An Efficient Gradient Boosting Approach for PVTAware Estimation of Leakage Power andPropagation delay in CMOS/FinFET Digital Cells

by

Deepthi Amuru, Salman Ahmed, Zia Abbas

Report No: IIIT/TR/2020/-1

Centre for VLSI and Embeded Systems Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
October 2020

An Efficient Gradient Boosting Approach for PVT Aware Estimation of Leakage Power and Propagation delay in CMOS/FinFET Digital Cells

Deepthi Amuru, Mohammed Salman Ahmed, Zia Abbas
Center for VLSI and Embedded Systems Technology (CVEST)
International Institute of Information Technology, Hyderabad (IIIT-H) Hyderabad, India - 500032
Email: deepthi.amuru@research.iiit.ac.in, salman.ahmed@research.iiit.ac.in, zia.abbas@iiit.ac.in

***Abstract*—In this paper, we propose an accurate and computationally efficient Gradient Boosting approach for the estimation of statistical variations aware leakage power and propagation delay in the CMOS/FinFET standard digital cells. The proposed model estimates the leakage power and propagation delay w.r.t variations in process, temperature ($-55^{\circ}C$ to $125^{\circ}C$) and supply voltage($\pm 10\%$ variations). The distinguishing feature of the proposed approach is its compatibility with both CMOS and FinFET technologies. Moreover, the performance of the proposed model is consistent with various technology nodes. Exhaustive tests report an average error of <1% in 16nm CMOS and FinFET standard digital cells w.r.t analog HSPICE simulations with several orders increase in computational speed. Further, the complex cell estimation can be carried out through pre-characterized standard cells abstaining longer simulations.**

***Index Terms*—Machine learning (ML), Gradient Boosting Model (GBM), Leakage power, Propagation delay, CMOS, FinFET, VLSI**

I. Introduction

The challenges in accurate estimation of power and delay in digital VLSI circuits are increasing in the nanometer regime. The aggressive scaling down of CMOS/MGFETs devices has led to a profound increase in process variability in turn, an enormous increase in leakage power. The existence of process variations in both devices and interconnect networks increases the randomness in the behavior of the digital circuits and affect their performance. This phenomenon is building difficulties in the device characterization of VLSI circuits and in turn, could result in the parametric degradation of manufacturing yield. The process variations together with the environmental variation sources affect the electrical behavior of the circuit leading to circuit deviation and deterioration of the overall performance of the chip [1]. Although, FinFETs are more tolerant to leakage than CMOS devices, the performance degradation due to the process variation is inevitable [2].

Therefore, the accurate and efficient estimation of leakage and delay of CMOS/FinFET devices under the influence of random variations in process, temperature and supply voltage is emphasized in the nanometer regime prior to fabrications of ICs. A common approach for the estimation of leakage and delay is through Monte-Carlo simulations from HSPICE [3]. The HSPICE Monte-Carlo simulations guarantee accurate results at the cost of a huge simulation time [4]. That makes them unfeasible for the estimation of complex devices. One solution for avoiding long simulation runs is to develop surrogate models that approximate the performance of a circuit as a function of statistical parameters. The surrogate models can then replace the simulator in large Monte-Carlo simulations thus, reducing the computational time.

We propose a computationally efficient surrogate model for PVT variations aware leakage and delay estimation in digital CMOS/FinFETs of standard digital logic cells using the Gradient Boosting algorithm. To our knowledge, this is the first paper developing a statistical black-box model based on gradient boosting for the estimation of leakage and delays in CMOS/FinFET circuits. We also presented a comparative analysis of results from other machine learning techniques reported in literature; linear, polynomial regression [5] and support vector machine (SVM) [6].

II. Related works

Several surrogate techniques are proposed in the literature for the estimation of leakage and delay in digital circuits. Kim $et\ al.$ [4] presented a calculation model for estimating propagation delay in digital circuits converting an arbitrary logic gate into an equivalent inverter. The model does not require any pre-simulation step but it shows 5% error w.r.t HSPICE. D.Helms $et\ al.$ [7] introduced RT level leakage macro models for the gate level leakage estimation which reported an error of 2.1% (for 16nm LP) 6.8% (for 65-nm bulk) w.r.t HSPICE simulations. The paper considers 90000 training samples and limited the process variations to 10 due to complexity reasons. Garg $et\ al.$ [6] presented a stack-based macro model for statistical aware gate-level leakage power estimation (CMOS) using support vector machines (SVM) with a 17x improvement in speed. In [8], Xingsheng Wang $et\ al.$ discussed the statistical variability and reliability of FinFETs taking into consideration the impact of different design and process parameters. But the paper does not address the effect of these variations on the leakage and delay.

A more accurate estimation through the proposed surrogate models is crucial for the realistic estimation of the fabrication yield. A model that does estimations on par with HSPICE is entailed. Our paper reports a negligible average error rate of $<1\%$ with 15000 PVT variations. Moreover, the model is an

978-1-7281-3320-1/20/$31.00 ©2020 IEEE

efficient black-box technique on CMOS and FinFET nanoscale digital circuits at different technology nodes.

III. The Proposed Methodology

Process variations can be broadly characterized into global (inter-die) and local (intra-die) variations. Global variations (e.g. t_{ox}) are chip-to-chip or wafer-to-wafer variations and therefore have long-range influences and affect every device of the same type in identical fashion [2]. Local process variations may occur due to Random Dopant Fluctuations (RDF), Line Edge Roughness (LER) and other intrinsic fluctuations affecting every device in a chip individually, i.e. variability between two devices may look identical to each other. The effect of RDF causes random threshold voltage (V_{th}) that lead to exponential changes in device currents especially in sub-threshold [9].

Estimating the effect of process variations through process corners is not effective at present technology nodes. The design needs to be verified over the entire range of variations. The proposed algorithm exploits the strength of statistical approach by intermittently varying its worst-case process variations and it is efficient in this sense when compared to conventional worst-case method [2].

A. SIMULATION SETUP

The training data for the gradient boosting algorithm is generated as a vector of random values from the Gaussian distribution of each process parameter in the targeted technology node considering $\pm 10\%$ variations at 3σ. Along with these statistical distributions, random samples of temperature ranging from $-55°C$ to $125°C$ and supply voltage with a $\pm 10\%$ deviation from the nominal value ($0.8V$) are included. These variations are combined with the standard cell CMOS/FinFET netlists through Monte-Carlo simulations for training samples genearation .

In this work, we consider 14 process parameters (PMOS and NMOS) for CMOS - Channel Length, Transistor Width, Physical and Electrical equivalent of oxide thickness, nominal gate oxide thickness, Source/Drain junction depth, Channel doping concentration [10], [11], [12]. Keeping into account the major performance variations in FinFETs [13], [14], 16 FinFET process variations (PFET and NFET) - Gate length, Fin height, Body thickness, Physical equivalent of oxide thickness, equivalent gate dielectric thickness, Conduction band density of states, Intrinsic carrier concentration of channel, Source/Drain doping concentration are considered [15], [16]. The model is capable of handling more process parameter variations than forth mentioned as its complexity is less affected by the number of process parameters.

We analyzed the impact of number of training samples on the accuracy of the proposed model and reported in Fig.1. The error metrics - Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R2Score) are the statistical measures for the model evaluation [17]. The hyperparameters of gradient boosting algorithm are tuned such that $RMSE$ and MAE are less than the order of E^{-6} with an R2Score > 0.95. The accuracy of the training improves with the number of training samples. The proposed work considers 15000 training samples from HSPICE Monte-Carlo simulations as an optimum solution w.r.t accuracy and training time. The model shows an impressive speed-up over HSPICE as reported in Fig.2. For 15K samples, the proposed model takes an average training time of 40secs and a testing time of < 1 sec. This paper targets 16nm CMOS/FinFET high-performance technology models. Successful estimations performed on different technology nodes - 32nm, 22nm, 16nm CMOS and 16nm, 7nm FinFETs (Fig.3) validate the proposed model to be an efficient black-box technique.

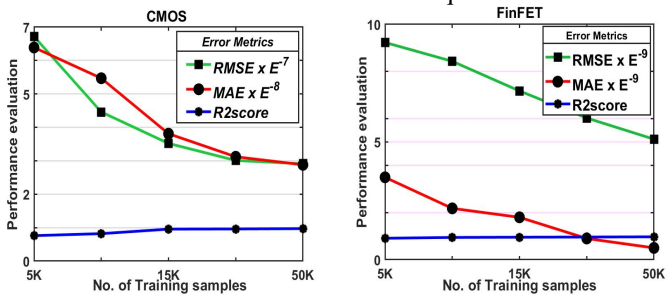

Fig.1. Impact of training samples on MODEL performance. (shown for 16nm Full adder Max. Leakage Power)

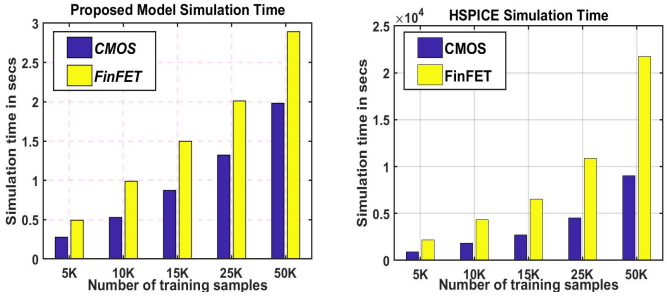

Fig.2. Comparison of Monte-Carlo simulation time & MODEL (reported for 16nm Full Adder Max. Delay) Simulated on 8GB RAM i5 CPU

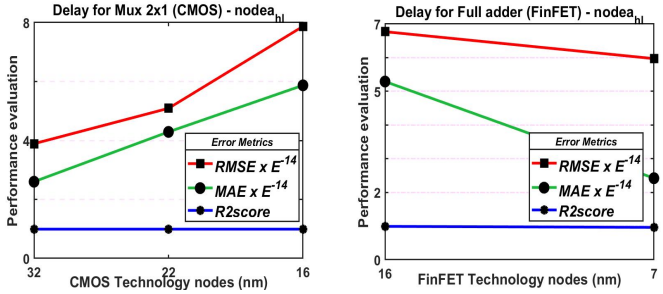

Fig. 3. Performance measures of the proposed model on different technology nodes

B. LEAKAGE and DELAY MODELING

In the proposed model, leakage power and propagation delays are calculated in HSPICE subjecting to statistical variations in process parameters $X_p = [X_{p,1}X_{p,2}X_{p,3}...X_{p,k}]^T \epsilon R^k{}_{xp}$ along with the random variations in operating parameters (such as temperature and supply voltage) $X_r = [X_{r,1}X_{r,2}]^T \epsilon R_{xr}$.

Besides, propagation delay varies with the load capacitance and slew time.

$$t_{pd} = f(C_L, V_{DD}, W, t_s) \quad (1)$$

The delays are also affected by the associated parasitic capacitance of a digital circuit whose voltage switches with the input transitions.

$$C_{Total} = C_{LOAD} + C_{parasitic} \quad (2)$$

The parasitic capacitances contributing to the total input capacitance, C_{Total} of a cell vary with input voltage transitions [18]. As the proposed approach estimates the delay of complex cells through pre-characterized cells, an accurate estimation of the total capacitance is necessary with the coupling of the driver and driven cells. Hence, the proposed approach computes equivalent gate capacitance values (equation 2) for all possible combinations of the standard logic cells (table II) with all possible input transitions and static input combinations (1/0) on other inputs of a driven cell as shown in Fig 4. Along with process variations, range of equivalent C_{Load} values between smallest and highest possible load combinations are randomly applied as inputs to HSPICE while computing training samples for the accurate estimation of delays. We considered slew time (t_s) to be is a constant at 10 pico-sec in this work. In future, slew time can be incorporated in our model as a linear interpolation model [3].

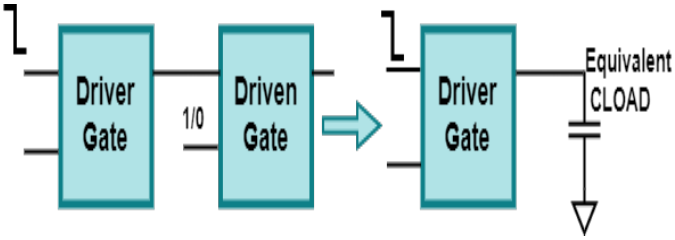

Fig. 4. Computation of equivalent Load Capacitance

The model calls the pre-computed standard cell leakages/delays for appropriate input combinations and capacitance values for each driver and driven cell combination in a complex cell for total leakage and delay estimations respectively [19].

This paper considers 12 digital cells for standard library characterization. Nevertheless, the proposed approach can be extended to more number of digital cells. We have considered a single transition on input nodes in this work wherein other transitions can be accommodated in the proposed approach.

C. GRADIENT BOOSTING ALGORITHM

The proposed Gradient Boosting Algorithm (GBM) is a learning-based statistical estimation approach using ensemble learning. Gradient boosting adaptively combines a large number of relatively simple tree models to produce an accurate predictive rule. The design of each decision tree is simple and speeds up the process of training. Boosting improves the accuracy of the estimation in each iteration by reducing the residue of the previous training model (decision tree) [20]. In each iteration, extra weights are given to the incorrectly predicted points by earlier predictors. Each iteration takes closer to the global optimization value using gradient descent [21].

Algorithm: Statistical Leakage/Delay Estimation using Gradient Boosting

1: **while** $i < N$
2: **while** $j < k$
3: $X_p \leftarrow random(\pm 10\%$ variations at 3σ of j_{th} process)
4: **endwhile**
5: $X_r \leftarrow random(temperature, \pm 10\%$ variations of $V_{sup})$
6: $X_l \leftarrow random(\mathrm{C}_{Load}, t_s)$ for delay estimation
7: **endwhile**
8: $CallHSPICE()$
9: **while** $i < N$
10: $(y\{target\}, \mathrm{X}\{\mathrm{X}_p, X_r\}) \leftarrow$target$({X_p}^k, {X_r}^k)$
11: **endwhile**
12: Extract feature set $X\{X_p, X_r\}$
13: Define target variable $y\{target\}$
14: **while** $j < k$
15: Calculate Pearson coefficient, r for all process
16: Sensitivity analysis of leakage to variations
17: **endwhile**
18: Determine effective feature set $X_e\{X_{se}, X_{pe}\}$
19: Call **GradientBoosting**$(y\{target\}, X_e, hyperparameters)$
20: $F_0(x) = \binom{argmin}{\gamma} \sum_{i=1}^{N} L(y_i, \gamma)$
21: **for m=1 to M do:**
22: $w_{im} = -[\frac{\partial L(y_i F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}; for\ i = 1, ..., N$
23: $R_{jm} = \binom{argmin}{\gamma} \mathrm{R}_j, \eta \sum_{i=1}^{N} [w_{im} - \eta I(x_i, R_j)]^2$
24: $\gamma_{jm} = \binom{argmin}{\gamma} \sum_{x_i \epsilon R_{jm}}^{N} L(y_i, F_{m-1}(x) + \gamma); i = 1, .N$
25: $F_m x = F_{m-1}(x) + \eta \sum_{j=1}^{jm} \gamma_{jm} I(x \epsilon R_{jm})$
26: **end for**
27: **end** GradientBoostingAlgorithm
28: Calculate *RMSE, MAE, R2Score*
29: **If** *R2Score* $<$ 0.95 & *RMSE* $>$ the order of E^{-6}
30: Hyper-parameter fine tuning
31: *Repeat* steps 14 to 26 by tuning hyperparameters
32: **until** *R2Score* $>$ 0.95
33: **end if**
34: *return* trained *GradientBoostingModel*
35: Call **GradientBoostingModel**$(new\ test\ \mathrm{X}\{X_s, X_p\})$
36: return *estimated target (leakage/delay)*
37: CallHSPICE$(new\ \mathrm{X}\{X_s, X_p\})$)
38: return *actual leakage*
39: Determine $\%error$ {actual,estimated}leakages/delays
40: if $\%error > 1\%$
41: increase N *by* 5%
42: *Repeat step* 1 *to* 37
43: **until** $\%error < 1$

In the proposed approach, N represents number of training samples and k is the number of process parameters. The Gradient Boosting Algorithm is described between lines 20

to 26. The parameters of the algorithm are minimum value of the loss function, γ, residue from the previous tree,w_{im} for i_{th} sample of m_{th} tree among M trees (M is chosen to be 100 in this work), R_{jm} is the terminal region for j_{th} leaf node of m_{th} tree and learning rate,η. In each iteration from m=1 to M=100, the loss function is minimized and getting closer to the optimum solution. A learning rate of 0.1 is chosen as an optimum value in the proposed model with 100 leaf nodes.

IV. Model Accuracy Verification

A two-fold analysis is performed to test the efficiency of the proposed model. Each dataset is split into training and testing datasets with N samples and M=n-N samples respectively where n represents the total number of samples in the dataset. M is chosen to be 10% of the total number of samples. $RMSE$ and MAE values reported quite a minimum for all the standard cells which demonstrates the efficiency of the proposed estimation model.

Table I Test cases at nominal values of process

Temp($^{\circ}C$)	Supply(V)	$C_{total}(fF)$
25	0.8	0.2
40	0.8	0.33
-25	0.85	1.2
120	0.76	0.2

Table II $\%Error$ w.r.t HSPICE for test cases(table I) (Reported for max combinations)

LOGIC CELL	TEST	COMBI.	MODEL CMOS	MODEL FinFET
NOT	3	$leak_0$	0.912	0.764
		tp_{hl}	0.583	-1.55
2-input NAND	1	$leak_{10}$	-0.864	0.922
		$nodea$ tp_{hl}	1.13	-1.06
2-input NOR	2	$leak_{00}$	0.859	0.949
		$nodea$ tp_{hl}	0.866	-1.024
2-input AND	4	$leak_{11}$	-0.641	-0.542
		$nodeb$ tp_{hl}	1.34	0.283
2-input XOR	3	$leak_{00}$	-0.972	-0.766
		$nodea$ tp_{hl}	-0.871	-1.587
3-input NAND	1	$leak_{110}$	1.105	1.212
		$nodeb$ tp_{lh}	0.658	0.564
3-input NOR	2	$leak_{001}$	-0.957	1.01
		$nodea$ tp_{hl}	1.15	0.491
3-input AND	4	$leak_{111}$	1.006	1.058
		$nodea$ tp_{hl}	0.906	-0.883
MUX 2x1	3	$leak_{101}$	-0.500	-0.029
		$nodea$ tp_{hl}	-0.72	0.606
FULLADDER	1	$leak_{001}$	0.997	-0.836
		$nodea$ tp_{hl}	-0.247	1.106
AOI12	2	$leak_{001}$	-0.641	-0.166
		$nodea$ tp_{hl}	0.578	0.160
AOI31	1	$leak_{1010}$	-1.063	-0.968
		$nodeb$ tp_{lh}	-0.595	0.035

As an example, we have also examined the estimation accuracy of the proposed model for new arbitrary chosen test data listed in Table I and the estimation results are reported in Table II. The test data is chosen such that it spans across the limits of temperature and voltage variations for the nominal process. And different values of equivalent load are considered for the delay estimations.

A comparative analysis of the proposed model is made with polynomial regression [5] and support vector regression [6] in Fig. 5 (demonstrated for FinFET AOI12 cell; Delay at nominal process, voltage, and temperature). The proposed model depict higher accuracy compared to other reported models. The same is again validated in the Fig.6, illustrating the mean and standard deviation distribution of different ML models with reference HPSICE distribution. It is clearly visible that GBM distribution completely overlaps the HSPICE estimations for both the leakage and delays estimations. Table III and IV show the leakage power and delays estimation in complex circuits using pre-modelled basic cells.

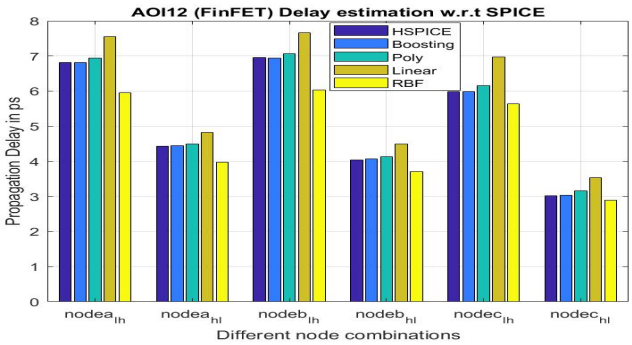

Fig. 5 Comparison of proposed model accuracy w.r.t HSPICE and models [5], [6]

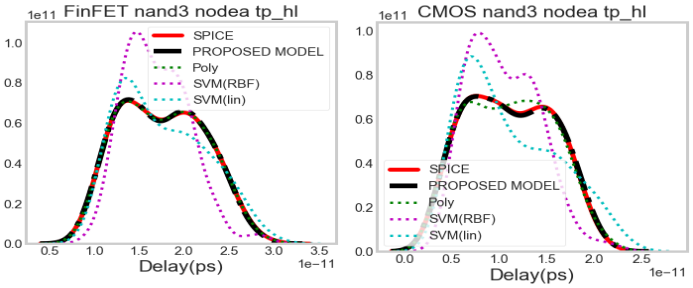

Fig.6. Statistical analysis of proposed GBM Model Vs. HSPICE and other Models [5], [6]

Table III Complex cell Leakage estimations (Reported for max. leakage x $E^{-7}watts$)

Complex cell	CMOS		FinFET	
	SPICE	MODEL	SPICE	MODEL
C17	1.904	1.908	0.585	0.595
4bit adder	8.475	8.461	3.04	3.1
4x4 multiplier	27.93	27.81	11.87	11.81
Parity checker	8.789	8.873	2.926	2.952
Interrupt Controller	9.158	9.20	3.163	3.198

Table IV Complex cell Delay estimations (Reported for max. delay X $E^{-11}secs$)

Complex cell	CMOS		FinFET	
	SPICE	MODEL	SPICE	MODEL
C17	1.357	1.342	1.20	1.189
4bit adder	6.96	6.86	5.19	5.178
4x4 multiplier	9.162	9.06	8.343	8.31
Parity checker	7.788	7.902	7.52	7.33
Interrupt Controller	5.973	6.04	2.42	2.35

V. Conclusion

We propose a computationally efficient machine learning based estimation model using Gradient Boosting Algorithm for an accurate estimation of PVT variations aware leakage power and propagation delays in CMOS/FinFET digital VLSI circuits. The obtained results show negligible average error $< 1\%$ and $< 3\%$ for basic standard cells and complex circuits respectively w.r.t SPICE simulation value with several order of faster estimations, even better for complex cells due to the utilization pre-estimated basic cell characterization. For future work, we aim to develop a robust estimation tool incorporating input and output slew times in delays estimations.

REFERENCES

[1] A. Zjajo, *Stochastic Process Variation in Deep-Submicron CMOS*. Springer, 2016.

[2] Z. Abbas and M. Olivieri, "Optimal transistor sizing for maximum yield in variation-aware standard cell design," *International Journal of Circuit Theory and Applications*, vol. 44, no. 7, pp. 1400–1424, 2016.

[3] M. Olivieri and A. Mastrandrea, "Logic drivers: A propagation delay modeling paradigm for statistical simulation of standard cell designs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 6, pp. 1429–1440, 2013.

[4] D.-W. Kim and T.-Y. Choi, "Delay time estimation model for large digital cmos circuits," *VLSI Design*, vol. 11, no. 2, pp. 161–173, 2000.

[5] A. Stillmaker and B. Baas, "Scaling equations for the accurate prediction of cmos device performance from 180 nm to 7 nm," *Integration*, vol. 58, pp. 74–81, 2017.

[6] L. Garg and V. Sahula, "Variability aware support vector machine based macromodels for statistical estimation of subthreshold leakage power," in *2012 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, pp. 253–256, IEEE, 2012.

[7] D. Helms, R. Eilers, M. Metzdorf, and W. Nebel, "Leakage models for high-level power estimation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 8, pp. 1627–1639, 2017.

[8] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale finfets," in *2011 International Electron Devices Meeting*, pp. 5–4, IEEE, 2011.

[9] S. Shukla, S. S. Gill, N. Kaur, H. Jatana, and V. Nehru, "Comparative simulation analysis of process parameter variations in 20 nm triangular finfet," *Active and Passive Electronic Components*, vol. 2017, 2017.

[10] B. Model, "http://www-device.eecs.berkeley.edu/bsim/."

[11] P. T. Model, "http://ptm.asu.edu/."

[12] Z. Abbas and M. Olivieri, "Impact of technology scaling on leakage power in nano-scale bulk cmos digital standard cells," *Microelectronics Journal*, vol. 45, no. 2, pp. 179–195, 2014.

[13] H. Mahmoodi, S. Mukhopadhyay, and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nanoscale cmos circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1787–1796, 2005.

[14] D. Bhattacharya and N. K. Jha, "Finfets: From devices to architectures," *Advances in Electronics*, vol. 2014, 2014.

[15] B.-C. Model, "http://bsim.berkeley.edu/models/bsimcmg/."

[16] 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, 2014.

[17] D. Amuru, A. Zahra, and Z. Abbas, "Statistical variation aware leakage and total power estimation of 16 nm vlsi digital circuits based on regression models," in *International Symposium on VLSI Design and Test*, pp. 565–578, Springer, 2019.

[18] S.-M. Kang and Y. Leblebici, *CMOS digital integrated circuits*. Tata McGraw-Hill Education, 2003.

[19] V. Champac and J. G. Gervacio, *Timing Performance of Nanometer Digital Circuits Under Process Variations*. Springer, 2018.

[20] A. Keprate and R. C. Ratnayake, "Using gradient boosting regressor to predict stress intensity factor of a crack propagating in small bore piping," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1331–1336, IEEE, 2017.

[21] A. B. Nassif, "Short term power demand prediction using stochastic gradient boosting," in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, pp. 1–4, IEEE, 2016.