Silicon Demonstration of Statistical Post-Production Tuning

Yichuan Lu, Kiruba Subramani The University of Texas at Dallas Richardson, TX 75080 He Huang, Nathan Kupp Yale University New Haven, CT 06520 Yiorgos Makris The University of Texas at Dallas Richardson, TX 75080

Abstract—Towards meeting the continued demand for higher performance in analog/RF ICs, the semiconductor industry has resorted to aggressive design styles. In conjunction with the increased variation of modern processes, aggressive design jeopardizes yield and, by extension, profitability of the design. In an effort to facilitate aggressive high-performance design without compromising production yield, modern designs incorporate tuning knobs which are used after manufacturing to individually calibrate the performances of each chip, as well as some type of non-volatile memory to store the selected knob settings. In this work, we discuss a statistical post-production tuning method, which rapidly selects the appropriate knob positions using low-cost measurements, and we demonstrate it on a tunable Low Noise Amplifier which we designed and fabricated in IBMs 130nm RF CMOS process.

Keywords—knob tuning, low noise amplifier (LNA), postproduction calibration, regression, testing, yield

I. INTRODUCTION

In analog/RF design, circuits are typically designed conservatively in order to ensure high yield. Indeed, an aggressive design approach would increase the risk of device performances falling below the circuit specifications due to the impact of process variation. As fabrication technologies continue to scale down, the impact of process variation on the performances of analog/RF circuits has become more prominent. Accordingly, circuit designers are compelled to sacrifice more performance in order to ensure a reasonable yield. On the other hand, the semiconductor industry continuously demands higher-performing ICs without compromising yield in order to remain competitive. As a solution towards facilitating aggressive high-performance design as well as high yield, modern designs incorporate tuning knobs which are used after manufacturing to individually calibrate the performances of each device.

The idea behind post-production tuning is that an aggressive design would have built-in knobs which can calibrate the performances of the circuit. Through the change of knob value during a post-production testing phase, we can effectively compensate the effects of process variation on device performances individually on every fabricated IC. By selecting an appropriate knob setting, the device can be made to operate within its specification range, even if initially it fails to abide by them. Various methods exist for selecting an appropriate knob setting. Once a knob setting that brings device performances within the design specification is selected, the corresponding value is stored in non-volatile memory before the IC is deployed [1]. Thereby, both aggressive design and high yield are achieved, while the customer is oblivious that each IC may have been tuned to a different knob setting.

Tuning knobs are designed to affect device performances, but the interplay between the impact of process variation and the effect of knob tuning complicates the task of measuring and calibrating device performances. Since tuning is performed during test time, while a device is interfaced with Automatic Test Equipment (ATE), each additional test performed increases the cost of the device. In addition to the number of tests performed, the cost of a single test also affects the overall test cost (e.g. the cost of taking a high-frequency measurement is much higher than taking DC measurements). While this introduces challenges for selecting the optimal knob value, it also provides fertile ground for leveraging low-cost test techniques such as alternate test.

Several previous publications attack the problem of performance calibration in analog/RF devices. These approaches fall mainly into two categories, iterative [2], [3], [4], which select a knob setting through a test-&-tune sequence of iterations and one-shot [5], [6], [7], which only test and tune once. One-shot methods have an advantage over the iterative ones due to lower tuning time and cost, though iterative methods can potentially be slightly more accurate. Among the existing one-shot methods, [7] assumed that the knob effects can be approximated as first order models. These models require a strict knob design, so that the knobs have no effect on each other, however this comes with increased design difficulty. This assumption is revoked in [6], in a method called mid-point alternative test-based calibration, where the new assumption is that the impact of process variations and knob positions is orthogonal. A similar method is proposed in [5], however it employs non-intrusive sensors which have the advantage of not affecting the signal path.

In this paper, we demonstrate the application of mid-point alternate test-based calibration using measured data from a tunable Low Noise Amplifier (LNA) which we designed and fabricated in IBM's 130nm RF CMOS process. The remainder of this paper is structured as follows. Existing knob selection and performance calibration methods are discussed in section II. Section III introduces the fabricated LNA design that is used to demonstrate the concept of tuning knobs. The proposed method is, then, experimentally evaluated in section IV with the help of results from the fabricated chips.

II. KNOB SELECTION AND PERFORMANCE CALIBRATION

A. Specification test

As the industrial standard testing method, specification test is the most accurate and highest costing approach to determining pass/fail of a device. Each performance of the device under test (DUT) is directly measured through the ATE. By comparing measurements to specifications, a pass/fail decision is made. Even though this method is accurate, employing it for the purpose of knob tuning, wherein a device may have a



Configuration	on Total Testing Cost		Variable	Definition
	Test Set Term	Training Set Term	N test	Number of devices in test set.
Exhaustive Specification Test	$C=N_{test}*N_k*P$		$N_{\it train}$	Number of devices in training set.
Exhaustive alternate Test	$C = N_{test} *N_k *A$	$+ N_{train} *N_k *(A+P)$	N_k	Number of knob settings
Iterative approaches	$C = N_{test} *N_i *A$	$+ N_{train} *N_k *(A+P)$	N_{i}	Number of Iternation needed
One-shot approaches	C= N test *A	$+ N_{train} *N_k *(A+P)$	P	Relative cost for measuring all performances
			A	Relative cost for measuring all alternate tests

Fig. 1: Cost model for test and performance calibration methods

large number of settings for each knob, is impractical. Indeed, selection of the optimal knob setting requires specification testing for every performance and for every knob setting, the cost of which is prohibitive.

B. Alternate test

Alternate test is naturally suitable to address the problem of selecting an appropriate position for the tuning knobs while maintaining a low cost. In alternate test, instead of measuring the circuit performances which incurs high cost, a set of low-cost measurements are taken either externally or through on-chip sensors. These measurements are designed to correlate well with the circuit performances, while requiring far fewer test resources in order to be collected. Alternate test is a two-phase process. First, for an initial batch of devices, both the high-cost specification tests and the low-cost measurements are collected. From this data, regression models are then trained to correlate the low-cost measurements with the circuit performances. In the post-production phase, only the low-cost measurements are taken; performances are then predicted through the trained regression models and compared to the specifications in order to determine a pass/fail decision for the device. For a knob-tunable device, each knob setting would have its corresponding high-cost and low-cost measurements. A regression could then be trained to predict device performances for each knob setting based on the low-cost measurements for this or for all knob settings. This method can substantially reduce the overall test cost by replacing high-cost with low-cost measurements in the production phase.

C. Iterative Approaches

Following the general concept of alternate test, iterative methods incorporate a directed search approach with the help of low-cost measurements, in order to determine the optimal knob setting. Once regression models are trained in a post-production phase, these methods select a knob setting and perform an alternate test to determine the performances. The knob setting selection is performed iteratively, while the search function is optimized through each iteration. In short, a cycle of test-&-tune operations is repeated until a satisfying knob position is selected. If a device does not contain a passing knob setting, the algorithm stays in this test-&-tune loop until it determines that there is no passing knob setting or until a prespecified iteration limit is reached. These methods reduce the cost of calibration from a factor of all possible knob settings to the number of iterations needed to converge or the limit.

D. One-shot approaches

The aim of these methods is to determine the optimal knob setting without repeating a test-&-tune cycle. These methods rely on a set of low-cost measurements for only a single knob setting, based on which they make a decision on the optimal knob though statistical learning methods. By collecting low-cost measurements for a single knob setting, they minimize the time an integrated circuits spends on the ATE for the purpose of knob tunning. Herein, we implement one such one-shot calibration approach, namely mid-point alternate test, and we demonstrate its effectiveness using silicon data.

E. Cost Model

As we have discussed, one of the major obstacles of post-production calibration is the cost of testing/tunning. If the benefit of yield recovery does not cover the cost of testing/tunning, the calibration method should not be implemented. In Figure 1, we compare the cost model for different calibration methods we previously mentioned. Since all approaches except exhaustive specification utilize a regression model, the cost is divided into a test set term and a training set term. These models require a small training set where both alternate and specification tests are performed. The training set data is used to train regressions to correlate low-cost measurements with performances. Thereby, we eliminate the high-cost performance measurement P in the test set term. In high volume environments, $N_{test} \gg N_{train}$, therefore adaptation of alternate test substantially lowers the total test cost. For the test set term, one-shot approaches have the lowest testing cost due to the lack of need for a test-&-tune loop, and the ability to predict with one set of alternate tests.

F. Mid-point alternate test

Mid-point alternate test-based calibration [6] is the focus of this work, towards demonstrating statistical knob tunning. It is one of the prominent approaches in the one-shot calibration category due to its low test cost and high prediction accuracy. This method makes the assumption that the impact of process variations and tuning knobs on device performances are orthogonal. The overall performance model is:

$$\hat{P} = f(A, K) = \hat{\beta}_0 + A^T \hat{\beta}_a + K^T \hat{\beta}_k + \varepsilon \tag{1}$$

where $\hat{\beta}_0$ denotes the performance of a variation-free device, K^T is the vector of knob settings and all pairwise interaction terms, and $\hat{\beta}_k$ is the knob effect parameter estimated by:

$$K = (\underbrace{K_1, K_2, ..., K_p}_{\text{main effects}}, \underbrace{K_1 K_2, K_1 K_3, ... K_{p-1} K_p}_{\text{interaction terms}})^T \qquad (2)$$

$$\hat{\beta}_k = \underbrace{(\beta_1, \beta_2, ..., \beta_p, \beta_{1:2}, \beta_{1:3}, ..., \beta_{(p-1):p})}_{\text{interaction terms}}^T$$
(3)

The term $A^T \hat{\beta}_a$ represents the correlation between low-cost measurement and process variations. Since the test is based on alternate test, we observe the process variation in terms of low-cost measurement. Each alternate test gives a direct measurement of the magnitude of process variation effects, and all of the alternate tests improve the estimation of the

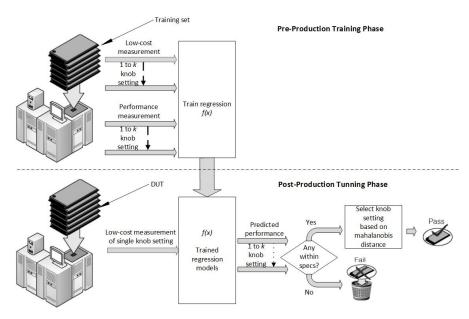


Fig. 2: Mid-point alternate test calibration flow

performances. As we can see in Equation 1, the term for reflecting the impact of process variations and tuning knobs are separate and assumed orthogonal. Following this model, we can generate regressions in order to predict performances.

With a basic understanding of how we model device performances, we can now proceed to the mid-point alternate test flow. Figure 2 demonstrates the complete test and calibration procedure. In the pre-production phase, the goal is to train a regression for each performance vector. Both low-cost and performance measurements for all knob settings are taken. The low-cost measurement are then used along with these functions in order to predict the performances of the device. The effect of changing the knobs on performance is modeled as an offset to the variation-free device performance. The trained regression takes low-cost measurements and knob settings as predictors, and gives a performance prediction for each knob setting as a response. In the post-production phase, low-cost measurements for the mid-point knob setting are taken from the DUT. The regressions then predict the performances for all knob settings of the device. If the predicted performances for all knob settings fail the design specifications, the device is classified as a failing device. If there exist more than one knob settings that would bring the device within its design specifications, the optimal knob setting is selected among them by using as a metric the normalized Mahalanobis distance, which maximizes distance from the specification planes.

III. TUNABLE DEVICE DESIGN

To demonstrate post-production tuning in silicon, we designed and fabricated a 2mm*2mm test chip in IBM's 130nm RF CMOS technology through the fabrication service provided by MOSIS. The chip contains RF tunable circuits, low-cost measurement sensors, a stimulus generator, and various peripheral and protection circuits. The chip is housed in a custom-designed evaluation board, which provides connection to external measurement instruments. Figure 3 shows a microphotograph of the fabricated die and Figure 4 shows the custom board housing the designed chip.

A. RF circuit

The RF circuit in this chip design is a cascode common source Low Noise Amplifier (LNA), which operates at a frequency of 1.575GHz. The LNA plays a critical role as it is the first component in the receiver chain so its noise figure is directly added to the receiver's overall noise, and the contribution of other components to receiver noise is scaled down by the gain of the LNA. As a result, a high performance LNA is a crucial part of high-end receiver design. This attribute makes the LNA an ideal candidate for demonstrating the tuning process in analog/RF circuits. We chose the cascode topology as our initial architecture, as it is one of the most universal RF LNA designs. The LNA schematic, as shown in Figure 5 (a), has three active devices: one amplifying common source transistor, one bias transistor, and one cascode transistor. The common source transistor provides the signal amplifying ability of the LNA. Its optimized width generates the minimal noise figure achievable for this fabrication technology for the given power consumption constraint. A cascode transistor of the same width is connected to the common source transistor to isolate the input from the load, therefore simplifying output matching. The bias voltage needed by the common source transistor is provided by the bias transistor, while the DC noise is minimized at the conjunction by a large resistor. A spiral inductor is placed at the drain of the cascode transistor along with other passives, for output stabilization and matching to a 50 ohm load. Input matching of 50 ohm is achieved through source degeneration provided by taking advantage of the parasitic bond-wire inductance. The specifications of the circuit are shown in Figure 5 (a).

B. Knob design

In the above circuit schematic, we were able to identify three key bias voltages as knobs that can be used for varying the performances of the LNA. Each of these three knobs has a different effect on the LNA performances and they are not orthogonal to each other. In other words, performance \hat{y} increases proportionally with the voltage on $knob_1$ and

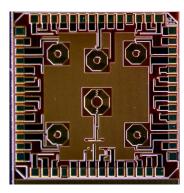


Fig. 3: Micro-photograph of fabricated die

 $knob_3$ individually. Yet the simultaneous increase of voltage on both $knob_1$ and $knob_3$ will decrease the performance \hat{y} . This property provides more freedom for knob selection and simplifies the knob design process, as the knobs are no longer constrained by the need for orthogonality. However, it also calls for more sophisticated tuning algorithms. The IBM 130nm fabrication technology has a default supply voltage of 1.2V and a breakdown voltage of 1.6V. To avoid high voltage that may cause damage to the devices, we set the upper limit of the knob voltage as 1.4V. The lower limit of knob voltage is bound by the device's ability to meet the specifications. Evidently, we selected the range to be from .8V to 1.4V, with a step size of .1V to provide sufficient granularity and range in the choice of knob positions. With 3 knobs and 7 knob positions, we have $7 \times 3 = 343$ total sets of knob combinations.

C. Low cost measurement sensor

The low-cost measurement sensor is implemented in the form of a peak detector. The peak voltage measurement has been shown to be highly correlated with LNA performances. Its input is connected to the output of the LNA in order to measure the positive peak voltage of the signal. The sensor captures the amplified signal at the output of the LNA. A capacitor acts as a high-pass filter, blocking the DC value and allowing the AC value to go through. The positive swing of the filtered signal charges the output capacitor to the peak value of the AC signal swing; as a result, the peak voltage is stored. A mirrored sensor circuit provides a measurement of the output capacitor voltage when there is no signal input at the sensing point. The final peak voltage is measured as the difference between the voltage of the two output capacitors.

D. Stimulus generator

A simple sine wave is the stimulus that we use for test. We generate this signal with an on-die voltage-controlled oscillator circuit. A cross-coupled PMOS-NMOS topology helps in achieving symmetry of the rising and failing times of the oscillation. The initial oscillation is created by a parallel resonant LC-tank. It contains a spiral inductor and two MOS varactors, which provide a resonant frequency range between .9GHz and 1.6GHz through the tunning of the MOS varactors. To sustain the generated oscillation, the active devices behave as "negative resistance" and counterbalance the LC tank's internal resistance. An input voltage is applied to the varactors to tune the output frequency of the sine wave. The generated differential signal is, then, connected to a differential-output to single-ended output conversion stage, which can be connected to the RF input of an LNA as stimulus in the desired frequency.

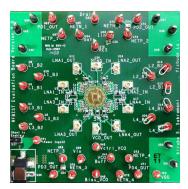


Fig. 4: PCB housing chip and interface connections

E. Chip and board overview

Each chip contains 4 LNA circuits, 4 peak detector circuits and one VCO circuit. Each LNA has a dedicated peak detector. The LNAs are replicas of each other, with rotation in the x and/or y axis. Each LNA and peak detector combination occupies a corner of the chip to ensure that the routing and the distances to chip pins are the same. With this setup, we are able to minimize the difference between each LNA on the same die, avoid the introduction of design variation, and generate more samples for testing purposes. The chips are fabricated in IBMs 130nm RF CMOS process through MOSIS. A total of 40 chips are fabricated with the above setup, which gives us 160 LNAs. Each chip is then individually mounted on a custom printed circuit board with chip-on-board technology. With the chip-on-board structure, we are able to remove the introduction of parasitics and variation by board sockets and IC packages. The boards are designed to ensure minimal variation. They provide a simple interface between the die and the test equipment, as well as a matching network. After extensive testing, we concluded that out of 40 boards, 4 were defective; therefore, we have 36 boards, or 144 LNA samples for test purposes.

F. Data collection

Each of the 36 boards is tested through two different setups. In setup 1, we collect the device's actual performances at every knob setting. Three DC power supplies each connect to a knob of the LNA, providing the tunning voltage. The input and output ports of the LNA are connected to a network analyzer to collect the S parameters (S11, S12, S21, and S22). As the power supplies sweep through each of the 343 knob settings, the S parameters of the LNA are measured and stored into a computer. Setup 2 collects the device's low-cost measurement data at every knob setting. The on-die VCO functions as a stimulus generator and provides a sine wave to the LNA input, while the dedicated peak detector of each LNA functions as a peak measuring sensor. The DC power supplies function in the same way as in setup 1. Another DC supply changes the VCOs control voltage, thereby varying the stimulus frequency. The output voltage of the peak detector is recorded for each frequency and for every knob position. Using the above two testing setups, we populated the data sheet shown in Figure 6.

IV. TESTING AND TUNING

Using all the collected data, we conducted various experiments to evaluate the effectiveness of alternate testing and statistical post-production tuning. In all cases, we split the data

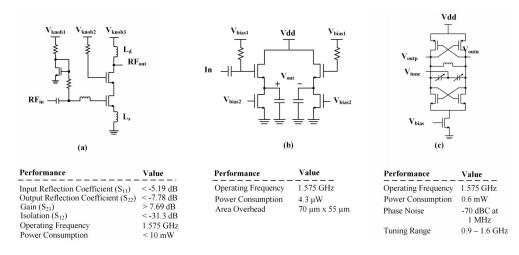


Fig. 5: Circuit design (a) LNA schematic (b) Peak detector schematic (c) VCO schematic

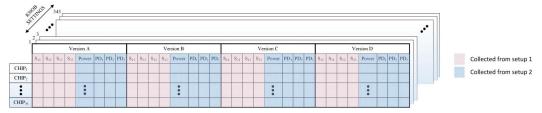


Fig. 6: Graphical depiction of dataset

into training and testing sets. Of the total of 144 LNAs, 80 were used for training and 64 were used for testing. To avoid any variation caused by the LNA position/orientation on the die, the training and the testing sets had a balanced number of LNAs from each of the four positions. For example, among the 80 LNAs in the training set, 20 were located in each of the top right, top left, bottom right, and bottom left corners of the die. To ensure statistical significance of the reported results, we performed 10 cross-validations for each experiment and averaged the results.

A. Specification test

The purpose of specification testing in our set of experiments is to emulate a knob free device, and define a baseline in order to compare the results of our statistical post-production calibration method. Specification testing reports the yield achieved by the process for this device, assuming that the design is knob-free. We emulate the knob-free device by using the center knob setting $(knob_1 = 1.2V, knob_2 = 1.2V, knob_3 = 1.2V)$. The performances of each device are then compared to the design specifications. If the device does not meet its specifications, it is defined as a failing device. The pass/fail ratio of our 144 LNAs in the center knob setting is as follows:

As indicated by this result, without calibration capabilities the LNA production would experience a yield loss of 31.1%. We expect the result to improve as we incorporate knobs in the design to recover this yield loss.

B. Exhaustive specification test

Exhaustive specification test is performed to collect the performances of a device in every knob setting and compare to the design specifications. If a device has at least one knob setting that would make it operate within the design specifications, then the device is healable. If all knobs of a device fail, then the device is unhealable. This result, which is shown below, provides the ceiling for any post production tuning method, as it is the best achievable result by any possible test.

Unhealable	Healable
3.8%	96.2%

As may be seen in the above results, the minimal yield loss that can be achieved for our LNA circuit through exhaustive specification testing is 3.8%. Below, we also calculate the mean and standard deviation of the number of knob settings per device that make a device pass its specifications.

Out of a total of 343 choices, each device has on average 245 passing knob settings. On the other hand, the large standard deviation reflects a high amount of variation among different devices.

C. Exhaustive alternate test

This test evaluates the capability of alternate testing by utilizing all low-cost measurements from every possible knob setting. To predict the performances of the chip through low-cost measurements, we train a regression function for

each dimension of the specification vector (S11,S21,S12,S22). In our case, we trained four MARS regressions, each with the low-cost measurements and the perspective knob settings as predictors. Then, the low-cost measurements are used to predict the performances for each knob setting for each LNA in the test set. Subsequently, among all possible knob settings for which all predicted specifications of the device are satisfied, we use the Mahalanobis distance in order to pick the optimal setting. This optimizes prediction robustness as it seeks to maximize the distance of the predicted device performances from their specifications and, thereby, tolerate prediction error. Finally, we compare the actual performances of every device for the chosen knob setting to their specifications, and we obtain the following confusion matrix.

		Actual	
		Fail	Pass
Predicted	Fail	0%	5.2%
	Pass	3.8%	90%

The results indicate that 90% of the devices are healed in this way, while 5.2% are healable but the method does not select a correct knob and 3.8% are unhealable yet the method still selects a setting which it predicts that it heals the device.

D. Mid-point alternate test

Evolved from the exhaustive test method, in mid-point alternate test we still train MARS regressions with low-cost measurements and perspective knob settings to predict each performance vector. However, in the testing phase, only the low-cost measurement for the center knob setting $(knob_1 =$ 1.2V, $knob_2 = 1.2V$, $knob_3 = 1.2V$) is used to predict the performances for all knob settings of the device. This saves significant time in tuning, since only the mid-point alternate tests are collected, rather than alternate tests for every one of the 343 knob settings. The rationale of this method is that alternate tests for a single knob position are sufficient to reflect the impact of process variations on performances, while the impact of knob variation on performances can be statically learned and is orthogonal to the impact of process variations. The results of mid-point alternate test are reported in the following confusion matrix.

		Actual	
		Fail	Pass
Predicted	Fail	0%	5.5%
Fredicted	Pass	3.8%	89.7%

As can be observed, mid-point alternate test is almost equally effective with respect to yield loss after statistical post-production tuning as the exhaustive alternate test approach, while being two orders of magnitude faster (i.e. only one set of alternate measurements instead of 343 need to be obtained and evaluated through regression models).

Figure 7 contrasts the effectiveness of the aforementioned methods for one device. The Y-axis shows the Mahalanobis distance between the performances of the device for a given knob setting and the optimal choice, while the X-axis lists the 343 knob settings, rank-ordered based on the corresponding Mahalanobis distance. As may be observed, exhaustive specification testing picks the knob setting that maximizes

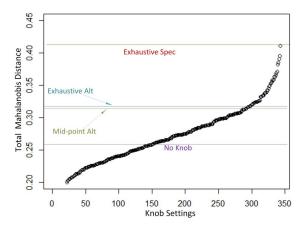


Fig. 7: Comparison of the four methods

that distance, thereby optimizing robustness of the device. Exhaustive alternate test and mid-point alternate test both select a knob setting that is within the top-25% of the available choices. The Mahalanobis distance for these two methods is very close. In contrast, the neutral knob setting represented by the "No Knob" line results in a far worse choice.

V. CONCLUSION

In this paper, we demonstrated a statistical post-production tuning method using actual silicon data. For the purpose of this research, a 1.575 GHz LNA with tuning knobs was fabricated in IBM's 130nm RF/CMOS process. Through extensive characterization, the performances of the device and the sensors were recorded for all possible 343 knob settings. Our experimental results demonstrated that the mid-point alternate test is a very cost-effective way of calibrating an analog/RF device after its production. Specifically, this method operates in one shot, avoiding costly test-&-tune iterations, thereby reducing cost, while achieving similar yield recovery results as the more expensive exhaustive alternate test approach.

REFERENCES

- [1] T. Das and P. R. Mukund, "Self-calibration of gain and output match in LNAs," in *IEEE International Symposium on Circuits and Systems*, 2006, pp. 4–9.
- [2] V. Natarajan, A. Banerjee, S. Sen, S. Devarakond, and A. Chatterjee, "Yield recovery of RF transceiver systems using iterative tuning-driven power conscious performance optimization," *IEEE Design and Test of Computers*, vol. 32, no. 1, pp. 61–69, 2015.
- [3] S. Sun, F. Wang, S. Yaldiz, X. Li, L. Pileggi, A. Natarajan, M. Ferriss, J.-O. Plouchart, B. Sadhu, B. Parker et al., "Indirect performance sensing for on-chip self-healing of analog and RF circuits," *IEEE Transactions* on Circuits and Systems, vol. 61, no. 8, pp. 2243–2252, 2014.
- [4] S. K. Devarakond, V. Natarajan, S. Sen, and A. Chatterjee, "BIST-assisted power aware self healing RF circuits," in *IEEE Mixed-Signals, Sensors, and Systems Test Workshop*, 2009, pp. 1–4.
- [5] M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "One-shot calibration of RF circuits based on non-intrusive sensors," in ACM/IEEE Design Automation Conference, 2014, 2014, pp. 1–6.
- [6] N. Kupp, H. Huang, P. Drineas, and Y. Makris, "Post-production performance calibration in analog/RF devices," in *IEEE International Test* Conference, 2010, pp. 1–10.
- [7] D. Han, B. Kim, and A. Chatterjee, "DSP-driven self-tuning of RF circuits for process-induced performance variability," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 18, no. 2, pp. 305–314, 2010