On Extracting Reliability Information from Speed Binning

Najafi-Haghi, Zahra Paria; Klemme, Florian; Amrouch, Hussam; Wunderlich, Hans-Joachim

Proceedings of the 27th IEEE European Test Symposium (ETS'22); Barcelona, Spain; 23-27 May 2022

doi: https://doi.org/10.1109/ETS54262.2022.9810443

Abstract: Adaptive Voltage Frequency Scaling (AVFS) is an important means to overcome process-induced variability challenges for advanced high-performance circuits. AVFS requires and allows determining the maximum speed F max (V dd) reachable under a set of certain operation voltages V dd . In this paper, it is shown that the F max (V dd) measurements contain relevant data to identify some hidden defects in a chip which are reliability threats and can cause device failures, but pass the speed binning procedure within the given specifications.Static Timing Analysis (STA) is applied to a circuit designed by using standard cell libraries in which the underlying transistors along with process variations have been carefully calibrated against industrial 14nm FinFET measurement data, and in-stances with and without injected small resistive open defects are generated. From the slope of the function F max (V dd), a machine learning procedure can identify some defects with high precision and few false positives. These chips can be then discarded without any further need and cost for testing. It has to be noted that this reliability information comes for free from the data which is already generated, and does not need any additional measurements.

Preprint

On Extracting Reliability Information from Speed Binning

Zahra Paria Najafi-Haghi, Florian Klemme, Hussam Amrouch, Hans-Joachim Wunderlich
Institute of Computer Architecture and Computer Engineering, University of Stuttgart, Germany
Email: {najafi-haghi, klemme, amrouch, wu}@informatik.uni-stuttgart.de

***Abstract*—Adaptive Voltage Frequency Scaling (AVFS) is an important means to overcome process-induced variability challenges for advanced high-performance circuits. AVFS requires and allows determining the maximum speed $F_{max}(V_{dd})$ reachable under a set of certain operation voltages V_{dd}. In this paper, it is shown that the $F_{max}(V_{dd})$ measurements contain relevant data to identify some hidden defects in a chip which are reliability threats and can cause device failures, but pass the speed binning procedure within the given specifications.**

Static Timing Analysis (STA) is applied to a circuit designed by using standard cell libraries in which the underlying transistors along with process variations have been carefully calibrated against industrial 14nm FinFET measurement data, and instances with and without injected small resistive open defects are generated. From the slope of the function $F_{max}(V_{dd})$, a machine learning procedure can identify some defects with high precision and few false positives. These chips can be then discarded without any further need and cost for testing. It has to be noted that this reliability information comes for free from the data which is already generated, and does not need any additional measurements.

***Index Terms*—Resistive open defects, small delay faults, reliability, speed binning, static timing analysis, machine learning.**

I. Introduction

Small Delay Fault (SDF)s are considered as a reliability threat even if the affected circuit is still operating within the specified margins [1] [2] [3]. They tend to increase during the circuit lifetime significantly faster than the aging mechanisms usually expected. Well-known examples are gate-oxide defects where the holes may increase or resistive opens, which lead to an increased current density and therefore faster electromigration [4]. Delay faults are especially hard to detect since the test generation algorithms usually provide rather large test pattern sets with relatively low fault coverage, which also leads to rather expensive test applications [5].

In this paper, we show that resistive open defects on the critical path of a circuit can be already detected during the device characterization and before an expensive complete manufacturing test, even if the behavior of the defective device is still within the specified margins. The main challenge of this task is to distinguish timing effects of maleficent defects from effects due to benign process-induced variability. While the appearance of the FinFET technology reduced the process-induced variability to some extent, it is still a significant concern, and the hard to control pure random variability may become pre-dominant. Root causes of process-induced variability are metal-gate-granularity (MGG), line-edge-roughness (LER) which consists of fin-edge-roughness (FER) and gate-edge-roughness (GER) [6] [7], less impact have random-dopant-fluctuations (RDF) and oxide-thickness-variation (OTV). MGG has impact on the metal work function (MWF), and hence on both the sub-threshold swing as well as the threshold voltage, while LER mainly causes variations of the threshold voltage. Both lead to changes in the timing behavior, but are usually not considered as reliability threats in contrast to resistive defects, even if both introduce a similar additional delay.

Assume D is the set of possible instances of a circuit with different Process-induced variability parameters. Variations makes it mandatory to characterize each device $d \in D$ and determine the maximum frequency F_{max} that device d can operate properly with. Usually, this process of speed-binning is performed after some initial parameter measurements and before an expensive and comprehensive manufacturing test, by executing a brief functional program with varying frequencies [8]. These programs are usually designed to sensitize a large number of critical paths in the combinational circuit [9]. As a result, the devices $d \in D$ are classified into fast, slow, and perhaps other categories.

The binning process becomes even more complicated with the advent of Adaptive Voltage Frequency Scaling (AVFS). In leading-edge systems, AVFS is applied for power and energy optimization, aging reduction, and lifetime increase, as well as for allowing robust operation under variations. AVFS designs require to execute the binning process at additional operating points. Let V_{op} be a set of supply voltages the circuit is intended to operate with. For each voltage $V_{dd} \in V_{op}$ and each device $d \in D$ we have to measure the maximum frequency $F^d_{max}(V_{dd})$ the circuit is operating correctly. The set of measurements

$$M_d = \{F^d_{max}(V_{dd}) \| V_{dd} \in V_{op}\} \tag{1}$$

can be the subject of further investigations. Hence, the goal of the paper at hand is a data analysis procedure to find anomalies in M_d, which indicate the existence of a resistive open on a critical path with high accuracy. Such devices need not to be passed to the expensive manufacturing test and can be discarded before.

The analysis is performed by applying Static Timing Analysis (STA) techniques on circuits [10] [11], which are synthesized using standard cell libraries. The underlying transistor models, as well as the process-induced variability, have been carefully calibrated against industrial 14nm FinFET measurements obtained from Intel [6] [12]. Instances following the specified variability distributions are generated, and in half of them, resistive open defects are injected. The defect size is selected such that the device is still within the specifications and speed-binning would not report a fail. The STA results of all the instances can be used to form a training set

$\{M_d \| d \in D\}$ for a machine learning classifier, here a Random Forest [13]. The classifier is validated as usual by additional instances and shows high precision. The overall flow can be seen in Fig. 1, which contains STA to generate $F^d_{max}(V_{dd})$ and machine learning, to identify defective devices based on F_{max} behavior.

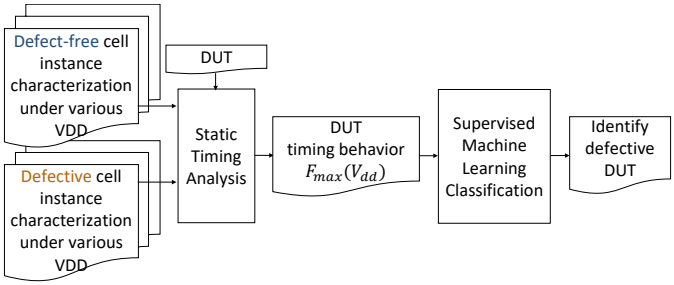

Fig. 1. Overall flow of defect identification based on STA

It should be noted that the classifier trained by STA data is not meant to be directly applied in practice. Instead, the training data can be directly taken from speed-binning , resulting in an even higher accuracy.

While there are publications on the use of STA for yield estimation [14], and critical-reliability paths [3], to the best knowledge of the authors, this is the first paper in literature which applies data analysis techniques on speed binning results to identify reliability threats by resistive open defects.

II. Background on Resistive Open Defects

In this section, we analyze the behavior of a cell with and without a resistive open under variations. We show that the differences are large enough for an accurate classification based on simulation data. These findings are based on the material published in [15] [16], and they are included here to keep the paper self-contained.

Let us illustrate the problem to be solved with the help of a NAND-gate obtained from FinFET Free PDK 15 [17] as shown in Fig. 2 which contains a resistive open, and all transistors are subject to process-induced variability.

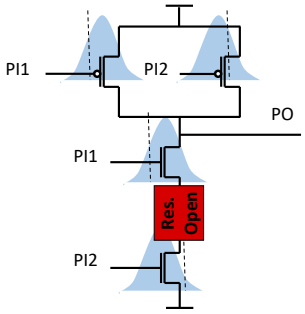

Fig. 2. NAND cell with a resistive open injected

In Fig. 3 we select two specific corner cases and a defective cell from the results of a Monte-Carlo HSPICE simulation. The lower black curve shows the delays of the NAND gate under varying supply voltages V_{dd}, the delays decrease with increasing V_{dd}. The applied random parameters are within the specifications and can lead to a fast instance of this gate. A slow NAND gate behavior is reflected in the dashed green upper curve. If we insert the open defect of Fig. 2 with an appropriate resistance into the fast NAND gate, the result is the dotted red curve. It can be observed that for each operation voltage V_{dd}, this curve is within the specified performance range of the slow and fast NAND instances, and testing at a single voltage value cannot detect this defect. However, the shape of the red defect curve differs from the two defect-free cases sufficiently to allow a machine learning-based classification between defective and defect-free cells.

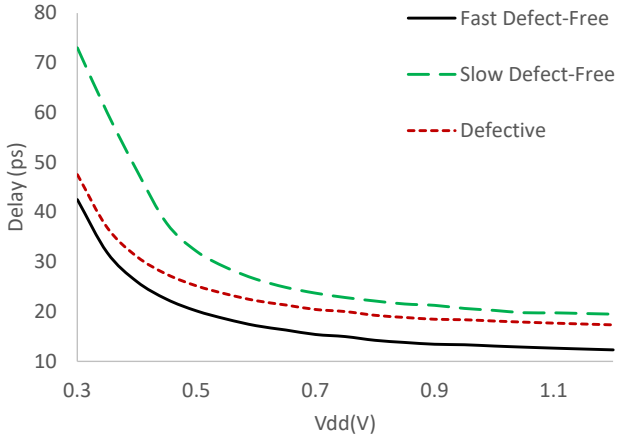

Fig. 3. Timing behavior of two corner instances of a NAND gate vs. a defective cell

This effect can be analyzed further just with textbook information. The delay τ depending on the supply voltage V_{dd} of a CMOS transistor in the absence of transistor short-channel effect, can roughly be presented by

$$\tau(V_{dd}) = Const \frac{V_{dd}}{(V_{dd} - V_t)^\alpha} \tag{2}$$

Here, V_t is the threshold voltage, and the details for the constant $Const$ can be found in textbooks, e.g., [18]. Since circuit timing is mainly affected by the threshold variations, the effects can be investigated by assuming $V_t' = V_t + \delta$. In [16] it is shown that the new delay will be

$$\tau'(V_{dd}) = \tau(V_{dd} - \delta) + Const \frac{\delta}{(V_{dd} - \delta - V_t)^\alpha} \tag{3}$$

The term $\tau(V_{dd} - \delta)$ corresponds to a shift of the black curve to the right in Fig. 3, while adding $Const \frac{\delta}{(V_{dd}-\delta-V_t)^\alpha}$ shifts the curve upwards in a scaled way and finally results in the dashed green curve.

A resistive open, however, will mainly affect the effective resistance R_{eff} in the usual RC-model, and for $R'_{eff} = R_{eff} + \delta$ we get the new delay $\tau' = C_L * (R_{eff} + \delta)$. Hence, a resistive defect of size δ will shift the black curve of Fig. 3 upwards by $C_L * \delta$ resulting in the dotted red curve. This simulation example shows that it is not possible to detect a small resistive open just by a single measurement at only one voltage value, yet measurements at multiple voltages may allow to reconstruct the curve $\tau(V_{dd})$ and to classify a cell as defective or defect-free.

If such a cell is deeply embedded in a combinational circuit, its exact timing behavior may be masked or at least blurred during its propagation along a sensitized path. While the exact timing measurements of an arbitrary path may be difficult during delay testing, the delay of the longest sensitizable, i.e., critical path corresponds to F_{max} determined during the characterization. The next section describes, how to estimate

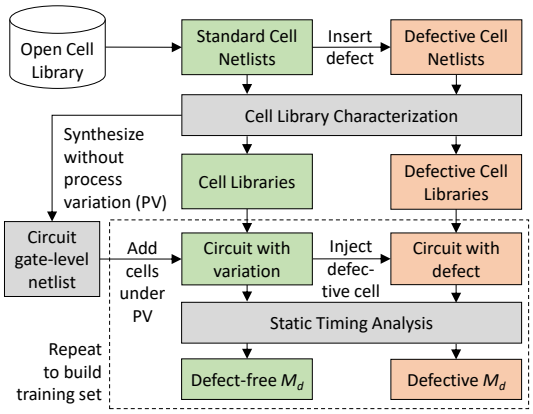

Fig. 4. Overview of our training data generation flow using standard cell characterization and STA.

$F^d_{max}(V_{dd})$ by using STA for instances without defect and instances where defects are injected such that they will be part of the critical path.

III. Training Data Generation using Cell Characterization and Static Timing Analysis

To model the impact of small delay faults in a chip under process variation, we employ a framework of SPICE simulations and STA. An overview of the workflow is depicted in Fig. 4. All cells of the library are characterized considering their parameter distributions for variance, and instances of circuit netlists for each $d \in D$ are generated and evaluated by using STA, which delivers the tuples M_d containing the $F_{max}(V_{dd})$ for multiple $V_{dd} \in V_{op}$. Finally, in some of the netlists, proper defects are injected and STA is repeated to also provide the M_d for defect devices.

To model standard cells under the impact of process variation, Monte-Carlo SPICE simulations are performed while the corresponding model parameters are varied at the same time. For our investigations, we require standard cells with individual instances of process variation applied, so that the same variation can be repeatedly evaluated for different voltages later in STA. Therefore, we build large standard cell libraries containing hundreds of cells of the same type, only with different process variations applied. Each cell version is characterized for a range of $V_{dd} \in V_{op}$ and stored into the corresponding cell library for that V_{dd}.

To build a database of defect-free $F_{max}(V_{dd})$ curves, we annotate the gate-level circuit netlist to employ a different cell version with regard to process variation for each instance of a cell. Then, STA is performed once for each $V_{dd} \in V_{op}$ while linking the corresponding cell library and the annotated circuit. The resulting $\{F^d_{max}(V_{dd}) \| V_{dd} \in V_{op}\}$ models one set of measurements M_d as it would be observed in the speed binning process. The above process is repeated multiple times with new, random annotations in the circuit netlist to reflect multiple devices $d \in D$.

For the database of defective $F_{max}(V_{dd})$ curves, defective cells are first prepared by injecting a single resistive open defect into the defect-free netlist. Its location is defined with regard to the inductive fault analysis (IFA) [19] of the corresponding cell on the most vulnerable nets at the transistor level. After building all defective cell models, standard cell characterization is repeated in the same way as for the defect-free cells. For each $F_{max}(V_{dd})$ curve including a defect, one cell instance in the circuit netlist is replaced with a defective version of that cell. The resulting timing reports reflect the impact of the injected defect in the reported critical path delay.

IV. Machine Learning to Identify Defects

A. Statistical Learning Scheme

To identify defects by using $F_{max}(V_{dd})$, a precise supervised classification scheme is needed. Tree-based schemes are one of the most commonly used supervised learning methods. They can empower classification models with high accuracy, stability, and ease of interpretation [20].

Random Forest scheme builds up several tree-based models on sub-samples of the dataset and report the final results based on the majority voting. It is implemented as follows:

1) Create multiple datasets from original training data.
2) Build multiple tree-based classifiers.
3) Combine classifiers and report the majority voting.

Random Forest is a strong tree-based supervised classification scheme, which is selected as the machine learning classification technique of the work at hand.

B. Evaluation

The classification quality for the reports from Random Forest (RF) is evaluated with respect to a standard metric used in statistical learning [21]. The **Precision**

$$\frac{|TP|}{|TP| + |FP|} \tag{4}$$

denotes the ratio of the correct prediction of a defect (True Positive) versus all defect predictions (True Positive+False Positive). Respectively, *1-Precision* is the ratio of false alarms.

V. Experimental Results

A. Circuits under Investigation

Three multiplier circuits for 8, 16, and 32 bit operands are synthesized with a commercial synthesis tool and investigated before place and route. Hence, the reported area and capacitances are only estimations but sufficient for the "proof of concept" discussion below. Table I describes the most relevant parameters of these circuits. They use seven standard cells from the Nangate $15\,\mathrm{nm}$ Open Cell Library [22] where the transistor models and variability parameters are carefully calibrated against industrial $14\,\mathrm{nm}$ FinFET measurements obtained from Intel [6] [12].

TABLE I
Circuit parameters

		Mul.8	Mul.16	Mul.32
Cell instances		440	1987	7444
Cell area		648	2888.5	10808
2-input gate equivalence		853	3814	14500.5
Longest path	Cell Instances	14	21	30
	Cell area	5.6033	9.4863	11.2067
	2-input gE	42.5	69.5	79

Table I shows the area information for both the entire circuits and the area or length of the respective longest paths.

Circuit Mul.32 with a path length of 30 and 79 in 2-input gate equivalence(gE) could be implemented more efficiently (e.g., using Wallace trees), but it is utilized here to show if the approach is robust against extremely long paths.

B. Data Generation

For each cell, 2000 Monte-Carlo samples of process variation are generated. We assumed, speed-binning was performed for at most 13 voltages in a range of 0.4V to 1.0V in steps of 50mV. With seven cells, 13 supply voltages and the 2000 instances, a total of 182 000 cells had to be characterized. For the defective libraries, we reduced the number of Monte-Carlo samples for process variation to 200, resulting in a total of 36 400 characterized cells. With these libraries, for each of three circuits, 10 000 STA runs for defect-free samples d generated the correct sets M_d, and similarly for the defective samples. Finally, for each circuit a training set of 20 000 samples was generated.

C. Circuit Classification

The tuples $M_d = \{F^d_{max}(V_{dd}) \| V_{dd} \in V_{op}\}$ for defective circuits contain at least one voltage V_{dd} where the defect changes $F^d_{max}(V_{dd})$. Table II presents the precision (Pre) results and the path coverage (PC). Path coverage is the percentage of the circuit covered in total by gates of the critical paths which may change from voltage to voltage.

TABLE II
CLASSIFICATION RESULTS

	PC(%)	Prec_{13}	Prec_{7}	Prec_{4}	$\text{Prec}_{\text{range}}$
Mul.8	10.0	0.79	0.82	1.0	0.80
Mul.16	3.0	0.74	0.75	0.75	0.75
Mul.32	1.0	0.72	0.72	0.72	0.71

We investigated the precision obtained by measurements at 13, 7, and 4 voltages, and by restricting the range to 7 measurements in the voltage range $[0.55, 0.85]$.

The results are amazingly robust against the reduction of features. Although, the reduction of the voltage range is more challenging, it gives still acceptable results.

D. Application

The classification results presented above show roughly a ratio $TP : FP = 0.8 : 0.2$, which means if we discard 5 chips due to this analysis, one correct chip is discarded as well. This pays off, if testing the five chips is more expensive than manufacturing one die. The benefit will be even higher, if additional integration steps are required or when test escapes of the hard to detect delay faults are considered.

In production, the STA-based data and model generation will give an initial estimate of the benefits of the presented approach. However, after some time in production, real measured data from binning and diagnosis will improve the models.

VI. CONCLUSION AND FURTHER WORK

Information about $F^d_{max}(V_{dd})$ for different voltages can be used for identifying resistive open defects in a circuit. This data will come from speed binning without any further measurement.

Exact library cell models and static timing analysis can be used for initial data generation and train a classifier based on Random Forest. The obtainable precision of around 0.8 allows substantial savings by discarding weak chips before testing and further processing.

ACKNOWLEDGEMENTS

This work is part of the project grant WU 245/19-1 (FAST) funded by the German Research Foundation (DFG), and is partly supported by Advantest as part of the Graduate School "Intelligent Methods for Test and Reliability" (GS-IMTR) at the University of Stuttgart.

REFERENCES

[1] S. Hellebrand, T. Indlekofer, M. Kampmann, M. Kochte, C. Liu, and H. Wunderlich, "FAST-BIST: Faster-than-at-Speed BIST Targeting Hidden Delay Defects," in *Proc. IEEE Int'l Test Conf. (ITC)*, Oct. 2014, pp. 1–8.

[2] C. Liu, E. Schneider, M. Kampmann, S. Hellebrand, and H. Wunderlich, "Extending Aging Monitors for Early Life and Wear-Out Failure Prevention," in *Proc. IEEE Asian Test Symp. (ATS)*, Oct. 2018, pp. 92–97.

[3] J. Chen, S. Wang, and M. Tehranipoor, "Critical-reliability path identification and delay analysis," *J. Emerg. Technol. Comput. Syst.*, vol. 10, no. 2, mar 2014.

[4] Y. M. Kim, Y. Kameda, H. Kim, M. Mizuno, and S. Mitra, "Low-Cost Gate-Oxide Early-Life Failure Detection in Robust Systems," in *Proc. IEEE VLSI Circuits Symp. (VLSIC)*, Jun. 2010, pp. 125–126.

[5] M. Kampmann, M. A. Kochte, C. Liu, E. Schneider, S. Hellebrand, and H. Wunderlich, "Built-In Test for Hidden Delay Faults," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Syst. (TCAD)*, vol. 38, no. 10, pp. 1956–1968, Oct. 2019.

[6] H. Amrouch, G. Pahwa, A. D. Gaidhane, C. K. Dabhi, F. Klemme, O. Prakash, and Y. S. Chauhan, "Impact of variability on processor performance in negative capacitance finfet technology," *IEEE Trans. on Circuits and Systems*, vol. 67, no. 9, pp. 3127–3137, 2020.

[7] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale finfets," in *Intl. Electron Devices Meeting*, 2011, pp. 5.4.1–5.4.4.

[8] B. Cory, R. Kapur, and B. Underwood, "Speed binning with path delay test in 150-nm technology," *IEEE Design and Test of Computers*, vol. 20, no. 5, pp. 41–45, 2003.

[9] A. Datta, S. Bhunia, J. H. Choi, S. Mukhopadhyay, and K. Roy, "Speed binning aware design methodology to improve profit under parameter variations," in *Proc. IEEE Asia-South Pacific Design Automation Conf. (ASP-DAC)*, ser. ASP-DAC '06. IEEE Press, 2006, p. 712–717. [Online]. Available: https://doi.org/10.1145/1118299.1118466

[10] J. Bhasker and R. Chadha, *Static timing analysis for nanometer designs: A practical approach*. Springer Science & Business Media, 2009.

[11] A. Devgan and C. Kashyap, "Block-based static timing analysis with uncertainty," in *ICCAD-2003. International Conference on Computer Aided Design (IEEE Cat. No. 03CH37486)*. IEEE, 2003, pp. 607–614.

[12] S. Natarajan *et al.*, "A 14nm logic technology featuring 2¡sup¿nd¡/sup¿-generation finfet, air-gapped interconnects, self-aligned double patterning and a 0.0588 µm¡sup¿2¡/sup¿ sram cell size," in *IEEE Intl. Electron Devices Meeting*, 2014, pp. 3.7.1–3.7.3.

[13] Tin Kam Ho, "Random Decision Forests," in *Proc. of Int'l Conf. on Document Analysis and Recognition*, vol. 1, 1995, pp. 278–282 vol.1.

[14] A. Gattiker, S. Nassif, R. Dinakar, and C. Long, "Timing yield estimation from static timing analysis," in *Proc. of the IEEE Intl. Symposium on Quality Electronic Design*, 2001, pp. 437–442.

[15] Z. P. Najafi-Haghi, M. Hashemipour Nazari, and H.-J. Wunderlich, "Variation-Aware Defect Characterization at Cell Level," in *Proc. IEEE European Test Symp. (ETS)*, May 2020, pp. 1–6.

[16] Z. P. Najafi-Haghi and H.-J. Wunderlich, "Resistive Open Defect Classification of Embedded Cells under Variations," in *Proc. IEEE Latin-American Test Symposium (LATS)*, 10 2021, pp. 1–6.

[17] M. Martins, J. M. Matos, R. Ribas, A. Reis, G. Schlinker, L. Rech, and J. Michelsen, "Open Cell Library in 15nm FreePDK Technology," in *Proc. Int'l Symp. on Physical Design (ISPD)*, Apr. 2015, pp. 171–178.

[18] J. García Gervacio and V. Champac, *Timing Performance of Nanometer Digital Circuits under Process Variations*, Apr. 2018.

[19] J. P. Shen, W. Maly, and F. J. Ferguson, "Inductive Fault Analysis of MOS Integrated Circuits," *IEEE Design and Test*, vol. 2, no. 6, pp. 13–26, Dec. 1985.

[20] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - a survey," *IEEE Trans. on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 4, pp. 476–487, 2005.

[21] S. Marsland, *Machine Learning: An Algorithmic Perspective*, 2nd ed. Chapman and Hall/CRC, 2014.

[22] NanGate Inc. (2017) NanGate15nm Open Cell Library. [Online]. Available: http://www.nangate.com/