Hierarchical Analog Circuit Reliability Analysis using Multivariate Nonlinear Regression and Active Learning Sample Selection

Elie Maricau, Dimitri De Jonghe and Georges Gielen
K.U. Leuven, ESAT-MICAS, B-3001 Heverlee, Belgium
Email: georges.gielen@esat.kuleuven.be

Abstract—**The paper discusses a technique to perform efficient circuit reliability analysis of large analog and mixed-signal systems. The proposed method includes the impact of both process variations and transistor aging effects. The complexity of large systems is dealt with by partitioning the system into manageable subblocks that are modeled separately. These models are then evaluated to obtain the system specifications. However, highly expensive reliability simulations, combined with nonlinear output behavior and the high dimensionality of the problem is still a very challenging task. Therefore the use of fast function extraction symbolic regression (FFX) is proposed. This allows to capture the high-dimensional nonlinear problem with good accuracy. Also, an active learning sample selection algorithm is introduced to minimize the amount of expensive aging simulations. The algorithm trades of space exploration with function nonlinearity detection and model uncertainty reduction to select optimal model training samples. The simulation method is demonstrated on a 6 bit Flash ADC, designed in a 32nm CMOS technology. Experimental results show a speedup of 360x over existing aging simulators to evaluate 100 Monte-Carlo samples with good accuracy.**

I. Introduction

For over three decades, scientists have been scaling CMOS devices to increasingly smaller feature sizes to meet requirements on speed, complexity, circuit density and power consumption demanded by many advanced applications. However, going to these ultra-scaled CMOS devices also comes at a cost. Guaranteeing circuit reliability over the entire lifetime of a system is one of the major challenges designers are faced with today [1], [2]. Circuit reliability issues can be categorized into spatial and temporal unreliability effects [2]. The former are related to process variability and are fixed in time and visible right after production. These effects depend on circuit layout, neighboring environment and process conditions, impact the geometry and structure of the circuit and can lead to yield loss. Temporal unreliability effects, on the other hand, are time-varying and change depending on operating conditions such as operating voltage, temperature, switching activity, presence and activity of neighboring circuits.

All of these effects interact with each other and can have a large impact on the performance of the entire system. At the same time, a full system simulation with respect to reliability is highly expensive to evaluate [3]. Therefore a hierarchical approach is mandatory, which in turn requires accurate modeling of each subblock. However, the parameterization of a subblock towards reliability analysis poses some tedious problems. Stochastic process and aging parameters combined with deterministic circuit inputs imply a high number of dimensions (> 10). Also, strongly nonlinear circuit behavior is expected because of large signal inputs. Above all, circuit reliability simulations tend to be highly expensive, which severely limits the amount of function evaluations. Standard Response Surface Modeling (RSM) based on classic Design of Experiments (DoE) can only direct one the abovementioned problems at the same time [4]. Miranda et al. [5] proposed a method to assess process variability in digital systems but they did not cope with expensive simulations and strong nonlinear dynamics. For analog circuits, reliability simulation including transistor aging and variability is still limited to the analysis of small blocks (< 100 transistors) [3].

978-3-9810801-8-6/DATE12/©2012 EDAA

This work tackles the problems described above and efficiently builds accurate system subblocks based on simulation results from an analog circuit reliability simulator that considers process variations, mismatch and ageing phenomena. The method includes:

1) Use of **fast function extraction symbolic regression** (FFX, [6]) allows to cope with the high number of dimensions and the nonlinear circuit behavior.
2) Use of a new and innovative **active learning sample selection** algorithm. This algorithm trades of space exploration with function nonlinearity detection and model uncertainty reduction to select optimal model training samples and to limit the amount of expensive aging simulations.

The presented method is demonstrated on a 6-bit flash Analog-to-Digital Converter (ADC) with over 1000 transistors, resulting in a 360x speed-up when compared to conventional analog circuit reliability simulators. Experimental results show how the average effective number of bits (ENOB) of this system reduces from 6.7 bits to 5.8 bits in 1 year time due to asymmetric stress at the input of each comparator.

Section II discusses the hierarchical simulator in more detail. Next, section III overviews different regression techniques and proposes a set of suitable regressors. Section IV then explains an active learning sampling algorithm to minimize the

number of function evaluations and demonstrates its effectiveness when compared to more traditional sampling techniques. The hierarchical reliability simulator is demonstrated on a 6-bit flash ADC circuit in section V. Finally, conclusions are drawn in section VI.

II. Hierarchical Reliability Simulation

Process variability and transistor aging are emerging problems in ultra-scaled ($<$ 90 nm) technologies. Ideally, one would like to extract performance parameters at system level ($\mathbf{p}_{sys}$) as a function of reliability parameters. In this work we consider deterministic input parameters $\mathbf{u}(t)$ (e.g. input amplitude), stochastic process parameters σ_p (e.g. $\sigma_{\Delta VTH,0}$) as well as stochastic aging parameters σ_{age} (e.g. $\sigma_{\Delta VTH,\text{aged}}$). The bold notation denotes a vector of parameters or observations.

A. Computational Complexity

The computation of a single combination of input parameters (i.e. one sample) becomes highly expensive when evaluating the circuit reliability due to the inherent complexity of transistor aging prediction [3]. The simulation of a small circuit subblock (e.g. 10 devices) for a certain combination of aging- and process variability related parameters and input waveform parameters easily takes a few minutes. The computational complexity increases exponentially when doing a full factorial analysis of large-scale systems with a rich set of input stimuli. When looking at a commercial mixed-signal design flow, this even becomes infeasible.

A DoE method for linear systems requires at least $2K + 1$ samples, where K equals the number of explanatory parameters. To model weakly nonlinear circuit behavior, more additional experiments are needed. Without prior knowledge, the dimensionality of the problem therefore grows at least linearly with the number of parameters. Dimensions of order 10 to 50 are commonly encountered in practice [3]. The strongly nonlinear dynamics of the deterministic input parameters even introduce an exponential growth in the number of samples. When only considering the amplitude and frequency of M_i single-tone input signals for the subblock models, a two-level full factorial DoE would require in the order of $(2M_i)^{2K+1}$ samples.

B. Hierarchical Simulator Setup

The problem associated with circuit reliability analysis of large and complex systems can be dealt with as follows. Firstly, the system is partitioned in local subblocks of manageable size ($10-30$ devices) with only a few input terminals. These subblocks are typically identified manually by the designer according to the hierarchy in the design database (e.g. opamp stages, comparators, filters, etc...), although automatic subblock detection could also be included in the flow. The performance parameters at system level $\mathbf{p}_{sys}$ are now a function of the performance parameters of each subblock $\mathbf{p}_{block}$ (see Fig. 1):

$$\mathbf{p}_{sys} = f\left(\mathbf{p}_{block}(\mathbf{u}(t), \sigma_p, \sigma_{age})\right) \tag{1}$$

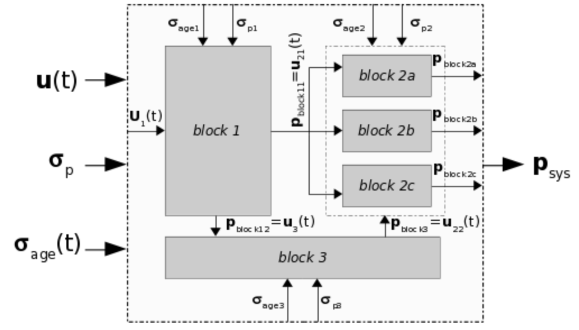

Fig. 1. Conceptual representation of a system with local subblocks. The performance of each subblock $\mathbf{p}_{block}$ is determined by deterministic input parameters $\mathbf{u}(t)$ and two sets of stochastic input parameters, σ_p and σ_{age} respectively.

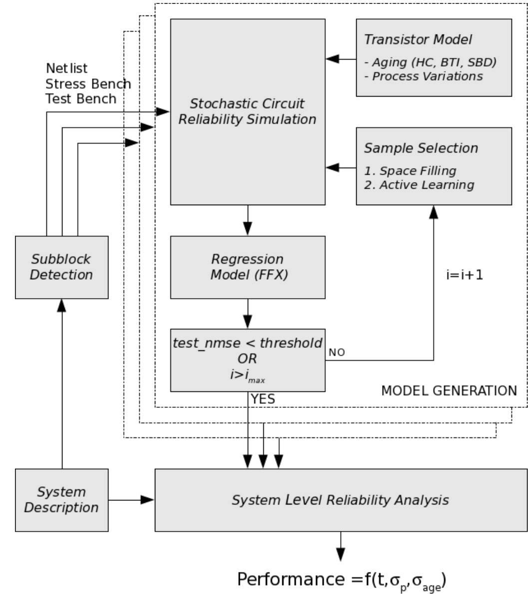

Fig. 2. The hierarchical system reliability simulation flow.

Typically, subblock performance measures (e.g. offset voltages ΔV, gain-bandwidth, delay, etc.) can be modeled as a weakly nonlinear function of the stochastic aging and process parameters σ_{age} and σ_p [3]. The referenced experiments are conducted and validated for a fixed set of input waveforms defined in a circuit stress bench. However, to use such a subblock model in a hierarchical system analysis flow, the input parameter space also needs to be included in the model. This increases the dimensionality of the problem even further and a more strongly nonlinear behavior can be expected.

The solution, presented in this work, to hierarchically simulate a large mixed-signal circuit or system is schematically represented in Fig. 2. First, the system is divided into subblocks. Then, every subblock is modeled using a *stochastic aging simulator* as described in [3]. The simulator uses transistor aging models for hot carrier injection [7], bias temperature instability [8], and soft breakdown [3] and evaluates the performance of a system subblock instance over time. One subblock instance corresponds to a sample, taken from the parameter space with deterministic input parameters

and stochastic aging and variability parameters. The high-dimensional parameter space and expensive circuit reliability computation require the amount of experiments, needed to model the subblock behavior, to be minimized. This is done by implementing an efficient *sample selection* algorithm (see Fig. 2 and section IV). The behavior of the circuit is modeled with a *fast function extraction (FFX)* symbolic regressor (SR), as explained in section III. After a new model has been generated, a new sample is selected based on the spatial and model uncertainty. Finally, the overall system performance is evaluated using the models for each subblock. To reduce simulation time, different circuit subblocks can be modeled in parallel on different computer cores.

III. Multivariate Nonlinear Regression

The regression problem for each subblock can be written as a least-squares minimization problem:

$$arg \min_{\alpha_i} |\mathbf{p}_{block}(\mathbf{u}(t), \sigma_p, \sigma_{age}) - \hat{\mathbf{p}}_{block}(\alpha_i, \mathbf{u}(t), \sigma_p, \sigma_{age})|^2 \tag{2}$$

where $\mathbf{p}_{block}(.)$ is obtained from subblock level simulations, $\hat{\mathbf{p}}_{block}(.)$ represents the regression model with the model parameters α_i. Following multidimensional regression approaches were considered for this work: multivariate adaptive regression splines (MARS) [9], least-squares support vector machines (SVM) [10] and a recently developed deterministic SR technique, fast function extraction (FFX) [6]. Interpolation algorithms are not considered due to their poor extrapolation performance for high-dimensional problems.

A recent comparison between these multidimensional regression techniques has been presented in [6]. Here, it was shown that the evolutionary based SR CAFFEINE [11] and modern feedforward neural networks (FFNN) [12] are less suitable regressors for high-dimensional test cases due to unreasonably long building times or by being too inaccurate (i.e. test error $> 100\%$).

The performance of the remaining regression techniques (SVM, MARS and FFX) is compared for the 2-dimensional test case shown on the left of Fig. 3. In this comparison, samples are progressively added[1] to the known set of samples and a model is built for 75% training samples and 25% test samples. The prediction ability of each regressor is tested by plotting the sum of the test and training normalized mean square error (NMSE) at each generation in the right part of Fig. 3. The error for MARS and FFX easily drops below 1% when more than 10 samples are available, while the error of SVM stays at approximately 10%. This is mainly due to the internal regularization paths of MARS and FFX. Here, the regression objective is biased toward cross-validation and minimization of the error on the test samples, which prevents overfitting of the data and ill-conditioned model parameters. Moreover, FFX generates a Pareto-optimal set of models that trade off model complexity with test error by ramping up the coefficients in the elastic net formulation [6]. To avoid overfitting even more, a weighted model evaluation can be used by selecting the

[1]Sample selection is done with the algorithm proposed in section IV.

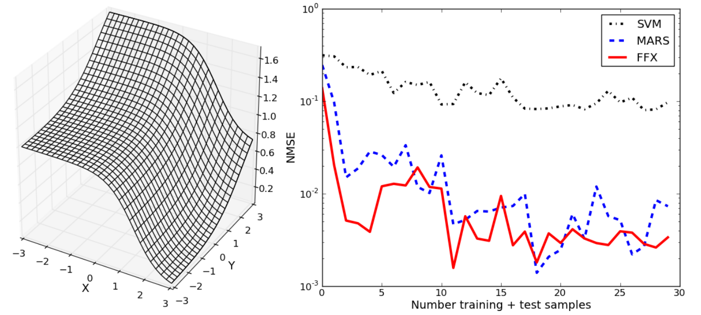

Fig. 3. Left: Test function: $\frac{1.0}{1.0+exp(2.0(x-1.5))} + \frac{1.0}{1.0+exp(-1.0(y-2.2))}$; Right: Train + test error (NMSE) of SVM, MARS and FFX for a progressively increasing amount of data points, selected by the active learning strategy discussed in section IV.

weights inverse proportional to the test error. The weighted and normalized model formulation $\overline{f}_{FFX}(.)$ for K Pareto-optimal models then becomes:

$$\overline{f}_{FFX}(.) = \sum_{k=1}^{K} w_k \cdot f_{FFX,k}(.), \tag{3}$$

with

$$w_k = \frac{(\text{nmse}_k)^{-1}}{\sum_{j=1}^{K}(\text{nmse}_j)^{-1}}. \tag{4}$$

The experiments demonstrated further on in this paper are implemented using the FFX regressor. Of course, a straightforward extension can be made by building multiple regressors of a different class simultaneously and to vote or average between them.

IV. Active Learning Sample Selection

The expensive simulation times and high dimensionality of reliability simulations render a sparse dataset. A full exploration of the parameter space requires the selection of the next sample that is added to the dataset to be chosen in such a way that the density of the samples is uniformly distributed. This is the philosophy behind space filling sampling algorithms such as LHS, uniform random sampling, FF designs, etc. [4]. Progressive sampling strategies such as Monte-Carlo random sampling (MCS) do not necessarily consider previously generated samples and thereby ignore any knowledge about the global sample density and the correlation between the samples. In addition, strongly nonlinear behavior is expected for parameterized signal inputs in the circuit stress bench. Sharp transitions or steep edges in the performance space are preferably sampled more densely than flat or weakly nonlinear regions. *Active learning* or *co-evolution* is a supervised machine learning technique where the selection of new inputs is controlled such that the added value of newly gathered information is optimal [13]. In statistics literature this is described as optimal experimental design [14].

The basic setup of active learning sample selection is to predict, for every new generation, at which locations in the input parameter space one would expect the model to have the highest uncertainty. The uncertainty predictor $D(.)$ is estimated by a distance metric. Such a metric compares inputs

to inputs, outputs to outputs, and models to models. In this work, the distance metric between two points is expressed as the Euclidian distance or 2-norm $\| \, . \, \|_2$.

A distance measure for the input space is declared as follows. Consider $\mathbf{x_L} \in \mathbb{R}^N$ as the collection of L data samples of the *known* N-dimensional dataset, i.e. the points that already have been simulated. The distance of a newly selected point $\mathbf{x}^* \in \mathbb{R}^N$ to the nearest point in the known dataset $\mathbf{x_L}$ is then expressed as:

$$D_x(\mathbf{x}^*) = \min \| \, \mathbf{x}^* - \mathbf{x_L} \, \|_2 \, . \tag{5}$$

Taking the minimum distance to the known samples forces the algorithm only to look at the nearest neighbor $\mathbf{x_n} \in \mathbf{x_L}$. A typical space filling sample selection algorithm maximizes the distance function $D_x(\mathbf{x}^*)$ such that newly selected samples are chosen as far as possible from previously visited places. The nonlinear behavior of the performance measures as a function of large-swing inputs is accounted for by defining two additional distance functions D_y and $D_{\text{var}(\hat{y})}$.

Abrupt changes in the L output performance values $y_L \in \mathbb{R}$ are predicted by the relative distance of the model output $\hat{y}_L(\mathbf{x}^*)$ to the known output of the nearest neighbor in the parameter space $y_n = y(\mathbf{x_n}) \in y_L$:

$$D_y(\mathbf{x}^*) = \| \, \frac{y_n - \hat{y}(\mathbf{x}^*)}{\max(y_L) - \min(y_L)} \, \|_2 \, . \tag{6}$$

When the distance between the model output and the nearest known output is large, steep edges tend to occur. Adding samples at those locations refines the model by extracting more information at those places. Different output specifications are combined into a single distance measure by taking the mean value over all computed relative output distance measures. An example of output distance active learning is illustrated on two shifted 2-dimensional $sinc$ functions in Fig. 4. It can be seen that more samples are inserted where the peaks (indicated by the contours in Fig. 4) occur.

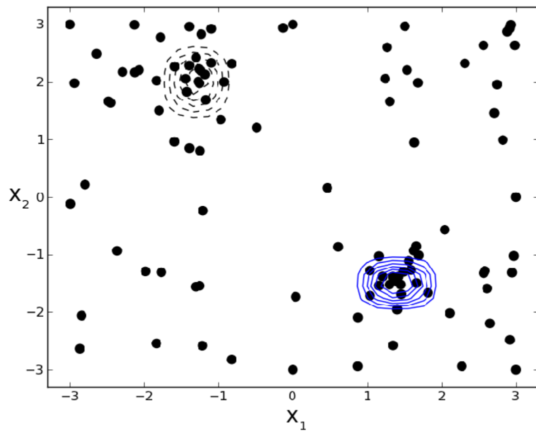

Fig. 4. Output distance sampling for the test function: $\{f_1(\mathbf{x}) = 4 \cdot 10^5 \cdot sinc(1.7(x_0 - 1.4)) \cdot sinc(2(x_1 + 1.5)); f_2(\mathbf{x}) = 3 \cdot 10^{-4} \cdot sinc(2(x_0 + 1.3)) \cdot sinc(1.4(x_1 - 2.0))\}$

As a third predictor, the variance of the model is computed by means of bootstrapping [15]. Bootstrapping provides a direct computational method of assessing the model uncertainty. Several models $\mathbf{\hat{y}_m}$ are built for different random permutations of training and test samples. A point in the parameter space where a large variance between the models occurs, corresponds to a large disagreement between the models. This is illustrated in Fig. 5. The model variance is normalized to the variance of the median of all deviation models:

$$D_{\text{var}(\hat{y})}(\mathbf{x}^*) = \frac{\sigma^2(\mathbf{\hat{y}_m}(\mathbf{x}^*))}{\sigma^2 \left[\mu_{1/2}(\mathbf{\hat{y}_m}(\mathbf{x}^*))\right]} \tag{7}$$

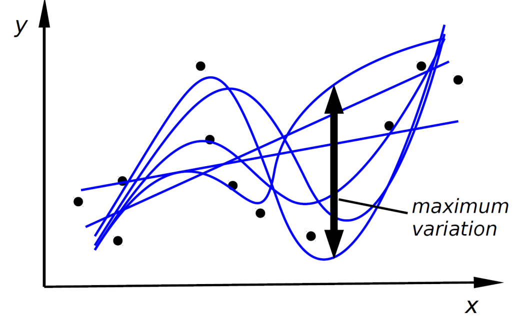

Fig. 5. Six bootstrap models for the same dataset. The maximal variance of the models is also plotted.

Finally, the total distance function we used is a combination of the distance functions (5), (6) and (7). Note how the total distance function is forced to reach a minimum value when the input distance function equals zero (i.e. when the sample is already included in the known dataset):

$$D_{tot}(\mathbf{x}^*) = D_x(\mathbf{x}^*) \cdot [1 + D_y(\mathbf{x}^*)]^\alpha \cdot \left[1 + D_{\text{var}(\hat{y})}(\mathbf{x}^*)\right]^\beta \tag{8}$$

The exponent parameters α and β skew the weight of the distance function towards exploration ($\alpha = \beta = 0$) or towards nonlinearity sampling ($\alpha = \beta = 1$).

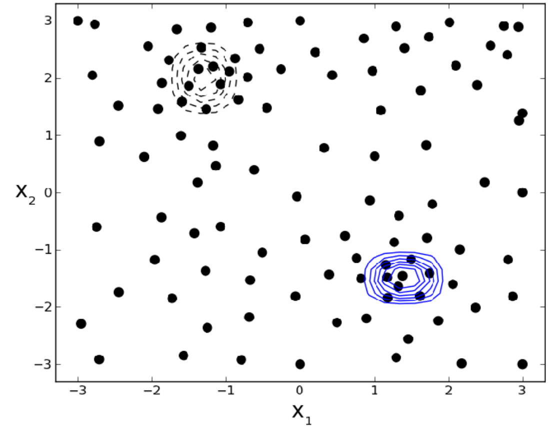

Fig. 6. Active learning sample selection using the total distance function $D_{tot}(\mathbf{x}^*)$ on the test function shown in Fig. 4.

The next best sample, given a known dataset $\mathbf{x_L}$, corresponds to the point where the distance function reaches a maximum:

$$arg \max_{\mathbf{x}^*} D_{tot}(\mathbf{x}^*) \tag{9}$$

This optimum can be found with a common-purpose global optimization engine such as the Multi-Objective Evolutionary Algorithm (MOEA) or with Simulated Annealing approaches [16], [17]. An example of sample selection using the total distance function $D_{tot}(\mathbf{x}^*)$ is plotted in Fig. 6. It can be seen that the sample density is optimally distributed in space and that more dense sampling is encountered at the function peaks. The proposed active learning sample selection algorithm is also compared to MCS for the test function depicted in Fig. 3. Fig. 7 shows the NMSE of the FFX regressor after each generation using both sampling strategies. On average, the NMSE using the active learning selection drops faster than the 75 percentile of the random sample selection. This demonstrates that, of all possible samples, the proposed algorithm selects one of the best samples to further reduce the model error.

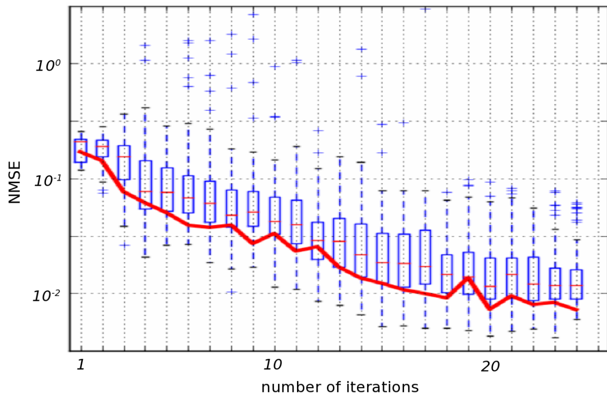

Fig. 7. Active learning sample selection versus Monte Carlo for the test case of Fig.3 using FFX models as a function of sampling iteration.

The proposed active learning sample selection strategy finds the next best sample based on information about samples that are already present in the known dataset. To start the model building algorithm, an initial dataset is needed. To build this dataset a space filling DoE such as LHS can be used. Note how the size of this initial dataset does not have to increase with the number of dimensions N. It only needs to contain sufficient samples to get the active learning algorithm started.

V. Experimental Results

The proposed hierarchical simulation flow of Fig. 2 has been applied to a 6-bit flash ADC test circuit, designed in a predictive 32nm CMOS technology with a 1V supply voltage [18] (see Fig. 8). The analog part of the circuit consists of more than 1000 circuit elements. The ADC contains 63 clocked comparators, each comparing the input voltage to a different reference voltage. The comparator is identified and modeled as a one-system subblock, with the reference voltage as a deterministic input that can vary between the ground and supply voltage. As an input to the ADC, a full-scale sinewave of fixed frequency and amplitude was applied. The hierarchical model was then built as a function of the deterministic input-reference voltage and the stochastic process and aging parameters. Evaluation of this model within an aging simulator

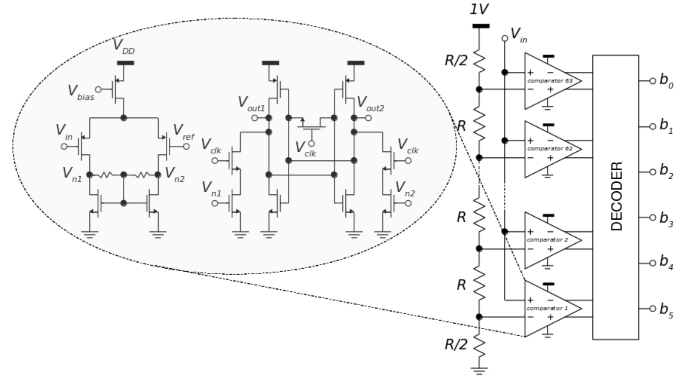

Fig. 8. Schematic representation of the demonstrator 6-bit flash ADC. The ADC is designed in a 32nm predictive technology and uses clocked comparators to compare the reference voltages with the input.

returns a tuple of time-dependent input-referred offset voltages between 0 and 1 year of operation. The accuracy of an ADC is typically described by the effective number of bits (ENOB), which is in turn determined by the integral and differential linearity of the converter (INL and DNL respectively [19]):

$$\text{ENOB} = \log_2 \left[\frac{V_{\text{in,min}} - V_{\text{in,max}}}{\max\left(2 \cdot \text{INL}, \text{DNL}\right)} \right] \tag{10}$$

Both the INL and DNL are mainly determined by mismatch between the resistors of the reference ladder and by the input-referred offset of each comparator. Right after production, both are only determined by process variations. Mismatch can however change over time due to the NBTI effect [8]. Fig. 9 depicts the input-referred offset for each comparator after 1 year of stress and for 100 Monte-Carlo samples, all derived from the comparator subblock model. Comparators at the top and the bottom of the reference ladder are particularly sensitive to transistor aging since they suffer from large asymmetric voltage stress. The bottom comparator for example (i.e. comparator 1 in Fig. 8), is at one side stressed by a very low reference voltage, while the other side sees the ADC input (i.e. the sinewave signal). Since NBTI is exponentially dependent on the magnitude of the gate voltage stress, this results in a large threshold voltage mismatch between the input transistors (on average $\Delta V_{TH} = 17mV$ at 1 year for comparator 1). A similar effect can be observed for comparators at the top of the reference ladder (e.g. comparator 63 in Fig. 9). The input offset increases over time and results in a reduction of the ENOB. Fig. 10 shows a normal probability plot of the ENOB right after production, after 1 month of operation and after 1 year. The solid lines are the ENOB as computed by the hierarchical models of the comparators, while the markers represent the ENOB calculated from a full system aging simulation. From Fig. 10 it can be seen how process variations cause a large initial spread on the ENOB, while the graph shifts towards lower values due to aging effects. It is clear how the results of the hierarchical simulation method presented here (i.e. solid lines) agree very well with results predicted by an existing aging simulator not using models for each system subblock. A good correspondence between the model and full system

simulations is observed at the time of interest (i.e. after 1 year of operation). Moreover, the logarithmic time dependence of the NBTI effect is observed [8]. The discrepancy at initial time is assumed to originate from active learning sample selection emphasis at long stress times.

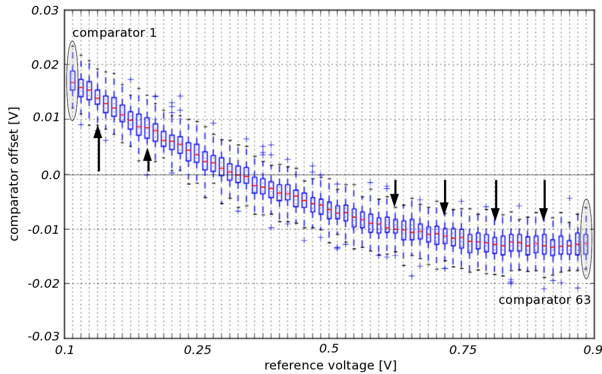

Fig. 9. The input-referred offset voltage for each flash ADC comparator after 1 year of stress and for 100 Monte-Carlo samples, all evaluated with the comparator subblock model.

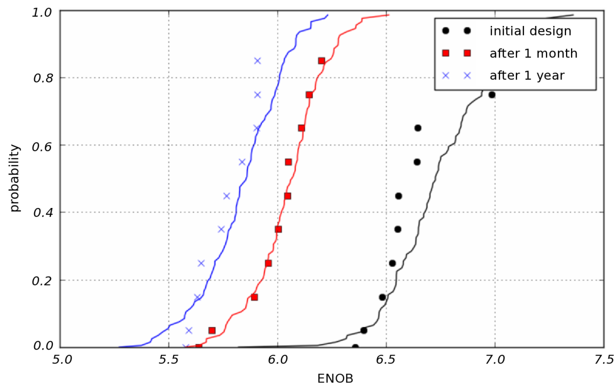

Fig. 10. A normal probability plot of the effective number of bits for 100 Monte-Carlo samples evaluated with the proposed hierarchical simulator (solid lines) and 10 samples evaluated with full system aging simulations (markers). The ENOB represents the static accuracy of the converter and decreases over time due to transistor aging effects.

The demonstrator circuit has been simulated on a dual-quad core 2.8GHz Intel Xeon processor with 8GB of RAM. Model build time for the comparator subblock took 31 minutes, while evaluation of the entire converter took 1 minute and 41 seconds for 100 Monte-Carlo samples. Evaluation of just one Monte-Carlo sample, using a traditional aging simulator took 1 hour and 55 minutes. This results in a speedup of 360x when evaluating 100 Monte-Carlo samples.

VI. Conclusion

A hierarchical bottom-up approach to perform efficient reliability simulations of large analog and mixed-signal systems has been presented. The proposed method includes the impact of both process variations and transistor aging effects. The proposed method first models each system subblock with a multivariate regression model, to capture the high-dimensional nonlinear subblock behavior. Also, an active learning sample selection algorithm has been proposed to minimize the amount of expensive aging simulations. Next, the simulator combines the regression models of the local system subblocks to analyze the overall system performance over time. The simulation method has been demonstrated on a 6-bit flash ADC. Experimental results show a speedup of 360x over existing aging simulators when evaluating 100 Monte-Carlo samples, while keeping a similar accuracy.

Acknowledgment

The authors acknowledge the financial support of FWO, IWT and ON Semiconductor Belgium BVBA in the frame of the projects CATRENE/GoldenGates. The work is also supported in part by IWT SBO Elixir.

References

[1] "International Technology Roadmap for Semiconductors (ITRS)," http://www.itrs.net/Links/2009ITRS/Home2009.htm, 2009.

[2] L. Lewyn *et al*, "Analog Circuit Design in Nanoscale CMOS Technologies," *Proceedings of the IEEE*, 2009.

[3] E. Maricau and G. Gielen, "Stochastic Circuit Reliability Analysis," in *Design, Automation Test in Europe Conference (DATE)*, 2011.

[4] D. Montgomery, "Design and analysis of experiments," 2009.

[5] M. Miranda, P. Zuber, P. Dobrovolny, and P. Roussel, "Variability Aware Modeling for Yield Enhancement of SRAM and Logic," in *Design, Automation Test in Europe Conference (DATE)*, 2011.

[6] T. McConaghy, "Ffx: Fast, scalable, deterministic symbolic regression technology," *Genetic Programming Theory and Practice IX, Edited by R. Riolo, E. Vladislavleva, and J. Moore, Springer*, 2011.

[7] E. Maricau *et al*, "An analytical model for hot carrier degradation in nanoscale CMOS suitable for the simulation of degradation in analog IC applications," *Microelectronics Reliability*, 2008.

[8] E. Maricau, L. Zhang, J. Franco, P. Roussel, G. Groeseneken, and G. Gielen, "A Compact NBTI Model for Accurate Analog Integrated Circuit Reliability Simulation," in *European Solid-State Device Research Conference*, 2011, Accepted for Publication.

[9] J. Friedman, "Multivariate adaptive regression splines," *The annals of statistics*, pp. 1–67, 1991.

[10] J. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[11] T. McConaghy and G. Gielen, "Template-free symbolic performance modeling of analog circuits via canonical-form functions and genetic programming," *TCAS*, vol. 28, no. 8, pp. 1162 –1175, aug. 2009.

[12] N. Ampazis and S. Perantonis, "Two highly efficient second-order algorithms for training feedforward networks," *Neural Networks, IEEE Transactions on*, vol. 13, no. 5, pp. 1064–1074, 2002.

[13] H. Lipson and J. Bongard, "An exploration-estimation algorithm for synthesis and analysis of engineering systems using minimal physical testing," in *ASME Design Automation Conference (DAC04)*. Citeseer, 2004, pp. 1087–1093.

[14] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[15] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, *The elements of statistical learning: data mining, inference and prediction*. Springer, 2005, vol. 27, no. 2.

[16] K. Deb, *Multi-objective optimization using evolutionary algorithms*. Wiley, 2001, vol. 16.

[17] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, p. 671, 1983.

[18] Arizona State University Nanoscale Integration and Modeling (ASU NIMO) Group, "32nm predictive technology model," *http://ptm.asu.edu/*, 2011.

[19] R. Van De Plasche, *Integrated Analog-to-Digital and Digital-to-Analog Converters*. Kluwer Academic Publishers, 1994.