Modeling the Dependency of Analog Circuit Performance Parameters on Manufacturing Process Variations with Applications in Sensitivity Analysis and Yield Prediction

Elena-Diana Şandru, *Student Member, IEEE*, Emilian David, Ingrid Kovacs, Andi Buzo, Corneliu Burileanu, *Life Senior Member IEEE*, and Georg Pelz

***Abstract*— There is a consistent dependence between integrated circuits performance parameters and manufacturing process variations and capturing it at an early development phase represents a major ongoing topic in the semiconductor industry. Typically, this is addressed by the means of Monte Carlo simulations, where the device model parameters are randomly instantiated according to the technology variations based on a pre-defined nominal process. Thus, the resulted simulation data can only capture the effect of these variations. This offers little or no insight on the performance's sensitivities to specific process variations or on the effect of altered statistical technology properties, as it may be the case of process drift or fab-to-fab migration. This paper proposes a methodology for modeling the dependency of the device performances (i.e. Electrical Parameters - EPs) with the influential technology parameters (i.e. Process Control Monitor parameters - PCMs), at an early stage (pre-Silicon). Using a set of standard Monte Carlo co-simulations of PCM structures and the circuit schematics (to maintain consistent process variation), it employs a feature selection step to choose the influential PCMs and it trains a Machine Learning regression algorithm. Both are wrapped up in a Bayesian Optimization framework to find the optimal feature set and the regression hyperparameters. The obtained regression model can explain the functional dependency of the EP on the influential PCMs. Thus, it directly enables sensitivity analysis to process variation and parametric yield prediction of the integrated circuit, as it will be illustrated for the case of an experimental Infineon Technologies product.**

***Index Terms* — Integrated circuits yield, yield prediction, Machine Learning, Process Control Monitor (PCM), sensitivity analysis**

I. Introduction

As the Integrated Circuits (ICs) complexity is growing exponentially, ensuring a high production yield has become a major challenge for the semiconductor manufacturers. Managing the inherent manufacturing process variation, defined as the deviation of parameters from their nominal specifications [1], represents one of the critical success factors in achieving a high yield. Nonetheless, due to Moore's law, as well as the More-than-Moore technologies development [2], the chip performance became more sensitive to the smallest variation of the process. Process variations can be illustrated as uncertainties linked to the electrical parameters of ICs. In order to keep it under control, the process variation is constantly being supervised during the manufacturing process, using Process Control Monitor (PCM) parameters or technology parameters.

In recent years, the PCM parameters proved their value to the semiconductor industry in terms of controlling and monitoring the manufacturing process, from a statistical point of view. The PCM structures (defined as narrow electrical structures located on the wafer among productive dies, used to measure the PCM parameters) are manufactured together with the production chips; therefore, they undergo the same technological steps. Consequently, the PCM parameters can be used in alternative ways (not only for monitoring), as follows: yield estimation and prediction [3]-[5], yield detractors detection [6], defects detectability [7], final test outcome prediction [8], etc.

Linking the circuit performance parameters (also known as Electrical Parameters – EPs) and the process variation (monitored through the technology parameters PCMs) may address the following problems:

1) Problem 1 – The approximation effect of the device models: In time, the pre-Silicon (preSi) yield estimation methodologies have been improved. Yet, it is common for a passed circuit in preSi to fail after the Design of Experiments (DoE) analysis. One of the many reasons is the approximating effects of the process parameters' distribution, as for complex technologies it is very difficult to fit with maximum accuracy a covariance matrix with tens of parameters [9]. In addition, it is not an easy

Elena-Diana Şandru is with the Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest, Romania and also with Infineon Technologies Romania & CO SCS, 20337 Bucharest, Romania (e-mail: diana.sandru@upb.ro).

Corneliu Burileanu is with the Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest, Romania (e-mail: corneliu.burileanu@upb.ro).

Ingrid Kovacs is with the Faculty of Electronics, Telecommunications and Information Technology, Technical University of Cluj-Napoca, Romania (e-mail: ingrid.kovacs@staff.utcluj.ro).

Emilian David is with Infineon Technologies Romania & CO SCS, 20337 Bucharest, Romania (e-mail: emilianconstantin.david@infineon.com), Andi Buzo and Georg Pelz are with Infineon Technologies AG, 85579 Neubiberg, Germany (e-mails: andi.buzo@infineon.com and georg.pelz@infineon.com).

task to state upfront the impact of a small inaccuracy in PCM modeling on the circuit performances, mainly because the same technology is employed for a very wide range of products. Therefore, even an approximation of more than 92% can have an unpredictable impact.

2) Problem 2 – Fab-to-fab migration: When migrating a product from one fabrication plant to another, the circuit performances might deteriorate, although the same technology is used. This may be caused by apparently small differences between the two fabrication processes. It is desirable and valuable to provide a solution capable of estimating upfront the impact of the fab-to-fab migration.

3) Problem 3 – Technology-specific DoE plan: Often, during the DoE test plan, the process handling is quasi-standard and does not consider all the product's particularities. In other words, the process may not always be varied in the most unfavorable corners for the considered product. Thus, the most vulnerable circuit performances with process variation must be identified, as well as the PCMs that impact them and cause out-of-specs issues. Moreover, each circuit performance (EP) is influenced by a limited set of PCM parameters, possibly unknown in advance to its full extent.

4) Problem 4 – Root cause analysis: When the yield problems are observed, the root cause analysis is performed, in order to determine whether the technological process or the design is responsible. In the case of small and dedicated to non-critical applications chips/products, there are no dedicated structures for each chip. Instead, there are few PCM structures per wafer (some of the times only 5), compared to thousands of productive dies. Furthermore, the number of wafers produced for testing purposes, before the actual production (DoE analysis) is limited (25-50), due to cost considerations. Hence, modeling the technological impact on purely measurement basis (measured PCMs) on the circuit performances (EPs) as:

$$\mathrm{EP} = f(\mathbf{PCM}) \quad (1),$$

would involve sub-optimal sampling of PCMs, leading to low accuracy. To obtain a proper function, a 1-to-1 correspondence between the EP and the PCMs samples is necessary, in order to avoid the resampling errors.

Moreover, all of the above-mentioned problems are observed and handled during the post-Silicon (postSi) phase and naturally induce production costs. Consequently, the solution might consist in properly and accurately model (1), by exploring the relationship between the EPs and the PCMs, at an early stage, during the preSi verification, as it is highly desirable to get an early overview of the impact of the technology variations on the circuit performances. Taking all of these into account, in this paper we address this demand by developing a methodology for modeling the relationship between the EPs and the PCMs, at an early stage, using a ML algorithm on preSi data, for ICs without dedicated PCM structures on every productive die. This implies co-simulating the models of the analog circuit along with the PCM structures schematics, provided that the accuracy of the device models is ensured. Therefore, we obtain a direct regression model to estimate the impact of the technology parameters on the circuit parameters, expressed as a mathematical relationship, which can also conditionally map the EP distribution dependency on the PCM distribution. Besides that, our method can highlight if the defined set of PCMs can accurately explain a certain EP.

Our approach for modeling (1) is not a trivial solution due to several reasons; firstly, the technology parameters' dataset is quite large and there is no a priori information regarding which one affects the behavior of a certain circuit EPs, thus leading to the need for an efficient feature selection step in order to eliminate the non-impacting PCMs. Secondly, very often the circuit performance parameters cannot be entirely explained by the PCMs; this calls for an adequate and optimized ML algorithm able to extract the maximum information out of the existent PCMs. We aim to overcome these problems through a novel solution, that performs the feature selection and the ML regression training adaptively using an optimization algorithm (Bayesian Optimization), employed to find the combination of PCMs and algorithm hyperparameters able to minimize the model error. Nevertheless, using the preSi simulation data will provide us with more control and flexibility to vary all the elements in the available process window and discover all the process variation, as well as ensuring a 1-to-1 correspondence between the performance parameters and the manufacturing process variation. This represents an advantage over using postSi measurements, as most of the time, the number of PCM structures is much smaller compared to the productive dies; yet, it must be mentioned that there are ICs produced with embedded PCM structures [10].

The automated methodology can be further used to perform sensitivity analysis and yield prediction, as it only requires the preSi simulations (for sensitivity analysis) and additional PCMs postSi measurements related to the targeted process window (for yield prediction). In particular, the previously-presented problems 1 and 2 employ the proposed framework for parametric yield prediction, while for problems 3 and 4, the methodology is used for sensitivity analysis to process variation. More specific, after obtaining the best ML regression model, the sensitivity analysis to process variation represents a direct benefit of the methodology, as it is provided directly by the regression training procedure. The set of resulted PCMs are the influencing factors, while the metric used for their selection quantifies their influence strength. As for the parametric yield, it can be accurately estimated even in the cases where the PCMs' distributions vary from the nominal case, because instead of directly modeling the yield (which can bring in an arbitrary character, as the specifications can be easily modified), we model the distribution of the EPs. Thus, the yield results from integrating this distribution over the specification limits. It is important to highlight that the proposed methodology can be applied only for parametric yield prediction, when the chip is functionally correct and the yield loss is caused by the failure of meeting some performance specifications criteria. The ICs total yield loss can occur due to several other causes (e.g. packaging problems, design error, process contamination, or even random defects) that fall into catastrophic yield loss category.

To prove its accuracy, the proposed preSi-based fitted function is validated using postSi data (measured chip parameters and PCM parameters), thus displaying a 92-95%

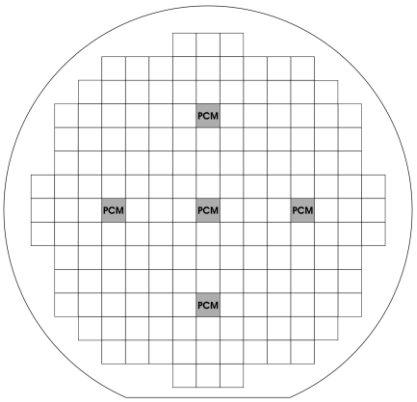


Fig. 1. Wafer placement of 5 PCM structures among the productive dies.

prediction accuracy. The methodology is applied on an experimental product designed by Infineon Technologies, along with the two use cases proposed for it – sensitivity analysis to process variation and parametric yield prediction.

The remainder of this paper is organized as follows: the related work is reviewed in Section II, along with our motivation. In section III, the problem is formulated mathematically and further, the proposed methodology for modeling the functional and statistical dependencies between chip and technology parameters is highlighted in section IV. Section V presents the experimental results obtained by applying the methodology on Low-Dropout Voltage Regulator parameters, along with its validation and the two use cases for the proposed methodology – sensitivity analysis and yield prediction. Finally, the conclusions are drawn in Section VI.

II. Related Work

A. Circuit Verification and PCM Parameters

Whether we are discussing preSi verification or postSi validation, their main goal is to verify the correctness of the circuit and confirm that the fabricated product meets the desired requirements. On one hand, preSi verification techniques are more varied and mature, but most of the time, they provide insufficient coverage, making it impossible to remove all the bugs from the design stage [11] – practically, the design is tested in virtual environments but corner cases are prone not to be tested enough. Basically, the preSi stage lacks the parasitic effects and the freedom of capturing the variations in an acceptable amount of time. On the other hand, postSi validation aims at certifying the correct behavior of the produced chip over a series of specified operating conditions [14]; the chip must be in full and perfect compliance with its specifications. However, compared to the preSi case, it lacks when it comes to controllability, observability and easy debugging. Bridging the two phases is still an ongoing topic [11]-[13].

Since the production of ICs is a complex process, the Statistical Process Control (SPC) was introduced, in order to detect, as early as possible, the process variation. PCM parameters are used to characterize and control the technology in reference to the technology specification [14]. As presented in Section I, the PCM test structures are located beside the chip area on the wafer at specific locations (as it can be observed in Fig. 1), but they undergo the same technology steps in order to verify the wafer from the technologist point of view. The PCM structures are dispersed across the entire wafer, offering a spatial overview of the process variation. Consequently, the PCM parameters proved to be useful not only in SPC, and few steps were made towards introducing them in the preSi verification analysis, by simulating their equivalent circuit schematics [15], [16].

B. Sensitivity Analysis

The sensitivity analysis (SA) concentrates on identifying the subset of input variables of a multivariate system that displays the strongest impact on its outputs. The sensitivity analysis methodologies can be classified into two major approaches – variable-based methods where the factors' influence on the output are quantified directly using correlation-like metrics; model-based methods where the functional dependencies between input factors and output are derived and the model is further used to quantify the influence.

Out of the first category, the variance-based methods [17]-[19] (Sobol's indices, Jansen method, FAST, EFAST) are the most popular, since they are able to compute main effects and higher order effects, but they impose a certain design of experiments, that is a limitation and can also be computationally expensive. Similarly, the entropy-based methods do not require a specific experimental setup [20], but they depend upon a significant number of runs compared to the number of input variables (at least 20 times higher). Model-based methods also proved to be a great alternative for costly functions, when it may be hard to run sensitivity analysis methods directly on them [21].

Traditionally and to the best of our knowledge, the sensitivity analysis of ICs with respect to the process variations is based either on prior knowledge of the circuit designer or on first order correlations. Our approach aims at quantifying this at a very early stage, offering the designer a glimpse into the future postSi dependencies.

C. Parametric Yield Prediction

Parametric yield prediction often implies forecasting by generating samples in preSi stage, using Monte Carlo (MC) or its enhancements variants simulations and further applying failure count methods [22]-[24] (simulate-and-count approaches); nevertheless, in order to lower the variance in estimation, at least hundreds of thousands of samples are required, thus becoming computationally expensive.

Introducing diversified measures of the process variation in the yield prediction process and combining it with ML algorithms has demonstrated even greater results in terms of accuracy and computational effectiveness [4], [25]-[28]. In [4], a new methodology entirely on silicon measurements (e-tests performed on PCMs structures and probe-tests performed on dies) is introduced to predict yield for production migration of products from one facility to another and transition to a new design generation. Nevertheless, it involves the need of high-volume silicon data, which may not be available for every manufacturer. In [26], the authors take advantage of the wafer-level spatial correlations between e-tests measurements and the

probe-test measurements (equivalent to the EPs) to predict probe-test measurements for the remaining die locations on the wafer, using Gaussian Process-based regressors. The methodology is useful when few wafers are available, based on spatiotemporal progressive sampling, but due to GP it can be computationally expensive. On the other hand, we have proposed in [28] a proof-of-concept preliminary idea of yield prediction based on the relationship between EP and PCMs, for normal distributions.

Consequently, predicting the yield based on the relationship between ICs and the process variation represents a very important state-of-the-art ongoing topic and to the best of our knowledge there are no studies involving parametric yield prediction based on postSi PCMs values. More efforts should be invested towards minimizing the costs associated with the high-volume measured data.

III. Problem Formulation

Our problem is related to the general ML regression framework [29], [30]; considering that the PCM parameters properties are reflecting process variations, our hypothesis is that they are correlated with circuit performances and this relationship can be tackled as early as possible, in preSi stage, thus serving as basis for yield prediction or sensitivity analysis.

Given a set of N data points pairs $\{(pcm_1, ep_1), \ldots, (pcm_N, ep_N)\}$ provided by circuit simulations, as observations of two random variables, a vector of influential PCMs, $\boldsymbol{PCM} \in \mathbb{R}^D$ and a scalar $EP \in \mathbb{R}$, we want to find the model function that accurately maps the two variables as:

$$EP = f(\boldsymbol{PCM}) + \varepsilon \quad (2),$$

where ε is a zero mean error independent of $\boldsymbol{PCM}$ and minimized in least square sense [29]. For our problem, the EP represents the electrical parameter and $\boldsymbol{PCM}$ represents the influential PCMs vector, both characterizing an IC device. Each element d is statistically dependent with EP and satisfies the following relation:

$$P(PCM_d, EP) \neq P(PCM_d)P(EP) \quad (3),$$

where $P(PCM_d, EP)$ stands for the joint probability density function of the two variables, while $P(PCM_d)$ and $P(EP)$ are the corresponding marginal probability densities. By influential PCMs we understand variables belonging only to a subset of all available PCM through simulation.

Consequently, we want to obtain an accurate model $f(\boldsymbol{PCM})$, that expresses the dependency between EP and the subset of influential PCMs. This will provide us the information regarding the dependence (sensitivity) of the electrical parameters (and implicitly of the IC device) on the technology parameters. By using some common probabilities properties [31], the EP distribution can be further modeled as a conditional dependence on the technology parameter distribution, as follows:

$$P(EP) = \int P(EP, \boldsymbol{PCM})\, d\boldsymbol{PCM} = \int P(EP \mid \boldsymbol{PCM})\, P(\boldsymbol{PCM})\, d\boldsymbol{PCM} \quad (4)$$

By knowing the specification limits of the EP parameter, we can also derive the parametric yield, defined as the probability of the variable of being within the specification limits *LSL* and *USL*, as follows:

$$yield = \int_{LSL}^{USL} P(EP) dEP \quad (5)$$

The regression model expressed in (2) provides us the conditional probability density function $P(EP|\boldsymbol{PCM})$ as an estimate [29], while $P(\boldsymbol{PCM})$ can also be estimated using the available samples. Nevertheless, without simplifying the assumptions on the functions form and on the distributions involved, (5) may not be analytically tractable in order to provide a convenient closed form solution.

A possible way to overcome this inconvenience is to make use of MC integration/sampling techniques [31], [32], by randomly generating samples according to the modeled distributions PCM and ε to generate samples of EP using (2). The percentage of EP samples that fall into the specification's limits will provide a proper estimate of (5), having the accuracy depending on the involved model's accuracy and the number of generated samples. This procedure is also known as the counting technique.

An algorithm providing a solution to estimate (5) will be presented in section IV-F; it is worth noting that (5) allows us to estimate the distribution of the EP and consequently the yield, even when the PCMs distribution differs from the simulation case. The only condition is that simulation variables ranges to provide coverage for the new distribution variables; consequently, this specifically allows yield prediction in the case of technological parameters drift of fab-to-fab migration.

IV. The Proposed Methodology

In this section, we will present in detail the automated methodology for modeling the relationship between the design parameters (EPs) and the technology parameters (PCMs) that capture the process variation. Moreover, we will provide insight and motivation of the chosen metrics and algorithms involved in the entire process.

A. The Methodology Flow

The goal of the proposed methodology is to express an optimal relationship of the form (1) through a ML model, by only using preSi Monte Carlo simulation and no additional prior knowledge regarding the possible dependencies between the variables. The data result by co-simulating the circuit and PCM structures schematics. The simulations are conducted in such way that they reflect the postSi measurements procedures; the operating conditions (e.g. temperature, voltage, load) are varied in a similar manner to the silicon measurement procedure.

A schematic representation of our proposed methodology is presented in fig. 2. Firstly, the input of the methodology, represented by the entire set of PCMs and the EP for which the analysis is conducted, is split into train and test subsets. Then, the PCMs undergo a feature selection step where PCMs variables are selected according to their correlation metric value with the EP. In this way, only those which exceed a certain threshold are further included in the analysis. Next, training set of the influential PCMs (as input) and the EP (as output) is fed into the ML regression algorithm, and a prediction model is

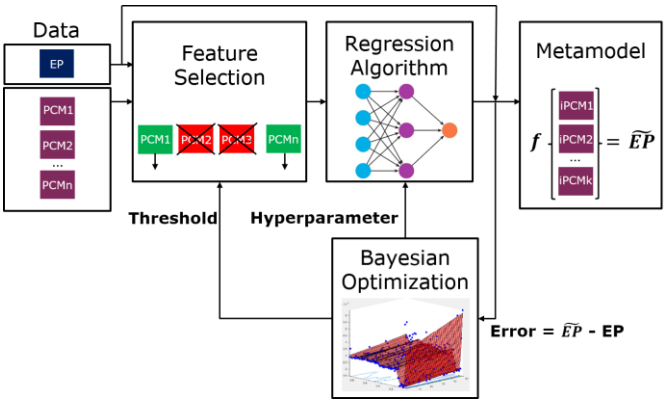


Fig. 2. Schematic representation of the proposed methodology for modeling the relationship between the circuit performance parameters (EP) and the process variation (monitored through the PCMs).

fitted. The Bayesian Optimization (BO) framework controls both the above steps. The BO optimally selects the regression hyperparameter and the correlation metric threshold, in an iterative manner, with the goal of minimizing the model error computed on the test set. The final optimization outcome is the regression model ("*metamodel*"), that incorporates the influential PCMs subset and has the optimal hyperparameters with respect to the test set error. In our previous work [33], we have presented a reliable proof-of-concept methodology for unified feature selection and hyperparameter BO for ML-based regression.

B. Model Improvement through Feature Selection

Typically, in order to capture most of the process variation, many PCMs are introduced and measured; however, previous studies of the authors showed that not all of them influence a particular EP. Consequently, introducing many input variables that do not influence the output in the regression algorithm might affect the accuracy of the *metamodel*. Similarly, adding PCMs that do not explain the EP when building the regression model should only add computational costs, with no benefits. Deciding on a PCM subset that certainly influences and explains an EP is a feature selection problem that is well covered by the literature [34] and a common practice in the semiconductor industry [35]. Using feature selection when training any ML algorithm helps in obtaining better learning accuracy with less computational cost.

In this paper, we employed a correlation metric-based feature selection method. For comparison, we propose two correlation metrics: Distance Correlation [36] and Maximal Information Coefficient [37]. Their performance for our problem will be quantified, as they represent very recent advances in the field and both are claimed as being among the best metrics for quantifying the dependence between any two random variables (PCMs and EPs in our case).

However, as it generally happens with any metric, it is almost impossible to specify in advance the value of their threshold that clearly indicate the (in)dependence of any possible two variables. Nevertheless, we overcome this disadvantage by using the Bayesian Optimization that optimally finds this threshold with respect to the regression error.

1) (Brownian) Distance Correlation – DistCorr

The distance correlation ($\mathcal{R}(X,\ Y)$) measures the degree of independence between two variables [36]. The empirical sample $\mathcal{R}_n^2(X,\ Y)$ is calculated as presented in (6), with values in [0, 1]:

$$\mathcal{R}_n^2(X,\ Y) = \frac{\nu_n^2(X,Y)}{\sqrt{\nu_n^2(X)\nu_n^2(Y)}} \quad (4),$$

where $\nu_n^2(X,\ Y) = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}B_{kl}$, namely the distance covariance, while $A_{kl} = a_{kl} - \overline{a_{k.}} - \overline{a_{l.}} + \overline{a_{..}}$, computed as: $a_{kl} = X_k - X_l$, $\overline{a_{k.}} = \frac{1}{n}\sum_{l=1}^{n} a_{kl}$, $\overline{a_{..}} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$ and $\overline{a_{l.}} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}$.

In a similar manner, B_{kl} is computed as A_{kl} with the difference that involves variable Y instead of X, while $\nu_n^2(X) = \nu_n^2(X,\ X)$ and $\nu_n^2(Y) = \nu_n^2(Y,Y)$ are called distance variances.

One of the most important distance correlation's properties is that $R^2(X,\ Y) = 0$ if and only if the variables are independent, i.e. $p(X,Y) - p(X)p(Y) = 0$. This metric proved to be consistent against all types of dependent alternatives with finite second moments [36]. Some of its advantages include: robustness to noise, low computational cost, and sensitivity to non-linear dependencies. *DistCorr*, as any other metric, depends on the variables' number of samples and on the form of their dependence; in our experience, a 0.1 value indicates a noticeable dependence between two variables sampled for at least few hundreds of times.

2) Maximal Information Coefficient – MIC

The Maximal Information coefficient (*MIC*) is a mutual information-based metric that measures the degree of correlation between two variables. It is based on the idea that a drawn grid of the scatter plot of two variables [37] can encapsulate the relationship between the variables.

$$MIC = max\left\{\frac{I(X,Y)_{|n_X,n_Y}}{log_2\ min\{n_X,n_Y\}}\right\} \quad (5)$$

where $I(X,Y)_{|n_X,n_Y}$ is the mutual information computed on each grid partitions of the two axes, determined by n_X and n_Y:

$$I(X,Y)_{|n_X,n_Y} = H(X)_{|n_X,n_Y} + H(Y)_{|n_X,n_Y} - H(X,Y)_{|n_X,n_Y} \quad (6)$$

Considering that $n_X * n_Y < B$, where B is a function of sample size, while $p(X_i)$ and $p(Y_j)$ are marginal distributions and $p(X_i,Y_j)$ is the joint distribution, (8) becomes:

$$I(X,Y)_{|n_X,n_Y} = \sum_{i=1}^{n_X} p(X_i)\ log_2\frac{1}{p(X_i)} + \sum_{j=1}^{n_Y} p(Y_j)\ log_2\frac{1}{p(Y_j)} - \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} p(X_i,Y_j)\ log_2\frac{1}{p(X_i,Y_j)} \quad (7)$$

Same as *DistCorr*, *MIC* takes values between zero and one, and it has two main properties: equitability and generality. The first one refers to obtaining the same values for the metric, when applying it to equally noisy functions, while the second one involves capturing a wide range of functional relationships between the two variables, given a sufficiently large dataset. The advantages over mutual information estimation were presented in [38]. Similarly, *MIC* has the ability to cope with nonlinear relationships between the two variables. An open-source MATLAB and Python implementation is available in [39].

C. Regression Models

The main goal of this research is to estimate the dependency between EP (treated as a measure of circuit design) and PCMs, under a certain accuracy. This can be regarded as a typical supervised learning problem – given a set of input features (the PCMs) and the desired output (the EP), discover the relationship between them. Several regression algorithms can be employed for this particular example: Support Vector Machine (SVM), Gaussian Process (GP), several types of regressions (Lasso, logistic), and Neural Networks (NN) [30].

In this paper, we use Multilayer Perceptron (MLP) Neural Network (NN), which is a feedforward NN. In our previous experiences, SVM and GP proved to be reliable candidates as well; nevertheless, tuning their hyperparameters led to an increased training time and computational cost when compared to NNs. The decision revolved around a compromise between a slightly better accuracy (SVM and GP) and training time (NN); we decided to use NN as they proved to be up to an order of magnitude faster. Our implementation uses Levenberg-Marquardt (LM) backpropagation algorithm to modify the weights on each neuron. The LM algorithm solves nonlinear least-squares problems and is presented as an interpolation between Newton's method (characterized by a rapid coverage of the minimum, but may encounter divergence issues) and Gradient descent (characterized by assured convergence, but slower) [30]. Consequently, LM represents the fastest backpropagation algorithm (due to its rapid convergence), thus becoming our top choice for the proposed methodology.

We build NN *metamodels* using PCMs as input variables and an EP at a time as outcome variable. The employed NN consists of three layers neurons: an input layer, a hidden layer and an output layer.

D. Bayesian Optimization

Among others, ML performances depend on their hyperparameters choice; choosing them optimally is not a trivial task, even for experienced practitioners. Besides grid and random search [40] (employing too many iterations to cover the entire variables space), there are optimization algorithms that iteratively use a surrogate function in order to map the error dependence on the hyperparameters and acquisition function in order to assess the most promising next point in this space to be evaluated. These algorithms can still be slow when many variables are involved, but they are still faster than classical grid when dealing with expensive-to-evaluate functions.

Fundamentally, the BO framework [41] is a sequential sampling-based optimization solution, that finds the minimum of multimodal black-box function defined over a bounded domain:

$$f(x): X \rightarrow \mathbb{R},\ X \subseteq \mathbb{R}^n,$$
$$x_m = \underset{x \in X}{argmin} f(x) \tag{10}$$

BO involves two different components: first, the probabilistic surrogate function, consisting of a prior distribution that captures beliefs about the behavior of the objective function (function evaluations data) and the observation model (posterior distribution), used to describe the manner of data generation. Second, the acquisition function, formulated based on the posterior distribution used to guide the search, by indicating the next query point. For the surrogate function, besides the parametric models (Beta-Bernoulli, Linear, Generalized Linear) with limited applicability, there are the non-parametric approaches [41]. Flexible and powerful, they marginalize away the weights in the Bayesian linear regression, by applying a kernel based on the Gaussian Process and its derivatives [41]. As for the acquisition functions, there are several strategies to use the posterior model for guiding the sequential search, that must be chosen based on the particularities of the problem; there is no strategy to clearly outperform the others.

Algorithm 1: Proposed Methodology

Require: EP, $\boldsymbol{PCM}$, $maxNN$, $itNumber$

1: Split train-test (EP_{train}, EP_{test}, $\mathbf{PCM}_{train}$, $\mathbf{PCM}_{train}$)

2: $FSmetrics = DistCorr(\mathrm{EP}_{train}, \mathbf{PCM}_{train})$ (v1)
$FSmetrics = MIC(\mathrm{EP}_{train}, \mathbf{PCM}_{train})$ (v2)

3: $neuronsNumber \leftarrow optimizableVariable([2, maxNN])$

4: $FSthreshold \leftarrow optimizableVariable(\begin{bmatrix}\boldsymbol{min}(FSmetrics),\dots \\ \dots \boldsymbol{max}(FSmetrics)\end{bmatrix})$

5: $\boldsymbol{OptResults} = BayesianOptimization(\mathrm{EP}_{train}, \mathrm{EP}_{test}, \mathbf{PCM}_{train}, \dots$
$\dots \mathbf{PCM}_{test}, FSmetrics, neuronsNumber, FSthrehold, itNumber)$

6: $\mathbf{PCM}_i = \mathbf{PCM}(FSmetrics \geq \boldsymbol{OptResults}.FSthreshold)$

7: $metamodel =$
$NNtrain(\mathrm{EP}_{train}, \mathbf{PCM}_{i-train}, \boldsymbol{OptResults}.neuronsNumber)$

We use BO in our methodology to properly select the number of neurons of the MLP NN hidden layer. In addition, we employ this framework to decide on the threshold value of the feature selection metric, in order to dynamically choose the subset of influential PCMs. Our implementation is derivative free and uses Gaussian Process as a surrogate function, due to its inherent ability to cope with the stochastic nature of the train and test errors of NNs, as minimizing their cost function involving the regression error is a non-convex optimization problem. Furthermore, it employs *expected-improvement-plus* function as the acquisition function, as it is able to return to space exploration when a region becomes over-exploited, thus avoiding local optimum. The optimization algorithm attempts at finding the minimum value of the test error function (the *metamodel* prediction error computed on the test set).

E. Summary

The detailed procedure of our methodology using the elements presented in sub-sections IV-B-D is given in Algorithm 1. Both the values of the EPs and the PCMs are of several different orders of magnitude; therefore, our implementation requires data normalization. For this, the Min-Max feature scaling is employed, resulting a [-1,1] interval:

$$x' = 2\frac{x - min(x)}{max(x) - min(x)} - 1 \tag{11}$$

The proposed methodology requires setting the maximum number of neurons for the hidden layer – $maxNN$, in advance. In order to find the best *metamodel* associated for each EP in what regards its generalization capability, the input dataset is split into train and test, by randomly sampling the entire distribution (line 1). This way, the BO aims at iteratively minimizing the regression error computed on the test set for a specified number of iterations ($itNumber$) by adaptively

adjusting the neurons number and correlations thresholds (line 5). For this, the *DistCorr* or *MIC* coefficients (depending on the version) are computed between the studied EP and each PCM (line 2), and two optimizable variables are declared (lines 3 and 4). The final *metamodel*, incorporating the functional dependency between the analyzed EP and the influential technology parameters, is re-trained based on the optimal hyperparameters that result after final iteration (lines 6 and 7).

Compared to other optimization methods, we consider the choice of BO framework to be optimal, due to its ability to cope with the stochastic nature of NNs training error. To the best of our knowledge, there are no similar approaches found in the literature, as highlighted in [42]. Consequently, employing BO in jointly optimizing feature selection problems that involve state-of-the-art feature selection metrics (*DistCorr* or *MIC*), represents a natural choice. Nonetheless, there are also other approaches that address very similar semiconductor industry problems. In [37], the authors introduce a feature selection method that minimizes the regression model training error residuals by iteratively including more and more features, by employing the *DistCorr* metric, but they do not address the NN hyperparameters optimization.

F. Use Case 1 – Parametric Yield Prediction

Our literature survey indicates that the use of PCM parameters improves the yield prediction performances. Our approach for yield prediction is based on modeling the EP conditional distribution on the PCMs distribution using preSi data through a ML-based regression. Thus, we are able to accurately model the marginal EP distribution given any PCM distribution, especially in the case when this distribution is different than the one obtained through MC simulation. Consequently, the yield prediction using the modeled marginal EP distribution will automatically inherit its accuracy. The presented concept relies on the common approach for production yield computing based on counting; the measured value of each EP (on every chip) is checked against its specification limits:

$$LSL \leq \mathrm{EP} \leq USL \tag{12},$$

where *LSL* and *USL* symbolize the lower and the upper specification limits of the EP. The parametric yield is further estimated by the out-of-specs (OOS) metric, expressed as follows:

$$OOS_{\mathrm{EP}} = \frac{no.of\ samples\ out\ of\ spec}{total\ no.of\ samples} \tag{13}$$

As highlighted in section II-C, the use of MC simulations for yield estimation is an expensive method, even under enhanced forms. In the same time, OOS implies a high estimation variance and it can be regarded as a reliable metric only if a large number of samples is available. Our concept surmounts this issue by artificially generating as many PCM samples as needed, based on their distribution, having the same statistical properties as the simulated samples. Then, instead of simulating the circuit responses, namely the EPs, we estimate their values based on the previously trained *metamodels*. Finally, the estimated EPs values are verified against the specification limits.

Each *metamodel* encapsulates the subset of PCMs that influence the studied EP. Ideally, the fitting error is zero and

Algorithm 2: Parametric Yield Prediction

Require: $\widetilde{\mathrm{EP}} = f(\mathbf{PCM}_i) + \varepsilon$, $\mathbf{PCM}_i$, ε, LSL_{EP}, USL_{EP}

1: $\varepsilon_{gen} \sim distFit(\varepsilon)$

2: $\mathbf{PCM}_{i-gen} \sim distFit(\mathbf{PCM}_i)$

3: $\widetilde{\mathrm{EP}} = f(\mathbf{PCM}_{i-gen}) + \varepsilon_{gen}$

4: Count $OOS_{\widetilde{\mathrm{EP}}}$

$\widetilde{\mathrm{EP}} = \mathrm{EP}$. Nevertheless, all *metamodels* have fitting errors, mostly because the available PCMs cannot entirely explain the EP's behavior. The EPs are also influenced by other factors (like operating conditions) that were intentionally not introduced in the regression model. Their effects are included in the *metamodels*' fitting errors. Therefore, the EP is expressed as a superposition between a deterministic component expressed as a PCM function and a stochastic one represented as a sample drawn from the modeled error distribution:

$$\widetilde{\mathrm{EP}} = f(\mathbf{PCM}_{\mathrm{i}}) + \varepsilon \tag{14}$$

The detailed procedure for parametric yield prediction case is given in Algorithm 2, for a single modeled EP variable. The multivariate N-dimensional case is treated similarly, by expressing (15) in vector form:

$$\widetilde{\mathbf{EP}} = \begin{bmatrix} f(\mathbf{PCM}_{i1}) \\ \dots \\ f(\mathbf{PCM}_{i\mathrm{N}}) \end{bmatrix} + \varepsilon \tag{15}$$

As prerequisites, the EP *metamodel* is needed, under the form of (14), along with the available data for the influential PCM set for the analyzed design parameter – $\mathbf{PCM}_i$, the *metamodel* fitting error distribution ε and the parameter specification limits *LSL* and *USL*.

In order to model the fitting error, we apply the *distFit* method presented in [44], a multivariate distribution fitting approach that can be further used for generating data according to the modeled distribution. The *distFit* method verifies whether the fitting error displays a Gaussian distribution or not, by employing the Anderson-Darling normality test. As seen in [45] Anderson-Darling is comparable to Shapiro-Wilk in terms of power, for more than 500 samples distributions. If the test is positive, it models the fitting error as a Gaussian distribution; otherwise, it models the fitting error as a non-normal distribution.

The same approach can be applied to model any PCM (preSi or postSi) distribution and artificially generate as many samples necessary that resemble the original PCM statistic properties. For the most of the cases, the PCMs are characterized by normal distributions; nevertheless, in order to artificially generate the required samples for yield prediction, we considered *distFit* method, as it can cope with a broader range of distributions (line 2). The corresponding EP values are estimated based on the *metamodel*, the synthetic generated PCMs values and the modeled fitting error, resulting $\widetilde{\mathrm{EP}}$. Finally, the OOS metric is computed for the studied EP based on (12) and (13) (line 4).

In conclusion, the yield prediction procedure outlined above works as follows: given the *metamodels* obtained using the simulation data and the corresponding fitting error model, any set of PCMs can be used to estimate the EPs and the yield, respectively. By any set of PCMs, we refer (but are not limited)

to: preSi simulations, postSi measurements samples or their artificially generated counterparts according to their distribution model. It is important to highlight that one should expect a high relative variance in the yield prediction when the specifications fall on low EP distribution probability regions (i.e. distribution tails). This is due to the limited number of preSi simulations or postSi measurements that may not produce enough samples in order to densely cover these tails; moreover, it will reflect a high relative estimation variance on the tails of any distribution modeled using limited data.

In addition, the constructed *metamodels* will exhibit a low relative accuracy for variables belonging to these low probability tails. While there are workarounds to avoid this issue, they fall into the importance sampling category [43], these are beyond the scope of this paper, as it concentrates on modelling the overall EP conditional distributions on PCMs distributions and the yield prediction is considered an immediate benefit. If higher yield accuracy is necessary when very low probability distribution tails are involved, the prediction methodology remains the same. However, supplementary effort is required to ensure the models accuracy in those areas.

G. Use Case 2 – Sensitivity Analysis to Process Variation

Our methodology is able to reveal and model the dependency between ICs performance and process variation, under the form of a *metamodel*. The captured relationships can be further used to assess the sensitivity analysis problem, based on the manner in which the PCMs (input variables) are influencing the EP of interest (output variable). More precisely, in our case, the sensitivity of the EP towards the PCMs is straightforward, since it can be quantified directly using the feature selection metrics, that can be regarded as inherent variable-based methods. The sensitivity analysis to process variation is derived promptly from the methodology, since the EP is expressed as a function of influential PCMs, as follows:

$$\widehat{\mathrm{EP}} = f(\mathbf{PCM}_i) \tag{16}$$

The entire training optimization procedure only includes the influential PCMs as they provide the best dependency *metamodel* with the minimum training error. Consequently, the sensitivity measures, even independently of the *metamodel*, are validated by its accuracy. Finally, the sensitivity index, for a considered *EP* , based on *DistCorr* or *MIC*, can be expressed as:

$$S_j = DistCorr(\mathrm{EP}, \mathrm{PCM}_{i-j}) \tag{17}$$

$$S_j = MIC(\mathrm{EP}, \mathrm{PCM}_{i-j}) \tag{18}$$

where PCM_{i-j} denotes the j^{th} influential PCM (PCM_i) selected by the methodology. Nevertheless, further sensitivity analysis methods can be applied on the *metamodel*, but the ranking provided by the above-discussed correlation metrics was sufficient for our problem.

V. Experimental Results

A. Case Study and Dataset

In order to experimentally evaluate the methodology for modeling the relationship between the circuit performance parameters (EPs) and the process variation (monitored through PCMs), along with the two proposed use-cases (sensitivity analysis and yield prediction), we use a dataset from an Infineon Technologies experimental design of a Low Dropout Voltage Regulator (LDO).

For this purpose, we have co-simulated a setup that includes both the circuit model and the PCM structures schematics, by using an Infineon proprietary simulation tool. This way, when the process is modified, the same process variation is reflected in both EPs and PCMs measurements. The setup measures 92 PCMs and 3 important EPs (due to the limitations of the industrial setup), through a classical scheme of 1000 random MC simulations. The obtained dataset will be further referred to as $\mathrm{EP}_{k-preSi}$ (for $k = [1,\dots,3]$) and $\mathrm{PCM}_{j-preSi}$ (for $j = [1,\dots,92]$). It is worth mentioning that the nominal circuit's model reflects the behavior of the circuit to external stimuli after some experimental results, and it aims to fit device level parameters such that the preSi PCM distributions are as close as possible to postSi distributions. In order to achieve this, the methodology from [9] was used, able to ensure a 92% accuracy. Despite the high accuracy, when several devices are connected in a circuit, the impact of this discrepancy on the EP can be attenuated or amplified.

As presented in sections IV-A and IV-D, the optimization for finding the best regression model associated to each EP aims at minimizing the generalization error (evaluated on a test set); thus, the initial preSi dataset was split into an 80-20 percentage train and test independent datasets - $n_{train} = 800$ samples and $n_{test} = 200$ samples. In order to assess the methodology accuracy, we validate it using the product postSi dataset, consisting of two lots totalizing $w_{postSi} = 44$ wafers, referred as $\mathrm{EP}_{k-postSi}$ (for $k = [1,\dots,3]$) and $\mathrm{PCM}_{j-postSi}$ (for $j = [1,\dots,92]$). Each wafer has 5 PCM structures and approximately 3750 dies per wafer. The considered EPs are some typical LDO's wafer measurable performances (threshold voltages for enabling and reset functionalities) that were not entirely confined by the specification limits, while the PCMs cannot be disclosed in this paper. The specification limits are also available; hence, we can compute the parametric yield of each EP, as well as the overall yield of the product. Nevertheless, in all the presented experiments, besides the MC simulation data, no a priori information regarding the dependency between EPs and PCMs was used. In the same time, the circuit designer validated our conclusions.

B. Regression Fitting

In order to quantify the methodology results, as proposed in section IV-A, we performed three independent experiments. One for each feature selection metric aimed to improve the model accuracy (experiments 2&3), using the metrics discussed in section IV-B, and one without employing the feature selection, to be considered the benchmark (experiment 1). After the data preprocessing phase, described in the above sub-section (normalization, train-test split), for each of the studied EP, the corresponding *metamodel* is trained using the proposed framework, for experiments 2 and 3. For experiment 1, the feature selection block is not used and the BO was employed to optimally select only the neurons number.

The algorithm optimizes the generalization (prediction) error, namely the Mean Squared Prediction Error:

$$MSPE = \frac{1}{n_{test}} \sum_{i=n_{train}+1}^{n_{train}+n_{test}} (\widehat{\mathrm{EP}}_{k-preSi}^{i} - \mathrm{EP}_{k-preSi}^{i})^2 \tag{19}$$

TABLE I
METHODOLOGY TRAINING METRICS AND RESULTS FOR: CASE 1) NO FEATURE SELECTION, CASE 2) FEATURE SELECTION WITH *DISTCORR* METRIC, CASE 3) FEATURE SELECTION WITH *MIC* METRIC

Parameter	**No Feature Selection**			***DistCorr***				***MIC***			
	MSPE	$\rho_{\widetilde{\mathrm{EP}}-\mathrm{EP}}$	$\rho_{\varepsilon-\widetilde{\mathrm{EP}}}$	*MSPE*	$\rho_{\widetilde{\mathrm{EP}}-\mathrm{EP}}$	$\rho_{\varepsilon-\widetilde{\mathrm{EP}}}$	$\mathbf{PCM}_i$	*MSPE*	$\rho_{\widetilde{\mathrm{EP}}-\mathrm{EP}}$	$\rho_{\varepsilon-\widetilde{\mathrm{EP}}}$	$\mathbf{PCM}_i$
EP_1	0.088	0.385	0.529	0.041	0.820	-0.026	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}, \mathrm{PCM}_{29}$	0.041	0.823	-0.077	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}, \mathrm{PCM}_{29}$
EP_2	0.181	0.289	-0.649	0.083	0.607	0.022	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}$	0.083	0.592	-0.079	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}$
EP_3	0.187	0.273	-0.615	0.086	0.619	-0.030	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}$	0.087	0.592	0.019	$\mathrm{PCM}_2, \mathrm{PCM}_4, \mathrm{PCM}_{27}, \mathrm{PCM}_{29}$ PCM_{92}

where n_{test} is the number of samples in the test set (that were not used in estimating the *metamodel*), while $\mathrm{EP}^i_{k-preSi}$ and $\widetilde{\mathrm{EP}}^i_{k-preSi}$ are the predicted and the actual EP values of the k^{th} electrical parameter, respectively.

Besides *MSPE*, we consider two more metrics; first-order correlations computed using Pearson's correlation coefficient [46], as follows:

$$\rho_{\widetilde{\mathrm{EP}}-\mathrm{EP}} = \frac{cov(\mathrm{EP}_{k-preSi}, \widetilde{\mathrm{EP}}_k)}{\sigma_{EP_{k-preSi}} \sigma_{\widetilde{EP}_k}} \quad (20)$$

$$\rho_{\varepsilon-\widetilde{\mathrm{EP}}} = \frac{cov(\varepsilon_k, \widetilde{\mathrm{EP}}_k)}{\sigma_{\varepsilon_k} \sigma_{\widetilde{\mathrm{EP}}_k}} \quad (21)$$

where $\widetilde{\mathrm{EP}}_k$ and $\mathrm{EP}_{k-preSi}$ represent the predicted parameter values and the real parameter values, respectively, while $\varepsilon_k = \widetilde{\mathrm{EP}}_k - \mathrm{EP}_{k-preSi}$, cov is the covariance and σ is the standard deviation.

Table I presents the methodology results and accuracy metrics for the described feature selection methods – *DistCorr* and *MIC*, as well as the case with no feature selection. As it can be seen, the prediction accuracies obtained when the feature selection step is employed are much higher compared to no feature selection step. The improvement is clear; on one hand, it confirms that PCM parameters carry enough information regarding the EP behavior, thus, they can be successfully used for EP prediction. On the other hand, incorporating the feature selection block to reduce the noise introduced by the non-influential PCMs results in better prediction accuracy.

Considering the experiments employing feature selection, from *metamodel* fitting point of view, the prediction errors are relatively large. As it was previously stated, not only the PCMs are influencing the circuit performances, there are additional statistical influences affecting the EPs that are not incorporated in the *metamodels* variables set. Nevertheless, some other influencing factors can be added to the analysis (temperature, voltages, etc.), but our main goal was linking the EPs distributions only to the technology parameters and disregarding the other factors.

The obtained *metamodels* are reliable, based on the correlation between the predicted values and the target values ($\rho_{\widetilde{\mathrm{EP}}-\mathrm{EP}}$), with values between 0.59 and 0.82, as well as the lack of correlation between the fitting error and the predicted values ($\rho_{\varepsilon-\widetilde{\mathrm{EP}}}$) (having values almost equal to 0). These metrics demonstrate that our methodology estimation is optimal, considering the orthogonality principle (in the sense of least squares), stating that the residual error of the optimal estimator is always orthogonal to any estimation [47].

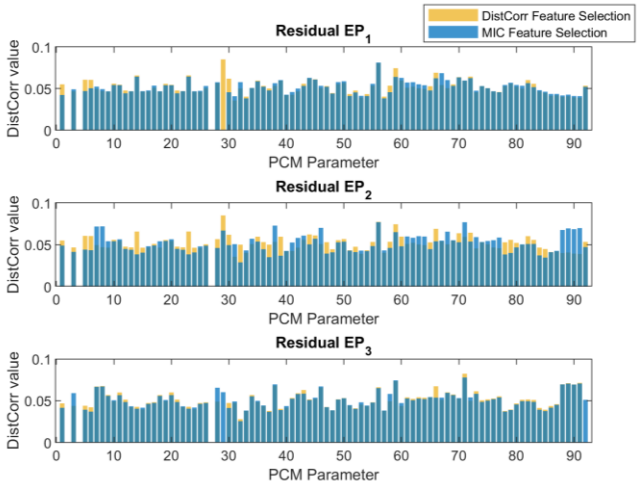


Fig. 3. The distance correlation between the *metamodel* residual error and the non-influential PCMs, for the three studied EPs.

As a sanity check for each EP *metamodel*, we computed the distance correlation (*DistCorr*) between the error residuals and all the non-influential PCMs that were not included in the regression model by the feature selection block. The choice of this metric was influenced by our literature survey in what regards feature selection metrics and methods, indicating that *DistCorr* is among the best in assessing the dependency of two variables, as it can easily cope with nonlinear relations between their noisy observations. Fig. 3 depicts the obtained results. As it can be observed, the distance correlations obtained are lower than 0.085, that reinforces the statement that the error residuals no longer contain any additional useful information that could explain the three EPs by the non-influential PCM variables.

Considering all of the above, we have the confirmation that our regression model's accuracy is optimal in what regards the available PCMs. The results for both implementations (*DistCorr* and *MIC*) are quite similar. Nevertheless, for the rest of the experiments, all *metamodels* are trained using *DistCorr* feature selection block, since we observed that for our problem this metric displayed lower computational costs.

C. Distribution Modeling

Another objective is to assess the *metamodels* distribution modelling capability; therefore, we are interested to quantify the empirical distributions of our generated EP samples prediction. Nevertheless, the produced ones are sufficient to be

TABLE II
SIMILARITY METRICS BETWEEN THE ELECTRICAL PARAMETERS DISTRIBUTIONS:
1) PRESI – PRESI ESTIMATED, 2) POSTSI – POSTSI ESTIMATED, 3) PRESI – POSTSI

Parameter	Similarity metric	$EP_{preSi} - \widetilde{EP}_{preSi}$	$EP_{postSi} - \widetilde{EP}_{postSi}$	$EP_{preSi} - EP_{postSi}$
EP_1	Bhattacharyya distance	0.0025	0.0017	0.0491
	Wasserstein metric	0.0123	0.0120	0.2208
EP_2	Bhattacharyya distance	0.0030	0.0026	0.0265
	Wasserstein distance	0.0139	0.0144	0.1451
EP_3	Bhattacharyya distance	0.0040	0.0032	0.0317
	Wasserstein distance	0.0158	0.0146	0.1496

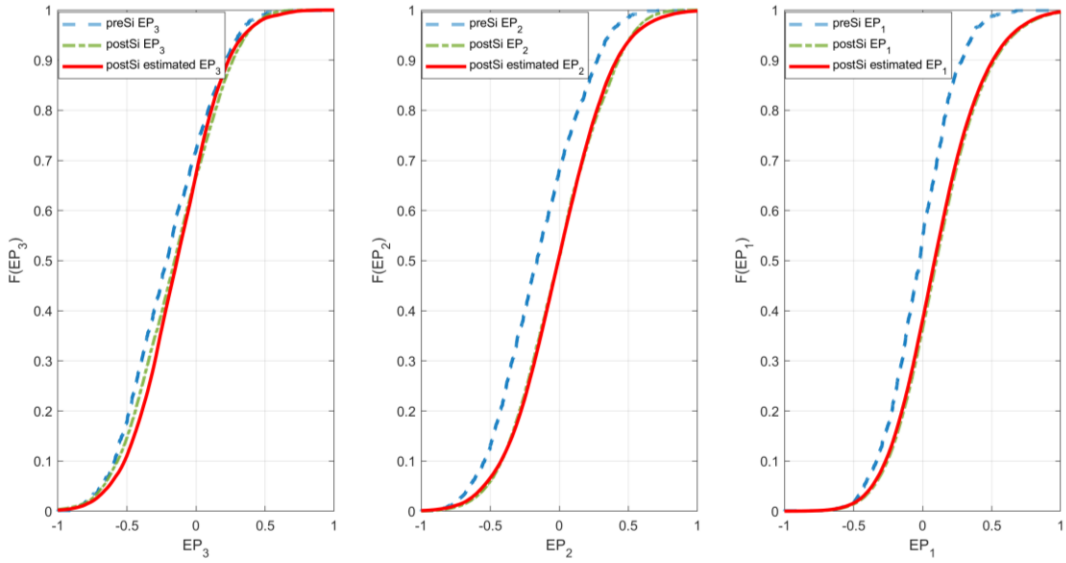


Fig. 4. *cdfs* of the three studied EPs in simulation (preSi), production (postSi) and *metamodel* prediction based on postSi samples (postSi estimated).

compared to their counterparts available from preSi simulation and postSi measurements. This can be obtained by generating a number of EP samples using (15), based on available PCMs data from preSi or postSi, followed by a comparison between the two distributions (estimated and existing).

In order to quantify the quality of the distribution modeling, we use two multidimensional similarity distribution metrics to measure the statistical distance between the input distribution and the estimated distribution based on our methodology. The first metric is Bhattacharyya distance [48], a divergence measure aimed to determine the similarity of two probability distributions, by computing the relative closeness (amount of overlap) of each two samples. The second metric is Wasserstein distance, also known as Earth's mover distance (EMD) [49], a metric arising from the optimal transport, but widely used to compare the probability distributions of two variables, due to its symmetry.

In table II are listed the metrics computed for each studied EP between the following combinations of distributions: preSi – preSi estimated, postSi – postSi estimated, and preSi – postSi. As it may be observed, the similarity between the simulation distribution and its predicted counterpart is high for all electrical parameters. This result was expected, since the *metamodels* are trained based on preSi data. Inspecting the second column's results from table II, a high degree of similarity can also be observed between the postSi modeled distribution and the postSi distribution. Moreover, the most important aspect is that the postSi estimated distribution is approximately an order of magnitude closer to the measured

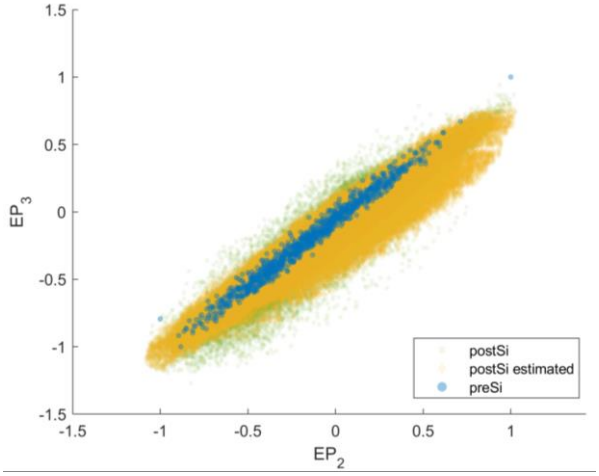


Fig. 5. Visualization of EP_2 vs EP_3 distributions for preSi, postSi and metamodel prediction based on postSi samples (postSi estimated).

postSi distribution, compared to the preSi distribution. This is clearly indicated by considering the table's last column. This fact is better illustrated in a visual manner by EPs marginal cumulative density functions plots (*cdfs*) represented in fig. 4, where the distribution modeled by our methodology using (measured) postSi PCM samples (red line) accurately resembles the real postSi distribution (green line). Fig. 5 illustrates the same fact when multimodal (joint) EPs distributions are considered for EP_2. Consequently, in terms of EP distribution modeling, our methodology shows a remarkable sense of adaptation to the technology parameters distribution changes compared to the one available through classical MC simulation.

TABLE III
RESULTS OF YIELD LOSS PREDICTION OF THE THREE STUDIED EPS – THE BENCHMARKS AND THE TWO EMPLOYED METHODS

Method		EP_1	EP_2	EP_3	Product
OOS EP_{postSi} [%] (164,129)		0.33	0.36	0.36	0.72
OOS EP_{preSi} [%] (1,000)		0.1	0.3	0.3	0.5
METHOD1 ***OOS*** $\widehat{EP}$ $\widehat{\mathbf{PCM}}_{postSi}$ (164,129)	μ [%]	0.13	0.28	0.47	0.67
	σ [%]	0.02	0.07	0.11	0.13
METHOD2 ***OOS*** $\widehat{EP}$ $\widehat{\mathbf{PCM}}_{postSi}$ (164,129)	μ [%]	0.47	0.51	0.54	0.78
	σ [%]	0.01	0.01	0.02	0.03

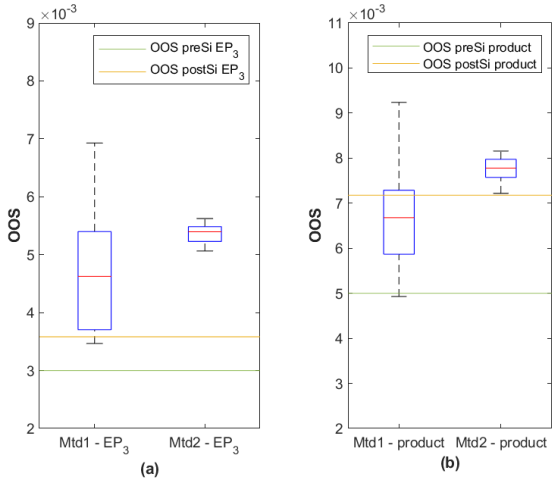


Fig. 6. Box plots of EP_3 (a) and the entire product (b), for the two methods.

D. Parametric Yield Prediction

The methodology proposed in Section IV can be directly used for yield prediction, by only using the modeled relationship between each EP and the PCMs, based on preSi data, and the information regarding the true postSi PCMs distribution. The fitted function is most helpful in postSi, where a limited number of wafers can be produced for yield prediction. Nevertheless, the produced ones are sufficient to validate the *metamodel*. Thus, the validated *metamodel* is used to predict the yield when more wafers are produced at different settings of PCMs. The PCMs input distributions can come in two forms: measurements samples (whether is historical data or PCM corners suggested by technology experts) or distribution models that can be used to produce the necessary samples.

In order to quantify and attest the proposed approach (Algorithm 2), we applied two methods, considering arbitrary specifications limits for each EP, and the results are presented in table III. To report robust results and assist with the interpretation of the overall results, we perform 10 iterations for each experiment and we compute the corresponding means (*mean*) and variances (*std*). Since our aim is to predict the

TABLE IV
SENSITIVITY ANALYSIS COEFFICIENTS AND INFLUENTIAL PCMS' RANKING FOR THE THREE STUDIED EPS

Parameter	Influential PCMs	*DistCorr* Coefficient
EP_1	PCM_2	**0.4567**
	PCM_{27}	0.4532
	PCM_4	0.4068
	PCM_{29}	0.2816
EP_2	PCM_{27}	**0.2252**
	PCM_2	0.2235
	PCM_4	0.2183
EP_3	PCM_4	**0.1901**
	PCM_{27}	0.1573
	PCM_2	0.1401

parametric yield, not to enhance it, the first step involved establishing a benchmark value for the experiments results. Therefore, we used the available EP_{postSi} data of 164,129 samples and EP_{preSi} data of 1,000 samples to compute the OOS for each studied EP and for the entire product.

Algorithm 2 involves modeling the PCMs distributions and the fitting error and this translates as two possible error sources when predicting the yield. The first applied method (*Mtd1*) implied computing the OOS on a dataset obtained as follows: we estimated the EP (220 values) using the corresponding *metamodels* and the available 220 measured samples of $\mathbf{PCM}_{postSi}$, followed by the modeling of the fitting error to reach the desired number of EP samples (164,129 samples). During the second method (*Mtd2*), we applied Algorithm 2 on the $\mathbf{PCM}_{postSi}$ samples; after modelling the postSi PCMs distributions to generate 164,129 PCMs samples, with the purpose of quantifying the PCMs distribution modeling accuracy, we estimated the EPs using the *metamodels* and computed the OOS on these samples, for each EP and the entire product.

When inspecting table III's results, it can be easily observed that every yield prediction (disregarding the particular estimation type) has the tendency to be closer to the considered ideal reference (i.e. postSi measurements OOS), than to the simulation-based-OOS. The visual representation in fig. 6 better illustrates the obtained results for a particular EP (a) and for the multivariate (product) case (b). This was expected; when modelling a certain distribution, any outcome involving it will inherit its accuracy properties – in our situation, the yield prediction as defined in (5). In other words, the closer the modeled distribution is to the real one based on postSi measurements, the better the yield prediction. Nevertheless, this statement does not apply entirely to the case of high yield products when the distribution accuracy has to be especially precise on low probability areas, e.g. distribution tails. This case of high yield products is a special case of our methodology outcome that needs to be addressed separately by tuning our methodology's distribution modelling, and it will constitute the subject of our future research.

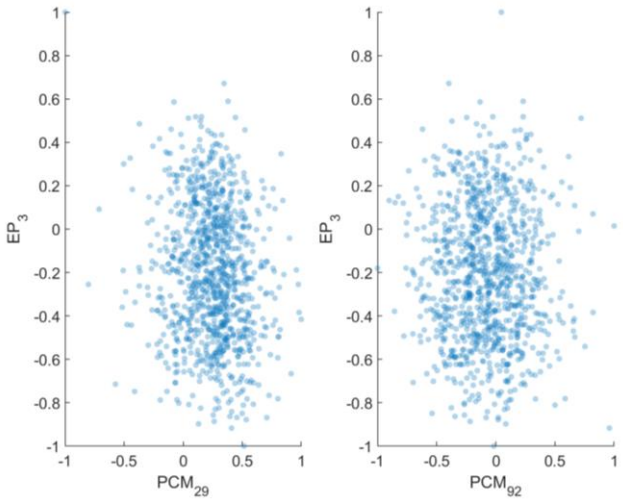


Fig. 7. Scatter plots between EP_3 and the additional influential PCMs selected by the *MIC* metric (PCM_{29} and PCM_{92}).

E. Sensitivity Analysis to Process Variation

The sensitivity analysis of the circuit performances with the process variations is a direct benefit of our methodology. It relies on the feature selection metric and the Bayesian Optimization. After the best *metamodel* is fitted for each EP, we can perform the SA by simply analyzing the set of influential PCMs selected by the BO and their corresponding distance correlation metrics.

In table IV is listed the ranked set of influential PCMs for each EP and their corresponding *DistCorr* coefficient is listed as well. When connecting this data with the results from table I, it can be easily observed that this metric's overall magnitude for each EP is strongly correlated with the regression models' accuracies. For comparison purposes, we inspected the two additional PCMs chosen as influential for EP_3 when *MIC* was employed (table I), without decreasing the *MSPE* value. Fig. 7 illustrates the scatter plots between EP_3 and PCM_{29} and PCM_{92}; the visual axis symmetry proves the statistical independence between the output (EP) and the variables (PCMs). Consequently, the *DistCorr* implementation can be validated as being more suitable as a feature selection method than *MIC* for this particular case. Moreover, the obtained results were validated by the designer.

VI. CONCLUSION

In this paper, we proposed an efficient and automated methodology for modelling the relationship between the circuits performance parameters (EPs) and the technology process variations (measured through the PCM parameters), in preSi stage, for ICs without dedicated PCM structures for every chip. On a broader scale, our contribution can be viewed as a bridge between variational-aware preSi verification and postSi validation, to overcome the challenges of these two worlds. During preSi stage, it is impossible to foresee the effects when the technological process suffers a certain degree of alteration as can be the case with process drift or fab-to-fab migration, while in postSi these process alterations can only be observed but they cannot be directly linked to the circuit performances in the absence of the 1-to-1 measurements correspondences.

We address these at preSi level, using the MC simulation data resulted by co-simulating the PCM structures schematics together with the circuit model. Based on the resulting correspondence between PCMs and EPs values, we are able to fit optimal ML models by employing a feature selection step (to select the influential PCMs) along with a regression algorithm (Neural Networks), everything wrapped up in the Bayesian Optimization framework to find the optimal feature selection metric threshold and the NN hyperparameter. Consequently, the reliable EPs regression *metamodels* are able to express both the functional and the statistical influence of the technology parameters on the circuit performances. This way, the proposed methodology can adapt to PCMs distribution deviations directly, without the need of adapting the statistical device parameters properties and re-simulating the circuit. Practically, it enables an instant and accurate snapshot on the circuit performances behavior when technological changes occur, without the need of additional simulations or postSi measurements. Nevertheless, the presented approach does not eliminate the need of postSi verification and product monitoring, that still stand as the ultimate product specifications compliance validation.

The outcome of our methodology enables the solutions to the four problems of the semiconductor industry stated in the Introduction. A large difference between the estimated performances distribution and the actually measured ones may provide a strong indication of the device models accuracy issues. The solution for fab-to-fab migration problem is straightforward; the methodology can be applied on historical PCMs characterizing the process window for the target fabrication plant, thus resulting upfront the circuit performances' distributions of the migrated product. The fitted metamodels are able to provide the necessary information for enabling a product-specific DoE test plan, as the sensitivity to process variation can be directly derived from the methodology. Similarly, by using the resulting *metamodels*, the process parameter influencing the decrease of performances (usually reported as parametric yield loss due to process variation) can be easily determined, thus solving the root cause analysis problem.

As for our approach for yield prediction, it displays the advantage of requiring less measurement data compared to the state-of-the-art methodologies, as the function that links the EPs and the PCMs is based entirely on preSi data and the postSi data is used to predict the yield when only few wafers are available. Moreover, the sensitivity analysis to process variation is a fast and computationally low outcome of our methodology, able to provide a ranking based on the employed correlation metrics.

The effectiveness of the proposed methodology and the two direct outcomes – sensitivity analysis to process variations and parametric yield prediction – were evaluated on an experimental LDO from Infineon Technologies and validated using the product postSi dataset. The results confirmed the accuracy and reliability of the proposed approach.

This data-driven methodology is applicable to almost any circuit that involves continuous variables, as long as it is possible to co-simulate the analog circuit model along with the PCM structures schematics, because it does not impose new experiments, but only standard Monte Carlo simulations. Although the simulation of complex circuits is costly, there is a

wide range of circuits can be simulated quickly where the proposed methodology is applicable. Moreover, our approach may prove its applicability in analog circuits calibration (tuning the design towards technology variations), as the performances in production test can be represented as a function of process parameters and the tuning knobs.

ACKNOWLEDGMENT

The authors would like to thank Petruț-Andrei Stănescu, Bogdan Dimitriu, Ozan Serpedin and Iulian Cioarcă from PSN department, Infineon Technologies Romania for their constant support with this project and for providing domain expertise.

This work was partially supported by the projects iDev40. The project iDev40 has received funding from the ECSEL Joint Undertaking under grant agreement No. 783163. The JU receives support from the European Union's Horizon 2020 research and innovation program. It is co-funded by the consortium members, grants from Austria, Germany, Belgium, Italy, Spain and Romania. The information and results set out in this publication are those of the authors and do not necessarily reflect the opinion of the ECSEL Joint Undertaking.

REFERENCES

[1] S. Mittal, "A Survey of Architectural Techniques for Managing Process Variation", *ACM Computing Surveys (CSUR)*, vol. 48, no. 4, art. 54, Feb. 2016.

[2] G. Moore, "VLSI: Some Fundamental Challenges", *IEEE Spectrum*, vol. 16, no. 4, pp. 30-20, Apr. 1979, DOI: 10.1109/MSPEC.1979.6367975.

[3]D. An *et al*., "A semiconductor yields prediction using Stepwise Support Vector Machine", presented at IEEE ISAM2009, South Korea, Nov. 2009, DOI: 10.1109/ISAM.2009.5376916.

[4] A. Ahmadi *et al*., "Yield Forecasting Across Semiconductor Fabrication Plants and Design Generations", *IEEE Trans. Comput-Aided Design Integ. Circuits Syst.*, vol. 36, no. 12, pp. 2120-2133, Dec. 2017.

[5] M. Karilahti, "Neural net analysis of integrated circuit yield dependence on CMOS process control parameters", *Microelectronics Reliability*, vol. 43, no. 1, pp. 117-121, Jan. 2003.

[6]B. Nehrer, K. Doong, and D. Ciplickas, „New visions for IC yield detractor detection", presented at IEEE EDTM2017, Japan, Mar. 2017, 10.1109/EDTM.2017.7947529.

[7] F. Domart *et al*., "An efficient methodology for Electrical Monitoring in manufacturing environment", IEEE/SEMI ASMC2008, USA, May 2008, DOI: 10.1109/ASMC.2008.4529036.

[8] N. Kupp, M. Slamani, and Y. Makris, "Correlating inline data with final test outcomes in Analog/RF devices", *Proc. IEEE/ACM DATE2011*, Grenoble, France, pp. 7–12.

[9] K.W. Pieper and E. Gondro, "An effective method for solving the covariance equation for statistical modeling", in *Proc. of SCD*, Dresden, Germany, 2011, pp.1-4.

[10] L. Abdallah *et al.*, "Experiences with non-intrusive sensors for RF built-in test", *Proc. IEEE ITC2012*, Anaheim, CA, USA, 2012.

[11] A. Adir *et al.*, "A unified methodology for pre-silicon verification and post-silicon validation", *Proc. IEEE/ACM DATE2011*, Grenoble, France, 2011, pp. 1–6.

[12] A. Nahir *et. al*., "Bridging pre-silicon verification and post-silicon validation", *Proc. DAC2010*, Anaheim, CA, USA, 2010, pp. 94-95.

[13] S. Ray and W. A. Hunt Jr., "Connecting pre-silicon and post-silicon verification", presented at FMCAD2009, Austin, TX, USA, 2009, DOI: 10.1109/FMCAD.2009.5351128.

[14] S. Mitra, S. A. Seshia, and N. Nicolici, "Post-silicon validation opportunities, challenges and recent advances", *Proc. DAC2010*, Anaheim, CA, USA, 2010, pp. 12-17, DOI: 10.1145/1837274.1837280.

[15] E.-D. Șandru et al., "Modeling the Dependencies between Circuit and Technology Parameters for Sensitivity Analysis using Machine Learning Techniques", *Proc. SMACD2019*, Lausanne, Switzerland, 2019, pp. 237-240.

[16] U. Sobe, K.-H. Rooch and D. Mörtl, "Simulation and Analysis of Analog Circuit and PCM (Process Control Monitor) Test Structures in Circuit Design", presented at Dresden Workshop on Circuit and System Design 2007, [Online] https://nbn-resolving.org/urn:nbn:de:swb:ch1-200700919.

[17] A. Saltelli *et al*., *Sensitivity Analysis in Practice: A guide to Assessing Scientific Models*, Chichester, UK, 2004, John Wiley & Sons.

[18] P.A. Ekstrom, "Eikos- A Simulation Toolbox for Sensitivity Analysis", Degree project, UTH, Uppsalla Univ., Sweden, 2005.

[19] K. Chan, A. Saltelli, and S. Tarantola, "Winding Stairs: A sampling tool to compute sensitivity indices", *Journal of Statistics and Comp*., vol. 10, pp. 187-196, Jul. 2000, https://doi.org/10.1023/A:1008950625967.

[20] I. Kovacs *et al*., "A novel entropy-based sensitivity analysis approach for complex systems", *Proc. of IEEE SSCI2016*, Athens, Greece, 2016, pp. 1-6.

[21] N. Durrandeab, D. Ginsbourgerc, O. Roustantb, and L. Carrarod, "ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis", *Journ. Of Multivar*. Anal., vol. 115, no. 1, pp. 57-67, March 2013.

[22] F. Gong, Y. Shi, H. Yu, and L. He, "Variability-aware Parametric Yield Estimation for Analog/Mixed-signal Circuits: Concepts, Algorithms and Challenges", *IEEE Design & Test*, vol. 31, no. 4, pp. 6–15, Aug. 2014.

[23] H.-G. Stratigopoulos and S. Sunter, "Fast Monte Carlo-based estimation of analog parametric test metrics," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 33, no. 12, pp. 1977–1990, Dec. 2014.

[24] A. Singhee and R.A. Rutenbar, "Why quasi-Monte Carlo is better than Monte Carlo or Latin Hypercube Sampling for statistical circuit analysis", *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 29, no. 11, pp. 1763-1776, Nov. 2010.

[25] D. An *et al*., "A Semiconductor Yields Prediction using Stepwise Support Vector Machine", IEEE ISAM2009, South Korea, Nov. 2009, 10.1109/ISAM.2009.5376916.

[26] C. Xanthopoulos, "Gaussian Process-Based Wafer-Level Correlation Modeling and Its Applications", in *Machine Learning in VLSI Computer-Aided Design*", 1st ed., I.M. Elfadel, D.S. Boning, X. Li, Cham, Switzerland, Springer, 2019, ch. 5, pp. 119-173.

[27] B. Liu, V. Fernandez and G. G. E. Gielen, "Efficient and Accurate Statistical Analog Yield Optimization and Variation-Aware Circuit Sizing Based on Computational Intelligence Techniques", *IEEE Trans. Comput-Aided Design Integr. Circuits Syst*., vol. 30, no. 6, pp. 793-805, June 2011.

[28] E.-D. Șandru, E. David, and G. Pelz, "Pre-Silicon Yield Estimation using Machine Learning Regression", *Proc. IEEE ICECS2019*, Genoa, Italy, 2019, pp. 103-104.

[29] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning. Data Mining, Inference, and Prediction", 2nd ed., New York, NY, USA, Springer, 2009, ch. 3, pp. 43-94.

[30] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed., New Jersey, USA, Pearson, 2009, ch. 4&6.

[31] M.E.J. Newman, G.T. Barkema, "Monte Carlo Methods in Statistical Physics", 1st ed., New York, NY, USA, Clarendon Press, 1999.

[32] C.P. Robert and G. Casella, "Monte Carlo Statistical Methods", 2nd ed., Springer, 2004, ch. 3, https://doi.org/10.1007/978-1-4757-4145-2.

[33] E.-D. Șandru and E. David, "Unified Feature Selection and Hyperparameter Bayesian Optimization for Machine Learning based Regression", *Proc. ISSCS2019*, Iasi, Romania, 2019, pp. 1-5.

[34] J. Miao and L. Niu, "A survey on feature selection", *Procedia Comp. Science*, vol. 91, pp. 919 – 926, 2016.

[35] F. Cilici *et. al*., "Assisted test design for non-intrusive machine learning indirect test of millimeter-wave circuits", *Proc. ETS2018*, Bremen, Germany, 2018.

[36] G. J. Székely and M. L. Rizzo, "Brownian distance covariance", *Ann. Appl. Stat.*, vol. 3, no. 4, pp. 1236–1265, 2009.

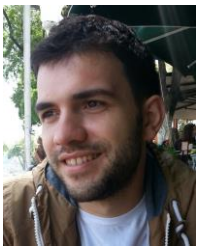
[37] D.N. Reshef *et al*., "Detecting Novel Associations in Large Data Sets", *Science*, vol. 334, pp. 1218-1224, Dec 2011.

[38] D.N. Reshef, Y Rechef, M. Mitzenmacher, and P. Sabeti, "Equitability Analysis of the Maximal Information Coefficient, with Comparisons", arXiv:1301.6314v2 [cs.LG].
[39] D. Albanese, M. Filosi, R. Visintainer, S. Riccadonna, G. Jurman and C. Furlanello. "minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers", *Bioinformatics*, vol. 29, no. 3, pp. 407-408 2013.
[40] J.S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization", in *Advances in Neural Info. Proc. Syst.*, pp. 2546-2554, 2011.
[41] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian Optimization", *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.
[42] J. Snoek, H. Larochelle, and R. Prescott Adams, "Practical Bayesian optimization of machine learning algorithms", *Proc. NPIS2012*, vol. 2, pp. 2951-2959, 2012.
[43] A.A. Bayrakci, A. Demir, and S.Tasiran, "Fast Monte Carlo Estimation of Timing Yield with Importance Sampling and Transistor-Level Circuit Simulation", IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 29, no. 9, pp. 1328-1341, Aug. 2010.
[44] I. Kovacs, M. Topa, A. Buzo, and G. Pelz, "An Accurate Yield Estimation Approach for Multivariate Non-normal Data in Semiconductor Quality Analysis", Proc. IEEE SMACD2017, Giardini Naxos, Italy, 2017, pp. 1-4.
[45] R. Nornadiah and W.Y. Bee, "Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests", *Journ. of Statis. Model. and Anal.*, vol. 2, no. 1, pp. 21–33, 2011.
[46] R.A. Fisher, "Statistical Methods for Research Workers" in *Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics)*, 1st ed., S. Kotz, N.L. Johnson, New York, NY, USA, Springer, 1992, ch. 4.
[47] T.K. Moon, W.C. Stirling, "Mathematical Methods and Algorithms for Signal Processing", New Jersey, USA, Prentice-Hall, 2002, ch. 3, pp. 130-137.
[48] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions", *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99-109, 1947.
[49] C. Villani, "Optimal Transport, Old and New", Springer, 2009, pt. 1, ch. 6, pp. 93-111, DOI:10.1007/978-3-540-71050-9.

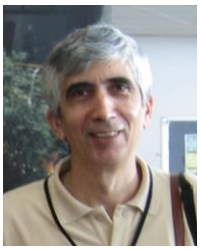
Elena-Diana Șandru (Student M'17) received the B.Sc. degree in electronics and telecommunications, in 2014, from University Politehnica of Bucharest, and the M.Sc. degree in 2016, after an internship at Télécom ParisTech (Multimedia department). She is currently pursuing the Ph.D. degree within the Doctoral School of Electronics, Telecommunications and Information Technology (University Politehnica of Bucharest), in partnership with Infineon Technologies. Her research interests include machine learning and artificial intelligence applied in semiconductor industry and integrated circuits yield prediction and adjustment.

Emilian David holds a Ph.D. degree in electronics and telecommunications from the Faculty of Electronics and Telecommunications, "Gh. Asachi" Technical University of Iași, Romania. He has 20 years of academic and industry research experience in applications of Machine Learning, Signal and Image Processing algorithms with 30 publications and 5 patents. He joined Infineon Technologies with the main focus of improving development methodologies of integrated circuits by using ML algorithms.

Ingrid Kovacs holds a Ph.D. degree in engineering sciences from the Faculty of Electronics and Telecommunications, at Technical University of Cluj-Napoca, Romania. Her focus is on machine learning algorithms applied for integrated circuits' yield estimation and prediction. She has more than 15 publications and two filed patents.



Andi Buzo holds a Ph.D. degree in engineering sciences from the Faculty of Electronics and Telecommunications, at University Politehnica of Bucharest. His focus is the improvement of methodologies for the efficient development of integrated circuits. He has 10 years of research experience in applications of machine learning algorithms, with more than 60 publications and five filed patents.



Corneliu Burileanu is professor at University Politehnica of Bucharest since 1995. He received the Ph.D. degree in electronics and telecommunications from the same university, in 1986. Its main fields of research are speech and voice recognition, machine learning, vocal signal analysis and coding, spoken terms recognition and natural language processing. He has participated in or led over 15 research grants, holds five national patents and has published over 150 books, book chapters and articles.

Georg Pelz received a diploma degree in computer science (University of Dortmund, 1988), a Ph.D. in CAD for physical design verification of microelectronic circuits (University of Duisburg, 1993), a habilitation in modeling and simulation of mechatronic systems (2001) and a honorary professorship in the field of automotive electronics (University of Duisburg, 2012). He authored and co-authored two books and 80+ scientific papers.