

Edge-over-Erosion Error Prediction Method Based on Multi-Level Machine Learning Algorithm

Daisuke FUKUDA^{†,††a)}, Member, Kenichi WATANABE^{†††}, Naoki IDANI^{†††}, Yuji KANAZAWA[†], Nonmembers, and Masanori HASHIMOTO^{††}, Member

SUMMARY As VLSI process node continue to shrink, chemical mechanical planarization (CMP) process for copper interconnect has become an essential technique for enabling many-layer interconnection. Recently, Edge-over-Erosion error (EoE-error), which originates from overpolishing and could cause yield loss, is observed in various CMP processes, while its mechanism is still unclear. To predict these errors, we propose an EoE-error prediction method that exploits machine learning algorithms. The proposed method consists of (1) error analysis stage, (2) layout parameter extraction stage, (3) model construction stage and (4) prediction stage. In the error analysis and parameter extraction stages, we analyze test chips and identify layout parameters which have an impact on EoE phenomenon. In the model construction stage, we construct a prediction model using the proposed multi-level machine learning method, and do predictions for designed layouts in the prediction stage. Experimental results show that the proposed method attained 2.7~19.2% accuracy improvement of EoE-error prediction and 0.8~10.1% improvement of non-EoE-error prediction compared with general machine learning methods. The proposed method makes it possible to prevent unexpected yield loss by recognizing EoE-errors before manufacturing.

key words: Edge-over-Erosion, CMP, manufacturability, machine learning

1. Introduction

Copper (Cu) interconnect is widely applied at sub-90 nm process technologies, because of its lower resistance as compared to aluminum. Instead of dry-etch process used in aluminum process, Cu interconnect structure is constructed by damascene process. In this process, interconnect trenches and via holes are etched after interlayer dielectric (ILD) deposition. Then a thin barrier metal layer, which facilitates Cu film generation, is deposited as a seed layer. Next, Cu is deposited to fill up the trenches and holes on the whole wafer surface by Electro-Chemical Plating (ECP) process. Finally, Cu outside the trenches and holes is removed to generate interconnect patterns.

Chemical Mechanical Planarization (CMP) is a technique to remove redundant Cu and to planarize the surface of wafers. However, CMP process causes undesirable wafer surface nonuniformity [1], [2]. This Cu and ILD thickness

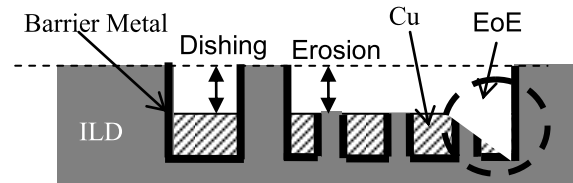


Fig. 1 Cross sectional view of interconnect layer.

variations strongly depend on thickness variations in ECP process, chip layout pattern, and polishing rate of each material [3], [4]. Figure 1 shows the cross sectional view of thickness variations. Dishing is the height difference between Cu and neighbor ILD region, and erosion is the difference in ILD height between pre- and post-CMP processes. Thickness variations produce chip performance degradation due to an increase in wire resistance and capacitance, and may cause open/short errors. Furthermore, thickness variations are propagated to upper layers, and the accumulated variations could cause an excess of depth-of-focus in photolithography and short errors in the worst case [4], [5]. This thickness variation is getting severer according to device miniaturization, which imposes more precise planarization on CMP process. In recent technologies, thickness variations due to CMP process are major cause of yield loss [6]–[9].

Recently Edge-over-Erosion error (EoE-error) is frequently observed [10] in addition to dishing and erosion. Figure 1 also illustrates the cross section of an EoE-error. This error occurs at selective CMP step. At this step, multiple materials are polished simultaneously, where the removal rate of Cu is much higher than that of barrier metal. At the location where an EoE-error occurs, the unexpected overpolishing error can be more than ten times larger than an estimate which is predicted from material-dependent removal rates. EoE-errors cause open errors, and furthermore may cause short errors at its upper layer. Although several works investigated the root cause of EoE phenomenon [10], [11], little is known about the mechanism of this problem.

To avoid EoE induced open and short errors, we need to mitigate EoE-errors. However, it is too costly to modify chip layout to mitigate EoE-errors after manufacturing and testing. Another approach for EoE mitigation is to tune some CMP process parameters, such as slurry, polishing pad, rotation speed, pressure, etc. [11], but it involves a comprehensive and consequently expensive tuning because CMP pro-

Manuscript received March 13, 2014.

Manuscript revised June 30, 2014.

[†]The authors are with FUJITSU LABORATORIES LTD., Kawasaki-shi, 211-8588 Japan.

^{††}The authors are with FUJITSU SEMICONDUCTOR LTD., Kuwana-shi, 511-0192 Japan.

^{†††}The authors are with the Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University, Suita-shi, 565-0871 Japan.

a) E-mail: d-fukuda@jp.fujitsu.com

DOI: 10.1587/transfun.E97.A.2373

cess is very sensitive to various parameters and their interdependency. CMP simulation is an effective method to predict thickness variations before manufacturing and nowadays has become an essential step to optimize wafer surface uniformity in chip design flow [12], [13]. However, no tools take into account EoE-error problem explicitly. Therefore, it is highly demanded to develop a prediction method that systematically estimates EoE-errors in design time to avoid unexpected yield loss.

Motivated by this, we developed a systematic EoE prediction method aiming at mitigating EoE-errors in design time. Contributions of this work include the followings:

- This is the first work that presents an EoE-error prediction method. Because of the high sensitivity of EoE phenomenon to CMP process condition, there is a certain amount of noise peculiar to individual chips, and hence an overfitting problem easily happens with normal machine learning algorithms when pursuing high accuracy. We thus explored and applied multi-level machine learning algorithm suitable for EoE-error prediction.
- We present a procedure that extracts model parameters which should be included as variables in machine learning for EoE model construction. By analyzing test chips, we find the layout parameters which really have an impact on the ambiguous EoE phenomenon, and screen out non-influential parameters which degrade accuracy as noise sources.
- We assessed the accuracy of the proposed method with industrial chip data.

The rest of the paper is organized as follows: In Sect. 2 we provide the overview of the proposed method consisting of four stages; error analysis stage, layout parameter extraction stage, model construction stage and prediction stage. Then we introduce error analysis and parameter extraction stages in Sect. 3 and model construction and prediction stages in Sect. 4. Section 5 presents the results and analysis of the proposed method. Finally, Sect. 6 concludes this paper.

2. Concept and Overview

In this section, we explain the concept and overview of the proposed method.

2.1 Concept of EoE-Error Prediction Method

As mentioned in the previous section, the mechanism of EoE occurrence is complicated and is not understood well enough to build a physical EoE model. Thus, instead of constructing a physical EoE model, the proposed method employs machine learning techniques to predict EoE-error locations with measurement data of real chips. Here, machine learning technique is a general method for statistical data analysis and a powerful tool for finding regularities in the dataset.

The proposed method first selects an appropriate set of

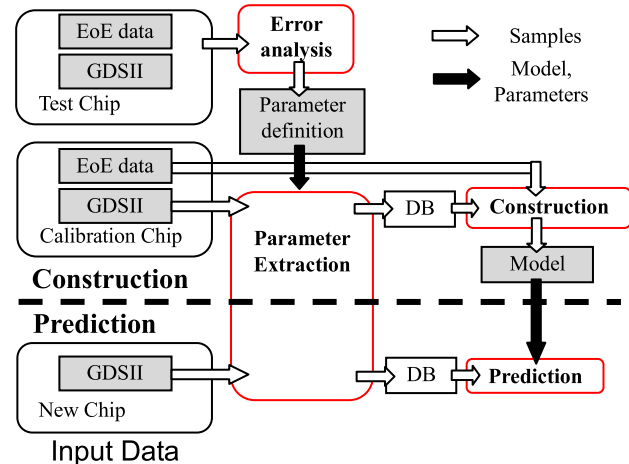


Fig. 2 Overview of proposed EoE-error prediction method.

layout parameters to model EoE-error by analyzing measurement data of the test chip designed for this EoE modeling purpose, because at the beginning there is little information about phenomenon in the process technology of interest. Then the proposed method constructs a model which has these layout parameters as variables using another measurement data of a calibration chip which is designed for a real product and includes various layout patterns.

2.2 Overview of EoE-Error Prediction Method

Figure 2 shows the flow of EoE-error prediction method. This method consists of four stages: error analysis stage, parameter extraction stage, construction stage and prediction stage. We first analyze the EoE-error measurement data of the test chips to clarify which layout parameter should be included as variables in the prediction model. Next, layout parameters selected in the previous error analysis are extracted from the calibration chips. Then, model construction process is carried out with machine learning methods and a prediction model is constructed. Finally, the constructed model is applied to new designs and we predict EoE-errors before manufacturing.

To find parameters which affect EoE-errors, a detailed analysis is executed with the surface measurement data of the test chip. Figure 3 shows the details of the test chip. We define two terms in Fig. 3 as follows:

- **Module:** A module is filled with a set of regular wires. Each module has parameters: wire width, metal density, and module size. Metal density is defined as the ratio of Cu wires area to the module area. Within the module, the wire width and space are uniform. A module is filled with Cu only when metal density is 100% and filled with dielectric only when metal density is 0%. The area outside modules is filled with dummy metals.
- **Array:** An array consists of modules. Each array has its own target parameter to investigate the EoE dependency on the parameter. In each array, the modules have different values of the target parameter, while other parameters

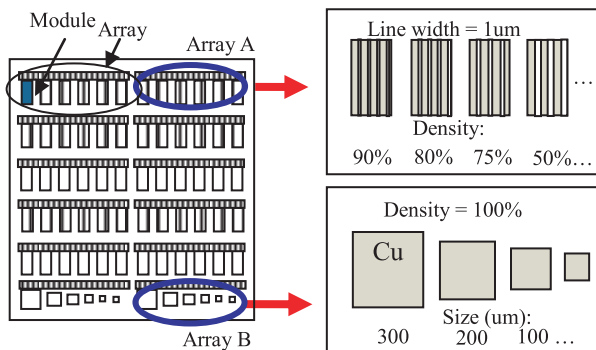


Fig. 3 Test chip structure.

are set to the same value in all the modules.

For example, each module in array A has the same module size, the same wire width, and different metal densities. In array B, module size is various but other parameters are the same. Analyzing the post-CMP surface of such patterns enables us to roughly recognize parameters which affect EoE-errors.

Then, layout parameters of interest, e.g. metal density and line width, are extracted from the original chip data and converted to new database. Hereafter, layout parameters mean these parameters. Generally, the physical chip data is recorded in GDSII (Graphic Data System 2) form at or OASIS (Open Artwork System Interchange Standard) format. These databases have a large file size (more than tens of gigabytes) since they have the entire chip information, and it costs much to get layout parameters directly from the original database. To reduce the size of database and calculation cost, a whole chip is discretized into small tiles and related parameters are extracted and recorded for each tile.

In the model construction stage, we use an industrial chip as a calibration chip and build the EoE prediction model that has the layout parameters selected in error analysis stage as variables. In contrast with the test chip mentioned above, a wide range of layout parameter values and more complex combinations of multiple layout parameters are included in a real design. Therefore the training with an industrial chip is suitable for evaluating the effect of each parameter quantitatively. Besides, the EoE-error area is generally very small (< 1% of whole chip area). The number of EoE-error tiles, which are tiles that include EoE-error, is much smaller than that of non-EoE-error tiles in a chip. When we build a prediction model, the training dataset from industrial chips becomes imbalanced, i.e. the numbers of EoE-error and non-EoE-error tiles in the training dataset become imbalanced, which causes poor performance of machine learning algorithms. Instead, in order to construct a precise model, the training dataset which includes EoE-error and non-EoE-error tiles with an appropriate ratio (e.g. 50%) must be prepared by non-uniform sampling and given to machine learning algorithms.

After constructing the EoE prediction model with machine learning algorithms, we predict EoE-errors of new

chips in prediction stage. Note that this prediction model can be applied to the chips which will be manufactured under the same process condition. If the process condition is changed, model construction for the new condition needs to be executed.

3. Error Analysis and Parameter Extraction Stages

This section explains the error analysis and parameter extraction stages in which we extract layout parameters with analyzing the post-CMP surface data of the test chip.

The test chip includes modules with various values of line width, density, and module sizes. The space between modules is filled with dummy metal patterns. For the purpose of data size reduction, the whole chip is divided into small tiles, as mentioned before. The prediction model is built as a function of average parameters of adjacent tiles instead of individual metal segments. The tile size has an impact on trade-off relation between computational cost and estimation accuracy, and a tile size of 10–40 μm is often used in ECP and CMP process simulation for sub-100 nm processes [14], [15]. In this work, the tile size was set to 10 × 10 μm thinking much of the accuracy. Considering the impact on wire parameter variation and Cu residue of upper layer, we define the EoE-error as the place at which a height of erosion is larger than 40% of wire height.

Figures 4 and 5 show the cross section of some modules in the test chip after CMP. In all cases erosion is observed at the high metal density side of the boundary between module and inter-module area, where the inter-module area is filled with dummy metals. More importantly, in the cases of (a) of both figures, EoE-errors are observed. The height of EoE-error is as tall as wire height and then open error occurs. With analyzing these data carefully, the following layout parameters seem to have a relation to EoE-errors.

- (1) Metal density

We first examine Figs. 4(a) and (b). EoE-errors are observed at the place where the metal density is higher than its adjacent area. When the difference in metal density between adjacent areas is not sufficient, EoE-errors are not observed (Fig. 4(b)). On the other hand, not only the difference between adjacent areas, but the absolute value of metal density plays an important role. In Fig. 4(c), the metal density difference is larger than that of case (a), but no EoE-errors are observed. For these reason, we use metal density of the tile and max/min metal density of adjacent tiles.
- (2) Effective density

In spite that metal density is 0% in the module area in both cases of Fig. 5, EoE-errors are observed only in case (a). This difference suggests that the metal density variation within a small region is filtered out and high frequency components of metal density in space need to be eliminated for EoE-error prediction. For this purpose, we introduce a parameter called “effec-

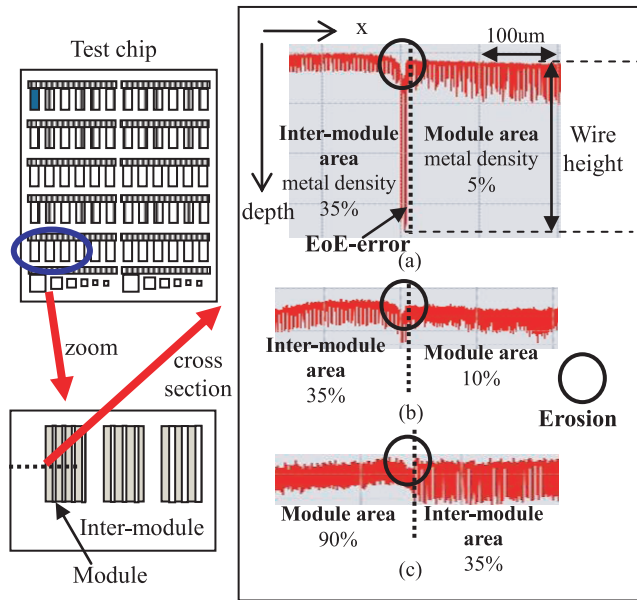


Fig. 4 Relationship between EoE and module density.

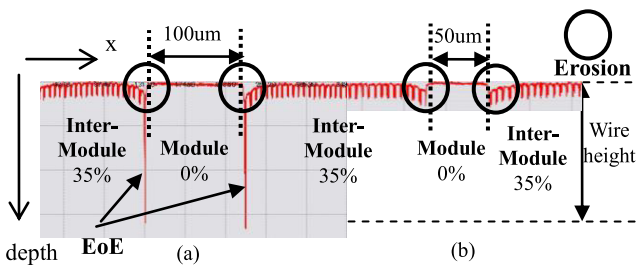


Fig. 5 Relationship between EoE and module size.

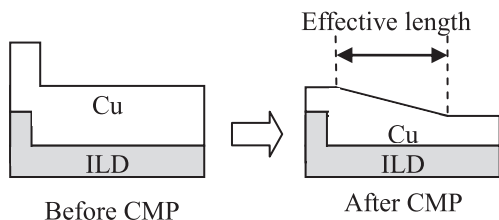


Fig. 6 Definition of effective length.

“effective length”. Figure 6 shows the definition of effective length. The effective length is the distance in which a feature influences planarization in polishing process. This parameter is also called “planarization length” or “interaction length”, and an appropriate value of effective length is determined by CMP process modeling methodology in each process condition [2], [5], [15], [16]. Using this effective length, we define effective density of each tile such that effective density is the average metal density within the range of effective length from the tile of interest (Fig. 7).

With further observations from the test chip, it turns out that EoE-errors occur at the place where the metal den-

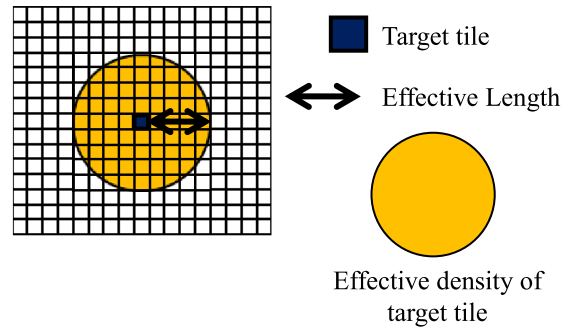


Fig. 7 Region for effective density computation.

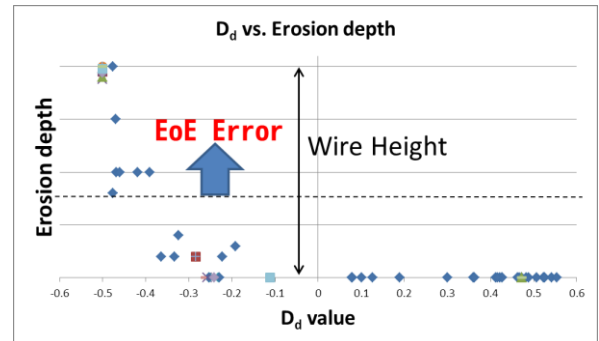


Fig. 8 Density deviation vs. EoE-error occurrence in a 65 nm process.

sity is higher than the effective density. We define a parameter of density deviation D_d as follows:

$$D_d = \frac{D_e - D}{D} \quad (1)$$

where D is metal density and D_e is effective density. Here, metal density D is defined as the average metal density within the tile of interest. Figure 8 illustrates the relationship between D_d and erosion depth (EoE occurrence) at the edge of various modules in a 65 nm technology node. This result shows that D_d is a good indicator for EoE-error.

(3) Line width

Wider metal lines are likely to become a cause of EoE-errors. The EoE-error of narrow lines ranges over multiple materials, while that of wider lines is mainly due to the disappearance of Cu metal (Fig. 9). Generally, the polishing rate of barrier metal is much smaller than that of other materials, and hence we have to consider the line width.

We then make a database including following parameters for each discretized tile: metal density, max/min metal density of adjacent tiles, effective density, density deviation, and line width. This database will be used in the next model construction stage.

For this 65 nm technology node, we select these 6 layout parameters to model EoE-errors. For another technology node, such as advanced technology, the layout parameters that have impact on EoE-errors may change. On the other

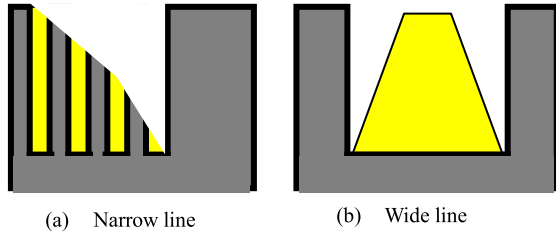


Fig. 9 Detailed shapes of EoE cross section.

hand, it is expected that error analysis stage identifies influential layout parameters on EoE-errors at a particular technology of interest, since a test chip fabricated in the technology, which includes various layout patterns and covers wide range of parameters, is newly analyzed. Once the influential parameters are identified, we construct a prediction model that have the influential parameters as input variables and use the model for EoE-error prediction.

It should be noted that these parameters are affected by process variation. Especially etching and lithography processes have a great influence on line width variation [17], [18]. Additional process might be required to eliminate the impact of process variation if the impact is not negligible.

4. Model Construction and Prediction Stage

In model construction stage, we build a prediction model that has the layout parameters selected in Sect. 3 as variables with binary classification method using machine learning algorithms. In each tile of the chip, a prediction model with layout parameters of each tile mentioned in Sect. 3 can predict whether EoE-error occurs or not.

Before explaining details of the prediction model construction, an accuracy metric of the prediction, which is considered in this paper, is introduced. As mentioned before, the dataset of industrial chips is imbalanced. In imbalanced dataset, the model performance cannot be expressed in terms of the average accuracy.

Table 1 shows the confusion matrix. Each column of the matrix represents the instances of prediction class, and each row represents the instances of an actual class. For example, when 1% samples are EoE-error and others are non-EoE-error, 99% accuracy is achieved by the model that all samples are judged as non-EoE-error ((a+d)/(a+b+c+d)). In this case, we cannot identify the samples that are likely cause EoE-error with such a model even while the accuracy is 99%. Considering this fact, we use geometric mean (g-mean) as a metric to evaluate the accuracy of the prediction model. G-mean g is defined as:

$$g = \sqrt{P_{err} \times P_{ok}}$$

$$P_{err} = \frac{a}{a+b}, P_{ok} = \frac{d}{c+d} \quad (2)$$

where P_{err} is the rate of correctly predicted EoE-errors from all EoE-error samples, P_{ok} is that of non-EoE-error samples, a, b, c, and d is the number of instances in Table 1. Because

Table 1 Confusion matrix.

		Predicted:	
		EoE-error	Non-EoE-error
Actual:	EoE-error	A	B
	Non-EoE-error	C	d

accuracy is calculated on the majority class and minority class separately, g-mean is suitable for evaluating the accuracy of imbalanced data classification problems [19]. In the previous case, g-mean value is 0 because P_{ok} value is 100% and P_{err} value is 0%.

4.1 Machine Learning Algorithms

This subsection briefly summarizes machine learning kernels we employed in this work.

4.1.1 RPART (Recursive Partitioning)

RPART [20] is a classification method using a 2-stage procedure, and provides resulting models represented by binary trees. This technique splits the samples using one input variable, i.e. a layout parameter in this paper, with a threshold value which makes the gain of splitting index maximum. This routine is applied to each separated group recursively until the subset size reaches to the minimum threshold or until no improvement can be obtained. In this work, we used Gini index as the splitting index. Gini index is given as follows:

$$I(g) = 1 - \sum_{i=1}^2 p_i^2 \quad (3)$$

where p_i is the fraction of samples belonging to class i (error or not) at a given node. This index reaches 0 when all the samples belong to a single class. Larger Gini index improvement indicates better sample splitting.

4.1.2 RF (Random Forest)

RF method [21] is an ensemble learning method for classification aiming to improve prediction ability and stability of RPART. RF method consists of a number of decision trees and performs classification by majority vote of all the trees. This method is processed with the following steps:

- Step 1 N sets of bootstrapped samples are extracted from the original data.
- Step 2 For each set, a tree is built by RPART method with m variables randomly selected out of M variables, where the variables correspond to layout parameters in this paper.
- Step 3 In prediction process, a new sample is classified by individual trees, and the majority result is selected as the classification result.

4.1.3 SVM (Support Vector Machine)

Suppose that each tile m can be described as a vector of n layout parameters $x = (f_1, \dots, f_n)$. SVM [22] constructs hyperplanes that optimally classify the data with these training vectors. Hyperplanes are set so as to attain the largest separation margin, where separation margin is the distance to the nearest training data. The vectors which form the boundary are called as support vectors.

On the other hand, there are a number of data sets which cannot be well separated linearly. For such data sets, kernel trick [23] provides improved separability. The kernel trick maps the original samples into a higher dimensional space, and it provides a method to nonlinearly separate the data set. Besides, there are several popular kernels, and this work used RBF (radial basis function) kernel. RBF kernel is defined as follows:

$$K(x, x_j) = \exp\left(-\sigma \|x - x_j\|^2\right) \quad (4)$$

where x and x_j are feature vectors, and σ is a free parameter.

Soft margin method [24] is also applied to our SVM prediction model. When error vectors are not separable due to EoE-error complexity and/or noise, slack variable ξ is introduced to allow mislabeled samples by paying violation penalty. The optimization problem is:

$$\begin{aligned} \min_{a,b,\xi} \quad & \frac{1}{2} \|a\|^2 + C \sum_{i=1}^m \xi_i \\ \text{Subject to} \quad & \xi_i \geq 1 - y_i (a \cdot x_i + b) \\ & \xi_i \geq 0 \quad (i = 1, 2, \dots, m) \end{aligned} \quad (5)$$

where a and b are parameters of hyperplane, y_i is a sign function of $(a \cdot x_i + b)$, and C is a parameter of soft margin to control the weight of penalty.

4.2 Multi-Level Machine Learning Algorithm

When pursuing accurate prediction of EoE-errors with the above algorithms, models tend to be more and more complex. In other word, the size of decision trees becomes large in RPART method and the number of support vectors increases in SVM method.

Figure 10 shows the complexity of SVM model in 2-dimensional graph. Figure 10(a) shows a simple model composed of two support vectors. All vectors above the dotted line are regarded as EoE-errors. A complex model is shown in Fig. 10(b), where the number of support vectors is increased. The error region is smaller than that of Fig. 10(a) and the number of mislabeled samples is decreased.

While a complex model improves the value of Eq. (2), an increase in model complexity may cause an overfitting problem. Overfitting degrades the generality of the model, which results in a bad performance in predicting new data in spite that the prediction for known data is accurate.

To achieve high accuracy without degrading general-ity, we introduce a multi-level machine learning algorithm

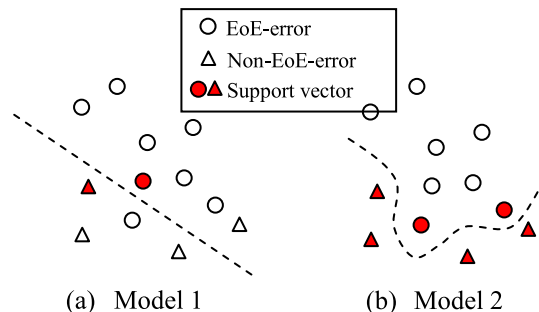


Fig. 10 Model complexity of SVM method.

(MML). In this method, we apply multiple trainings to the data in sequence. MML consists of screening and brushup steps. The aim of screening step is to reduce a number of non-EoE-error samples and outliers. This step is helpful for complexity reduction of the model which will be built at the next brushup step. In other words, this step reduces error classification patterns and outlier samples to be considered at brushup step to avoid overfitting problem. At brushup step, prediction model is constructed with samples labeled as an EoE-error in screening step. Details of each step will be explained in the following.

4.2.1 Screening Step

At screening step, we apply the first training and predict EoE-errors. The samples labeled as an EoE-error at this step go to next step and the others are regarded as non-EoE-error samples. Because the purpose of this step is the screening of non-EoE-error samples, high P_{err} value in Eq. (2) is required in the model constructed at this step.

4.2.2 Brushup Step

The samples labeled as an EoE-error at screening step include many false errors, which is the non-EoE-error samples misjudged as EoE-error, since improving P_{ok} is scarcely considered at screening step. This brushup step aims to reduce false errors for attaining high g-mean value in Eq. (2). Besides, each learning method has individual features (ensemble/single classifier, linear/non-linear classification, for example), and samples which are poorly classified by one method may be accurately predicted by another method. This multi-step classification is thus expected to attain higher accuracy, since advantages of both methods can be exploited while concealing disadvantages.

We attempted various combinations (strictly speaking various permutations since the order also affects the accuracy) of learning methods and numbers of learning steps. Experimental setup was the same with that in Sect. 5, and chip C1 data was used here. The detail will be explained later. We first tested 2-level permutations that include RF, where RF attained the highest accuracy among single-level prediction methods explained in Sect. 4.1, and found that those permutations degraded the accuracy compared to

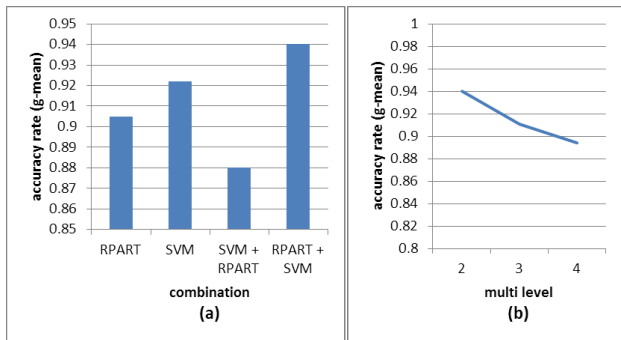


Fig. 11 Comparison of Accuracy rate of (a) combinations of learning method and (b) number of multi-level.

single-level RF. We thus excluded the permutations that included RF. We then evaluated the combinations of RPART and SVM. Figure 11(a) shows the accuracy rate of Eq. (2), and we can see that the permutation of RPART as the first stage and SVM as the second stage improved the accuracy, where the accuracy rates of single-stage RPART and SVM were 90.5% and 92.2%. We also tested RPART+RPART and SVM+SVM, but these could not improve the accuracy.

This result indicates that a better classification result can be expected when RPART method is applied to screening step and SVM method is applied to brushup step. We think there are two reasons for this result. 1) RPART is a simple method and this feature matches the aim of screening. 2) RPART is based on decision tree algorithm. SVM method may compensate weak points of this algorithm in EoE-error classification problem. Moreover, we attempted 2, 3, and 4-level learning steps. RPART was applied to the 1st step and SVM was applied to the other steps. Figure 11(b) shows the result. This result shows that the accuracy metric of Eq. (2) starts to degrade when the number of learning steps is 3 and more. We thus concluded that RPART and SVM methods should be used at screening and brushup steps respectively, and two-step classification with screening and brushup steps was reasonable.

4.3 Model Construction and Prediction Flow

Figure 12 shows the detailed flow of model construction and prediction stages in Fig. 2. In model construction stage, we build two prediction models with an industrial chip (calib-chip). The overall EoE-error prediction model of MML algorithm consists of Models 1 and 2 constructed in screening stage and brushup stage, respectively.

First, we construct input database that includes layout parameters and the EoE-error information for each discretized tile.

Next, EoE-error and non-EoE-error tiles are sampled from the database as a subset1, which is used as the training dataset for Model 1 construction. As previously mentioned, EoE-error/non-EoE-error class distribution in the database is imbalanced. Sampling is a common practice to improve classifier performance and numerous methods are proposed

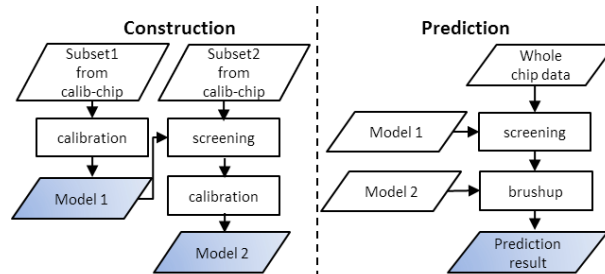


Fig. 12 Detailed flow of construction and prediction stages with multi-level machine learning method.

[19], [25]–[28]. According to a comparison of various sampling methods [29], random under sampling (RUS) method is one of the best sampling techniques for the purpose of the learning from imbalanced data. In RUS method, samples of the majority class are randomly discarded and the training dataset becomes balanced. We apply RUS method to non-EoE-error samples and make subset1 which includes all EoE-error samples and reduced non-EoE-error samples, and construction of Model 1 is carried out with this training dataset.

In the proposed MML method, samples labeled as non-EoE-error at screening step with Model 1 are discarded before brushup step and EoE-error/non-EoE-error ratio changes. Thus, Model 2 construction is carried out with new training dataset subset2. This dataset is constructed from samples labeled as an EoE-error with Model 1. Because EoE-error/non-EoE-error distribution of the data passing through Model 1 is still imbalanced, we apply RUS method to non-EoE-error samples again to make subset2 which includes all EoE-error samples and reduced non-EoE-error samples. Then the prediction model of brushup step (Model 2) is constructed with subset2.

In prediction stage, EoE-errors of new chips are predicted with Model 1 and Model 2 in sequence.

5. Experimental Results

Here we present experimental results to validate the proposed method. We obtained EoE-error data from three industrial chips. Note that the silicon measurement to identify EoE-error coordinates requires huge cost, which motivated us to develop EoE-error prediction model. We used one chip to construct the prediction models and other two chips to validate the efficiency to unknown data.

Table 2 lists the details of calibration data (C1) and data for validation (V1, V2) with 65 nm technology node. The tile size of each data was set to $10 \times 10 \mu\text{m}$. The number of EoE-error tiles was measured from actual chips whose CMP process had been completed. We extracted layout parameters explained in Sect. 3, and obtained model parameters with industrial chip C1. Chip V1 and V2 cover wide ranges of the tile-by-tile layout parameter values and have different distribution shapes of the parameters.

We implemented RPART, RF, SVM, and the proposed

Table 2 Details of calibration/validation chips.

Chip Name	C1	V1	V2
Chip size (mm ²)	5.9x5.9	7.5x7.4	7.8x6.2
Average density (%)	34.5	30.2	30.8
# of EoE-error tiles	1045	2431	2604
# of non-EoE-error tiles	343K	544K	489K
EoE-error tile ratio (%)	0.3	0.45	0.53

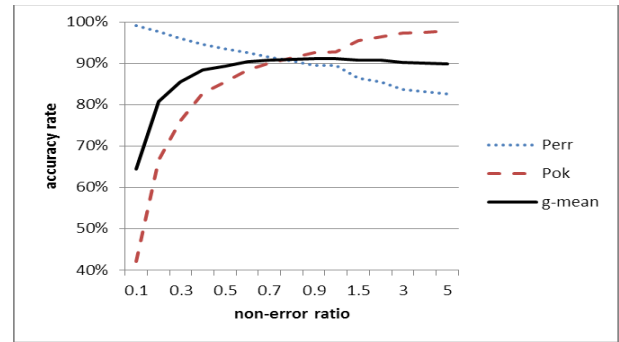
Table 3 Prediction performance comparison.

Chip		RPART	RF	SVM	MML	
					SC	SC+BU
C1	P_{err}	89.6%	100.0%	91.1%	97.6%	94.5%
	P_{ok}	91.4%	98.1%	93.2%	65.7%	93.5%
	g -mean	90.5%	99.0%	92.2%	80.1%	94.0%
	ratio	0.5	4.0	1.0	0.2	0.9
V1	P_{err}	92.3%	82.2%	92.2%	97.2%	87.1%
	P_{ok}	85.2%	91.1%	81.8%	57.9%	92.1%
	g -mean	88.7%	86.5%	86.8%	75.0%	89.6%
V2	P_{err}	86.3%	74.3%	84.9%	97.8%	88.6%
	P_{ok}	82.5%	90.0%	82.4%	44.4%	90.7%
	g -mean	84.4%	81.8%	83.7%	65.9%	89.7%

MML methods in R language [30]. In each method, training dataset consists of all EoE-error samples and proper amount of non-EoE-error samples selected with RUS method. In addition to model parameters, the balance of non-EoE-error/EoE-error class heavily affects prediction performance. Therefore we calibrate model parameters and non-EoE-error/EoE-error sample ratio of the training dataset in each method. At construction step, we use default values of $N = 500$, $M = 6$, and $m = 2$ in RF method of R library [30]. C and σ in SVM method are set as calibration parameters.

Table 3 shows the performance of each method for calibration chip C1. SC and BU represent the screening and brushup steps, respectively. The meanings of P_{err} , P_{ok} , and g -mean are the same with Eq. (2). “Ratio” denotes the ratio of non-EoE-error samples to EoE-error samples in training dataset which achieved the best value of g in C1. Sample ratio affects the relationship between EoE-error and non-EoE-error sample’s misclassification cost. Figure 13 shows the value of P_{err} , P_{ok} , and g -mean with various sampling ratio of SVM method ($C = 10$). When sample ratio value becomes large, the misclassification cost of EoE-error sample decreases and that of non-EoE-error sample increases, which results in decrease of P_{err} and increase of P_{ok} .

RF method shows good performance in the calibration chip. RF is the only method that did not miss the actual EoE-errors. MML method also attained high g -mean value. At screening step of MML, RPART method is applied and it uses smaller sample ratio (0.2) than that of single-level RPART method (0.5). In MML method, samples labeled as non-EoE-error at screening step are discarded before brushup step. High P_{err} and low P_{ok} value at screening step caused by lower sample ratio means that only outlier EoE-error samples and obvious non-EoE-error samples are

**Fig. 13** P_{err} , P_{ok} , and g -mean with various sampling ratio in SVM.

removed from dataset. Detailed classification is processed in brushup step. According to this step, the complexity of classification is reduced and the g -mean value after brushup step which is processed with SVM method is higher than that of single-level RPART and SVM method. According to the result of calibration, D_d was the most influential parameter to EoE-errors in this 65 nm technology node.

Table 3 also shows the performance of each method for the validation chips. While it achieved the highest performance for calibration chip C1, RF method shows the worst performance in both V1 and V2. This is due to the lowest P_{ok} value even though P_{err} value is higher than other single-level methods. It is considered that overfitting problem occurs in RF method.

In chip V1, similar g -mean values are observed in all the method. In contrast, compared to chip V1, the performance of single-level methods degraded in chip V2 although the proposed MML method kept up its accuracy. A possible reason why the performance in chip V1 was better than that in V2 is that chip V1 had some similarities with calibration chip C1, such as the average density. Compared with other methods, MML shows the best performance in both P_{err} and P_{ok} . MML improved P_{err} by 2.7~19.2% and P_{ok} by 0.8~10.1%. In MML method, we apply multiple simple models aiming to achieve high accuracy without sacrificing generality. This concept prevents overfitting problems in construction process and contributes to sustaining the classification performance to new chip data.

Even while P_{err} attained high percentages of 87.1% and 88.6% in chips V1 and V2, the number of non-EoE-error samples labeled as an EoE-error is more than ten times as large as the numbers of EoE-error samples labeled correctly, since the rate of EoE-error/non-EoE-error samples is imbalanced as listed in Table 2. However, the proposed method is still useful in design phase with the following three reasons.

Firstly, it is too costly in both mask cost and time to modify layout after fabrication and measurement than to modify layout in pre-manufacturing phase. This proposed method is the first solution to predict EoE-errors before manufacturing.

Secondly, potential EoE-errors are likely to exist in mislabeled non-EoE-error samples. We here define poten-

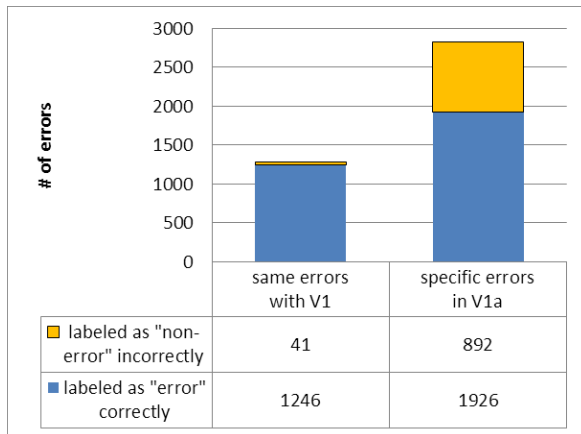


Fig. 14 Details of EoE-errors in chip V1a.

tial EoE-error as a non-EoE-error sample which is EoE-error in other chip because of inter-chip variations. According to various process variations due to, for example, CMP, ECP and wafer location, EoE-error location and EoE-error numbers are different between chips. To clarify this, we measured EoE-error of chip V1a, which was fabricated on the same wafer of chip V1 and has exactly the same layout with chip V1. Figure 14 shows the details of EoE-errors of chip V1a. The total EoE-error number is 4105 and the numbers of EoE-errors observed at the same locations with chip V1 is 1287. Other 2818 EoE-errors are potential EoE-errors of chip V1. They are treated as non-EoE-error samples in chip V1, but 1926 samples are labeled as EoE-error in prediction of chip V1. Further evaluation on potential EoE-errors in prediction model is a future work.

Finally, EoE-error reduction cost in design phase is not expensive. In general, dummy metal optimization technique is used to planarize wafer surface and several methods are proposed [3], [9], [13], where dummy metal modification does not affect the logic function of the circuit and wire topology. Guided by the proposed prediction model, dummy metal patterns can be modified so that tiles labeled EoE-error are altered to tiles labeled non-EoE-error.

6. Conclusion

In this paper we proposed the first EoE-error prediction method with powerful learning algorithms. It consists of error analysis stage, layout parameter extraction stage, model construction stage and prediction stage. In error analysis and layout parameter extraction stages, we define and extract layout parameters having an impact on EoE phenomenon with analysis of the test chip. In model construction and prediction stages, we use multi-level machine learning method which can predict EoE-error locations accurately. This method makes it possible to prevent yield loss with recognizing EoE-error before manufacturing.

References

[1] G. Nanz and L.E. Camilletti, "Modeling of chemical mechanical

polishing: A review," *IEEE Trans. Semicond. Manuf.*, vol.8, no.4, pp.382–389, Nov. 1995.

- [2] J. Luo and D.A. Dornfeld, "Integrated Modeling of Chemical Mechanical Planarization for Sub-Micron IC Fabrication," Springer-Verlag, Berlin, Germany, 2004.
- [3] S. Sinha, J. Luo, and C. Chiang, "Model based layout pattern dependent metal filling algorithm for improved chip surface uniformity in the copper process," *Proc. IEEE Asia South Pacific Design Automation Conference*, pp.1–6, 2007.
- [4] T.H. Park, *Characterization and Modeling of Pattern Dependencies in Copper Interconnects for Integrated Circuits*, Ph.D. thesis, Massachusetts Institute of Technology, 2002.
- [5] T.E. Gbondo-Tugbawa, *Chip-Scale Modeling of Pattern Dependencies in Copper Chemical Mechanical Polishing Processes*, Ph.D. Thesis, Massachusetts Institute of Technology, 2002.
- [6] C. Chiang and J. Kawa, "Three DFM challenges: Random defects, thickness variation, and printability variation," *IEEE Asia Pacific Conference Circuits and Systems*, pp.1099–1102, 2006.
- [7] M. Cho, D.Z. Pan, H. Xiang, and R. Puri, "Wire density driven global routing for CMP variation and timing," *Proc. IEEE/ACM International Conference on Computer-Aided Design*, pp.487–492, 2006.
- [8] L. He, A.B. Kahng, K. Tam, and J. Xiong, "Design of IC: Interconnects with accurate modeling of CMP," *International Society for Optical Engineering Symposium on Microlithography*, pp.109–119, 2005.
- [9] A.B. Kahng and K. Samadi, "CMP fill synthesis: A survey of recent studies," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol.27, no.1, pp.3–19, 2008.
- [10] J.H. Park, D.W. Park, H.K. Hwang, J.D. Lee, C. Hong, W.S. Han, and J.T. Moon, "Study of over-polishing at the edge of a pattern in selective CMP," *Proc. Electrochemical Society International Symposium on Chemical Mechanical Planarization*, pp.283–289, 2003.
- [11] C. Euvrard, C. Perrot, A. Seignard, and F. Dettoni, "Slurry selectivity influence on STI and POP processes for RMG application," *Proc. International Conference on Planarization/CMP Technology*, pp.1–6, 2012.
- [12] J. Luo, Q. Su, C. Chiang, and J. Kawa, "A layout dependent full-chip copper electroplating topography model," *Proc. IEEE International Conference on Computer Aided Design*, pp.133–140, 2005.
- [13] D. Fukuda, T. Shibuya, N. Idani, and T. Karasawa, "Full-chip CMP simulation system," *Proc. International Conference on Planarization/CMP Technology*, pp.187–194, 2007.
- [14] X. Zhang, L. He, V. Gerousis, L. Song, and C.C. Teng, "Case study and efficient modelling for variational chemical-mechanical planarization," *Circuits, Devices & Systems, IET 2.1*, pp.30–36, 2008.
- [15] J. Luo, Q. Su, C. Chiang, and J. Kawa, "A layout dependent full-chip copper electroplating topography model," *Proc. IEEE/ACM International Conference on Computer-Aided Design*, pp.133–140, 2005.
- [16] D.O. Ouma, *Modeling of chemical mechanical polishing for dielectric planarization*, Ph.D. Thesis, Massachusetts Institute of Technology, 1998.
- [17] M. Orshansky, S.R. Nassif, and D.S. Boning, *Design for manufacturability and statistical design: A constructive approach*, Springer, 2008.
- [18] A.B. Kahng, "Key directions and a roadmap for electrical design for manufacturability," *Proc. European Solid-State Circuits Conference*, pp.83–88, 2007.
- [19] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," *Proc. 14th International Conference on Machine Learning*, pp.179–186, 1997.
- [20] T.M. Therneau and E.J. Atkinson, "An introduction to recursive partitioning using the RPART routine," *Technical Report 61*, Section of Biostatistics, Mayo Clinic, Rochester, 1997.
- [21] L. Breiman, "Random forests," *Machine Learning*, vol.45, no.1, pp.5–32, 2001.

- [22] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol.2, no.2, pp.121–167, 1998.
- [23] K.R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol.12, no.2, pp.181–201, 2001.
- [24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol.20, no.3, pp.273–297, 1995.
- [25] R. Barandela, R.M. Valdivinos, J.S. Sanchez, and F.J. Ferri, "The imbalanced training sample problem: Under or over sampling?," in *Structural, Syntactic, and Statistical Pattern Recognition*, pp.806–814, Springer Berlin Heidelberg, 2004.
- [26] N.V. Chawla, L.O. Bowyer, and W.P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artificial Intelligence Research*, pp.321–357, 2001.
- [27] H. Han, W.Y. Wang, and B.H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," *Advances in Intelligent Computing*, Springer Berlin Heidelberg, pp.878–887, 2005.
- [28] T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts," *ACM SIGKDD Explorations Newsletter* 6.1, pp.40–49, 2004.
- [29] J.V. Hulse, T.M. Khoshgoftaar, and A. Napolitano, "Experimental perspectives on learning from imbalanced data," *Proc. 24th International Conference on Machine Learning*, pp.935–942, 2007.
- [30] "The R Project for Statistical Computing," <http://www.r-project.org/>



Daisuke Fukuda received the B.E. and M.E. degrees in Communications and Computer Engineering from Kyoto University, Japan, in 1999, and 2001, respectively. Since 2001, he has been with the Fujitsu Laboratories Ltd., Kanagawa, Japan. His research interest includes data mining for design and manufacturing, and design for yield/manufacturing.



Kenichi Watanabe received the B.S. and M.S. degrees in electronics engineering from Tokyo University of Agriculture and Technology, Tokyo, Japan, in 1995 and 1997, respectively. In 1997, he joined Advanced Process Integration Department, Electronic Devices Group, Fujitsu Ltd., Kawasaki, Japan, where he has been engaged in the development of advanced LSI and process integration for a multi-layered interconnect technology. His work also includes a development of highly reliable process integration technologies for ULSI devices. He is a member of the

Japan Society of Applied Physics (JSAP).



Naoki Idani received the B.E. and M.E. degrees in Engineering from Osaka University, Japan, in 1990, and 1992, respectively. He joined Fujitsu in 1998 and worked on dielectric, W, and Cu-CMP at Mie Plant. He is currently engaged in CMP development of CMP process.



Yuji Kanazawa received the B.E. degree in Mathematical Engineering and Information Physics, and the M.E. degree in Information Engineering from the University of Tokyo, Japan, in 1988 and 1990, respectively. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1990 and has since been engaged in research and development of operating systems and CAD for digital systems. He is a member of the Information Processing Society of Japan (IPSJ).



Masanori Hashimoto received the B.E., M.E. and Ph.D. degrees in Communications and Computer Engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2001, respectively. Since 2004, he has been an Associate Professor in Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University. His research interest includes computer-aided design for digital integrated circuits, and high speed and low power circuit design.

Dr. Hashimoto served on the technical program committees for international conferences including DAC, ICCAD, ITC, Symposium on VLSI Circuits, ASP-DAC, DATE, ISPD and ICCD. He is a member of IEEE, ACM and IPSJ.