Enrichment of Limited Training Sets in Machine-Learning-Based Analog/RF Test

Haralampos-G. Stratigopoulos*, Salvador Mir* and Yiorgos Makris†
*TIMA Laboratory (CNRS-Grenoble INP-UJF), 46 Av. Félix Viallet, 38031 Grenoble, France
†Department of Electrical Engineering, Yale University, 10 Hillhouse Ave., New Haven, CT 06520, USA

***Abstract*— This paper discusses the generation of information-rich, arbitrarily-large synthetic data sets which can be used to (a) efficiently learn tests that correlate a set of low-cost measurements to a set of device performances and (b) grade such tests with parts per million (PPM) accuracy. This is achieved by sampling a non-parametric estimate of the joint probability density function of measurements and performances. Our case study is an ultra-high frequency receiver front-end and the focus of the paper is to learn the mapping between a low-cost test measurement pattern and a single pass/fail test decision which reflects compliance to all performances. The small fraction of devices for which such a test decision is prone to error are identified and retested through standard specification-based test. The mapping can be set to explore thoroughly the trade-off between test escapes, yield loss, and percentage of retested devices.**

I. Introduction

Testing the RF components of integrated circuits (ICs) in high-volume manufacturing results in added costs, which according to anecdotal evidence can amount up to 40%. This high test cost originates from the specialized test instrumentation and custom-made test strategies, the test complexity, the lengthy test application times, etc [1]. The test cost is ever increasing as more functionality and protocols are integrated into a single IC. This calls for immediate test solutions, which challenge the standard specification-based test approach in terms of incurring test cost, yet which meet the defective PPM goals aimed by industry.

To reduce test cost, it is evident that we need to rely on test measurements which are extracted rapidly on simple test configurations and which correlate well with the original specification-based tests. To this end, machine learning is a powerful approach to derive such correlations. For example, it has been shown that a committee of neural network classifiers can be trained with low-cost test measurements to predict the test decision of a full specification test set [2]. Furthermore, individual specified performances can be predicted by training regression functions to establish a quantitative relationship with the low-cost test measurements. This approach is commonly referred to as alternate test (see for example [3], [4], [5]). While the above results show great promise, the predictive accuracy has so far only been validated using small sets that typically comprise a few hundred devices. In fact, the accuracy of any predictive model, whether based on classifiers or regression functions, depends on the information richness of the training set which is used in modeling.

Ideally, we would like to have available a large training set from real devices that is collected across different lots during a long period of time, so as to be representative of the manufacturing process. In particular, we would like to have data from a large population of "critical" devices that fall marginally-in-the-specification and marginally-out-of-the-specification bounds. However, such data sets are typically not available during the test development phase. Brute-force Monte Carlo analysis is not a viable option either, since it samples only the statistically likely cases, thus it will hardly generate any "critical" devices in a few simulation runs [6].

In this work, the objective is to generate synthetic data, which, nevertheless, have the structure of true data. This is achieved by first fitting a non-parametric model to provide an estimate of the joint probability density function of test measurements and performances [7]. Then, the estimate is sampled to rapidly generate data corresponding to as many "critical", faulty, and good devices as necessary and, thereby, to create (a) information-rich training sets and (b) a separate (e.g. independent) arbitrarily-large validation set that can be used to estimate test metrics of interest, such as yield loss and test escapes, in PPM accuracy.

In this paper, we focus on the classification-based test approach [2]. Next, we provide a brief overview of this approach and we discuss the shortcomings of having small and unbalanced training sets. In section III, we revisit the non-parametric density estimation technique in [7]. In section IV, we present the device under test and the experimental data set. In section V, this data set is enriched to improve training and to estimate test metrics in PPM accuracy.

II. Machine-Learning-Based Test

The approach consists of two linked steps. In the first step, a low-cost alternate test configuration is identified which gives rise to an alternate measurement pattern. The term "alternate" serves to indicate that this measurement pattern differs from the standard measurements obtained in specification-based test. In the second step, a predictive model is learned that implicitly maps this alternate measurement pattern to one of three possible classifications: the device is good, the device is faulty or a decision based solely on this alternate test is prone to error. In the latter scenario, the device is retested through the standard specification-based test approach. The two steps are linked in the sense that the input stimuli involved in the alternate test configuration and the output alternate

978-3-9810801-5-5/DATE09 © 2009 EDAA

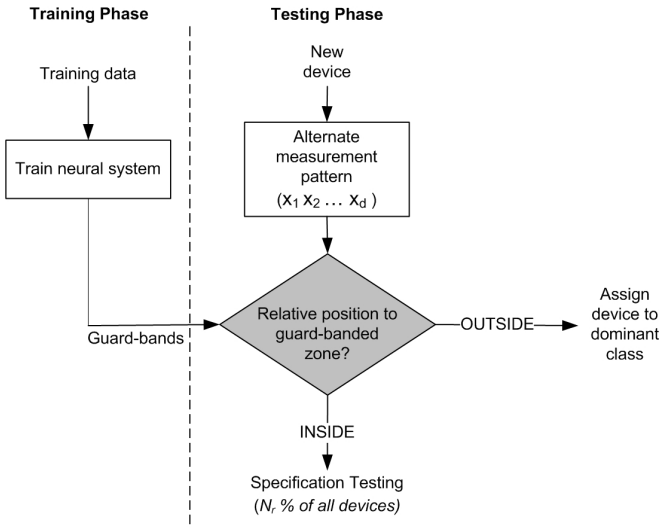

Fig. 1. Test flow after successful training.

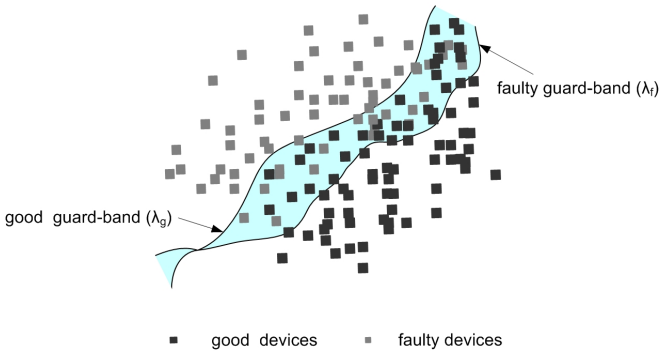

Fig. 2. Guard-bands placement in an alternate measurement space.

measurements are optimized jointly to achieve a minimum error in future predictions. The test cost savings stem from the fact that the majority of the devices are classified directly using the alternate measurement pattern.

The predictive model consists of a committee of two ontogenic neural networks [2], which are trained to allocate guard-bands in the alternate measurement space, as shown in the left-hand part of Fig. 1 and in Fig. 2. The training phase makes use of training data from a population of device instances. These data include the alternate measurement pattern and the performance values of each instance. The performance values are compared to the specifications in the data-sheet to label each instance as good or faulty. Fig. 2 illustrates the projection of both good and faulty instances in an imaginary 2-dimensional alternate measurement space. Training aims to allocate two guard-bands as shown in Fig. 2: the good guard-band that "guards" the good population, i.e. it has all good training instances on one side, and the faulty guard-band that "guards" the faulty population, respectively. In essence, the guard-bands create a trichotomy in the alternate measurement space: two regions outside the guard-banded zone that are dominated by either good or faulty training instances and the guard-banded zone which contains a mixed population.

In the testing phase, the guard-bands are used to test new devices that come out of the fabrication line, as shown in the right-hand part of Fig. 1. Instead of carrying out all specification tests, one needs only to obtain the alternate measurement pattern and examine the position of its footprint with respect to the guard-banded zone that has been learned during the training phase. If the footprint falls outside the zone, then the device is assigned to the dominant class, otherwise, if it falls within the zone, the device is considered to have ambivalent status. These ambivalent devices ($N_r\%$ of all devices) are forwarded to standard specification-based testing, in order to reach a final accurate decision.

Predicting the label of new devices through the guard-bands entails two types of errors: the pattern of a faulty device might be interwoven in the "good" region giving rise to test escapes, or the pattern of a good device might be interwoven in the "faulty" region giving rise to yield loss. It is evident that the wider the guard-banded zone is, i.e. the more the guard-bands are pushed away of each other into the "clean" regions, the less the prediction error and, of course, the larger the percentage N_r of retested devicess. Therefore, there is a trade-off between the prediction error, denoted hereafter by ϵ_r, and the percentage N_r. The trade-off can be explored by two parameters, namely λ_g and λ_f, which define the position of the good and faulty guard-band, respectively [2]. When both are zero, the two guard-bands collapse to a single classification boundary (i.e. $N_r = 0$). When λ_g (λ_f) increases, then the good (faulty) guard-band is pushed into the "faulty" ("good") region, thus widening the guard-banded zone and decreasing yield loss (test escape).

It is evident that the prediction accuracy heavily depends on the information that is available during the training phase. Indeed, the training set should include a balanced population between faulty and good devices such that one population does not overshadow the other (statistically speaking the last changes of the position of the guard-bands during training will be instigated by misclassified devices that belong to the class with the larger number of samples). Second, the training devices must be representative of the manufacturing process such that they "fill up" the alternate measurement space. This is necessary because when the pattern of a new device falls in a subspace which was empty of patterns during training, then the device will be assigned a random pass, fail or retest label (the guard-bands have been curved randomly in this subspace due to the lack of instances). Most importantly, the training set must contain "critical" devices that are marginally-in-the-specification and marginally-out-of-the-specification bounds such that the true separation boundaries are approximated. The problem lies in that such training sets with numerous faulty and "critical" devices are not readily available during the test development phase.

To this end, in the following section, we discuss a method to generate synthetic data that respect the distribution of experimental data. The resulting enhanced data set contains all the relevant information needed for training and, in addition, its volume can be arbitrarily large, allowing us to express the prediction error in PPM accuracy.

III. Synthetic Data Generation

The underlying idea is to estimate the joint probability density function, $f(\mathbf{x})$, of the random vector $\mathbf{x} =$ [performances, alternate measurement pattern] and, subsequently, sample the estimate to construct a sequence of independent observations, $\mathbf{y}_k$, from it. Each observation $\mathbf{y}_k$ corresponds to a new device instance.

Specifically, we will use the non-parametric density estimation technique that we have applied in [7] in the context of evaluating a built-in self-test technique. The density estimate is given by

$$\tilde{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K_e \left(\frac{1}{h} (\mathbf{x} - \mathbf{X}_i) \right), \tag{1}$$

where $\mathbf{X}_i$ is an observation of $\mathbf{x}$, $i = 1, .., n$, d is the dimensionality of $\mathbf{x}$, h is a parameter called bandwidth, and K_e is the Epanechnikov kernel

$$K_e(\mathbf{t}) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2) \left(1 - \mathbf{t}^T \mathbf{t}\right) & \text{if } \mathbf{t}^T \mathbf{t} < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where $c_d = 2\pi^{d/2} / (d \cdot \Gamma(d/2))$ is the volume of the unit d-dimensional sphere. The bandwidth is chosen using the rule of thumb [8]:

$$h = \left\{ 8 c_d^{-1} (d+4) (2\sqrt{\pi})^d \right\}^{1/(d+4)} n^{-1/(d+4)}. \tag{3}$$

Details about the consistency of the estimate $\tilde{f}(\mathbf{x})$ at a single point (i.e. convergence of $\tilde{f}(\mathbf{x})$ to the true density $f(\mathbf{x})$ in probability as $n \to \infty$) and the uniform consistency (i.e. convergence of $\sup |\tilde{f}(\mathbf{x}) - f(\mathbf{x})|$ to zero as $n \to \infty$) can be found in [8].

Notice that the estimate in (1) is expressed as a sum of bumps centered at the observations $\mathbf{X}_i$. The kernel defines the shape of the bump and the bandwidth defines its half-width. Thus, a sample $\mathbf{y}_k$ from $\tilde{f}(\mathbf{x})$ can be generated by sampling a vector $\mathbf{t}$ from the probability density function $K_e(\mathbf{t})$ and then transforming $\mathbf{t}$ back to $\mathbf{y}_k$ for a random $\mathbf{X}_i$:

Step 1 Choose I uniformly with replacement from $\{1, ..., n\}$.

Step 2 Generate $\mathbf{t}$ to have probability density function $K_e(\mathbf{t})$.

Step 3 Set $\mathbf{y}_k = \mathbf{X}_I + h\mathbf{t}$.

In order to sample $K_e(\mathbf{t})$, we use the acceptance-rejection method [9]. In particular, let $U(\mathbf{t})$ be the probability density function of the uniform distribution in $[-1, 1]^d$ and notice that $K_e(\mathbf{t}) \leq c \cdot U(\mathbf{t})$, $c = c_d^{-1}(d+2)/2$, $\forall \mathbf{t} \in R^d$. The acceptance-rejection method is as follows:

Step 2a Generate $\mathbf{t}$ from U.

Step 2b Generate u from a uniform distribution in $[0, 1]$.

Step 2c If $c \cdot u \leq K_e(\mathbf{t})$ accept and return $\mathbf{t}$, otherwise return to step 2a.

To construct the training set, we are interested in sampling "critical" and faulty instances. The performance values in $\mathbf{y}_k$ can be compared to the specifications in the data-sheet to conclude whether this observation corresponds to a "critical" or faulty instance. Sampling can be repeated as often as necessary, in order to generate the required $\mathbf{y}_k$.

Performances	Mean	StdDev
LNA NF (dB)	2.422	0.235
LNA IIP_3 (dBm)	0.483	0.523
LNA GAIN (dB)	12.5	0.187
CAS IIP_3 (dBm)	-8.481	0.724
CAS GAIN (dB)	20.155	0.425
CAS NF (dB)	6.55	0.264
MIX NF (dB)	13.478	0.153
MIX IIP_3 (dBm)	2.956	0.741
MIX GAIN (dB)	10.278	0.258
MIX ISWR (dB)	7.094	0.306
LNA ISWR (dB)	1.696	0.060
LNA OSWR (dB)	2.357	0.074
LNA RevIso (dB)	18.672	0.845

TABLE I

Statistics of the available data set.

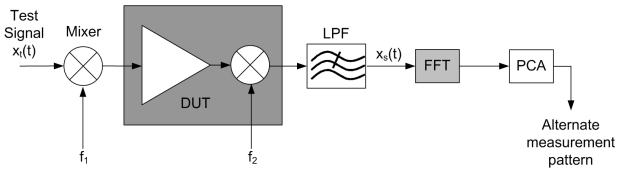

Fig. 3. Single alternate test configuration for the ultra-high frequency receiver front-end [4].

IV. Case Study

The device under test is an ultra-high frequency receiver front-end, which comprises a dual-band low noise amplifier (LNA) and a balanced mixer. For the sake of simplicity, only one band (e.g. 850 MHz) is considered, resulting in the 13 performances listed in Table I. Testing the compliance of these performances to the corresponding specifications through the standard specification-based test approach requires a total of 7 test configurations that involve different test equipment and pin connections.

The selected single alternate test configuration is shown in Fig. 3. The tester provides a test stimulus $x_t(t)$ that consists of 7 tones with a step of 1 MHz and a central frequency of 177 MHz. The amplitudes of the tones range between -10.5 dBmb and -20 dBm. The test stimulus is first up-converted to 850 MHz using an external mixer placed on the load board. The up-converted signal excites the LNA, its response is then down-converted to a baseband signal of frequency 50 MHz using the mixer of the device, passed through a low-pass filter, and logged by a sampling scope. Since the mixers are driven by unsynchronized free running local oscillators, the effect of their phase difference is removed by computing a 1024-point FFT of the output $x_s(t)$. We consider 28 output tones whose amplitudes are above a certain noise floor. Subsequently, we perform a principal component analysis (PCA) to reduce the dimensionality while retaining 99.5% of the total variation in the data. This results in a 3-dimensional alternate measurement pattern $[x_1, x_2, x_3]$. In comparison to the 7 test configurations used in the standard specification-based testing, this single test

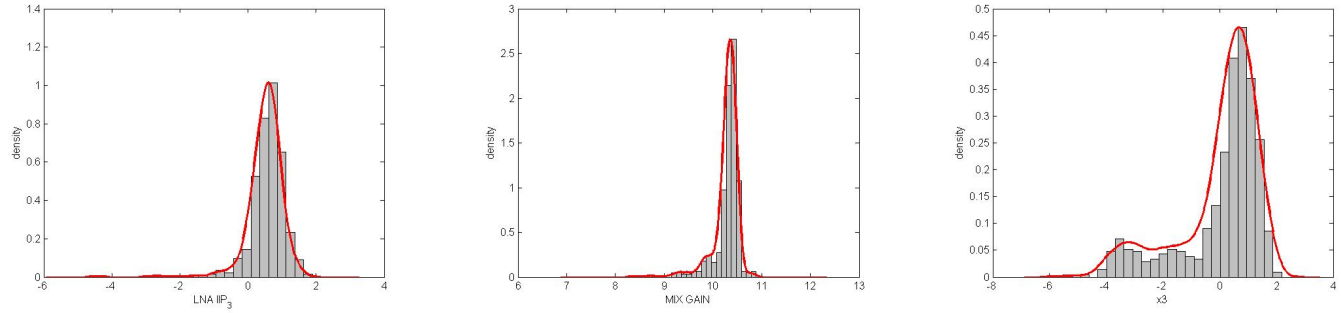

Fig. 4. 1-dimensional marginals of the estimated non-parametric density.

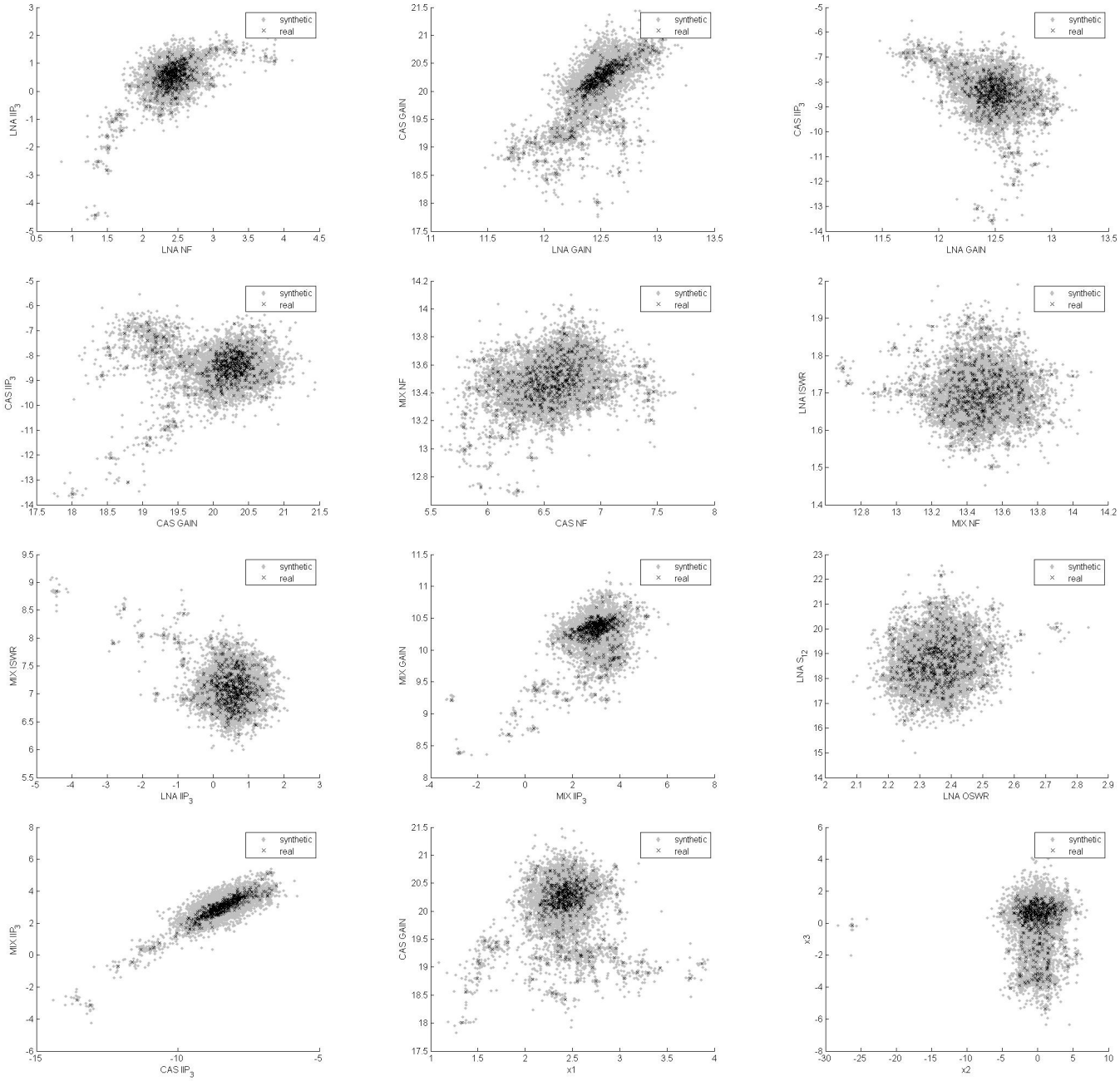

Fig. 5. Real and synthetic devices projected in 2-dimensional spaces.

configuration offers an overall reduction of 36% in testing time while the needed test instrumentation costs 48% less.

We use a real data set which contains the values of the 13 performances and the 3-dimensional alternate measurement pattern for a total of 541 devices. The experiment to obtain this data set is explained in detail in [4]. The specification limits are set to Mean$\pm$7·StdDev (+ or -, depending on whether the performance has a lower or upper limit). The values of Mean and StdDev are computed across the set of 541 devices and are listed in Table I. For example, the passing limits for the LNA IIP$_3$ are $[0.483 - 7 \cdot 0.523, +Inf] = [-3.178, +Inf]$. Notice that all 541 devices pass the specification tests.

V. Results

A. Synthetic Data Generation

The sample data from $n = 541$ devices form the initial observations $\mathbf{X}_i$ which are used to estimate the density in (1). The dimensionality is $d = 13 + 3 = 16$. Fig. 4 shows example 1-dimensional marginals of the estimated density. The figure also illustrates the scaled histograms which are plotted using the 541 devices. It can be deduced that marginals are far from being normal. The advantage of the non-parametric density estimation technique is that it does not make any assumption regarding the true parametric form (i.e. normal, log-normal, gamma, etc.). Instead, it allows the data to speak for themselves. As can be seen, the densities fit the histograms very well.

Once the density is estimated, it can be readily sampled using the algorithm of section III to obtain new instances. 1 million new instances can be obtained approximately in 30 minutes using Matlab [1] on an Intel(R) Core(TM)2 2.40-GHz PC. The defect level is 3795ppm. The example scatter plots of Fig. 5 project the real 541 devices together with 10^4 randomly generated synthetic devices in 2-dimensional spaces formed by pairs of performances and alternate test measurements. As can be observed, the distribution of synthetic devices respects the structure of the real distribution very well. Fig. 5 illustrates the full potential of the method to generate synthetic data that are, in practice, indistinguishable from real data.

B. Machine-learning-based test

1) Training: To train the good and faulty guard-bands we use a sample set that comprises $3 \cdot 10^4$ synthetic devices, of which 1/3 are "critical", 1/3 faulty, and 1/3 good. The "critical" devices marginally pass the specification test: they have at least one performance which falls within the 5·StdDev and 7·StdDev specification limits. The faulty devices violate at least one specification and, statistically, they fall close to the 7·StdDev specification limit, i.e. they marginally fail one or more specifications. The good devices are statistically an image of the real distribution. Their performances fall actually far from the specification limits, nevertheless they are used in training, in order to retain a population of devices that

[1]MATLAB is a trademark of the MathWorks Inc., http://www.mathworks.com.

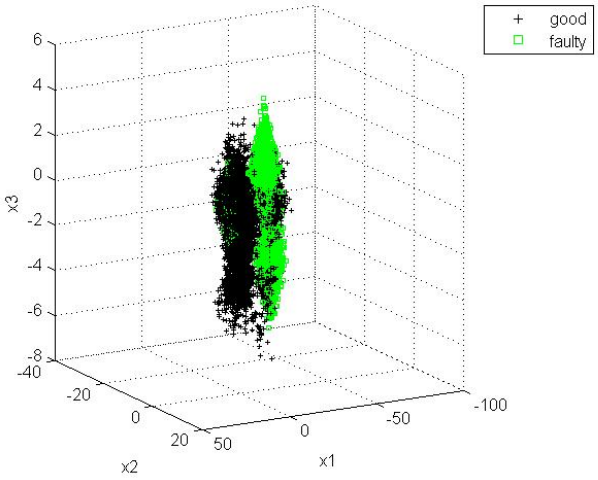

Fig. 6. Synthetic devices projected in the 3-dimensional alternate measurement space. A good discrimination is observed.

spans the manufacturing process. Fig. 6 projects the training devices in the 3-dimensional alternate measurement space (i.e. the top 3 principal components). Note that the above set consists of devices with process variations due to drifts in the manufacturing process. Devices containing catastrophic faults are not considered during training since a catastrophic fault is a random event: a basic principle of machine learning is that an unknown dependency can be learned only when the data is drawn from a fixed distribution $f(\mathbf{x})$. However, if the guard-bands correctly classify marginally-out-of-the-specifications devices, then, intuitively, they will also be capable of correctly classifying devices with catastrophic faults since it is expected that their footprints in the alternate measurement space are scattered far-off from the core of devices with nominal process variations.

2) Validation: Training is repeated for different values of the parameters λ_g and λ_f, in order to examine different placements of the guard-bands and, thereby, different widths of the guard-banded zone. As explained in section II, this allows us to explore a trade-off between test escapes, yield loss, and percentage of retested devices.

The estimated density is sampled 10^6 times to predict test escapes (T_E) and yield loss (Y_L) with PPM accuracy using relative frequencies:

$$T_E = P\left(C_g^c | T_p\right) = \frac{P\left(C_g^c \bigcap T_p\right)}{P\left(T_p\right)} \approx \frac{N_{C_g^c \bigcap T_p}}{N_{T_p}} \qquad (4)$$

$$Y_L = P\left(T_p^c | C_g\right) = \frac{P\left(C_g \bigcap T_p^c\right)}{P\left(C_g\right)} \approx \frac{N_{C_g \bigcap T_p^c}}{N_{C_g}}, \qquad (5)$$

where C_g is the event that a device is good and T_p is the event that a device passes the test. The complementary events, C_g^c and T_p^c, correspond to a device being faulty and a device failing the test, respectively.

Table II shows the results in % (i.e. one needs to multiply by 10^4 to obtain the values in PPM). The third column shows

λ_g	λ_f	ϵ_r (%)	N_r (%)	T_E (%)	Y_L (%)
0	0	3.054	0	0.012	3.054
0	0.5	3.053	0.105	0.011	3.054
0	1	3.028	0.95	0.008	3.032
0	1.5	3.043	15.929	0.002	3.053
0	2	3.043	22.481	0.001	3.054
0.5	0	2.111	0.98	0.012	2.107
0.5	0.5	2.114	1.077	0.011	2.111
0.5	1	2.099	1.902	0.008	2.099
0.5	1.5	2.123	16.862	0.002	2.13
0.5	2	2.129	23.404	0.001	2.136
1	0	0.708	2.517	0.011	0.699
1	0.5	0.707	2.621	0.011	0.699
1	1	0.705	3.422	0.008	0.699
1	1.5	0.699	18.443	0.002	0.699
1	2	0.698	24.998	0.001	0.699
1.5	0	0.011	3.411	0.011	0
1.5	0.5	0.011	3.515	0.011	0
1.5	1	0.008	4.316	0.008	0
1.5	1.5	0.002	19.338	0.002	0
1.5	2	0.001	25.892	0.001	0

TABLE II

PREDICTION ERROR, RETESTED DEVICES, TEST ESCAPES, AND YIELD LOSS AS A FUNCTION OF THE POSITION OF THE GUARD-BANDS.

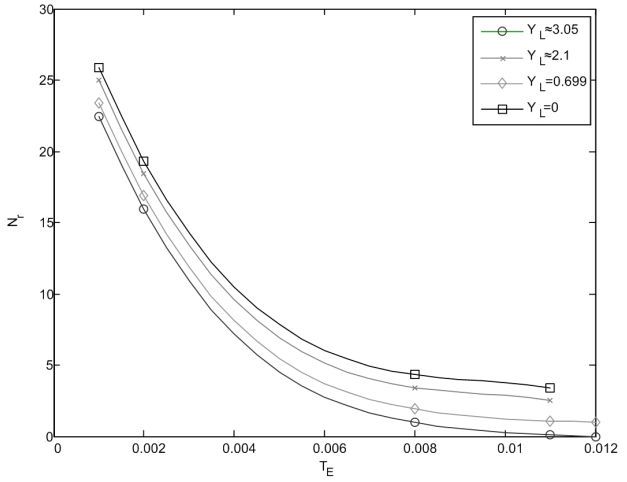

Fig. 7. Test escape as a function of retested devices for constant yield loss (all three metrics are expressed in %).

the overall prediction error, i.e. the percentage of devices that are misclassified. Fig. 7 plots N_r versus T_E for approximately constant Y_L. It can be seen that, for constant λ_g, T_E drops as λ_f increases. For constant λ_f, Y_L drops as λ_g increases. N_r increases as λ_g and/or λ_f increase. One can identify interesting trade-off points, for example $T_E = 110$ppm and $Y_L = 0$ppm if we choose to retest $N_r = 3.4\%$ of the devices through specification-based test. An estimate of test cost savings can be found using the following simplified model:

$$C = C_a T_a + N_r C_s T_s, \tag{6}$$

where C_i denotes test cost per second, T_i denotes test time, and subscripts "a" and "s" correspond to the alternate test measurements and the standard specification-based test, respectively. Since $C_a T_a = (1 - 0.48)(1 - 0.36) C_s T_s = 0.333 C_s T_s$, equation (6) becomes $C = (0.333 + N_r)\, C_s T_s$, which yields $C = 0.367 C_s T_s$ for $N_r = 3.4\%$.

VI. CONCLUSIONS

Machine-learning-based test of analog/RF circuits offers a low-cost alternative to standard specification-based test. However, to date it has not been possible to corroborate the claim that the two approaches are equivalent in terms of test error. To this end, we presented a technique to (a) improve the training phase of learning machines and, thereby, to ameliorate their ability for future predictions, and (b) estimate the correctness of the predictions, in terms of resulting test escape and yield loss, with PPM accuracy. The key benefit of this technique is that it permits us to obtain a good estimate of the prediction accuracy at an early stage of the test development phase. Thus, it can also be used to choose among an array of different alternate test configurations and measurement patterns. In this paper, the technique was employed to train a committee of neural network classifiers to explore the trade-off among test cost, test escapes, and yield loss for a receiver front-end device.

ACKNOWLEDGMENTS

The authors would like to thank José Torres and Tim Swettlen from Intel Corporation for providing the ultra-high frequency receiver front-end data.

This research was supported by the Commission of European Communities (MIRG-CT-2007-209653), the National Science Foundation (NSF ECS-0622081 & CCF-0702522), and the Semiconductor Research Corporation (SRC 1632).

REFERENCES

[1] A. Abdennadher and S. A. Shaikh, "Practices in mixed-signal and RF IC testing," *IEEE Design & Test of Computers*, vol. 24, no. 4, pp. 332–339, 2007.

[2] H.-G. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine-learning-based Analog/RF testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339–351, 2008.

[3] S. Cherubal, R. Voorakaranam, A. Chatterjee, J. Mclaughlin, J. L. Smith, and D. M. Majernik, "Concurrent RF test using optimized modulated RF stimuli," in *IEEE International Conference on VLSI Design*, 2004, pp. 1017–1022.

[4] S. S. Akbay, J. L. Torres, J. M. Rumer, A. Chatterjee, and J. Amtsfield, "Alternate test of RF front ends with IP constraints: Frequency domain test generation and validation," in *IEEE International Test Conference*, 2006, pp. 4.4.1–4.4.10.

[5] R. Voorakaranam, S. S. Akbay, S. Bhattacharya, S. Cherubal, and A. Chatterjee, "Signature testing of analog and RF circuits: Algorithms and methodology," *IEEE Transactions on Circuits and Systems - I*, vol. 54, no. 5, pp. 1018–1031, 2007.

[6] A. Singhee, J. Wang, B. H. Calhoun, and R. A. Rutenbar, "Recursive statistical blockade: An enhanced technique for rare event simulation with application to SRAM circuit design," in *IEEE International Conference on VLSI Design*, 2008, pp. 131–136.

[7] H.-G. Stratigopoulos, J. Tongbong, and S. Mir, "A general method to evaluate RF BIST techniques based on non-parametric density estimation," in *Design, Automation and Test in Europe*, 2008, pp. 68–73.

[8] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, 1986.

[9] J. E. Gentle, *Random Number Generation and Monte Carlo Methods*, Springer, 2nd edition, 2004.