PAPER

Efficient Wafer-Level Spatial Variation Modeling for Multi-Site RF IC Testing

Riaz-ul-haque MIAN[†], **Tomoki NAKAMURA**[††], **Masuo KAJIYAMA**[††], **Makoto EIKI**[††], ***Nonmembers***, ***and*** **Michihiro SHINTANI**[†††a)], ***Member***

SUMMARY Wafer-level performance prediction techniques have been increasingly gaining attention in production LSI testing due to their ability to reduce measurement costs without compromising test quality. Despite the availability of several efficient methods, the site-to-site variation commonly observed in multi-site testing for radio frequency circuits remains inadequately addressed. In this manuscript, we propose a wafer-level performance prediction approach for multi-site testing that takes into account the site-to-site variation. Our proposed method is built on the Gaussian process, a widely utilized wafer-level spatial correlation modeling technique, and enhances prediction accuracy by extending hierarchical modeling to leverage the test site information test engineers provide. Additionally, we propose a test-site sampling method that maximizes cost reduction while maintaining sufficient estimation accuracy. Our experimental results, which employ industrial production test data, demonstrate that our proposed method can decrease the estimation error to 1/19 of that a conventional method achieves. Furthermore, our sampling method can reduce the required measurements by 97% while ensuring satisfactory estimation accuracy.

key words: *wafer-level spatial characteristic modeling, Gaussian process regression, LSI test*

1. Introduction

Currently, large-scale integrated circuits (LSIs) are being embedded in all modern products, thereby ensuring smooth functioning of daily life. In addition to automobiles, healthcare, and aerospace, which are industries essential for daily life, LSIs are utilized in critical infrastructure that support our daily lives, such as computer networks, power transmission systems, and transportation control systems. However, with the spread of LSIs, their reliability has emerged as a crucial issue, and faulty LSIs not only interrupt the services of systems reliant on these circuits but also severely impact our society.

To ensure the reliability of LSI products, multiple test items are tested and/or measured under various conditions during several stages of LSI manufacturing. With increasing scale and multi-functionality of LSIs, an increasing number of items must be tested, resulting in test-cost inflation. However, as the test cost accounts for a majority of the cost incurred in LSI manufacturing, this inflation has emerged as a major challenge.

Manuscript received September 19, 2023.
Manuscript publicized November 16, 2023.
[†]Shimane University, Matsue-shi, 690-8504 Japan.
[††]Sony Semiconductor Manufacturing, Isahaya-shi, 854-0065 Japan.
[†††]Kyoto Institute of Technology, Kyoto-shi, 606-8585 Japan.
a) E-mail: shintani@kit.ac.jp (Corresponding author)
DOI: 10.1587/transfun.2023EAP1115

Various test-cost reduction methods have been proposed that apply data analytics, machine-learning algorithms, and statistical methods [1]–[3]. In particular, the wafer-level characteristic modeling method based on a statistical algorithm is one of the most promising candidates that reduces the test cost, that is, the measurement cost, without impairing the test quality [4]–[11]. In these studies, a statistical modeling technique was used to predict the entire measurement on a wafer from a small number of sample measurements. As the estimation eliminates the need for measurement, it not only reduces the cost of measurement but can be used to reduce the number of test items and/or change the test limits, which is expected to improve the efficiency of adaptive testing [12]–[14]. In [4], the expectation-maximization (EM) algorithm [15] was leveraged to predict the measurement. Moreover, in [5]–[8], a statistical prediction method, named *virtual probe*, based on compressed sensing [16] was proposed. Notably, the *Gaussian process* (GP)-based method [17] yields more accurate prediction results [9]–[11]. Furthermore, the use of GP modeling entails another benefit. This method calculates the confidence of a prediction; accordingly, the user can confirm whether the number and location of the measurement samples are sufficient, which is a significant advantage from a practical viewpoint.

Most of these methods assume that the device characteristics on the wafer gradually change with wafer coordinates; however, this assumption does not hold for the measurement of radio frequency (RF) circuits under multi-site testing [18]–[20], in which a probe card is adapted to simultaneously probe multiple devices under test (DUTs). The contact of the probe card with the DUTs to be tested is named the *touchdown* and, the position of the needles in a touchdown is called a *site*. During the measurement of the RF circuit, a calibration circuit for impedance matching is added to the probe card, causing a significantly greater variation than the spatial variation owing to its parasitic components, as illustrated in Fig. 1.

Figure 1(a) presents the histograms of the characteristics of an industrial RF circuit, which is fabricated using a 28-nm process technology, measured through a multi-site test with 16 sites per measurement in the first fabrication lot. The histograms are presented in different colors corresponding to each site. Although each histogram exhibits a low variance, significant differences, that is, differences in sites among the histograms are evident. Most existing methods fail to model this measurement result because of

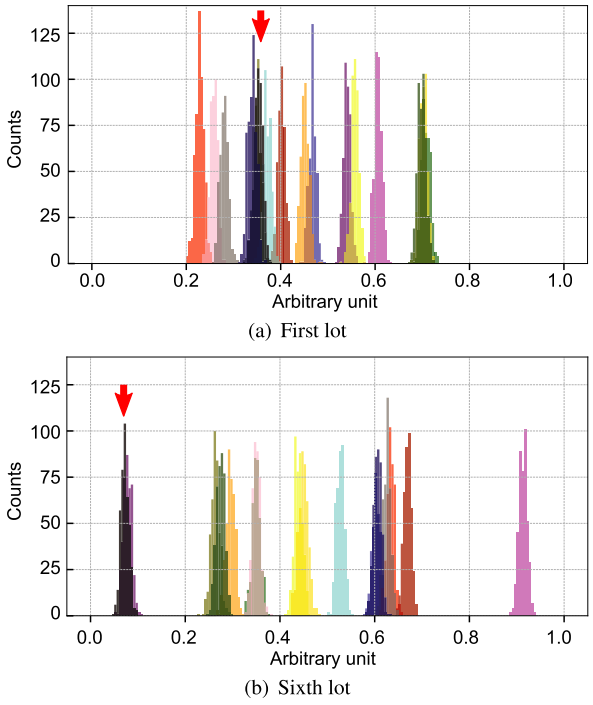


Fig. 1 Histograms of measured characteristics of an industrial RF circuit on a wafer, measured by multi-site testing with 16 sites. The histograms of each site are shown with different colors, that is, 16 histograms are presented herein. The significant variations in the histogram between sites are clearly visible. Furthermore, the locations of the black histograms in the early and latest lots are considerably different. The horizontal axis is expressed in an arbitrary unit.

discontinuous changes between sites.

Only the work in [11] attempted to solve this issue of the discontinuous change in wafer-level variation modeling. In [11], a two-step modeling method using k-means clustering [21] and GP was proposed. In the first step, all dies on a wafer are explicitly measured, and subsequently, the k-means clustering algorithm is applied to divide the measurements into k measurement groups. In addition, the wafer coordinates are also clustered according to the k measurement groups. In the second step, for the subsequently fabricated wafers, GP is applied to each cluster individually. Because the spatial variation is modeled according to the partitioned magnitude of the measured value, discontinuous changes can be accurately reproduced. In addition, a method has been reported [22] that addresses variations between sites. Reportedly [22], to set outlier limits in a test, there is a method of eliminating the variation between sites by normalizing each site [22]. However, setting the normalization constant appropriately in small samples is challenging; thus, applying site normalization to wafer-level modeling is a difficult task.

However, this method of [11] relies heavily on the assumption that the k-means clustering results obtained from the first wafer are applicable to all the subsequent lots. Notably, certain site histograms drastically changed in the latest fabrication lot, as displayed in Fig. 1(b), which presents the histogram in the sixth lot. For example, to achieve an accurate prediction, the highlighted black histogram should correspond to a cluster different from that shown in Fig. 1(a). Although the possibility of recalibrating k-means clustering has been briefly described, no specific solution has been reported as yet [11].

Herein, we propose a novel wafer-level spatial-variation modeling method for RF circuits under multisite testing. Generally, test engineers possess site information for probing; thus, we exploited this aspect as a cluster in the proposed method to predict spatial variation through hierarchical GP modeling of each site. Therefore, the proposed method does not require a clustering algorithm and measurement corresponding to the first step. In particular, the use of site information is straightforward yet efficient under multisite testing. Because the characteristics measured within one cluster possess additional parasitic components identical to those of the calibration circuit, only the spatial changes on the wafer are modeled. Consequently, the proposed method allows for accurate modeling across wafers. Moreover, we propose an active sampling method based on *active learning* [23], while considering the measurement of multisite testing. Through the active sampling method utilizing the predictive variance of each site, the proposed method achieved optimal estimation with a small number of measured samples.

This manuscript is based on our previous work [24]. While our previous evaluation insufficiently uses only 6 wafers, a more practical evaluation is provided to show the effectiveness of the proposed method in a real production test environment by increasing the number of wafers to all wafers of six lots, i.e., totally 143 wafers. Furthermore, we evaluate the effectiveness of the proposed sampling method by comparing the sampling method proposed in [10].

The main contributions of this work are summarized as follows:

- **Hierarchical GP modeling using site information:** Our proposed method enables precise modeling of the wafer's spatial correlation, even when measuring RF circuits with discontinuous changes for any lot. This is accomplished by applying the GP separately to the appropriate clusters obtained through the use of site information.
- **Active sampling algorithm under multi-site testing environment:** We propose an efficient sampling algorithm based on the predictive variance of the estimation to determine the sample location.
- **Comparison with the conventional method using industrial production data:** Our experimental results confirm that the assumption made in the two-step modeling method [11], which applies that the k-means clustering result can be used for subsequent wafers, is not valid for a more miniaturized fabrication process. Furthermore, we demonstrate that our proposed method is capable of reducing the prediction error to an average of $1/19.4$ compared to the prediction error obtained through the two-step modeling method.
- **Thorough evaluation of the proposed active loca-**

tion selection algorithm: The experimental results reveal that the proposed sampling method successfully reduces the number of touchdowns compared to the random sampling method and aggressive sampling method [10] without sacrificing the prediction accuracy. To the best of our knowledge, our study is a pioneering effort that successfully demonstrates spatial variation modeling in a multi-site testing environment.

The remainder of this paper is organized as follows. Section 2 briefly illustrates the GP, which plays a central role in the proposed method. In addition, we reviewed the existing wafer-level spatial variation modeling based on the two-step approach [11], as a previous work. Section 3 proposes a hierarchical GP based on site information and an active sampling method for multisite testing. The experimental results using industrial production test-data of an RF IC fabricated via a 28 nm process technology are presented in Sect. 4, and the effectiveness of the proposed method is quantitatively evaluated against that of conventional methods, as presented in this section. Finally, the conclusions drawn based on the findings are presented in Sect. 5.

2. Preliminaries

2.1 Gaussian Process

First, we quickly review a GP [17], which is an integral part of the conventional method [11] and our method. The GP model is employed to estimate the function $y = f(x)$ based on the input variable, x, to the output variable y, which is generally used for regression. In the GP model, the function f is assumed to follow a multidimensional normal distribution and is expressed as $f \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Z})$ using a kernel matrix $\boldsymbol{Z}$. The primary benefit of this model is its ability to address nonlinear estimation problems. Another important advantage is the utilization of Bayesian inference [25]. Because the estimated function is obtained as a distribution of functions, not as a single function, the uncertainty of the estimation can be expressed as predictive variance.

The outline of GP-based multiple regression is summarized in Algorithm 1. We consider $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}}) = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_N, y_N)\}$ and $\boldsymbol{X}_{\text{test}} = (\boldsymbol{x}^*{}_1, \boldsymbol{x}^*{}_2, \cdots, \boldsymbol{x}^*{}_M)$ as the training and test datasets, respectively, where $M \gg N$. In addition, a kernel function, f_{kern}, is given as an input. Using the predicted model, f, calculated based on $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$, the algorithm returns the mean values and variances of the predicted $\boldsymbol{y}^* = (y_1^*, y_2^*, \cdots, y_M^*)$ for $\boldsymbol{X}_{\text{test}}$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_M)$ and $\boldsymbol{v} = (v_1, v_2, \cdots, v_M)$.

Algorithm 1 Gaussian process regression

Input: Training dataset: $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$, Test dataset: $\boldsymbol{X}_{\text{test}}$, Kernel function: f_{kern}
Output: Mean and variance of predicted values: $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_M)$ and $\boldsymbol{v} = (v_1, v_2, \cdots, v_M)$
1: **for** $n = 1$ to N **do**
2: **for** $n' = 1$ to N **do**
3: Calculate (n, n')-th element of a kernel matrix $\boldsymbol{Z}$ as $f_{\text{kern}}(\boldsymbol{x}_n, \boldsymbol{x}'_n)$
4: Calculate fitting parameters of f_{kern} to fit $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$
5: **for** $m = 1$ to M **do**
6: **for** $n = 1$ to N **do**
7: Calculate n-th element of $\boldsymbol{z}_*$ as $f_{\text{kern}}(\boldsymbol{x}_n, \boldsymbol{x}^*{}_m)$
8: $\boldsymbol{z}_{**} = f_{\text{kern}}(\boldsymbol{x}^*{}_m, \boldsymbol{x}^*{}_m)$
9: Append $\mu_m = \boldsymbol{z}_*^T \boldsymbol{Z}^{-1} \boldsymbol{y}_{\text{train}}$ to $\boldsymbol{\mu}$
10: Append $v_m = \boldsymbol{z}_{**} - \boldsymbol{Z}^{-1} \boldsymbol{z}_*$ to $\boldsymbol{v}$

In lines 1 to 5, the kernel matrix $\boldsymbol{Z}$ of the training dataset is calculated for each element of $\boldsymbol{X}_{\text{train}}$ using the kernel function. Subsequently, in lines 7 to 14, the probability density function of the predicted y_m^* corresponding to $\boldsymbol{x}^*{}_m$ is derived by modeling a multidimensional normal distribution, as follows:

$$\begin{aligned} & p(y_m^* | \boldsymbol{x}^*{}_m, \boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}}) \\ &= \mathcal{N}(\boldsymbol{z}_*^{\mathrm{T}} \boldsymbol{Z}^{-1} \boldsymbol{y}_{\text{train}}, \boldsymbol{z}_{**} - \boldsymbol{z}_*^{\mathrm{T}} \boldsymbol{Z}^{-1} \boldsymbol{z}_*), \end{aligned} \tag{1}$$

where $\boldsymbol{z}_*$ and $\boldsymbol{z}_{**}$ denote the covariances between the training and test datasets and between the test datasets, respectively. As demonstrated in Eq. (1), the mean value and variance of y_m^* can be derived analytically. The expected values are utilized in the prediction; nevertheless, the variances can also be used to confirm the uncertainty of the prediction.

Several kernel functions are available, such as linear, squared exponential, and Matérn kernels. For example, the radial basis function (RBF) kernel is expressed as follows [26]:

$$f_{\text{kern}}(\boldsymbol{x}, \boldsymbol{x}') = \theta_1 \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}')^2}{\theta_2}\right), \tag{2}$$

where θ_1 and θ_2 represent the fitting parameters calculated using an iterative optimization routine, as expressed in line 6. As $\boldsymbol{Z}$ is a variance-covariance matrix, when $\boldsymbol{x}$ and $\boldsymbol{x}'$ are close, $f_{\text{kern}}(\boldsymbol{x}, \boldsymbol{x}')$ becomes large, and consequently, $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ are also close.

The predictive mean, $\boldsymbol{\mu}$, is leveraged in wafer-level characteristics modeling. Expectedly, GP regression can be incorporated into the wafer-level spatial variation modeling in IC characteristics with high affinity, as the characteristics of adjacent dies on the wafer are similar because of the systematic components of process variation [27], [28].

2.2 Related Work

Owing to intensive research on wafer-level spatial variation correlation modeling, the prediction accuracy of the spatial measurement variation has been improved, thereby, enabling the successful reduction of measurement costs in production tests [4]–[11]. Among others, in [11], a two-step modeling approach has been proposed to handle the discontinuous effect induced via multi-site testing and reticle shots in wafer-level modeling.

The objective of the first step is to partition the wafer into k groups, which reflect the k levels of wafer measurements induced by discontinuous effects. For this purpose, a k-means algorithm is exploited as follows:

$$\boldsymbol{y} = \{\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \cdots, \boldsymbol{y}^{(k)}\}, \tag{3}$$

where $\boldsymbol{y}$ represents the vector of the measured characteristics of all the dies on the wafer. Consequently, $\boldsymbol{X}$ corresponding to $\boldsymbol{y}$ is partitioned as follows:

$$\boldsymbol{X} = \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \cdots, \boldsymbol{X}^{(k)}\}. \tag{4}$$

Note that Eq. (4) indicates that the coordinates on the wafer are divided according to the measured characteristics. Once the k clusters are identified, in the second step for subsequent wafers, the GP is applied to each cluster individually based on Algorithm 1. As the changes in each $\boldsymbol{y}^{(k)}$ can be expected to be smooth, the GP regression will function successfully; thus, the two-step approach can handle discontinuous changes.

The determination of optimal k is not trivial. Although several conventional methods, such as silhouette value [29] and the elbow method [30], determine the optimal k value, in this modeling approach [11], k is determined based on the following equation:

$$k = \arg\max_{g} CH(g), \tag{5}$$

where $CH(g)$ indicates the Calinski and Harabasz index when g clusters are considered [31].

However, the two-step modeling is not always applicable because this approach keeps using k clusters for subsequent wafers, assuming that the content of the clusters will not change for other wafers/lots. Because the experiment in [11] used the industrial data fabricated using a relatively mature process technology, the assumption might hold true; in contrast, for our production data on immature process technology, the process is inapplicable, as shown in Fig. 1.

Another observation of Fig. 1 is also represented in Fig. 2. In this figure, the measured characteristics for each site are depicted as functions of the lot ID from the first to sixth lot. The first wafer is used for each lot, and the lines and shaded regions represent the mean and three standard deviations, respectively, for each site. Observably, the distributions within the site are comparatively maintained up to the first two lots; however, they fluctuate significantly from the third lot onwards. This result suggests that the two-step modeling may function optimally up to the first two lots; however, the clusters must be recalibrated for the third through sixth lots, thereby adding to the measurement costs. Additionally, early stage lots generally exhibit low production yields; consequently, applying the two-step modeling method is a difficult task.

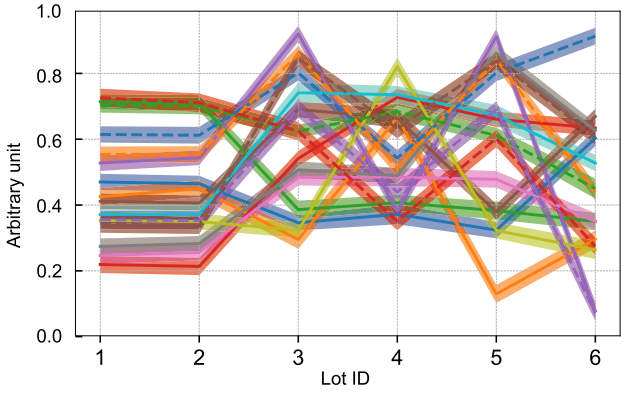


Fig. 2 Measured characteristics of 16 sites from the first lot to the sixth lot. The solid lines and shaded regions represent the means and the three standard deviations of the variations, respectively. The vertical axis denotes arbitrary units.

3. Wafer-Level Variation Modeling for RF IC under Multi-Site Testing

We propose a novel spatial variation model based on the site information provided by test engineers; this model yields the correct cluster without applying clustering algorithms. In particular, GP-based prediction is hierarchically performed for each site cluster. Site-to-site variations during multi-site testing are attributed to the parasitic components of the calibration circuit. Ideally, these variations should be eliminated during the measurement, which is an impractical approach because of the design and manufacturing costs of the probe card. Although this issue can be solved by testing them one at a time, the benefits of multi-site testing will not be realized. Semiconductor manufacturing engineers have knowledge of the systematic discontinuous fluctuations in a manufacturing environment owing to the manufacturing recipes and measurement items [32]. In the proposed method, we fully employ it to improve the modeling accuracy. The proposed method applies hierarchical GP modeling through clustering using site information, thereby yielding a highly accurate modeling performance while considering the actual measurement environment. As observed in Fig. 2, the measurements at the same site exhibit a minor deviation. Thus, site-based hierarchical clustering is expected to be an optimal model without recalibration.

In addition, to achieve a small sampling ratio, we propose an active sampling algorithm based on the variance computed via GP-based regression in a multi-site testing environment. In contrast to all the existing studies that assume sequential sampling, the proposed algorithm can effectively reduce the measurement cost.

3.1 Modeling Based on Site-Based Hierarchical GP

Algorithm 2 presents the proposed spatial correlation modeling through a site-based hierarchical GP in detail. We as-

Algorithm 2 Site-based hierarchical spatial variation modeling

Input: $\boldsymbol{\mu}$ and $\boldsymbol{v}$ Training dataset: $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$ measured under multi-site testing, Test dataset: $\boldsymbol{X}_{\text{test}}$, Kernel function: f_{kern}, site information

Output: Mean and variance of predicted values: $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_M)$ and $\boldsymbol{v} = (v_1, v_2, \cdots, v_M)$

1: Cluster $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$ and $\boldsymbol{X}_{\text{test}}$ into S groups according to the site information

2: **for** $s = 1$ to S **do**

3: $\quad \boldsymbol{\mu}^{(s)}, \boldsymbol{v}^{(s)} = \text{gpr}\left((\boldsymbol{X}^{(s)}_{\text{train}}, \boldsymbol{y}^{(s)}_{\text{train}}), \boldsymbol{X}^{(s)}_{\text{test}}, f_{\text{kern}}\right)$

4: Concatenate all $\boldsymbol{\mu}^{(s)}$ and $\boldsymbol{v}^{(s)}$ into $\boldsymbol{\mu}$ and $\boldsymbol{v}$

sume that the measurements are conducted using multi-site testing. Compared to the conventional method, clustering is performed according to the site information in a single touchdown, as listed in line 1. Essentially, in the conventional method, the characteristics of the initial wafer for the first production lot need to be measured entirely, whereas the proposed method eliminates the need for such measurements during clustering. In Algorithm 2, S represents the number of the sites in a single touchdown, and the training and test datasets are grouped into S groups, as follows:

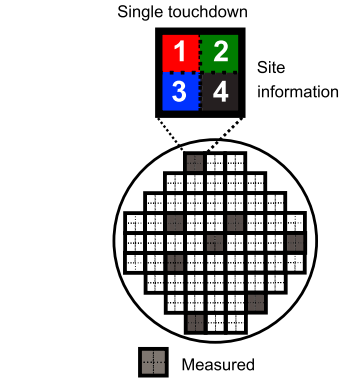
$$(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}}) = \{(\boldsymbol{X}_{\text{train}}^{(1)}, \boldsymbol{y}_{\text{train}}^{(1)}), (\boldsymbol{X}_{\text{train}}^{(2)}, \boldsymbol{y}_{\text{train}}^{(2)}), \cdots, (\boldsymbol{X}_{\text{train}}^{(S)}, \boldsymbol{y}_{\text{train}}^{(S)})\} \tag{6}$$

and

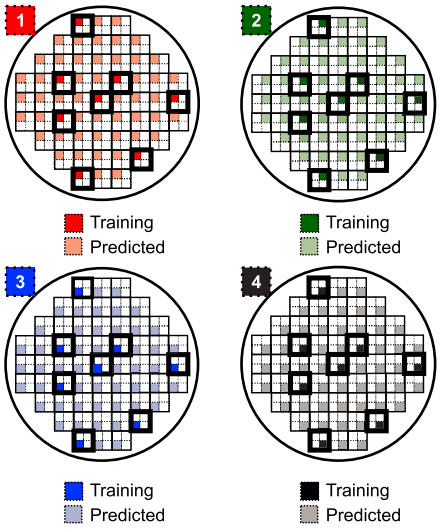
$$\boldsymbol{X}_{\text{test}} = \{\boldsymbol{X}_{\text{test}}^{(1)}, \boldsymbol{X}_{\text{test}}^{(2)}, \cdots, \boldsymbol{X}_{\text{test}}^{(S)}\}, \tag{7}$$

respectively. The GP-based regression is performed individually by hierarchically modeling each site, as listed in lines 2 to 4, based on the *gpr* function listed in Algorithm 1, through which the mean and variance of the prediction for the test dataset are returned. Finally, the prediction result for the entire wafer is obtained by concatenating each prediction result.

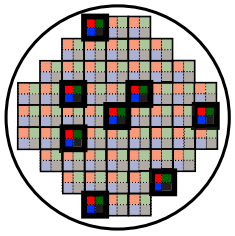
An example of the proposed modeling method with $S = 4$ (sites 1 to 4) is depicted in Fig. 3. Initially, eight positions are chosen and measured as the training data, as shown in Fig. 3(a), resulting in the measurement of 32 dies ($= 8 \times 4$) using eight touchdowns. Next, GP-based modeling and prediction are individually applied according to the site, as shown in Fig. 3(b). For instance, for site 1, the measured value belonging to site 1 is used as the training data to construct a GP model, and the measured value of the unmeasured die belonging to site 1 is predicted. The measurements and predictions for the other sites follow a similar procedure. The complete prediction result is obtained through concatenation, as shown in Fig. 3(c).



(a) Eight randomly sampled touchdown locations in multi-site testing



(b) Modeling and prediction using hierarchical GP regression



(c) All-site concatenation

Fig. 3 Example of the site-based hierarchical GP regression wherein a single touchdown features four sites.

3.2 Active Sampling under the Multi-Site Testing

In wafer-level spatial modeling, inputting a small training dataset is advantageous for minimizing the measurement cost. In [10], an aggressive sampling method was proposed: this method preferentially measures the location with the largest predictive variance calculated via GP regression. In addition, in [7], a Latin hypercube sampling approach [33] was employed to select random sample points evenly over the entire wafer. However, these methods are simplistic in their approach, and most importantly, they do not account for a multi-site testing environment.

An optimal model should exhibit minimal error between the model and the actual measurement. The mean squared error (MSE) against the test dataset can be expressed as:

$$E_{\text{MSE}} = ||\boldsymbol{v}|| + ||\boldsymbol{\mu} - \boldsymbol{y}_{\text{true}}||^2, \tag{8}$$

where $||\cdot||$ denotes the Euclidean norm, and $\boldsymbol{y}_{\text{true}}$ indicates the correct value at the location of $\boldsymbol{X}_{\text{test}}$ and unknown, that is, the unmeasured value. Assuming that the model is correct, the contribution of the second term in Eq. (8) to E_{MSE} is small compared with the variance contribution, that is, the first term. Thus, to minimize E_{MSE}, $\boldsymbol{X}$ must be selected such that the overall variance of the estimator is minimized [23].

Based on the aforementioned premise, we propose an active sampling method, as outlined in Algorithm 3, which is incorporated into the site-based hierarchical spatial modeling

Algorithm 3 Active location selection with site-based hierarchical Gaussian process regression

1: $\boldsymbol{\mu}, \boldsymbol{v} = \text{hgpr}((\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}}), \boldsymbol{X}_{\text{test}}, f_{\text{kern}})$
2: **for** $p = 1$ to P **do**
3: $\quad \boldsymbol{\mu}_p, \boldsymbol{v}_p = \text{hgpr}(\boldsymbol{X}_{\text{train}} + \boldsymbol{X}_{\text{add}}^{(p)}, \boldsymbol{X}_{\text{test}}, f_{\text{kern}})$
4: $\quad$ Calculate the Euclidean distance between $\boldsymbol{v}$ and $\boldsymbol{v}_p$ as $\Delta_{\text{var}}^{(p)}$
5: Select $\boldsymbol{X}_p$ with the largest $\Delta_{\text{var}}^{(p)}$ as the next touchdown location

(*hgpr*) presented in Algorithm 2. The proposed sampling method focuses on the Euclidean distance between the prior and post measurements. The proposed method proceeds as follows: The numbers on the left indicate the line numbers corresponding to Algorithm 3.

1) Calculate $\boldsymbol{\mu}$ and $\boldsymbol{v}$ through the hierarchical GP regression using $(\boldsymbol{X}_{\text{train}}, \boldsymbol{y}_{\text{train}})$ and $\boldsymbol{X}_{\text{test}}$ as shown in Algorithm 2. $\boldsymbol{X}_{\text{train}}$ can be obtained via multi-site testing.
2) Repeat steps 3) and 4) for all touchdown location candidates. At this step, the p-th touchdown candidate has $\boldsymbol{X}_{\text{test}}^{(p)} = \{\boldsymbol{x}_1^{*(p)}, \boldsymbol{x}_2^{*(p)}, \cdots, \boldsymbol{x}_S^{*(p)}\}$ with the S sites.
3) Add the touchdown candidate $\boldsymbol{X}_{\text{add}}^{(p)}$ by assuming it to be measured and performing a hierarchical GP regression. Note that as this parameter is not actually measured, we assume that the mean values are measured as $\boldsymbol{X}_{\text{add}}^{(p)} = \{(\boldsymbol{x}_1^{*(p)}, \mu_1^{(p)}), (\boldsymbol{x}_2^{*(p)}, \mu_2^{(p)}), \cdots, (\boldsymbol{x}_S^{*(p)}, \mu_S^{(p)})\}$, where $\mu_S^{(p)}$ denotes the predicted mean corresponding to $\boldsymbol{x}_S^{*(p)}$, and is one of the elements of $\boldsymbol{\mu}$ yielded by hgpr in step 1). In this step, $\boldsymbol{\mu}_p$ and $\boldsymbol{v}_p$ are obtained as in step 1).

4,5) Calculate the Euclidean distance of $\boldsymbol{v}$ and $\boldsymbol{v}_p$ as $\Delta_{\text{var}}^{(p)}$. Note that steps 2) to 5) are iterated for all the touchdown candidates.

6) Select $\boldsymbol{X}_p$ with the largest $\Delta_{\text{var}}^{(p)}$ as the next measurement location.

The mentioned procedure is iterated until an exit condition is satisfied; for example, a sufficient number of iterations are obtained. As the reduction of the whole deviation for the test dataset is compared in step 6), a more accurate modeling can be expected with a smaller number of measurements compared to simply checking the location of the highest variance, as reported in [23].

4. Numerical Experiments

4.1 Setup

To demonstrate the effectiveness of the proposed method, we conducted experiments using an industrial production test dataset of a 28 nm analog/RF device. Our dataset contains 143 wafers from six lots. A single wafer features approximately 6,000 DUTs. In this experiment, we utilized a measured character for an item of the dynamic current test, in which site-to-site variability was noticeable owing to the multi-site test is noticeably observed, as demonstrated in Figs. 1 and 2. A heat map of the full measurement results for the first wafer of the sixth lot is displayed in Fig. 4. For the ease of experimentation, the faulty dies were removed from the dataset, and the number of sites in a single touchdown was 16, that is, $S = 16$. The form of a single touchdown is presented in Fig. 5. This is different from the rectangular touchdown illustrated in Fig. 3, which prevents interference on the probe of the impedance-matching circuits. Consequently, a special pattern emerges during the multi-site test, as depicted in Fig. 4. To fully measure all DUTs on a single wafer, approximately 600 touchdowns were needed.

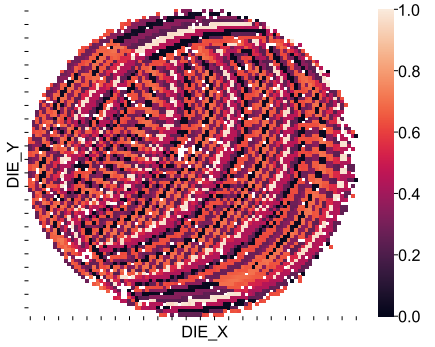


Fig. 4 Heat map of fully measured characterization. The measured values are normalized.

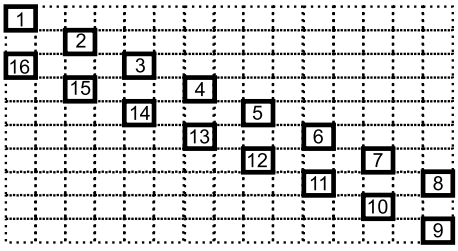


Fig. 5 Single touchdown with 16 sites in our multi-site testing.

All experiments were implemented in the Python language using the packages, GP [34] and scikit-learn [35], for the GP and k-means clustering, respectively. The RBF kernel was used as the kernel function, f_{kern}, for the GP-based regression. The experiments were conducted on a Linux PC with an Intel Xeon Platinum 8160 2.10 GHz central processing unit using a single thread.

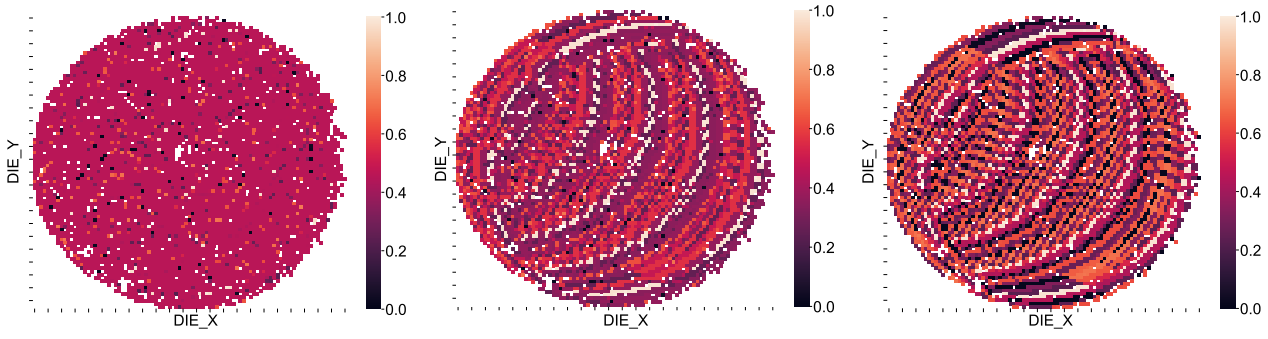
To quantitatively evaluate the modeling accuracy, we defined the error ($\boldsymbol{\delta}$) between the correct ($\boldsymbol{y}_{\text{true}}$) and the predicted mean ($\boldsymbol{\mu}$) normalized using the maximum and minimum values of $\boldsymbol{y}_{\text{true}}$ as follows:

$$\boldsymbol{\delta} = \frac{\boldsymbol{\mu} - \boldsymbol{y}_{\text{true}}}{d_{\text{spec}}}, \tag{9}$$

where d_{spec} indicates the range between the minimum and maximum values of the fully measured characteristics illustrated in Fig. 4.

4.2 Experimental Results on Site-Based Hierarchical Spatial Modeling

First, we evaluated the site-based hierarchical spatial modeling presented in Algorithm 2. For comparison, a naive GP



(a) Naive GP (b) 2-step GP (c) Ours

Fig. 6 Heat maps of the predicted characteristics obtained using the naive GP, 2-step GP, and the proposed method at a spatial sampling rate of 10%. Observably, the prediction results are closer to the actual measurements in the order of naive GP, 2-step GP, and the proposed method. The measured values are normalized.

regression-based approach (hereafter called *naive GP*) [9] and a two-step approach (hereafter called *2-step GP*) [11] were also applied. Notably, we did not consider touchdown, that is, a one-by-one measurement was conducted. The experimental results based on touchdown are presented in Sect. 4.3.

For the 2-step GP method, the first wafer of the first lot was used to obtain k clusters through k-means clustering. For the subsequent wafers, k clusters were used to predict the device characteristics. In the experiment, the optimal k was determined using the silhouette value and elbow method [29], [30], instead of Eq. (5), resulting in the seven clusters.

In Fig. 6, the prediction results for the first wafer of the sixth lot using each method are presented. These results were predicted using randomly sampled values at a spatial sampling rate of 10%. Clearly, the naive GP method failed to capture the site-to-site variation, as presented in Fig. 6(a). In contrast, the specific pattern depicted in Fig. 4 caused by the site-to-site variation can be visually confirmed in the 2-step GP method and our method, as illustrated in Figs. 6(b) and 6(c). Figure 7 displays the violin-plots of $\boldsymbol{\delta}$ using Eq. (9) for each method, as follows: In the figure, the top and bottom of the lines represent the maximum and minimum values, respectively. The average is indicated by the dot. The distribution of $\boldsymbol{\delta}$ generated via a kernel density estimation is also presented herein. The average errors of $\boldsymbol{\delta}$ corresponding to the naive GP, 2-step GP method, and our method are 18.59%, 13.43%, and 0.69%, respectively. Moreover, the proposed method can also drastically minimize the variance in the predictions. These results conclusively demonstrate the proposed method reduced the average error by approximately 5.13% compared with the 2-step GP method, that is, 19.46 times (= 13.43/0.69) more accurately.

Figure 8 plots the averages of $\boldsymbol{\delta}$ as a function of the spatial sampling rate using the three methods for the first wafer of the sixth lot. Notably, the prediction methods are not applied when the spatial sampling rate is 100% and the sampling rate is incrementally increased; that is, the measured locations at the 10% sampling rate are invariably contained at

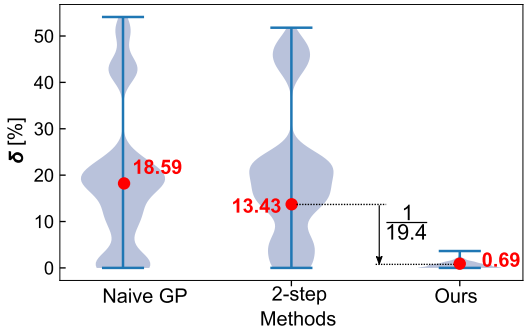


Fig. 7 Violin-plots of $\boldsymbol{\delta}$ for each method, where the distributions of $\boldsymbol{\delta}$ are shown.

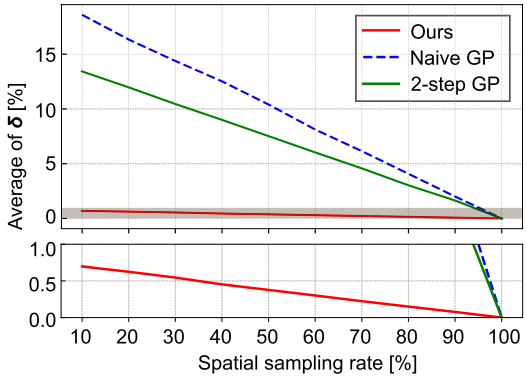


Fig. 8 Averages of $\boldsymbol{\delta}$ obtained using the naive GP, 2-step GP, and the proposed method at various sampling rates. An enlarged view of the gray part is presented under the main figure.

subsequent rates. As the spatial sampling rate increases, the averages of all the methods decrease monotonically. Moreover, we find that the average errors of the proposed method always achieve better prediction results for all the sampling rates.

The averages of $\boldsymbol{\delta}$ for the wafers at the 10% sampling rate as a function of the wafer ID from the first lot to the sixth lot, comprising a total of 143 wafers, are illustrated in Fig. 9. Observably, the proposed method achieves the best

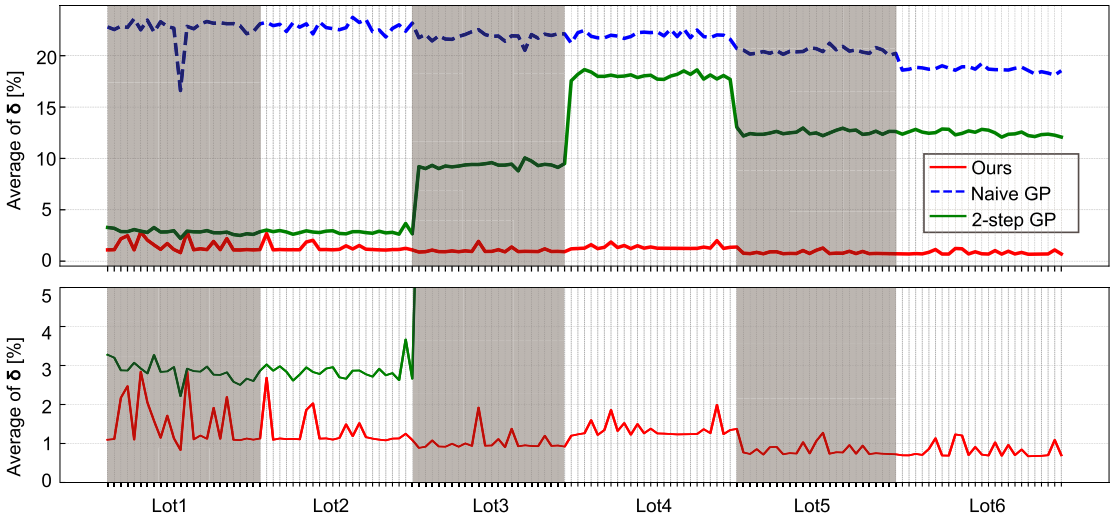


Fig. 9 Changes of the averages of δ for the 143 wafers of the six lots. An enlarged view of the area below 5% is presented under the main figure.

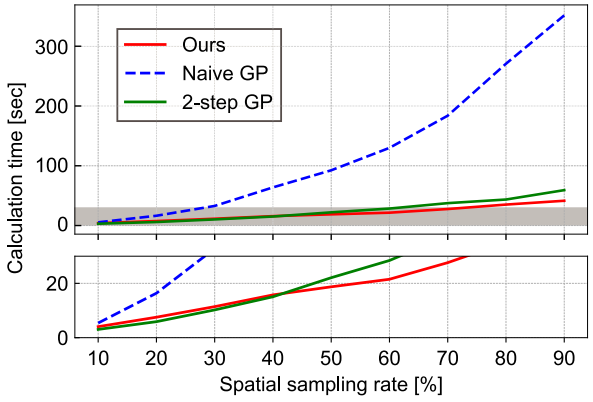


Fig. 10 Calculation time. The gray part is enlarged at the bottom of this figure.

estimation results for all the lots among the three methods. The prediction performance of the 2-step GP degrades as the production lot progresses, whereas the proposed method maintains a low prediction error below 3% regardless of the lot and wafer. This implies that the k-means clustering result obtained in the first lot is inappropriate for subsequent lots.

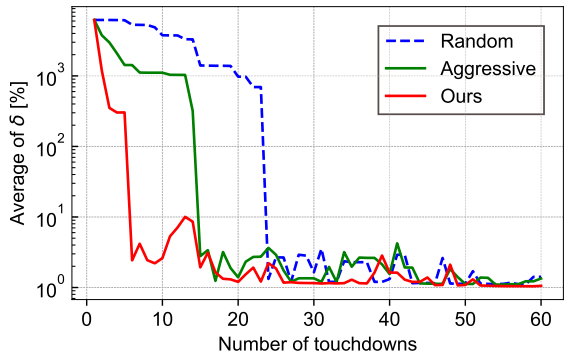
The calculation time of the prediction for each method was evaluated. Figure 10 summarizes the calculation time for each method for the first wafer of the sixth lot at various sampling rates. We can see that the proposed method and 2-step method can significantly reduce the calculation time compared to the naive GP method. This improvement is attributed to the additional benefits of hierarchical GP modeling approaches. In general, the inference time of GP is $O(N^3)$ scaled because of the computation of the matrix inverse [36], [37]. In the proposed method, GP modeling is conducted for each site individually; thus, the calculation time can be drastically reduced because the training samples are reduced to N/S in each GP modeling, where S is 16 for the proposed method in this experiment. The reduction becomes N/k for the 2-step method, where $k = 7$. Note that this calculation was conducted using a single thread. Therefore, the calculation time can be reduced further through implementing parallel processing.

4.3 Experimental Result under Multi-Site Testing

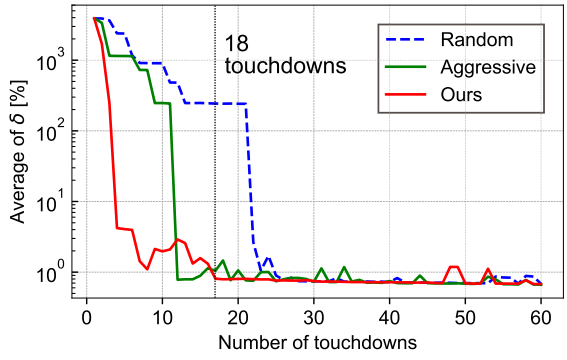
In the evaluation presented in the previous section, the sample dies were randomly selected one by one, and thus, the multi-site test environment in which measurements were conducted per site unit was not considered. We evaluated the sampling method listed in Algorithm 3 in a multi-site test environment. Herein, it is assumed that the touchdown shown in Fig. 5 is performed in a single measurement with 60 touchdowns, which corresponds to approximately 10% of the touchdowns of the full measurement. In all the existing researches on the wafer-level variation modeling, sampling is assumed to be one DUT at a time; and thus, this work is the first to consider a multi-site testing environment for wafer-level variation modeling. In this experiment, the random sampling and the aggressive sampling proposed in [10] were used for comparison. First, we measured one randomly sampled touchdown and subsequently selected the next touchdown using each method.

Figure 11 plots the average δ as a function of the number of touchdowns that incrementally increased. The first wafers in the first and sixth lots are used as examples. Although not displayed herein, similar results were obtained for almost all other wafers. Although errors of over several thousand percent are observed for all the methods at the first touchdown, the error decreases for all the methods. However, the random and aggressive sampling methods converge slowly, whereas the proposed method converges rapidly.

The average δ of the 143 wafers for each touchdown is shown in Fig. 12 to evaluate the prediction errors of each method for all the 143 wafers. As shown in Fig. 11, the random sampling converges slowly, whereas the aggressive and



(a) Lot-1, Wafer-1



(b) Lot-6, Wafer-1

Fig. 11 Averages of $\boldsymbol{\delta}$ as a function of the number of the touchdowns. The vertical axis is depicted in log scale. The proposed method converges more quickly.

active sampling methods converge quickly. However, while the random and aggressive sampling methods converge to 4.46% and 4.38%, respectively, the proposed method converges to 0.96%. In addition, the error at which the random and aggressive sampling methods converge is achieved in 18 touchdowns for the proposed method. We emphasize that the 18 touchdowns correspond to approximately 3% of the number of the touchdowns for the full measurement. Moreover, the proposed method reduces the average of $\boldsymbol{\delta}$ by 3.42% when the 60 touchdowns were conducted. The results plotted in the figure clearly indicate that the proposed sampling method can successfully reduce the number of necessary touchdowns while achieving a better prediction accuracy for

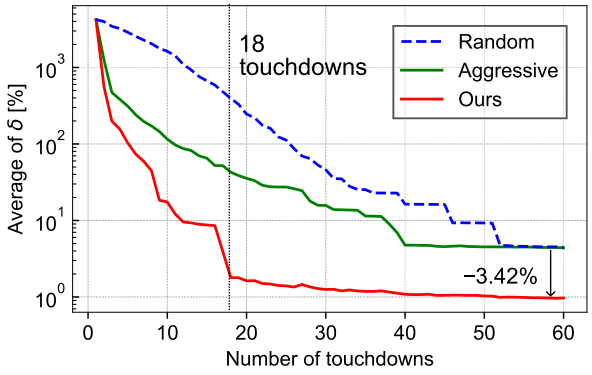
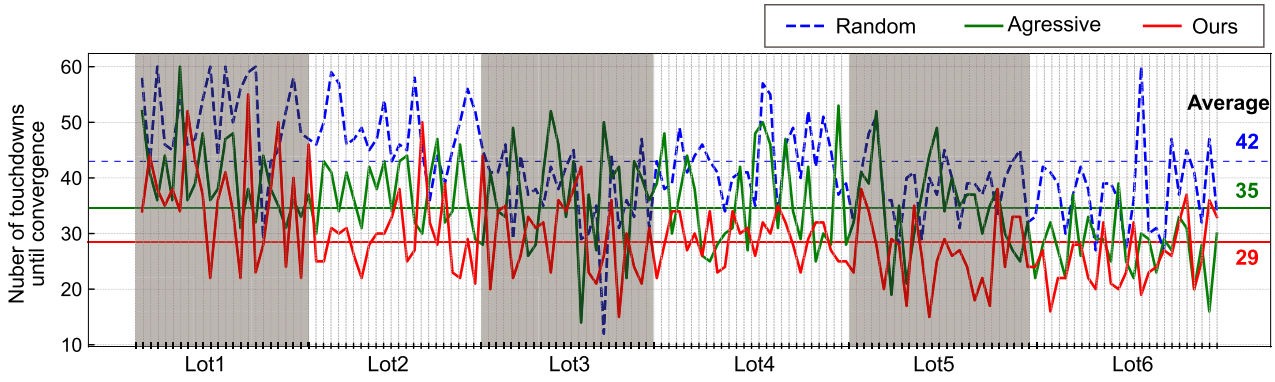
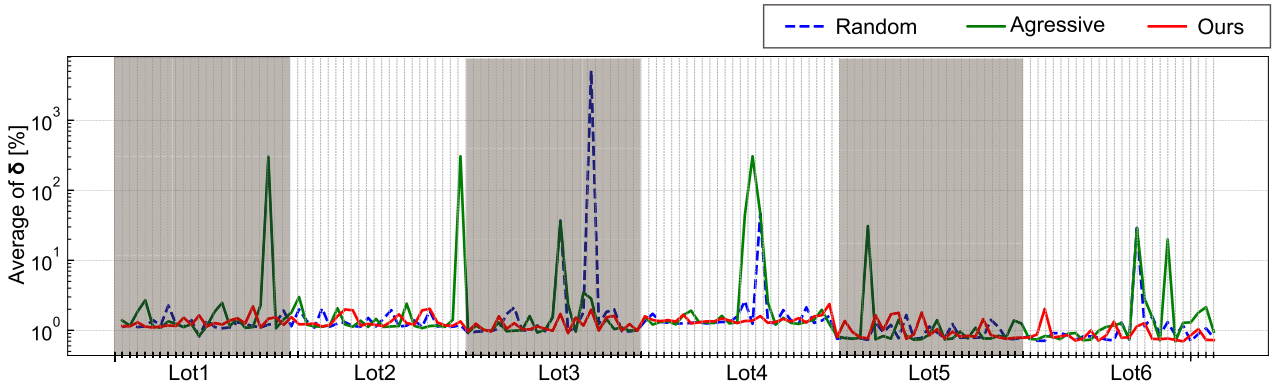


Fig. 12 Comparison of the average $\boldsymbol{\delta}$ of all the wafers per touchdown.

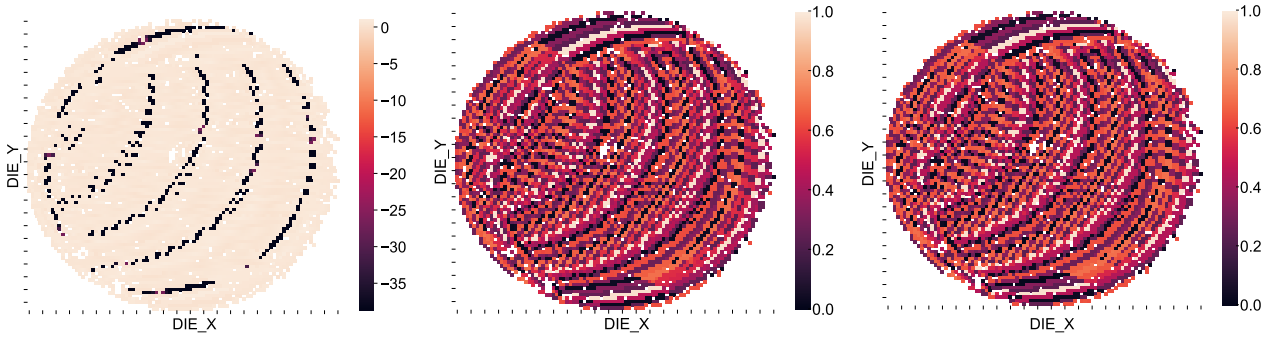


(a) Number of iterations



(b) Average of $\boldsymbol{\delta}$

Fig. 13 Changes of the number of iterations and the averages of $\boldsymbol{\delta}$ for the 143 wafers of the six lots.



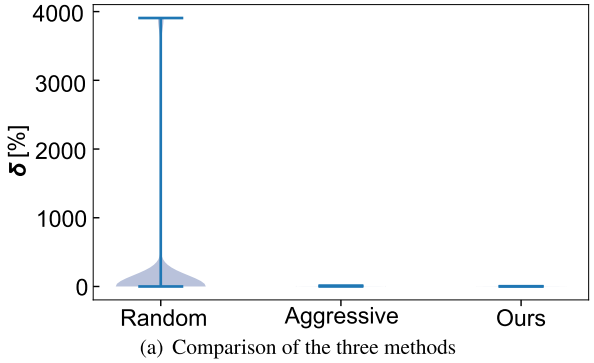
(a) Random sampling (b) Aggressive sampling (c) Ours

Fig. 14 Heat maps of the predicted characteristics for the 18 touchdowns.

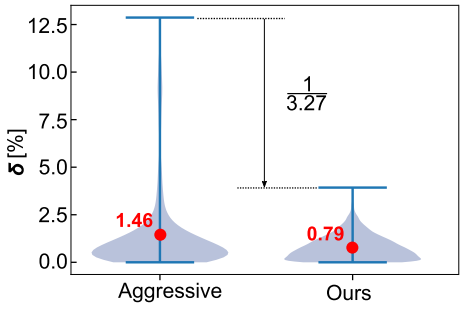
all wafers.

To demonstrate the convergence ability of each sampling method, the averages of $\boldsymbol{\delta}$ and the number of touchdowns until convergence as a function of the wafer ID of the 143 wafers are illustrated in Fig. 13, for which we consider a convergence as an error when the change in error is 1.0% or less for ten consecutive counts. As shown in Fig. 13(a), the proposed method exhibits the fewest touchdowns for the convergence condition. The average touchdowns of the random, aggressive, and proposed sampling methods are 42, 35, and 29 times, respectively. Notably, wafers showing 60 touchdowns represent a case when the convergence condition is not achieved within 60 touchdowns. The proposed method converges optimally, whereas aggressive and random sampling do not converge in some wafers, as depicted in Fig. 13(b).

The prediction results are displayed for each method at the 18-th touchdown of the first wafer of lot 6. Observably, the prediction results of random sampling are not sufficient compared to those of the other two methods, greatly exceeding the range of the normalized range (that is, 0 to 1). In contrast, excellent agreement is observed between the aggressive sampling and proposed sampling methods compared with Fig. 4, because they are visually very similar. To quantitatively evaluate the prediction results in Fig. 14, the violin plots for the three methods for the 18 touchdowns in Fig. 11(b) are presented in Fig. 15. The results plotted in Fig. 15(a) clear indicates that not only the average δ but also the variance of the estimation errors can be reduced using the proposed method. The maximum and average δ values of the aggressive method are 12% and 1.46%, respectively, while those of the proposed method are 3.92% and 0.79%, respectively, as depicted in Fig. 15(b). Compared with aggressive sampling, the proposed method halves the average of δ and improves the maximum of δ by 3.27 times. These results conclusively demonstrate that the proposed method successfully yields highly accurate wafer-level variation modeling even in a multisite test environment.



(a) Comparison of the three methods



(b) Comparison of the aggressive and our sampling methods

Fig. 15 Violin-plots of $\boldsymbol{\delta}$ for each method, where the distributions of $\boldsymbol{\delta}$ are shown. $\boldsymbol{\delta}$ of the random sampling is widely distributed compared to those of other methods.

5. Conclusion

We proposed a new technique for wafer-level spatial correlation modeling in multi-site RF IC testing. Our method employs GP regression, which is a statistical modeling method that predicts the value of an unmeasured point using a small amount of sampling data. We apply GP individually by partitioning the die location on a wafer according to the site information provided by test engineers. We also proposed an active sampling method that uses the predictive variance calculated via GP to achieve better prediction results while minimizing measurement costs. Our experimental results using an industrial production test-dataset demonstrated that the proposed method has a prediction error 19.46 times smaller than that of the conventional method. Furthermore, the proposed sampling method provides an equivalent prediction accuracy to the conventional method with only 18 touchdown measurements, which is only 3% of the number of

touchdowns required for full measurement. By contrast, random sampling requires over 60 measurements for equivalent prediction accuracy. Moreover, we confirmed that the proposed method reduces the average error by 3.42% with 60 touchdowns. Our method achieves better prediction results under multi-site testing by considering an actual touchdown, unlike all existing methods that evaluate using only one DUT measurement.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant No 22K11954.

References

[1] L.C. Wang, "Experience of data analytics in EDA and test — Principles, promises, and challenges," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol.36, no.6, pp.885–898, 2017.
[2] H.G. Stratigopoulos, "Machine learning applications in IC testing," Proc. IEEE European Test Symposium, 2018.
[3] M. Shintani, M. Inoue, and Y. Nakamura, "Artificial neural network based test escape screening using generative model," Proc. IEEE International Test Conference, p.9.2, 2018.
[4] S. Reda and S.R. Nassif, "Accurate spatial estimation and decomposition techniques for variability characterization," IEEE Trans. Semicond. Manuf., vol.23, no.3, pp.345–357, 2010.
[5] X. Li, R.R. Rutenbar, and R.D. Blanton, "Virtual probe: A statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," Proc. IEEE/ACM International Conference on Computer-Aided Design, pp.433–440, 2009.
[6] W. Zhang, X. Li, and R.A. Rutenbar, "Bayesian virtual probe: Minimizing variation characterization cost for nanoscale IC technologies via Bayesian inference," Proc. ACM/EDAC/IEEE Design Automation Conference, pp.262–267, 2010.
[7] W. Zhang, X. Li, F. Liu, E. Acar, R.A. Rutenbar, and R.D. Blanton, "Virtual probe: A statistical framework for low-cost silicon characterization of nanoscale integrated circuits," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol.30, no.12, pp.1814–1827, 2011.
[8] S. Zhang, F. Lin, C.K. Hsu, K.T. Cheng, and H. Wang, "Joint virtual probe: Joint exploration of multiple test items' spatial patterns for efficient silicon characterization and test prediction," Proc. IEEE Design Automation and Test in Europe, 2014.
[9] N. Kupp, K. Huang, J.M. Carulli, Jr., and Y. Makris, "Spatial correlation modeling for probe test cost reduction in RF devices," Proc. IEEE/ACM International Conference on Computer-Aided Design, pp.23–29, 2012.
[10] A. Ahmadi, K. Huang, S. Natarajan, C. John M., Jr., and Y. Makris, "Spatio-temporal wafer-level correlation modeling with progressive sampling: A pathway to HVM yield estimation," Proc. IEEE International Test Conference, p.18.1, 2014.
[11] K. Huang, N. Kupp, C. John M., Jr., and Y. Makris, "Handling discontinuous effects in modeling spatial correlation of wafer-level analog/RF tests," Proc. IEEE Design Automation and Test in Europe, pp.553–558, 2013.
[12] E.J. Marinissen, A. Singh, D. Glotter, M. Esposito, J.M. Carulli, A. Nahar, K.M. Butler, D. Appello, and C. Portelli, "Adapting to adaptive testing," Proc. IEEE Design Automation and Test in Europe, pp.556–561, 2010.
[13] K.R. Gotkhindikar, W.R. Daasch, K.M. Butler, J.M. Carulli, Jr., and A. Nahar, "Die-level adaptive test: Real-time test reordering and elimination," Proc. IEEE International Test Conference, p.15.1, 2011.
[14] E. Yilmaz, S. Ozev, O. Sinanoglu, and P. Maxwell, "Adaptive testing: Conquering process variations," Proc. IEEE European Test Symposium, 2012.
[15] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the *EM* algorithm," Journal of the Royal Statistical Society. Series B (Methodological), vol.39, no.1, pp.1–22, 1977.
[16] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, 2006.
[17] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[18] N. Sumikawa, L.C. Wang, and M.S. Abadir, "An experiment of burn-in time reduction based on parametric test analysis," Proc. IEEE International Test Conference, p.19.3, 2012.
[19] T. Lehner, A. Kuhr, M. Wahl, and R. Brück, "Site dependencies in a multisite testing environment," Proc. IEEE European Test Symposium, 2014.
[20] P.O. Farayola, S.K. Chaganti, A.O. Obaidi, A. Sheikh, S. Ravi, and D. Chen, "Quantile — Quantile fitting approach to detect site to site variations in massive multi-site testing," Proc. IEEE VLSI Test Symposium, 2020.
[21] D. Steinley, "K-means clustering: A half-century synthesis," British Journal of Mathematical and Statistical Psychology, vol.59, no.1, pp.1–34, 2006.
[22] K.M. Butler, A. Nahar, and W.R. Daasch, "What we know after twelve years developing and deploying test data analytics solutions," Proc. IEEE International Test Conference, 2016.
[23] S. Seo, M. Wallat, T. Graepel, and K. Obermayer, "Gaussian process regression: Active data selection and test point rejection," Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks, pp.241–246, 2000.
[24] M. Shintani, R.U.H. Mian, T. Nakamura, M. Kajiyama, M. Eiki, and M. Inoue, "Wafer-level variation modeling for multi-site RF IC testing via hierarchical Gaussian process," Proc. IEEE International Test Conference, pp.103–112, 2021.
[25] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[26] M.G. Genton, "Classes of kernels for machine learning: A statistics perspective," The Journal of Machine Learning Research, vol.2, pp.299–312, 2001.
[27] S. Ohkawa, M. Aoki, and H. Masuda, "Analysis and characterization of device variations in an LSI chip using an integrated device matrix array," IEEE Trans. Semicond. Manuf., vol.17, no.2, pp.155–165, 2004.
[28] S. Saxena, C. Hess, H. Karbasi, A. Rossoni, S. Tonello, P. McNamara, S. Lucherini, S. Minehane, C. Dolainsky, and M. Quarantelli, "Variation in transistor performance and leakage in nanometer-scale technologies," IEEE Trans. Electron Devices, vol.55, no.1, pp.pp131–pp144, 2008.
[29] P.J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational & Applied Mathematics, vol.20, pp.53–65, 1987.
[30] C. Goutte, P. Toft, E. Rostrup, F.A. Nielsen, and L.K. Hansen, "On clustering fMRI time series," NeuroImage, vol.9, no.3, pp.298–310, 1999.
[31] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," Communications in Statistics, vol.3, no.1, pp.1–27, 1974.
[32] T. Nagao, T. Nakamura, M. Kajiyama, M. Eiki, M. Inoue, and M. Shintani, "Wafer-level characteristic variation modeling considering systematic discontinuous effects," Proc. IEEE/ACM Asia and South Pacific Design Automation Conference, pp.442–448, 2023.
[33] B. Tang, "Orthogonal array-based latin hypercubes," Journal of the American Statistical Association, vol.88, no.424, pp.1392–1397, 1991.
[34] GPy, "GPy: A Gaussian process framework in Python," http://github.com/SheffieldML/GPy, since 2012.
[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,

O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and ÉÉ. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol.12, no.85, pp.2825–2830, 2011.

[36] S. Park and S. Choi, "Hierarchical Gaussian process regression," Proc. Asian Conference on Machine Learning, pp.95–110, 2010.

[37] D.T. Nguyen, M. Filippone, and P. Michiardi, "Exact Gaussian process regression with distributed computations," Proc. ACM/SIGAPP Symposium on Applied Computing, pp.1286–1295, 2019.

Riaz-ul-haque Mian received B.E. degree from Islamic University Bangladesh and M.E degree from Bangladesh University of Engineering and Technology (BUET) in 2009 and 2017 respectively. He received his Ph.D. degree from Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Japan in 2021. He was with Synesis IT Ltd. and GIZ from 2009 to 2017. Currently, he is an assistant professor of the department of Information System Design and Data Science at Shimane University, Japan. His research interests include computer architecture and decimal computing. He is a member of IEEE.

Tomoki Nakamura received B.E from Nagasaki University, Japan in 2007. After graduating, he joined Sony Semiconductor Manufacturing Corporation. As a product test engineer, he engaged in the development and mass production of System Level Tests for various advanced process SoCs. He is also responsible to implement software and embedded systems to apply AI technology to semiconductor testing, and promotes internal use.

Masuo Kajiyama received B.E. and M.E. degrees from Hiroshima University, Hiroshima, Japan in 2002 and 2004, respectively. He joined Sony Semiconductor Manufacturing Corporation, Nagasaki, Japan, since 2004. He engaged in mass production of various advanced process SoCs as a product technology engineer and manager, and promoted productivity improvement through mass production test data analysis using AI technology.

Makoto Eiki received B.E from from Kumamoto University, Japan in 2004. After graduating, he joined Sony Semiconductor Manufacturing Corporation. He engaged in mass production of various advanced process SoCs as a product technology engineer and manager, and promoted in-house activities to apply AI technology to semiconductor testing. He is also currently a member of the doctoral course at Nara Institute of Science and Technology as a working doctoral student.

Michihiro Shintani received B.E. and M.E. degrees from Hiroshima City University, Hiroshima, Japan, and a Ph.D. degree from Kyoto University, Kyoto, Japan, in 2003, 2005, and 2014, respectively. He was with Panasonic Corporation, Osaka, Japan, from 2005 to 2014; with Semiconductor Technology Academic Research Center (STARC), Yokohama, Japan, from 2008 to 2010; with Kyoto University, Kyoto, Japan, from 2014 to 2017; and with Nara Institute of Science and Technology, Ikoma, Japan, from 2017 to 2022. In 2022, he joined the Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto, Japan, where he is currently an Associate Professor. His research interests include reliability-aware LSI design, device modeling, and circuit simulation.