15V	2.26	2.23
OVS-	2	<-2.35V	-2.4	-2.4
PSRR+	8	>70dB	103.1	95.4
PSRR-	8	>70dB	76.1	76.4
CMRR	8	>70dB	74.4	74.4
P_{diss}	-	<5mW	3.37	3.78
Yield	-	-%	96.71	96.14
Avg. test time	-	-	26.41	14.48

Table 2. Design variables after circuit sizing for the two-stage op-amp.

Design variable	Constraints	Yield-driven sizing	TC & yield-driven sizing
$W_1(=W_2)$	[1, 150] µm	125.1 µm	101.6 µm
$W_3(=W_4)$	[1, 150] µm	12.3 µm	12.9 µm
W_5	[1, 150] µm	16.6 µm	12.3 µm
W_6	[1, 150] µm	93.9 µm	79.7 µm
W_7	[1, 150] µm	110.5 µm	143.4 µm
W_8	[1, 150] µm	70.2 µm	92.3 µm
R_C	[200, 5K] Ω	1 KΩ	1 KΩ
C_C	[200, 5K] Ω	2.05 pF	3.04 pF
I_{bias}	[10, 500] µA	241 µA	356µA

Table 3. Circuit sizing results for the folded-cascode op-amp.

Spec.	Test time	Spec. constraints	Yield-driven sizing	TC & yield-driven sizing
Av(0)	7	>60dB	64.9	64.7
GB	12	>20MHz	52.7	49.5
PM	14	>60°	67°	65°
Offset	2	<2mV	-0.848	-0.915
SR+	11	>6V/μs	16.05	15.17
SR-	11	<-6V/μs	-16.51	-15.66
EIN	9	10nV/$\sqrt{Hz}$	5.22	5.19
OVS+	2	>1.9V	2.08	2.02
OVS-	2	<-1.9V	-2.06	-2.16
PSRR+	8	>60dB	79.7	79.8
PSRR-	8	>60dB	68.7	68.61
CMRR	8	>66dB	87.9	85.9
P_{diss}	-	<5mW	2.00	1.96
Yield	-	-%	93.56	92.58
Avg. test time	-	-	5.21	3.31

Table 4. Design variables after circuit sizing for the folded-cascode op-amp.

Design variable	Constraints	Yield-driven sizing	TC & yield-driven sizing
$W_1(=W_2)$	[1, 150] μm	91.3 μm	75.5 μm
W_3	[1, 150] μm	39.2 μm	37.9 μm
W_4	[1, 150] μm	10.0 μm	8.5 μm
W_5	[1, 150] μm	22.0 μm	20.7 μm
$W_6(=W_7=W_8)$	[1, 150] μm	89.3 μm	55.2 μm
$W_9(=W_{10}=W_{11})$	[1, 150] μm	140.4 μm	121.4 μm
$W_{12}(=W_{13}=W_{14}=W_{15})$	[1, 150] μm	39.7 μm	144.8 μm
R_0	[200, 5K] Ω	1.69 KΩ	1.58 KΩ
R_1	[200, 5K] Ω	4.91 KΩ	4.39 KΩ
I_{bias}	[10, 100] μA	36.1μA	30.3μA

CHAPTER 3

ADAPTIVE RESPONSE SURFACE MODELING-BASED METHOD FOR CIRCUIT SIZING

In this chapter, a modeling-based circuit sizing method is presented that is capable of significantly reducing the computational cost of a circuit sizing process via adaptive response surface modeling. In the area of circuit synthesis and sizing, distinct research directions, based on evaluation techniques of a circuit, can be found in the literature [43] such as an equation-based approach and simulation-based method. The key advantages of the simulation-based approach are 1) to be applied to any circuit since simulation tools such as SPICE can be used, and 2) trustworthy to circuit designers with full accuracy of SPICE simulation. However, it incurs intensive computations due to iterative simulations in circuit sizing. For that reason, research in this area has focused on the reduction of simulation cost, which is based on traditional ideas such as task parallelization and knowledge-based algorithms.

Alternatively, response surface modeling-based methodology becomes a viable solution in applications such as circuit sizing and synthesis, which involve expensive evaluation of cost functions. Previously, modeling-based approaches have been proposed for the design of magnetic devices [55] and RF circuits [56]. The method for the design of magnetic devices is based on simulated annealing coupled with a number of expensive function evaluations, which increase exponentially with the number of input parameters.

Therefore, the method in [55] is only applicable to design problems having a smaller number of input parameters. The approach for RF circuits in [56] is based on the use of genetic algorithms with embedded SPICE simulations and the use of design knowledge that causes lack of generality.

3.1 Conceptual Structure

The basic idea of the proposed method is to use an accurate cost model to evaluate cost values within a limited region of the design space. Assuming that a modeling process occupies a fraction of the circuit sizing process in terms of the computational cost, the modeling-based method has a computational advantage by partially eliminating the use of expensive SPICE simulations. However, generating the cost model itself requires a number of SPICE runs. This can be more expensive than running multiple SPICE simulations to evaluate the cost function in circuit sizing problems. As a result, the modeling-based method is only favorable when (1) the number of samples to generate the cost model is much smaller than the number of function evaluations during the circuit sizing process, and (2) the quality of the final solution obtained is not affected by the use of the cost model.

The basic structure of the proposed algorithm that satisfies these conditions is shown in Figure 20. The proposed algorithm is based on the selective evaluation of the cost model, coupled with numerical SPICE simulations and the adaptive update of the cost model for accuracy. An effective sampling scheme utilizes two criteria. The first one provides sufficient samples for enhancing model accuracy, whereas the other prevents over-sampling of the design space after accuracy of the model is saturated.

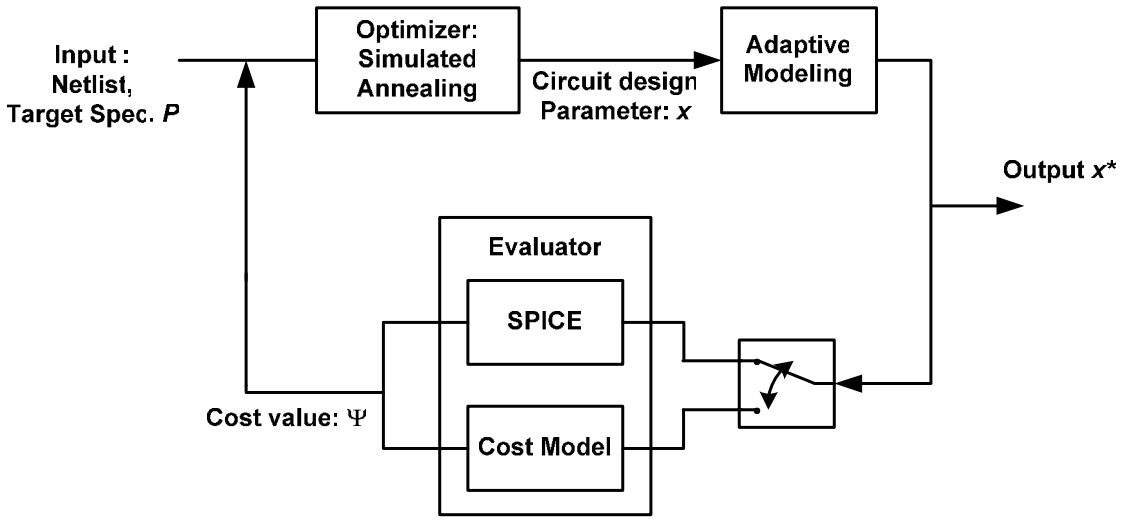


Figure 20. Modeling-based circuit sizing.

3.2 Adaptive Cost Function Modeling

To generate a model, the input and output space needs to be efficiently sampled for reducing the number of sampling without compromising model accuracy. Conventionally, there are several well-known methods for design space sampling, such as the fractional factorial design [58], Latin hypercube [59], and Taguchi methods [60]. These methods provide a sampling mechanism of the design space to map the input space onto the corresponding output space with a smaller number of samples. Note that there is a trade-off between accuracy and complexity of a model. Therefore, highly accurate and fully covered models require large sample sets. Furthermore, increased model complexity, coupled with large sample sets, incurs a computationally intensive modeling process.

Our goal is to perform as few SPICE runs as possible to obtain a cost model with the objective of using the model to drive a circuit design toward an optimal solution. For

this, it is not necessary to have an accurate cost model over the entire input and output domains. In this work, to increase model accuracy and also reduce modeling time, a local model that is valid only for the current candidate is made at every step. The proposed algorithm shown in Figure 21 uses the simulated annealing algorithm as an optimizer and the additional components, shown with stress in Figure 21.

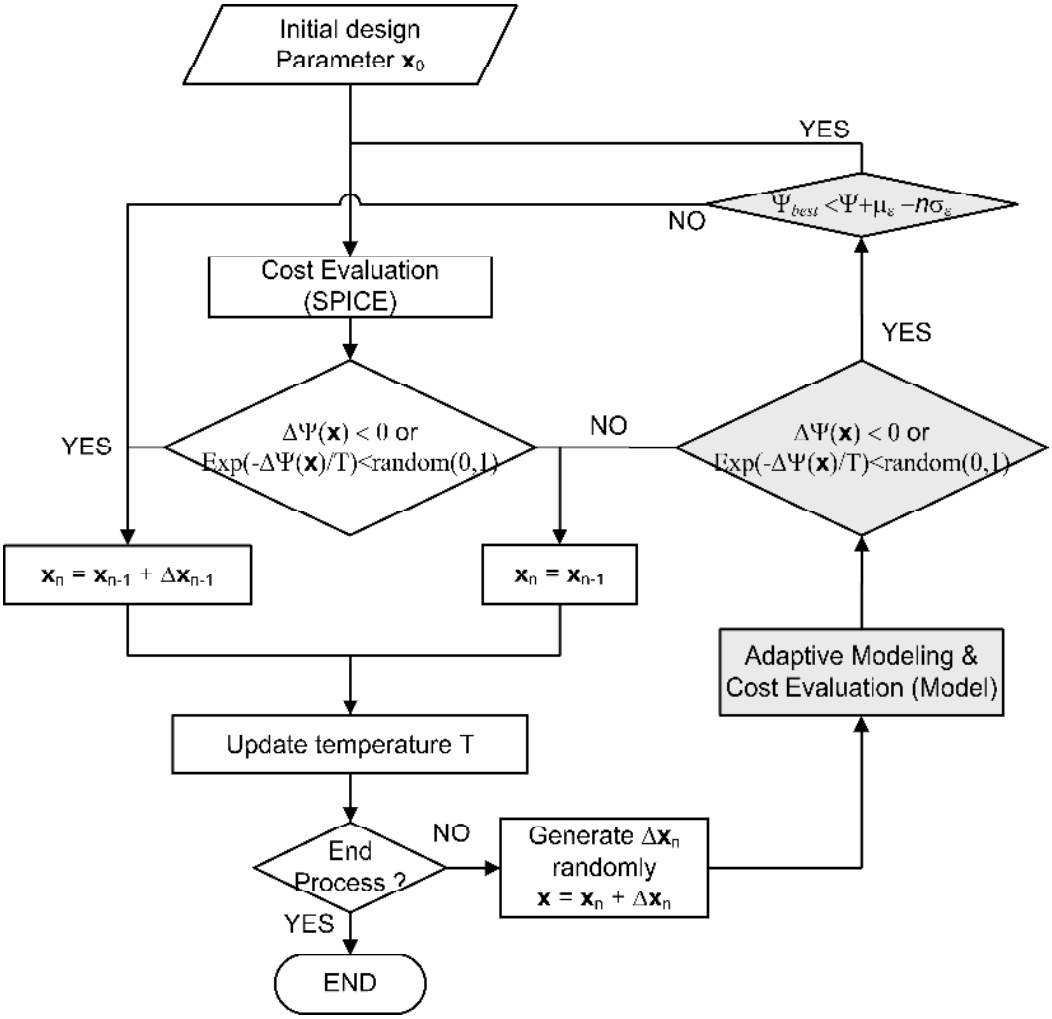


Figure 21. Adaptive sampling procedure.

Three components of adaptive modeling, and sampling criteria I and II, are added to the simulated annealing algorithm. At each step, interpolation using a local sample set around the current candidate can be done using a nonlinear mapping function. The key benefit of the local model is to use a smaller number of samples corresponding to the limited region. Considering that the smaller valid region of the model reduces the complexity of modeling compared to a global model, the local model has higher accuracy within the limited region. Furthermore, the required modeling time is also reduced. To sample the design space adaptively and achieve enough accuracy to converge to the optimum solution, two sampling criteria are employed in the proposed method.

3.2.1 Sampling Criterion I

The first criterion determines that a predicted cost value using the local cost model is acceptable. This process makes use of the statistical acceptance laws of simulated annealing, a statistical algorithm for a global optimization problem. Each step of the simulated annealing algorithm accepts the current candidate by a probability that depends on the difference between the corresponding function values, and a global parameter T called temperature. The temperature T is gradually decreased during the process. The dependency is such that the current candidate is accepted randomly when T is larger, but is increasingly downhill as T goes to zero. The allowance for uphill moves saves the algorithm from becoming stuck at local minima.

Before running SPICE simulations to get a cost value for the current candidate, the local cost model is evaluated to predict the cost value. If the predicted cost value is accepted from the acceptance law of the simulated annealing algorithm, the next sampling criterion is evaluated to consider whether SPICE simulation needs to be done.

Otherwise, the result obtained from the local cost model is accepted, and SPICE simulations can be avoided. At initial steps in the sizing process, T is high enough to gather global data over the entire space. As the process iterates and T is reduced, the sizing process focuses on a local space to converge an optimum solution. Hence, more sampling of the design space is done around the optimum solution. By contrast, sampling criterion I conducts sparse sampling of the other space. Finally, the local model can be accurate around the optimum solution with decent accuracy for the other design space. Note that this is only valid for an optimization problem where its response surface is smooth without deep valleys in the surface. After enough sampling, the cost model reaches a state in which it needs no additional samples for accuracy. To avoid over-sampling at the state, sampling criterion II is used.

3.2.2 Sampling Criterion II

The relations between the cost function Ψ and the model $\hat{\Psi}$ for the design parameters x can be stated as

$$\Psi(x) = \hat{\Psi}(x) + \varepsilon \tag{13}$$

where ε denotes the error between Ψ and $\hat{\Psi}$. Assuming that ε has a normal distribution with mean μ_ε and standard deviation σ_ε, the range of Ψ can be estimated from $\hat{\Psi}$ as

$$\hat{\Psi}(x) + \mu_\varepsilon - n\sigma_\varepsilon < \Psi(x) < \hat{\Psi}(x) + \mu_\varepsilon + n\sigma_\varepsilon \tag{14}$$

where n determines the confidence interval of Ψ. For example, the probability that Equation (14) is valid is 68% for n=1. Based on the current best cost value Ψ_{best} and the

statistics of the last M consecutive error values, the condition that needs SPICE simulation for the current candidate is defined as

$$\Psi_{best} > \hat{\Psi}(x) + \mu_{\varepsilon} - n\sigma_{\varepsilon} \quad \textbf{(15)}$$

The condition in Equation (15) implies that the current candidate can be a best candidate with a certain probability based on the last M consecutive predicted cost value, where the variable n determines the probability. If the current candidate has at least the specified probability to be a best candidate, SPICE simulation is performed. Otherwise, the result from the cost model is accepted. In the case that σ_{ε} is large at the initial state, cost evaluations are mainly determined by sampling criterion I. However, as the process iterates and sufficient samples are logged from SPICE simulations, sampling criterion II plays a major role in selecting a proper evaluator. Finally, when the cost model is accurate within the limited region of the design space, most of the function evaluations are done via the use of the cost model. Therefore, expensive SPICE simulations can often be avoided.

3.3 Numerical Results

The proposed method is tested on the Dixon-Szegö test functions [61]. The characteristics of each function are listed in Table 5. These test functions are not really expensive to evaluate but they share some important features with real cost functions. Hence, the relative performance of optimization algorithms on these test functions is expected to mimic performance on expensive functions with similar shapes.

Table 5. Characteristics of test functions.

Problem	Dim.	No. of local minima	No. of global minima	Domain
Branin	2	3	3	$[-5,10]\times[0,15]$
Hartman3	3	4	1	$[0,1]^3$
Hartman6	6	4	1	$[0,1]^6$
GP	2	4	1	$[-2,2]^2$

Table 6 shows the results of the proposed method with that of typical simulated annealing and differential evolution (DE). In the numerical evaluations, all optimization parameters such as initial temperature, cooling schedule, etc. are identical for the proposed method and simulated annealing. The results for DE in Table 6 are reported in [61].

Table 6. Average number of function evaluation.

Function	Proposed method	SA	DE
BR	66	471	1190
H3	133	219	476
H6	173	465	7220
GP	220	444	1018

The simulated annealing algorithm is based on Cauchy annealing. The sample size for the initial cost model is set to $20\times N$ and 3σ is used for sampling criteria II, where N is the input dimension of a function. The algorithms are stopped when the relative error $\left|f-f_{opt}\right|/\left|f_{opt}\right|$ becomes smaller than 1%, where f is the current cost value and f_{opt} is the global optimum value.

Assuming that function evaluation is expensive, the number of function evaluations implies overall time required to converge to an optimum solution. As can be seen in Table 6, the proposed method is two to seven times faster than simulated annealing. For example, the proposed method requires 66 function evaluations and is seven times faster than the result of simulated annealing for the test function BR. Also, the proposed method has much better performance in terms of convergence speed than does DE for all the test functions. Numerical experiments using the stopping criterion reported in [3] is also conducted to consider the case that the optimum cost value f_{opt} is unknown. Circuit sizing problems typically correspond to this case. Table 7 shows the number of function evaluations and the relative error of the proposed method and simulated annealing. The speedup in terms of function evaluation count is roughly between 4 and 45. Along with the speedup, final solutions are very accurate comparable to the results obtained with simulated annealing.

Table 7. Comparison with SA.

Function	Proposed method		SA	
	No.	Error	No.	Error
BR	337	$5.09 \cdot 10^{-6}$	6560	$2.54 \cdot 10^{-6}$
H3	842	$1.06 \cdot 10^{-5}$	4057	$7.20 \cdot 10^{-5}$
H6	926	0.0017	4189	0.0014
GP	383	$1.1 \cdot 10^{-5}$	18133	$0.9 \cdot 10^{-5}$

3.4 Practical Case Studies

As a practical case, two case studies are performed to evaluate the proposed method using a two-stage CMOS op-amp and fully differential class AB op-amp. The schematics of the circuits are depicted in Figure 22 and Figure 23. The objective of circuit sizing is to find an optimum set of design parameters to meet all the target specifications and minimize power consumption. For the case studies, the design parameter set includes transistor geometric dimensions, passive component values, and bias currents with all the transistor lengths set to the minimum. The circuit sizing process is done using Cadence SPECTRE simulator with BSIM3V3 transistor models. All the design parameter values after circuit sizing are shown in the figure as well.

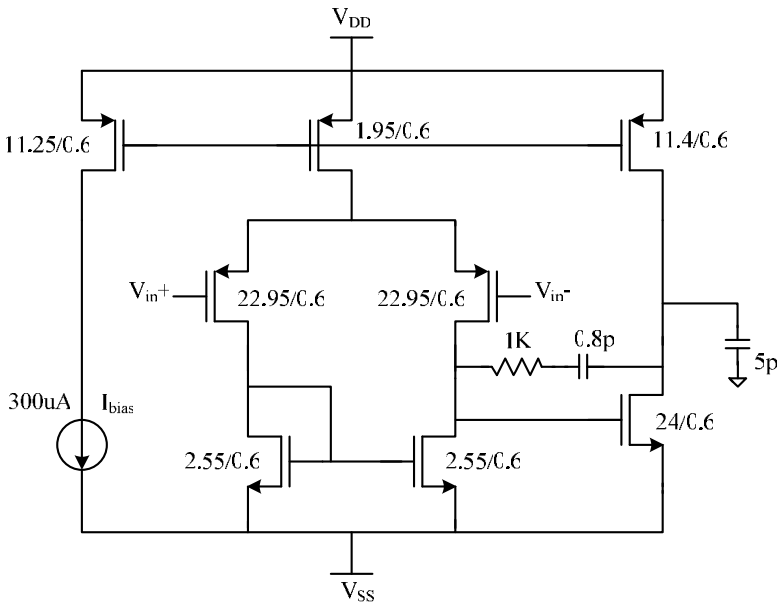


Figure 22. Two-stage CMOS op-amp with sizing results (width[μm]/length[μm]).

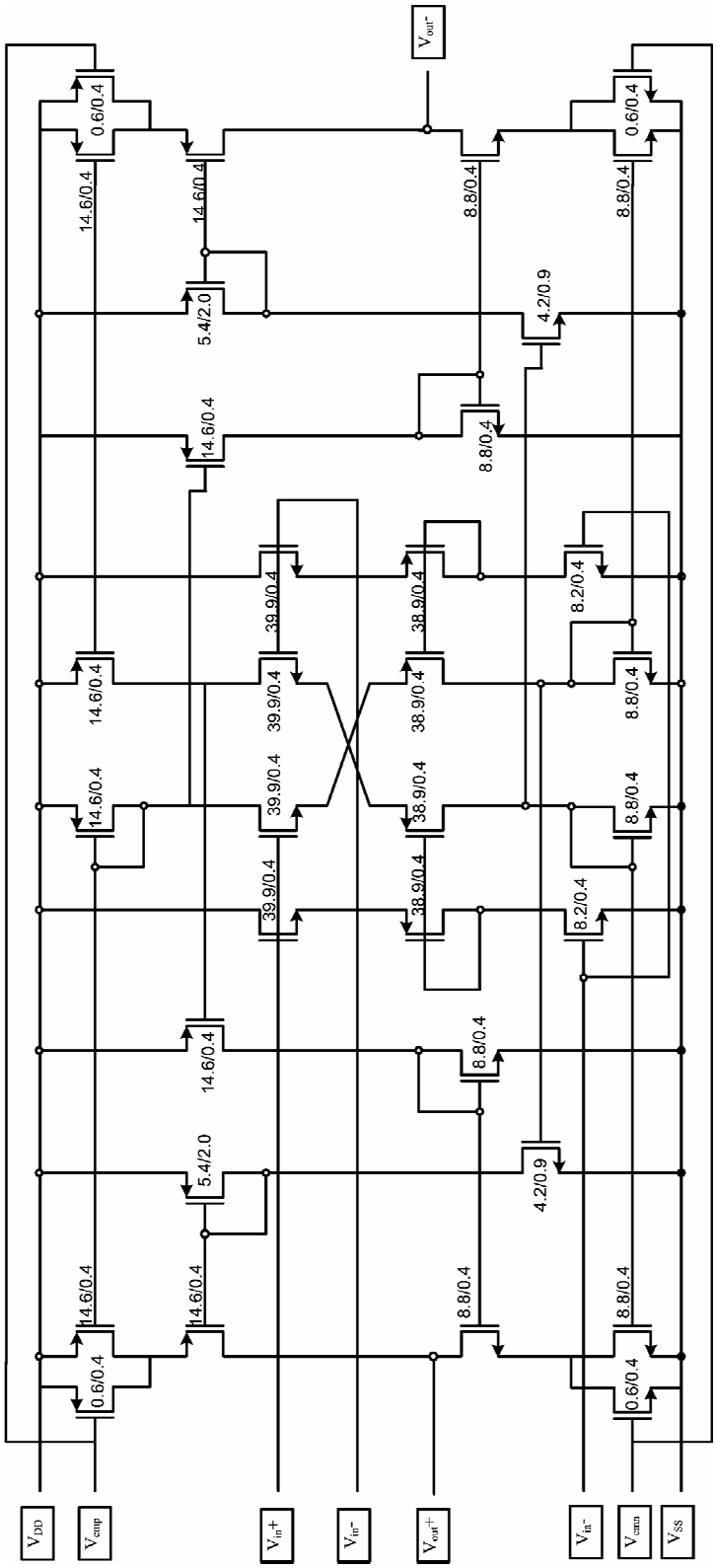


Figure 23. Fully-differential class AB op-amp with sizing results (width[μm]/length[μm]).

Table 8 shows the circuit sizing results for the given target specifications. The number of SPICE simulations for the two-stage CMOS op-amp is 1783, while the total iteration number is 7781. Assuming that the total number of iteration obtained from the proposed method is equal to that of the SPICE simulation-based circuit sizing, 77 % of SPICE simulations are avoided. Similarly, the case of the class AB op-amp shows 71% elimination of SPICE simulations.

Table 8. Circuit sizing results for two case studies.

	Two-stage op-amp		Class AB op-amp	
	Spec.	Result	Spec.	Result
Av	≥ 60dB	60.3dB	≥ 60dB	60dB
UGF	≥ 40MHz	40.2MHz	≥ 20MHz	28MHz
PM	≥ 60º	58.2º	≥ 70º	82º
SR	≥ 15 V/μs	15V/μs	≥ 160 V/μs	161V/μs
Output	≥ 4V	4.1V	≥ 6V	6.3V
Power	minimize	3.4mW	minimize	1.9mW
# simulation		1783		2855
# iteration		7781		9881

CHAPTER 4

LOW-COST PARAMETRIC TEST AND DIAGNOSIS OF RF SYSTEMS USING RESPONSE ENVELOPE DETECTION

To reduce production test costs, it is necessary to be able to measure all the test specifications using a low cost external test system since the test cost for RF circuitry is dominated by expensive automatic test equipment (ATE). In this work, we propose a low cost test and diagnosis scheme for integrated RF systems in which small sensors are *designed-into* the ATE load board or the device itself to facilitate manufacturing test and diagnosis. When the sensors are designed-into the ATE load board, very accurate performance testing and diagnosis of RF systems is possible with minimal impact on the performance of the device. In addition, using measurements made on the observable system outputs, it is possible to *predict* the performances of the embedded modules (LNA, mixer, PA) fairly accurately. When the sensors are designed into the RF circuit itself, more accurate test and diagnosis is possible for the embedded RF modules in addition to the capabilities already available for the RF systems. In both situations, it is possible to perform test and diagnosis of the RF front end with little or no support from an external tester via the software running on the transceiver baseband processor.

In the following, the objectives and approach of this work are presented. Next, prior work on test and diagnosis of RF circuits and systems is discussed. A theory for test and diagnosis of RF systems using RF sensors is then developed. This is followed by a

discussion of the use of sensors on the load board and sensors designed into the RF circuit under test. The pros and cons of both techniques from a design, test and calibration perspective are discussed. Finally, experimental results are used to show the practical viability, potential and usefulness of the proposed test and diagnosis methodology.

4.1 Production Tests for RF ICs

Although testing and failure diagnosis of analog circuits has been a major field of research and is well established [64]-[67], most of the prior research has focused on analog/mixed-signal systems as opposed to RF test and diagnosis. The core problem with RF test is that, high frequency (multi-GHz) signals need to be applied to the DUT and observed for test and diagnosis purposes, thereby requiring the use of a high-speed external tester and test access to embedded RF modules. Due to increasing circuit speeds and high levels of device integration, this has become increasingly difficult and expensive to perform in a high-volume manufacturing environment. In addition, where on-chip test access is possible, the electrical losses involved in transport high frequency signals from the chip to the external tester have made accurate test measurement a very hard problem to solve.

RF test equipment such as spectrum analyzers makes use of highly accurate mixers, frequency synthesizers, filters and power detectors for accurate RF measurements. While it is difficult to replicate such accurate measurement circuitry on a load board, several load board test structures have been proposed to down-convert RF signals to DC values for measuring RF circuit specifications such as gain, NF, IIP3, ACPR, and phase

noise [15]. The approach in [27] employs mixers for up- and down-conversion of the test stimulus and its response, respectively, hence eliminating the need for RF signal handling at the external tester, and providing a mechanism for extracting multiple RF test specifications using *alternate test* methodology. The work of [15], [27], however, only considers testing of discrete RF circuits. In addition, since the load board test circuitry (i.e., mixer, frequency synthesizer, etc) is too complex to be deployed inside a DUT, the work is not appropriate for performing embedded circuit testing.

Authors in [70] have looked at failure diagnosis of RF circuits for catastrophic faults. The approach, however, lacks a general method to determine fault models and does need RF test equipment to perform such tests. The diagnosis method proposed in [71] attempts to isolate and classify parametric and catastrophic failures in embedded RF circuits. The authors use a series of specification measurements via standard RF circuit test techniques to enable failure diagnosis. Even though a high probability of correct parametric and catastrophic RF fault identification via behavioral simulations in MATLAB is demonstrated, the method does not resolve the problem of test and diagnosis of parametric failures under *simultaneous multi-parameter perturbations*. Moreover, to perform a set of complicated RF specification tests for diagnosis, multiple test configurations and expensive ATE are necessary, resulting in long test times and high cost for failure diagnosis. Loop-back based transceiver diagnosis methods have also been proposed in [72], [73]. Pseudo-random bit sequences were used as test stimulus and the test data was used to extract the specifications of the transmitter and receiver. However, the problem of determining the test specifications of the RF system components from the observed response of the transmitter and receiver was not explored.

In this context, the use of embedded sensors that are designed-into the RF DUT for test purposes is a viable method for accessing internal RF nodes. Several built-in test (BIT) schemes have been proposed in the past that rely on the use of embedded peak, RMS, and power detectors for testing discrete LNA modules and RF transceivers [74]-[78]. The main limitations of these methods are that they either require the use of dedicated tests geared towards a few targeted RF test specifications and specific RF devices, or do not deliver the test measurement accuracy necessary for detection of parametric failures. Moreover, all the BIT methods in utilize the DC output of the detectors. To overcome the limited amount of information from a single DC value extracted by a detector, they deploy multiple detectors at an internal node and/or apply multiple test stimuli, thus incurring area/power/performance design penalty due to the complex designs of the sensors.

4.2 Basics: Envelope Detection Based Response Sensors

In this section, we describe how information can be extracted from the envelope of a test response and its implementation. For this purpose, we consider three types of test stimuli employed in RF tests and analyze the response of an ideal envelope detector to each of those.

4.2.1 Functionality of Envelope Detector

4.2.1.1 Envelope of a multi-tone sinusoid

The envelope of a multi-tone sinusoid can be defined in a closed form [79]. Supposed that the output waveform $y(t)$ of a DUT is in the form

$$y(t) = \sum_n c_n \cos(w_n t + \varphi_n) \tag{16}$$

and $y(t)$ is then derived with respect to a frequency w_q called the "midband frequency"

$$y(t) = I_c \cos(w_q t) - I_s \sin(w_q t)\,,$$

where

$$I_c = \sum_n c_n \cos[(w_n - w_q t) + \varphi_n] \tag{17}$$

$$I_s = \sum_n c_n \sin[(w_n - w_q t) + \varphi_n]$$

Then, the envelope $g(t)$ for $y(t)$ can be formulated by Rice's formulation as

$$g(t) = [{I_c}^2 + {I_s}^2]^{1/2} \tag{18}$$

For a memoryless nonlinear system, the input-output relationship can be approximated with a polynomial [80]. For simplicity, our analysis is limited up to the 3rd order term with the assumption that the higher terms can be negligible. Then, a system can be expressed with input $x(t)$ and constant a_i as

$$y(t) \approx a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) \tag{19}$$

In Equation (19), for example, a_1 is directly related to the gain of the system and input third order intercept point (IIP3) as a measure of DUT nonlinearity can be expressed as

$$IIP3 = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|} \quad \textbf{(20)}$$

4.2.1.2 Test response envelope analysis

Based on the envelope analysis and DUT definition described above, one can extract the envelope of a signal. First, consider the response of the DUT for a single-tone sinusoidal waveform, which is employed in standard tests to extract the specification gain and P_{1dB} and also utilized in the BIT schemes of [75]-[77]. If $x(t) = A\cos(w_c t)$, the resulting envelope of the DUT response is

$$g(t) = (a_1 A + \frac{3}{4} a_3 A^3) \approx a_1 A \quad \textbf{(21)}$$

where the frequency term at the RF input tone and its harmonic terms are assumed to be filtered out since RF detectors output a DC value or a signal at much lower frequencies compared to the operating frequency. In general, we can assume that $A<1$ and $a_1>>a_3$ considering the limited input dynamic range and degree of nonlinearity for RF circuits. Consequently, the term a_3A^3 in Equation (21) can be negligible and the equation can be approximated in terms of the coefficient a_1 of the system and the input signal amplitude A. If a test sensor is deployed at the output of a DUT, one can extract only the coefficient a_1 of the system. Hence, as addressed in [77], these approaches can detect only the specification gain and P_{1dB}, which are directly related to the coefficient a_1.

Consider the case of a two-tone sinusoidal test stimulus and is

$$x(t) = A\cos[(w_c - w_b/2)t] + A\cos[(w_c + w_b/2)t] \quad \textbf{(22)}$$

which is a typical test stimulus to measure the specification IIP3 and gain of the system. Then, the DUT outputs

$$\begin{aligned} y(t) &= b_1\cos[(w_c - w_b/2)t] + b_1\cos[(w_c + w_b/2)t] \\ &\quad + b_2\cos[(w_c - w_b 3/2)t] + b_2\cos[(w_c + w_b 3/2)t] \end{aligned} \quad \textbf{(23)}$$

where a constant b_k is defined in terms of the system coefficient a_k and the input signal amplitude A. In Equation (23), the coefficient b_1 is directly related to the DUT gain when the amplitude of $x(t)$ is adjusted in such a way that inter-modulation products are negligible. The coefficient b_2 is the amplitude of the third-order inter-modulation (IM) product. In standard tests, the coefficient b_1 and b_2 need to be measured to extract the specification gain and IIP3. The envelope of the test response in Equation (23) is written as

$$g(t) = \left| 2b_1\cos(\frac{1}{2}w_b t) + 2b_2\cos(\frac{3}{2}w_b t) \right| \quad \textbf{(24)}$$

Assuming that the frequency ω_c is the center frequency of the DUT and ω_b is at a much lower frequency (e.g., ω_c at 2.4GHz and ω_b at 10kHz), the envelope is a function of the coefficient b_1 and b_2, which can be measured at the much lower frequencies, thereby being able to replace RF signal measurements with that of the envelope.

For an AM test stimulus, the DUT outputs

$$
\begin{aligned}
y(t) &= b_0 \cos[w_c t] \\
&\quad + b_1 \cos[(w_c - w_b)t] + b_1 \cos[(w_c + w_b)t] \\
&\quad + b_2 \cos[(w_c - 2w_b)t] + b_2 \cos[(w_c + 2w_b)t]
\end{aligned}
\tag{25}
$$

where ω_c is a carrier frequency and ω_b corresponds to the modulated base-band signal frequency in AM scheme. Note that we consider only the main frequency tones and the other small amplitude tones are neglected for simplicity. The output of the envelope detector is

$$
g(t) = | b_0 + 2b_1 \cos(w_b t) + 2b_2 \cos(2w_b t) | \tag{26}
$$

As can be seen in Equation (26), the envelope of the at-speed test response is placed at much lower frequencies compared to the frequency ω_c.

In summary, Figure 24 depicts the envelope spectrum for the three test stimuli analyzed above. It can be observed from the figure that the extracted envelope has the characteristics of the input signal at the much lower frequencies (i.e., $w_b << w_c$). In addition, the down-conversion functionality of the envelope detector can be observed in Figure 24 (c). Even though the amplitude at each frequency tone in Figure 24 (b) is not directly corresponding to that of the input signal, each coefficient is derived in terms of the coefficient b_1 and b_2 of the input signal. Consequently, one can say that the envelope of a particular signal provides down-converted characteristics of the input signal. However, note that it is not valid for any arbitrary input signals. For example, frequency-shift keying (FSK) signals have no time-varying envelope. Any characteristics of the signal can not be extracted at the lower frequency band through envelope detection.

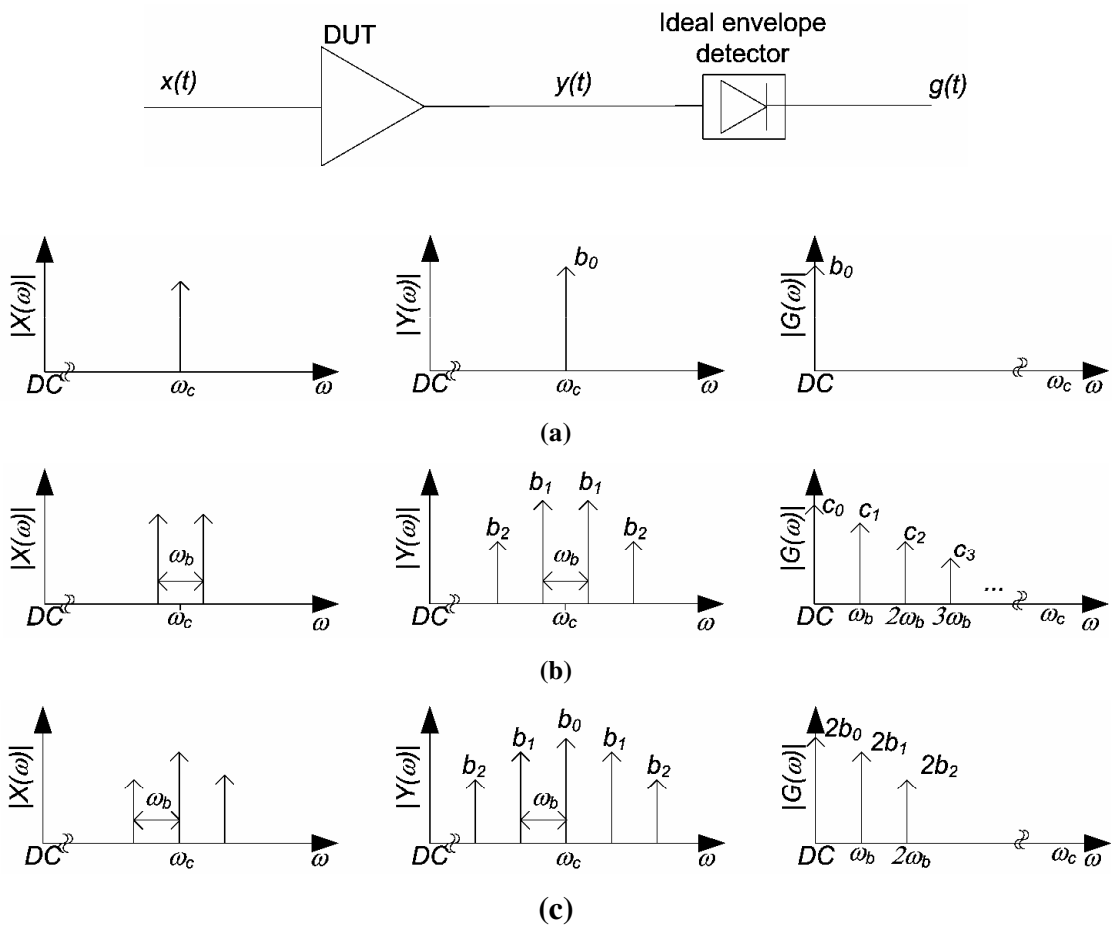


Figure 24. Envelope for each test response of (a) single-tone sinusoid, (b) two-tone sinusoid, and (c) AM signal.

4.2.2 Hardware Scheme

The envelope detector is a well known circuit for AM demodulation, composed of a diode, a resistor, and a capacitor. Through proper adjustment of the output *RC* constant value, envelope detection can be performed.

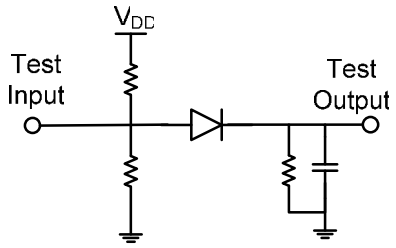


Figure 25. Simplified schematic of the envelope detector.

It is essential to make the deployed BIT circuitry less sensitive to process variations, thereby reduce impact on the quality of the test performed. The envelope detector performance is mainly determined by the *RC* time constant of the circuit shown in Figure 25, which is designed to detect the envelope of the signal and filter out the RF frequency signals. The value of the *RC* time constant can be set to be $f_o << 1/RC << f_c$, where f_o is the frequency of the envelope and f_c is the carrier frequency. Considering that the above two frequencies have large separation, the *RC* time constant can be picked to make the decoded envelope less sensitive to process variations relating to the R and C. Assuming that both the small signal resistance r_d and parasitic capacitance of the diode are much less than the values of *R* and *C*, the variations in the diode will have much less impact on the overall performance of the envelope detector.

The input impedance of the envelope detector depends mainly on the bias resistors and the capacitance of the diode. The bias resistors are relatively large compared to the typical 50Ω RF matching impedance. During the normal operating mode, we assume that the power of the envelope detector is turned off using a switch or the output of the envelope detector is connected to the power supply, thereby deactivating the envelope detector. In this case, the diode behaves like an open switch. Moreover, the input capacitance of a diode is several tens of femto-Farads for on-chip

implementation. Therefore, the input impedance of the envelope detector has negligible effect (loading) on the RF DUT output and its overall performance.

In summary, the BIT scheme shown in Figure 25 has the following advantages: 1) low die area overhead due to the simple circuit structure, 2) robustness to process variations, and 3) negligible impact on DUT performances.

4.3 Alternate Test Approach Using Test Response Sensors

In the proposed test method, the alternate test methodology [23]-[28] is used to predict the system and sub-module specifications from the extracted envelope of the test response. Alternate test works on the principle of finding a set of alternate test measurements **M** (different from and much more simpler than conventional specification tests) such that the set of multi-parameter process perturbations **S,** which cause device parametric failure (i.e., directly impact the set of performance metrics **P** of the device under test), also affect the alternate test measurements **M** made on the DUT. If the measurements **M** and device performance metrics **P** show strong statistical correlation under arbitrary process perturbations ***S***, then a non-linear regression-based mapping function f:**M**→**P** can be built for predicting the specifications **P** from the measurements **M.** Moreover, the original specification tests can be fully or partially replaced by the alternate test measurements if strong statistical correlation between **M** and **P** can be established. The key benefit of the alternate test methodology is that a single test configuration and a single test stimulus can be used to compute all the test specifications of interest, thus reducing the complexity of implementing truly autonomous built-in self-test.

Practical implementation of the alternate test procedure [23] involves time domain sampling of the obtained test response, signal processing (digital) of the obtained data and prediction of the performance metrics **P** of the device under test using regression-based mapping functions. For RF circuits, time domain sampling of the output response is difficult due to the very high sampling rates required. In the proposed test approach, the output of the envelope detector is a low frequency signal that can be easily digitized and analyzed by the on-board processor or embedded resources. The alternate test methodology with the test sensor utilized in this work is depicted in Figure 26. Let $\mathbf{M}'$ be the sampled time-domain signal at the output of the envelope detector. Then, the proposed approach relies on the use of a mapping functions f':$\mathbf{M}' \rightarrow \mathbf{P}$, for predicting the specifications **P** of the device under test.

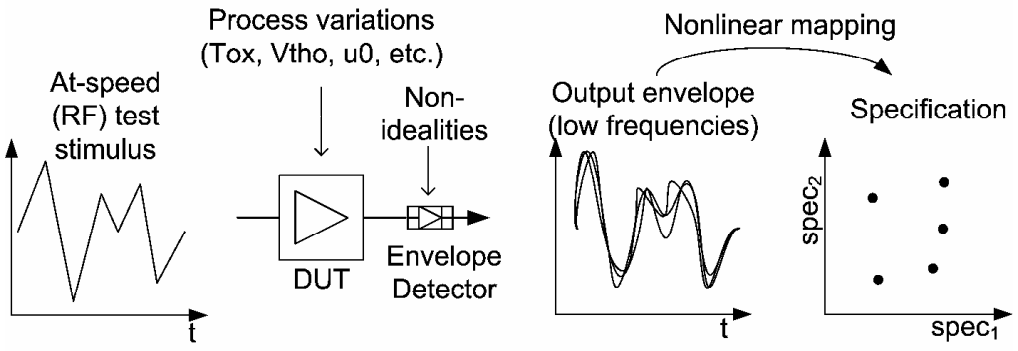


Figure 26. Alternate test methodology with a test sensor.

In general, the non-idealities inherent in the envelope detector need to be accommodated by the alternate test response analysis procedure. For example, a diode or diode-connected transistor in the envelope detector shows non-ideal switching behavior for the applied voltage over the diode and works in the linear or square-law region of its

operating characteristics depending on the input signal level. Furthermore, since the diode is a non-linear device, inter-modulation products are introduced into the detected envelope, thereby causing distortions. These non-idealities can, in general, be subsumed by the mapping function f' if the calibration procedure (discussed later) used to build the mapping f' is modified appropriately. This allows the accuracy of the alternate test procedure to be maintained even when the diode of Figure 26 is non-ideal.

4.4 Test Stimulus Optimization

Test stimulus optimization methods presented in the past literature have been mainly driven by iterative transistor-level simulations [81]. However, since simulation for RF transceivers is computationally expensive, their direct use in RF testing is difficult. In this work, the optimization process for diagnostic test stimuli is based on the concepts present in [24], where behavioral models are employed to generate a test stimulus for the receiver. This procedure relies on the fact that even though a test response obtained from the behavioral models is not identical to that of the actual devices, the measurements obtained are correlated to the measurements made on the same. For this reason, a test stimulus can be optimized based on behavioral models of the actual DUT. Finally, in order to compensate for the difference between behavioral models and actual devices, fine tuning of the optimized stimulus and calibration of the test measurements needs to be performed via hardware experiments.

The goal of the diagnostic stimulus optimization is to determine an optimal stimulus, which maximizes the statistical correlation between the measurements **M** and

the test specifications **P** as described in Section 4.3. We utilize a multi-tone sinusoidal waveform $x_i(t)$ as the test stimulus waveform,

$$x_i(t) = \sum_{k=1}^{N} V_k \sin(2\pi f_k t + \theta_k) \quad \textbf{(27)}$$

where the optimization process searches for the optimum V_k, f_k, and the number of different tones N. Since phase is not controlled by the test procedure, it is assumed to be zero for all tones.

In the optimization, the ranges of the test stimulus variables are determined by the characteristics of the DUT and external tester (i.e., data acquisition system). The maximum level of V_k is limited by the absolute maximum rating of the DUT's input, whereas the minimum input level is determined by the DAC's performance metrics – bit resolution and maximum output swing. The frequency set $\{f_k\}$ determines the period of the test response's envelope. As the period is increased, the data acquisition time increases and the bandwidth requirement of the ADC reduces. Therefore, the frequency bounds for $\{f_k\}$ can be set by the allowable maximum testing time and digitizing ADC's bandwidth.

The test stimulus optimization procedure is shown in Figure 27 and works as follows. Consider the set of test specifications **TS** = $\{T_1, T_2, \ldots, T_N\}$ for which an alternate test stimulus needs to be determined. Using statistical sampling techniques [82], a set of behavioral model parameter perturbations **BP** is created in such a way that all the target specifications in the set **TS** are perturbed from their nominal values to their test limits (i.e. the set **BP** contains as many test corners as possible) and the entire test specification parameter space is sampled as uniformly as possible.

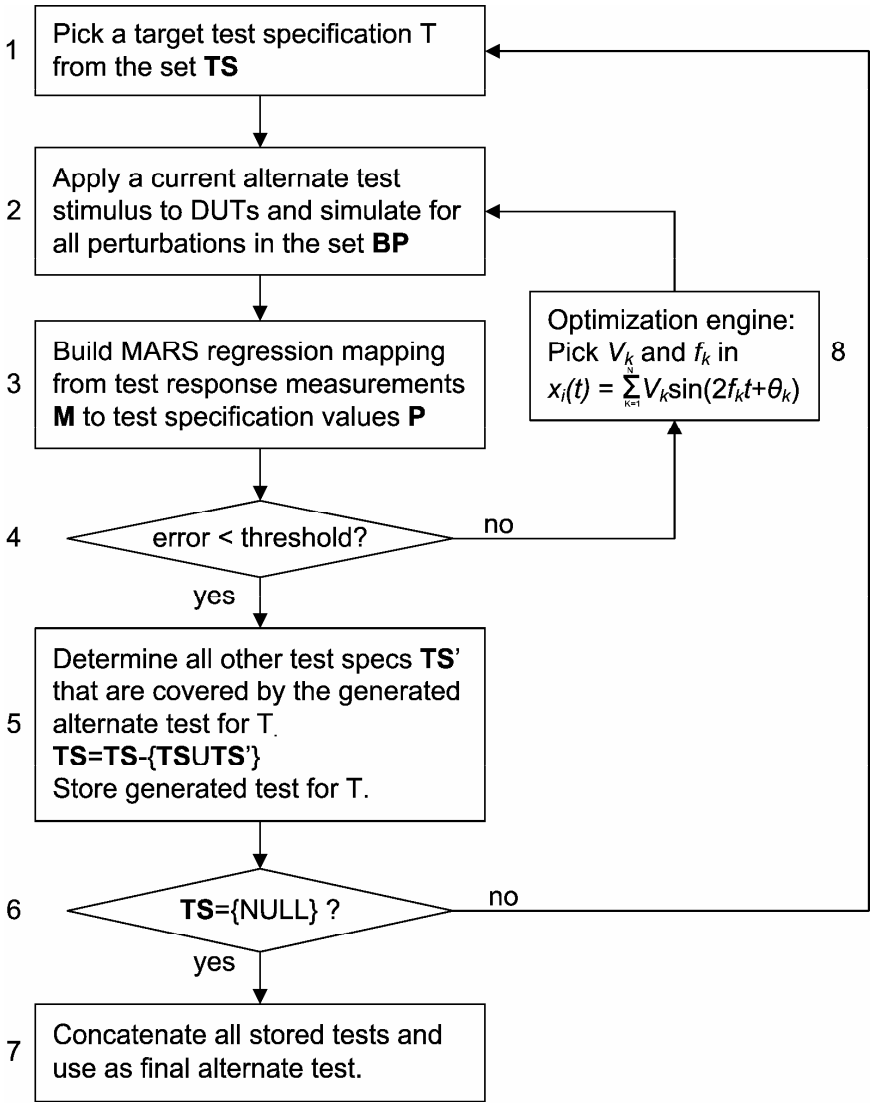


Figure 27. Diagnostic test stimulus optimization.

First, a target specification T is selected from the set of specifications **P**. The selection heuristic picks a test specification that shows *maximum correlation* with all other test specifications in the set **P** under random behavioral model parameter perturbations. Next, a set of test tone amplitudes V_k and frequencies f_k are selected by the optimization engine and applied to DUT models (called instances) corresponding to each

of the parameter perturbations in the set **BP** and simulation is performed. The resulting measurement response set **M** is stored. Each response in the set **M** consists of the digitized envelope obtained at the output of the envelope detector. The complete response set **M** consists of a set of such digitized responses. For each of the parameter perturbations in **BP**, the value of the target specification T is determined a priori using simulation. This results in a corresponding set of test specification values **P**. Note that the cardinality of the set **M** and the set **P** is the same as the cardinality of the set **BP**. Also, for every DUT instance, there exists a digitized envelope in **M** and a corresponding value of the test specification T in **P**.

Next, a MARS regression mapping [30] is built from **M** to **P** using the data stored in the prior step. The accuracy of this mapping is then checked as follows. A set of DUT instances corresponding to random and extreme behavioral parameter perturbations is created and their test specification values are measured (in simulation). Then the selected test stimulus is applied to each DUT instance and the test specification value T corresponding to each such instance is predicted from the MARS model built earlier. The prediction error of the predicted test specification values is then computed from the data obtained (since the predicted value of T and the exact value of the same for each DUT instance are known, this can be computed easily). If the prediction error is larger than a prescribed threshold, the test is modified using the optimization engine. If not, then test generation for T is successful and in the next step, the suitability of the obtained test for predicting other test specification values is determined. This involves running steps 1, 2, 3 and 4 in Figure 27 for all remaining test specs in the set **TS**.

In step 5, the set **TS** is modified to include only those test specifications for which the current set of stored alternate tests is not sufficient. If the resulting set **TS** is not

{NULL}, then the process is repeated until alternate tests for the remaining test specifications in **TS** are found and stored. Else, the process is stopped and all the generated alternate tests are concatenated (i.e. applied in sequence) to become the final alternate test for the RF circuit.

In general, an RF circuit may have detectors inserted into more than one RF test node. In this case, each element of **M** consists of not just a *single* digitized envelope response but an *ensemble* of such responses. The MARS function, in this case, maps the outputs of all the envelope detectors (digitized envelope responses) to the target test specification T. Another key issue is that the MARS mapping functions must be constructed from hardware measurements for accuracy. Hence, measurements made on a large number of DUTs with different performance metrics (instead of DUT instances in simulation as described earlier) are necessary in order to properly calibrate the alternate test. Where embedded detectors are concerned, their outputs must be made externally observable on a test chips for calibration purposes. While this is tedious, it is a one-time cost that is necessary for accuracy of the test procedure.

4.5 Test and Diagnosis Framework

The overall test and diagnosis framework is illustrated with a RF transceiver in Figure 28, where RF- loopback in the RF path is employed to feed the transmitter output signal to the receiver, thereby eliminating the need for an external RF signal generator to test the receiver. An attenuator is deployed in the load board to adjust the signal power fed to the receiver by the transmitter. If necessary, a frequency shifter can also be used for frequency translation from transmitter frequency to receiver frequency.

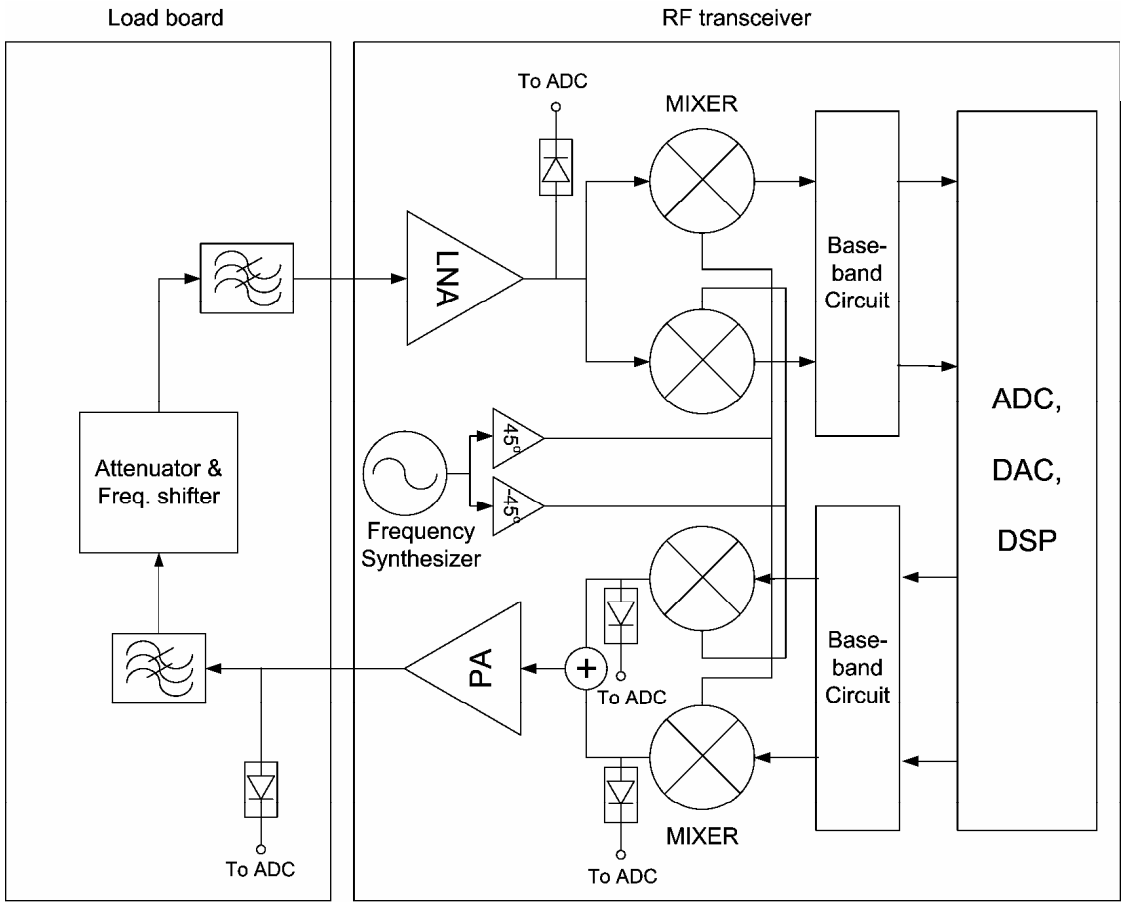


Figure 28. Proposed test and diagnosis structure.

Of the two possible built-in test configurations, one configuration uses a single sensor on the load board connected to the output of the transmitter (called '*external sensor-based method*') and the other uses sensors inside the die for increased test accuracy (called *'embedded sensor-based method'*). Figure 28 shows detectors placed at various DUT RF nodes.

In the first test configuration, only the detector at the output of the transmitter is used and the other sensors are unnecessary. A single test pattern is applied to the input of the mixers via the embedded DAC or DAC on the load board. The output of the envelope

detector placed on the load board is digitized using an embedded or external low sampling rate ADC. To test the transceiver, test response from the load-board envelope detector and the digitized baseband receiver output is used to predict the transmitter as well as the receiver test specification values. Calibration of the alternate test procedure is easily performed by measurements made on a set of transceivers with diverse test specification values. The test specification values of the embedded transceiver modules, such as mixers, low noise amplifier (LNA), and power amplifier (PA), are predicted from the measured test data but cannot be validated directly without access to the LNA and mixer outputs. However, once calibration is performed, no RF test instruments are necessary and the built-in test procedures can be used to predict the transceiver test specification values with great accuracy.

In the second test configuration, data from all the sensors are used to predict the transmitter, receiver as well as the individual module specifications (LNA, mixer, PA). It is assumed that for calibration, the respective module outputs are externally available on a set of test chips. While the accuracy of prediction of the transmitter and receiver specifications is the *same* as for the first test configuration, the increased test access allows very accurate prediction of the embedded RF module (LNA, mixer, PA) test specification values as well.

Note that even in the case of the first test configuration, accurate prediction of the transmitter and receiver test specifications values is not straightforward. This is because any problems with the transmitter can cause incorrect signals to be fed to the receiver (it is assumed that both the transmitter and receiver and any number of embedded RF modules can *simultaneously suffer from parametric defects/variations*). This can cause problems with receiver failure diagnosis. However, by analyzing data from the detector at

the output of the transmitter and the output of the receiver *simultaneously*, both the input and output of the receiver are monitored concurrently. This allows the effects of improper input signals to the receiver due to transmitter failure to be detected and compensated by the MARS regression model building procedure. Also, transmitter failures are diagnosed from the *same* data. The same arguments can be extended for the case of second test configuration where detectors are placed at the quadrature mixer outputs that feed the PA and a detector is placed at the output of the LNA that feeds the down-conversion mixers of Figure 28.

4.5.1 Post-Test Response Analysis

For both the configurations, the input to the RF system is a baseband signal applied to the up-conversion mixer that is compatible with an embedded DAC or low sampling rate DAC on the load board. Once a test stimulus is applied, the envelope of a test response can be observed via the test sensor(s), and digitized using the ADCs. To compute the specifications of the sub-modules and the system, the digitized transient outputs from the envelope detectors are fed into pre-built regression models. The envelope of the RF test response, however, is sampled in the presence of noise, which can significantly degrade the computational accuracy of the model evaluation. Therefore, a de-noising step is added during post processing on the DSP. First, the transient signal is sampled for multiple periods and time averaged. After that, one period of the transient envelope of each sensor is extracted as the final input parameter for the pre-built regression models.

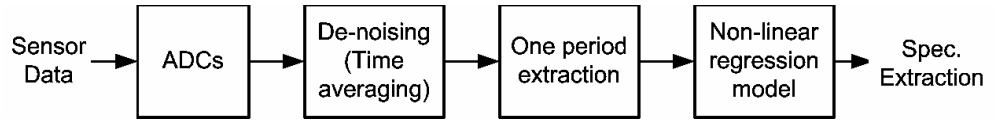


Figure 29. Post-processing steps performed on the envelope response.

4.5.2 Comparison of Two Schemes

Both the configurations can eliminate the need of RF equipment to capture RF test responses, thus significantly reduces equipment capital costs. However, the pros and cons of both techniques are different from a design, test quality, and test time perspective. The embedded sensors in the transceiver in Figure 28 involve additional die area, resulting in increased circuit design time and power consumption to incorporate sensors into the die. Moreover, the embedded sensors are also under the same process variations as the DUT, and its performance is varied across dies and wafers. To avoid the performance deviations of test sensors and corresponding test quality degradation, test sensors need to be tested and calibrated before performing DUT tests. For this purpose, a few calibration methods [77], [78] have been proposed, but are only applicable with the assumption that die-to-die variations are considerably dominant compared to within-die variations. However, this assumption is not valid as the feature size of transistors continues to shrink and thereby within-die variations keep increasing. On the contrary, since an external test sensor on the load board can be fully calibrated before DUT testing and its design is separated with the DUT design, the cons of the design and calibration issues for the embedded sensor-based method can be avoided.

On the other hand, the embedded sensor-based method provides a way to probe internal nodes, thus providing more information regarding embedded RF modules. To

achieve the same test quality as that, the external sensor-based method should utilize longer test pattern to extract similar amount of information. Otherwise, the test quality would be degraded significantly. Assuming that the one period of the envelope for the embedded sensor-based technique is 10 μsec, corresponding to 100 kHz signal, and 100 periods of the extracted envelope from a RF test response are sampled for de-noising, the electrical test time can then be estimated as the total measured duration (i.e., 1msec) plus the post-processing time depicted in Figure 29. Suppose that the external sensor-based method employs *n* times longer test pattern than that of the embedded sensor-based method to gather the same amount of information, the overall test time roughly increases by *n* times.

4.6 Experimental Results

In this section, the results from three case studies are described. The proposed test approach has been validated using (1) transistor-level simulation, (2) behavioral simulation, and (3) a hardware prototype of a 1.575GHz transceiver.

As a figure-of-merit for evaluation, root-mean-square (RMS) error was calculated with respect to the true specification values of interest. This is expressed as

$$RMS\ error = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{true,i} - P_{estimated,i})^2} \quad \textbf{(28)}$$

where N is the total instances used for validation. In Equation (28), $P_{ture,i}$ and $P_{estimated,i}$ are the true value and extracted specification value for the i^{th} specification using the proposed method, respectively.

4.6.1 Case Study I: 900MHz LNA

The efficiency of the proposed test structure was evaluated using a 900 MHz low noise amplifier (LNA) [83] shown in Figure 30. The specifications of interest are the power gain (S21), the noise figure (NF), and IIP3. Besides prediction accuracy as a figure-of-merit, the BIT circuit effects on the DUT performance metrics are also evaluated.

In order to make nonlinear mapping functions from the measurement space (transient envelope) to the specification space, 600 LNA instances were generated using Monte-Carlo simulation, which is performed in Cadence SpectreRF simulator. Multivariate adaptive regression splines (MARS) is utilized to make a nonlinear mapping [30]. The process variations in the Monte-Carlo simulation include the value of passive components (i.e., resistor and capacitor), and the BJT model parameters such as saturation current (I_s), forward current gain (β_f), forward early voltage (V_{af}), base resistance (r_b), and current corner for beta (I_{kf}). The parameter variations were assumed to be uniformly distributed within +/- 20% range around their nominal values. Note that the envelope detector employed inside the circuit is also under the same process variations.

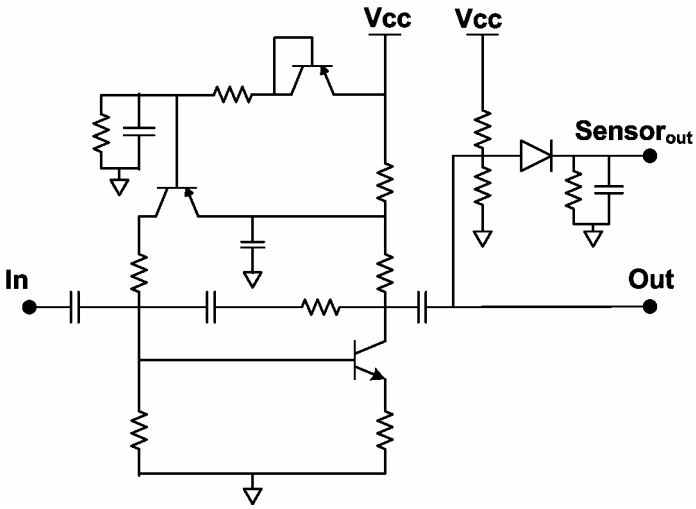


Figure 30. A 900MHz LNA.

Through the Monte-Carlo simulation, the specification values and the detected envelopes via the envelope detector were logged for 600 LNA instances. Among them, 500 instances were used to build the mapping functions and 100 instances were used to validate the accuracy of specification prediction. As test stimulus, two-tone signal was utilized and had 900MHz and 900.1MHz frequency with -10dBm power level. To digitize the detected envelopes, a 100MHz sampling rate ADC was used to sample 2048 points, which correspond to 20.48 μsec in the transient response.

4.6.1.1 BIT effect on the DUT performance

The performance comparison was made between the two designs, i.e., with and without the envelope detector to evaluate the effect of the BIT structure on the DUT performances. Table 9 shows the specification values for S11, S22, S21, S12, NF, and IIP3 for the DUT with and without the envelope detector.

Table 9. Loading effects of test sensor.

	S11 [dB]	S22 [dB]	S21 [dB]	S12 [dB]	NF [dB]	IIP3 [dBm]
without BIT	-8.5	-27.1	14.4	-31.6	4.04	1.81
with BIT	-8.5	-26.6	14.2	-31.7	4.04	1.75

Among the specifications, the specification S22 shows the worst deviation, and has 0.5dB degradation from the original specification value. This degradation mainly comes from the parasitic capacitances of BJT in the envelope detector, but considering the nominal specification value of -27.1dB, 0.5dB can be negligible. As evident from the table, none of the specifications were affected prohibitively after the envelope detector was added to the DUT.

4.6.1.2 Prediction accuracy

To emulate the effects of noise available in reality, a $1mV_{rms}$ Gaussian noise was added to the detected envelope of the 600 DUTs. Figure 31 shows the scatter plots comparing the actual and predicted specifications for LNA S21, NF, and IIP3, and 45 degree straight line corresponds to the ideal prediction line. The scatter plots for S21 and IIP3 prove close matching between the actual and predicted specification values, whereas the predicted NF values show relatively loose matching with the ideal 45 degree line. Considering the specification distribution range of NF is very narrow (i.e., <1dB) and the typical measurement error 0.1dB, one can say that the NF prediction shows the good prediction.

Table 10 shows the prediction accuracy in terms of RMS error and relative error, where the relative error is defined as RMS error over the specification distribution range. The RMS errors for three specifications over the validation set are 0.053 dB, 0.158 dB, and 0.086 dB, respectively.

Table 10. Test accuracy for case study I.

	S21	NF	IIP3
RMS error	0.053 dB	0.158 dB	0.086 dBm
Relative error	1.20 %	13.52 %	1.57 %

Finally, it needs to be mentioned that even though the envelope detector is under the same process variations as the DUT, one can achieve considerable prediction accuracy without calibration.

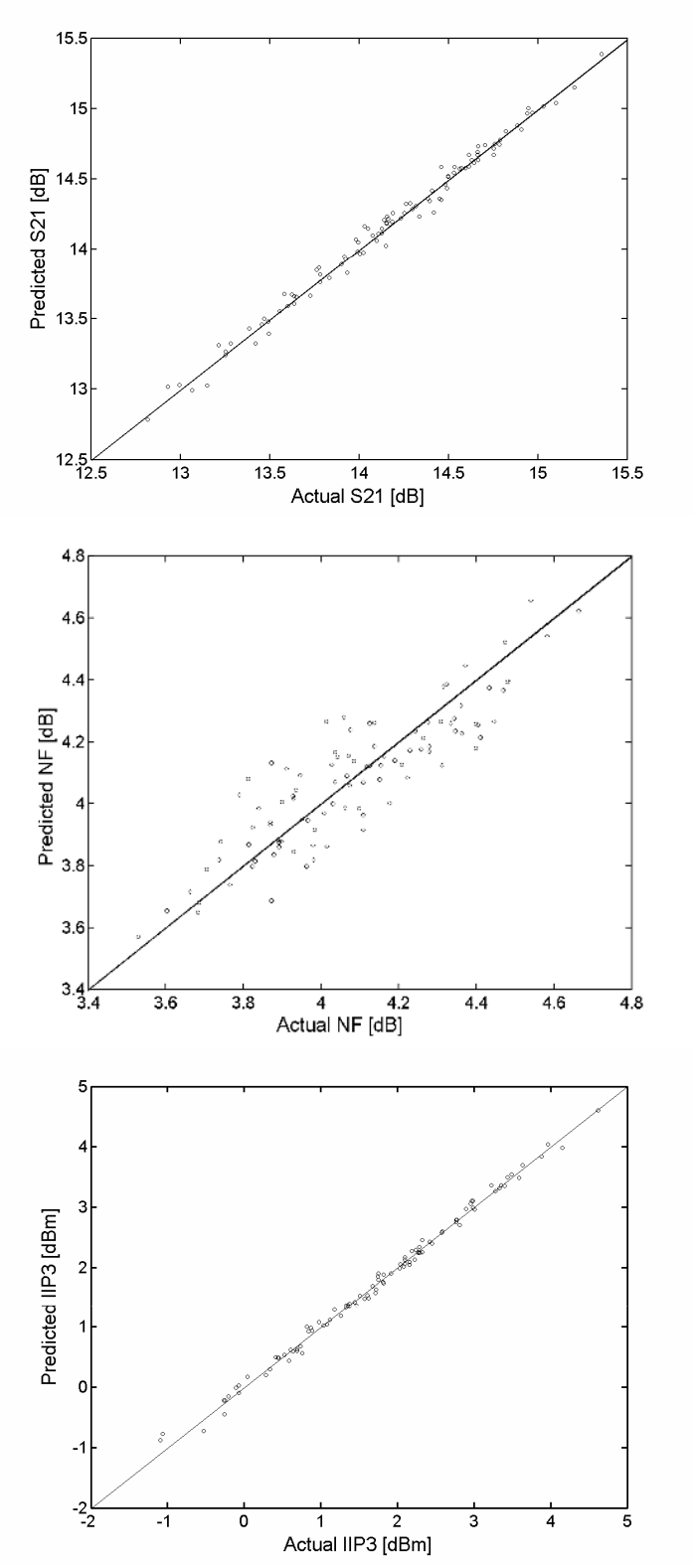


Figure 31. Specification predictions with 1 mV_{rms} white noise for S21, NF, and IIP3.

4.6.2 Case study II: Behavioral Transmitter

The proposed method was validated using behavioral simulations. A simple transmitter shown in Figure 32 was built using the behavioral models in Cadence rfLib, which are parameterized by a set of circuit performance metrics such as gain, IIP3, etc. The envelope detectors were implemented in MATLAB.

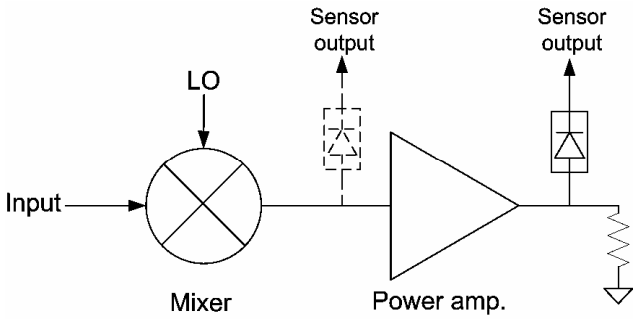


Figure 32. Transmitter with behavioral models.

To make multiple instances of the system with different performance metrics (i.e., to induce parametric faults), the specification S21, IIP3, and P_{1dB} of the PA and mixer were perturbed simultaneously with no correlation between their variations, and all of them were also served as the target specifications to be extracted. Due to the lack of statistical information regarding the correlation between the specifications, no correlation was assumed. However, from test and diagnosis perspective, it is one of the tough cases since the number of independent variables (i.e., the specifications to be extracted) increases in the circuit. Hence, the test stimulus needs to be optimized in a larger number of dimensions and hence, the optimization becomes a tougher problem. Each specification was assumed to have a uniform distribution around its nominal value with a

maximum deviation of ±3dB (or dBm, as applicable). Assuming that access is available to each module output, two envelope detectors were placed at the outputs of the PA and mixer.

A set of diagnostic test stimuli were optimized for the transmitter shown in Figure 32. Single and two-tone sinusoidal stimulus showed the best performance of test quality for all the specifications. Each specification extraction was optimized with a different input power level. Table 11 shows the optimized stimuli used in the simulations for the external sensor-based method. For the embedded sensor-based method, a single sinusoidal waveform was applied at 3MHz and 4MHz with -20dBm power level, respectively.

Table 11. Test stimulus used for case study II.

Input power	PA		Mixer	
	S21	P_{1dB}	S21	IIP3
@3MHz [dBm]	-32	4	-26	-36
@4MHz [dBm]	-26	-	-26	-23

Monte-Carlo simulation was performed to generate 300 instances in which each instance had different specification values. Among 300 instances, 250 instances were used to calibrate the non-linear regression models. Validation of the proposed method was performed on the remaining 50 instances.

Table 12 shows the test accuracy for the transmitter shown in Figure 32. It should be noted that for the specifications of the mixer, the embedded sensor at the mixer output provides significant increase of test quality. The external sensor case employed four

different test stimuli, whereas the embedded sensor case used only a single test stimulus to provide almost the same test quality as that of the single senor case.

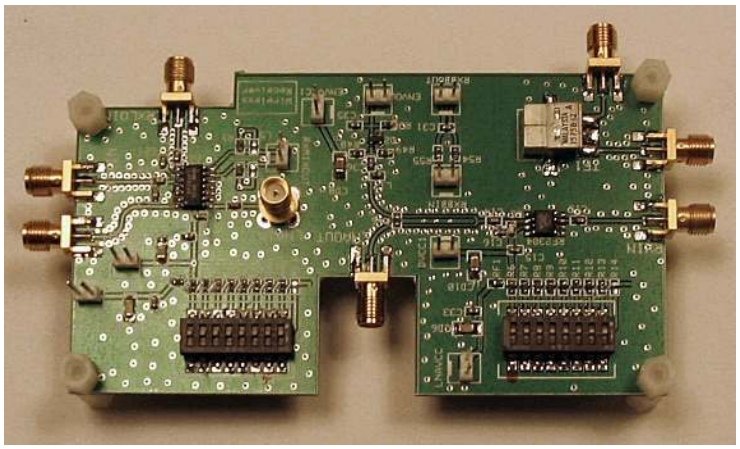
Table 12. Test accuracy for case study II.

RMS Error	PA		Mixer	
	S21 [dB]	P_{1dB} [dBm]	S21 [dB]	IIP3 [dBm]
External sensor	0.11	0.13	0.12	0.155
Embedded sensor	0.169	0.118	0.003	0.037

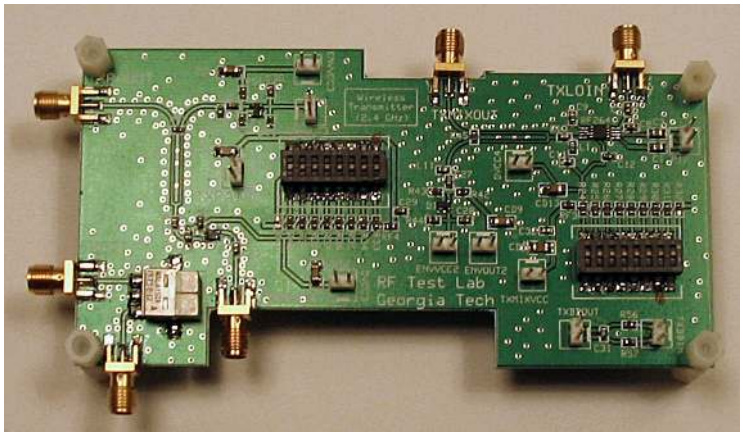
4.6.3 Case Study III: 1.575GHz Transceiver Prototype

For the second case study, a 1.575GHz transceiver front-end shown in Figure 33 was implemented using off-the-shelf components and used as a test vehicle.

The transmitter consisted of a mixer, PA, and band-pass filter. The PA was implemented using NEC μPC2763, RFMD RF2641 for the up-conversion mixer, and 1575B-12 for the band-pass filter. In the receiver chain, a LNA, down-conversion mixer, and band-pass filter were implemented via RFMD RF2304, RFMD RF2411, 1575B-12, respectively. Three envelope detectors were deployed in the prototype. As described earlier, the detectors were placed at the output node of the PA, mixer, and LNA. They were implemented using a HSMS-2865 diode. The RC time constant of all the envelope detectors was set to 0.1 μsec with 20kΩ resistor and 5pF capacitor.



(a)



(b)

Figure 33. Prototype of a 1.575GHz transceiver front-end with (a) receiver and (b) transmitter.

4.6.3.1 Test setup

The overall test configuration including the prototype is shown in Figure 34. Loop-back mode was employed to test the LNA without an external RF source. A PC with a NI-DAQ card PCI-6115 was used to run the test, i.e. test stimulus feed and capture of the test response. The NI-DAQ card emulates embedded data acquisition circuitry, and can be performed up to the sampling rate 4Msamples/sec with 12 bit resolution on 4

ADC channels and 2 DAC channels. Note that no RF equipment was deployed in the test configuration to measure complicated RF specifications.

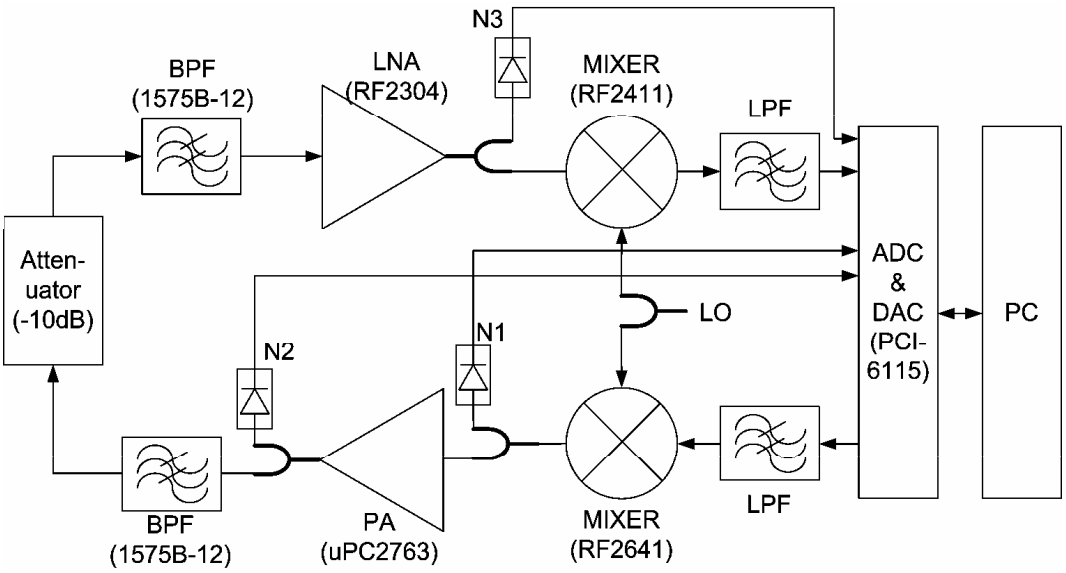


Figure 34. Test setup for the prototype.

To emulate multiple instances of the system, individual power supply level for each component was perturbed simultaneously and independently. As the power supply level is perturbed, each device exhibits different performance metrics. Ten instances for the PA and ten instances for the mixer were generated by sweeping the power supply level. Hence, the total number of the transmitter instances obtained is 100(=10 PA instances × 10 mixer instances). For all these 100 instances, true specification values were measured using standard test equipment. These specification values are the output parameters of the regression models and were used to evaluate test quality. Among them, 60 instances were used as a training set to generate the nonlinear regression models. As a validation set, the remaining 40 instances were selected. For the LNA in the receiver, 25

and 8 instances were generated for a training and validation set, respectively. The specifications S21 and third-order-intercept (TOI) were considered for the transmitter, PA, mixer, and LNA as test specifications. In the experiment, two configurations were evaluated similar to the case study using the behavioral simulations. The embedded sensor-based method utilized all the sensors (i.e., N1, N2, and N3 in Figure 34) deployed in the prototype, whereas the sensor N2 at the output of the transmitter was used to extract all the specifications of the transmitter, PA, and mixer in the experiment to evaluate the external sensor-based method.

It should be noted that off-the-shelf components commercially available are fault-free circuits that already passed production tests. Therefore, no faulty circuits could be explored in this experiment if multiple instances are tested with a RF socket. Instead, sweeping the power supply levels changes the performance metrics, thereby making the circuits behave as parametric failures that emulate a real production test scenario.

4.6.3.2 Evaluation of the envelope detector

The envelope detector built on the prototype was characterized. The performance of the envelope detector was then compared to simulation results. A two-tone signal was used as a stimulus with varying input power to extract dynamic range as shown in Figure 35.

The RMS value of extracted envelope is used as a performance metric. The plot shows good match with the simulation results. Note that the RMS output values are saturated for the input signals below -40dBm, which is limited by the ADC resolution (i.e., 12 bits and +/-1V input range of NI-DAQ) that we used. Even in the presence of the performance limitations of the ADC, the envelope detector can extract RMS values up to

-40dBm two-tone signal. The manufactured envelope detector consumes 50uA at 3V power supply, which is very low power consumption compared to RF circuits.

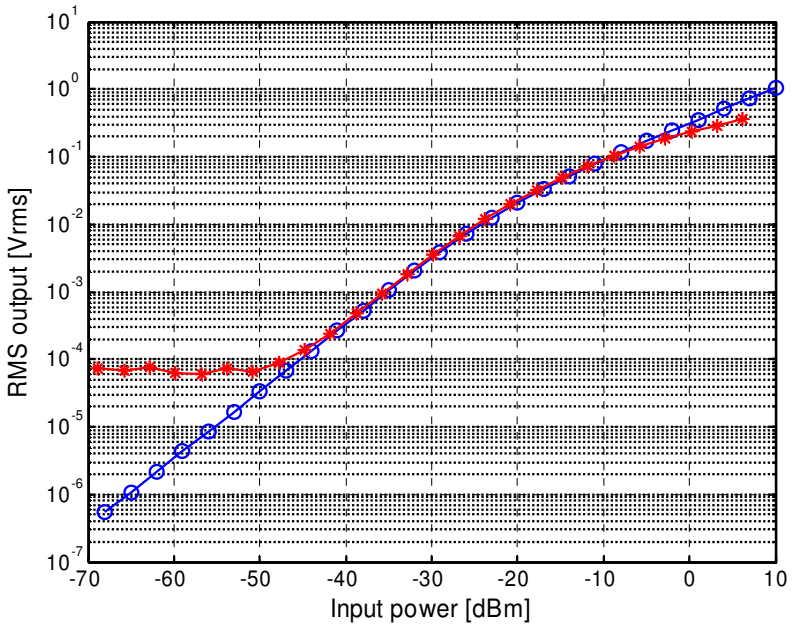


Figure 35. Comparison between simulation and hardware measurements of the envelope detector.

4.6.3.3 Conventional specification based tests

The specifications of the transceiver and sub modules such as gain and IIP3 were measured using the conventional test setup and test equipments for the purposes of training and evaluation of the proposed test methodology. An Agilent spectrum analyzer E4407B was used to capture the RF output spectrum. A HP signal generator 8648D is used to provide LO to the up- and down-conversion mixer modules of the RF transceiver prototype. Keithley source meters are used to provide the required supply voltages to the component modules of the prototype.

To measure gain of each instance, a single tone test input was provided to the prototype by a signal generator 8648D. The corresponding spectrum output for each instance is observed with the help of the spectrum analyzer.

For TOI measurements, two signal generator2 8648D were utilized to provide two input tones with closely placed frequencies and identical amplitudes. These tones were combined using a Wilkinson power combiner designed to operate at 1.6 GHz. The RF spectrum at the output of each instance is observed with the help of the spectrum analyzer and the corresponding TOI values are computed from the observed RF spectrum.

4.6.3.4 Hardware measurements

For the embedded sensor-based method, a single-tone sinusoidal waveform was applied at 40 kHz with −15dBm power level to the input of the transmitter, whereas for the case that a single test sensor was employed, different multi-tone waveforms specialized for each target specification were stimulated sequentially. These results are summarized in Table 13.

Table 13. Test stimulus used for case study II.

Input power	Transmitter		PA		Mixer	
	S21 (1)	TOI (2)	S21 (3)	TOI (3)	S21 (4)	TOI (4)
@40KHz [dBm]	-20	-20	-30	-30	-25	-25
@20KHz [dBm]	-	-20	-30	-30	-	-

The test response envelops from the test sensors were captured simultaneously for 0.24msec time duration, which corresponds to 100 periods of the envelope. Therefore, the

total data capturing times are 0.96msec and 0.24msec for the external and embedded sensor-based method, respectively. After performing the post-processing described in Section 4.5, one period of the averaged envelope was extracted for each stimulus.

Figure 36 shows all the envelopes of each sub-module via N1 and N2 in which each envelope corresponds to a different instance (generated by perturbing the power supply). All the envelopes extracted at the output of the PA are given in Figure 37 for four different stimuli listed in Table 13. Finally, one period of the envelope from each envelope detector was fed into the pre-built MARS models, and the specifications of interest were computed via the models.

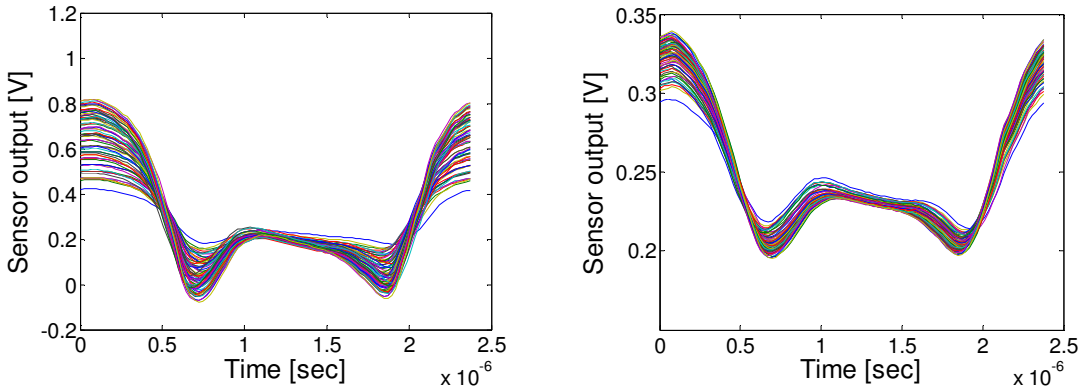


Figure 36. Extracted Envelopes for the validation set at (a) N2 and (b) N1 in the multi-sensor based method.

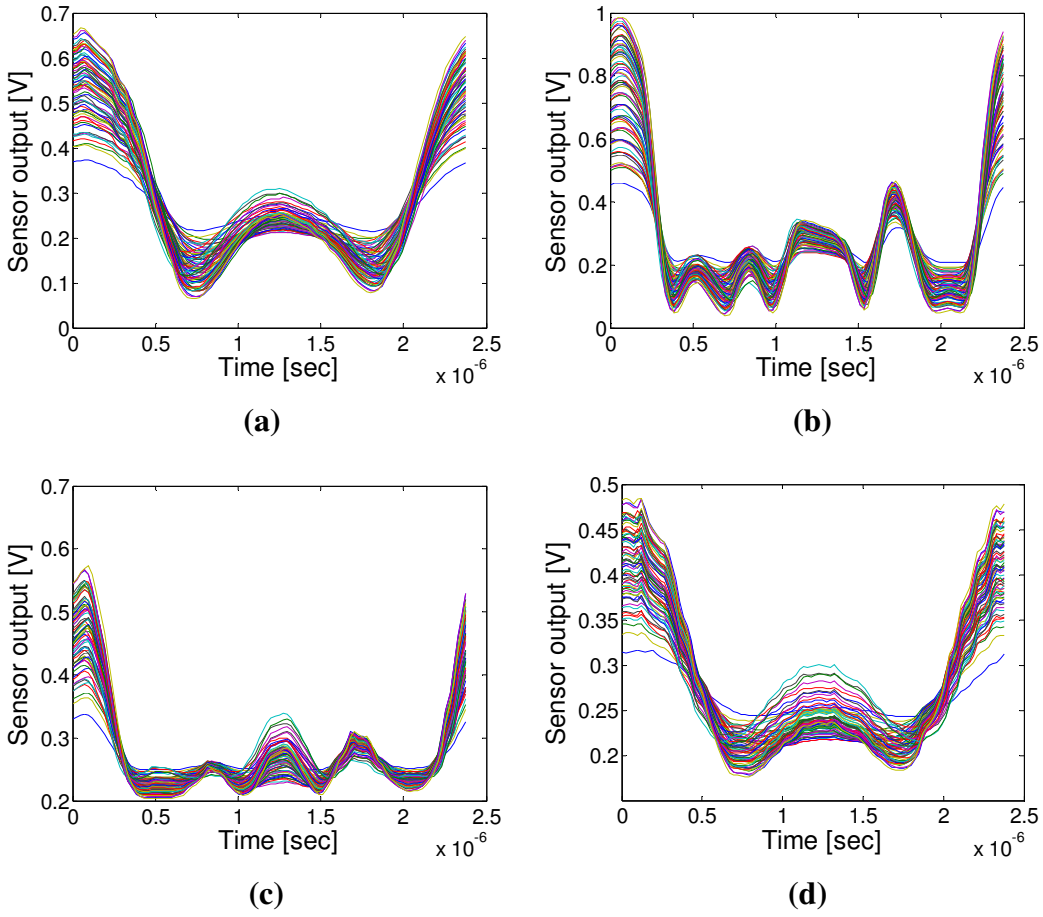


Figure 37. Extracted envelopes for the validation set at N2 in the single sensor based method for the stimulus of (a) 1, (b) 2, (c) 3, and (d) 4 from Table 13.

Table 14 shows the test quality in terms of RMS error for the transmitter and the sub-modules with the comparison of that of the BIT scheme in [25], which utilizes RMS and peak DC values of test responses. In this experiment, the RMS and peak values for the scheme in [25] were computed from the extracted envelope and the same test stimulus used for the proposed method was applied. As can be seen from Table 14, the proposed method shows high degree of test accuracy for all the specifications. For example, the RMS error for the transmitter S21 is 0.049dB and 0.017dB for the external sensor and embedded sensor case, respectively. In addition, the test quality shown in Table 14 is

within the measurement noise expected from standard specification measurements using RF instrumentation. Compared to the method in [25], the proposed methods provide better test quality for all the target specifications. Moreover, the specification TOI of the mixer fails to be extracted via the method in [25]. It proves that the transient envelope features provide more information getting better test quality. Using loop-back mode, the LNA specifications were also extracted with high test quality.

Table 14. Test accuracy for the transmitter and the sub-modules.

RMS Error	Transmitter		PA		Mixer		LNA	
	S21 [dB]	TOI [dBm]	S21 [dB]	TOI [dBm]	S21 [dB]	TOI [dBm]	S21 [dB]	TOI [dBm]
external sensor	0.049	0.104	0.029	0.136	0.020	0.124	-	-
embedded sensor	0.017	0.082	0.012	0.035	0.020	0.063	0.032	0.187
[25]	0.175	0.396	0.100	0.284	0.182	-	-	-

All the scatter plots for the cases listed in Table 14 are depicted in Figure 39, Figure 40, and Figure 38, comparing the actual and predicted target specifications. As can be seen, the scatter plots show close matching between the actual and predicted specification values.

Testing of the discrete RF circuits should require the proper termination at the input and output such as 50Ω termination. However, for the fully integrated system, the embedded circuit is cascaded, and the performance is also affected by the input and output impedance variations of the following and followed device. This effect can be seen in the case of the mixer S21 shown in Figure 39 and Figure 40, where a scatter

group has the same mixer instance (i.e., the same S21 value at a particular power supply level) with different PAs (i.e., different power supply levels). Note that the specifications of each module are measured with 50Ω input and output termination such as a standard measurement approach for a discrete RF DUT. Hence the measured specification values involve no variation of the cascaded impedance. The proposed method, however, provides the specification values of the embedded circuit including the impedance variations of the following and followed sub-modules.

It is observed that the PA and LNA specifications were extracted accurately in the presence of performance variations of the mixer and PA. It shows that the proposed scheme is suitable for computing specifications of embedded sub-modules, where performances get perturbed simultaneously and affect the following sub-modules' response. In addition, even though the variation range of the specification S21 of the mixer and LNA is around 1dB, the proposed method can extract the specification very accurately within measurement noise level.

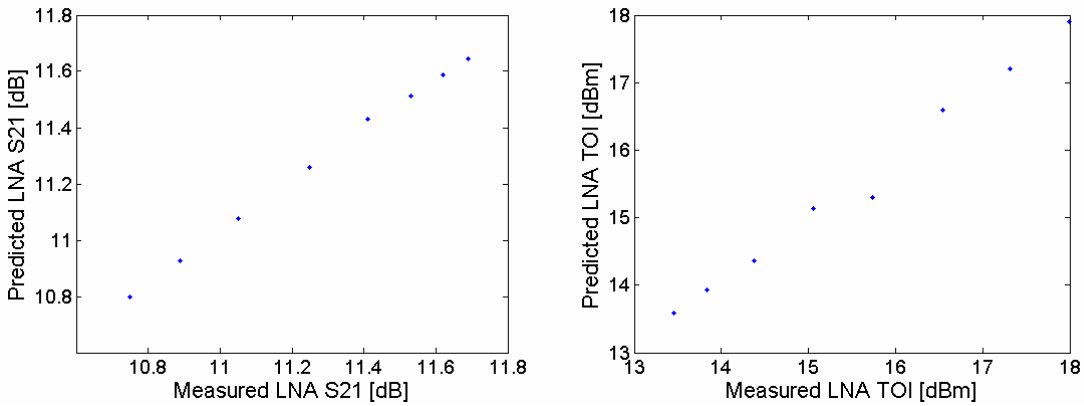


Figure 38. Extracted LNA Specifications

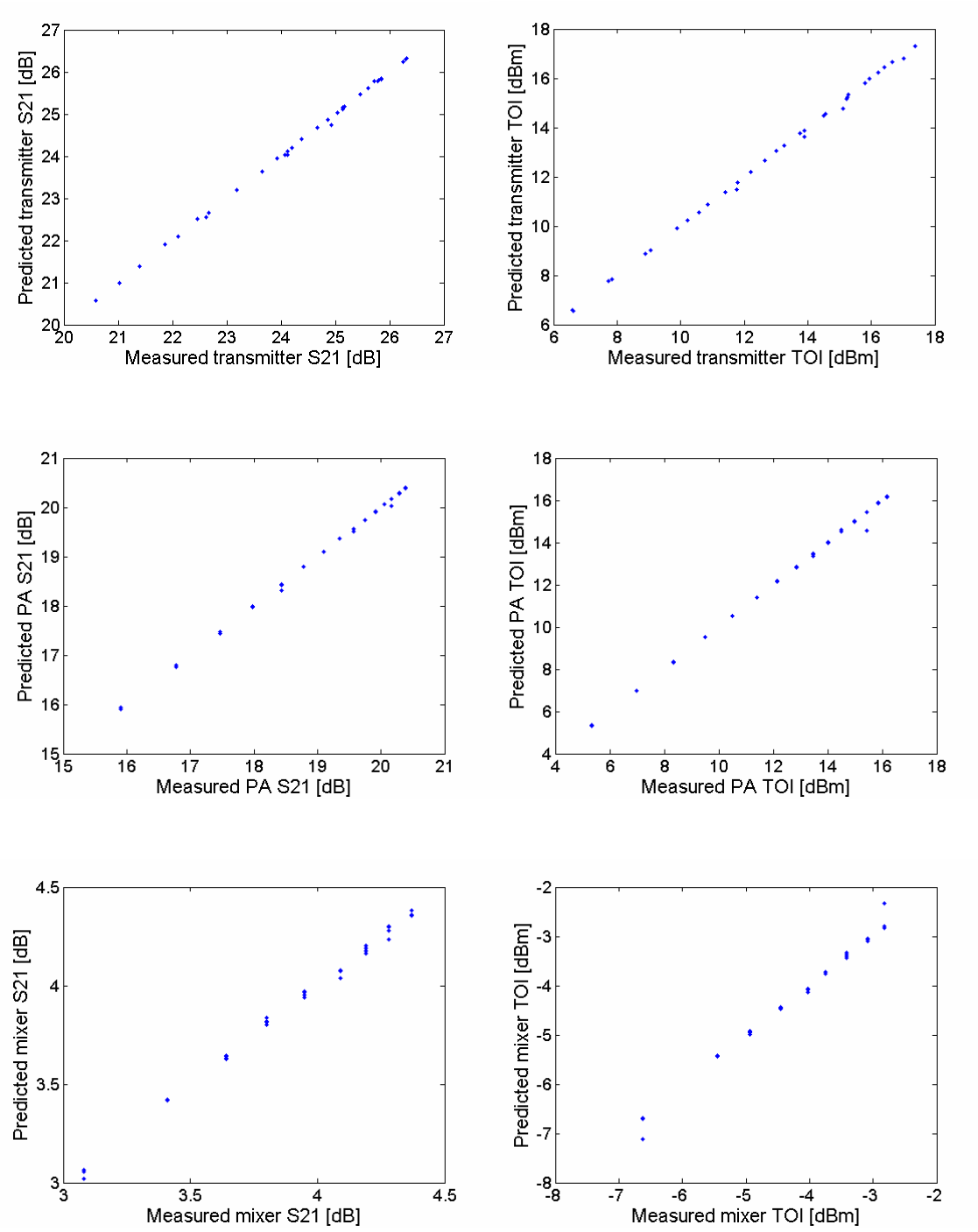


Figure 39. Predicted specs vs. measured specs of external sensor-based method.

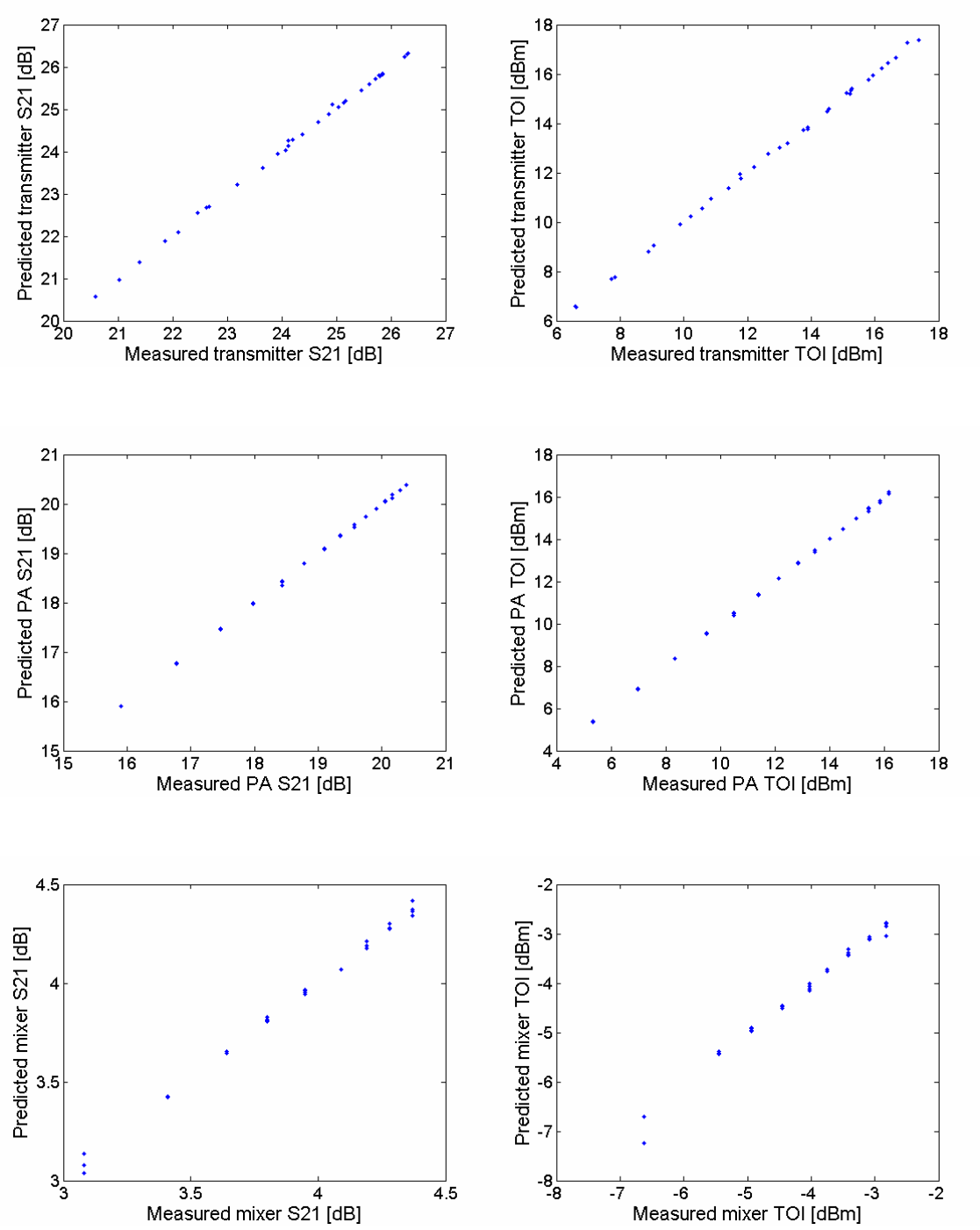


Figure 40. Predicted specs vs. measured specs of embedded sensor-based method.

4.6.4 Impact on Test Cost and Quality

Finally it should be noted the efficiency of the proposed method in terms of test economics. A widely accepted cost-of-test (COT) formula as follows:

$$COT = \frac{Capital + Operating\ Costs}{Yield \times Throughput} \quad \textbf{(29)}$$

where *Capital* implies the depreciation of capital equipment such as testers, probers, handlers and possible facilities, and *Operating Costs* consists of maintenance, labor, consumables, and any additional overhead. *Yield* is the ratio of the total number of good devices to the total number of manufactured devices, and *Throughput* is the number of devices tested for a given unit of time. The proposed method significantly impacts on *Capital* and *Throughput* in COT.

Compared to typical ATE based production tests (i.e., standard test method), the proposed method is feasible for self-test and diagnosis using on-chip resources such as ADC, DAC, and DSP module. Moreover, RF test stimulus generator is substituted with the transmitter to test the receiver in loop-back mode. Therefore, the RF front-ends of the system can be tested without any external test resources, coupled with fully embedded loop-back circuitry such as a circuit present in [84]. Finally, the term *Captical* and *Operating Costs* can be reduced using on-chip facilities.

From a test throughput perspective, the proposed method requires below 1msec data capturing time and data processing time in the DSP module for a DUT testing. Considering the high computing power in DSPs or general CPUs, the computation for post-processing and evaluation of the non-linear regression models is not computationally intensive. The overall test time is significantly reduced compared to typical production

test times on RF ATE. It is due to the fact that all the specifications for the transceiver as well as the sub-modules are extracted simultaneously from a single test pattern for both the cases. For example, in [15], the lowest RF test time is set to 280msec (a 200 msec handler index time plus an 80 msec electrical test time). Finally, the total electrical test time is improved by several orders of magnitude over the reported test times, resulting in negligible test time as compared to 200msec handler index time.

CHAPTER 5

SOFTWARE-IN-THE-LOOP SELF-CALIBRATIN OF RF CIRCUITS FOR MULTI-PERFORMANCE VARIABILITY

In this chapter, a global RF *specification-aware compensation methodology* is described that can trade-off performance specifications against one another in a preferred way while performing *simultaneous multi-parameter compensation*. In order to determine *how* to change embedded tuning knobs such as bias current/voltage, capacitance, and inductance, it is necessary to first know *which* of the circuit specifications have been affected and by *how much*. Then corrective action is taken by *adapting* tuning knobs to the drifted process variations. The procedure is iterated until all the specifications are within the acceptance limits for parametric faulty DUTs. Suppose that a standard test and diagnosis method is employed for self-calibration, even though the method provides no way to diagnose embedded modules in RF systems. Then, self-calibration significantly increases overall test times on expensive ATE under stringent requirements of test costs. This results post-silicon calibration to be infeasible. As described in the previous chapter, the envelope of a test response provides enough features to calculate complex at-speed specifications. The proposed self-calibration technique uses the fact on digital feedback loop.

In the following, limitations of prior work are discussed, followed by the proposed self-calibration methodology and alternate control law to trim tuning knobs.

Simulation results and hardware measurement using fabricated CMOS 1.9GHz LNAs are presented.

5.1 Limitations of Previous Work

In the area of analog and digital circuits for reducing circuit performance variability, several methods have been presented in the literature. Self-calibration techniques are an integral part of analog-to-digital and digital-to-analog converter designs [86], [87]. To reduce device mismatches, these methods trim reference voltages or capacitance. Alternative solutions have been proposed that are inherently robust to particular process variations. One such example is the circuit proposed in [88] that compensates for the mismatch of current mirrors. The feedback circuit in [89] tracks effects induced by predefined process variations such as transistor thresholds, body effect, and channel length modulation. However, it has been difficult to apply similar techniques to RF circuits due to the following reasons: 1) all the previous methods described above are application oriented, and 2) issues arising from high speed operations of RF circuits are not explored. For example, loading effect due to additional circuitry can significantly degrade overall RF circuit performances.

On the other hand, self-repair or self-healing techniques for digital circuitry have been proliferated [90], [91]. For example, memory circuits employ redundant cells as well as test circuitry. Once performing tests by embedded test circuitry, detected faulty cells are replaced with redundant cells via adjusting on-chip circuit topology. These methods have concerns on catastrophic faults in digital circuits without consideration of parametric failures.

A few calibration techniques for RF circuits have also been presented. For example, pre-distortion linearization has been employed as a common technique for RF power amplifiers [96]. This approach compensates for the presence of nonlinearities by applying pre-distorted input signals. On the other hand, analog feedback-based compensation techniques have focused on a specific performance deviation such as gain loss, I-Q mismatch, and DC offset errors [39]-[42]. The embedded sensor-based technique has been presented to address the specification S11 drifts. However, it hasn't considered other performance specifications as well. Consequently, these approaches regarding RF circuits only compensate for a single performance deviation at a time and are not universal to any RF circuits and topologies. Moreover, the range of performance variation over which compensation can be performed is limited by the fact that the compensation circuitry is itself exposed to the same process variations as the RF circuits. In addition, these approaches require significant processing and die area overheads to incorporate self-calibration into RF circuits.

In this work, we propose a universal RF self-calibration scheme to compensate for multi-performance variability. The performance evaluation scheme is specification centric so that the deviations are compensated when they affect the end specifications. In addition, the *compensation is performed via digital algorithms* that are "immune" to thermal and process effects.

5.2 Cost Function Formulation for Self-Calibration

Self-calibration involves an optimization process to search for optimal adjustment of multiple tuning knobs and in turn converge to an optimum solution. In order to achieve

good compensation for performance drifts, an optimization process needs to be driven by an objective called cost function. Depending on the constraints of the circuit performances, various formulations are feasible. For example, suppose that the requirement for the noise figure (NF) specification of a circuit is stringent, whereas the other specifications have enough performance margins with respect to the specification limits. In this case, the calibration process focuses mainly on the noise figure correction.

Without loss of generality, a cost function employed in the calibration process can be stated as

find $x^* \in D$ such that $\Phi(f(x^*)) \le \Phi(f(x))$, $x \in D$,

where $\mathbf{D} \subset \Re^n$ denotes the space of tuning knobs, $f(\cdot)$ is a set of functions derived from performance specifications, which, in general, are measured during the test and diagnosis phase. For instance, a cost function $\Phi(x)$ for tuning knobs x can be defined to minimize performance variability as

$$\Phi(x) = \left\| [w_1 f_1(x), w_2 f_2(x), \cdots, w_n f_n(x) \right\|_2 \quad \textbf{(30)}$$

$$f_i(x) = S_i - S_{i\_nom} \quad \textbf{(31)}$$

where w_i is a weighting factor, and S_i and S_{i_nom} denote the *i-th* specification value and its expected value, respectively. In this scheme, the optimization process search for a solution minimizing the difference with respect to the nominal specifications with weighting factors. Therefore, the minimum value of the cost function corresponds to minimal performance variability for each performance metric.

For example, a CMOS LNA in Figure 41 shows one application with two tuning knobs denoted as $knob_1$ and $knob_2$, which control the bias points of the transistors. These tuning knobs need to be adjusted to compensate for any performance drifts.

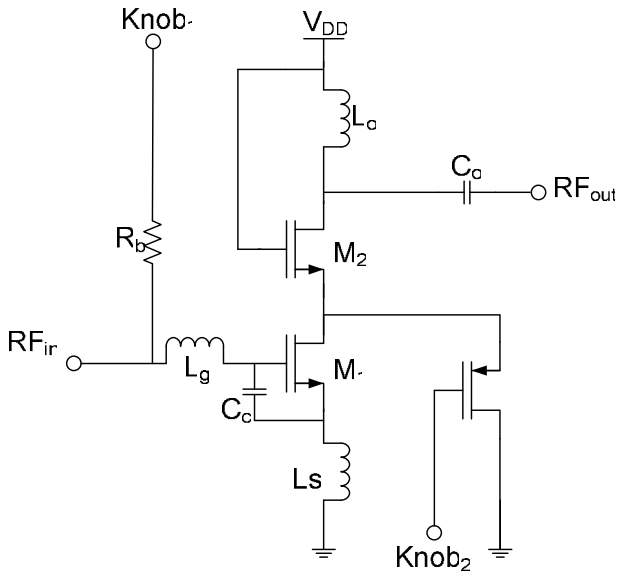


Figure 41. CMOS LNA using folded PMOS IMD sinker [97].

Figure 42 depicts a contour plot of response surface of the cost function for an instance of the CMOS LNA, which has process parameter drifts. The weighting factor is set to the replica of the nominal value for each specification, which thereby is normalized each other. In the plot, the mark * denotes a cost value for a particular DUT which performances are deviated. The solution with optimal trimming of tuning knobs is marked as '+'. Proper trimming of two tuning knobs enable the instance to converge to the optimum solution as shown in Figure 42. Suppose that only one of the tuning knobs is deployed. Then, one can observe that the instance shows no convergence to the optimum due to limited controllability of performance. To maximally compensate for performance

drifts, multiple tuning knobs are preferred to increase the degree of freedom (DOF). However, it leads to a higher complexity optimization problem in searching for optimum tweak of tuning knobs.

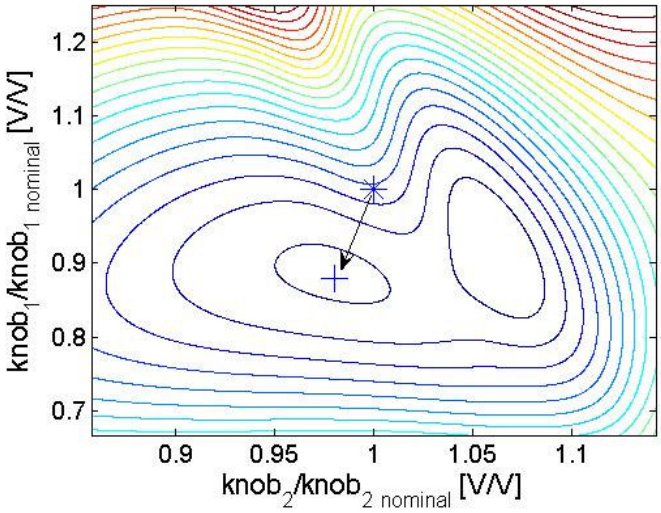


Figure 42. Contour plot of the cost function.

For the given contour plot, the optimization process is performed to converge to the optimum. In general, local and global optimizers can be employed in view of finding a solution. Though the iteration number required to converge to the optimum depends on the response surface complexity of a cost function and optimization engine employed, it takes at least a few hundred iteration numbers in global optimizers. Local optimizers can be employed to reduce the number of iteration. However, compromise with performance variability after calibration may stem from the non-optimal tuning of the knobs. As mentioned already, the number of iteration is critical to minimize the overall self-calibration time. Further, it directly increases overall manufacturing costs.

5.3 Alternate Control Law for Tuning Knobs

This section describes a one-shot control law for adapting multiple tuning knobs simultaneously. The key feature of the method is to predict the optimum solution from the extracted test response, thereby eliminating a number of iteration of the optimization process.

A control mechanism called *alternate control law* is proposed that makes use of the principals of alternate test methodology [23]. In the past, there has been significant work on applying alternate test ideas to analog, mixed-signal and RF circuits. The concept can be also employed for self-calibration in test and diagnosis phase to figure out the direction to trim tuning knobs.

Process variations inherent to the manufacturing process affects the circuit specifications as well as the measurement space such as transient responses obtained from a test stimulus. Suppose that particular process drifts are induced and thereby specifications are also deviated correspondingly. Optimal tuning knob control can be determined for the given their drifts. Figure 43 illustrates the effect of the variation of one such parameter in the process parameter space **P** on the optimal tuning knob space **T** and the corresponding variation of a particular measurement data in **M**. For any point *p* in the parameter space **P**, a mapping function onto the optimal tuning knob space **T**, *f*:**P**→**T** can be computed. Similarly, for the same point *p* in **P**, another nonlinear mapping function onto the measurement space in **M**, *f*:**P**→**M** can be computed. Suppose that a specially crafted stimulus is selected in such a way that the test response is *strongly correlated to all the optimum trim of the tuning knobs*. Alternatively, the mapping function *f*:**M**→**T** could be constructed for the optimal tuning knobs from all measurements in the

measurement space **M** using nonlinear regression mapping tools [30]. The test stimulus can be optimized via special alternate test generation algorithms [23], [24] to exploit these correlations.

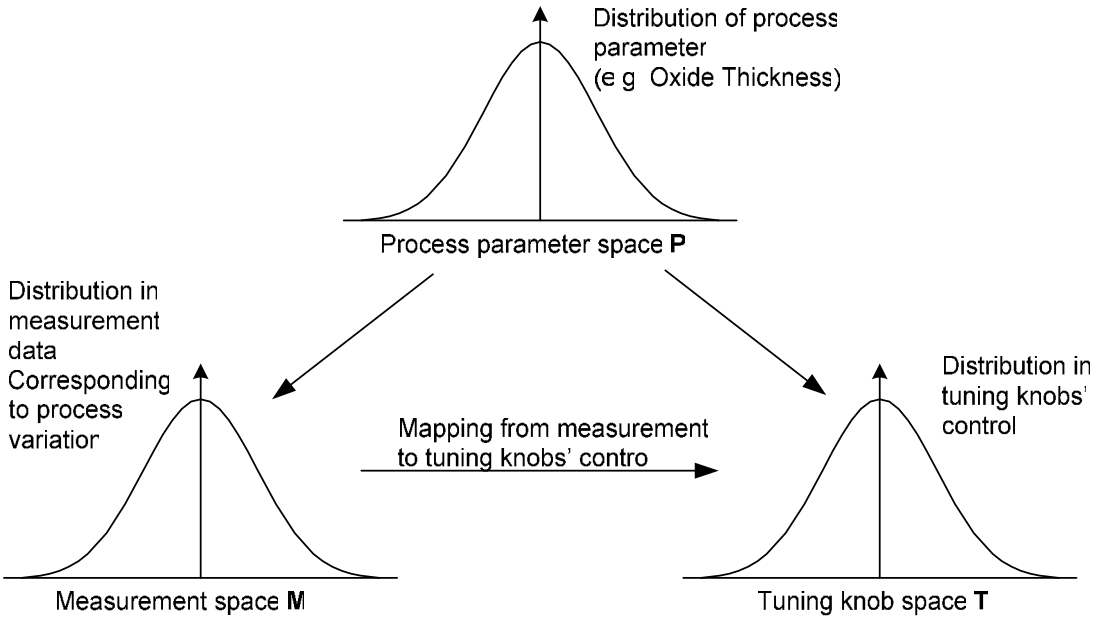


Figure 43. Effects of process variations on measurement and tuning knob control.

The overall procedure of the alternate control law is depicted in Figure 44. A specially generated test stimulus is applied to a DUT. The envelope of its test response as a feature is extracted via the embedded detector and is placed at much lower frequencies compared to the at-speed test stimulus. Finally optimal tuning knobs are predicted via pre-built regression models. Note that the generation of mapping functions is one time process in the characterization phase before production test. Prediction of the optimal control from the extracted low-frequency signature via an embedded detector provides the following benefits. First, it is performed on a single stimulus and test configuration.

Second, iterative optimization process can be eliminated via one-shot prediction, thereby reducing the overall time to perform self-calibration procedure. Last, this control law has no limitation on the number of tuning knobs and results in enhancing controllability of performance metrics.

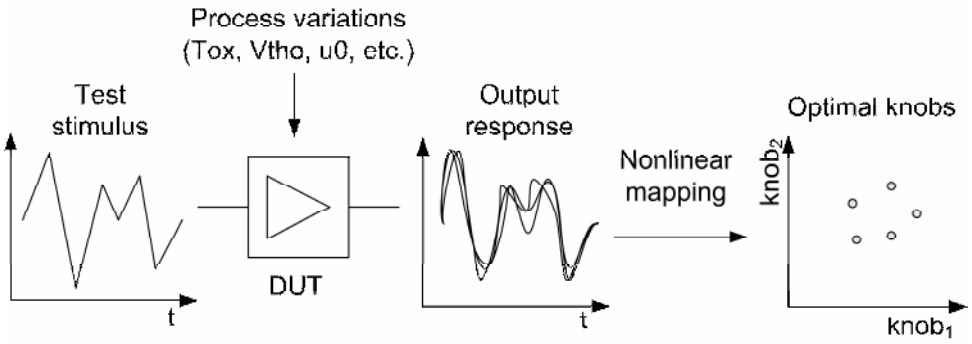


Figure 44. Alternate control law.

5.4 Self-Calibration Structure

The overall procedure of the proposed self-calibration method is depicted in Figure 45 and works as follows. After manufacture, production tests are conducted to ensure the proper functionality of a DUT. If the current DUT is determined as a faulty DUT with catastrophic faults such as open or short of a node, the procedure ends since the catastrophic faults typically result topology changes of a circuit and is almost infeasible to compensate without replacing it with designated redundant circuitry. The self-calibration mode is turned on if the current DUT has parametric fault(s). Then, test stimulus generation process is activated and switches device inputs accordingly if needed. The embedded detector is used to convert the test response into a low frequency test

response “signature”. Based on the pre-defined regression models (described later in details), these signatures are mapped into the optimum knob controls. Finally the DSP processor adjusts the digital control bits to trim them. The trimming process can be iterated until the optimum tuning is obtained based on a predefined *optimization rule*.

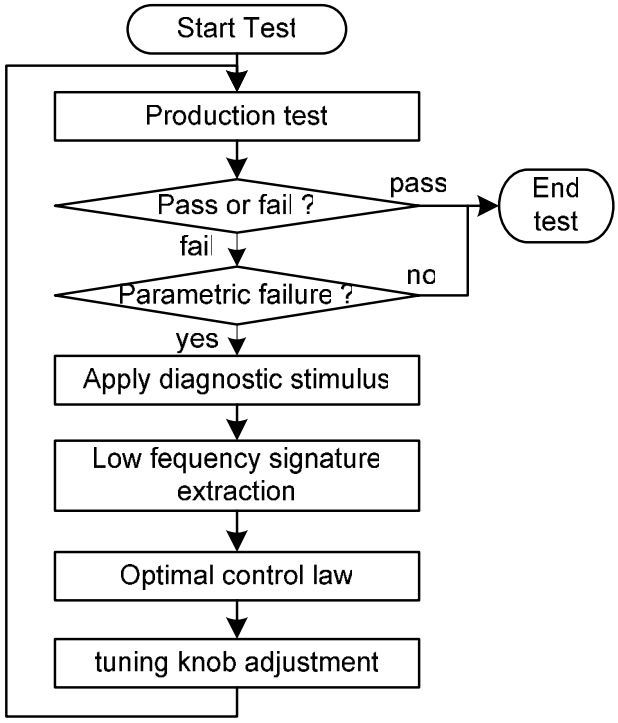


Figure 45. Self-calibration procedure.

The self-calibration procedure is done using the hardware configuration depicted in Figure 46, mainly composed of feature extraction, ADC, DSP, and tuning knob modules. As can be seen from the configuration, the procedure can be portable into embedded resources in RF systems. For example, an ADC in the receiver chain can be employed to quantize low frequency signatures, and a DSP module can be utilized to

analyze all the extracted signatures and control tuning knobs. Thereby, the proposed method eliminates the use of any external RF testers to measure RF test responses and is suitable for on-chip implementation.

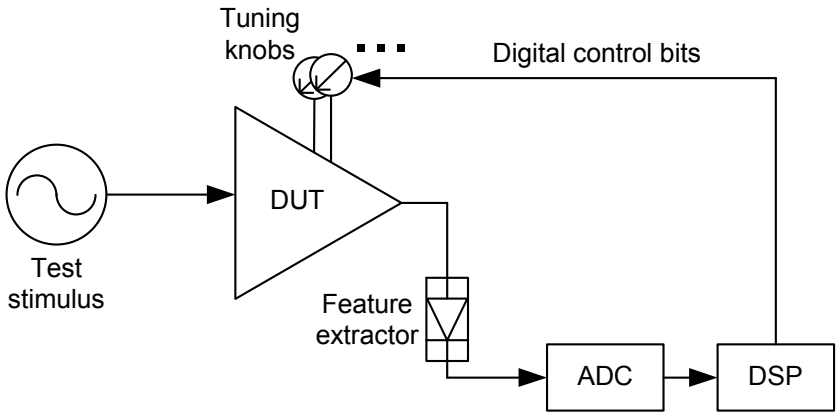


Figure 46. Hardware configuration.

As a part of the self calibration scheme, bias current/voltage, capacitance, and inductance can be trimmed via DAC control or switch on/off operations to compensate for performance deviations from the expected. For example, the method in [42] trims inductance via a set of switches. The schematic shown in Figure 41 shows one application with two tuning knobs. These tuning knobs can be controlled via programmable bias circuitry. Such a circuit is depicted in Figure 47, which shows bias circuitry where digital switches S0-S4 control the total current of a designated transistor.

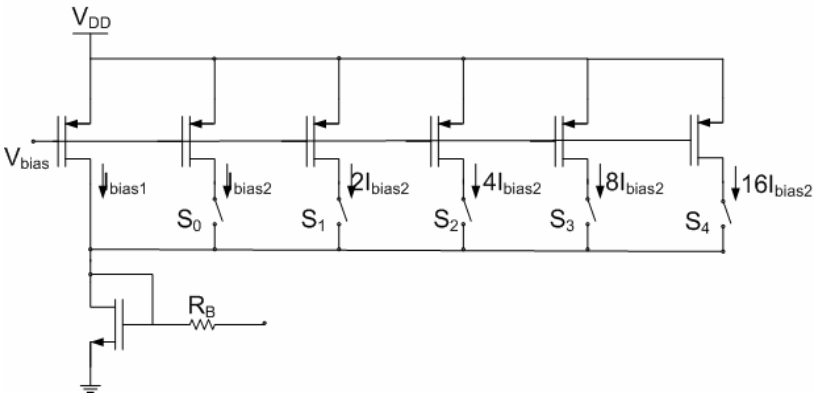


Figure 47. Programmable bias circuit.

The objective of the embedded detector is to extract much lower frequency signature which is suitable for reuse of the on-chip ADC. Moreover, the extracted signature needs to be strongly associated with the performance deviations and/or cause of the deviations. For this purpose, envelope detector is employed as a feature detector. It is well-known circuitry for amplitude modulated (AM) signal demodulation and is composed of a diode, a resistor and a capacitor. Compared to a single DC value from other RF detectors such as power and RMS detectors, the envelope detector outputs the envelope of a test response that provides much more information if time-varying envelopes exist in the test response. In the literature, the envelope detector has been employed for test purpose, and proved to provide features to extract complicated at-speed RF specifications [98].

5.5 Results

The proposed self-calibration structure has been applied to a source-degenerative CMOS LNA and partially implemented on TSMC 0.25μm CMOS technology. Hardware

measurements on the fabricated CMOS LNAs were evaluated. As mentioned earlier, the proposed method is based on statistics (i.e., generation of nonlinear regression mapping functions), which should be extracted to explore all the relationship among parametric variations, transient signatures, and optimal knob controls. However, the number of fabricated dies is not enough to imitate real production environment in industry. Hence, a case study based on transistor-level simulations was also considered to ensure the effectiveness of the proposed method.

The compensation for performance variability was driven by the cost function defined in Equation (30), which gears toward minimizing performance variability with respect to the normal specification values. In the equation, w_i was set to the reciprocal of the i-th nominal specification value to normalize each specification's variability with respect to the expected. The proposed method is, however, no limitations on the definition of a cost function. As a figure of merit, the variability degree after calibration was evaluated by standard deviation. For successful calibration, the standard deviation of performance after calibration should be reduced compared to that before calibration.

5.5.1 Case Study I: 1.9GHz CMOS LNA with Two Tuning Knobs

For transistor-level simulation, the topology of a CMOS LNA with folded PMOS IMD sinker [97] was used as a test vehicle and was designed in TSMC 0.25um CMOS technology. Its simplified schematic is shown in Figure 48. Two tuning knobs were employed to control the bias of the transistor M_1 and M_p. Programmable bias control circuitry was implemented via current and voltage sources in ADS library. The sources were routed to each tuning knob, and were set to a range of ±40% and ±2% with respect to the nominal with 5 bit resolutions.

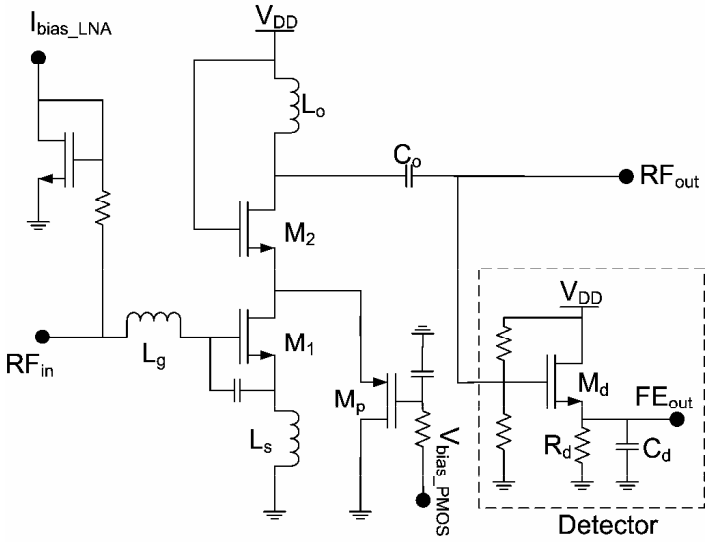


Figure 48. CMOS LNA with PMOS IMD sinker and feature detector.

To induce the effects of process variations and therefore generate instances with parametric faults as well as fault-free instances similar to the situation in real productions, Monte-Carlo simulation was performed via Agilent ADS simulator. All the perturbation parameters in simulation were assumed to be independent and normally distributed with 5% standard deviation. The parameters perturbed in Monte-Carlo simulation are as follows: zero-bias threshold voltage of p/n-channel transistors, channel doping concentration, low field mobility of p/n-channel transistors, value of passive components such as resistors, capacitors, and inductors. It should be noted that the embedded detector was also under the same process variations such as the situation in reality.

For each instance generated via Monte-Carlo simulation, all the specifications of interest were measured with the default tuning knobs. Also, the optimum specification values after calibration were logged for each instance by a brute-force method (i.e., sweeping both the tuning knobs within the pre-defined tuning range). The expected

optimal values were compared with the results from the proposed method. Optimal knob controls were also logged for output parameters used for nonlinear regression mapping. Similarly, the transient responses through the embedded detector were captured as input parameters. The total 318 instances were generated via Monte-Carlo simulation. Among them, 200 instances were used for generating the regression mapping functions and 118 instances were used as a validation set.

Two-tone sinusoidal waveform was utilized as a test stimulus with -20dBm at ± 5MHz offset from the center frequency 1.9GHz. Hence, the fundamental frequency of the envelope response was placed at 10MHz, which was set to reduce the transient simulation time. In real applications, the fundamental frequency term can be adjusted to be fit for an on-chip ADC available. From the extracted envelope responses, one period of the envelopes for each instance was used as an input parameter for mapping. The specification of interest includes noise figure (NF), S21, third-order-intercept (TOI), and quiescent current consumption (Idd).

Figure 49 shows the goodness of fit for predicting the optimal tuning for each knob. For perfect prediction, the scatter plots should be placed on the 45 degree straight line. As can be seen from the scatter plots, high prediction accuracy can be obtained from the proposed method. In terms of standard deviation of the difference between actual optimum knob values and the predicted one, both cases show 0.03mA and 0.6mV for I_{bias_LNA} and V_{bias_PMOS}, respectively. These degrees of standard deviation correspond to 3% and 1.1% over the entire variable range of each tuning knob, respectively.

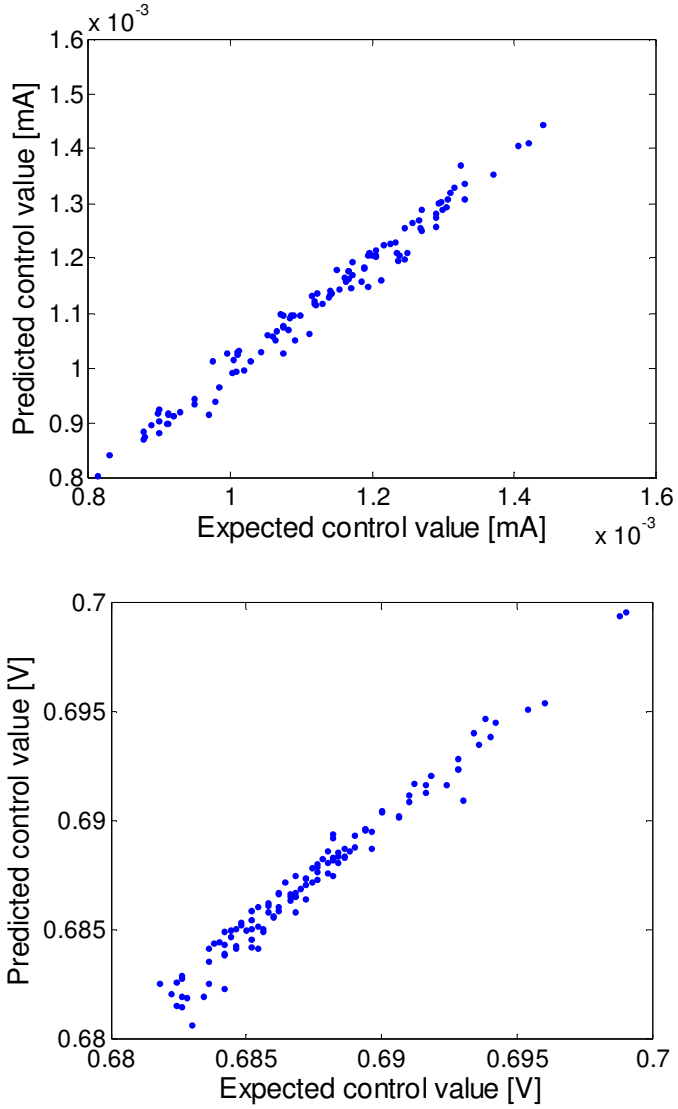


Figure 49. Goodness of fit of predicting optimal tuning knob control for I_{bias_LNA} and V_{bias_PMOS}.

Based on the predicted tuning knobs' values, the changes of each specification were evaluated. Figure 50 shows the distribution of each specification before and after calibration for the validation set. The performance variability is significantly reduced without changes of the mean value of each specification. The changes of each specification are listed in Table 15 in terms of mean μ and standard deviation σ. For

example, the standard deviation of the specification TOI shows 2.8612dBm before calibration and is reduced to 0.8236 dBm after calibration. Note that the quiescent current consumption (i.e., power consumption) is also reduced without compromising the performance metrics. Considering the power consumption of mobile systems is one of the most critical metrics, this feature significantly contributes to reduce the overall power consumption without loss of performance.

Table 15. Changes of the specification after calibration.

	Before		After		Ratio
	μ_b	σ_b	μ_a	σ_a	(σ_a /σ_b)
TOI [dBm]	15.76	2.86	15.77	0.86	0.30
NF [dB]	1.72	0.07	1.72	0.03	0.43
S21 [dB]	12.64	0.62	12.68	0.46	0.74
Idd [mA]	7.10	0.14	7.09	0.07	0.5

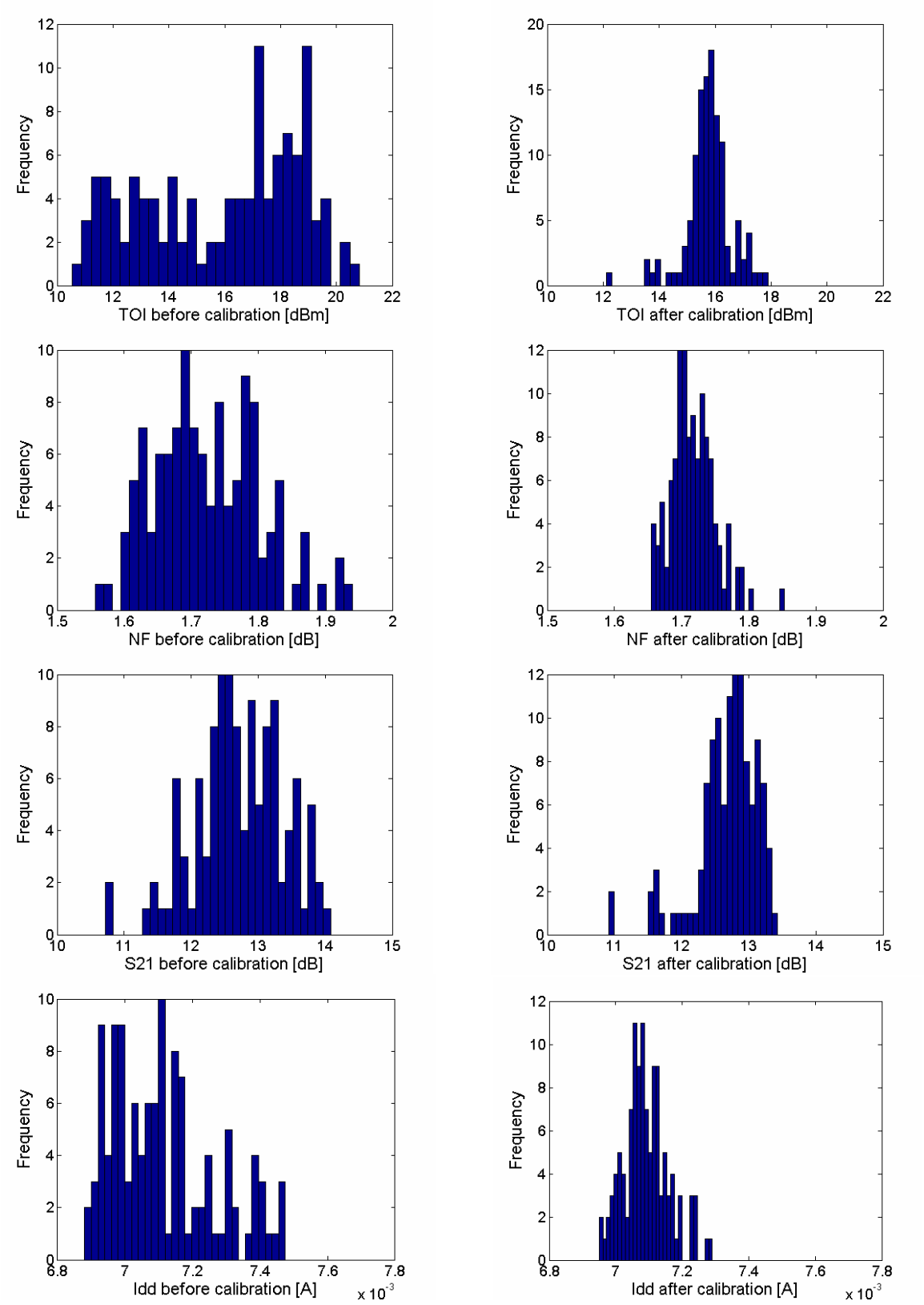


Figure 50. Specification distribution before and after calibration for case study I.

5.5.2 Case Study II: Fabrication on TSMC 0.25µm Technology

The proposed self-calibration structure has been applied to a source-degenerative CMOS LNA and partially implemented on TSMC 0.25µm CMOS technology. The simplified schematic depicted in Figure 51 shows a LNA with an embedded envelope detector at the output of the LNA. As a tuning knob, the bias current of the transistor M_1 was considered through the current mirror. All the components are integrated on the die except the gate inductance L_g and the control circuitry to adjust the bias current I_{bias_LNA}.

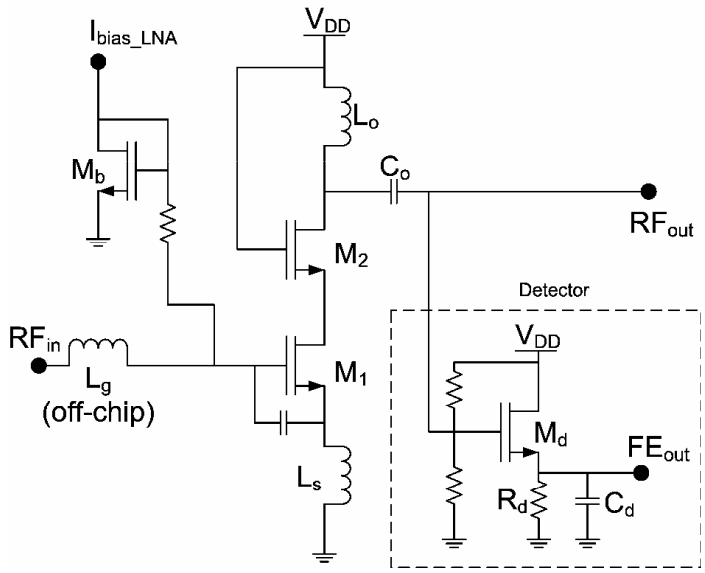


Figure 51. Fabricated circuit in TSMC 0.25µm CMOS technology.

Figure 52 displays the photomicrograph of the fabricated chip. The physical chip layout area including bond pads is approximately 1mm×0.7mm where the embedded detector occupies a design space of 0.13mm×0.11mm. The die area overhead for the embedded detector over the entire die area is 2%.

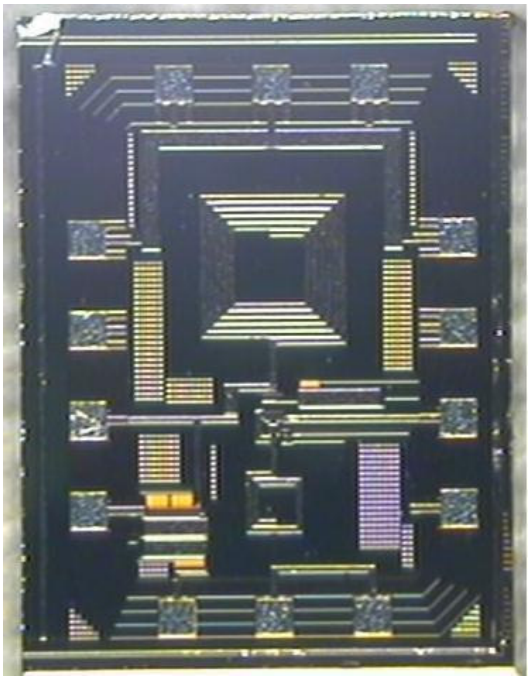


Figure 52. Photomicrograph of the fabricated chip.

The RF detectors present in the literature [74], [77], [100] are listed in Table 16 to compare die area overheads. As can be seen, the envelope detector occupies small die area comparable to other RF detectors.

Table 16. Die area of published detectors.

Ref.	Detector type	Die area	Technology
This work	Envelope det.	0.0143mm^2	TSMC 0.25μm
[74]	*RMS det.	0.66mm^2	IBM 6HP ICMOS
[100]	*Power det.	0.33 mm^2	IBM 6HP BICMOS
[77]	RMS det.	0.0135 mm^2	TSMC 0.35um

(* the area includes bond pads and the chip edge)

All the measurements were performed on the modules fabricated in a chip-on-board (COB) process. An off-the-self component for the gate inductor was employed on the module, and two DC-block capacitors are placed on both the input and output of the LNA. Figure 53 shows the S-parameters, which are measured by a network analyzer Agilent E8363B.

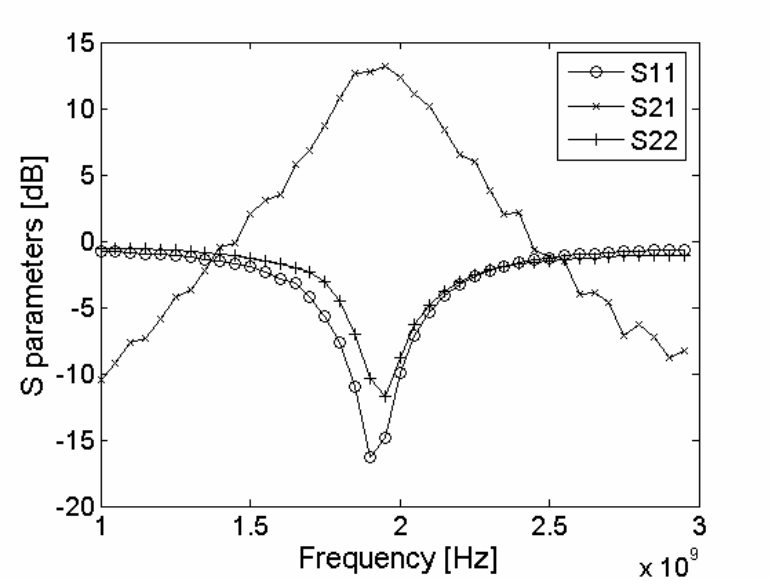


Figure 53. Measured S-parameters on the fabricated chip.

5.5.2.1 Test setup

As the self-calibration configuration is shown in Figure 46, Control circuitry for the tuning knob, ADC, and DSP modules were emulated via a DC source meter Keithley 2400, and NI-DAQ card PCI-6115, and PC, respectively. The source meter is controlled via GP-IB connection using the PC. The overall test setup is depicted in Figure 54.

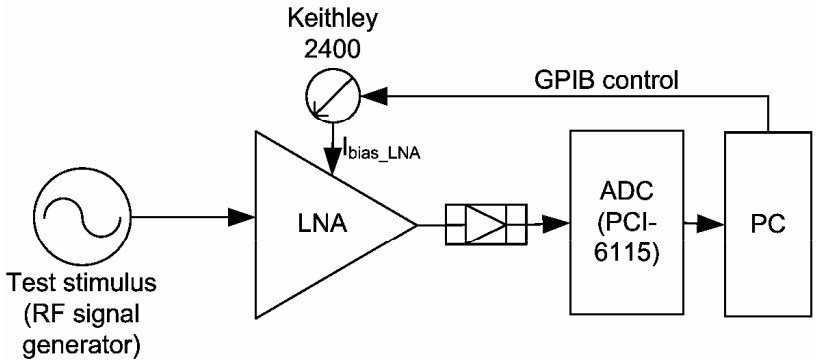


Figure 54. Measurement setup.

As described earlier, at least a few hundred dies are required to explore statistical relationship between the transient response and the optimal control knobs. However, three available dies given are not enough for this purpose. In addition, the fabricated dies show very narrow performance deviations from each other such as 0.2dB deviation for the specification S21, which is not a set of good samples to investigate DUTs with large process variations. Therefore, to make multiple instances as well as large performance deviations expected in very deep sub-micron technologies, the main power supply level was swept from 1.3V to 2.5V. Though this experiment scheme does not fully emulate the cases expected in real production, it can show feasibility of the proposed method along with the simulations described in the previous section. Through sweeping the power supply level for three dies, 75 instances were totally generated. Similar to the test scheme used for the simulation, the optimum tuning knob's values were searched for and logged through sweeping all possible tuning knobs' values for each instance. Among 75 instances generated, 50 instances were utilized for generating nonlinear regression mapping functions and the remaining 25 instances were utilized as a validation set.

As a test stimulus, a two-tone sinusoidal waveform was utilized with -12dBm at ±10KHz offset from the center frequency 1.9GHz. The envelope response via the embedded detector has 50usec period length, and was captured during 5msec corresponding to 100 periods of an envelope response. Then, one reprehensive period was extracted after time-averaged to reduce noise.

5.5.2.2 Experimental results

Figure 55 shows the scatter plots of the predicted tuning knob values versus the actual optimum ones for the validation set, where one dot corresponds to one instance.

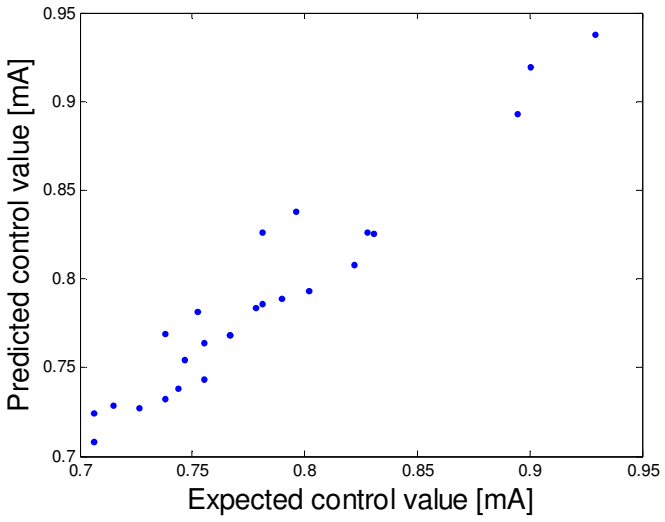


Figure 55. Actual control values versus predicted control values.

Though 50 instances are not enough to explore all the inherent statistics and thereby generate mapping functions, the figures show proper goodness of fit with 0.0157mA standard deviation of the difference between the actual and predicted ones.

Based on the predicted tuning knob's value via the proposed method, each DUT was calibrated, and the specifications of interest were measured using standard test setup and RF equipment. Figure 56 shows the distribution of each specification with respect to the normalized distribution. In the figure, the distribution of each specification is normalized with respect to its mean and standard deviation values before calibration. The dotted line shows the distribution before calibration as a reference, and we assume that all the specifications are normally distributed. To efficiently compensate for loss of performance, the variability of each distribution needs to be less than the normalized distribution (dotted line).

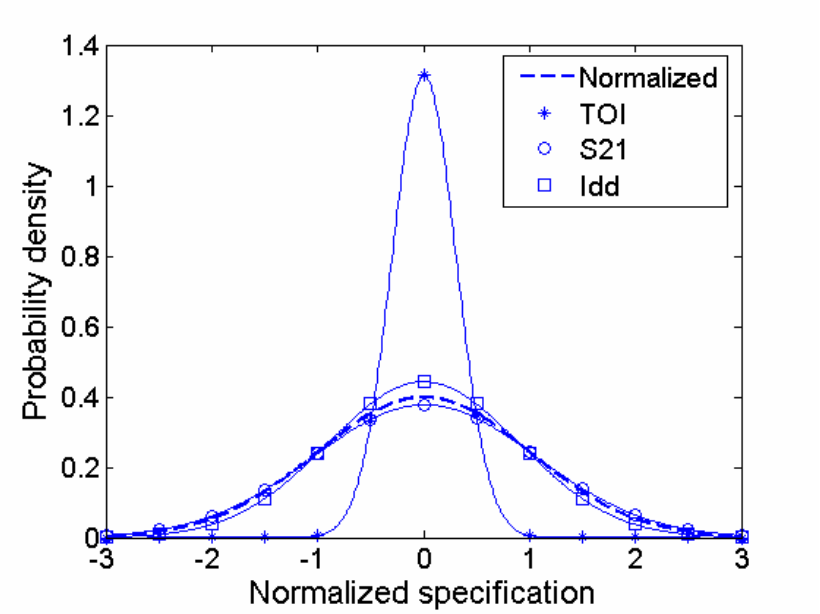


Figure 56. Specification distribution after calibration for case study II.

All the mean and standard deviation values before and after calibration are listed in Table 17. As can be observed, the variability of the specification TOI is significantly

reduced from 0.5633dBm to 0.1730dBm in terms of standard deviation while the variability of the specification S21 and the quiescent current consumption were almost unchanged.

Table 17. Specification variability after calibration for the fabricated die.

	Before		After		Ratio
	μ_b	σ_b	μ_a	σ_a	(σ_a /σ_b)
TOI [dBm]	6.08	0.56	6.05	0.17	0.30
S21 [dB]	11.64	0.35	11.31	0.37	1.05
Idd [mA]	4.10	0.10	4.10	0.09	0.9

5.6 Impact on Parametric Yields

From the results through two case studies described above, impact of the proposed method on parametric yield enhancement is explored. For this, we assume that parametric failure of each specification arises if its value is deviated greater than ±10 percent with respect to the expected, which is set to the mean value of each specification. All the specifications are assumed to be distributed normally. Table 18 lists parametric yields before and after calibration, Y_{before} and Y_{after} of each specification. As can be inferred from the results in Table 15 and Table 17, significant increments of parametric yields are achieved via the proposed method. For example, parametric yield of the specification TOI in case study I is 41.8% before performing calibration and is improved by 57.99% after calibration.

Table 18. Impact on parametric yield for two case studies.

	Case study I				Case study II		
	TOI	NF	S21	Idd	TOI	S21	Idd
Y_{before}[%]	41.8	98.6	95.8	100	72.2	99.9	100
Y_{after}[%]	93.3	100	99.4	100	99.9	98.8	100
Increment [%]	+57.9	+1.4	+3.9	0	+27.8	-1.3	0

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of Results

This dissertation has addressed production test and self-calibration issues to provide a low-cost production testing solution and enhance parametric yield. First, a novel cost metric for analog circuit sizing is proposed that significantly reduces the production test cost for specified process statistics. The sizing problem is solved using a cost metric that is formulated in terms of the correlation values between the different test specifications, the test time corresponding to each specification, and the manufacturing yield of the specification (percentage of ICs that pass the test). The proposed cost metric can be incorporated into existing optimization tools for manufacturing yield enhancement without incurring significant computational costs, resulting in a circuit sizing methodology that simultaneously improves both yield and test cost while guaranteeing that all target design specifications are met. Through the case study for two types of op-amps, the performance and feasibility of test cost-driven circuit sizing has been verified, showing 45 % and 36% reduction of test times as compared to typical yield-driven sizing with little compromise of manufacturing yields.

For highly integrated and high frequency circuits, the built-in self-test approach is proposed and incorporated into the alternate test methodology. Due to aggressive

technology scaling and multi-GHz operating frequencies of RF devices, parametric failure test and diagnosis of RF circuitry is becoming increasingly important for reduction of production test cost and faster yield ramp-up. In this research, a low-cost test and diagnosis method is presented for multi-parametric faults in wireless systems that allow accurate prediction of the end-to-end specifications as well as the specifications of all the embedded RF modules. The low frequency envelope of the RF transient response is shown to provide the features for specification prediction without external RF ATE support. Moreover, the proposed scheme just employs a single test pattern on a single test configuration to extract all the specifications of the system and its sub-modules. Simulations and hardware measurements performed on the RF transceiver and its sub-modules have demonstrated that considerable test accuracy can be achieved using a single test pattern, comparable to measurement noise level obtained via standard RF instrumentation.

After the design and production test phase, the self-calibration technique for RF circuitry has been described to compensate for large performance variability due to process variations. The embedded detector makes on-chip resources used for self-calibration. Moreover, one-shot optimization procedure eliminates long iterations to search for optimum trim of tuning knobs in typical optimization engine. As a result, the RF circuits can perform diagnosis and control multiple tuning knobs without external assistance from test equipment. In turn, multiple performance variability can be adjusted simultaneously with multiple tuning knobs. Through transistor-level simulation and experiments on the fabricated 1.9GHz CMOS LNAs, it is shown that the performance variability is significantly reduced, and thereby the parametric yields are enhanced up to 58 %.

6.2 Future Research Directions

The research presented in this dissertation introduces the framework addressing built-in self-test and calibration for RF circuits, coupled with a design methodology incorporating production test costs into circuit design. In this section, a number of future research directions are outlined.

Compensation for power-driven reconfigurable RF front-ends

This dissertation has studied the design of adaptive RF modules that adapt dynamically to process variations. Such dynamic adaptation can involve simultaneous V_{dd} modulation of RF circuit modules, modulation of ADC and DAC word size and sampling rate, and feedback mechanisms between a transmitter and receiver pair to enable both of them to operate at the minimum power level necessary for maintaining communication. The objective is to ensure that the receiver error vector magnitude (EVM, a measure of received signal quality) is always below a specified critical value. Clearly, as reconfiguration of the transceiver is performed dynamically to minimize power, appropriate tuning/compensation will need to be performed to ensure "best" transceiver performance under the current reconfigured transceiver operating environment. This can, perhaps be accomplished by treating supply voltage Vdd, for example, as an additional input to the MARS regression models.

Jitter and phase noise measurement technique using envelope response

The simple diode-based circuitry has shown the down-conversion functionality in the research. Such feature can be extended to jitter and phase noise measurements, which are typically done on expensive standard testers such as signal analyzer and high-speed

oscilloscope. In high volume manufacturing, the envelope detector-based test circuitry can be designed to accomplish jitter expansion. The low-speed output of the test detector can be routed to an external tester for analysis without the attendant signal integrity problems that occur with transmission of very high speed signals. The expanded (reconstructed) jitter of the low speed signal can be analyzed with a conventional time interval analyzer (TIA) or a low-speed oscilloscope. The jitter measurement approach has significant potential to offer the following benefits over existing jitter measurement techniques:

(a) *High resolution and accuracy*: A duty-cycle resolution of 0.1% of the UI corresponds to timing accuracy resolution of 1 psec for 1GHz signal. The jitter measurement technique holds the promise of sub-picosecond jitter measurement capability for multi-Gbps signals without the expense of precision external test equipment.

(b) *Simplicity/robustness of jitter expansion circuitry:* The jitter expansion circuit is easy to implement and is robust to component variations and device nonlinearities.

(c) *Ability to calibrate for reference signal noise:* The jitter measurement technique requires the use of a sinusoidal waveform with low phase noise value. In case a "perfect" source is not available, it is possible to calibrate for the phase noise of the reference signal (within reasonable bounds) to improve overall jitter measurement capability of the hardware.

REFERENCES

[1] S. Director, W. Maly, and A. Strojwas, *VLSI Design for Manufacuring: Yield Enhancement.* Norwell, MA: Kluwer, 1990.

[2] J. Turino, "Test economics in the 21st century," *IEEE Design & Test*, vol. 14, pp. 41-44, 1997.

[3] R. Kramer, "Test throughput for mixed-signal devices," *IEEE Instrumentation and Measurement Mag.*, vol. 8, pp. 12-15, 2005.

[4] S. R. Nassif, "Modeling and analysis of manufacturing variations," In *Proc. IEEE Conf. on Custom Integrated Circuits, 2001*, pp. 223-228.

[5] S. R. Nassif, "Design for variability in DSM technologies," in *Proc. Int. Symp. Quality Electronic Design*, 2000, pp. 451-455.

[6] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. John Willy & Sons, Inc., 4th edition, 2001.

[7] W. Maly, "Computer-aided design for VLSI circuit manufacturability," *Proc. IEEE*, vol. 78, pp. 356-392, 1990.

[8] M. Soma, "Challenges in analog and mixed-signal fault models," *IEEE Circuits and Devices Mag.*, pp. 16-19, Jan. 1996.

[9] W. Lindermeir, H. Graeb, and K. Antreich, "Analog testing by characteristic observation inference," *IEEE Trans. Computer-Aided Design*, vol. 18, no. 9, pp. 1353-1368, Sept. 1999.

[10] L. Milor and A. L. Sangiovanni-Vincentelli, "Minimizing production test time to detect faults in analog circuits," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 796-813, 1994.

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

[12] S. L. Hurst, *VLSI Testing: Digital and Mixed Analogue/Digital Techniques.* Technology & Industrial Arts, 1998.

[13] W. M. Lindermeir, "Design of robust test criteria in analog testing," In *Proc. Int. Conf. Computer-Aided Design*, 1996, pp. 604-611.

[14] S. Ozev, C. Olgaard, and A. Orgiloglu, "Testability implications in low-cost integrated radio transceivers: a Bluetooth case study," In *Proc. Int. Test Conf.*, 2001, pp. 965-974.

[15] J. Ferrario, R. Wolf, S. Moss, and M. Slamani, "A low-cost test solution for wireless phone RFICs," *IEEE Communication Magazine*, vol. 41, pp. 82-89, 2003.

[16] S. Ozev, C. Olgaard, and A. Orailoglu, "Multi-level testability analysis and solutions for integrated bluetooth transceivers," *IEEE Design & Test of Computers*, vol. 19, pp.82-91, 2002.

[17] J. Brockman and S. Director, "Predictive subset testing: optimizing IC parametric performance testing for quality, cost, and yield," *IEEE Trans. Semiconductor Manufacturing*, vol. 2, no. 3, pp. 104-113, 1980.

[18] S. Benner and O. Boroffice, "Optimal production test times through adaptive test programming," In *Proc. Int. Test Conf.*, 2001, pp. 908-915.

[19] S. D. Huss and R. S. Gyurcsik, "Optimal ordering of analog integrated circuit tests to minimize test time," In *Proc. IEEE Design Automation Conf.*, 1991, pp. 494-499.

[20] W. Jiang and B. Vinnakota, "Defect-oriented test scheduling," In *Proc. IEEE VLSI Test Symp.*, Nov. 1992, pp. 145-150.

[21] K. M. Butler and J. Saxena, "An empirical study on the effects of test type ordering on overall test efficiency," In *Proc. Int. Test Conf.*, 2000, pp. 408-416.

[22] *International Technology Roadmap for Semiconductor (ITRS)*, the ITRS 2006 Update. [Online]. Available: http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm (Accessed Jan. 03, 2007).

[23] P.N. Variyam, S. Cherubal, and A. Chatterjee, "Prediction of analog performance parameters using fast transient testing," *IEEE Trans. Computer-Aided Design*, vol. 21, pp. 349-361, 2002.

[24] A. Haldar, S. Bhattacharya, and A. Chatterjee, "Automatic multitone alternate test generation for RF circuits using behavioral models," In *Proc. Int. Test Conf.*, 2003, pp. 665-673.

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

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

[27] R. Voorakaranam, S. Cherubal, and A. Chatterjee, "A signature test framework for rapid production testing of RF circuits," In *Proc. Design Automation and Test in Europe*, 2002, pp. 4-8.

[28] S. S. Akbay, A. Halder, A. Chatterjee, and D. Keezer, "Low cost test of embedded RF/analog/mixed-signal circuits in SOPs," *IEEE Trans. on Advanced Packaging*, vol. 27, pp. 352-263, 2004.

[29] D. Han and A. Chatterjee, "Robust built-in alternate test of RF ICs using envelope," In *Proc. International Mixed-Signals Test Workshop*, 2005.

[30] J. H. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics*, vol. 19, pp. 1-141, 1991.

[31] E. Felt, S. Zanella, C. Guardiani, and A. Sangiovanni-Vincentelli, "Hierarchical statistical characterization of mixed-signal circuits using behavioral modeling," In *Proc. Int. Conf. Computer-Aided Design*, 1996, pp. 374-380.

[32] H. Yoon, P. Variyam, A. Chatterjee, and N. Nagi, "Hierarchical statistical inference model for specification based testing of analog circuits," In *Proc. IEEE VLSI Test Symp*., 1998, pp. 145-150.

[33] K. Anteich, H. Graeb, and C. Wieser, "Circuit analysis and optimization driven by worst-case distances," *IEEE Trans. on Computer-Aided Design*, vol. 13, no. 1, pp. 57-71, 1994.

[34] C. Guardiani, P. Scandolara, J. Benkoski, and G. Nicollini, "Yield optimization of analog IC's using two-step analytic modeling methods," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 7, pp. 778-783, 1993.

[35] J. M. Wojciechowski and J. Vlach, "Ellipsoidal method for design centering and yield estimation," *IEEE Trans. on Computer-Aided Design*, vol. 12, no. 10, pp. 1570-1579, 1993.

[36] T. Mukherjee, and L. R. Carley, "Rapid yield estimation as a computer aid for analog circuit design," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 3, pp. 291-299, 1991.

[37] G. Debyser and G. Gielen, "Efficient analog circuit synthesis with simultaneous yield and robustness optimization," In *Proc. International Conf. Computer-Aided Design*, 1998, pp. 308-318.

[38] T Mukherjee, L. R. Carley, and R. A. Rutenbar, "Efficient handling of operating range and manufacturing line variations in analog cell synthesis," *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems*, vol. 19, no. 8, pp. 825-839, 2000.

[39] J.K. Cavers and M.W. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," *IEEE Trans. Vehicular Technology*, vol. 42, no. 4, pp. 581-588, Nov. 1993.

[40] M. Faulkner, T. Mattson, and W. Yates, "Automatic adjustment of quadrature modulators," *Electronic Letter*, vol. 27, no. 3, pp. 214-216, 1991.

[41] M. Elmala and S. Embabi, "A self-calibration technique for mismatches in image-reject receivers," In *Proc. IEEE Custom Int. Circuits Conf.*, 2002, pp. 251-254.

[42] T. Das, A. Gopalan, C. Washburn, and P. R. Mukund, "Self-calibration of input-match in RF front-end circuitry," *IEEE Trans. Circuits and Systems*, vol. 52, no. 12, pp. 821-825, Dec. 2005.

[43] G. Gielen and R. A. Rutenbar, "Computer-aided design of analog and mixed-signal integrated circuits," *Proc. of the IEEE*, vol. 88, no. 12, pp. 1825-1852, 2000.

[44] W. Daems, G. Gielen, and W. Sansen, "Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 5, pp. 517-533, May 2003.

[45] F. Medeiro, F. V. Fernández, R. Domínguez-Castro, and A. Rodríguez-Vázquez, "A statistical optimization-based approach for automated sizing of analog cells," In *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 1994, pp. 594-597.

[46] R. Phelps, M. Krasnicki, R. A. Rutenbar, L. R. Carley, and J. R. Hellums, "Anaconda: simulation-based synthesis of analog circuits via stochastic pattern search," *IEEE Trans. Computer-Aided Design*, vol. 19, pp.703-717, June 2000.

[47] T. Mukherjee, L. R. Carley, and R. A. Rutenbar, "Efficient handling of operating range and manufacturing line variations in analog cell synthesis," *IEEE Trans. Computer-Aided Design*, vol. 19, no. 8, pp. 825-839, 2000.

[48] C. M. Berrah, "Parametric yield estimation for a MOSFET integrated circuit," In *Proc. IEEE Symp. Circuits and Systems*, 1990, pp. 2260-2263.

[49] T. K. Yu, S. M. Kang, J. Sacks, and W. J. Welch, "Parametric yield optimization of CMOS analogue circuits by quadratic statistical circuit performance models," *Int. Journal Circuit Theory and Application*, vol. 19, pp. 579-592, 1991.

[50] T. K. Yu, S. M. Kang, J. Sacks, and W. J. Welch, "An efficient method for parametric yield optimization of MOS integrated circuits," In *Proc. IEEE Int. Conf. Computer-Aided Design*, 1989, pp. 190-193.

[51] A. Papoulis, *Probability and Statistics.* Prentice-Hall, 1990.

[52] S. A. Aftab and M. A. Styblinski, "IC variability minimization using a new Cp and Cpk based variability/performance measure," In *Proc. IEEE Int. Symp. Circuits and Systems*, vol.1, 1994, pp. 149–152.

[53] E. Felt, S. Zanella, C. Guardiani, and A. Sangiovanni-Vincentelli, "Hierarchical statistical characterization of mixed-signal circuits using behavioral modeling," In *Proc. Int. conf. Computer-Aided Design*, 1996, pp. 374-380.

[54] D. B. Ribner and M. A. Copeland, "Design techniques for cascade CMOS op amps with improved PSRR and common-mode input range," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 919-925, Dec. 1984.

[55] P. Alotto, A. Caiti, G. Molinari, and M. Repetto, "A multiquadrics-based algorithm for the acceleration of simulated annealing optimization procedures," *IEEE Trans. Magnetics,* vol. 32, pp.1198-1201, 1996.

[56] P. Vancorenland, C. De Ranter, M. Steyaert, and G. Gielen, "Optimal RF design using smart evolutionary algorithms," In *Proc. Design Automation Conference*, 2000, pp. 7-10.

[57] W. Daems, G. Gielen, and W. Sansen, "Simulation-based automated generation of signomial and posynomial performace models for analog integrated circuit sizing," In *Proc Int. Conf. Computer-Aided Design*, 2001, pp.70-74.

[58] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experiments: An Introduction to Design, Data Analysis, and Model Building*. John Wiley & Sons, 1978.

[59] M. McKay, R. Beckman, and W. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, pp. 239-246, 1979.

[60] G. Taguchi, *Introduction to Quality Engineering.* Dearborn, MI: Distributed by American Supplier Institute, Inc., 1986.

[61] L. C. W. Dixon and G. P. Szegö, "The global optimization problem: an introduction," *Towards Global Optimization 2*, North-Holland, Amsterdam, I-15.

[62] D. Vanderbilt and S. G. Louie, "A Monte Carlo simulated annealing approach to optimization over continuous variables," *Journal of Computational Physics*, vol. 56, pp. 259-271, 1984.

[63] H. M. Gutmann, "A radial basis function method for global optimization," *Journal of Global Optimization*, vol. 19, pp. 201-227, 2001.

[64] J. W. Bandler and A. E. Salama, "Fault diagnosis of analog circuits," *Proc. of the IEEE*, vol. 73, no. 8, pp. 1279-1325, 1985.

[65] S. S. Somayajula, E. Sanchez-Sinencio, and J. Pineda de Gyvez, "A power supply ramping and current measurement based technique for analog fault diagnosis," In *Proc. VLSI Test Symposium*, April 1994, pp. 234-239.

[66] S. Cherubal and A. Chatterjee, "Parametric fault diagnosis for analog systems using functional mapping," In *Proc. Design, Automation and Test in Europe Conf. and Exhibition*, March 1999, pp. 195-200.

[67] P. Wang and S. Yang, "A new diagnosis approach for handling tolerance in analog and mixed-signal circuits by using fuzzy math," *IEEE Trans. Circuits and Systems I*, vol. 52, no. 10, pp. 2118-2127, 2005.

[68] C. Alippi, M. Catelani, A. Fort, and M. Mugnaini, "Automated selection of test frequencies for fault diagnosis in analog electronic circuits," *IEEE Trans. Instrumentation and Measurement*, vol. 54, no. 3, pp.1033-1044, 2005.

[69] R. Voorakaranam, A. Cherubal, and A. Chatterjee, "A signature test framework for rapid production testing of RF circuits," In *Proc. Design Automation and Test in Europe Conf. and Exhibition*, March 2002, pp. 186-191.

[70] E. Acar and S. Ozev, "Defect-based RF testing using a new catastrophic fault model," In *Proc. IEEE Int. Test Conf.*, Nov. 2005, paper 17.3.

[71] E. Acar and S. Ozev, "Diagnosis of the failing components in RF receivers through adaptive full-path measurements," In *Proc. VLSI Test Symposium*, May 2005, pp. 374-379.

[72] M. S. Heutmaker and D. K. Le, "An architecture for self-test of a wireless communication system using sampled IQ modulation and boundary scan," *IEEE Communications Magazine*, vol. 37, no. 6, pp. 98–102, 1999.

[73] A. Halder, S. Bhattacharya, G. Srinivasan, and A. Chatterjee, "A system-level alternate test approach for specification test of RF transceivers in loopback mode," In *Proc. Int. Conf. VLSI Design*, Jan. 2005, pp. 289-294.

[74] A. Yin, W. R. Eisenstadt, R. M. Fox, and T. Zhang, "A translinear RMS detector for embedded test of RF ICs," *IEEE Trans. Instrumentation and Measurement*, vol. 54, no. 5, pp. 1708-1714, 2005.

[75] J. Ryu, B. C. Kim, and I. Sylla, "A new low-cost RF built-in self-test measurement for system-on-chip transceivers," *IEEE Trans. Instrumentation and Measurement*, vol. 55, no. 2, pp. 381-388, 2006.

[76] A. Gopalan, T. Das, C. Washburn, and P. R. Mukund, "An ultra-fast, on-chip BiST for RF low noise amplifiers," In *Proc. Int. Conf. VLSI Design*, Jan 2005, pp. 485-490.

[77] A. Valdes-Garcia, R. Venkatasubramanian, R. Srinivasan, J. Silva-Martinez, and E. Sanchez-Sinencio, "A CMOS RF RMS detector for built-in testing of wireless transceivers," In *Proc. VLSI Test Symposium*, May 2005, pp. 249-254.

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

[79] J. Dugundji, "Envelope and pre-envelopes of real waveforms," *IEEE Trans. Information Theory*, vol. 4, no. 1, pp. 53-57, 1958.

[80] B. Razavi, *RF Microelectronics.* Prentice Hall, 1998.

[81] K. S. Kundert, "Introduction to RF simulation and its application," *IEEE Journal of Solid State Circuits*, vol. 34, no. 9, pp. 1298-1319, 1999.

[82] R. Hooke and R. A. Jeeves, "Direct search solution of numerical and statistical problems," *Journal Assoc. Comp.*, vol. 8, pp. 212-229, 1961.

[83] Cadence SpectreRF Simulator User Guide.

[84] J. Yoon and W. R. Eisenstadt, "Embedded loopback test for RF ICs," *IEEE Trans. Instrumentation and Measurement*, vol. 54, no. 5, pp. 1715-1720, 2005.

[85] R. Goering and R. Wilson, Yield, packages hang up design below 100nm, EE Times, March 31, 2003.

[86] J. Fattaruso, S. Kiriaki, M. Wit, and G. Warwar, "Self-calibration techniques for a second-order multibit sigma-delta modulator," *IEEE Journal of Solid-State Circuits*, vol. 28. no. 12, pp. 1216-1223, 1993.

[87] K. Hagaraj, "Area-efficient self-calibration techniques for pipe-lined algorithmic A/D converters," *IEEE Trans. Circuits and Systems-II: Analog and Digital Processing*, vol. 43, no. 7, pp.540-544, 1996.

[88] S. Bandi and P. R. Mukund, "A compensation technique for transistor mismatch in current mirrors," In *Proc. System-On-Chip Conf.*, 2003, pp. 320-323.

[89] G. Gramegna, M. Paparo, P. Erratico, and P. Vita, "A sub-1-dB 2.3-kV ESD-protected 900-MHz CMOS LNA," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1010-1017, 2001.

[90] J. Li, J. Yeh, R. Huang, and C. Wu, "A built-in self-repair design for RAMs with 2-D redundancy," *IEEE Trans. VLSI Systems*, vol. 13, no. 6, pp. 742-745, June 2005.

[91] T. Chen and G. Sunada, "Design of a self-testing and self-repairing structure for highly hierarchical ultra-large capacity memory chips," *IEEE Trans. VLSI Systems*, vol. 1, no. 2, pp. 88-97, June 1993.

[92] J. K. Cavers and M.W. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," *IEEE Trans. Vehicular Technology*, vol. 42, no. 4, pp. 581-588, Nov. 1993.

[93] M. Faulkner, T. Mattson, and W. Yates, "Automatic adjustment of quadrature modulators," *Electronic Letter*, vol. 27, no. 3, pp. 214-216, 1991.

[94] M. Elmala and S. Embabi, "A self-calibration technique for mismatches in image-reject receivers," In *Proc. IEEE Cutom Int. Circuits Conf.*, 2002, pp. 251-254.

[95] T. Das, A. Gopalan, C. Washburn, and P.R. Mukund, "Self-calibration of input-match in RF front-end circuitry," *IEEE Trans. Circuits and Systems*, vol. 52, no. 12, pp. 821-825, Dec. 2005.

[96] W. Woo, M. Miller, and J. Kenney, "A hybrid digital/RF envelope predistortion linearization system for power amplifiers," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, no. 1, pp. 229-237, 2005.

[97] T. Kim and B. Kim, "Post-linearization of cascade CMOS low noise amplifier using folded PMOS IMD sinker," *IEE Microwave and Wireless Components Letters*, vol. 16, no. 4, pp. 182-184, April 2006.

[98] 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, 2007* (to appear).

[99] T. Lee, *The design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2001.

[100] T. Zhang, W. R. Eisenstadt, and R. M. Fox, "A novel 5GHz RF power detector," In *Proc. International Symposium on Circuits and Systems*, vol. 1, 2004, pp. 897-900.