Alternate Test of RF Front Ends with IP Constraints: Frequency Domain Test Generation and Validation

S. Sermet Akbay*, José L. Torres**, Julie M. Rumer**, Abhijit Chatterjee*, Joel Amtsfield**

*School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
**Intel Corporation, Phoenix, AZ, USA
{sermet,chat@ece.gatech.edu}, {jose.atced.l.torres,julie.m.rumer,joel.amtsfield@intel.com}

Abstract

This paper summarizes an alternate test methodology that enables significant reduction in testing time and tester complexity for RF circuits without the need for low-level simulation models. Traditionally, alternate test makes use of circuit and process-level models to analyze the sensitivity of datasheet specifications to the variations in process parameters. In this paper, we demonstrate a "gray-box" approach by creating a high-level simulation model from datasheet information and simple hardware measurements. This model is used together with a customized behavioral simulator to enable efficient search of an alternate test stimulus that is optimal in terms of tester constraints, test time and specification prediction accuracy. The specific example is a third party RF front-end chip, for which 13 specifications including S-parameters, intermodulation products and noise figures are measured with both conventional and alternate methods. The results are compared in terms of testing time, tester cost and accuracy.

1. Introduction

With the explosion in wireless applications, the last decade witnessed an ever-increasing test challenge for radio-frequency (RF) circuits. While the design community has pushed the design envelope far into the future, the test barriers have not kept pace with the test requirements of high speed, integrated wireless and wired communications designs. Consequently, testing such devices became a major bottleneck in high-volume production further driven by the growing need for tighter quality control.

The benchtop test equipment used throughout prototyping is very precise yet specialized for a subset of functionalities. During the high-volume manufacturing, the projected performance parameters are measured one-by-one for each device and tested against defined limits called specifications. The set of tests required for each product differs greatly in terms of the equipment required and the time taken to test individual devices. Together with signal integrity, precision and repeatability concerns, the initial cost of a RF ATE is prohibitively high. As more functionality and protocols are integrated into a single RF device [1], the required number of specifications to be tested also increases proportionally, adding to the overall cost of testing both in terms of the initial cost and the recurring operating cost.

1.1. Previous Work in RF Test Automation

There is significant published research on test cost reduction for RF devices. Previous methods have used a reduced set of tests [2], and customized ATE for reducing test costs [3]. In [4], a built-in self-test (BIST) methodology is proposed for RF circuits; similarly [5] explores design-for-testability (DfT) techniques. However, the high sensitivity of RF circuit nodes to parasitic components and the area overhead required to incorporate test circuits for each specification have made the implementation of BIST and DfT difficult. Recent BIST proposals [6][7][8] can only handle a limited number of specifications or otherwise incur significant area overhead.

1.2. Previous Work in Alternate Test

In recent years, a growing number of publications on "alternate test" have proposed new solutions for the analog/RF test bottleneck. In the alternate test approach a suite of sequential specification tests is replaced with a *single test*, consisting of a carefully crafted test stimulus applied to the device-under-test (DUT). The response of the DUT to the applied alternate test stimulus can be mapped to all specifications-under-test concurrently [9], thereby allowing significant test time savings. Simultaneously, the cost of the external tester can also be reduced since a simpler test setup can be used to measure all the different test specification values of the DUT. The first implementations of alternate test for low-frequency analog circuits made use of a time-domain oversampling procedure to generate the response samples. However, oversampling poses a great challenge in the RF domain. Subsequently, another variant based on the use of upconversion/downconversion mixers to change the frequency range of the stimulus and response [10] was proposed. In order to address the sampling problem, [11] proposes the use of sensors that convert the obtained RF response to DC-level signatures. A combination of built-in (BIT) and built-off (BOT) alternate

1-4244-0292-1/06/$20.00 © 2006 IEEE

methodologies is explored in [12], especially for SOPs and SoCs, where observability and controllability are the major problems. Use of generic alternate test modules enables limited on-chip hardware to perform the test functionality of many specialized modules. This idea is further explored in [13][14] for BIST implementations and in [15] for on-chip self-calibration.

1.3. Motivation and Key Contributions

Built-in and sensor-based extensions of alternate test make use of DfT features, which is only possible when the test engineer can influence the design process. Similarly most other applications require access to a detailed device netlist and statistical information about the variations of process variables. *However, such information is not always available especially when the device-under-test includes third party intellectual property (IP) blocks. Similarly, the simulation models for complicated RF systems are generally very complex and time consuming for repeated Monte Carlo simulations necessary during the alternate test stimulus generation process.* Although the use of high-level behavioral device models have been proposed in the past to improve simulation speed, the complexity of these models limit the search space [16] or forces computationally cheaper yet sub-optimal greedy search [17].

In this paper, we demonstrate a new variation of the alternate test methodology that makes no use of device netlists or process parameter distributions. We propose a "gray-box" approach suitable for devices with IP blocks by creating a high-level simulation model from datasheet information and simple hardware measurements. This model is used together with a customized behavioral simulator to enable efficient search of an alternate test stimulus that is optimal in terms of tester constraints, test time and specification prediction accuracy. The specific example is a third party RF front-end chip, for which 13 specifications including S-parameters, intermodulation products and noise figures are measured with both conventional and alternate methods.

The main contributions of this paper are:

- A new frequency domain simulator is used for test generation. This uses a simple model created with datasheet information and limited hardware measurements to simulate the responses of a sample set of devices with different specification values; the models are only detailed enough to have accurate results for the alternate stimulus search, hence they do not impose limitations on the initial search domain or force greedy algorithms.
- A flexible automatic test pattern generation (ATPG) structure for which the individual constraints of an ATE can be incorporated into the optimization process as parameters.
- A genetic algorithm (GA) based ATPG method that codes the frequency content of the candidate stimuli in efficient gene formations.
- A comparison of the alternate test and standard specification test in terms of accuracy, testing time and equipment cost both on a benchtop setup and on ATE.

2. Alternate Test with IP Constraints: Theory

In *specification-based alternate tests*, the data sheet specifications of a DUT are predicted by analyzing its response to a specific input pattern. This stimulus is carefully crafted to yield a significant correlation between the response and the specification variations. The DUT response can be considered as a signature of the effects of process variations on that particular DUT instance. Nonlinear statistical multivariate regression analysis [18] allows one to construct mapping functions such that for a given set of measurements, these functions generate predictions for the values of each specification [9].

The key to accurate specification prediction can be summarized in three principles: (i) the alternate test stimulus is selected to maximize the correlation between the response and specifications-under-test, hence requires elaborate ATPG algorithms; (ii) the response signature provides a robust basis for the mapping functions to convert the single signature into many specification values; (iii) the mapping functions are generated by a supervised learning process on a sample set of training devices, for which the specifications-under-test are measured with conventional test methods.

The alternate test generation flow can be studied in three steps: (1) craft an alternate test stimulus; (2) calibrate the mapping functions using hardware measurements on a sample set of devices; (3) apply alternate test in high volume manufacturing (HVM) together with the recalibration of mapping functions when process screening indicates a significant shift from the characteristics of the sample set of devices used in initial calibration (step 2).

2.1. Search for an Alternate Test Stimulus

Figure 1 shows the typical scenario in crafting an alternate test stimulus: In Part I, DUT netlists, semiconductor device models and related process parameters are considered together to create a sample set of DUT models. Monte Carlo techniques are used to generate N instances which reflect the changes under given process parameter variation statistics. Then, these DUT models are simulated with conventional simulators such as SpectreRF and their individual specification values are recorded. These *actual specification values* will form the reference for comparison to alternate test results. In Part II, the alternate test stimulus is determined by a greedy search loop. As in any greedy algorithm, the initial starting point is very

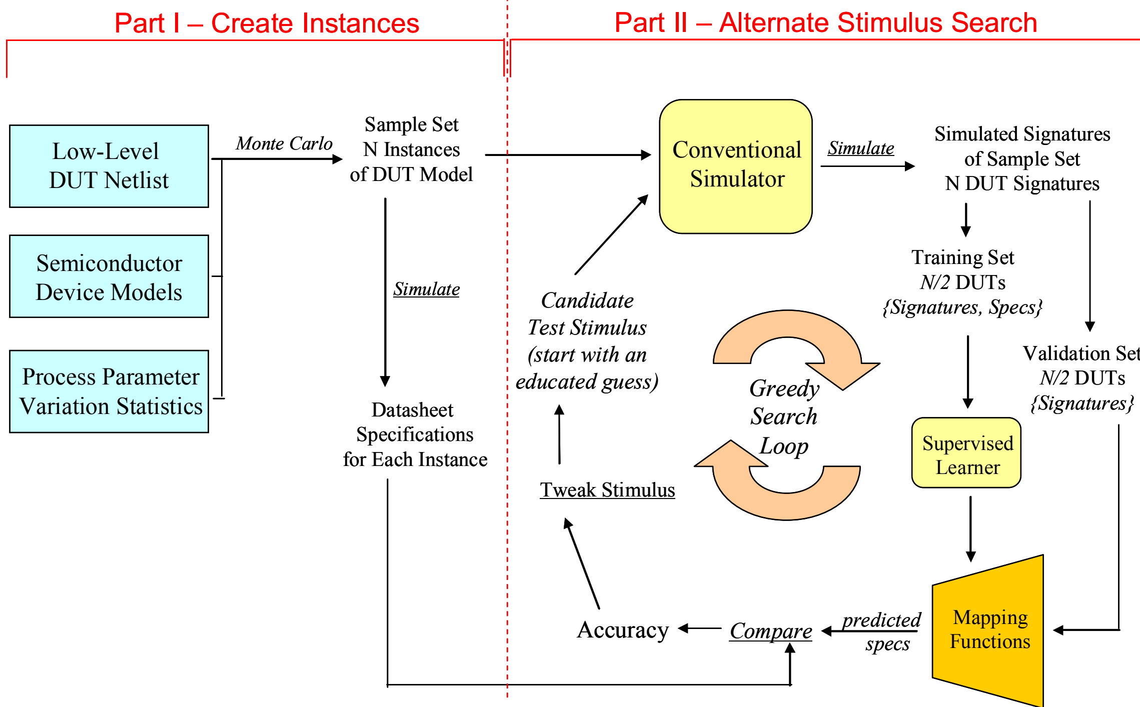

Figure 1: Crafting an alternate test stimulus; Part I: create sample set instances and measure reference/actual specification values, Part II: greedy search for alternate test stimulus.

likely to have an important impact on the final stimulus. In this case, the engineer makes an educated guess for the initial test stimulus. Using this stimulus, the corresponding signatures are obtained through a conventional simulator for each member of the sample set of DUT models which is further divided into a training set and a validation set. Then the response signatures and the actual specification values for the training set are fed into a supervised learning algorithm which creates *mapping functions*. When the response signatures of the validation set are input to these mapping functions, the result is a full set of *predicted specification values* for the validation set. These predicted specifications are compared with the actual specifications obtained in Part I and the result gives a figure of *test accuracy* for the candidate test stimulus. If this accuracy satisfies the termination criteria, then the final test stimulus is found; else, the candidate stimulus is tweaked according to some local rules and the search loop runs again with the new stimulus.

The computational complexity of Part I is not significant since it is processed only once. The computational complexity of the methodology is dominated by the loop in Part II: $O(Q)O(Nn_l)$ where N is the size of the sample set, n_l is the total number of candidate stimuli considered throughout the search loop, and $O(Q)$ represents the computational complexity of one simulation. For a conventional simulator, it is also safe to assume that the total simulation time dominates over the time spent for generating mapping functions.

Complicated RF systems have proven difficult in terms of low-level simulation models. Considering that simulations will be run on a statistically sound set of devices, the computational complexity of the search algorithm becomes prohibitive. This is mainly due to $O(Q)$; the circuit simulation time is cubically proportional to the number of nodes and to the number of voltage and current variables [19]. By moving from transistor level netlists to higher-level behavioral models, the simulation time for a complex RF device is cut roughly by two orders of magnitude [20]. One can further reduce total simulation time by replacing the simulation of N-element sample set with a small number of simulations for sensitivity analysis. In this method [21], the circuit is simulated k times, k being the number of process parameters, and a linear sensitivity matrix is generated from process parameters to response signatures by singular value decomposition.

When IP blocks are present in the DUT, the netlist and statistics of process parameter variations are not available. This situation renders conventional simulators and sensitivity analysis useless. We handle the problem by building a high-level model of the device using a gray-box approach, where a set of N behavioral-level instances are created from hardware measurements on N devices. Consequently, stages in Part I of Figure 1 are replaced with a hardware-based measurement. With no models to apply Monte Carlo analysis, the specifications of N devices are measured one-by-one by using classical test

equipment. When compared to the original alternate test flow, these hardware measurements do not present an overhead in terms of test development time or cost; because the exact same measurements are already required for hardware calibration of the mapping functions (step 2 in alternate test flow). *N* devices are divided into training and validation sets such that each set is close to being a representative sample of the overall specification distributions.

Since the computational complexity heavily depends on the number of instances generated, $O(Q)$ needs to be kept at a minimum. Therefore the alternate test stimulus and the response signature are based on a minimal set of simulation features. Frequency domain features prove more efficient in this case, where only the propagation of a number of tones needs to be simulated.

Although greedy algorithms can cut down the computational complexity by reducing n_l, that is by considering a single candidate stimulus per execution of the optimization loop, their application is limited to certain domains [22]. Preliminary studies and experience shows that the RF stimulus domain presents a neither monotonous nor well behaved topology, former due to saturation and intermodulation effects and latter due to isolated frequency bands from filter effects. As a result, greedy algorithms easily get stuck in local minima. Considering the vast size of the domain, brute force approaches are also not feasible. Genetic algorithms (GA) have been shown to be very powerful for similar applications [23]. GA emulates the natural selection process: a large number of individuals form a population; each individual is evaluated for its fitness and the ones with higher fitness values have a higher probability of bearing children; the less fit individuals are replaced with these children so survival of the fittest strains are guaranteed; on the same while a small amount of mutation rate creates children with new fitness conditions; the population evolves slowly but in a steady pace. GA-based search algorithms cannot guarantee global optimum in finite time, yet they tend to easily move away from local minima. The downside is that GA is computationally more demanding, because n_l becomes a product of n_p, the number of individuals in a population and n_g, the number of generations required to satisfy optimization constraints.

2.2. Mapping Functions using Regression Analysis

The specification values for each device are obtained by mapping the signatures into the specification space. These mapping functions are constructed by a supervised learning process called *multivariate adaptive regression splines (MARS)*. The final functions can be visualized as a weighted sum of basis functions made of splines, which span values for each of the independent variables. The mapping function *f* for a specification *y* with *M* elements in the signature can be represented as:

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m B_{km}(x_{v(k,m)}) \quad \text{eq. (1)}$$

where the summation is over the *M* independent variables and β_0 and β_m are parameters of the function. The basis function *B* is defined as:

$$B_m(x_{v(k,m)}) = \prod_{k=1}^{K_m} H[s_{km} \cdot (x_{v(k,m)} - t_{km})] \quad \text{eq. (2)}$$

where $x_{v(k,m)}$ is the kth independent variable of the mth product, K_m is the number of splits that gave rise to B_m, s_{km} can be ±1 depending on the right or left sense of the step function *H*, and t_{km} are the knot locations for these step functions. MARS uses an initial recursive partitioning during training to gradually add these basis functions using forward stepwise placement; then, a backward procedure is applied and the basis functions associated with the smallest increase in the least squares fit are removed [18].

The quality of a mapping function can be judged by checking the correlation coefficient R^2 between the actual and predicted specification values of a validation set. Since MARS is a supervised learning tool, the final mapping functions will only be as good as the quality of the data used for the training set. In this case, the two components of the training data are (i) the actual specification values coming from hardware measurements, and (ii) the response signatures from the simulator. Looking at the big picture, the accuracy of alternate test acc_P is limited by the repeatability of the actual specification measurements rep_A. In this paper, we use the difference $\Delta AR = acc_P - rep_A$ as a figure of merit for alternate test quality together with R^2.

2.3. The Behavioral Simulation with IP Constraints

In order to make the simulation process as fast as possible, we have written a customized behavioral simulator in Matlab. The simulator is based on propagation of frequency domain tone information through a series of elements representing the devices, transmission lines, equipment interfaces and cables in the actual test system. It can model typical RF module behavior including amplification, generation of intermodulation products, frequency translation, compression, filtering, harmonic distortion and feed-through. Phase noise from the local oscillators (LO) can also be represented by a bunch of spurs around the fundamental LO tone.

The simulator only uses magnitude information of the frequency content. The phase is neglected in a controlled manner, because: (i) it yields faster simulation times by reducing the information to be processed at each step; (ii) it simplifies the modeling process, models can be generated by using a spectrum analyzer and a multi-tone signal generator; (iii) the inaccuracies resulting from addition of two tones with different phases is

compensated by not considering the resultant tone in the final signature, this generates a robust alternate signature; (iv) it can make use of free running local oscillators for up-down conversion.

The proposed simulator is fast enough so that the simulation time for a total of N instances is as short as the time it takes MARS to create corresponding mapping functions. Hence, the complexity of the methodology has to be corrected as $O(Q'+M)O(Nn_g n_p)$, where $O(Q'+M)$ is the total complexity of the proposed simulator and the creation of mapping functions.

During modeling, the test path is divided into a series of elements. Any element with a lookup table is interpolated linearly for missing values and extrapolated with the exact same end value closest to it. The elements are determined in such a way that:

- Loss is represented by a *filter-element* made up of a lookup table for frequency-loss value pairs.
- Any input (LO and other RF signal sources including stimuli) is represented by an *input-element*, a group of tones corresponding to the harmonics and their phase noise tails.
- Frequency translators are represented by a *mixer-element* made up of a normalized matrix and correction tables. The normalized matrix represents the output corresponding to a single unit amplitude tone as a result of a predefined number of –defined by the model- LO and input harmonics varied by input frequency. Hence, the normalized matrix generates LO±IN, LO±2IN, LO±3IN, … 2LO±IN, 2LO±2IN, …, LO feedthrough (LO, 2LO, …), and input feedthrough (IN, 2IN, …). Once these tones are generated, three correction lookup tables scale the output with respect to LO frequency, LO power and input power.
- Amplifiers are represented by *amplifier-elements* made up of three lookup tables: frequency-gain value pairs, gain compression, and intermodulation (IMD). Intermodulation products are considered up to an order defined by the model (usually 3rd).

The frequency domain is quantized in terms of both frequency and amplitude, the minimum quantization step is determined by a simulation parameter. If two tones with the same frequency is generated by an amplifier-element or a mixer-element, than these tones superpose and their powers add up as if they are in phase. If one tone is significantly larger than the other –a limit defined by a simulation parameter- the error coming from the phase difference will not be significant. If the two tones are comparable in magnitude then the resultant sum is marked by a *"dirty bit"*, in this case the dirty bit propagates into latter elements and marks related tones and IMD products as dirty. At the very end of the chain, only the tones with clean bits are considered for the alternate signature.

Cables, transmission lines, socket interfaces, and all kinds of filters are directly converted into filter-elements. A passive mixer is converted into a mixer-element followed by a filter-element. An active mixer is a mixer-element followed by an amplifier-element. The modeling process is simply composed of dividing the test path into a logical chain of these elements and deriving the corresponding matrices and lookup tables using a spectrum analyzer and two signal generators –one for input and one for LO-. Then, each component is defined by a software structure in Matlab containing the appropriate tables and matrices. The elements are connected in the form of a graph notation. The coding of the simulator is performed such that it makes use of the efficient matrix operations in Matlab by combining many instances of the model into a single matrix and computing responses simultaneously.

2.4. The Gene Notation and Fitness Function

The quality of the GA implementation depends mainly on two factors: (i) a concise gene notation which transforms the information each individual possesses into a code for exchanging that information without breaking its meaning; (ii) a fitness function which can identify the better individuals from the lesser without losing diversity by over-penalizing others.

For this study, we have experimented with a number of fitness functions which are based on different definitions of test accuracy. Although different functions yielded similar optimized stimuli at the end, the one based on root-mean-square (rms) error with 0.05 significance interval required less number of generations to produce the same level of quality.

The gene notation is also unique in the sense that it speeds up the convergence process. Since the candidate stimuli are essentially multi-tone signals, each individual is represented by a D-bit gene sequence. Each b consecutive bits, called as gene-bytes, represents a quantized amplitude value for equally-spaced frequency components with spacing Δf_{min}. The simulator interprets any amplitude below A_{border} as a no-tone, which is an infinitely small tone not represented in an input-element. The whole input frequency domain is divided into K overlapping regions, limited by the bandwidth available on the signal generator and defined by the center frequency component.

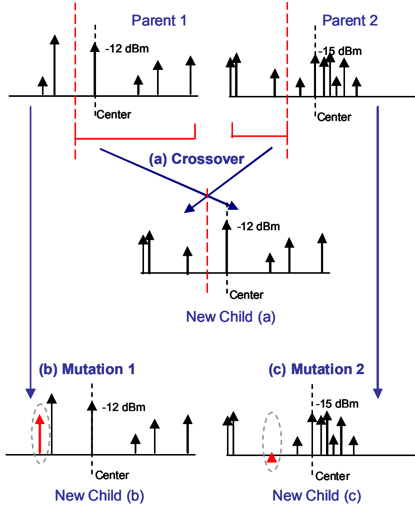

Figure 2: Crossover (a) and mutation (b, c) in GA with the proposed gene notation.

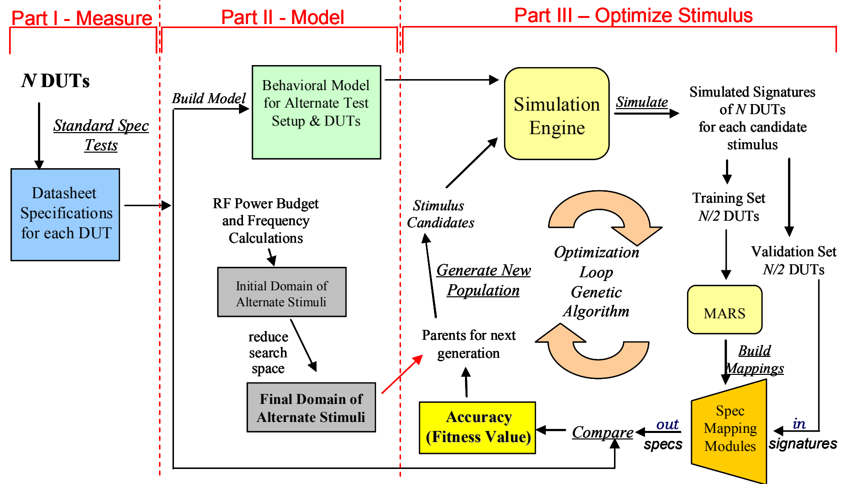

Figure 3: Proposed alternate test methodology.

Figure 2 shows the mechanisms for generating new children. In crossover (denoted with *a*), a random gene-byte location is selected (red dotted line); the first parent's bits from the right of the selected location are combined with the second parent's bits from the left of the location. Each region designated by the center frequency component can only mate parents within; this way the newborn (new child *a*) does not violate the bandwidth limit. The other main mechanism, mutation, makes a random amplitude difference in one of the quantized frequency locations in such a way that a tone already smaller than A_{border} becomes larger (new child *b*) or vice versa (new child *c*). The GA implementation also lets a small percentage of *elites*, individuals with best fit scores, propagate to the next generation without any modification. The use of the elites guarantees that best fitness value will not get worse from one generation to the next.

We also implemented a migration feature in the GA. Every n_{mg} generations, copies of evolved individuals from different regions are subjected to circular migration by changing their center frequency with that of the region being migrated to. This way a frequency-amplitude pattern that thrives in one region has a chance to mate with another leading pattern in another region.

The reader can refer to [23] for the theory behind these features.

Figure 3 shows the proposed approach for alternate test. When compared to Figure 1: in Part I, the actual specifications are measured on hardware as opposed to performing Monte Carlo simulations; then there is an extra modeling step shown by Part II; finally in Part III, the GA-based search replaces the greedy approach.

3. Experimental Results

The proposed methodology is validated on a RF front-end chip from RF Micro Devices. RF2411 [24] is a dual-band low-noise amplifier (LNA) followed by a mixer with pins for access to both components individually. For the sake of simplicity, only one band (850MHz) is considered for 13 specifications: LNA gain, input standing wave ratio (ISWR), output standing wave ratio (OSWR), reverse isolation, input third order intercept point (IIP3), noise figure (NF); mixer gain, ISWR, IIP3, NF; and cascade (LNA + Mixer) gain, IIP3, and NF.

We have used a sample set of 541 ICs. Although different batches of ICs are bought over a time interval of two months, the resulting set only partially represents the lot-to-lot variation and all parts pass RFMD's high volume manufacturing screening.

The experiments can be studied in three steps: (1) standard specification tests, these benchtop tests are performed to evaluate the specifications of each IC with the conventional methods and obtain equipment cost, testing time and repeatability measures for the standard tests; (2) benchtop alternate tests, the standard specification test is replaced with an alternate counterpart and the results are compared; (3) alternate test on ATE, the benchtop alternate test is migrated onto a commercial ATE platform for a feasibility study.

In order to demonstrate the equipment cost reduction resulting from alternate test, the target ATE is selected as a mixed-signal open architecture [25] tester instead of a RF tester. The tester can create 12-bit arbitrary waveforms (AWG) at 800Msps yet limited to 200MHz bandwidth due to clock bleeding. It can also coherently sample at 800Msps with 12-bit accuracy.

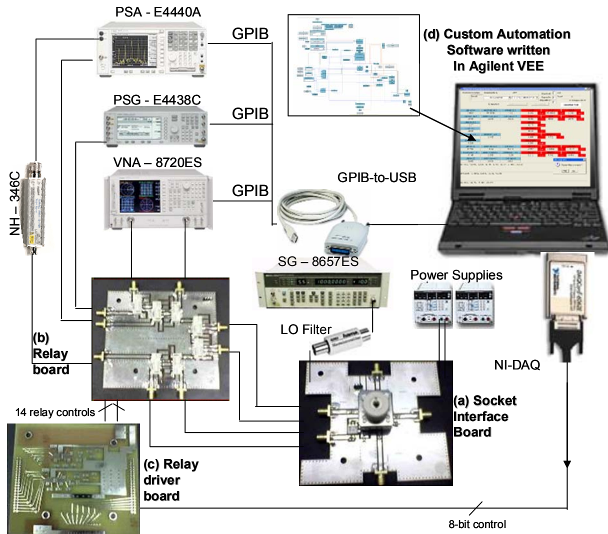

Figure 4: Standard specification test setup.

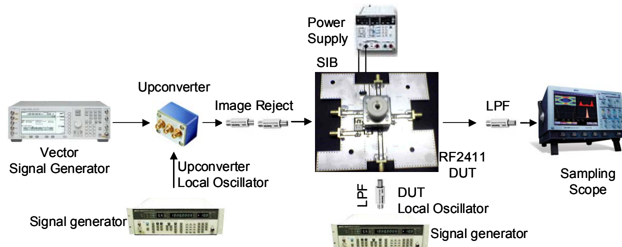

Figure 5: Benchtop alternate test setup.

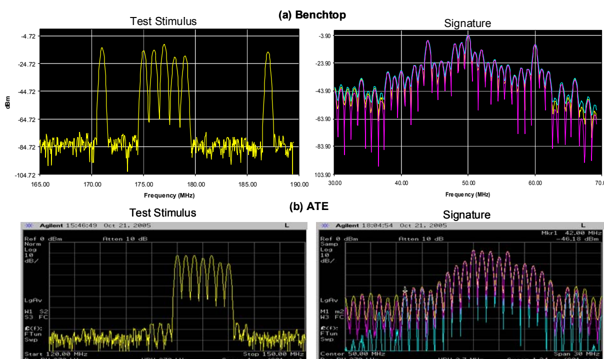

Figure 6: Alternate test stimulus and response for (a) benchtop setup, (b) ATE setup.

TABLE I: Alternate test stimulus tone locations and power levels

Bench Freq. MHz	171	175	176	177	178	179	187
Bench Pwr. dBm	-13.5	-15.5	-15	-10.5	-20	-19.5	-16.5
ATE Freq. MHz	135	136	137	138	139	140	141
ATE Power dBm	-12	-12	-12	-14.5	-12	-18.5	-19.5

3.1. Standard Specification Tests

Testing the listed thirteen specifications of a single IC requires seven test configurations involving different equipment and pin connections. In order to minimize the repeatability problem, we designed a load board that can automatically cycle through these configurations with a single IC insertion. The various test equipments connect to the load board and routed to the proper I/O of the chip through relays. Figure 4 shows this setup: the IC is interfaced through a socket on the *socket interface board (SIB)*, which has five RF connectors for LNA input, LNA output, mixer input, mixer output and LO input; the test equipment consists of a spectrum analyzer, a network analyzer, two signal generators, a noise head, and a power supply; all the equipment connect to the *relay board (RB)*, and routed to the proper input and output ports on the SIB; the *relay driver board (RDB)* –Figure 4 shows the unpopulated stage- hosts relay controller chips, for which the digital control is interfaced to the laptop using a NI-DAQ card; the whole process is automated by a custom written Agilent VEE program, which controls the test equipment through GPIB.

The partitioning of the load board into three parts provides major benefits: (i) the SIB can be reused in alternate test setups, hence standard and alternate test results can be compared directly; (ii) the separate RDB isolates the sensitive RF paths on RB from the digital control. The three different boards are vertically connected with cables and encapsulated into a Faraday cage, which is necessary for noise figure measurements.

The accuracy of the proposed standard specification tests are also measured by a number of repeatability experiments. These experiments are designed in such a way to characterize the contributions of relay switching, equipment drift and socketing. The socketing experiments, which dominates the overall repeatability error, includes periodical re-testing of a comparison IC throughout the testing of 541 ICs and instant re-socketing of specific ICs in each batch of 25.

3.2. Benchtop Alternate Test Experiment

The experiment starts with defining the constraints for the alternate stimulus optimization. These constraints also depend on the capabilities of the ATE, because the end goal is to develop a benchtop alternate test which can be migrated. Since the AWG is limited to 800Msps with 200MHz BW, the at-speed test of the 850MHz DUT requires an up/down conversion scheme [10]. Figure 5 shows this setup: The center frequency of the optimized input stimulus is upconverted to 850MHz with a passive mixer driven by a free running LO. The upconversion mixer and the following image reject filter is characterized and embedded into the behavioral model. The response is downconverted to 50MHz with the mixer in the DUT driven by another free running LO. The

response is passed through a low-pass filter and logged by a sampling scope. After a 1024-point FFT, the magnitudes of 12 robust tones constitute the signature.

The top part of Figure 6 shows the optimized stimulus and a sample response signature for the benchtop alternate test. The stimulus is made up of seven tones around 177MHz ranging from -10 to -20dBm in amplitude as given in the top part of Table I. This multitone signal is implemented on a vector signal generator using Agilent signal studio software [26]. *All 13 specifications are predicted by this single stimulus.*

3.3. Alternate Test on ATE

The ATE alternate test setup is similar to Figure 5, except that the vector signal generator and the sampling scope are replaced with the arbitrary waveform generator and the digitizer on the ATE. Since it is a mixed-signal tester, the external LO sources are still required for up/down conversion. The integration of these LO sources is simple due to the fact that response signatures are composed only of FFT magnitudes, hence the LO sources are free running and require no synchronization [10].

Originally, the alternate test stimulus optimized for benchtop was to be used with the ATE setup. However, initial runs showed the AWG has an internal clock-bleed filter, which generates a large roll-off for tones close to 200MHz. Also differential to single ended conversion imposed an additional limit on the maximum amplitude of any tone. As a result, the filter was added to the behavioral model and we re-optimized the stimulus with the new constraints. Bottom part of Figure 6 shows the optimized stimulus and a sample response signature. The stimulus is made up of seven tones around 138MHz ranging from -12 to -20dBm in amplitude as listed in the bottom part of Table 1. As one can see, the additional constraints and the filter model have a significant effect on the optimized tones, which are further away from the 200MHz roll-off and smaller in amplitude when compared to the older stimulus.

3.4. Experiment Results and Comparison

The results from standard and alternate experiments show that alternate testing time is an order of magnitude smaller than the standard method on benchtop. This is mainly due to the sequential nature of standard tests, for which the testing time almost linearly scales with the number of specifications. The setup times are dominated by the GPIB-USB communications and add up to ~50 seconds for 13 specifications; whereas in alternate test, there is only a single setup. Also, the test equipment for alternate test requires a signal source, a sampling scope and two LO sources, which overall costs ~$144K, 60% less than standard test equipment used(~$340K).

These benefits do not directly scale up for the ATE setup. For one reason, standard ATE setup times are far more optimized; for another, the system cost of an ATE is already high regardless of its functionalities. In order to make a healthy comparison, we used figures from a low-end commercial RF tester. This particular class of RF tester is comparable to the mixed-signal tester we have used in terms of settling time yet still has the minimal functionality to perform the required standard tests. We estimated the total standard specification testing time by multiplying the average per-specification time with the number of specifications. For 13 specifications, the proposed alternate test stimulus provided a significant 36% reduction in testing time. Furthermore, the mixed-signal tester together with two LO sources costs 48% less than the RF tester.

These cost benefits are significant only if alternate test is as accurate as the original specification test. In order to compare the accuracy, we use two figures: R^2 or the variance explained, as commonly used in statistics, and ΔAR as described in Section 2.2.

Figure 7 shows the ISO graphs for four specifications: LNA NF, mixer gain, cascade IIP3 and LNA OSWR. The x-axis shows the original specifications measured by standard tests while y-axis shows the specification values predicted by alternate test on the benchtop setup. The blue lines show the $\pm 3\sigma$ repeatability error for standard tests. The 7th and 8th rows of Table II list R^2 values for all specifications, for which LNA OSWR is the smallest with 74%. On the other hand, the ISO graph for LNA OSWR shows that all prediction errors are still smaller than the 3σ repeatability error shown in the 1st row of the table. As a matter of fact, out of 270 devices in the validation set, only 3 devices in benchtop and 4 devices in ATE fall out of the 3σ limits for any specification. The 2nd and 3rd rows in the table list the maximum prediction errors for each specification whereas 5th and 6th rows show the rms prediction errors. All rms prediction errors are smaller than the rms repeatability errors listed in the 4th row. Hence, ΔAR is very close to zero in terms of maximum errors and always negative in terms of rms errors. These results, together with high R^2 values, show that the implemented alternate test is as accurate as the original specification tests.

4. Limitations and Discussion

For very complex systems such as transceivers, the speed of the proposed simulator can still pose a bottleneck. However, it is possible to further improve the speed. If one considers the propagation of tones from the stimulus to the response, new tones are created through amplifiers and mixers while existing tones are crippled by filters and samplers. Hence, not all the tones generated by an earlier element in the chain show up in the response. In other

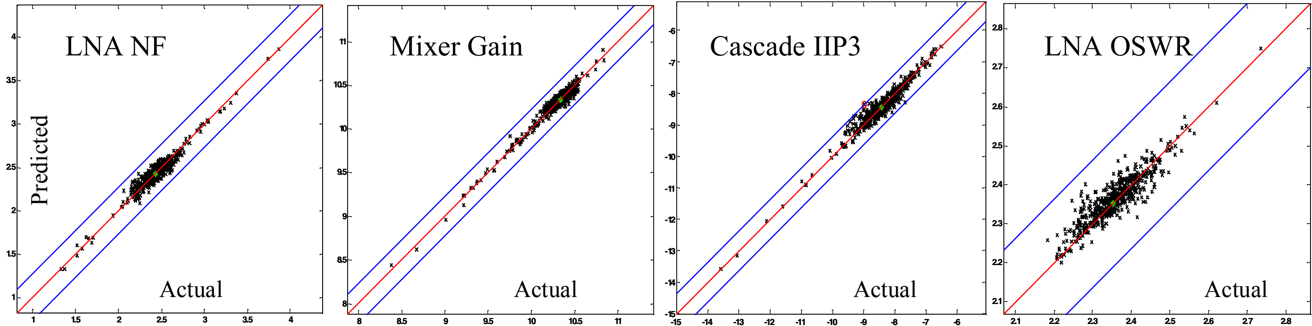

Figure 7: Selected actual versus predicted specification values for the validation set of ICs, x-axis corresponds to the standard specification test measurements while y-axis shows corresponding benchtop alternate test predictions. From left to right: LNA noise figure, mixer gain, cascade IIP3, LNA OSWR.

TABLE II:
Comparison of standard specification repeatability errors with alternate test prediction errors; (1) 3σ maximum repeatability error for standard, (2) maximum prediction error for alternate benchtop, (3) maximum prediction error for alternate ATE, (4) rms repeatability error for standard, (5) rms prediction error for alternate benchtop, (6) rms prediction error for alternate ATE, (7) R^2 correlation metric for alternate benchtop, (8) R^2 correlation metric for alternate ATE.

		Low-Noise Amplifier (LNA)						Mixer				Cascade (LNA + Mixer)		
#	Error	NF	Gain	IIP3	ISWR	OSWR	RevIso	NF	Gain	IIP3	ISWR	NF	Gain	IIP3
1	Standard Spec 3σ	0.27	0.29	0.62	0.11	0.16	1.94	0.33	0.23	0.82	0.60	0.48	0.42	0.62
2	Alternate Bench Max	0.20	0.27	0.54	0.13	0.12	1.27	0.28	0.17	0.82	0.45	0.40	0.33	0.64
3	Alternate ATE Max	0.35	0.26	0.54	0.07	0.12	1.37	0.23	0.24	0.66	0.35	0.54	0.69	0.99
4	Standard Spec Std	0.14	0.18	0.39	0.08	0.12	1.36	0.18	0.07	0.28	0.41	0.31	0.26	0.40
5	Alternate Bench Std	0.06	0.07	0.14	0.02	0.03	0.33	0.07	0.05	0.20	0.12	0.09	0.10	0.17
6	Alternate ATE Std	0.07	0.06	0.14	0.02	0.04	0.33	0.06	0.06	0.17	0.10	0.09	0.14	0.22
7	Alternate Bench R^2	94%	88%	93%	88%	85%	85%	79%	97%	97%	85%	86%	95%	94%
8	Alternate ATE R^2	94%	87%	94%	88%	74%	83%	85%	93%	95%	92%	86%	89%	92%

words, the proposed simulator works with a "forward-only" data flow model. An additional feature can be added by back-propagating the filtering constraints from the last element to the first in the chain. At the beginning of the process the system will be studied only once, and valid frequency bands will be assigned for the outputs of each element.

Also for complex systems such as transceivers, the specifications-under-test are usually end-to-end meaning that individual specifications of a LNA or a mixer are not critical. From this point of view, the proposed scheme presents additional benefits for diagnosis in the sense that the proposed end-to-end stimulus/response pair (stimulus applied to LNA input, response collected at mixer output) is used to predict individual specifications of the LNA and mixer as well as the cascade specifications. Note that for the diagnosis scenario to be effective, the complex system needs to have taps for corresponding component input/outputs so that relevant component specifications can be measured for training of the supervised learner.

It is hard to define a figure of merit for the accuracy of alternate test. In Section 3, we have used two figures: ΔAR and R^2. In HVM, the real figure of merit is in defects-per-million (DPM) together with some figure of yield loss resulting from guardbands. In our case study, the DUT was a 3rd party product, hence all the samples were well within guardbands induced by the manufacturer, in other words they were "perfectly passing" samples. As a result, the device sample set is not really representative of the manufacturing line, and DPM or yield cannot be measured without imposing artificial specification boundaries and guardbands. Since any boundary introduced in an artificial manner should be taken with a pinch of salt, we preferred using the rather generic metrics defined in Section 3. On the other hand, these two metrics are over-pessimistic in the sense that they are designed to work for any given pass-fail boundary within the specification range. For example, with the proposed metrics, all specification predictions outside the repeatability error band degrade the accuracy measure for alternate test; however, as long as they are far away from the guard-band, most of these predictions would still be accurate enough when evaluated with a single pass-fail boundary. A similar penalty would not be assigned in the DPM count.

Since there is no defined pass-fail boundary, the mapping functions are generated by the regression tool called MARS. For a system with hard pass-fail decisions, one can use classifiers [27] instead of regression tools with greater accuracy in terms of DPM. Support vector machines (SVM) and artificial neural networks (ANN) [28] are also considered for similar applications. Comparison of these different approaches for test applications is not yet available in literature.

5. Conclusions

Alternate test provides significant testing time savings and can be completed with simpler testers. These advantages reduce test costs and also increase the life time of available ATE. Although originally tailored for test operations with full access to netlist and process data, the proposed modification can be applied to IP blocks using a gray-box approach. This method results in similar test accuracy for the cost of test development time and the low-volume equipment used in calibration of the mapping functions.

6. Acknowledgement

The authors would like to acknowledge David Kruppa, Darrin Zinda, Chris Nelson, Bob Miller, Tim Swettlen, Jin Pan, Kevin Redmond and Marc Berube from Intel Corporation.

7. References

[1] "Radio Frequency and Analog/Mixed-Signal Technologies for Wireless Communications," *ITRS*, 2005.

[2] K.M. Mackay, "Testing Wireless Local Area Network ICs at 5GHz," *Proc. Int. Test Conference*, 2002, pp. 1146-1150.

[3] M. Slamani, J. Ferrario, R. Wolf, H. Ding, "Moving from Mixed Signal to RF Test Hardware Development," *Proc. Int. Test Conference*, 2001, pp. 948-956.

[4] M. Hafed, N. Abaskaroun, G.W. Roberts, "A Stand-Alone Test Core for Time and Frequency Domain Measurements," *Proc. Int. Test Conference*, 2000, pp 1031-1040.

[5] S. Ozev, C. Gaard, A. Orailoglu, "Testability Implications for Low-Cost Integrated Radio Transceivers: A Bluetooth Case Study," *Proc. Int. Test Conference*, 2001, pp. 965-974.

[6] M.M. Hafed, N. Abaskharoun, G.W. Roberts, "A 4-GHz effective sample rate integrated test core for analog and mixed-signal circuits," *IEEE Trans. Solid-State Circuits*, vol. 37, pp. 499–514, Apr. 2002.

[7] A.P. Jose, K.A. Jenkins, S.K. Reynolds, "On-chip spectrum analyzer for analog built-in self test," *Proc. VLSI Test Sym.,* 2005, pp. 131- 136.

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

[9] P.N. Variyam, S. Cherubal, A. Chatterjee, "Prediction of Analog Performance Parameters Using Fast Transient Testing," *IEEE Trans. Computer-Aided Design of Integrated Circuits*, Vol. 21, No. 3, March 2002, pp. 349-361.

[10] R. Voorakaranam, S. Cherubal, A. Chatterjee, "A Signature Test Framework for Rapid Production Test of RF Circuits," *Proc. Design, Automation and Test in Europe,* 2002, pp. 186-191.

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

[12] S.S. Akbay, A. Halder, A. Chatterjee, D. Keezer, "Low Cost Test of Embedded RF/Analog/Mixed-Signal Circuits in SOPs," *IEEE Trans. on Advanced Packaging*, vol. 27, issue 2, pp. 352-363, May 2004.

[13] S.S. Akbay, A. Chatterjee, "Feature extraction based built-in alternate test of RF components using a noise reference," *Proc. VLSI Test Sym.*, 2004, pp. 273- 278.

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

[15] D. Han, S.S. Akbay, S. Bhattacharya, A. Chatterjee, W.R. Eisenstadt, "On-chip self-calibration of RF circuits using specification-driven built-in self test (S-BIST)," *Proc. Int. On-Line Testing Sym.,* 2005, pp. 106- 111.

[16] S. Bhattacharya, G. Srinivasan, S. Cherubal, A. Halder, A. Chatterjee, "System-level testing of RF transmitter specifications using optimized periodic bitstreams," *Proc. VLSI Test Sym.,* 2004, pp. 229- 234.

[17] A. Halder, A. Chatterjee, "Automated test generation and test point selection for specification test of analog circuits," *Proc. Int. Sym. Quality Electronic Design,* 2004, pp. 401- 406.

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

[19] J.E. Bracken, V. Raghavan, and R.A. Rohrer, "Extension of the asymptotic waveform evaluation technique with the method of characteristics," *Tech. Dig. ZEEE Int. Conf. Computer-Aided Design*, 1992, pp. 71-75.

[20] J. Wood, X. Qin, A. Cognata, "Nonlinear microwave/RF system design and simulation using Agilent ADS 'system - data models'," *Proc. Int. Workshop Behavioral Modeling and Simulation,* 2002, pp. 75- 79.

[21] R. Voorakaranam, A. Chatterjee, "Test generation for accurate prediction of analog specifications," *Proc. VLSI Test Sym.*, 2000, pp. 137 – 142.

[22] Vafaie A., Imam I., "Feature Selection Methods: Genetic Algorithms vs. Greedy-like Search," *Proc. Int. Conf. Fuzzy and Intelligent Control Systems*, 1994.

[23] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989.

[24] RF Micro Devices, "Dual-Band Front-End RF2411, LNA & Mixer Datasheet," http://www.rfmd.com.

[25] Semiconductor Test Consortium, *Openstar ATE System Platform Specifications*, November 2004.

[26] Agilent, *Signal Studio for Enhanced Multitone Manual*, http://www.agilent.com.

[27] J.H. Friedman, "An Overview of Computational Learning and Function Approximation", *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, Springer-Verlag, 1994.

[28] H.G.D. Stratigopoulos, Y. Makris, "Constructive derivation of analog specification test criteria," *Proc. VLSI Test Sym.*, 2005, pp. 395- 400.