A Highly Accurate Machine Learning Approach to Modelling PVT Variation Aware Leakage Power in FinFET Digital Circuits

by

Zia Abbas, Shirisha Gourishetty, Harshini Mandadapu, Andleeb Zahra

Report No: IIIT/TR/2019/-1

Centre for VLSI and Embeded Systems Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
November 2019

A Highly Accurate Machine Learning Approach to Modelling PVT Variation Aware Leakage Power in FinFET Digital Circuits

Shirisha Gourishetty, Harshini Mandadapu, Andleeb Zahra, Zia Abbas
Center for VLSI and Embedded Systems Technology (CVEST)
International Institute of Information Technology, Hyderabad (IIIT-H) Hyderabad, India - 500032
gourishetty.shirisha@research.iiit.ac.in, h.mandadapu@students.iiit.ac.in, andleebiitd@gmail.com, zia.abbas@iiit.ac.in

***Abstract*—Due to the advent of deep sub-micron technologies, statistical (in addition to temperature and supply voltage) variations aware estimation of leakages power has become prominent. Also, estimation of leakage currents at SPICE level guarantees the most accurate results, however not feasible means in high complexity ICs. This performs adversely for Monte-Carlo iterations for statistical analysis. In this paper we introduced an accurate machine learning technique to model statistical and operating variation aware estimation from Artificial Neural Network and regression based Multivariate Polynomial Regression which exhibits innately faster computation and attained error less than 1% for the targeted 16nm FinFET technology node although model is black box for any technology. The accuracy of the proposed technique has been tested over several basic cells and estimation of the complex circuits have been carried out utilizing the pre-modelled basic cells.**

***Index Terms*—Machine Learning, Neural Networks, Polynomial Regression, Statistical Variations, Leakage Power, FinFET, VLSI**

I. Introduction

Due to relentless scaling of technology nodes beyond 22nm, Complementary Metal Oxide Semiconductor (CMOS) circuits has got disadvantages like high power density and enormous leakage power [1] [2]. Many solutions have been proposed to enhance the performance or to replace CMOS devices. Fin Field Effect Transistor (FinFET) is one such solution [3]. FinFET is multi-gate device i.e. incorporates more than one gate into a single device and thus have better electrostatic control over the channel [4]. Even then small variation during fabrication can largely affect the device performance and ultimately impact not only yield but also performance in circuits. In addition, temperature and supply voltage variations always exist to make device unreliable due to external conditions.

SPICE level Monte Carlo analysis is a prevalent way to analyze the robustness of circuit by randomly generating the parameters according to the distribution model. Although, large number of simulations from Monte Carlo analysis is highly time consuming. SPICE guarantees most accurate results and therefore can be used to compare results obtained from computationally efficient models. However, the generation of reliable models which ensure high accuracy and high computational efficiency is a challenging task.

One technique is to create a model which approximates the circuit behaviour and less time consuming. Machine learning algorithms have proven to be efficient in mocking the behaviour when trained efficiently. Here, we propose two machine learning techniques namely Polynomial Regression and Artificial Neural Network models to appropriate the task of SPICE, but computationally efficient than SPICE. The time complexity reduced by several order compared to SPICE and with a reliable error rate of $\leq 1\%$.

II. Previous Work

Many techniques were proposed to model the leakage power of the circuits. In [5], gate leakage estimation is done using pattern independent probabilistic analysis. In [6], VHDL-based technique to estimate leakage power of a design considering the state-dependency of the leakage power is proposed. In [7], pattern dependent steady-state leakage estimation technique was proposed that accounts for major leakage components, namely, gate, band-to-band-tunneling, and sub-threshold leakage using Newton Raphson Method. In [8], linear models are used for leakage estimation by taking number of gates and inputs as inputs to the model. In [9], a transistor is modelled as the sum of current sources (SCS) using compact current model, which is used to calculate total leakage of logic circuits. In [10], author explained the implementation of HDL model, which can calculate the gate tunneling, sub-threshold and reverse junction leakages separately with good accuracy. Authors in paper [4], designed model based on characterisation of voltages at internal nodes and also characterised the leakage currents for single FET devices, inaddition authors explored the feasibility of statistical leakage estimation under process variations. In [11], author proposed abstract models to predict leakage and delay behaviours in FinFET digital cells, based on response surface methodology. In the paper [12], author has developed a adjusted 2D model for 3D structure of FinFET using gate under lap adjustment methodology to speed up simulation time with compromise in accuracy up to 20%. In [13], surrogate models are generated based on two machine learning models and used to calculate performance of standard cells. In [14], regression models are trained to calculate leakages. In most of the above mentioned works the accuracy is compromised for time complexity or vice-versa. Therefore, we present a methodology which is trouble-free, highly accurate and efficient in computation time.

III. Simulation Setup

In the proposed work, process parameters such as body doping concentration ($nbody$), source/drain doping concentration (nsd), conduction band carrier concentration at 300 K temperature ($nc0subn$), physical oxide thickness ($toxp$), height of fin ($hfin$), intrinsic carrier concentration at 300K temperature ($ni0subn$), equivalent gate dielectric thickness (eot), physical gate length (lg), thickness of fin ($tfin$) have been considered referring to targeted High Performance 16nm FinFET technology with $\pm 10\%$ variations at 3-sigma. In addition, a wide range (-55°C to 125°C) of temperature ($temp$) and $\pm 5\%$ in supply voltage (V_{DD}) variations are also considered, as depicted Table I. All the simulation level characterization have been carried out in HSPICE using Predictive Technology Model-Multi Gate Transistors (PTM-MG) based on BSIM-CMG [15] [16].

A hybrid method of sampling, which includes both Gaussian and Latin Hypercube Sampling (LHS) is used inorder to get better distribution of training data set. Latin Hypercube Sampling technique is another simulation based reliability evaluation method. In this method, the domain of each random variable decomposed into interval of same probability. The number of intervals depends on how many samples would be generated for each variable. One value from each interval is selected at random with respect to the probability density in the interval and to satisfy Latin Hypercube requirements. In this paper, we have confined the results reported to 16nm FinFET cells. However, the model can be extended to further scaled FinFET technology nodes and even to CMOS technology by appropriately training the model.

TABLE I: Considered Process and Operating values (16nm FinFET Technology)

Sr. No.	Process parameter	Device	Lower deviation	Nominal	Higher deviation
1	$nbody$	both nFet and pFet	0.97e+23	1e+23	1.03e+23
2	nsd	both	2.9e+26	3e+26	3.1e+26
3	$nc0sub$	both	2.765e+25	2.86e+25	2.955e+25
4	$toxp$	both	1.305e-09	1.35e-09	1.395e-09
5	$hfin$	both	2.513e-08	2.6e-08	2.687e-08
6	$ni0sub$	both	1.063e+16	1.1e+16	1.137e-16
7	eot	both	7.73e-10	8e-10	8.27e-10
8	lg	both	19.33e-09	20e-09	20.67e-09
9	$tfin$	both	1.16e-08	1.2e-08	1.24e-08
10	V_{dd}	both	0.76V	0.8V	0.84V
11	$temp$	both	-55	25	125

IV. Methodology

Two different kinds of machine learning approaches have been developed and compared. Fig.2 depicts the major steps involved in model development.

A. Data Generation and Pre-processing:

Data sets are generated using SPICE by considering inputs as process parameters, supply voltage (V_{dd}), temperature ($temp$) and input signal of the respective circuit. Generation of process parameters have discussed in Section III. Voltage and temperature are generated randomly between the specified range in Table I. Data is spread over a wide range and hence the outputs also distribute over a wide range. In general learning algorithms perform better by standardisation of the data set. Standardisation can be also explained as mean removal and variance scaling. The standard score for a sample x is:

$$z = (x - u)/s \tag{1}$$

where u is mean and s is standard deviation of the data. z represents standardised data.

B. Data Organisation

Data is divided into training and test data sets. This is done in the ratio of 80:20 out of generated samples. Which implies a part of generated data is used for validating the trained model in terms of accuracy and generalisation factor.

C. Training Algorithms:

1) Polynomial Regression: Polynomial Regression is a form of regression analysis algorithm. Regression analysis is a form of predictive modelling technique which uses the relationship between dependent (Y) and independent variables (X) to find best fit regression equation that can be used for predictions. If the relation is non-linear, we use polynomial regression to model Y as an nth degree polynomial equation in X. The nth order polynomial model can be represented as:

$$Y = a_0 + a_1 X + a_2 X^2 + .. + a_n X^n + \epsilon \tag{2}$$

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_{11} X_1^2 + a_{22} X_2^2 + a_{12} X_1 X_2 + \epsilon \tag{3}$$

is the polynomial model with two independent and one dependent variable with emphasizing correlation of independent variables. Where a_0,a_1,a_{11},..,an are regression coefficients and ϵ represents unobserved random error with mean zero. Polynomial regression can fit wide range of curvature. Degree of polynomial is used according to the complexity of function. Using of more complex function for about linear or less degree will leads to Over-fitting. Over-fitted model works very accurate for training set but worse for samples out of training set.

2) Artificial Neural Networks: Artificial Neural Networks are inspired by the biological structure of neurons and their working. ANNs are useful in diverse applications and efficient in accurately determining outputs even in conditions of large input parameters and high degree of non-linearity. An ANN consists of number of highly interconnected processing elements which process information to state response of external inputs. Multi-Layer Perceptron (MLP) is the widely used ANN structure [17]. MLP belongs to the feed-forward class of Neural Networks. It used supervised learning technique called Back Propagation for training. It has input layer, output layer and hidden layer which basically transforms inputs to outputs, shown in Fig.1. Each neuron consists of activation function. The purpose of activation function is to introduce non-linearity into output of neuron. It generally decides whether to activate

a neuron or not by calculating weighted sum and adding bias at each neuron. Number of nodes in input and output layers are

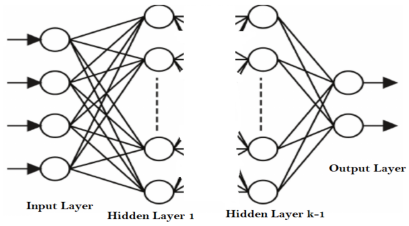

Fig. 1: ANN structure

according to the number of inputs and outputs respectively. Number of hidden layers and neurons in each hidden layer are the parameters to be tuned. Optimised parameters of ANN architecture suitable for training standard cells are shown in Table II.

D. Evaluation:

R^2 score is used to determine how well the model fits the data. It evaluates the scatter of data points around fitted regression function. R^2 score is also called as determination coefficient. In addition, error is also calculated between sample results from SPICE and Model. R^2 score is calculated from python library and error is calculated by Eq.(4):

$$error = \frac{|SPICEValue - ModelValue|}{SPICEvalue} * 100 \quad (4)$$

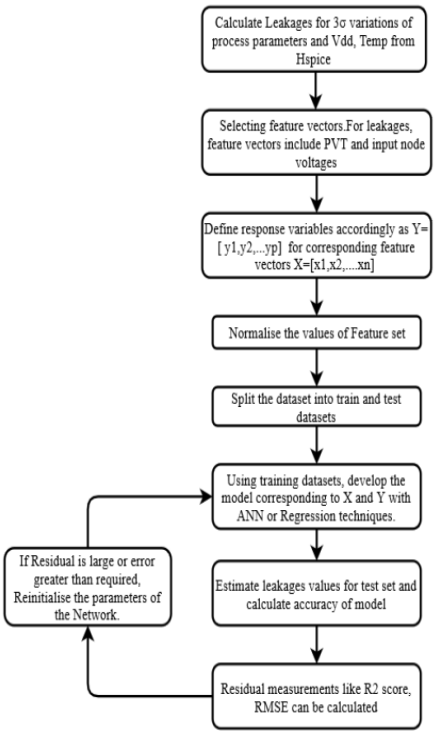

Fig. 2: Flowchart for training Model for a Cell

TABLE II: Architecture of ANN

Parameter	Optimised
Architecture	Feed forward MLP
Training Algorithm	Back propagation
No. of hidden Layers	4
Neurons in each Hidden Layer	15
Activation Function for Hidden Layers	logistic
Maximum Epochs	1000000

V. RESULTS AND DISCUSSION

As mentioned earlier, the data set is split into training and testing sets in the ratio of 80:20 respectively. Test set is used to validate the data. The comparison of results from model and SPICE for AND3 gate are shown in Table III for ANN and polynomial regression. The values reported are for the sample, in which process parameters are at nominal and temperature ,V_{DD} at 80°C, 0.8V respectively.

TABLE III: Comparison of AND3 Leakage Results between SPICE and both Models

Leakages	SPICE	Regression Model		ANN Model	
		Model	Error	Model	Error
Leak_000	$9.034nW$	$8.67nW$	4.0	$9.08nW$	0.51
Leak_001	$9.63nW$	$9.05nW$	6.0	$9.72nW$	3.55
Leak_010	$9.67nW$	$9.07nW$	6.2	$9.72nW$	0.52
Leak_011	$20.65nW$	$21.5nW$	3.9	$20.54nW$	0.53
Leak_100	$9.93nW$	$9.46nW$	4.7	$10.01nW$	0.80
Leak_101	$20.95nW$	$21.79nW$	3.8	$20.93nW$	0.09
Leak_110	$21.25nW$	$22.12nW$	3.9	$21.41nW$	0.93
Leak_111	$5.184nW$	$3.12nW$	3.9	$5.21nW$	0.50

It is observed that artificial neural networks are more suitable for training leakage power models than polynomial regression. We observed that polynomial regression is less suitable for developing leakage model when trained for all input pattern together in the respective circuit. However, the model accuracy is high, if trained with one leakage combination.

The distribution of the data set results from SPICE is compared with the distribution of outputs from ANN Model by mean and standard deviation (Std. Div) values, which depicts the accuracy for Monte Carlo analysis. In Table IV, comparison of statistical aware leakage power are reported for standard cells in Nano-Watts. Due to limited place and to decrease fussiness, results listed only include for maximum leakage. Although results are tested for other combinations for all standard cells. It can be seen from Table IV, that mean and standard distribution from SPICE and ANN Model are almost equal.

R^2 score of each standard Cell is reported for both models in Table V. We can observe that for ANN, R^2 score is between 99 to 100 which specifies that model is best fitted. For polynomial leakages R^2 score is comparatively less which specifies that model is less fitted for those cells.

A. Modelling of Leakages in Complex Cells

The major advantage of this method is that, the results obtained by calculating leakages of complex circuits from pre-characterised and pre-trained basic cell models also has only infinitesimal error.

TABLE IV: Comparison of statistical aware leakage power

Standard cell	Comb.	Mean (μ)		Std. Div. (σ)	
		SPICE	MODEL	SPICE	MODEL
not	leak_max	4.92	4.88	2.85	2.86
AND2	leak_max	19.57	19.49	10.42	10.54
NAND2	leak_max	10.05	10.04	5.67	5.69
NOR2	leak_max	9.84	9.76	5.69	5.55
XOR2	Leak_max	4.42	2.49	4.41	2.48
NAND3	leak_max	15.39	15.36	8.48	8.39
AND3	leak_max	24.9	24.5	12.95	12.86
NOR3	leak_max	14.76	14.81	8.55	8.54
AO12	leak_max	24.5	24.4	13.0	12.8
AO22	leak_max	29.6	29.7	15.63	15.70
AO31	leak_max	29.8	29.8	15.61	15.50
FA	leak_max	68.71	69.12	37.01	36.79

TABLE V: R^2 Score to each standard cell for Polynomial Regression and ANN

Standard cell	Leakages	
	Neural	Poly-reg
NOT	99.929	98.01
AND2	99.789	95.814
NAND2	99.749	93.105
NOR2	99.904	96.080
XOR2	99.909	95.891
NAND3	99.906	94.719
AND3	99.898	90.11
NOR3	99.923	95.468
AO12	99.899	92.73
AO22	99.936	94.98
AO31	99.895	94.304
FA	99.891	91.223

The leakage power of a complex cell can be estimated as sum of all individual gate leakages present in that complex cell by neglecting loading effect.

$$Leakage_{Complex\_cell} = \sum_{i=1}^{N} Leakage_{Basic\_cell_i} \qquad (5)$$

where N is number of basic cells whose leakages should be added. The combination for which leakage is required is given to the primary inputs of complex cell and leakage of each cell is added according to the inputs present at the respective cell to obtain total leakage. The results for some complex cells are shown in Table VI. Due to insufficient space only maximum leakages are reported. Even by neglecting loading effect, results obtained has infinitesimal error.

TABLE VI: Comparison of SPICE and Model Values for Complex Gates

Complex Circuit	SPICE	Model	Error(in %)
C17	$69.99nW$	$69.84nW$	0.21
Multiplier	$1200.30nW$	$1200.19nW$	0.01
Parity Checker	$292.26nW$	$297.95nW$	1.92
Interrupt Controller	$2823.00nW$	$2869.71nW$	1.64
Ripple Carry adder	$304.195nW$	$309.48nW$	1.73

VI. Conclusion and Future Work

In this paper, we proposed two efficient machine learning regression techniques; Polynomial Regression and Artificial Neural Networks for the PVT variation aware estimation/prediction of the leakage power in VLSI digital circuits. The proposed model works as a black box i.e. independent of technology node, computationally very efficient and exhibit very good accuracy with a negligible error ($\leq 1\%$). For future work, we aim to develop modeling techniques incorporating the estimation/prediction of other performance figures (e.g. propagation delays), especially in complex circuits considering all possible factors affecting targeted performances.

References

[1] N. Z. Haron and S. Hamdioui, "Why is CMOS scaling coming to an end?" in *2008 3rd International Design and Test Workshop*, Dec 2008, pp. 98–103.

[2] Z. Abbas and M. Olivieri, "Impact of technology scaling on leakage power in nano-scale bulk CMOS digital standard cells," *Microelectronics Journal*, vol. 45, Jan 2013.

[3] M. Agostinelli, M. Alioto, D. Esseni, and L. Selmi, "Leakage − Delay tradeoff in FinFET logic circuits: A comparative analysis with bulk technology," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 2, pp. 232–245, 2010.

[4] Z. Abbas, A. Mastrandrea, and M. Olivieri, "A Voltage-based leakage current calculation scheme and its application to nanoscale MOSFET and FinFET standard-cell designs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 12, pp. 2549–2560, Dec 2014.

[5] R. M. Rao, J. L. Burns, A. Devgan, and R. B. Brown, "Efficient techniques for gate leakage estimation," in *Proceedings of the International Symposium on Low Power Electronics and Design*, 2003, pp. 100–103.

[6] A. Nourivand, Chunyan Wang, and M. O. Ahmad, "A VHDL-based technique for an accurate estimation of leakage power in digital CMOS circuits," in *The 3rd International IEEE-NEWCAS Conference*, 2005, pp. 47–50.

[7] A. Sanyal, A. Rastogi, W. Chen, and S. Kundu, "An efficient technique for leakage current estimation in nanoscaled CMOS circuits incorporating self-loading effects," *IEEE Transactions on Computers*, vol. 59, no. 7, pp. 922–932, 2010.

[8] J. Derakhshandeh, N. Masoumi, S. Aghnoot, B. Kasiri, Y. Farazmand, and Akbarzadeh, "A precise model for leakage power estimation in VLSI circuits," in *Fifth International Workshop on System-on-Chip for Real-Time Applications (IWSOC'05)*, 2005, pp. 337–340.

[9] A. Agarwal, S. Mukhopadhyay, C. H. Kim, A. Raychowdhury, and K. Roy, "Leakage power analysis and reduction: models, estimation and tools," *IEE Proceedings - Computers and Digital Techniques*, vol. 152, no. 3, pp. 353–368, 2005.

[10] Z. Abbas, V. Genua, and M. Olivieri, "A novel logic level calculation model for leakage currents in digital nano-CMOS circuits," in *7th Conference on Ph.D. Research in Microelectronics and Electronics*, July 2011, pp. 221–224.

[11] S. Chaudhuri, P. Mishra, and N. K. Jha, "Accurate leakage estimation for FinFET standard cells using the response surface methodology," in *25th International Conference on VLSI Design*, 2012, pp. 238–244.

[12] S. M. Chaudhuri and N. Jha, "3D vs. 2D device simulation of FinFET logic gates under PVT variations," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 10, pp. 1–19, April 2014.

[13] C. Vicari, M. Olivieri, Z. Abbas, and M. Khozoei, "Statistical variation aware ANN and SVM model generation for digital standard cells," June, 2016, pp. 419–428.

[14] D. Amuru, A. Zahra, and Z. Abbas, "Statistical Variation Aware Leakage and Total Power Estimation of 16nm VLSI Digital Circuits Based on Regression Models," in *International Symposium on VLSI Design and Test*. Springer, 2019, pp. 565–578.

[15] S. Venugopalan, M. A. Karim, D. Lu, A. Niknejad, and C. Hu, "Compact models for real device effects in FinFETs," in *The International conference on simulation of semiconductor processes and devices (SISPAD), Denver, CO, USA*, 2012.

[16] "Predictive technology model," http://ptm.asu.edu/.

[17] S. S. Haykin, *"Neural Networks and Learning Machines"*, 3rd ed. Pearson Education, 2009.