Advances in Variation-Aware Modeling, Verification, and Testing of Analog ICs

Dimitri De Jonghe∗, Elie Maricau∗, Georges Gielen∗, Trent McConaghy†, Bratislav Tasić‡, and Haralampos Stratigopoulos.§

∗K.U. Leuven, Heverlee, Belgium. georges.gielen@esat.kuleuven.be
†Solido Design Automation Inc., Canada. trentmc@solidodesign.com
‡NXP Semiconductor Inc., Eindhoven, The Netherlands. bratislav.tasic@nxp.com
§TIMA Laboratory, Grenoble, France. haralampos.stratigopoulos@imag.fr

Abstract—This paper describes novel scalable, nonlinear, and industrially-oriented approaches to perform variation-aware modeling, verification, fault simulation, and testing of analog/custom ICs. In the first section, Dimitri De Jonghe, Elie Maricau, and Georges Gielen present a new advance in extracting highly nonlinear, variation-aware behavioral models, through the use of data mining and a re-framing of the model-order reduction problem. In the next section, Trent McConaghy describes new statistical machine learning techniques that enable new classes of industrial EDA tools, which in turn are enabling designers to perform fast and accurate PVT / statistical / high-sigma design and verification. In the third section, Bratislav Tasić presents a novel industrially-oriented approach to analog fault simulation that also has applicability to variation-aware design. In the final section, Haralampos Stratigopoulos presents a new advance in extracting highly nonlinear, variation-aware behavioral models, through the use of data mining and a re-framing of the model-order reduction problem. In the next section, Trent McConaghy describes new statistical machine learning techniques that enable new classes of industrial EDA tools, which in turn are enabling designers to perform fast and accurate PVT / statistical / high-sigma design and verification. In the third section, Bratislav Tasić presents a novel industrially-oriented approach to analog fault simulation that also has applicability to variation-aware design. In the final section, Haralampos Stratigopoulos presents a new advance in extracting highly nonlinear, variation-aware behavioral models, through the use of data mining and a re-framing of the model-order reduction problem.

I. VARIATION-AWARE BEHAVIORAL MODELING OF ANALOG CIRCUITS

A. Motivation

We present a new technique for accurate, reliable, behavioral model extraction for large analog circuits under process variations. From a designer’s perspective, the ideal tool does one-click macromodel generation from a given circuit netlist, with sufficient speed and accuracy, and without having to go through procedural difficulties like model selection and training. Today’s analog and mixed-signal design flows however have a high degree of customization, and increasingly have variation / reliability issues when pushing the boundaries of technology scaling. EDA developers therefore face a challenge when trying to meet all of these requirements at once.

Nonlinear Model Order Reduction (MOR) techniques that extract the necessary data from the internal circuit description are becoming a standard for accurate and efficient model extraction [1]. In MOR techniques, data mining is applied on the Modified Nodal Analysis (MNA) matrices while the circuit is simulated in transient analysis [2]. Before the matrix samples are stored in a database, useful information is reduced to a minimum by matrix projection. The resulting model is a collection of low-rank / low-order local expansion points of the original system. During evaluation, the model points are retrieved from the database and stitched together by interpolation [3].

A recently developed approach, called Transfer Function Trajectories (TFT), applies strongly nonlinear regression techniques on the extracted MNA matrices by transforming them to a mixed state space/frequency description [4], [5]. The resulting models contain a relatively compact set of equations, which are more easily ported to any general purpose simulator than a database interpolant. These models prove to scale well, mimic SPICE behavior with a high degree of accuracy, and can be extracted automatically with a minimum amount of model training. Before these models can be deployed trustworthy in an industrial environment, process variations must be taken into account. In this paper, we exploit state-of-the-art machine learning techniques that fit the purpose, including advanced pathwise regression [6].

B. Variation Aware-Transfer Function Trajectories

1) MNA Based Modeling: A general expression for MNA-modelled circuits is given by the nonlinear branch equation:

$$\frac{d}{dt}q(v) + i(v) = Bi_{in}$$  

(1)

which is the typical representation of a system using a SPICE-based simulator. \(v = v(t) \in \mathbb{R}^N\) is the vector of unknown node voltages and inductor currents in the circuit and \(i_{in} = i_{in}(t) \in \mathbb{R}^M\) are the external inputs to the circuit. \(q(.)\) and \(i(.)\) are matrix-valued functions describing the charges and currents of nonlinear components. \(B \in \mathbb{R}^{N \times M}\) is a constant incidence matrix, which maps the inputs to the internal nodes of the circuit.

A general trajectory sampling method stores the linearized conductance and capacitance matrices that is returned by the Newton-Rhapson algorithm of the simulator during transient simulation at each time point \(k\) [1], [2]:

$$G^{(k)} = \nabla i(v)|_{v_k}, \ C^{(k)} = \nabla q(v)|_{v_k}$$  

(2)

After performing projection-based MOR, the reduced matrices can be used as a low-order expansion point in a piecewise
model [2]. Alternatively, a compact nonlinear model can be found by transforming the MNA samples into the frequency domain:

\[ H(k)(s) = \frac{V^{(k)}(s)}{I_{in}^{(k)}(s)} = \left( G^{(k)} + s \cdot C^{(k)} \right)^{-1} B \]  

(3)

Fig. 1. Transformation of the MNA trajectory samples into the frequency domain.

The data transformation is depicted more in detail in Fig. 1. The Transfer Function Trajectory approach models the state-dependent transfer functions as a hyperplane \( T(x^{(k)}, s) \) in a mixed state space/frequency domain [5]:

\[ H(k)(s) \approx T(x^{(k)}, s) = \sum_{p=1}^{P} \tilde{v}_p(x^{(k)}) \]  

(4)

Here, \( x^{(k)} = x[k] \) is a state estimator that links each state \( k \) to \( l_1 \) delayed inputs (FIR), \( l_2 \) delayed autoregressive outputs (AR) and PVT parameters such as process variations \( \sigma_p \), mismatch \( \sigma_m \) and temperature \( T \):

\[ x[t] = (i_{in}[t], ..., i_{in}[t-l_1], v[t-l_1], ..., v[t-l_2], \sigma_p, \sigma_m, T) \]  

(5)

The Vector Fitting algorithm is used to extract a common set of \( P \ll N \) stable poles \( a_P \) such that the system approximation (4) can be implemented with a single or parallel Hammerstein structure [5], [7]. The nonlinear Hammerstein blocks are derived from the indefinite integral of the residue data \( \{\tilde{v}_p(x[k])\} \). A suitable high-dimensional nonlinear regressor for these residue samples is explored in the next section.

2) Multivariate Nonlinear Residue Regression: The computational complexity of the nonlinear Hammerstein functions is determined both by the number of deterministic input parameters and stochastic process parameters. The dimensionality of the problem equals \( l_1 + l_2 + K + 1 \), where \( K \) equals the amount of explanatory PVT parameters. Moreover, strongly nonlinear dynamic behavior is expected for large signal input waveforms. We consider the following multidimensional regression approaches: multivariate adaptive regression splines (MARS) [8], least-squares support vector machines (LS-SVM) [9] and a recently developed deterministic symbolic regression technique, fast function extraction (FFX) [6]. Interpolation algorithms are not considered due to their poor extrapolation performance.

The model generation flow is shown in Fig. 2. Firstly, a set of PVT samples is generated with for example Monte-Carlo sampling. This can of course be replaced by a more clever active learning sampling strategy. The data is further fitted with Vector Fitting and Residue Fitting and finally translated to a Hammerstein structure.

3) Experimental Results: The above variation-aware modeling technique has been verified on the high-speed output buffer example that is described in detail in [5]. The average normalized error (NMSE) for FFX, MARS and LS-SVM is 1.3%, 2.7% and 2.5% respectively.

II. INDUSTRIAL VARIATION-AWARE DESIGN AND VERIFICATION OF CUSTOM ICs

As Moore’s Law progresses [10] and variation gets worse, the traditional approaches are becoming inadequate. We now describe the issues, and how modern variation-aware flows and tools can help manage PVT and statistical variation.

A. Fast PVT Design and Verification

Process, voltage, and temperature (PVT) variations are often modeled as a set of PVT corners. Traditionally, only a few were needed: \{FF, SS\} \times \{min V, max V\} \times \{min T, max T\} = 8. But modern processes have more modelsets, and tighter margins mean that more intermediate values must be considered. Consider the reference VCO from the TSMC AMS Ref. Flow 2.0 [11] on TSMC 28nm. A reasonable setup has 15 modelset values, 3 values for temperature, and 5 values for each of its three voltage variables, totalling 3375 corners. HSPICE \( T^M \) takes 70 s to simulate this, therefore it takes 66 hours to evaluate all corners.

Designers may cope by guessing which corners cause the worst-case performance, but a wrong guess could mean failure in testing (leading to a re-spin), or failure in the field. One
alternative is linear sensitivity analysis, but it will miss the worst-case corners whenever the response is not linear enough. Quadratic modeling is a bit more general, but will fail if the response is not quadratic enough.

A new alternative is FastPVT [12], which is both fast and accurate. The idea is to cast the PVT problem as a global optimization problem: minimize performance in the space of PVT variation; then to solve it reliably using adaptive machine learning with arbitrarily nonlinear models [13] and SPICE in the loop. On the VCO listed above, FastPVT found the worst-case PVT corner, with confidence, in 371 simulations (versus 3375, for a 9.1x speedup).

We have benchmarked FastPVT on a suite of 108 representative industrial problems. Figure 4 left shows the distribution of speedups. The average speedup was 11.3x, and the maximum speedup was 43.1x. For the 56 problems with >200 corners, the average speedup was 19.3x. A few problems had speedup of just 1.0x, which simply indicated that there was so little structure in the mapping from PVT variables to outputs that adaptive modeling could not help.

FastPVT is part of the rapid-iteration design flow shown in Figure 5 left. The idea is to first extract one or a few design-specific PVT corners; then to design against them from SPICE; and finally to verify. Design iterations are fast because there are so few corners. If verification succeeds, then the designer can go to layout; otherwise he/she adds the failed corners and goes back to the design step.

Fig. 4. Left: Distribution of FastPVT speedups on 108 industrial problems. Right: QQ plot of sense amp power distribution, comparing 1M MC samples (1M simulations) to 100M HSMC samples (<10K simulations).

B. Fast Statistical Design and Verification

Of course, PVT is not always the way. Some designers have access to sufficiently good statistical MOS models to consider doing statistical analysis, which is inherently more accurate than PVT. However, since Monte Carlo (MC) simulations are far too slow within the design loop, MC is traditionally run as a verification afterthought. For high-sigma, the challenge is even greater, since it is not feasible to do the 5 billion MC simulations to verify at 6 sigma yield. Finally, designers tend to think in terms of corners, not statistics.

There is a way for handle these challenges. The key is to extract statistical corners that actually bound the 3-sigma or 6-sigma output performances for the circuit, versus traditional MOS corners like “FF” which bound the performances of the device. Then, changing device sizes to improve performance on statistical corners will improve the whole performance distribution. This is followed up by verification. Figure 5 middle and right illustrates the statistical flows.

The statistical flows in Figure 5 require corner extraction / verification tools that are fast, accurate, and scalable. These lead to subproblems that can be addressed via advanced statistical, optimization, and machine learning techniques. 3-sigma corner extraction uses nonparametric density estimation [14] on a small set of MC samples to identify a 3-sigma target output value, then optimizes in process variable space to find a process point that meets the target value. 3-sigma verification relies on scalable low-discrepancy sampling [15] to minimize the number of samples to verify 3-sigma yield.

For high-sigma, importance sampling [16] and statistical blockade [17] are popular in the literature; unfortunately, those have yet to demonstrate scalability to more than 6-12 process variables. High-sigma MC (HSMC) [18] generates a huge set of MC samples (e.g. 5 billion), then uses adaptive machine learning to rapidly identify the MC samples that have extreme output values and simulate them. HSMC has bounded complexity because it operates on a finite set of samples; and has been applied to industrial problems with >1000 variables. Figure 4 right shows example results on a sense amp with 150 process variables; we see that HSMC is able to correctly identify the tail (top right hand corner) from 100M samples generated in <10K simulations.

These advanced techniques enable corner-based, accurate, and scalable flows for PVT, statistical, and high-sigma variation. As Figure 5 illustrates, the flows are are identical, except for the corner extraction / verification tool. The result is a unified, designer-friendly approach for variation-aware design.

III. FAST FAULT SIMULATION ALGORITHM FOR FAULT AND MONTE-CARLO SIMULATIONS

A. Introduction

Simulating modern Analog Mixed Signal circuits typically requires significantly large CPU time, especially for Defect Oriented Test (DOT) Fault and Monte-Carlo (MC) simulations, where one needs to perform a significant number of simulations per circuit, see [19]. Fault simulations for DOT involve calculations of the impact of probable defects (typically bridges or opens) that are injected into the netlist of the circuit and performing analog simulation (e.g. DC,
Transient, and AC) of the predefined test benches. Since the number of possible defects is large (typically 1000-5000 for moderately large IC) CPU time needed to complete the analog simulations is hardly feasible. Therefore, multiple ideas have been introduced to overcome this issue, see e.g. [20]–[22]. In [23], [24] we introduced applications of the novel simulation technique that provides a significant speedup for fault simulations. Here we provide an overview of the Fast Fault Simulation algorithm with the extension, results obtained and possible applications for statistical MC simulations.

B. Fast Fault Simulation Algorithm

The standard way of performing the fault simulations is launching a series of sequential simulation runs. The runs are unrelated, i.e. every fault run simulates the entire time interval without any reuse of the knowledge obtained from the golden (fault free) and previous fault runs. The main idea of the Fast Fault Simulation approach is the reuse of the already obtained simulation data by defining a parallel run during which all the circuits are simulated simultaneously, as shown in Figure 6.

![Fig. 6. Comparison between standard approach and Fast Fault Simulation](image)

The Fast Fault Simulation concept is based on usage of the already obtained golden circuit solution as the initial value of the Newton-Raphson iterative process (enabling faster convergence) and the bypassing of the unchanged circuit parts which are unaffected by the fault impact. The benefit of two speedup concepts can be illustrated by an example, shown in Figure 7, where an NXP automotive product with 61 faults has been analyzed by applying a transient test.

![Fig. 7. Reuse of golden solution and bypassing benefits of FFS algorithm (bypassing = 1 means the whole matrix is solved)](image)

Assuming that all fault simulations require similar effort as the golden simulation (workload corresponding to the first input on the horizontal axis), it is clear that the algorithm introduce benefits from the both aspects which are multiplied to provide the final speedup. In practical applications, the algorithm performs significantly better than the standard approach, as shown in Table I, where several designs have been analyzed. TJA1050 and TJA1055 are NXP products from the family of CAN Transceivers, while CDAC stands for Control Digital to Analog Converter, an IP block used in multiple NXP products.

![Table I: FFS SPEEDUP RESULTS ON SEVERAL NXP DESIGNS](image)

The results shown in Table I are produced by using NXP in-house simulator Pstar. However, the idea is independent of simulator or simulation technique, since the algorithm can be easily superposed to any analysis of choice.

C. Fault Sensitivity Analysis

Although FFS algorithm introduces a significant speedup, for the already mentioned DOT setups the performance is not always sufficient. Therefore, for most frequently used and time-consuming transient simulations we exploit two additional characteristics of the fault simulation: 1) the main objective of the DOT method is to establish the detection status of a defect and not to determine the actual output values of the faulty circuit, 2) the output value is often measured at a few time points and therefore it is sufficient to compute the output value only for these time points while many time points have to be computed in a standard transient analysis. Therefore, we introduce a notion of the numerical discrete bridge model that is present in the topology of the circuit at measurement points only, as shown in Figure 8.

![Fig. 8. Time points reduction: (a) standard model (b) bridge present at the measurements points only](image)

Of course, such an approach introduces significantly larger speedups, as well as inaccuracies of the fault solution. To keep, at least partly, the accuracy under control, we introduce the nonlinear correction technique, which is based on Newton-Raphson iterative correction steps. The algorithm has been applied on already mentioned TJA1055 with 18 tests and 38000 extracted faults. Matching vs. speedup results are shown in Table II. Since there are typically a large number of tests, overall accuracy is at the acceptable level for practical application. For example, total of 150 tests for TJA1055 has 95.1% average matching, which proved accurate enough when compared to 1.5M IC samples measured.
D. Monte-Carlo Simulations

Comparing MC and fault simulation problems, it is clear that there are similarities. For example, many MC samples are quite close to the nominal value or at least to each other. Therefore, combining multiple samples in a joined run would allow to exploit similar benefits as introduced for fault simulation problem. The main difficulty here is how to superpose the simulation algorithm to dynamic sampling techniques, i.e. the ones that choose new samples based on the outcome of the previous ones.

IV. ADVANCES IN VARIATION-AWARE TESTING OF ANALOG ICs

A. Analog Test Challenges in the SoC Era

Today’s circuit design trends, namely channel-length scaling, dense integration of heterogeneous systems onto a single die, reduction of the form factor, etc., as well as the requirement to meet stringent specifications, result inadvertently in increased process variability. This reality places an ever increasing emphasis on efficient test strategies to screen out and discard circuits that fail to meet the desired specifications. To this end, there has been a shift in the analog test paradigm in recent years towards full specification-based test suites. The current practice nowadays is to verify indiscriminately, one-by-one all the performances that are promised in the datasheet. As a result, the cost of high-volume testing has been increased dramatically in the recent years. According to recent industry reports [25], testing the analog, mixed-signal, and RF functions of a modern SoC may reach more than 50% of the overall production cost, including silicon and packaging costs. The test cost is mainly due to the sophisticated automatic test equipment that is required and the lengthy test times that are involved. Thus, the reduction of test cost can be achieved by minimizing the test time per device, by increasing the number of devices that are tested in parallel, and by alleviating the dependence on expensive test equipment.

B. Towards Analog Test Cost Reduction

1) Statistical test: A generic approach to reduce the test cost is to replace the specification-based tests by alternate measurements that can be extracted rapidly on a low-cost assortment of test equipment. Thereafter, the performances can be implicitly inferred from the alternate measurements, as long as there exists a high level of correlation between them. Typically, however, it is not possible to obtain a closed-form mathematical expression of the mapping between the alternate measurements and the performances. For this purpose, the mapping is derived through statistical learning which employs a representative sample of devices that is collected during the ramp-up phase across different wafers and lots. Once the mapping is derived, it can be readily used to test future devices based solely on alternate measurements.

There are two types of mappings that we can establish. One possibility is to perform a direct go/no-go test by learning the mapping \( f_1 : x \rightarrow \text{pass/fail} \), where \( x \) denotes the vector of alternate measurements. The simplest approach is to assign test limits individually to each of the alternate measurements which effectively results in examining the footprint of \( x \) with respect to the position of a hyper-rectangular in the alternate measurement space. However, a hyper-rectangular acceptance region is often a crude approximation and, thus, a more sophisticated approach is to allocate a non-linear hyper-surface [26]. The second possibility, known as alternate test, is to predict the values of the individual performances by learning mappings of the form \( f_2 j : x \rightarrow p_j \), where \( p_j \) is the \( j \)-th performance, [27]. In this case, we learn \( n \) regression functions, where \( n \) is the number of performances.

2) Built-in test: Built-in test consists of adding auxiliary circuitry into the device with the aim to perform part of the test on-chip and, thereby, to provide simple digital, DC or low-frequency measurements to the test equipment or even just the pass/fail information. Built-in test facilitates the test of embedded blocks, enables parallel test, reduces the complexity of test instrumentation, and helps to diagnose the source of failure so as to provide valuable information for yield enhancement.

One built-in test approach is to migrate some features of the test equipment into the circuit, i.e. build a miniature tester on-chip, in order to perform various curve tracing, oscilloscope, and spectrum analysis tasks. Several integrated test cores have been demonstrated aiming at characterizing the baseband frequency response [28], generating arbitrary band-limited waveforms [29], and digitizing arbitrary periodic analog waveforms [29].

Built-in test can also rely on reconfiguring the device into an easily testable form. For example, a loop-back connection can be established between the transmitter and the receiver chains of an RF transceiver, in order to use purely baseband test signals [30]. Another idea is to connect the device in a negative feedback loop during the test mode such that sustainable oscillations are produced at the output [31]. The amplitude and the frequency of the oscillation can be used to detect the presence of defects within the device.

Another built-in test approach is to integrate sensors into the device with the aim to extract digital, DC or low-frequency test signatures that nevertheless carry higher frequency information. Several such sensors have been demonstrated for RF device, including envelope detectors [32], current sensors [33], and process sensors [34]. Process sensors are particularly attractive because they are non-intrusive, i.e. they do not degrade the performance of the device, whereas the envelope detectors and current sensors necessarily require to be co-designed with the device. The process sensors are basic analog

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Worst match</th>
<th>Best match</th>
<th>Speedup</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>100%</td>
<td>100%</td>
<td>1</td>
<td>3 years</td>
</tr>
<tr>
<td>Linear FSA</td>
<td>72%</td>
<td>97%</td>
<td>1050</td>
<td>1 day</td>
</tr>
<tr>
<td>Nonlinear FSA (5 iter)</td>
<td>89%</td>
<td>97%</td>
<td>238</td>
<td>4.2 days</td>
</tr>
<tr>
<td>Nonlinear FSA (10 iter)</td>
<td>92%</td>
<td>97%</td>
<td>173</td>
<td>5.8 days</td>
</tr>
</tbody>
</table>

TABLE II
FAULT SENSITIVITY ANALYSIS PERFORMANCE
stages that mimic part of the architecture of the device (i.e. bias stage, current mirrors, etc.) or basic layout components (i.e. transistor, capacitor, etc.). They are laid out in close proximity to the device and they monitor it by virtue of being subjected to the same process variations. In particular, any degradation in the performances of the device reflects on the outputs of the sensors that shift away from their nominal values.

3) Adaptive test: Adaptive test is the dynamic adjustment of the test program (e.g. test limits, test content, and test flow) based on historical and real-time test data. In effect, the aim of adaptive test is to define a decision-tree structure that in the extreme could result in each device being uniquely tested by a flow dictated by the observed response and by historical test data. Adaptive test consists in (a) eliminating specification tests on a lot-to-lot basis [35], (b) adapting the test limits based on the test results of the preceding device in order to obtain better quality control [36], (c) changing the order of specification tests on-the-fly to move forward tests that have proven to detect many failing devices [37], (d) assessing on-the-fly the confidence of test decisions based on alternate measurements and, in case the test decision is deemed to be prone to error, forward the device to the standard specification test approach [38], (e) skip specification tests if their pass probability is beyond a confidence level threshold [39].

REFERENCES


