

Efficient Statistical Parameter Selection for Nonlinear Modeling of Process/Performance Variation

Hassan Ghasemzadeh Mohammadi, *Student Member, IEEE*, Pierre-Emmanuel Gaillardon, *Member, IEEE*, and Giovanni De Micheli, *Fellow, IEEE*

Abstract—With the growing number of process variation (PV) sources in deeply nano-scaled technologies, parameterized device and circuit modeling is becoming very important for chip design and verification. However, the high dimensionality of parameter space, for PV analysis, is a serious modeling challenge for emerging VLSI technologies. These parameters correspond to various interdie and intradie variations, and considerably increase the difficulties of design validation. Today’s response surface models and most commonly used parameter reduction methods, such as principal component analysis and independent component analysis, limit parameter reduction to linear or quadratic form and they do not address the higher order of nonlinearity among process and performance parameters. In this paper, we propose and validate a feature selection method to reduce the circuit modeling complexity associated with high parameter dimensionality. This method relies on a learning-based nonlinear sparse regression, and performs a parameter selection in the input space rather than creating a new space. This method is capable of dealing with mixed Gaussian and non-Gaussian parameters and results in a more precise parameter selection considering statistical nonlinear dependencies among input and output parameters. The application of this method is demonstrated in digital circuit timing analysis in both FinFET and Silicon Nanowire technologies. The results confirm the efficiency of this method to significantly reduce the number of required simulations while keeping estimation error small.

Index Terms—Circuit modeling and simulation, parameter reduction, process variation (PV), statistical analysis.

I. INTRODUCTION

THE CURRENT dimension shrinkage trend in CMOS technology has led to the development of various nano-devices such as Doped/Schottky barrier silicon nanowire FETs (SiNWFETs) [1], [2], carbon nanotube FETs [3], and graphene-based devices [4] exhibiting short-channel effect immunity, greater electrostatic control, and lower leakage. However, fabrication-induced process variations (PVs) on

device and circuit characteristics are a growing challenge with ongoing feature size downscaling. Geometrical and physical parameter variations, e.g., changes in transistor effective gate-length or threshold voltage (V_{th}), lead to considerable effects on performance and reliability of modern integrated circuits (ICs). Moreover, the performance sensitivity on each parameter can vary from a technology to the next. These parameter fluctuations may adversely affect the circuit performance. Therefore, variation analysis is becoming significant in circuit modeling and simulation.

PV analysis through simulation is the mostly realistic approach for comprehensive study of variation impacts for both circuit static timing and leakage power. Parametric variation analysis is performed by means of Monte Carlo (MC) simulation and is widely used in microelectronics industries, even if it is extremely time-consuming for large circuits. Considering the variety of local and global variations in device and circuit simulations would need up to some thousands or millions of variation variables to represent the distributions of the geometrical and physical parameter quantities [5]. Moreover, for practical reasons circuits are usually characterized with relatively small number of parameters through compact models. Scaling beyond 28 nm forces a transition from CMOS technologies to others like fully depleted silicon on insulator, FinFET, or SiNW, for which statistical compact models are inevitable for variability aware design. Statistical compact models exploit technology computer aided design (TCAD) to predict the impacts of fluctuations on device performance [6], [7]. The results of TCAD simulation can be fed to SPICE-like simulator for MC simulations of circuits. Nevertheless, the high dimensionality of the parameters space and the computational complexity of TCAD simulation make the PV analysis very costly and even sometimes infeasible. Therefore, new tools which speed up the variation analysis for deeply nano-scaled circuits are required.

The efficiency of current methods for performance analysis, e.g., statistical timing verification techniques, critically relies on the dimension of the parameter space [8]–[10]. Most of the existing techniques such as principal component analysis (PCA) and independent component analysis (ICA) use a linear transformation to reduce the number of input parameters by decorrelating the input space [11], [12]. In spite of their popularity, they are inherently limited because they only consider the relations among the input parameters and ignore

Manuscript received August 23, 2015; revised December 6, 2015; accepted March 10, 2016. Date of publication March 29, 2016; date of current version November 18, 2016. This work was supported by ERC-2009-AdG-246810. This paper was recommended by Associate Editor A. Raychowdhury.

The authors are with the Integrated Systems Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: hassan.ghasemzadeh@epfl.ch; pierre-emmanuel.gillardon@epfl.ch; giovanni.demicheli@epfl.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCAD.2016.2547908

the impact of each input on the circuit outputs. This limitation becomes important when either some critical parameters, which significantly affect the output, are ignored or a large set of transformed parameters may still be produced after redundancy removal. Moreover, although statistical methods, such as reduced ranked regression (RRR) and canonical correlation analysis (CCA), consider the correlation between the input parameters and the circuit outputs, they ignore the correlation among the input parameters [13]. Therefore, they may lead to a large set of correlated parameters while the input space can be compressed by considering interparameter correlation. Last but not least, the mentioned methods put strict assumptions on the distribution of the model parameters such as Gaussian distribution which limits their applicability to recently proposed nano-devices in which parameters have mixed Gaussian and non-Gaussian distributions.

In this paper, we introduce a novel multiobjective parameter selection method capable of addressing the aforementioned limitations. A preliminary version of this paper appeared in [14]. This method takes into account the interset (among inputs) and intraset (between input and output sets) correlations. The objective function is modified to be distribution free and minimize the error of output estimation. The major contributions of the method can be summarized as the following.

- 1) High precision by considering nonlinear dependencies between interset and intraset parameters.
- 2) Distribution free feature selection which can be used for any model or parameter set with unknown statistical distributions.
- 3) Feature selection in the input parameter space which preserves the meaning of the parameters and highlights the major contributors on device or circuit variability.

We show that such parameter selection approach leads to more feasible PV analysis of complex design. Therefore, parameterized models are built with a smaller set of statistically significant parameters.

To validate the technique, we performed two sets of experiments on two different target technologies. First, we use FinFET 20 nm technology as a contemporary IC technology. Based on that, we analyzed the delay variation of the longest path for a couple of ITC'99 and ISCAS benchmark circuits. Here, $5\times$ speed up in MC is obtained for timing variation analysis with the average variance error of 4.1% in presence of 5% variation on each parameter. Second, we use double-gate silicon nanowire FETs (DG-SiNWFETs) technology as a strong potential substitute for future silicon technologies [2]. The simulation results for timing analysis of the combinational logic ISCAS89 benchmark circuit s27, using this technology, prove the performance of this technique for selecting relevant parameters. Indeed, up to $2.5\times$ speed up in MC is obtained for timing variation analysis with the variance error of 11.7% in presence of 30% variation on each parameter.

The organization of this paper is as follows. Section II describes the motivation and background. In Section III, we explain the proposed methodology for fast variation analysis, including a nonlinear learning-based sparse parameter selection technique. Section IV validates the method using simulations, and finally Section V concludes this paper.

II. BACKGROUND AND MOTIVATION

In the nanoscale era, modeling and simulation of VLSI circuits have been facing a significant challenge called “curse of dimensionality.” Due to the extra process complexity required to build deeply scaled devices, the number of device parameters affected by interdie and intradie variations dramatically grows [15]. The variation modeling requires distinct variables for each physical and structural parameter in order to represent the effect of PV. Exploiting modeling techniques such as response surface model technique is not applicable anymore because the complexity of the model is exponential with respect to the number of parameters [16].

Fortunately, all the model parameters are not independent. Indeed, the correlations among these parameters can be exploited to get rid of redundant parameters. Parameter reduction methods are generally divided into two categories.

- 1) *Unsupervised Parameter Reduction*: In this category, parameter reduction is done only considering the correlation among models parameters (intraset correlation). Indeed the impact of the parameters on the model functionality is not considered. Thus, the parameters with a negligible impact on the functionality may be selected. Methods such as PCA and ICA are among the examples of unsupervised parameter reduction methods. These methods are favorable when finding the relation between model parameters and its functionality is not easy, nor is it cost-effective.
- 2) *Supervised Parameter Reduction*: The impact of a redundant parameter can be significant on the model outputs. Therefore, the correlation between model parameters and model outputs (intraset correlation) can be exploited for efficient feature reduction. This is done by using either an objective function that considers the intraset correlation (i.e., CCA) or a regression model that shows the relations between models parameters and the model output (i.e., ℓ_1 -norm regularization). Unlike the unsupervised category, the parameters are ranked based on their impact on the model output, and then the important ones are selected.

The mentioned methods and some of their modifications have widely used in VLSI applications. In the following, we discuss these methods in detail and review their potentials and their limitations. In Section II-D, we briefly explain the requirements of an efficient parameter reduction method for PV analysis, which provides the motivation for our research.

A. Principal Component Analysis

PCA has been widely used in the field of device compact modeling [11] and statistical static timing analysis [17]. The PCA performs a linear transformation through the conversion of correlated parameters into a smaller set of new uncorrelated parameters, called principal components. Indeed, the parameter space is transformed to new coordinates in which the largest variance of the data is projected to the first few principal components. Then, the principal components, which have

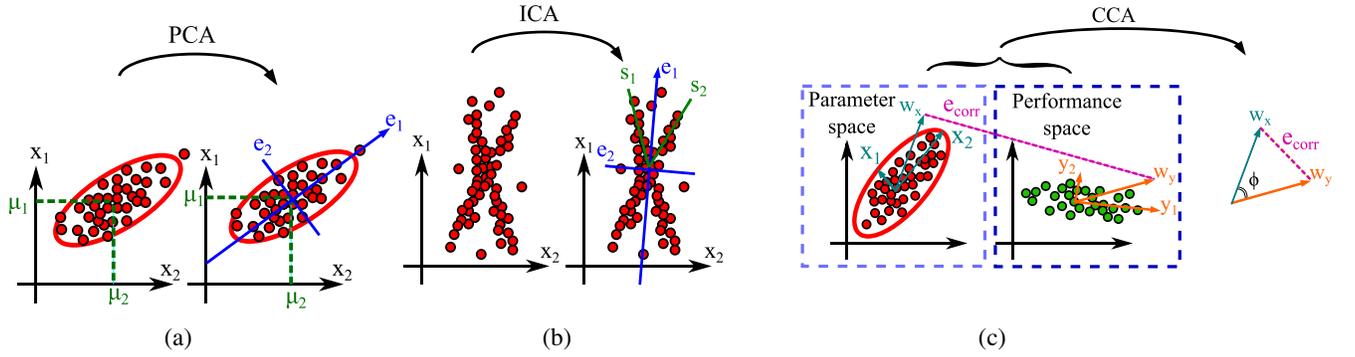


Fig. 1. Visual representation of the common parameter reduction techniques. (a) PCA. (b) ICA. (c) CCA.

the maximum variations in the parameter space, are selected as it follows.

Given an n -dimensional input set $\mathbf{x} = [x_1, x_2, \dots, x_n]$, which has zero mean and multivariate Gaussian distributions. Assume that the correlation of components in \mathbf{x} is represented by the covariance matrix Σ . Using eigen-decomposition procedure, PCA computes Σ as

$$\Sigma = \mathbf{E} \psi \mathbf{E}^T \quad (1)$$

where ψ is a diagonal matrix of Σ eigenvalues, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]$ contains the corresponding orthogonal eigen-vectors. Fig. 1(a) illustrates the projection of a multivariate Gaussian distribution for which the vectors \mathbf{e}_1 and \mathbf{e}_2 are corresponding orthogonal eigen-vectors computed by PCA. By including few eigen-vectors of \mathbf{E} that have the largest eigen-values into the projection matrix \mathbf{E}_{red} , the new parameter set that has a smaller dimension than that of the original set can be obtained by

$$\mathbf{x}_{\text{red}} = \mathbf{E}_{\text{red}}^T \mathbf{x}. \quad (2)$$

As a main limitation of the PCA, it only focuses on the correlation among the input parameters and discards the dependency between the input parameters and the corresponding outputs. Therefore, a set of parameters may be selected that have no considerable impact on the output of the model. Moreover, when the underlying statistical information about the distribution of the input parameters is unknown, PCA fails to select the relevant parameters that contribute to the model output. Last but not least, the maximum performance can be obtained when the distribution of input parameters is Gaussian [18].

B. Independent Component Analysis

For a Gaussian distribution, uncorrelatedness implies statistical independence which means that the principal components are also statistically independent. However, such a property does not hold for general non-Gaussian distributions. In (2), the random vector \mathbf{x} consists of correlated non-Gaussian random variables, and a PCA transformation would not guarantee statistical independence for the components of the transformed input parameters. Since the PCA technique focuses only on second order statistics, it can only ensure uncorrelatedness, and not the much stronger requirement of statistical independence.

ICA is a statistical technique that precisely transforms a set of non-Gaussian correlated parameters to a set of parameters that are statistically as independent as possible, through a linear transformation. Given a linear mixture of n independent components such as $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, that are the correlated non-Gaussian parameters, the n statistically independent components like $\mathbf{s} = [s_1, s_2, \dots, s_n]^T$ can be obtained as follows:

$$\mathbf{x} = \mathbf{A} \mathbf{s} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a transformation matrix. Similar to PCA, the independent components of vector \mathbf{s} are mathematical abstractions that cannot be directly observed. The ICA technique requires centering and whitening of the vector \mathbf{x} , leads to variables with zero mean and unit variance. The goal of ICA is to estimate the elements of unknown transformation matrix \mathbf{A} , and the samples of statistically independent components of vector \mathbf{s} given only the samples of the observed vector \mathbf{x} . Equation (3) can also be written as

$$\mathbf{s} = \mathbf{W} \mathbf{x} : s_i = \mathbf{w}_i^T \mathbf{x} = \sum_{j=1}^n w_{ij} x_j \text{ for } i = 1, \dots, n. \quad (4)$$

Here, $\mathbf{W} \in \mathbb{R}^{n \times n}$ is the inverse of the unknown mixing matrix \mathbf{A} . Fig. 1(b) represents a hypothetical multivariate distribution along with the corresponding independent components (s_1 and s_2). It is obvious that ICA has found the original components by relaxing the constraint that all the identified directions have to be orthogonal. However, PCA fails to estimate the major components for this data set, as it finds each uncorrelated and orthogonal component in the direction of highest variance (\mathbf{e}_1 and \mathbf{e}_2). Algorithms for computing ICA estimate the vectors \mathbf{w}_i that maximize the non-Gaussianity of $\mathbf{w}_i^T \mathbf{x}$ by solving a nonlinear optimization problem. This can be performed by using kurtosis, neg-entropy, and mutual information as typical methods measuring non-Gaussianity [19].

In contrary to PCA, ICA is used for feature reduction of non-Gaussian parameters. When more than two parameters follow the Gaussian distribution, ICA fails to find the constructive components [20]. ICA like PCA is output ignorant which means that the parameters with minor impacts on the outputs may be selected, and important information may be lost during the dimensionality reduction.

C. Canonical Correlation Analysis

As an output sensitive statistical method, CCA is capable of reducing the parameters which have major impacts on the output. Suppose that the relationship between model parameters $\mathbf{x} = [x_1, x_2, \dots, x_n]$, and model outputs $\mathbf{y} = [y_1, y_2, \dots, y_m]$, can be estimated by the following regression:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \varepsilon \quad (5)$$

where $\mathbf{X} \in \mathbb{R}^{n \times k}$ and $\mathbf{Y} \in \mathbb{R}^{m \times k}$ are matrices containing the samples of the \mathbf{x} and the corresponding \mathbf{y} , \mathbf{A} is a $(m \times n)$ matrix to project the n -dimensional parameter space onto an m -dimensional output space, and ε is matrix of a zero-mean random error of the regression. CCA computes two sets of basis vectors, $\mathbf{w}_x \in \mathbb{R}^n$ for \mathbf{X} and $\mathbf{w}_y \in \mathbb{R}^m$ for \mathbf{Y} , such that the correlations between the projections of the variables onto these basis vectors ($\mathbf{w}_x^T \mathbf{X}$ and $\mathbf{w}_y^T \mathbf{Y}$) are mutually maximized

$$\begin{aligned} \rho &= \arg \max_{\mathbf{w}_x, \mathbf{w}_y} \text{corr}(\mathbf{w}_x^T \mathbf{X}, \mathbf{w}_y^T \mathbf{Y}) \\ &= \frac{\mathbb{E}[\mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y]}{\sqrt{\mathbb{E}[\mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x] \mathbb{E}[\mathbf{w}_y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y]}} \\ &= \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}} \end{aligned} \quad (6)$$

where $\mathbf{C}_{xy} \in \mathbb{R}^{n \times m}$, $\mathbf{C}_{xx} \in \mathbb{R}^{n \times n}$, and $\mathbf{C}_{yy} \in \mathbb{R}^{m \times m}$ are in intersets and within sets covariance matrices.

Since the correlation is not influenced by rescaling, the CCA problem is formulated to maximizing the numerator subject to

$$\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x = 1, \quad \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y = 1. \quad (7)$$

The CCA method then is reduced to find the optimum ρ under the above constraints. This problem now can be solved by Lagrange multiplier method result in

$$\mathbf{C}_{xx} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{w}_x = \lambda^2 \mathbf{C}_{xx} \mathbf{w}_x. \quad (8)$$

Finally, the problem is reduced to a general eigen problem of the form $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$. Therefore, the sequence of basis vectors (\mathbf{w}_x s and \mathbf{w}_y s) is obtained by computing the eigenvectors of the (8). This leads to a new coordinate system that optimizes the correlations between input parameters and target outputs.

Now, the parameter reduction problem is reduced to form an r -dimensional space of the most important basis vectors so that $r \leq n$. The importance of each basis vector is determined according to their computed eigen values. Fig. 1(c) shows how CCA selects a basic vector in the input space which has the maximum correlation with the projected output space. The maximum correlation is obtained when the angle between two projected variables is reduced toward 0.

Similar to previous methods, CCA strictly requires a Gaussian distribution of input variables to significantly enhance the performance of feature reduction. However, variation analysis of deeply nanometer scaled technologies has revealed that the distribution of several parameters, such as V_{th} , does not follow a Gaussian distribution [21]. Thus, the

performance of feature selection may be considerably affected by the distribution of input parameters. Furthermore, in CCA like other linear models input parameters are considered independent, while several geometrical parameters of the transistor, e.g., gate length and V_{th} are correlated to one another [20].

D. Sparse Linear Regression via ℓ_1 -Norm Regularization

Sparsity via ℓ_1 -norm regularization is a learning-based feature selection method [22]. This method focuses on the cases where the number of samples is less than the number of coefficients. In this case, the solution (i.e., the model coefficients) is not unique, unless exploiting several additional constraints. As a result, sparsity can be used to uniquely determine the values. For a vector of input parameters such as \mathbf{w} , ℓ_1 -norm regularization technique is used to find the most important parameters subject to the following objective function:

$$\min L = \|\mathbf{w}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_1 \quad (9)$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ represent the ℓ_2 -norm and ℓ_1 -norm of a vector, respectively. The ℓ_1 -norm ($\|\mathbf{w}\|_1$) gives us the sum of the absolute non-zero elements of the \mathbf{w} . Indeed, it measures the sparsity of \mathbf{w} in the regression model. Therefore, ℓ_1 -norm regularization attempts to find a sparse solution that minimizes the least-square error. λ is a hyper parameter in (8) that controls the tradeoff between the sparsity of the input parameters and the minimal value of the loss function $\|\mathbf{w}\mathbf{x} - \mathbf{y}\|_2^2$. For example, a large λ value will result in a small error function, but it will increase the number of non-zero elements in \mathbf{w} . It is important to note that a small error function does not necessarily mean a small modeling error. Although this method can find linear dependencies between input and output parameters, it suffers from lack of modeling nonlinear relations among parameters.

E. Parameter Reduction for PV Analysis

In order to handle the high dimensionality of the circuits' models in presence of PV, parameter reduction is necessary to find the intrinsic dimensionality of the models. The intrinsic dimensionality of the models is the minimum number of PV parameters needed to account for variation analysis. In the framework of PV analysis, the applicable parameter reduction method needs to capture the nonlinearity among process parameters and performance parameters. The simple construction of process parameters from the reduced space is necessary for experimental simulations. Last but not least, the reduction method should be able to handle PV variables with different statistical distributions.

The methods mentioned above are linear. Considering nonlinear dependencies can remarkably increase the precision of parameter reduction. Many modifications [20] have been proposed to alleviate this problem, e.g., function driven component analysis, quadratic RRR, kernel PCA, and kernel ICA. kernel-based methods try to address this issue by using fixed nonlinear kernels, e.g., quadratic, polynomial, and exponential functions. They map the input space to a higher dimensional space, and then linearly relate the model to the output space. This has several limitations: it increases the dimensionality of

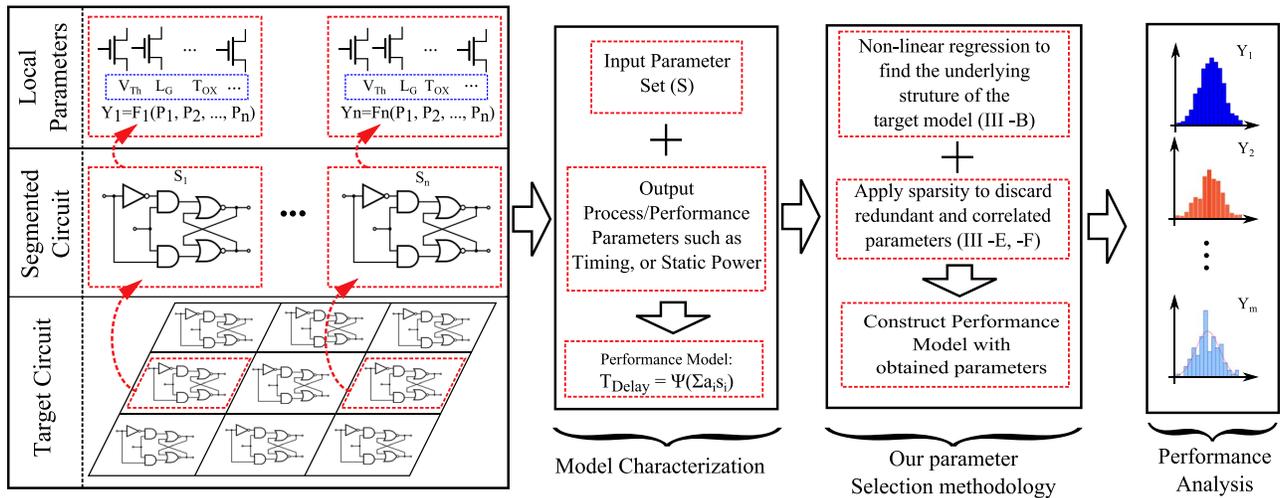


Fig. 2. General flow of the parameter reduction toward a fast and efficient PV analysis.

problem before reducing it, and, more importantly, it assumes a known nonlinear relationship between the input and the output spaces.

Moreover, these methods perform dimensionality reduction, meaning that the problem is transformed from an input parameter space to a reduced parameter space. Since these transformations change the meaning of physical parameters, either we need to reconstruct the original parameters from the reduced parameters, or modify the PV simulator to work with the new set of parameters. Modifying device and process simulators like TCAD simulators is very challenging. Moreover, due to the nonlinearity of these transformations, it is extremely costly to reconstruct the original parameters from the lower dimension space. While the above nonlinear methods increase the precision, but they can not be used efficiently in our applications. Therefore, a parameter selection method in the input space, that considers the nonlinear relation between the input and the output spaces, is then proposed in the following section. The method accelerates statistical PV analysis and addresses the major drawbacks of the previous work.

III. LEARNING-BASED PARAMETER REDUCTION FOR FAST VARIATION ANALYSIS OF EMERGING DEVICES

In this section, we present a learning-based feature selection method adapted to VLSI modeling and simulation. We overview the framework of parameter selection, and then discuss the method in detail.

In Section III-A, we introduce our general parameter reduction methodology for PV analysis. Next, we briefly review the nonlinear regression through feed forward neural network (FFNN) which is used to build the nonlinear regressor of our parameter reduction method (Section III-B). The training and validation algorithms of the nonlinear regressor are discussed in Sections III-C and III-D. In Section III-E, the mathematical background of the sparsity we use in our method is explained. Finally, the proposed parameter reduction method is introduced in Section III-F.

A. Parameter Reduction Toward Low Dimensional Device and Circuit Models

In order to achieve fast PV analysis for digital ICs, large designs have to be partitioned into a set of logic cells. The size of each logic cell should be small enough such that the parameter selection can be efficiently performed. After extracting the variation parameters, logic cells are hierarchically clustered to form the initial large circuit. Then, the parameter selection can be performed again on each cluster with the new reduced parameter set, to completely cover the targeted large circuit. In most cases, the circuits that we want to model are known to be structured in the sense that their physical parameters are highly correlated and therefore the associated models are compressible. Considering the correlation among parameters provides an opportunity by which the circuit functionality can be estimated with smaller number of parameters which leads to a lower computational complexity.

Fig. 2 illustrates the general flow of the proposed parameter reduction method for circuit PV analysis. First, input and output parameter sets are selected according to the hierarchy level at which the parameter reduction is performed. The input parameter set can be obtained from three different sources: 1) compact model parameters of the device; 2) parameters of the TCAD model; and 3) measured characteristics of the fabricated devices such as threshold voltage (V_{Th}), I_{on} , I_{off} , and subthreshold slope (SS). The output parameter set can also be selected among delay, power consumption, or any other functionality criteria of the logic cells and circuit blocks. In the next step, a learning-based statistical multivariate regression is used to predict the relations among the input and output parameter sets. The objective function of the regression is modified to minimize the error of the output prediction while discarding the unnecessary parameters. Here, training the regressor under the constraint of a limited error bound is the major step toward parameter reduction. Finally, the most significant parameters are only considered for the PV analysis of the target circuit, whereby increasing the evaluation speed.

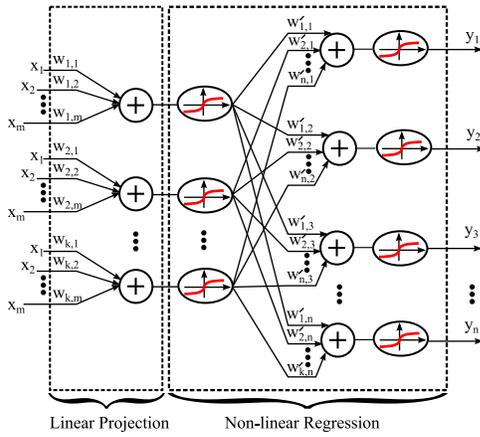


Fig. 3. Structure of nonlinear regression using FFNN.

B. Nonlinear Regression via Feed Forward Neural Network

FFNN is a powerful nonlinear regressor known to be a universal approximator by increasing the size of hidden layer [23]. We adopt FFNN here as our regressor to consider the nonlinear relations among the parameters. The regression model is formulated as

$$\mathbf{Y} = \mathbf{W}'\tanh(\mathbf{W}\mathbf{X}^T) + \epsilon \quad (10)$$

where $\mathbf{X} \in \mathbb{R}^{n \times m}$ is a matrix in which each row represents the sample values of the input set. The vector $\mathbf{Y} \in \mathbb{R}^{k \times n}$, represents the corresponding output. $\mathbf{W} \in \mathbb{R}^{k \times m}$ is a transformation matrix in which k is the size of the hidden layer. It transforms each input feature to a space formed by hidden units. $\mathbf{W}' \in \mathbb{R}^{k \times k}$ is a matrix that forms the output from the hidden layer. Vector ϵ represents the error of estimation in comparison with target objectives. The hyperbolic tangent (\tanh) is an activation function. In order to use an FFNN as a universal function approximator, the FFNN activation function needs to be nonlinear and continuously differentiable. Moreover, stable training through gradient-descent algorithms needs a monotonic, finite range, and 0-neighborhood identity activation function. The \tanh satisfies all these properties and is widely chosen for FFNNs. Fig. 3 represents the structure of the nonlinear regressor. In the case of multiple outputs, ($\mathbf{Y} \subset \mathbb{R}^{k \times n}$, $k > 1$), we handle each output independently.

To find the best fitting model we perform the following optimization over the objective function:

$$\underset{\mathbf{W}, \mathbf{W}'}{\operatorname{argmin}} L(\mathbf{W}, \mathbf{W}') = \frac{1}{2} \|\mathbf{Y} - \mathbf{W}'\tanh(\mathbf{W}\mathbf{X}^T)\|_2^2. \quad (11)$$

The above optimization minimizes the prediction error of the model using all m parameters.

C. Learning Algorithm for Nonlinear Optimization

The Levenberg–Marquardt (LM) algorithm is used for learning the parameters of the FFNN [24]. LM benefits the steepest descent (SD) and Gauss-Newton (GN) algorithms to avoid finding local optimums. The LM algorithm combines the SD and GN in the following manner:

$$\mathbf{x}_k = \mathbf{x}_{k-1} - (\mathbf{J}_{k-1}^T \mathbf{J}_{k-1} + \mu \mathbf{I})^{-1} \mathbf{J}_{k-1}^T \mathbf{e} \quad (12)$$

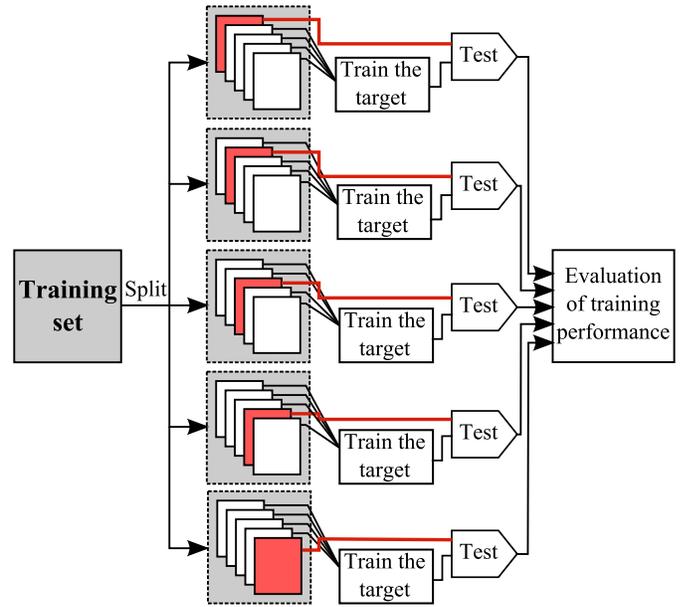


Fig. 4. Flow-chart of a fivefold cross-validation.

where \mathbf{J} is the Jacobian matrix contains first derivatives of the FFNN errors, \mathbf{e} is a vector of FFNN errors in the last step, \mathbf{x} is a vector of unknown parameters $w_{i,j}$, k_i , b_i that are obtained after training, and μ is a hyper parameter that offers a balance between SD and GN during the learning iterations. The Jacobian matrix can be computed via back-propagation technique that is much less complex than computing the Hessian matrix.

In Section III-F, we design a function to reward sparsity over input parameters and add that function to the above optimization. Thus, we can find the set of significant parameters that can predict the output precisely.

D. Validation via k -Fold Cross-Validation

Cross-validation is a frequently used method to avoid overfitting on training set and improve the quality of trained model [25]. In order to perform k -fold cross validation, the training set is randomly divided into k separate subsets of equal size. Then, training procedure is performed k times, each time discarding one set as a testing set, and the average error over all the runs is computed. Finally, the trained model with the lowest error is selected. This has the additional benefit of avoiding local optimums for the trained model.

Fig. 4 illustrates the an example of a fivefold cross-validation. Here, the training set is divided to five subsets. The training procedure iteratively is done five times. In each iteration, one of the subsets is selected as a testing set, and the remaining ones are exploited as a training set. The trained system with the lowest error is then selected. In our experiments we use tenfold cross-validation which is commonly used for the efficient training of the two layer FFNN.

E. Column-Wise Sparse Parameter Selection

Our proposed parameter reduction technique inspired by ℓ_1 -norm regularization method. If x represents one input

sample, we can reformulate the ℓ_1 -norm regularization loss function as the following:

$$\mathbf{Y} = \mathbf{W}' \tanh\left(\sum_{i=1}^m \mathbf{W}_i x_i\right) + \epsilon \quad (13)$$

in which the vector \mathbf{W}_i represents the column i of the matrix \mathbf{W} . The contribution of each input parameter corresponds to a column of the matrix \mathbf{W} . To select few number of parameters, we need to learn \mathbf{W} as a column-wise sparse matrix. If the matrix is column-wise sparse, it means that there are several columns of all zeros and the corresponding parameters do not have any contribution in the model. Consequently, the significant parameters are the ones with the corresponding non-zero columns.

To achieve the column-wise sparsity, we measure the sparsity on the vector consisting of the maximum of the columns: $\|\max(\mathbf{W}_1) \cdots \max(\mathbf{W}_m)\|_0$. If the entry with maximum value in a column is pushed toward zero, we expect all the other values in a column become zeros. It is a common practice to approximate norm-zero with norm-one to achieve a sparse answer while making the optimization easier. But, still the optimization is almost impossible because of the discrete max function applied on the columns. Big norm function provides a continuous approximation of the maximum function (infinity norm is equal to max). Therefore, we approximate the max function with the continues p -norm function ($p \geq 2$)

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^n |v_i|^p\right)^{\frac{1}{p}}. \quad (14)$$

We choose p large enough that achieves a column-wise sparse answer on a held-out data set.¹ Similarly in group lasso [26] combination of norm 1 and 2 is used to achieve a linear group-wise sparse model.

Fig. 5 schematically represents the concept of column-wise sparsity. The norm- p (p is selected reasonably big) is applied to \mathbf{W} in order to compute the maximum element of each column. Then, norm-one is applied to the vector of obtained values to impose the sparsity. Thus, the column-wise sparsity is measured by $\|\|\mathbf{W}_1\|_p \cdots \|\mathbf{W}_m\|_p\|_1$. In the following, we present how the column-wise sparsity is applied on an FFNN regressor to form a feature selector.

F. Nonlinear Column-Wise Sparse Parameter Selection

In order to find the reduced input set, the sparsity objective function is added to the regressor. Putting the FFNN regressor and column-wise sparsity together, the objective function becomes

$$\begin{aligned} \operatorname{argmin}_{\mathbf{W}, \mathbf{W}'} L(\mathbf{W}, \mathbf{W}') &= \frac{1}{2} \|\mathbf{Y} - \mathbf{W}' \tanh(\mathbf{W}\mathbf{X}^T)\|_2^2 \\ &+ \lambda \|\|\mathbf{W}_1\|_p \cdots \|\mathbf{W}_m\|_p\|_1. \end{aligned} \quad (15)$$

The first term of the objective function is called loss function and tries to minimize the error of regression. The second term

¹The held-out data set is referred to a set that is only used for the training or the evaluation purpose.

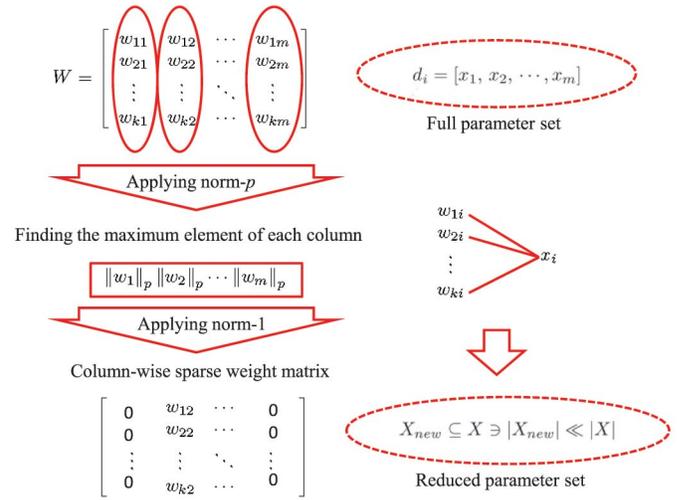


Fig. 5. Role of norm- p regularization in weight matrix for feature selection.

is called regularization term which controls the number of parameters in regression.

Feature selection can be used whenever the values of the \mathbf{W} and \mathbf{W}' are obtained. Algorithm 1 represents the steps of learning for the column-wise sparse feature selection method. In each iteration, the gradient of objective function is computed to update \mathbf{W} and \mathbf{W}' [Algorithm 1 (line 6)]. The algorithm continues either to reach the defined bound of error or to end at the maximum learning iterations [Algorithm 1 (line 3)]. Thus \mathbf{W} and \mathbf{W}' are learned during the training process. The λ and p are model hyper parameters. The λ value controls the number of parameters in the regression model. As the λ value increases, the objective function shrinks the weights in \mathbf{W} in a column-wise manner toward zero. Thus, the bigger λ value forces more parameters toward zero and reduces the parameter space.

G. Hyper Parameter Selection

The hyper parameters of the proposed method can be selected very efficiently in the following manner.

- 1) *The Value of Regularization Penalty (λ):* Binary search can be efficiently used to determine the λ value. When the λ value increases, the weights in the \mathbf{W} start to decrease until one of the columns becomes zeros. This means that the number of parameters remain constant for an interval of λ values. To find the desired λ value (which shows the number of parameters after reduction), we start from a big constant value. First the objective function (L in Algorithm 1) is computed to give us the number of remaining parameters. Next the objective function is recomputed with the value of $(\lambda/2)$. Similar to binary search, the desired interval for the λ value is selected. This procedure continues until we find an interval in which each λ value gives us the desired number of parameters.
- 2) *The Size of Hidden Layer:* This can be done with capacity saturation measurement of the target regressor. In

Algorithm 1: Nonlinear Multiobjective Parameter Selection

input : $\mathbf{x}_i = \{\text{Input vector}\}$, $\mathbf{y}_i = \{\text{Output vector}\}$, $\epsilon =$ Error bound, $\lambda =$ Regularization parameter, $\mathbf{M} =$ Maximum number of iterations

output: $\mathbf{W}, \mathbf{W}' =$ Matrix of transition weights

```

1: Initialize  $\mathbf{W}, \mathbf{W}'$ ;
2:  $Iter \leftarrow \emptyset$ ,  $E \leftarrow \emptyset$ ;
3: while  $|E| \geq \epsilon$  or  $Iter \leq M$  do
4:    $E \leftarrow \emptyset$ ;
5:   for  $i=1:n$  do
6:     Set the objective function
7:      $L \leftarrow \frac{1}{2} \|\mathbf{Y} - \mathbf{W}' \tanh(\mathbf{W}\mathbf{X}^T)\|_2^2 + \lambda \|\mathbf{W}_i\|_1$ ;
8:     Compute the error
9:     ( $E \leftarrow \frac{1}{2} \|\mathbf{Y} - \mathbf{W}' \tanh(\mathbf{W}\mathbf{X}^T)\|_2^2$ );
10:    Calculate the gradient of objective function in
11:    order to update weights;
12:     $\mathbf{W}, \mathbf{W}' \leftarrow$  Gradient-based optimization
13:    ( $\mathbf{W}, \mathbf{W}', \frac{\partial L}{\partial \mathbf{W}}, \frac{\partial L}{\partial \mathbf{W}'}$ );
14:    $Iter \leftarrow Iter + 1$ ;
15:    $E \leftarrow \frac{E}{n}$ ;
16: return  $\mathbf{W}$ ;
  
```

order to find the optimum number of nodes in the hidden layer, we performed the following procedure. In each iteration, hidden nodes are added incrementally to increase the FFNN learning capacity. This procedure continues until the performance of the FFNN on the testing set starts to decrease. After this step, increasing the number of hidden nodes causes the network to overfit the training set. Here, the training error is driven to a very small value, but when new data, as a testing set, is presented to the network the error becomes large. Fig. 6 depicts the error value of the network on the testing and the training sets for the ITC'99 b03 benchmark circuits. As shown in the figure, the minimum error on the testing set is achieved when hidden layer has 40 nodes ($err = 4.26 \times 10^{-6}$). The figure clearly depicts that how the performance of a larger network is degraded on the testing set.

IV. EXPERIMENTAL RESULTS

This section evaluates the proposed method by applying the column-wise feature selection to a set of combinational and sequential logic benchmark circuits in the context of regular and emerging technologies. The focus of our study is on the timing variation analysis. The use of this method is motivated by the lack of intuition that a skilled designer may have to identify the critical parameters of novel devices. We first look at FinFET as a cutting edge technology. Design verification in FinFET technology is complex because of the novel 3-D structure of the devices. We then look at a further interesting technology, silicon nanowires, that are also 3-D structures with specific features.

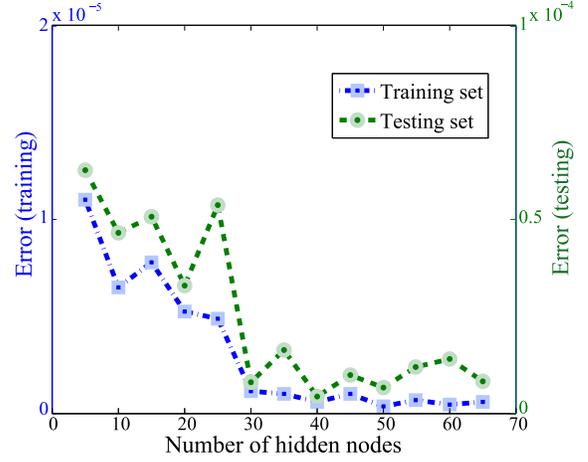


Fig. 6. Error of training and testing sets for a network with various number of hidden nodes.

A. PV Analysis for FinFET Technology

In this section we present the application of described method for timing variation of FinFET-based circuits.

1) *FinFET Technology*: FinFET technology is currently used as the cutting edge IC technologies owing to its remarkable scalability. FinFET as tree-dimensional structure uses a fin-shaped channel that efficiently providing more volume than a planar device for the same surface area. The device gate wraps around the channel that provides more efficient control over the channel. Thus, very little amount of leak current passes through the substrate when the device is in the off state. This, in turn, provides the opportunity of using lower threshold voltage leading to better performance and lower power consumption.

2) *Setup of Experiments*: To evaluate the proposed parameter selection technique, we exploit a number of combinational and sequential logic circuits from ITC'99 and ISCAS benchmarks. We study the timing variation of the longest path for each benchmark circuits. The longest path of each benchmark circuit is extracted using Synopsys PrimeTime [27] as exemplified in Fig. 7. In our analysis, the input parameter set includes the V_{Th} of each transistor within the circuit. Here, V_{Th} s are selected as they can significantly reflect the variation of physical parameters of each transistor on its performance. A transistor pool, which contains 5000 n -type and p -type transistors in 20 nm FinFET technology, is generated by applying a 5% Gaussian variation on the V_{Th} of each transistor. To build each circuit instance, the transistors are randomly selected from the pool and are added to the SPICE model of the target circuit. Using the obtained SPICE model, we can assess the timing variation through MC simulations but require tremendous amount of time. By applying the proposed parameter selection method, we show how this sampling space can be limited to the most important input parameters that mainly impact the timing of the circuits.

3) *Parameter Reduction and Simulation Speed-Up*: We performed 10000 MC simulations to extract the distribution of the delay for each benchmark by applying variation on all the parameters (the total number of parameters are listed in

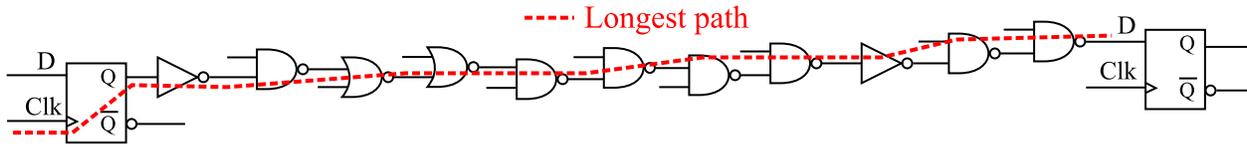


Fig. 7. Longest path in ITC'99 b03 benchmark.

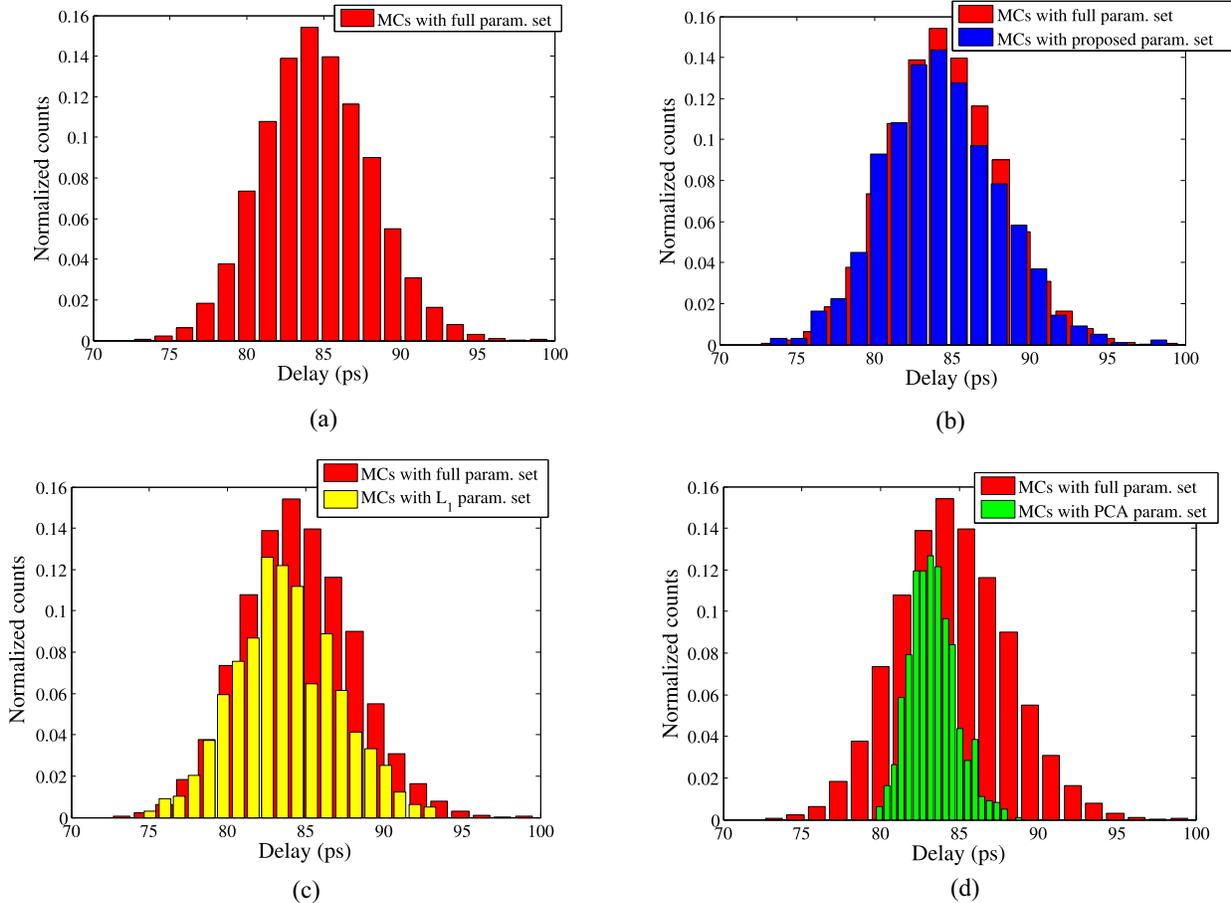


Fig. 8. Distribution of longest path delay for benchmark b03 before and after parameter selection. (a) Full parameter set. (b) Proposed method parameter set. (c) ℓ_1 -norm regularization parameter set. (d) PCA parameter set.

Table I). We then trained our parameter selection method over 1000 MC simulations to pick out the 20% most important parameters from each benchmark. In our method, the reduced set of input parameters is achievable through increasing the value of λ till reaching a desirable balance between the performance estimation accuracy and the number of selected input parameters. After extracting the important parameters, we perform 1000 SPICE simulations for each target benchmark by applying variations on the selected parameters. Table I demonstrates the mean and standard deviation of the longest path delay for each benchmark before and after parameter reduction. The results reveal average errors of 1.2% and 3.2% on the mean and the standard variation values, respectively.

The distributions of the longest path delay for ITC b03 benchmark are depicted in Fig. 8 before and after parameter reduction. The distribution in Fig. 8(a) is obtained through 10000 MC simulations over 88 variation parameters.

We reduced the size of the input parameter space using proposed parameter selection method, and two baselines ℓ_1 -Norm regularization and PCA. We did not perform our experiments using ICA and CCA baseline methods since ICA failed to find the reduced input parameters having Gaussian distributions, and ℓ_1 -norm regularization surpasses the CCA method [28]. The distributions in Fig. 8(b)–(d) are attained through 1000 MC simulations but using only 17 parameters selected by three mentioned parameter selection methods. The variation sampling using the most relevant parameters, obtained by our method, is capable of estimating the timing variation distribution with less amount of error ($\sigma = 3.68$ ps as compared to $\sigma = 3.54$ ps) compared to ℓ_1 -Norm regularization ($\sigma = 3.30$ ps as compared to $\sigma = 3.54$ ps) and PCA ($\sigma = 1.48$ ps as compared to $\sigma = 3.54$ ps). Therefore, the proposed parameter selection can efficiently reproduce the timing variation with a small subset of input parameters.

TABLE I
COMPARISON OF MEAN AND VARIANCE OF VARIOUS ITC'99 AND ISCAS BENCHMARKS ON THE DELAY VARIATION OF LONGEST PATH BEFORE AND AFTER APPLYING PROPOSED PARAMETER REDUCTION

	Benchmark	# of param.	Full-parameter set		Reduced-parameter set			
			Mean	Std	Mean	Std	Mean-error (%)	Std-error (%)
ITC'99	b01	74	5.372e-11	3.160e-12	5.335e-11	3.208e-12	0.69	1.52
	b02	80	5.506e-11	2.980e-12	5.457e-11	2.955e-12	0.88	0.86
	b03	88	8.451e-11	3.547e-12	8.429e-11	3.686e-12	0.26	3.91
	b04	116	1.237e-10	4.156e-12	1.226e-10	3.847e-12	0.84	7.44
	b05	96	9.031e-11	3.707e-12	8.970e-11	3.615e-12	0.68	2.49
	b06	76	6.611e-11	3.113e-12	6.528e-11	2.894e-12	1.25	7.05
ISCAS	C432	238	2.687e-10	1.529e-11	2.660e-10	1.433e-11	0.99	6.28
	C880	242	2.424e-10	7.569e-12	2.398e-10	7.262e-12	1.08	4.06
	S386	82	7.544e-11	7.733e-12	7.480e-11	3.722e-12	0.86	0.30
	S641	172	1.485e-10	1.369e-11	1.480e-10	1.272e-11	0.32	7.09

TABLE II
COMPARISON OF THE RUNTIME OF THE TIMING ANALYSIS FOR THE ITC'99 b03 BENCHMARK WITH AND WITHOUT PARAMETER REDUCTION (CPU: DUAL-XEON X5650, MEMORY: 24 GB)

	Without parameter reduction	with parameter reduction
Spice simulation per iteration	29 sec	29 sec
Total runtime	80.5 (hrs) [for 10000 MC simulation (80.5 hrs)]	16.38 (hrs) [for 2000 MC simulation (16.11hrs) + training (27 min)]

TABLE III
COMPARISON OF THE NUMBER OF THE ACCURACY VERSUS NUMBER OF PARAMETERS FOR THE ITC'99 b03

# of param.	benchmark ITC'99 b03			
	Mean	Std	Mean-error(%)	Std-error(%)
17	8.429e-11	3.686e-12	0.26	3.91
18	8.471e-11	3.659e-12	0.23	3.16
19	8.433e-11	3.450e-12	0.21	2.72
20	8.467e-11	3.446e-12	0.20	2.87
21	8.466e-11	3.451e-12	0.18	2.69
22	8.465e-11	3.452e-12	0.18	2.66

The number of required MC simulations is reduced by $5\times$ (2000 simulations for training and reproducing the distribution of timing variation versus 10000 MC simulations without parameter selection). Table II compares the runtime of finding the longest path delay for the ITC'99 b03 benchmark circuit with and without parameter reduction. Here, the parameter reduction result in $4.9\times$ speed up when it is compared with MC simulations.

In order to show the stability of the parameter reduction, we also repeated the experiment by selecting different number of parameters. Table III clearly shows that the prediction error is reduced when a larger set of parameters contribute in the longest-path delay estimation. Here, the increase in the std error with 20 parameters is due to the fact that the experiments for finding the distribution for each row is performed with a new random value for the parameters. So, it is possible that to see such an increase in the std value or even in the mean values.

B. PV Analysis for Double-Gate Silicon Nanowire Technology

Finally, we demonstrate the result of parameter selection for the variation analysis of a benchmark circuit in double-gate silicon nanowire technology.

1) *Double-Gate Silicon Nanowire Technology:* DG-SiNWFET technology is considered as a potential candidate for current CMOS technology thanks to its 1-D properties, lower short channel effect, and lower leakage [2]. DG-SiNWFETs are double independent gate devices whose polarity can be dynamically configured between n - and p -type through an additional terminal, called polarity gate (PG) [2]. In-field polarity reconfiguration property is interestingly used to realize compact exclusive or-based circuits [29]. Fig. 9 summarizes the geometrical structure of the DG-SiNWFET as well as the constructive device parameters, used in a TCAD model description. Fig. 10(a) also illustrates the different in-field reconfigurations of the device polarity. The p -type and n -type are realized by fixing the PG bias to system ground ("0") and V_{dd} ("1"), respectively.

To perform variation analysis, we first characterize a population of devices by TCAD simulation using a 30% Gaussian variation on each geometrical parameter ($\sigma = 30\%$). In our case study, 2500 3-D TCAD simulations were performed to provide statistical information of the DG-SiNWFET device. Fig. 11 depicts the distinctive analytical metrics of the device such as I_{on} , I_{off} , V_{Th} , and SS. Only the distribution of V_{Th} can be approximated by a Gaussian distribution contrary to the remaining metrics. This result highlights that the distributions

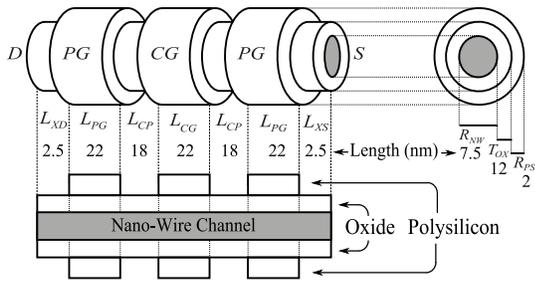


Fig. 9. DG-SiNWFET structure with related parameters.

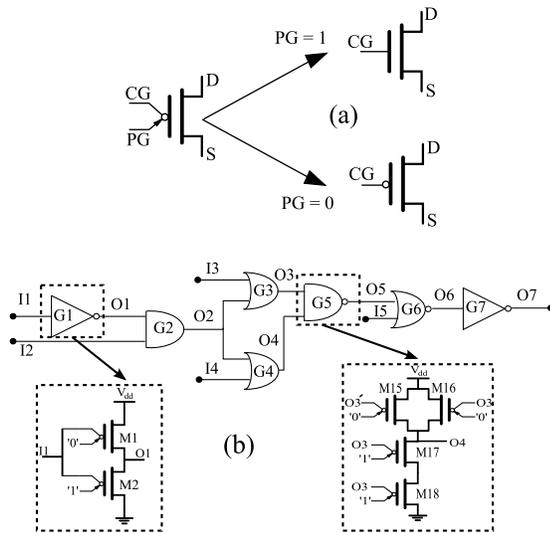


Fig. 10. (a) Use of DG-SiNWFET polarity control. (b) ISCAS89 benchmark circuit s27 using DG-SiNWFET technology.

of all parameters are not necessarily Gaussian. Thus, the distribution free parameter reduction techniques are required for future technologies.

2) *Setup of Experiments*: For evaluating the proposed parameter selection method, the small size benchmark circuit ISCAS89-s27 is selected as a case study. Without loss of generality, the method can be used for any other circuits. The main reason to select such a small size circuit is the long computation time of the TCAD simulations to produce the DG-SiNWFET device data set due to the lack of a mature compact model. In other technologies, compact models can be used to accelerate the data set generation. The schematic of the circuit is shown in Fig. 10(b). All the gates use DG-SiNWFET transistors. The PG of each transistor is appropriately configured to provide the correct functionality in the pull-up and pull-down of the gates. The considered circuit is comprised of 30 transistors leading to 300 geometrical parameters. Normal MC simulation to evaluate the performance variation requires tremendous amount of time, considering that no intuitions on the fundamental parameters can be done in the context of unconventional device mechanisms. By applying the proposed method, we show how this sampling space can be restricted to the main parameters that considerably affect the performance of the circuit.

Among various performance metrics, we select the delay of circuit to form the output set. For the sake of keeping a

reasonable complexity for the experiments, a reduced subset of geometrical parameters of the transistors (50 parameters) is randomly considered as the input set. Here, the goal is to determine how much the parameter reduction can improve the circuit performance evaluation, while the estimation error is bounded by a certain threshold.

To simulate the characteristics of the target circuit, the obtained $I - V$ curve of the transistors, are injected in a Verilog-A table model. This model is run with HSPICE to perform the MC simulations for the timing analysis purpose.

3) *Parameter Reduction and Simulation Speed-Up*: After applying column-wise sparse parameter selection, we can reduce the number of parameters to improve the computational complexity of the simulations. Decreasing the number of parameters can be obtained by increasing the λ value which results in larger delay estimation error. In this case, the performance of the circuit can be evaluated with a smaller number of parameters which really contribute to the MC simulations, but results in a higher performance estimation error. The capability of bounding the error by changing the numbers of parameters enables the designers to tradeoff evaluation precision with computation complexity. In our case study, reducing the number of parameters to 10 (from 50) is obtained with the variance of delay estimation error of 11.7%.

We compared the proposed technique with PCA as a well-known parameter reduction methods for estimating the delay of ISCAS89-s27. For PCA, 20% of the new features were selected according to their highest eigenvalues. To be able to perform the MC simulations without any change in the underlying model or simulator, the reverse of these transformations are applied to produce the exact values of the input space parameters. In our method, λ value was tuned to select the same number of parameters in input space. Using reduced input parameter sets obtained by PCA and the proposed method, we performed 1000 MC simulations for each set to estimate the delay distribution of ISCAS89-s27. The proposed method shows a better performance compared to its competitor with lower variance of delay estimation error (11.7% versus 13.5%).

To verify the accuracy and the performance improvement of doing such reduction, we evaluate the delay of the target circuit in the presence of variations. We perform the MC simulations in both cases of reduced and nonreduced input parameter set with 10 and 50 parameters, respectively. Fig. 12 represents the probability density function of the ISCAS89-s27 delay in both cases. The figure depicts a high correlation between two sets. We observe that the proposed column-wise sparsity is able to estimate the major parameters for delay variation analysis with tiny amount of error on each test samples ($\sigma = 8.91$ ps as compared to $\sigma = 10.10$ ps leading to a variance error of 11.7%). Thus, the method is able to efficiently evaluate the delay variation of the circuit, while reducing the number of parameters. A reduced input set results in less MC simulations which is very critical in the case of execution time. As we used 100 random samples for each parameter, the parameter reduction reduces the number of required MC runs by $2.5\times$ (5000 simulations without feature selection versus 2000 simulations for training and feature selection).

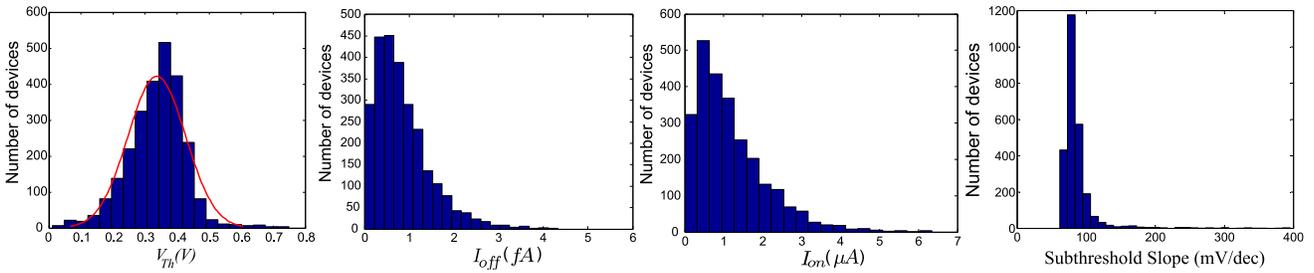


Fig. 11. Distribution of V_{Th} , I_{off} , I_{on} , and SS for DG-SiNWFET ($\sigma = 30\%$ for structural parameters). Only the variation of V_{Th} follows a Gaussian distribution.

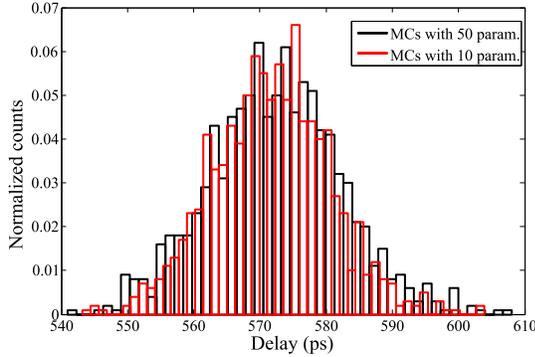


Fig. 12. Delay distribution comparison of the full and the reduced parameter models.

V. CONCLUSION

We introduced an efficient parameter selection method which can be used for performance evaluation of the emerging technologies like silicon nanowires. Using this method, we are able to accurately evaluate the PVs while reducing the computation complexity by utilizing the obtained reduced parameter set. This method is based on FFNN regression, and employs column-wise sparsity to reduce the size of parameters space. Unlike the widely used feature reduction methods, this method is able to take to account the mixed Gaussian and non-Gaussian parameters. Moreover, it considers the nonlinear dependencies between input parameters and outputs which lead to effective parameter reduction. We applied this method to a couple of FinFET-based combinational and sequential benchmarks from ITC'99 and ISCAS to study the variation of delay of the longest path for each circuit. In this case, experimental results show $5\times$ speed up and estimate the delay distribution with the average variance error of 4.1% in presence of 5% variation on each parameter. Applied to ISCAS89-s27 benchmark exploiting DG-SiNWFET technology as well, experimental results show $2.5\times$ speed up in timing analysis and estimation of the delay distribution with the variance error of 11.7% in presence of 30% variation on each parameter.

REFERENCES

- [1] S. Bangsaruntip *et al.*, "High performance and highly uniform gate-around silicon nanowire MOSFETs with wire size dependent scaling," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Baltimore, MD, USA, 2009, pp. 1–4.
- [2] M. De Marchi *et al.*, "Polarity control in double-gate, gate-around vertically stacked silicon nanowire FETs," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2012, pp. 8.4.1–8.4.4.
- [3] Y.-M. Lin, J. Appenzeller, J. Knoch, and P. Avouris, "High-performance carbon nanotube field-effect transistor with tunable polarities," *IEEE Trans. Nanotechnol.*, vol. 4, no. 5, pp. 481–489, Sep. 2005.
- [4] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nat. Mater.*, vol. 6, no. 3, pp. 183–191, 2007.
- [5] Z. Feng, P. Li, and Y. Zhan, "An on-the-fly parameter dimension reduction approach to fast second-order statistical static timing analysis," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 1, pp. 141–153, Jan. 2009.
- [6] R. Wang *et al.*, "Investigation on variability in metal-gate Si nanowire MOSFETs: Analysis of variation sources and experimental characterization," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2317–2325, Aug. 2011.
- [7] N. Moezi, D. Dideban, B. Cheng, S. Roy, and A. Asenov, "Impact of statistical parameter set selection on the statistical compact model accuracy: BSIM4 and PSP case study," *Microelectron. J.*, vol. 44, no. 1, pp. 7–14, 2013.
- [8] W. Hong *et al.*, "A novel dimension-reduction technique for the capacitance extraction of 3-D VLSI interconnects," *IEEE Trans. Microw. Theory Techn.*, vol. 46, no. 8, pp. 1037–1044, Aug. 1998.
- [9] Y. Zhan *et al.*, "Correlation-aware statistical timing analysis with non-Gaussian delay distributions," in *Proc. Design Autom. Conf. (DAC)*, Anaheim, CA, USA, 2005, pp. 77–82.
- [10] C. Wisweswariah *et al.*, "First-order incremental block-based statistical timing analysis," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 25, no. 10, pp. 2170–2180, Oct. 2006.
- [11] B. Cheng *et al.*, "Statistical-variability compact-modeling strategies for BSIM4 and PSP," *IEEE Design Test Comput.*, vol. 27, no. 2, pp. 26–35, Mar./Apr. 2010.
- [12] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, 2003, pp. 900–907.
- [13] Z. Feng and P. Li, "Performance-oriented parameter dimension reduction of VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 1, pp. 137–150, Jan. 2009.
- [14] H. G. Mohammadi, P.-E. Gaillardon, M. Yazdani, and G. De Micheli, "Fast process variation analysis in nano-scaled technologies using column-wise sparse parameter selection," in *Proc. IEEE/ACM Int. Symp. Nanoscale Archit. (NANOARCH)*, Paris, France, 2014, pp. 163–168.
- [15] K. Chopra, C. Zhuo, D. Blaauw, and D. Sylvester, "A statistical approach for full-chip gate-oxide reliability analysis," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, San Jose, CA, USA, 2008, pp. 698–705.
- [16] D. S. Boning and P. K. Mozumder, "DOE/Opt: A system for design of experiments, response surface modeling, and optimization using process and device simulation," *IEEE Trans. Semicond. Manuf.*, vol. 7, no. 2, pp. 233–244, May 1994.
- [17] D. Blaauw, K. Chopra, A. Srivastava, and L. Scheffer, "Statistical timing analysis: From basic principles to state of the art," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 4, pp. 589–607, Apr. 2008.
- [18] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [19] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, 2000.

- [20] L. Cheng, P. Gupta, and L. He, "Accounting for non-linear dependence using function driven component analysis," in *Proc. Asia South Pac. Design Autom. Conf. (ASP-DAC)*, Yokohama, Japan, 2009, pp. 474–479.
- [21] R. Huang *et al.*, "Variability investigation of gate-all-around silicon nanowire transistors from top-down approach," in *Proc. IEEE Int. Conf. Electron Devices Solid-State Circuits (EDSSC)*, Hong Kong, 2010, pp. 1–4.
- [22] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
- [23] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
- [24] J. J. Moré, "The Levenberg–Marquardt algorithm: Implementation and theory," in *Numerical Analysis*. Heidelberg, Germany: Springer, 1978, pp. 105–116.
- [25] M. Stone, "Cross-validation: A review 2," *Statist. J. Theor. Appl. Stat.*, vol. 9, no. 1, pp. 127–139, 1978.
- [26] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *J. Comput. Graph. Stat.*, vol. 22, no. 2, pp. 231–245, 2013.
- [27] (Feb. 2015). *Synopsys PrimeTime*. [Online]. <http://www.synopsys.com>
- [28] Y. Zhang, S. Sankaranarayanan, and F. Somenzi, "Sparse statistical model inference for analog circuits under process variations," in *Proc. Asia South Pac. Design Autom. Conf. (ASP-DAC)*, Singapore, 2014, pp. 449–454.
- [29] M. H. Ben-Jamaa, K. Mohanram, and G. De Micheli, "An efficient gate library for ambipolar CNTFET logic," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 2, pp. 242–255, Feb. 2011.



Pierre-Emmanuel Gaillardon (S'10–M'11) received the Electrical Engineer degree from CPE-Lyon, Lyon, France, in 2008, the M.Sc. degree in electrical engineering from INSA Lyon, Lyon, in 2008, and the Ph.D. degree in electrical engineering from CEA-LETI, Grenoble, France, and the University of Lyon, Lyon, in 2011.

He was a Research Associate with the Laboratory of Integrated Systems (Prof. G. D. Micheli), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He is an Assistant Professor with the Electrical and Computer Engineering Department, University of Utah, Salt Lake City, UT, USA, where he leads the Laboratory for NanoIntegrated Systems. He is a Visiting Research Associate with Stanford University, Palo Alto, CA, USA. He was a Research Assistant with CEA-LETI. His current research interests include development of reconfigurable logic architectures and digital circuits exploiting emerging device technologies, and novel EDA techniques.

Prof. Gaillardon was a recipient of the C-Innov 2011 Best Thesis Award and the Nanoarch 2012 Best Paper Award. He is an Associate Editor of the IEEE TRANSACTIONS ON NANOTECHNOLOGY. He has served as a TPC Member for several conferences, including DATE'15–16, DAC'16, Nanoarch'12–16, and is a Reviewer for several journals and funding agencies.



Giovanni De Micheli (S'79–M'79–SM'80–F'94) received the Nuclear Engineer degree from the Politecnico di Milano, Milan, Italy, in 1979, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California at Berkeley, Berkeley, CA, USA, in 1980 and 1983, respectively.

He was a Professor of Electrical Engineering with Stanford University, Stanford, CA, USA. He is currently a Professor and the Director of the Institute of Electrical Engineering with the Integrated Systems Centre, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, where he is also the Program Leader of the Nano-Tera.ch Program. He has authored or co-authored over 700 papers in journals and conferences, and a book entitled *Synthesis and Optimization of Digital Circuits* (McGraw-Hill, 1994), and co-authored and co-edited eight other books. He has an H-index of 87 citations (Google Scholar). His current research interests include design technologies for ICs and systems, such as synthesis for emerging technologies, networks on chips, 3-D integration, and heterogeneous platform design, including electrical components and biosensors, and data processing of biomedical information.

Prof. De Micheli was a recipient of the 2012 IEEE/Circuits and Systems Society (CAS) Mac Van Valkenburg Award for contributions to theory, practice, and experimentation in design methods and tools, the 2003 IEEE Emanuel Piore Award for contributions to computer-aided synthesis of digital systems, the Golden Jubilee Medal for outstanding contributions to the IEEE CAS Society in 2000, the D. Pederson Award for the best paper in the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS in 1987, and several Best Paper Awards, including the Design Automation Conference (DAC) in 1983 and 1993, the DATE in 2005, and the NANOARCH in 2010 and 2012. He has been with IEEE in several capacities, including as the Division 1 Director from 2008 to 2009, the Co-Founder and the President Elect of the IEEE Council on Electronic Design Automation from 2005 to 2007, the President of the IEEE Circuits and Systems Society in 2003, and the Editor-in-Chief of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS from 1987 to 2001. He has been the Chair of several conferences, including DATE since 2010, pHealth since 2006, International Conference on Very Large Scale Integration since 2006, DAC since 2000, and the International Conference on Computer Design since 1989. He is a fellow of ACM and a member of the Academia Europaea and Scientific Advisory Board of imec and STMicroelectronics.



Hassan Ghasemzadeh Mohammadi (S'16) received the B.Sc. degree in computer engineering from the Iran University of Science and Technology, Tehran, Iran, in 2005, and the M.Sc. degree in computer engineering from the Sharif University of Technology, Tehran, in 2008. He is currently pursuing the Ph.D. degree in computer science with the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

In 2011, he joined the École Polytechnique Fédérale de Lausanne. His current research interests include interesting problems in fault-tolerant design, circuit testing, and reliability.