Reliability Assurance in Early-Life-Failure Test through Improved Nearest Neighbor Regression

Tai Song, Huaguo Liang, and Zhengfeng Huang, and Jie Hou

Abstract During manufacturing test, researchers usually overlook the importance of process variation defects and marginal defects, which can seriously affect test results of Early-Life-Failure (ELF). Theoretically, machine learning classification methods can be used to identify these two defects. In fact, when features overfitting or data distribution overlap seriously, classifiers perform poorly, it will not achieve the desired results. This paper first-ever proposes a kind of data preprocessing method combines improved K-Nearest Neighbors (KNN) regression classifier, so that the classification results will be enhanced in terms of classification performance. Experiment results demonstrate that the predictive accuracy is 45% higher than the traditional logistic regression method. This proposed method will drive critical new requirements for fault modeling, test generation and test application, and implementing them effectively will require a new level of collaboration between process and product developers.

key words: Early Life Failure, Process Variation, Marginal Defects, Machine Learning (ML)

Classification: Integrated circuits (logic)

1. Introduction

As semiconductor manufacturing process moves into Fin field-effect transistor (FinFETs) era, more and more marginal defects are observed due to extremely small feature sizes and complex manufacturing processes [1]. Such defects may produce a small amount of extra delay to paths of a circuit, but it’s not large enough to be detected and pass the conventional manufacturing test[2]. Test without considering the impact of marginal defects will reduce reliability. Therefore, marginal defects become great challenges for ELF and traditional test methods become less effective [3].

These new defects often cause marginal behaviors, not gross failures, with subtle signatures that differ significantly from both traditional defects and from parametric process variations[4]. Marginal defects degrade further when they go in the field and make device fail after a short period of time [5][6]. The preferred approach is to eliminate defect mechanisms until manufacturing can deliver a product that meets customer quality requirements without screening, using only sample testing for quality assurance[7]. This leads to faster, cheaper manufacturing and a higher quality product. In order to meet this requirement, it is necessary to detect and screen defective units with machine learning classification methods. This technology has stood the test of time, particularly when extended through the traditional fault model to add data-driven fault coverage to structural tests.

However, some circuit delay behavior has been essential to distinguish between marginal defects and process variation defects. Circuit delay is one of the most important reliability indicators [5][6]. In particular, very small delay faults may indicate marginal hardware that can degrade further under stress. They can be “hidden” at nominal frequency and only be detected at higher frequencies (“faster-than-at-speed test” / FAST) [8]. Therefore, conventional approaches for testing reach their limitations and new methods must be investigated and developed [9]. Conventional fault models and test method may not be able to effectively handle such marginal defects with a small delay [10].

For more than a decade, intersecting trends in process variations have been driving an important shift in defect populations. This shift undermines the effectiveness of a simple stuck-at based test solution. Data-based classification algorithms have improved continuously with the change of defects, minimize yield loss [11-17], ML can be used to distinguish between marginal defects and process variation defects based on circuit delay, depend on different delay distribution. The classification results can be able to locate the defects and identify the root cause. High quality predictive results from classification algorithms can help the failure analysis engineer pick the marginal defects as soon as possible, thus improving the test success rate and reliability by reducing the turnaround time and cost.

Many works based on machine learning methods have been published in the field of testing. In [11-13], the authors discussed the usefulness of test data analysis based on their experiences and measurements, but did not consider process variations. In [12], critical parameters were focused on pass/fail characteristics of past test data, but did not detail correlation of performances and characteristics measured...
accuracy. In [14], hidden characteristics of test data that make a correlation between dies and test items were investigated, but did not consider data overfitting. In [16], a wafer test flow was optimized with a graphical model and in [17], defect characteristics by wafer mapping were investigated, but did not consider the marginal defects. A method in [18] predicts test results using Support Vector Machine (SVM) with a weighted dynamic time warping kernel function, but did not consider classification [19].

As we can see, most of previous works are based on conventional fault models [11-18], lack of research on marginal defects and process variation defects with machine learning classification methods. To bridge this gap and formally show that, in a quest to reduce ELF and increase reliability, the work presented in this paper combines the concepts of regression and KNN algorithm with data preprocessing method. This paper presents an innovative classification approach to optimize predict accuracy for process variation defects.

2. Background

With transistors per area doubling in each generation, the numbers make it ever more likely a few will be imperfect. Indeed, the line between marginal defect and process variation defect can already sometimes seem arbitrary [20]. Given all of these challenges, it should come as no surprise that defects due to small variations in placement, alignment and the shaping of complex features are on the rise[21]. Such lithography-based defects often occur within the boundaries of typical logic gates and standard cell libraries, at the level of transistor gate features and lower level interconnect, where geometries are smallest and lithography is most challenging, such defects often cause subtle parametric variations in circuit behavior[22]. These failures can be quite difficult to screen. They do not cause gross failures often associated with simple particles. The traditional stuck-at fault model may not be adequate for managing such defects[23].

The nature of these marginal defects is to have a more subtle impact on electrical behavior. They simply alter the silicon’s functions parametrically[24]. That is, marginal unit fails (or will in the future fail, or intermittently fails) at some voltage-temperature-frequency set point within the product's specified operating envelope, while success set points within the specified envelope[25]. Marginal defect poses a significant challenge to currently prevalent test method.

Marginal defects in chips which add a small amount of extra delay to the nominal path delay. This added delay is sometimes smaller than the available slack of the paths, therefore they will not make device fail at manufacturing test with nominal frequency[26]. However, since they are indicators of marginalities, when degraded, marginal defects will make device fail after a short period of working time at their early life. But process variation defects will remain the same and will not degrade further.

The problem here is that both of process variations and marginal defects will produce delay in the same order of magnitude. They will both pass the manufacturing test, which should be avoided for marginal defects group. This makes it critical to distinguish between these two groups.

Fig. 1 shows the electric level simulation which performed with HSPICE under different voltages to measure the propagation delay for both groups of cells. Cells with just process variations (fault-free cells) and cells with process variations plus marginal defects (faulty cells). The data distributed overfitting seriously, and they have a great chance to pass manufacturing test which make it great challenge to distinguish between these two groups.

In this paper, study has also been investigated the predictive accuracy in relation among ratio of the size of training data to that of test data, detail of the selection for training and test data is explained later. Main contributions of this paper are shown as follows:
1) A linear discriminant analysis (LDA) data preprocessing method is proposed.
2) The regularization method and the optimal training set and test set are selected.
3) KNN combines logistic regression methods for higher accuracy.

The remainder of the paper is organized as follows: The road map of the paper is briefly sketched in Fig.2. Section 3 is to clarify the LDA data preprocessing method and KNN regression method. Section 4 outlines the experiment results that are briefly reviewed. Section 5 summarizes this paper.

3. Proposed test flow and marginal defect detection

3.1 Data preprocessing

LDA is a dimension reduction technique which can maximize differences and minimize the same features [27]. It is used for dimension reduction as feature extraction, classification or data visualization. The main idea of LDA is to find the linear combination of features that separates data from two classes. By virtue of eigen decomposition, the resulting combination gives a more condensed representation of
the data, with discriminative information preserved for later classification, it can also be helpful to avoid overfitting by minimizing the error in parameter estimation. The proposed method is shown in the following five steps:

**Step 1**: Computing the d-dimensional mean vectors
In this first step, with a simple computation of the mean vectors $m_i$ (i = 1 and 2) of the 2 different classes. (fault-free = 1 and fault = 2). (Refer Eq. 1).

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x$$

**Step 2**: Computing the Scatter Matrices
There are 19 voltage properties (Fig 1) from 0.3 to 1.2. Compute the two 19x19-dimension matrices: within-class and the between-class scatter matrix.

- **Within-class scatter matrix $s_w$** is computed by the following equation: (Refer Eq. 2 and 3).

$$s_w = \sum_{i=1}^{c} s_i$$

$$s_i = \sum_{x \in D_i} (x - m_i) * (x - m_i)^T$$

Alternatively, we could also compute the class-covariance matrices by adding the scaling factor $\frac{1}{n_i}$ to the within-class scatter matrix, so that our equation becomes: (Refer Eq. 4 and 5)

$$\sum_{i} i = \frac{1}{N_i - 1} \sum_{x \in D_i} (x - m_i) * (x - m_i)^T$$

**Step 3**: Solving the generalized eigenvalue problem for the matrix $S_w^{-1} S_B$ (Refer Eq. 9).

$$\Sigma V = \lambda V$$

Here, $\lambda$ is the eigenvalue, and V is also an eigenvector that same eigenvalue, since

$$AV = \lambda V$$

Where

$$A = S_w^{-1} S_B$$

$$V = \text{Eigenvector}$$

$$\lambda = \text{Eigenvalue}$$

Next, we solve the generalized eigenvalue problem for the matrix $S_w^{-1} S_B$ to obtain the linear discriminants.

**Step 4**: Selecting linear discriminants feature subspace.
Sorting the eigenvectors by decreasing eigenvalues, not only merely projecting the data into a subspace that improves the class separability, but also reduces the dimensionality of our feature space, where the eigenvectors will form the axes of this new feature subspace.

**Step 5**: Transforming the samples onto the new subspace.
In the last step, the dimensional matrix W that computed to transform our samples onto the new subspace via the equation (Refer Eq. 10).

$$Y = X \times W$$

### 3.2 Regularization

In this section, two kinds of regularization metrics are briefly introduced highlighting the effect of interaction between variables. The performance of predictive model developed after applying LDA method. The root mean square (RMSE) values are statistically establish measures to quantify this difference.
MSE in mathematical statistics refers to the expected value of the square of the difference between the parameter estimate and the parameter true value, which is a convenient method to measure the "average error". (Refer Eq. 11). MSE can evaluate the degree of change of data.

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2 \]  \hspace{1cm} (11)

Where \( y_i \) is the real data, \( \hat{y} \) is the predicted data, \( n \) is the number of samples, when the MSE closer to 0, the better the model selection and fitting, and the more successful the data prediction. The next RMSE is the same effect as MSE (Refer Eq. 12). Which means the smaller the value of RMSE, the better the accuracy of the prediction model to describe the experimental data.

\[ \text{RMSE} = \sqrt{\text{MSE}} \]  \hspace{1cm} (12)

3.3 Improved KNN Classification Method

This paper intends to further pursue the accuracy, and more effective points were selected in prediction. The KNN combined logistic regression with regularization methods were used. Standardized \( \beta \) obtained in the simple logistic regression (SLR) indicates the importance in selection of the KNN points. 

The step is as follows: First perform the SLR with 3 kinds of regularization methods, and obtain the effective feature variables (as a weighting function). Then, using these values of \( \hat{\beta} \), we redefine the weighted distance \( d_1(a_i, b_j) \) between the two points as: (Refer Eq. 13).

\[ d_1(a_i, b_j) = \sqrt{\sum_{l=1}^{p} |\hat{\beta}_l| (a_{il} - b_{jl})^2} \]  \hspace{1cm} (13)

Where \( \hat{\beta} \) is the value obtained by regularization methods. This means that the important variables are actively adopted in searching for the KNN points. The results using this combination methods will be shown in Fig 7. We may further impose an optimal weight to each feature variable point, such as (Refer Eq. 14).

\[ d_N(a_i, b_j) = \sqrt{\sum_{l=1}^{p} N |\hat{\beta}_l| (a_{il} - b_{jl})^2} \]  \hspace{1cm} (14)

Where \( N \geq 0 \) can be an acceleration coefficient for weighting. To find the optimal value of \( N \), the obtained MSE and RMSE for \( N = 1, 3, 5, 10, 50, 100 \) are in Fig 8. It reveals that this method can be optimal for all the KNN regression methods.

4. Experiments and results shown

The data used in this experiment was generated from the NAND cell, FinFET 15nm technology, Monte Carlo simulation to make the fault-free class of 1000 cells, fault simulation by injecting marginal defect to make the faulty class of 1000 cells. Machine learning classification experiment platform adopted python 3.7 and scikit-learn library, test_size=0.2, random_state=0. Fault-free cells are cells only with process variation. To model these cells a population of cells from same type will be modeled in which every instance has a process parameter value in the acceptable range. The process parameters have gaussian distribution around their nominal values. FinFET has very lightly doped channel accordingly no significant random dopant fluctuation will happen. Therefore, variation in device geometry in particularly gate length (L) and gate width (W) will be focused in this work which will result into variation in threshold voltage. To model the fault-free cells, Monte Carlo simulation with 1000 iterations will be run, applying an average for standard deviation of 20% on both gate length (L) and gate width (W). Faulty cells are cells inject various marginal defect mechanisms are well known such as resistive shorts, resistive opens, gate-oxide pinholes, each marginal defect is injected into a population of fault-free cells with standard deviation over process parameters which makes a faulty group of cells. For each mechanism many possible placements are considered. Combination of mechanisms and placements, provides an acceptable coverage over the types of defects we are interested in. Therefore, there would be a package of faulty groups each corresponds to a different marginal defect mechanism and various placement of them.

4.1 Data analysis and preprocessing

According to the feature selection algorithm (LDA), the compare of result scatter plot as is shown in Fig.3 and Fig.4. Features of the same class are basically brought together, but they are still not completely separated and need further operations.

4.2 Traditional SLR Result

The training data are selected randomly but different from each other, and the test data are collected from the rest of the sampled data. The ratio of the size of training data to that of the test data are 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1.
The resampling is repeated 10 times. As is shown in Fig.5. The result might expect for larger the size of training data, the smaller the value of RMSE for test data. Fig.5. Shows the best value for RMSE is 0.45 at 9:1.

4.3 Regularization Method
With regularization method to make the RMSE minimum. Here, 1) lasso [28], 2) ridge [29], 3) elastic net [30] were investigated. Fig.6 shows that the value of RMSE is concentrated around 0.5 in any ratio, it indicates that the regularization method is more robust than SLR method.

4.4 KNN Ratio
After data preprocessing and using the regularized method, the accuracy rate experiment was performed and it was found that there is still improvement through combing KNN method. Fig. 7 shows that the value of RMSE is concentrated around 0.425 when the ratio is at 8:2, it is obvious to find that the ratio of 8:2 is more robust than that of 9:1.

4.5 KNN Combined with SLR Method
Through the equation (Refer Eq. 13) and (Refer Eq. 14), using KNN combined logistic regression, choose ratio at 8:2. Fig.8 shows the minimum RMSE value are available when 1NN is used, the value achieved 0.37, the results are better than previous.

4.6 Predictive Accuracy Comparison
In order to compare the accuracy of the proposed method, the traditional classic LSR and KNN were used. Fig.9 (a) shows that the predictive accuracy before and after data preprocessing (LDA) method for SLR is 50% and 85%, KNN after LDA is 75%, which shows that LDA can improve accuracy dramatically. Fig.9 (b) shows that the combined method (KNN+SLR+Regularization) achieves the highest accuracy up to 95%. Comparatively, the predictive accuracy is higher 45% than traditional SLR (which is 50%). Moreover, regularization reduces the dependence of training data and test data, which shows the robustness of our work.

5. Conclusion
While testing is challenged by process scaling, it is also called upon to create ever more inexpensive and effective test methods. Therefore, we first-ever proposes an improved
KNN regression method for overfitting data, the improvement accuracy up to 95%, which is 45% higher than traditional SLR and 20% than KNN. The optimum accuracy could be obtained stably when the ratio of size of training data to that of the test data is 8:2, the number of nearest neighbors is 1. This paper will reduce the early life failure and test costs, increase the reliability while keeping the yield high. It is apparent that high-quality predictive results are critical for a successful ELF identification.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under the Granted No. 61674048, 61874156 in part by Science and Technology on Electronic Test & Measurement Laboratory of China under Grant No. 61674048, and test costs, increasing the reliability while keeping the yield high. It is apparent that high-quality predictive results are critical for a successful ELF identification.

References

[1] H. H. Chen, “Data analytics to aid detection of marginal defects in system-level test,” 2016 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), Hsinchu, 2016, pp. 1-4.

[2] X. Qian et al., “Distinguishing Resistive Small Delay Defects from Random Parameter Variations,” 2010 19th IEEE Asian Test Symposium, Shanghai, 2010, pp. 325-330.

[3] M. Tehranipoor et al. “Test and diagnosis for small-delay defects,” New York [u.a.]: Springer, 2011.

[4] M. O. Simsir et al., “Fault modeling for FinFET circuits,” 2010 IEEE/ACM International Symposium on Nanoscale Architectures, Anaheim, CA, 2010, pp. 41-46.

[5] F. Hopsch et al., “Variation-Aware Fault Modeling,” 2010 19th IEEE Asian Test Symposium (ATS), Shanghai, pp. 87-93.

[6] H. Tang, T. Tai, W. Cheng, R. Benware and F. Hapke, “Diagnosing timing related cell internal defects for FinFET technology,” VLSI Design, Automation and Test (VLSI-DAT), Hsinchu, 2015, pp. 1-4.

[7] T. Song et al., “Novel Application of Deep Learning for Adaptive Testing Based on Long Short-Term Memory,” 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, CA, USA, 2019, pp. 1-6.

[8] M. Kampmann and S. Hellebrand, “X marks the spot: Scanflip-flop clustering for faster-than-at-speed test,” in 2016 IEEE 25th Asian Test Symposium (ATS), Nov. 2016, pp. 1-6.

[9] Y. Huang, “The Emerging Applications of Machine Learning in Testing,” International Test conference, 2017.

[10] F. K. Javvaji, et al., “Test set identification for improved delay defect coverage in the presence of statistical delays,” 2018 19th International Symposium on Quality Electronic Design (ISQED), Santa Clara, CA, 2018, pp. 14-19.

[11] Rob Aitken, “Opportunities in Machine Learning and Test,” International Test Conference, 2017.

[12] James Moynie, “Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing,” Processes, vol.5 issue 3, 39, 2017.

[13] Sounil Biswas, et al., “Machine learning and data analysis in test,” VLSI Test Symposium, 2014.

[14] Chih-Ying Tsai, et al., “Predicting Vt mean and variance from parallel Id measurement with model-fitting technique,” VLSI Test Symposium, 2016.

[15] Hsu Chan Kai, et al. “Variation and Failure Characterization Through Pattern Classification of Test Data From Multiple Test Stages,” International Test Conference, 2016.

[16] Ali Ahmadi, et al., “Harnessing process variations for optimizing wafer-level probe-test flow,” International Test Conference, 2016.

[17] Fatima Adly, et al., “Machine-Learning-Based Identification of Defect Patterns in Semiconductor Wafer Maps: An Overview and Proposal,” International Parallel & Distributed Processing Symposium Workshops Pages 420-429, 2014.

[18] Young-Seon Jeong, “Semiconductor Wafer Defect Classification Using Support Vector Machine with Weighted Dynamic Time Warping Kernel Function,” Industrial Engineering & Management Systems, vol. 16 no.3, pp. 420-426, 2017.

[19] H. Hirose, et al. “NNRMLR: A Combined Method of Nearest Neighbor Regression and Multiple Linear Regression,” 2012 IIAI International Conference on Advanced Applied Informatics, Fukuoka, pp. 351-356.

[20] P. G. Ryan, I. Aziz, W. B. Howell, T. K. Janczak and D. J. Lu, “Process defect trends and strategic test gaps,” 2014 International Test Conference, Seattle, WA, 2014, pp. 1-8.

[21] X. Qian, Chao Han and A. D. Singh, “Detection of gate-oxide defects with timing tests at reduced power supply,” 2012 IEEE 30th VLSI Test Symposium (VTS), Hyatt Maui, HI, 2012, pp. 120-126.

[22] C.-W. Tseng, R. Chen, P. Nigh und E. J. McCluskey, „MINVDD testing for weak CMOS ICs,“ in Proc. 19th IEEE VLSI Test Symp. VTS 2001.

[23] A. N. Bhoj, M. O. Simsir and N. K. Jha, “Fault Models for Logic Circuits in the Multigate Era,” in IEEE Transactions on Nanotechnology, vol. 11, no. 1, pp. 182-193, Jan. 2012.

[24] T. Wang, X. Cui, Y. Ni, K. Liao, N. Liao, D. Yu und X. Cui, “Reliability evaluation of high-performance, low-power FinFET standard cells based on mixed RBB/FBB technique,” Journal of Semiconductors, 2017.

[25] J. T. -. Chang und E. J. McCluskey, “Detecting delay flaws by low-voltage testing,” Proceedings International Test Conference 1996. Test and Design Validity, Washington, DC, USA, 1996, pp. 367-376.

[26] K. Chiang, et al. “Fault Simulation and Test Pattern Generation for Cross-gate Defects in FinFET Circuits,” 2015 IEEE 24th Asian Test Symposium (ATS), Mumbai, pp. 181-186.

[27] F. Ye, Z. Shi and Z. Shi, “A Comparative Study of PCA, LDA and Kernel LDA for Image Classification,” 2009 International Symposium on Ubiquitous Virtual Reality, Gwangju, 2009, pp. 51-54.

[28] Tibshirani, R., “Regression shrinkage and selection via the lasso,” The Royal Statistical Society, Vol. 58, pp. 267-288 (1996).

[29] Hoerl, A. E. et al., Ridge Regression: Biased Estimation for Correlated Variables and Model Selection, du Pont de Nemours & Co, Vol. 12, No. 1 (1970).

[30] Young-Seon Jeong, “Fault Models for Logic Circuits in the Multigate Era,” in IEEE Transactions on Nanotechnology, vol. 11, no. 1, pp. 182-193, Jan. 2012.

[31] F. Ye, Z. Shi and Z. Shi, “A Comparative Study of PCA, LDA and Kernel LDA for Image Classification,” 2009 International Symposium on Ubiquitous Virtual Reality, Gwangju, 2009, pp. 51-54.

[32] Tibshirani, R., “Regression shrinkage and selection via the lasso,” The Royal Statistical Society, Vol. 58, pp. 267-288 (1996).

[33] Hoerl, A. E. et al., Ridge Regression: Biased Estimation for Correlated Variables and Model Selection, du Pont de Nemours & Co, Vol. 12, No. 1 (1970).

[34] Young-Seon Jeong, “Fault Models for Logic Circuits in the Multigate Era,” in IEEE Transactions on Nanotechnology, vol. 11, no. 1, pp. 182-193, Jan. 2012.

[35] F. Ye, Z. Shi and Z. Shi, “A Comparative Study of PCA, LDA and Kernel LDA for Image Classification,” 2009 International Symposium on Ubiquitous Virtual Reality, Gwangju, 2009, pp. 51-54.