

Experimental Study of Pass/Fail Threshold Determination Based on Gaussian Process Regression

Daisuke Goeda¹ Tomoki Nakamura² Masuo Kajiyama²
Makoto Eiki² Takashi Sato³ Michihiro Shintani¹

¹Graduate School of Science and Technology, Kyoto Institute of Technology

²Nagasaki TEC, Sony Semiconductor Manufacturing Corporation

³Graduate School of Informatics, Kyoto University

Email: dgoeda@vlsi.es.kit.ac.jp, takashi@i.kyoto-u.ac.jp, shintani@kit.ac.jp

Abstract—As large-scale integrated circuits (LSIs) grow in size and complexity, improving LSI test quality without increasing test costs becomes challenging. LSIs manufactured with advanced technologies exhibit significant variation in characteristics. The variation makes it difficult to determine the pass/fail threshold that distinguishes good and bad chips. Therefore, the yield loss and test escape ratios are increasing. Particularly, automotive semiconductors must comply with test standards set by the Automotive Electronics Council (AEC), resulting in increased yield loss and test escape compared to carefully designed threshold. To address this issue, a method that utilizes Gaussian process regression to determine the pass/fail threshold with has been proposed for power MOSFETs [1]. This paper applies this approach to industrial LSI test data and confirms that it is equally effective for both power MOSFETs and LSIs. The method reduces yield losses and test escapes by 0.019% and 35.5% when compared to conventional methods in compliance with the AEC standard.

I. INTRODUCTION

The manufactured large-scale integrated circuits (LSIs) are tested with various test items under multiple environments before shipping. It is considered defective if the LSI fails to meet any of the requirements. In other words, only LSIs that meet all required specifications are shipped as good products [2]. However, due to variations in the manufacturing process, characteristics of the LSI vary greatly, making it challenging to differentiate between good and faulty products. “Yield loss,” in which faulty products are classified as faulty even if they are good products, and “test escape,” in which faulty products are shipped as good ones, have become a serious problem. Furthermore, automotive semiconductor products must comply with the test standard (AEC-Q001) established by the Automotive Electronics Council (AEC) [3]. AEC-Q001 utilizes the dynamic part average testing (DPAT) method to determine the threshold based on Six Sigma for each wafer. The DPAT method has limitations in capturing

chips with local deviation as it is based on wafer-wide distribution. This makes it difficult to test automotive LSIs without causing yield losses and test escapes.

Many studies have been conducted to appropriately set pass/fail thresholds. The example includes nearest neighbor residual (NNR) method [4]. The NNR method predicts trends based on the characteristics of adjacent chips, utilizing local trends on a wafer. However, when there are multiple failed LSIs present, the predictions may be biased due to the utilization of only a limited number of adjacent chips for the prediction. Recently, a new method has been proposed for setting pass/fail thresholds by learning the characteristics of good LSIs through machine learning [5,6]. However, the results of machine learning lack clear explanation why the chip is passed/failed, thus have limited application in actual testing environments.

This study reports the results of applying the method proposed in [1] to industrial LSI test data. This method utilizes Gaussian process regression (GPR) [7] to predict the measurement result of the target LSI and compare the with the actual measurement result. GPR calculates the prediction accuracy based on Bayes’ theorem, thus enabling probabilistic comparisons. In [1], silicon carbide (SiC) power MOSFET was targeted, and their threshold voltage and on-resistance were adopted as the test items. As there is no essential difference between SiC MOSFETs and silicon LSIs in terms of semiconductor manufacturing variations, the GPR-based method is also expected to be effective for LSIs. This evaluation indicates that yield losses and test escapes are reduced by 0.019% and 35.5%, compared to the DPAT and NNR methods, respectively. Furthermore, analysis of the LSIs that could be revealed that the LSIs that could be detected by the GPR-based method included those detected by the DPAT and NNR methods. It is expected that AEC-Q001 can be replaced by the GPR-based method.

The remainder of this paper is organized as follows: Sec. II briefly summarizes the GPR, which is the core technique in [1]. This section also provides an overview of the DPAT and NNR methods as conventional methods and discusses these issues. Sec. III describes the details of the GPR-based method [1]. In Sec. IV, we evaluated the

GPR-based method utilizing industrial LSI test data for comparison with the DPAT and NNR methods. Finally, we will conclude this paper in Sec. V.

II. PRELIMINARIES

A. Gaussian process regression

A Gaussian process is employed to estimate the function $y = f(x)$ from the input variable x to the output variable y [7]. The estimated function $f(\cdot)$ has nonlinear characteristics and can be predicted even if it is a complex model. In addition, the Gaussian process is based on the Bayes' theorem, and the estimated function is obtained not as a single function but as a distribution function, allowing the uncertainty of the estimation to be expressed as a predictive distribution.

Here, we will explain the GPR. The variables $(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N))$ and $\mathbf{X}_{\text{test}} = (\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_M^*)$ denote the training and test datasets, respectively, where $M \gg N$. Furthermore, the kernel function f_{kernel} is provided as an input. From these inputs, the GPR utilizes the calculated prediction model $f(\cdot)$ to return the mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_M)$ and variance $\boldsymbol{v} = (v_1, v_2, \dots, v_M)$ of the predicted values corresponding to the test input \mathbf{X}_{test} .

For each element of $\mathbf{X}_{\text{train}}$, the kernel matrix K of the training dataset is calculated with a kernel function. Then, by modeling the multidimensional normal distribution, as in

$$p(y_m^* | \mathbf{x}_m^*, \mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) = \mathcal{N}(\mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{y}_{\text{train}}, \mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{k}_*), \quad (1)$$

the probability density function of the predicted value Y_m^* corresponding to \mathbf{X}_m^* is derived. In Eq. (1), \mathbf{k}_* is the covariance matrix between the training and test datasets and \mathbf{k}_{**} is the covariance matrix of test datasets.

From Eq. (1), the mean and variance of Y_m^* can be derived analytically. The examples of the kernel function include linear kernels, squared exponential kernels, and Matern kernels [8]. The appropriate kernel can be selected based on the specific situation.

In Eq. (1), because \mathbf{K} is a covariance matrix, $f_{\text{kernel}}(\mathbf{x}, \mathbf{x}')$ becomes larger when \mathbf{x} and \mathbf{x}' have similar values. Consequently, $f(\mathbf{x})$ and $f(\mathbf{x}')$ exhibit similar values. In predicting the characteristic distribution of wafer space, the variance v is utilized to check the uncertainty of the prediction against the predicted mean μ . Neighboring chips on a wafer are known to have similar characteristics, and this characteristic makes Gaussian processes, which estimate the characteristic variance on a wafer in a Bayesian manner, highly compatible with the prediction of smoothly varying characteristics in the wafer space [9,10]. Furthermore, verifying the reliability of the predicted function through variance facilitates operation in an actual test environment.

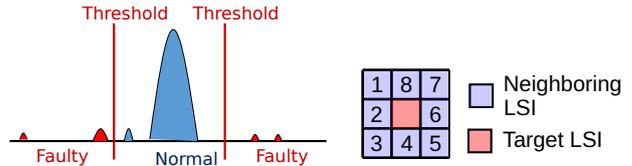


Fig. 1. Concept of the DPAT method.

Fig. 2. Eight neighbors for interpolation.

B. Related works

B.1. Dynamic part average testing (DPAT)

The DPAT method has been standardized as AEC-Q001 and must be applied to automotive semiconductors. As illustrated in Fig. 1, the DPAT method sets a single pass/fail threshold for all the chips on a wafer. The average is calculated from the measured values of all N chips on the wafer, and $\pm 6\sigma$, where σ is the standard deviation of the measured values for N chips, from the average is set as the threshold value p_d for determining whether the target chip is good or not as $p_d = \frac{1}{N} \sum_{i=1}^N p_i \pm 6\sigma$. In the DPAT method, p_d is determined for each wafer, and chips with measured value of i -th chip, p_i , exceeding this value are considered as bad chips.

However, ignoring the characteristic variations occurring within the LSI wafers manufactured with advanced technologies is no longer possible. In addition, the probability distribution of the measurements does not always follow a normal distribution. The DPAT method, which sets a single threshold for all chips on a wafer, is inappropriate to use without incurring yield losses and test escapees. LSI testing becomes more difficult as manufacturing processes become more complex and the relative magnitude of manufacturing variation increases. However, AEC-Q001 has not been updated for over a decade, and there is a strong need for new testing methods to supersede this standard.

B.2. Neighbor Nearest Residual (NNR) method

The NNR method is a test method that considers local variation trends on a wafer [4]. The estimations are made based on the measured values of the neighboring chips of the target chip, as shown in Fig. 2. The NNR method leverages the characteristics of a systematic variation component in the process variation of a manufactured LSI. The systematic variation component represents a gradual spatial change on the wafer and is modeled using a low-order polynomial function of the chip coordinates on the wafer. Therefore, the characteristic values measured at neighboring coordinates on the wafer are similar.

When the NNR method is applied to the measured value p , the NNR method estimates the average of the measured values obtained from the n_V chips around the target chip as illustrated in Fig. 2, as the measured value

p^* of the target chip, as follows: $p^* = \frac{1}{n_V} \sum_{i=1}^{n_V} p_i$. This takes advantage of the fact that neighboring chips have similar characteristics owing to the systematic variation component. Ideally, the estimated value p^* and the measured value p match. However, if the residual difference between p^* and p exceeds the threshold value, the target chip is determined to be faulty.

The NNR method does not consider the process variation across the entire wafer and only estimates the measured value of the target chip from the neighboring chips. If multiple faulty chips coexist among neighboring chips, p^* will be inaccurate, leading to a degradation in test accuracy.

III. LSI TEST BASED ON GAUSSIAN PROCESS REGRESSION

This section describes the method for setting test pass/fail threshold based on the GPR approach proposed in [1]. We will hereafter refer to it as the ‘‘GPR method’’. In this method, the spatial latent tendency of the chip characteristics is regarded as a latent function f , and its input \mathbf{X} is a vector of coordinates (x, y) on the measured wafer. Let \mathbf{y} be the measured values of their LSIs at the corresponding coordinates.

The concept of the proposed method is illustrated in Fig. 3. The measured value \mathbf{y} of a chip on a wafer contains both systematic and random variation components, which optimize the hyperparameters to describe the relationship between the chip coordinates \mathbf{X} and the measured value \mathbf{y} . Next, the posterior distribution $p(y^*)$ corresponding to each chip coordinate X^* is predicted with GPR. The mean of the distribution $p(y^*)$ is only one expression of the trend of y^* , which is represented by the distribution \mathbf{v} . The probability that a measurement is in that distribution implies the reliability of each chip based on the underlying characteristic trend. Thus, from the distribution $p(y^*)$, we can compute a $100(1 - \alpha)\%$ confidence interval, where α implies a rejection rate, which is utilized in this method to determine outlier (fail) chips. Those within confidence interval is regarded as good chips. In Fig. 3, although the measured values of Chips A and B are similar, the performance of Chip B deviates significantly from the predicted performance of Chip B. Therefore, it is considered that the chip contains latent defects and classified as fail.

In the GPR method, the ratio of chips that are determined to be faulty depends on the rejection rate α . Increasing α can eliminate possible outliers near the threshold value. However, setting α to be too large will result in chips that should be classified as pass being classified as fail. In [1], assuming the fault is a random process, the number of chips theoretically classified as outliers is compared with the number of chips actually classified as outliers while sweeping α . It is proposed to use the smallest α value that leads to a discrepancy between the theoretical value and the actual value.

Unlike the DPAT method, this method can be applied even when the distribution is non-normal. Furthermore, the systematic trend is estimated for the entire wafer and compared with actual measurements, unlike the NNR method. We can expect more accurate LSI testing with this method, as it is similar to well-established concept of systematic and random components verified with various measurements [11]. The obtained results are more explanatory than the neural-network based methods. Note that the GPR method is not done on the tester but in post-processing on a server for the measurements.

IV. EXPERIMENTS

Experiments were conducted with an industrial LSI test dataset. In the experiments, three methods, i.e., GPR, DPAT, and NNR, were applied to the dataset, and the yield loss and fault detection rates were compared.

A. Experimental setup

A set of industrial production test data with approximately 4,000 chips per wafer was provided by a semiconductor manufacturing company. Among multiple test items, we utilized a standby-current test in the following experiments. The test data includes nonfaulty (pass) and faulty chips. The faulty/nonfaulty labeling was carefully performed by LSI testing experts for all the wafers. The faulty chips include latent defects that conventional test items cannot detect. The measurement results for the first wafer of the first lot are presented in Fig. 4(a), where the systematic variation components are distributed in concentric circles.

All programs employed in the experiment were implemented in Python. GPR was performed with GPy [12] which is a Python package with a radial basis function (RBF) kernel [8] as f_{kernel} . Assuming we do not have prior information, the estimation was performed on all chips. The NNR method uses eight neighboring chips for estimation, as illustrated in Fig. 2.

To quantitatively evaluate the test performance, we defined *yield-loss ratio* and *fault-detection ratio* as follows.

Yield-loss ratio: Percentage of fault-free chips classified as faulty

Fault-detection ratio Percentage of faulty chips classified as faulty

Ideally, The yield-loss ratio should be 0%, and the fault-detection ratio should be 100%. Considering the semiconductor business, the yield-loss ratio should be less than 1%. In our experiment, we maximized the fault-detection ratio while maintaining a yield-loss ratio of less than 1%.

B. Estimation result

The NNR and GPR methods estimated the measured values of the LSI under testing. The results estimated

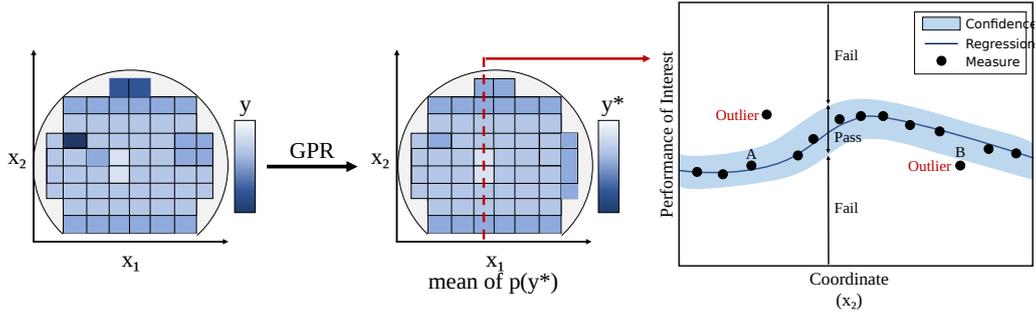


Fig. 3. Concept of a method for determining pass/fail threshold using Gaussian process regression [1]. The chips that fall outside the confidence interval (blue area) are determined as faulty.

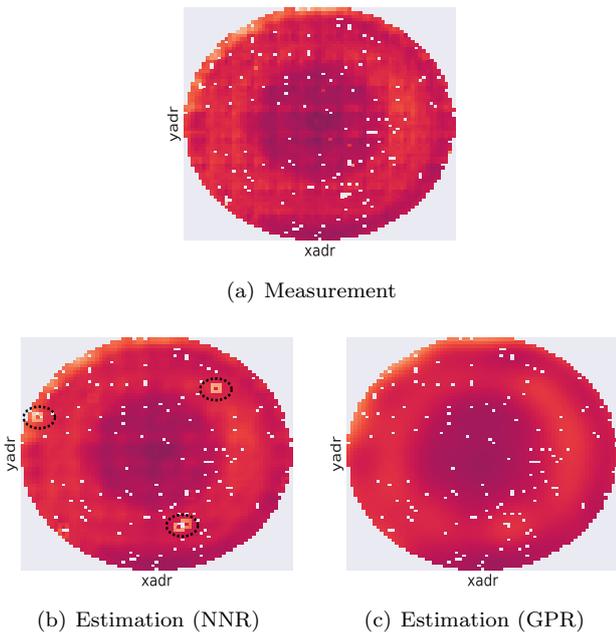


Fig. 4. Heat maps of the first wafer.

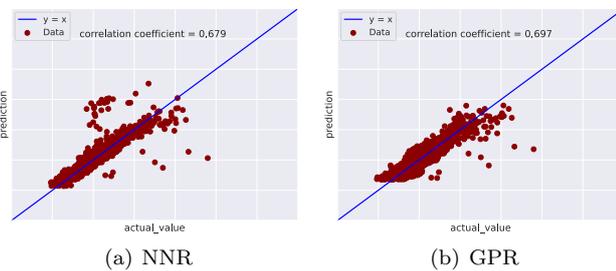


Fig. 5. Scatter plots between the measured and predicted values for the first wafer.

with each method are presented in Figs. 4(b) and 4(c). As described in Sec. II, p^* calculated with the NNR method gives biased estimations when there exist faulty chips in the neighborhood, as shown in Fig. 4(b). In contrast, in

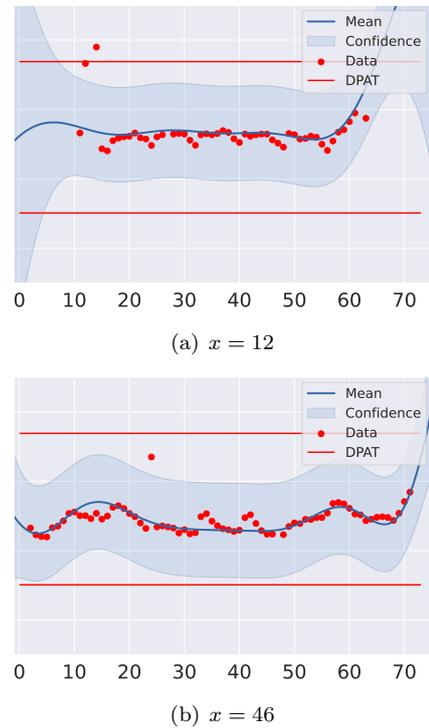


Fig. 6. Comparison of actual measurement data of the first wafer of the lot, estimations by GPR method and DPAT method.

the GPR method, a concentric trend in Fig. 4(a) was reproduced with the average value estimates, as shown in Fig. 4(c). Fig. 5 presents scatter plots comparing the measured and estimated values for a quantitative comparison. The correlation coefficients were 0.679 and 0.697 for the NNR and GPR methods, respectively. This result indicates that the Gaussian process method yields better prediction results.

Fig. 6 presents the y-coordinate plane with a fixed x-coordinate for the measured values and the estimated and DPAT method thresholds obtained by the GPR method for the first wafer. Here, the rejection rate α given to the GPR method is 0.0005, and the results at $x = 12$ and

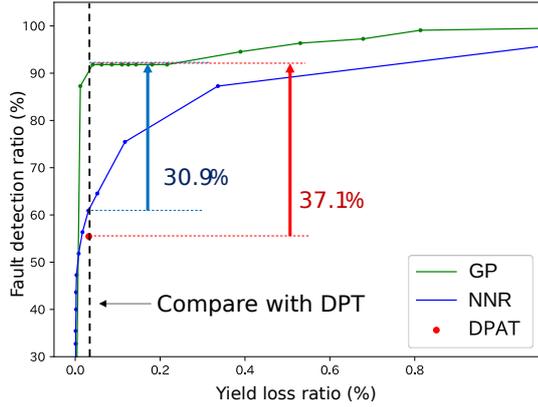


Fig. 7. ROC curve obtained from results for all wafers. Here, the yield-loss ratio is evaluated at 1% or less.

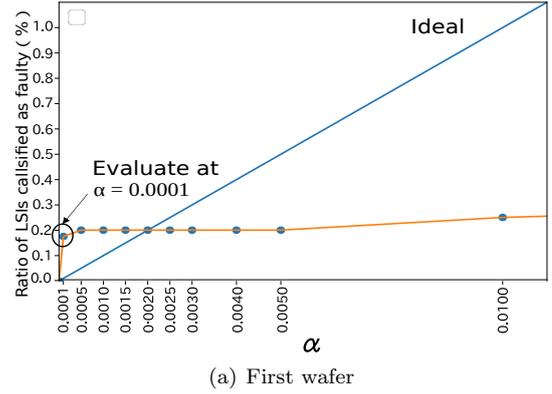
$x = 46$ are presented. As illustrated in the figure, the DPAT method provides a single pass/fail threshold for each wafer. In contrast, the GPR method can adaptively set the threshold for each chip, which is expected to provide a more flexible determination of pass and fail chips. In addition, the GPR method can identify faulty chips that deviate from the characteristic trend, which cannot be detected with the DPAT method. For example, in the results presented in Fig. 6, the chips with different trends deviating from the confidence interval can be confirmed. Because of page limitations, only the results for $x = 12$ and $x = 46$ on the first wafer are presented. Similar results were confirmed for the other wafers and coordinates.

C. Test performance comparison

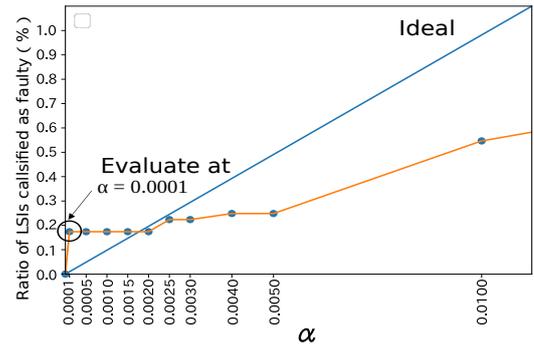
C.1. Comparison of the fault-detection ratio

To compare the test performance of each method, the results of the evaluation utilizing the receiver operating characteristic (ROC) curve are presented in Fig. 7. Here, we present the results for all the wafers. The horizontal and vertical axes represent the yield-loss and fault-detection ratios, respectively. When a curve is drawn in the upper left of the figure, it indicates better test performance result. In the figure, the result of DPAT is shown as a point because it determines the pass/fail threshold with Six Sigma. The GPR method has a higher fault-detection ratio than the DPAT and NNR methods. Specifically, the GPR method improved the fault-detection ratio by 37.1% compared with the DPAT method and by 30.9% compared with the NNR method when the yield-loss ratio of the DPAT method was considered, confirming an improvement in the detection performance of faulty chips.

The above evaluation does not discuss how to determine α , which is an important issue from a practical viewpoint. As mentioned in Sec. III, the threshold determination in the GPR method depends on the rejection rate α . Fig. 8 illustrates the relationship between the number of chips theoretically determined to be defective based on the re-



(a) First wafer



(b) Second wafer

Fig. 8. Comparison of theoretical and actual values of the number of chips determining defective products for each rejection rate α .

jection rate α (a straight line with $y = x$) and the number of chips determined to be faulty. Although the results for the first and second wafers are presented here as examples, similar trends have been observed for many other wafers. In [1], the usage of the smallest α that deviates from the theoretical value and the actual number was proposed because the number of fault detection chips increased along the theoretical value in the range where α was small. However, as shown in Fig 8, no similar trend was identified in this dataset. In this experiment, α is evaluated in the increments of 0.0001. The evaluation point alpha in the GPR method was 0.0001, which was the onset of a significant increase in the number of chips determined as defective, and the following evaluation was performed.

Based on the above discussion, the ROC curve at the rejection rate $\alpha = 0.0001$ is illustrated in Fig. 9. From this figure, by comparing the fault-detection rate of the GPR method with that of the NNR method, which yielded the same yield-loss ratio, we confirmed that the GPR method improved the detection rate by 35.5%. In addition, the GPR method improved the yield-loss ratio by 0.019% compared with the DPAT method, where this corresponds to an increase of ten good chips per wafer.

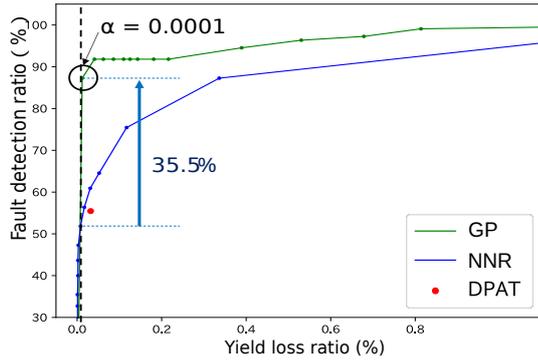


Fig. 9. Evaluation of the fault-detection ratio at evaluation points.

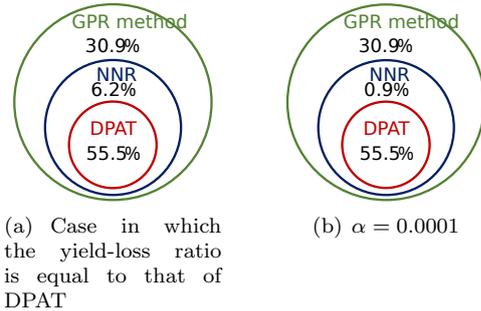


Fig. 10. Relationship between the number of the faulty chips detected by each method.

C.2. Classification of detected faults

Finally, detectable defective chips for each method were analyzed. Fig. 10 presents a Venn diagram illustrating the relationship between the chips identified as defective by each method. We present the ratio of the number of chips detected to the total number of failures. Additionally, the results of α are shown in Figs. 7 and 9. These results indicate an increase in fault-detection rate in the order of DPAT, NNR, and GPR methods. Furthermore, both the NNR and GPR methods can detect the same chip failures as the DPAT method, indicating that they have an inclusion relationship. The NNR method is included in the GPR method. When the yield-loss ratio of the DPAT method was equal to that of the GPR method, only 92.6% of all failed chips could be detected by applying the GPR method. This indicates that the GPR method can substitute the DPAT and NNR methods and potentially replace AEC-Q001, as a novel test standard for automotive semiconductors.

V. CONCLUSION

We applied the GPR-based adaptive test-threshold determination method proposed in [1] to industrial LSI test data. The conventional methods, such as the DPAT method, which automotive semiconductors need to comply with, determine the pass/fail threshold for each wafer

and cannot consider local process variation. In contrast, the GPR-based method utilizes GPR to estimate the systematic variation component on the wafer and evaluates the result between the measured and estimated values with confidence intervals, achieving accurate adaptive testing on a wafer.

The result of the measured value and the estimated value are evaluated with confidence intervals to determine whether the target LSI is pass or fail. When applied to industrial test data, the GPR method was confirmed to improve the fault detection ratio by 37.1% and the yield loss ratio by 0.019% compared with the DPAT method, a test method standardized in AEC-Q001. Furthermore, compared to the existing NNR method, the GPR method improved the fault detection ratio by 35.5%. By analyzing the LSIs detected by each method, the faulty LSIs were found to be detectable by the NNR and DPAT methods could be detected by the GP method, suggesting that this method can replace these methods.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grants, No. 22K11954 and 23H03362.

REFERENCES

- [1] K. Shimozato, M. Shintani, and T. Sato, "Adaptive outlier detection for power MOSFETs based on Gaussian process regression," in *Proc. APEC*, 2022, pp. 1709–1714.
- [2] L.-C. Wang, "Experience of data analytics in EDA and test—principles, promises, and challenges," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 36, no. 6, pp. 885–898, 2017.
- [3] Automotive Electronics Council, "AEC-Q001 Rev-D, Guidelines for part average testing," 2011, Online available: http://www.aecouncil.com/Documents/AEC-Q001_Rev_D.pdf.
- [4] W. Daasch, J. McNames, R. Madge, and K. Cota, "Neighborhood selection for IDDQ outlier screening at wafer sort," *IEEE Design & Test*, vol. 19, no. 5, pp. 74–81, 2002.
- [5] F. Lin and K.-T. Cheng, "An artificial neural network approach for screening test escapes," in *Proc. ASPDAC*, 2017, pp. 414–419.
- [6] M. Shintani *et al.*, "Artificial neural network based test escape screening using generative model," in *Proc. ITC*, 2018, p. 9.2.
- [7] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006, Online Available: <http://gaussianprocess.org/gpml/chapters/RW.pdf>.
- [8] D. Duvenaud, "The kernel cookbook," Online available: <https://www.cs.toronto.edu/~duvenaud/cookbook>.
- [9] K. Huang *et al.*, "Handling discontinuous effects in modeling spatial correlation of wafer-level analog/RF tests," in *Proc. DATE*, 2013, pp. 553–558.
- [10] M. Shintani *et al.*, "Wafer-level variation modeling for multi-site RF IC testing via hierarchical Gaussian process," in *Proc. ITC*, 2021, pp. 103–112.
- [11] S. Saxena *et al.*, "Variation in transistor performance and leakage in nanometer-scale technologies," *IEEE Trans. on Electron Devices*, vol. 55, pp. 131–144, 2008.
- [12] GPy, "GPy: A gaussian process framework in python," <http://github.com/SheffieldML/GPy>, since 2012.