Fast Process Variation Analysis in Nano-Scaled Technologies Using Column-Wise Sparse Parameter Selection

Hassan Ghasemzadeh Mohammadi*, Pierre-Emmanuel Gaillardon*, Majid Yazdani†, Giovanni De Micheli*

Integrated Systems Laboratory*, CLCL Laboratory†

École Polytechnique Fédérale de Lausanne* (EPFL), University of Geneva†

Switzerland

Email: hassan.ghasemzadeh@epfl.ch

Abstract- With growing concern about process variation in deeply nano-scaled technologies, parameterized device and circuit modeling is becoming very important for design and verification. However, the high dimensionality of parameter space is a serious modeling challenge for emerging VLSI technologies, where the models are increasingly more complex. In this paper, we propose and validate a feature selection method to reduce the circuit modeling complexity associated with high parameter dimensionality. Despite the commonly used methods such as *Principal Component Analysis* (PCA) and *Independent Component Analysis* (ICA), this method is capable of dealing with mixed Gaussian and non-Gaussian parameters, and performs a parameter selection in the input space rather than creating a new space. By considering non-linear dependencies among input parameters and outputs, the method results in an effective parameter selection. The application of this method is demonstrated in digital circuit timing analysis to effectively reduce the number of simulations. The experimental results on *Double-Gate Silicon NanoWire* FET (DG-SiNWFET) technology indicate $2.5\times$ speed up in timing variation analysis of the ISCAS89-s27 benchmark with a controlled average error bound of 9.4%.

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), *Carbon Nanotube* FETs (CNTFETs) and Graphene-based devices exhibiting short-channel effect immunity, greater electrostatic control, and lower leakage [1], [2], [3]. However, fabrication-induced *Process Variations* (PVs) on device and circuit characteristics are a growing challenge with ongoing feature size downscaling. Geometric and physical parameter variations, e.g., changes in transistor effective gate-length and *Threshold Voltage* (V_{th}), lead to considerable effects on performance and reliability of modern *Integrated Circuits* (ICs). Moreover, the output sensitivity on each parameter can vary from a technology to the next. For example, the V_{th} fluctuation in 16nm FinFET, due to the variation in effective oxide thickness, is $2\times$ less than that of the 16nm bulk CMOS [4]. 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 only realistic approach for comprehensive study of variation impacts for both circuit static timing and leakage power. 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. Utilizing compact models, 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. In the lack of mature compact models for emerging nano-devices, *Technology Computer Aided Design* (TCAD) modeling has been exploited to predict the impacts of fluctuations on device performance [6]. The results of TCAD simulation can be fed to SPICE-like simulator for MC simulation 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 [7], [8], [9]. Most of the existing techniques such as *Principal Component Analysis* (PCA) and *Independent Component Analysis* (ICA) use a linear transformation to decorrelate the parameter space [10], [11]. In spite of their popularity, they are inherently limited because they only consider the relations among the input parameters and ignore the impact of each input on the circuit outputs. This limitation becomes important when either some critical parameters, i.e., that 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 [12]. Therefore, they may lead to a large set of correlated parameters while the input space can be compressed by considering inter-parameter 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 multi-objective parameter selection method capable of addressing the aforementioned limitations. This method takes into account the inter-set (among inputs) and intra-set (between input and output sets) correlations. The loss 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:

- High precision by considering non-linear dependencies between inter-set and intra-set parameters.
- Distribution free feature selection which can be used for any models or parameter sets with unknown statistical distributions.
- 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 analyses of complex design where building-block parameterized models are built with a smaller set of statistically significant parameters.

To validate the technique, we use *Double-Gate Silicon Nanowire* FETs (DG-SiNWFETs) technology as a strong potential substitute for future silicon technologies [2]. The simulation results for 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 *Monte Carlo* (MC) is obtained for timing variation analysis with the average error bound of 9.4%.

The organization of this paper is as follows. Section II describes the motivation and background. Section III explains the proposed methodology for fast variation analysis, including a non-linear learning-based sparse parameter selection technique. Section IV validates the method using simulations, and finally Section V concludes the 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 inter-die and intra-die variations dramatically grows [13]. 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* (RSM) technique is not applicable anymore because the complexity of the model is exponential with respect to the number of parameters [14]. Fortunately, all of these parameters are not independent and therefore they can be partitioned to several sets of correlated parameters. By considering the correlation among parameters of each set and understanding the contribution of each set

on the output, it is possible to substantially reduce the model complexity by selecting the most statistically significant parameters. Thus, new methodologies are required to reduce the number of variables while keeping the estimation error fairly small.

To control the number of parameters, various feature selection and reduction techniques have been used by the research community [12]. *Principal Component Analysis* (PCA) has been widely used in the field of device compact modeling [10] and statistical static timing analysis [15]. The PCA performs a linear transformation through the conversion of correlated parameters into a smaller set of new uncorrelated parameters, called principal components. Then, the principal components, which have the maximum variations in the parameter space, are selected. 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. Moreover, the maximum performance can be obtained when the distribution of input parameters is Gaussian [19]. In contrary to PCA, *Independent Component Analysis* (ICA) is used for feature reduction of non-Gaussian parameters. However, when all the parameters follow the Gaussian distribution, ICA fails to find the constructive components [22]. Both mentioned methods are 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.

As an output sensitive statistical method, *Reduced Ranked Regression* (RRR) is capable of reducing the parameters which have major impacts on the output. Similar to previous methods, RRR strictly requires a Gaussian distribution of input variables to significantly enhance the result 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 [17]. Thus, the performance of feature selection may be considerably affected by the distribution of input parameters. Furthermore, in RRR 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 [22].

The methods mentioned above are linear, considering the non-linear dependencies can considerably increase the precision of parameter reduction. Many modifications have been proposed to alleviate this problem, e.g., *Function Driven Component Analysis* (FCA), quadratic RRR, Kernel PCA, and Kernel ICA [22]. These methods perform dimensionality reduction, means that the problem is transformed from an input parameter space to a reduced parameter space. To be able to use these methods in combination with PV simulators, either we need to reconstruct the original parameters from the reduced parameters, or modify the simulator to work with the new set of parameters (in reduced space). Modifying device and process simulators like TCAD simulators is very challenging. Moreover, due to the non-linearity of these transformations, it is not possible to reconstruct the original

parameters from the lower dimension space. Therefore, while the above non-linear methods increase the precision, but can not be used efficiently in our application.

To overcome the mentioned issues, we propose a feature selection method to accelerate statistical PV analysis. This method 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 in context of VLSI modeling and simulation. We overview the frameworks of parameter selection, and then discuss the method in detail.

A. Parameter Reduction towards 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 feasibly performed. After extracting the variation parameters, logic cells are clustered to form the initial large design with the new reduced parameter set. 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. 1 illustrates the general flow of the proposed parameter reduction 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: compact model parameters of the device, parameters of the TCAD model, or 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 statistical multivariate regression is used to predict the relation among the input and output parameter sets. The objective function of the regression is modified to minimize the error of output prediction while discarding the unnecessary parameters. Finally, the most significant parameters are only considered for the PV analysis of the target circuit, whereby increasing the evaluations speed.

B. Feed Forward Neural Network Regression

Feed Forward Neural Network (FFNN) is a powerful non-linear regressor known to be a universal approximator by increasing the size of hidden layer [18]. We adopt FFNN here as our regressor to consider the non-linear relations among the
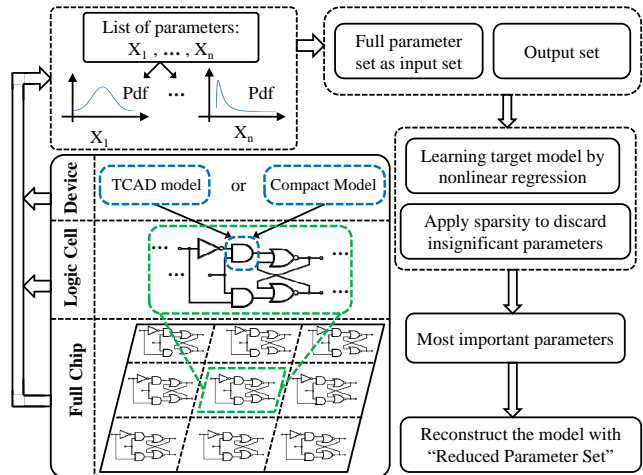

Fig. 1. General flow of the parameter reduction

parameters. The regression model is formulated as:

$$\mathbf{Y} = \mathbf{W}'tanh(\mathbf{WX}^T) + \epsilon \tag{1}$$

where $\mathbf{X}$ is a $n \times m$ matrix ($\mathbb{R}^{n \times m}$) in which each row represents the sample values of the input set. The vector $\mathbf{Y}$, of length n ($\mathbb{R}^{1 \times n}$), represents the corresponding output. $\mathbf{W}$ is a $k \times m$ transform matrix in which k is the size of the hidden layer. It transforms each input feature to a space formed by hidden units. $\mathbf{W}'$ is $1 \times k$ matrix that forms the output from the hidden layer. Vector ϵ represents the error of estimation in comparison with target objectives and $tanh$ is a non-linear activation function that is chosen conventionally.

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

$$\underset{\mathbf{W},\mathbf{W}'}{\operatorname{argmin}} L(\mathbf{W}, \mathbf{W}') = \frac{1}{2} \left\| \mathbf{Y} - \mathbf{W}'tanh(\mathbf{WX}^T) \right\|_2^2 \tag{2}$$

The above optimization minimizes the prediction error of the model using all m parameters. In the next step, we design a function to reward sparsity of used parameters and add that function to the above optimization. Thus, we can find the set of significant parameters that can predict the output precisely.

C. Column-Wise Sparse Parameter Selection

The contribution of each input parameter is in the columns of the matrix $\mathbf{W}$. If x represents one input sample, we can reformulate the above regression as the following:

$$\mathbf{Y} = \mathbf{W}'tanh(\sum_{i=1}^{m} \mathbf{W}_i x_i) + \epsilon \tag{3}$$

in which the vector $\mathbf{W}_i$ represents the column i 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
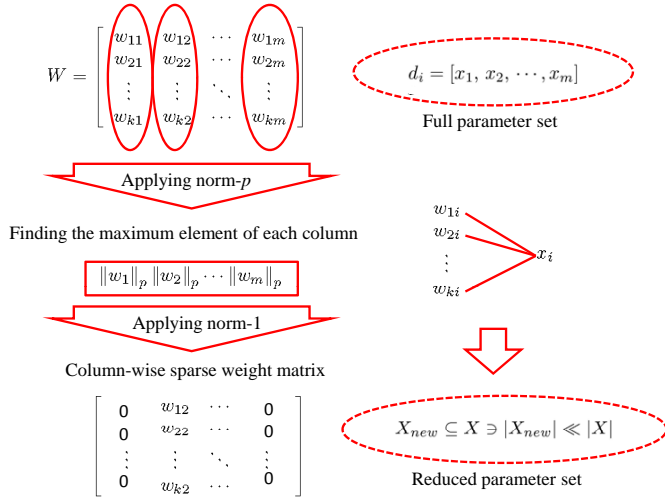
Fig. 2. The role of norm-p regularization in weight matrix for feature selection

Algorithm 1: Non-linear Multi-Objective 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	: W,W' = Matrix of transition weights

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

columns: $\|max(\mathbf{W}_1) \cdots max(\mathbf{W}_m)\|_0$. If the entry with maximum value in a column is pushed towards 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 ($p \geq 2$):

$$\|\mathbf{v}\|_p = (\sum_{i=1}^{n} |v_i|^p)^{\frac{1}{p}} \qquad (4)$$

We choose p large enough that achieves a column-wise sparse answer on a held-out data set. Similarly in group lasso [20] combination of norm 1 and 2 is used to achieve a linear group-wise sparse model.

Fig. 2 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.

D. Non-linear 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 loss function becomes:

$$\operatorname*{argmin}_{\mathbf{W},\mathbf{W}'} L(\mathbf{W}, \mathbf{W}') = \frac{1}{2}\left\|\mathbf{Y} - \mathbf{W}'tanh(\mathbf{WX}^T)\right\|_2^2 \qquad (5)$$
$$+\lambda\|\|\mathbf{W}_1\|_p \cdots \|\mathbf{W}_m\|_p\|_1$$

The first term of the objective function is called *least square error* and tries to minimize the error of regression. The second term 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 - l. 6). The algorithm continues either to reach the defined bound of error or to end at the maximum learning iterations (Algorithm 1 - l. 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 towards zero. Thus, the bigger λ value forces more parameters toward zero and reduces the parameter space.

IV. EXPERIMENTAL RESULTS

This section evaluates the proposed method by applying the column-wise feature selection to a combinational logic benchmark circuit in the context of 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 whose switching mechanisms are non-standard.

A. Target Technology

Double-Gate Silicon Nanowire FET (DG-SiNWFET) technology is considered as a potential candidate for current CMOS technology thanks to its 1D properties, lower *Short Channel Effect* (SCE), and lower leakage [2]. DG-SiNWFETs are *Double Independent Gate* (DIG) 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 XOR-based circuits [21]. Fig. 3 summarizes

the geometrical structure of the DG-SiNWFET as well as the constructive device parameters, used in a TCAD model description. Fig. 4a 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 GND ('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. 5 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.
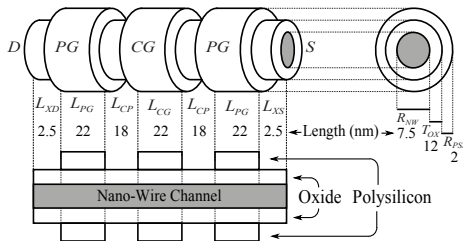

Fig. 3. DG-SiNWFET structure with related parameters

B. Setup of the Experiments

Circuit simulations are performed using the combinational logic ISCAS89 benchmark circuit s27. The schematic of the circuit is shown in Fig. 4b. 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 a tremendous amount of simulations, 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.

C. 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 freedom of tuning error by changing the number of parameters enables designers to compromise between evaluation precision and computational complexity. In our case, reducing the number of parameters to 10 (from 50) is obtained for a corresponding λ value of 8.957×10^3 with the average error bound of 9.4%.

In Table 1, the proposed technique is compared with other well-known feature reduction methods for estimating the delay of ISCAS89-s27. For PCA, ICA, and RRR, 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. Moreover, λ value was tuned to select the same number of parameter in input space. The value of entries in second and third columns denote the mean and variance of the delay estimation error for 1000 MC simulations on ISCAS89-s27 respectively. The proposed method shows a better performance rather than its competitors with lower mean and variance of delay estimation error (9.4% and 11.7%). To verify the accuracy and the performance

TABLE I
ERROR COMPARISON OF VARIOUS FEATURE REDUCTION METHODS IN CASE OF DELAY

	Average error	Error Variance
PCA	11.2%	13.5%
ICA	10.8%	13.1%
RRR	10.1%	12.3%
Proposed method	9.4%	11.7%

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 case of reduced and non-reduced input parameter set with 10 and 50 parameters respectively. Fig. 6 represents the *Probability Density Function* (PDF) of the ISCAS89-s27 delay in both case. The figure depicts a high correlation between the 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 ($\sigma = 8.9ps$ as compared to
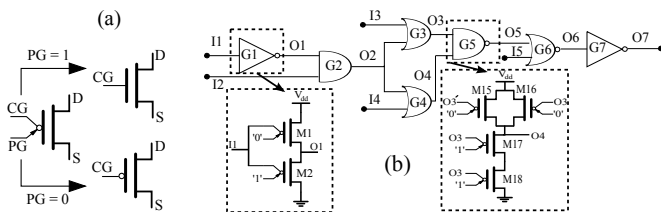

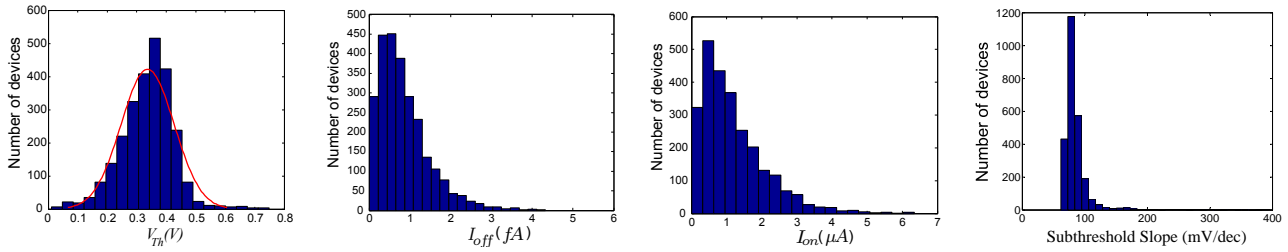Fig. 4. Use of DG-SiNWFET polarity control (a); ISCAS89 benchmark circuit s27 using DG-SiNWFET technology (b)

Fig. 5. Distribution of V_{Th}, I_{off}, I_{on}, and SS for DG-SiNWFET (σ=30% for structural parameters). Only the variation of V_{Th} follows a Gaussian distribution.
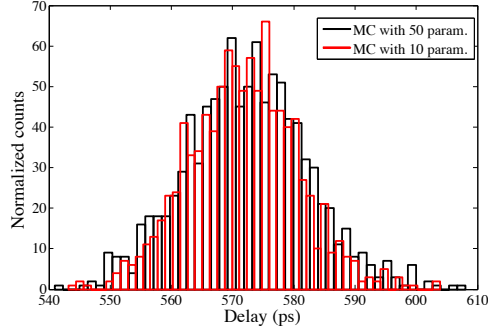

Fig. 6. Delay distribution comparison of the full and the reduced parameter models.

$\sigma = 10.1ps$ leading to an average error of 9.4%). 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 vs. 2000 simulations for training and feature selection).

V. CONCLUSIONS

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 process variations while reducing the computation complexity by utilizing the obtained reduced parameter set. This method is based on Feed Forward Neural Network 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 non-linear dependencies between input parameters and outputs which lead to effective parameter reduction. Applied to ISCAS89-s27 benchmark exploiting DG-SiNWFET technology, experimental results show $2.5\times$ speed up in timing analysis and estimation of the delay distribution with the average error bound of 9.4%.

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] S. Bangsaruntip, *et al.*, "High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling," IEDM Tech. Dig., 2009.

[2] M. De Marchi, *et al.*, "Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire FETs," IEDM Tech. Dig., 2012.

[3] S. Khasanvis, *et al.*, "Hybrid Graphene Nanoribbon-CMOS tunneling volatile memory fabric," NanoArch Tech. Dig., 2011.

[4] Y. Li, *et al.*, "Process-variation- and random-dopants-induced threshold voltage fluctuations in nanoscale planar MOSFET and bulk FinFET devices," Microelectronic Eng., 86(3):277-282, 2009.

[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. ICSCAD, 28(1):141-153, 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, 58(8):2317-2325, 2011.

[7] W. Hong , *et al.*, "A novel dimension-reduction technique for the capacitance extraction of 3-D VLSI interconnects," IEEE Trans. Microwave Theory & Tech., 46(8):1037-1044, 1998.

[8] Y. Zhan, *et al.*, "Correlation-aware statistical timing analysis with non-Gaussian delay distributions," DAC Tech. Dig., 2005.

[9] C. Visweswariah, *et al.*, "First-Order Incremental Block-Based Statistical Timing Analysis," IEEE TCAD, 25(10):2170-2180, 2006.

[10] C. Binjie, *et al.*, "Statistical-Variability Compact-Modeling Strategies for BSIM4 and PSP," IEEE Design & Test of Computers, 27(2):26-35, 2010.

[11] A. Agarwal, *et al.*, "Statistical timing analysis for intra-die process variations with spatial correlations," ICCAD Tech. Dig., 2003.

[12] H. Feng and P. Li, "Performance-Oriented Parameter Dimension Reduction of VLSI Circuits," IEEE TVLSI, 17(1):137-150, 2009.

[13] K. Chopra, C. Zhuo, D. Blaauw, and D. Sylvester, "A statistical approach for full-chip gate-oxide reliability analysis," ICCAD Tech. Dig., 2008.

[14] D.S. Boningy and P.K. Mozumder, "A System for Design of Experiments, Response Surface Modeling, and Optimization using Process and Device Simulation," IEEE Semiconductor Manufacturing, 7(2):233-244, 1993.

[15] D. Blaauw, K. Chopra, A. Srivastava, L. Scheffer, "Statistical Timing Analysis: From Basic Principles to State of the Art," IEEE TCAD, 27(4):589-607, 2008.

[16] Z. Feng and P. Li., "Performance-oriented statistical parameter reduction of parameterized systems via reduced rank regression," ICCAD Tech. Dig., 2006.

[17] R. Huang, *et al.*,"Variability investigation of gate-all-around silicon nanowire transistors from top-down approach," EDSSC Tech. Dig., 2010.

[18] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks, 2(5):359-366, 1989.

[19] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Inc., Secaucus, NJ, 2006.

[20] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," Journal of Computational and Graphical Statistics 22(2): 231-245, 2013.

[21] M.H. Ben Jamaa, K. Mohanram and G. De Micheli, "An Efficient Gate Library for Ambipolar CNTFET Logic," IEEE TCAD, 30(2):242-255, 2011.

[22] L. Cheng, P. Gupta, and L. He, "Accounting for non-linear dependence using function driven component analysis," ASP-DAC Tech. Dig., 2009.