On the limits of machine learning-based test: a calibrated mixed-signal system case study

Manuel J. Barragan
CNRS, TIMA, F-38000 Grenoble, France
Université Grenoble Alpes, TIMA, F-38000 Grenoble, France
e-mail: manuel.barragan@imag.fr

G. Leger, A. Gines, E. Peralias, A. Rueda
Instituto de Microlectrónica de Sevilla
CSIC-Universidad de Sevilla
Av. Américo Vespucio s/n, 41092 Sevilla, Spain.
e-mail: leger@imse-cnm.csic.es

***Abstract*—Testing analog, mixed-signal and RF circuits represents the main cost component for testing complex SoCs. A promising solution to alleviate this cost is the machine learning-based test strategy. These test techniques are an indirect test approach that replaces costly specification measurements by simpler signatures. Machine learning algorithms are used to map these signatures to the performance parameters. Although this approach has a number of undoubtable advantages, it also opens new issues that have to be addressed before it can be widely adopted by the industry. In this paper we present a machine learning-based test for a complex mixed-signal system –i.e. a state-of-the-art pipeline ADC– that includes digital calibration. This paper shows how the introduction of digital calibration for the ADC has a serious impact in the proposed test as calibration completely decorrelates signatures from the target specification in the presence of local mismatch.**

I. Introduction

Nowadays, production test standard practices for analog, mixed-signal and RF (AMS-RF) circuitry still rely on costly functional tests. These systems are tested on dedicated mixed-signal Automated Test Equipment (ATE) that applies precise test stimulus to the Device Under Test (DUT), acquires the DUT response and processes the test data to compute the set of functional specifications of the DUT. However, the high level of integration of current mixed-signal systems and the increasing performance of AMS-RF blocks make functional test a challenging task that requires costly dedicated ATEs. Even accessing the primary input and outputs of an AMS-RF block in a tightly integrated system may turn impossible.

Machine learning-based test, also known as indirect test or alternate test [1] is a promising strategy for overcoming this issue. The basic idea is to replace costly specification-based tests by a set of simpler (and cheaper) measurements, often called signatures, and then use machine learning algorithms to map these simple measurements onto the specification space. Conventional specification-based tests are then replaced at the production line by a set of low-cost indirect observations, and test results are then inferred by post-processing these observables.

The usual approach to machine learning-based test is based on supervised machine learning algorithms. The process is developed in two stages: a learning stage, and a testing stage. During the learning stage both performance parameters and signatures are measured from a set of training devices. A machine learning algorithm is then trained over the two sets of measurements to build a mapping model. In the testing stage, signatures are measured for each Device Under Test (DUT), and performances are inferred by using the mapping model obtained in the previous stage. Test result interpretation is the same as in conventional functional test. This is an elegant way to deal with the issue of valid ranges of acceptance, and also gives a valuable insight into reliability since we can estimate how far the circuit is from its nominal specification. In addition, machine learning algorithms are designed to handle complex multi-dimensional and non-linear relations like those between defects and specifications, and by extension, between signatures and specifications.

Machine learning-based test may lead to significant cost savings in production test, but this test strategy is not free of shortcomings that should have to be addressed before it can be confidently adopted by the industry. Thus, in the last few years, many researchers have pointed out the potential issues of the technique and proposed different strategies for overcoming them.

A key issue is finding an appropriate set of signatures to extract meaningful prediction models. Proposing appropriate signatures to run a machine learning-based test is often a matter of creativity based on a precise knowledge of the DUT. Thus, signature sets are usually *ad hoc* and sub-optimum: they may contain redundant information, non-relevant data, noisy signatures, etc. Different approaches have been proposed to select the most meaningful subset of signatures from a given set of signatures [2]–[4], and going even further to automatize the design of new signatures that target missing information in a given set of signatures [5].

Another issue that has gathered attention in the past few years is the *a priori* validation of machine learning test. That is, the assessment of its quality, usually in terms of some standard test metrics, before deploying the technique in the production line. Different simulation frameworks have been presented to evaluate the quality of alternative tests and compare it to traditional functional test [6]–[8], usually based on statistical modeling and Monte Carlo simulation.

Another fundamental issue of machine learning-based test derives from its model-oriented philosophy: any non-modeled errors [9] can lead to unexpected bad results. Model construction is always based on some premises that should ideally

978-3-9815370-8-6/17/$31.00

be verified. In the particular case of machine learning, it is assumed that the process that governs variability is stochastic and stationary. In this way, the statistic information learned from the training set is still valid for future devices. Clearly, process variations fit this description. Thus, the aim of machine learning-based test is to propose a set of signatures that are highly correlated to global process variations, with the tacit assumption that local uncorrelated variations (i.e. mismatch) will only add a small noise to the measurement that will be translated into a negligible error in the regression of the DUT specifications. Clear examples of this strategy are the test based on die-level process monitors and dummy circuits introduced in [10], [11] where mismatch between dummy circuits and DUT is neglected.

In this work we present a machine learning-based test strategy for estimating the ENOB of a state-of-the-art high-speed high resolution Analog to Digital Converter (ADC). It will be shown how the introduction of digital calibration for the ADC has a serious impact in the proposed test as calibration completely decorrelates signatures from the target specification, making the final specification dependent on random local variations instead of on global process variations. This paper presents a negative result in which the fundamental assumption that correlation to global process variations is enough for learning a regression model turns out to be incorrect due to the use of digital calibration. Although some methods for post-manufacture tuning (which can be seen as a form of calibration) using machine learning tools have been presented in [12]–[14], the goal of these works was not to predict the performance of the post-tuned circuit but rather to propose a cost efficient method to get the best possible performance out of the tuning knobs. To the best of our knowledge, this is the first time that the impact of digital calibration for complex system on a machine learning test framework has been analyzed and that this fundamental issue has been exposed.

The rest of the paper is organized as follows. Section II reviews the theoretical basis of machine learning-based test. Section III presents our case study, describes both the ADC under test and the proposed ENOB machine learning-based test technique. In Section IV we present and discuss the obtained results and finally Section V summarizes our main contributions.

II. Theoretical basis

As it was described in the introduction, machine learning-based test is built on the assumption that we can find a set of low-cost signatures highly correlated to the set of functional specifications of a given DUT, in such a way that we can replace the –often costly and complex– measurement of the specifications by these signatures. The set of specifications are then "predicted" from a machine learning model previously trained for extracting the regression function between signatures and specifications.

In a general scenario, and assuming that the DUT is free of catastrophic faults, the performance of the DUT is a function of its PVT (Process-Voltage-Temperature) coordinates. Assuming constant Voltage and Temperature, a performance P of a given instance j of the DUT can be expressed as

$$P = f\left(\mathbf{p_j},\, \mathbf{m_j}\right) \tag{1}$$

where $\mathbf{p_j}$ is the vector of global process parameters corresponding to this particular DUT instance, $\mathbf{m_j}$ is the vector of local variation parameters (i.e. mismatch) for each element in the DUT, and f is a scalar function –usually unknown– that links these technological process parameters to the specifications of the DUT. In the same way, the set of low-cost signatures should be specially crafted to be sensitive to the same degradation mechanisms as the performances. Let us consider without loss of generality a one-dimensional scenario with a single signature S. Again, assuming constant voltage and temperature conditions, the magnitude of signature S for instance j of the DUT can be expressed as

$$S = g\left(\mathbf{p_j},\, \mathbf{m_j}\right) \tag{2}$$

where, similarly to the previous case, g is an unknown scalar function.

In this scenario, machine learning-based test is built on the assumption that performance P and signature S are highly correlated, in such a way that we can develop (1) and (2) into

$$\begin{aligned} P &= \tilde{f}\left(\tilde{\mathbf{p}}_\mathbf{j}\right) + r_{j1} \\ S &= \tilde{g}\left(\tilde{\mathbf{p}}_\mathbf{j}\right) + r_{j2} \end{aligned} \tag{3}$$

where $\tilde{\mathbf{p}}_\mathbf{j}$ corresponds to the vector of global and local process parameters affecting both to P and S, r_{j1} and r_{j2} correspond to the contribution of the uncorrelated local variation parameters to P and S, respectively, and $\tilde{f}$ and $\tilde{g}$ are unknown scalar functions. From (3) if follows

$$P = \tilde{f}\left(\tilde{g}^{-1}\left(S - r_{j2}\right)\right) + r_{j1} \tag{4}$$

One fundamental assumption in machine learning-based test is that the contribution of the uncorrelated local variations, r_{j1} and r_{j2}, is considered to be small, in such away that the following identity holds

$$P = \tilde{f}\left(\tilde{g}^{-1}\left(S\right)\right) + o\left(r_{j1}, r_{j2}\right) \tag{5}$$

In this general scenario, a training set of DUT instances from which both P and S are extracted is used to learn a regression model using machine learning algorithms in order to approximate function $\tilde{f}\left(\tilde{g}^{-1}\left(\cdot\right)\right)$. The contribution of $o\left(r_{j1}, r_{j2}\right)$ is neglected and considered as a source of error in the proposed test strategy.

In this paper we will discuss a practical machine learning test scenario in which identity (5) does not hold because the contribution of uncorrelated variations becomes dominant.

III. Case study

A. The Device Under Test: 17-bit pipeline ADC with digital calibration

The Device under Test that we consider in this paper is a state-of-the-art pipeline ADC with a resolution of 17 nominal

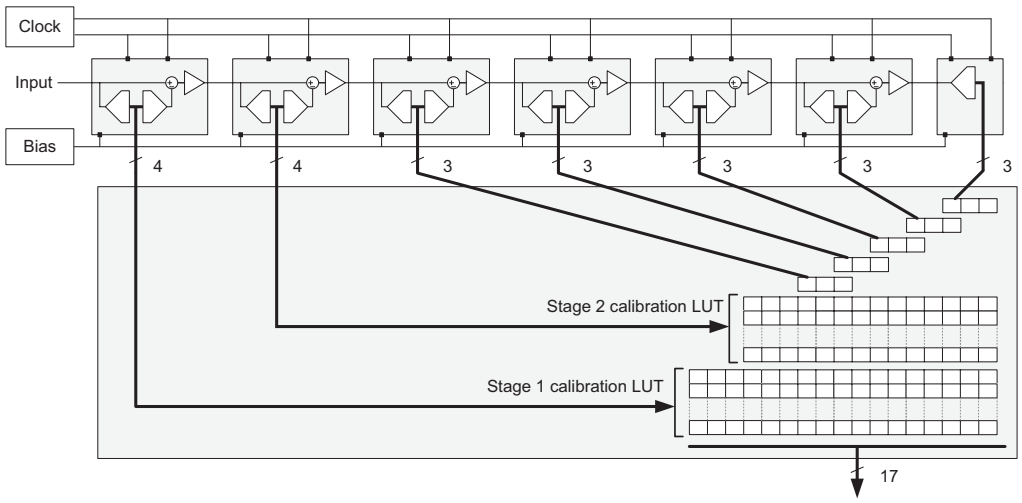

Fig. 1. Device under test: Pipeline ADC architecture

bits at 100MS/s built in a 1.8V $0.18\mu m$ CMOS technology. The target static effective resolution is 15 bits in nominal conditions, which poses an important test challenge. Indeed, it is broadly accepted that the tester performance should be at least a couple of bits above the target resolution of the device under test. In this case, an ATE with a signal generator above 17 effective bits at 100MHz would be necessary. Such an off-the-shelf signal generator can hardly be found, which implies the use of a complex load-board with adequate filtering.

Figure 1 shows the structure of the pipeline. It is composed of 6 stages plus a final Flash quantizer in a $[4433333]$bit configuration. Notice that this architecture does not make use of any front-end sample-and-hold in order to save power. The Multiplying DACs that generate and amplify the conversion residue for the next stage are switched-capacitor flip-around MDACs using two-stages operational amplifiers with Miller compensation. The same amplifier structure is used for all stages, conveniently scaled for power optimization.

As a usual practice, the stages incorporate an extra bit for redundancy such that small errors in the residue generation do not saturate the next stage. In addition, the first two stages (the ones that derive the 7 most significant bits) are self-calibrated in foreground. The calibration technique is performed in two steps based on [15]. In the measurement step, amplifier finite DC-gain and capacitor mismatches are estimated in the digital domain using the back end stages as measurement instrument. During the conversion mode, the calibration codes previously stored in the calibration registers (Look-Up Tables, LUTs in Fig. 1) are addressed by the stages under calibration to generate the final calibrated output.

B. Proposed machine learning test strategy for ENOB characterization

The purpose of our machine learning indirect test is to estimate the value of the Effective Number Of Bits (ENOB) of the calibrated pipeline ADC under test for a full-scale input sine-wave. The signatures that are used for the regression are the set of calibration codes (30 signatures, read from the calibration LUTs) and some simple measurements performed on dummy structures. These dummy structures are replicas of key circuit structures in the ADC that are only used in test mode and are disconnected from the circuit itself, as proposed in [10], [11]. In total, we use 3 dummy structures: an NMOS switch, a CMOS switch and a unit MiM capacitor. A total of 7 test signatures are extracted from these dummy structures: NMOS and CMOS switch ON resistance, ON and OFF time constants, and unit capacitor value. Summarizing, the proposed set of signatures has been crafted to contain the information on the calibration codes and simple measurements of the main dynamic limitations in the ADC under test. The calibration codes are read directly from the self-calibration routine and the measurements on the dummy structures are either DC or timing measurements and do not require a resolution anywhere close to the 17 bits of the ADC. The cost of acquiring the signatures is thus very reduced compared to that of measuring the ENOB at a 15 bit level.

The model that we train on that input space is an ensemble of perceptron neural networks [16]. Such an ensemble works as follows. The training set is randomly split into 5 equal-size partitions. One partition is set aside and a perceptron neural network is trained on the remaining 4 partitions using the conventional back-propagation algorithm. Its generalization error is estimated on the partition previously set aside. The same procedure is repeated 4 more times, changing the partition that is set aside. Notice that some initializing conditions of the network (e.g., the number of hidden neurons) can be randomly changed for each iteration. The best 3 models with respect to the generalization error are then selected and the final model output is computed as the average response of these 3 perceptrons.

In this study, the described modeling process with the same 37 signatures and the perceptron ensemble process is applied to predict the ENOB either before or after calibration in the context of global process variations only (also called inter-die variation) or with both global and local process variations

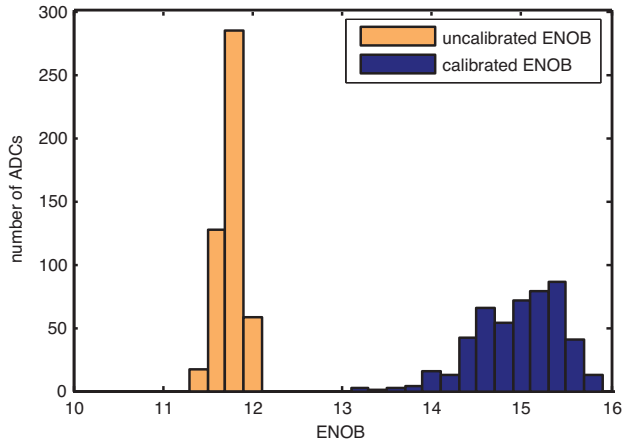

Fig. 2. ENOB distributions under process variations only

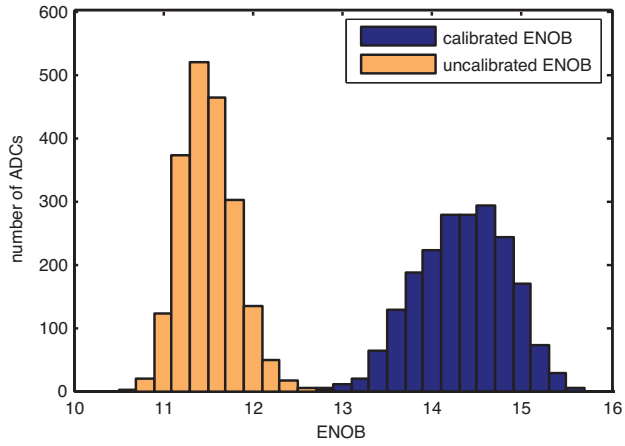

Fig. 3. ENOB distributions under both process variations and mismatch

(inter-die plus intra-die). This results in four different scenarios.

Calibration being a post-processing step, two Monte Carlo simulations had to be undertaken. For the first one, 500 instances were generated varying only process parameters. For the second one, both process and mismatch variations were contemplated. Since the dimensionality of the variation space is much greater than for the process-only simulation, the number of simulated samples was increased to 2000. In both cases, 10% of the instances were set aside to compute the generalization error of the ensemble model.

For all instances, the complete set of signatures and the ENOB were measured with and without calibration based on noise-free electrical simulations of the full transistor-level schematic of the core of the pipeline ADC. In particular, the core of the ADC includes op-amps, capacitor arrays, switches, comparators and common-mode feedback circuitry. The rest of auxiliary blocks in the converter (clock, bandgap, voltage/current references and digital logic) were modeled in VerilogA/Verilog to reduce computation time. To give an idea of the ADC and test setup complexity, the total number of devices in the simulations included 7.7k 1.8V $0.18\mu m$ transistors, 1.1k MiM capacitors and 0.7k polysilicon resistors. For the ENOB measurement, a 1.9Vpp amplitude sine-wave at a frequency of 7MHz was selected. The ADC clock speed was set at 100MHz as specified and an FFT was performed on 128 samples after discarding the first $0.7\mu s$ that are affected by settling.

Figures 2 and 3 show the distribution of the calibrated and uncalibrated ENOB for the Monte Carlo simulations with only process variations for the former and both process variations and mismatch for the latter. It can readily be observed how the introduction of mismatch variations increases the spread of the distributions, both for the calibrated and uncalibrated ENOB.

IV. Results

The first test scenario consists in predicting the uncalibrated ENOB in the case of a process-only Monte Carlo simulation.

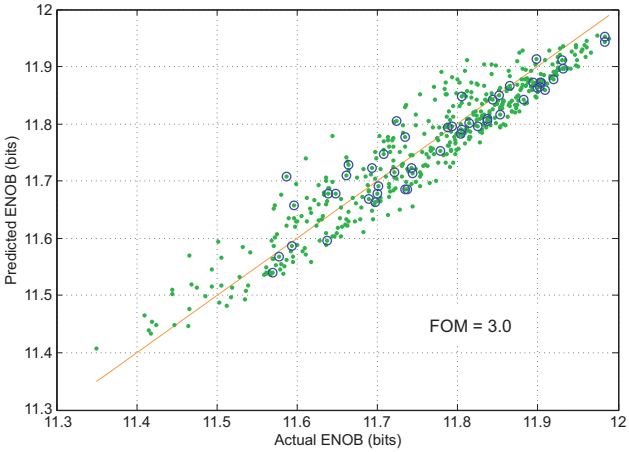

Fig. 4. Scatterplot of the predicted versus the true uncalibrated ENOB, for a simulation with process variation only

Figure 4 displays the obtained results: the green dots represent the full set of simulated circuits and the blue circles highlight those that have been used only in the test set to evaluate the generalization error of the model. In order to assess the quality of the obtained regression model, we use the Figure of Merit (FOM) defined in [17],

$$FOM = \sqrt{\frac{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2}{\sum_{i=1}^{N}\left(P_i^* - P_i\right)^2}} \tag{6}$$

where P_i and P_i^* are the real and predicted performance of device i, respectively and $\bar{P}$ is the average performance of all devices. The FOM is thus the standard deviation of the original cloud of point of the metric to be predicted (i.e. the uncalibrated ENOB in this case) divided by the generalization error. A FOM higher than 1 shows that the model brings more information than the naive predictor consisting in taking the expected (mean) value of the cloud of points. Here, a value of 3 is obtained which is quite high considering the small variation range of the uncalibrated ENOB (from 11.4 to 12 bits, with a standard deviation of 0.12bits).

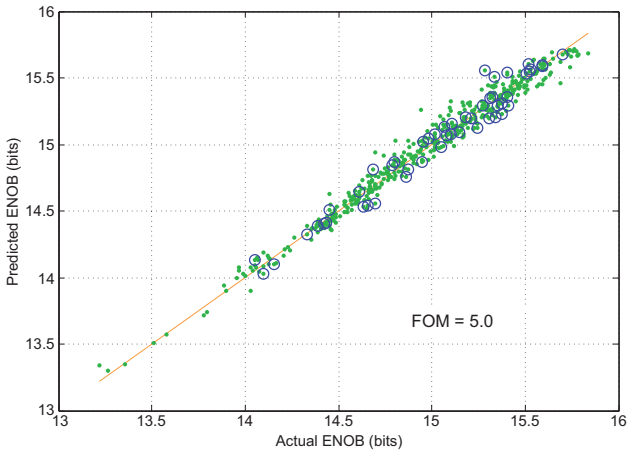

Fig. 5. Scatterplot of the predicted versus the true calibrated ENOB, for a simulation with process variation only

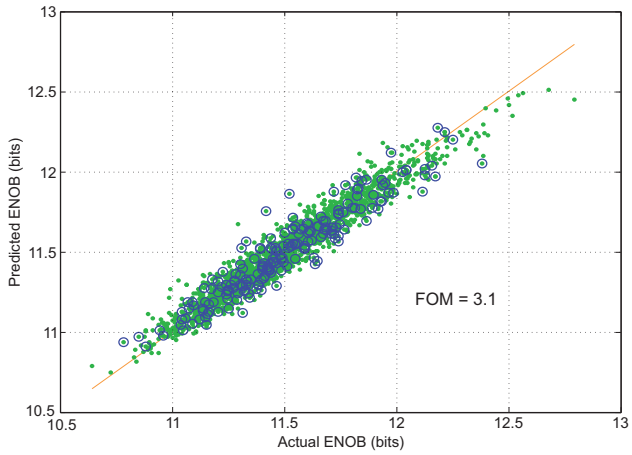

Fig. 6. Scatterplot of the predicted versus the true uncalibrated ENOB, for a simulation with both process and mismatch variations

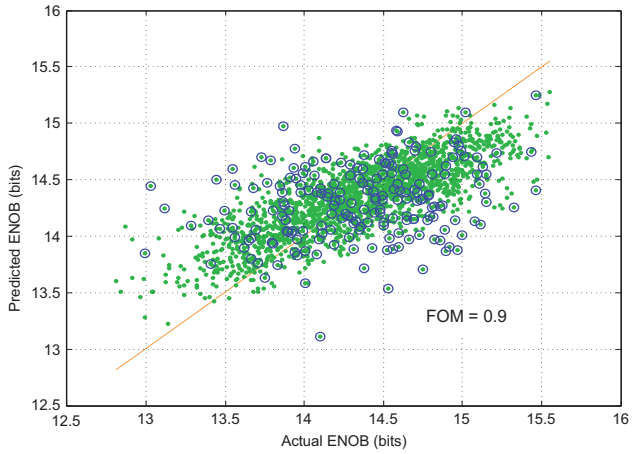

Fig. 7. Scatterplot of the predicted versus the true calibrated ENOB, for a simulation with both process and mismatch variations

The second scenario consists in predicting the calibrated ENOB for the same data, generated by a process-only Monte Carlo simulation. Figure 5 shows that impressive results can be obtained, with a FOM as large as 5. The generalization error on the prediction of the calibrated ENOB is only 0.08bits. This is a good result that is comparable to the reproducibility error of an FFT on 128 points.

We can thus assume that the calibrated ENOB variation is due to high-order phenomena that were masked by the first-order effect of the most significant stages static transfer function errors. Examples of phenomena that can affect the static transfer function and that are not calibrated out by the method in [15] are: non-linear settling (both op-amp and switches), amplifier non-linearity, insufficient calibration depth, etc.

Anyhow, stopping our analysis at that point may have led to the inadequate conclusion that this Alternate Test is a perfect substitute to the performance-driven test of a state-of-the-art pipeline ADC.

In order to make a faithful validation of the approach, we performed a Monte Carlo simulation of both process and mismatch variation. In a first place, Fig. 6 shows the prediction obtained for the uncalibrated ENOB. Similarly to the first scenario, a FOM of 3.1 is obtained which shows that a good precision can be reached.

However, the prediction of the calibrated ENOB shown in Fig. 7 is far from satisfying. The FOM of 0.9 actually tells us that it is almost useless. The prediction on the training set seems to point to a correlation but this is a known artifact due to the finite size of the training set. The generalization capability is seen on the blue circles (the validation set) and the obtained cloud of point shows no clear predictive power.

The calibrated ENOB is likely to be affected by global process parameters but, in the view of the obtained results, it seems that the dominant contribution to the calibrated ENOB comes from local mismatch parameters –information that is only partially present in the selected set of signatures. However, we could also hypothesize that it is the signatures themselves, and not the calibrated ENOB, that are severely affected by mismatch to the point of losing their correlation to global process variations, and hence losing all their predicting power. This is very unlikely, though, since the uncalibrated ENOB is well predicted both with and without mismatch variations. We can safely state that the uncalibrated ENOB is dominated by global process variations, so the information about global process variations must be present in the signatures and it is not obscured by mismatch.

In order to confirm that extent, we trained a new machine learning model including as new signatures the complete set of technological process parameters, that were saved during the Monte Carlo simulation. If the calibrated ENOB variability is due to global process variations, then such a model should show a good predicting power since the very root cause of these variations would be included (while the original signatures only contain indirect information). Nevertheless, Fig. 8 shows that no significant improvement is observed with respect to Fig. 7. So we can safely conclude that the variability of the calibrated ENOB is dominated by local mismatch.

Though we cannot draw a generic conclusion from this

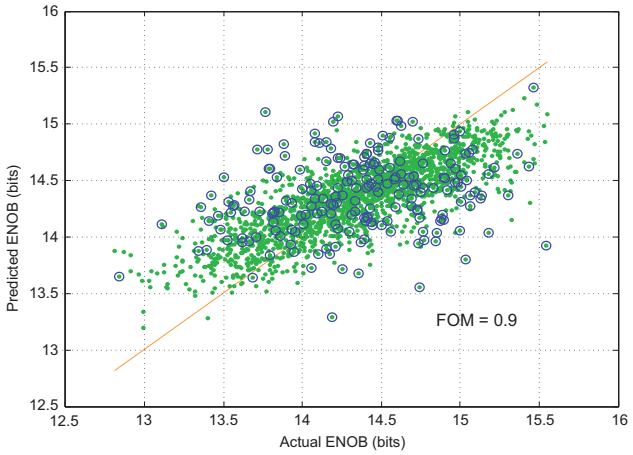

Fig. 8. Scatterplot of the predicted versus the true uncalibrated ENOB, using process parameters as additional signatures, for a simulation with both process and mismatch variations

single experiment, it actually makes much sense that calibrated systems handle the majority of performance deviations due to first order effects, both due to global and mismatch variations. Global variations usually have a stronger impact on circuits than local ones and are thus likely to be calibrated out, bringing local uncorrelated effects to the front of the causes of degradation.

A question obviously remains open: would it be possible to design some signatures that capture the impact of local variations so as to predict the calibrated ENOB? Our conjecture is that the set of signatures should share the same root causes of degradation as the ENOB. With local process variations, the dimensionality of the space of possible root causes literally explodes, which definitely makes a direct search an impractical approach.

V. Conclusions

In this paper, it has been shown that the performance of machine learning indirect test may be largely impacted by local process variations. In particular, it has been shown that the good prediction of the calibrated ENOB of a state-of-the-art pipeline ADC under global process variations is completely unrealistic. In the presence of local variations (which are unfortunately unavoidable), the calibrated ENOB prediction does not provide more insight than the naive predictor consisting in considering the expected mean value.

The usual assumption that local variations can be neglected should thus always be verified. The easiest way to proceed is to generate a small test set using a Process and Mismatch Monte Carlo simulation. This would not take into account the effects of mismatch on the model training (which is presumably equivalent to measurement noise) but would definitely rule out a strong sensitivity to local variations.

The conclusion of this paper should not be understood, though, as a criticism to the totality of machine learning indirect test. Indeed, it has also been observed that the low-cost signatures preserve the ability to predict the uncalibrated ENOB in the presence of mismatch. This obviously cannot replace the specification test, but these signatures could be good candidates for an efficient screening test. Future work will thus focus on validating an efficient defect-filter approach, as proposed in [18], based on the presented set of signatures.

Acknowledgment

This work has been partially funded by the Spanish Government project TEC2015-68448-R (co-funded by FEDER program) and supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir.

References

[1] P. N. Variyam and A. Chatterjee, "Enhancing test effectiveness for analog circuits using synthesized measurements," in *Proc. of the IEEE VLSI Test Symposium*, 1998, pp. 132–137.

[2] M. Barragan and G. Leger, "Efficient selection of signatures for analog/RF alternate test," in *Test Symposium (ETS), 2013 18th IEEE European*, May 2013, pp. 1–6.

[3] H. Ayari *et al.*, "Smart selection of indirect parameters for DC-based alternate RF IC testing," *VLSI Test Symposium*, pp. 19–24, 2012.

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

[5] M. Barragan and G. Leger, "A Procedure for Alternate Test Feature Design and Selection," *IEEE Design Test*, vol. 32, no. 1, pp. 18–25, Feb. 2015.

[6] H. G. Stratigopoulos, P. Faubet, Y. Courant, and F. Mohamed, "Multi-dimensional analog test metrics estimation using extreme value theory and statistical blockade," *Design Automation Conference (DAC), 2013 50th ACM / EDAC / IEEE*, pp. 1–7, May 29 2013-June 7 2013.

[7] E. Yilmaz and S. Ozev, "Adaptive-Learning-Based Importance Sampling for Analog Circuit DPPM Estimation," *IEEE Design Test*, vol. 32, no. 1, pp. 36–43, Feb. 2015.

[8] M. Barragan, H. G. Stratigopoulos, S. Mir, H. Le-Gall, N. Bhargava, and A. Bal, "Practical Simulation Flow for Evaluating Analog and Mixed-Signal Test Techniques," *IEEE Design & Test*, vol. 33, no. 6, pp. 46–54, 2016.

[9] G. N. Stenbakken, "Effects of nonmodel errors on model-based testing," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 384–388, 1996.

[10] A. Zjajo, M. J. Barragan, and J. P. de Gyvez, "Low-Power Die-Level Process Variation and Temperature Monitors for Yield Analysis and Optimization in Deep-Submicron CMOS," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2212–2221, August 2012.

[11] L. Abdallah, H. G. Stratigopoulos, S. Mir, and C. Kelma, "Experiences with non-intrusive sensors for RF built-in test," in *Proc. International Test Conference*, 2012.

[12] V. Natarajan, S. Sen, A. Banerjee, A. Chatterjee, G. Srinivasan, and F. Taenzler, "Analog Signature- Driven Postmanufacture Multidimensional Tuning of RF Systems," *Design & Test of Computers, IEEE*, vol. 27, no. 6, pp. 6–17, 2010.

[13] M. Andraud, H.-G. Stratigopoulos, and E. Simeu, "One-shot calibration of RF circuits based on non-intrusive sensors," in *Proc. Design Automation Conference (DAC)*. IEEE, 2014, pp. 1–6.

[14] G. Volanis, D. Maliuk, Y. Lu, K. S. Subramani, A. Antonopoulos, and Y. Makris, "On-die learning-based self-calibration of analog/RF ICs," in *Proc. VLSI Test Symposium (VTS)*. IEEE, 2016, pp. 1–6.

[15] A. N. Karanicolas, H.-S. Lee, and K. L. Barcrania, "A 15-b 1-Msample/s digitally self-calibrated pipeline ADC," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 12, pp. 1207–1215, 1993.

[16] J. D. Wichard, M. J. Ogorzalek, and C. Merkwirth, "Entool-a toolbox for ensemble modelling," in *Europhysics Conference Abstracts ECA*, vol. 27, 2003.

[17] M. J. Barragán *et al.*, "Alternate test of lnas through ensemble learning of on-chip digital envelope signatures," *Journal of Electronic Testing*, vol. 27, no. 3, pp. 277–288, 2011.

[18] H.-G. Stratigopoulos, S. Mir, E. Acar, and S. Ozev, "Defect filter for alternate RF test," in *Proc. European Test Symposium*, Sevilla, Spain, May 2009, pp. 101–106.