Enhanced Virtual Metrology on Chemical Mechanical Planarization Process using an Integrated Model and Data-Driven Approach

Yuan Di[1], Xiaodong Jia[1], and Jay Lee[1]

[1]*University of Cincinnati, Cincinnati, Ohio, 45221, US*

diyn@mail.uc.edu
jiaxg@mail.uc.edu
jay.lee@uc.edu

ABSTRACT

As an essential process in semiconductor manufacturing, Chemical Mechanical Planarization has been studied in recent decades and the material removal rate has been proved to be a critical performance indicator. Comparing with after-process metrology, virtual metrology shows advantages in production time saving and quick response to the process control. This paper presents an enhanced material removal rate prediction algorithm based on an integrated model and data-driven method. The proposed approach combines the physical mechanism and the influence of nearest neighbors, and extracts relevant features. The features are then input to construct multiple regression models, which are integrated to obtain the final prognosis. This method was evaluated by the PHM 2016 Data Challenge data sets and the result obtained the best mean squared error score among competitors.

1. INTRODUCTION

Chemical Mechanical Planarization (CMP) has been widely applied in the semiconductor industry as a crucial wafer polishing process (Steigerwald, Murarka & Gutmann, 2008). In order to evaluate the performance of the CMP and employ the process control, the material removal rate (MRR) is derived as a performance indicator by measuring the layer thickness of the wafer after the process. In recent decades, achieving the goals of the advanced process control (APC) (Moyne, Castillo & Hurwitz, 2001), there have been two major control scenarios: lot-to-lot (L2L) control which only samples one or several workpieces from the lot for the inspections, and wafer-to-wafer (W2W) control which measures each workpiece to fulfill the control purpose. Compared with the L2L control, the W2W control is obtaining more attentions since it can reduce the variability of the wafer processing and assure the product quality.

However, the W2W control needs gauging every wafer, which will increase the production cycle time significantly. Further, as an after-process measurement, the MRR calculation inevitably delays the control process and will degrade the APC performance.

To overcome the aforementioned issues, the Virtual Metrology (VM) has been presented (Kang et al., 2009) to predict the metrology results by the state of the process. Plenty of approaches have been proposed to implement the VM for various semiconductor processes and the VM on CMP has been further investigated due to its physio-chemical complexity. The mechanism of the CMP process was studied first and various physical based models were derived to explain the relationship between the state variables and the final MRR. Generally, the removing mechanism in CMP can be described in particle-scale (Luo & Dornfeld, 2001), die-scale (Stine et al., 1998) and wafer-scale (Hocheng, Tsai & Chen, 1997). Luo & Dornfeld (2003) reviewed these three models and investigated their characteristics respectively. In addition, the pad condition of CMP tool is essential in all three models and the MRR will decrease dramatically without proper pad conditioning. Tso & Ho (2007) studied factors which would influence the condition of the pad, and Yen & Chen (2010) further investigated the pad conditioning simulation and its relationship with the MRR. The physical model, however, is still difficult to describe the behavior of the equipment in practice since there exist unknown factors and ambient variables. Consequently, coefficients in the model have not been well determined. Therefore, currently VM on CMP is mainly employed by data-driven methods.

By applying data-driven approaches, the VM can be realized as a regression task which predicts the metrology variables through a machine learning model. Both linear and non-linear regression techniques have been developed and applied in the VM field. Partial Least-square Regression (PLR) (Geladi & Kowalski, 1986) and lasso regression have been widely used as linear regression approaches and there have been research efforts applying PLR (Hirai & Kano, 2015) and lasso regression (Park & Kim, 2016) to estimate the performance

in semiconductor manufacturing. A variant of PLR was derived (Hirai, Hazama & Kano, 2014) to predict the MRR. On the other hand, Neural Network (NN) and Support Vector Regression (SVR) are regarded as two major non-linear VM technologies (Su et al., 2008). Lenz & Barak (2013) investigated the dielectric layer thickness prediction for the Chemical Vapor Deposition (CVD) by SVR. A semi-supervised SVR was proposed to handle the label uncertainty in VM (Kang, Kim & Cho, 2016). Due to the complexity of the CMP process and unknown correlations between sensor data, the prediction has also attracted investigations using deep learning techniques. Recently, a deep belief network based approach was presented to deal with the MRR prediction (Wang, Gao & Yan, 2017). In addition to modeling techniques, there also exist two different modeling strategies: global modeling, which uses the entire dataset to construct a single global model, and local modeling, which only uses similar instances in the dataset to construct local models. Jebri et al. (2016) proposed that the local modeling, or Just-In-Time Learning (JITL) strategy, could resolve the bias issue associated with a global model, and obtain higher VM accuracy. Nevertheless, the local model relies heavily on the number of similar historical instances, and the performance will degrade if there are few similar instances available. Thus, the global model is still essential for most applications.

The VM for CMP, however, is still challenging due to high data complexity and process dynamics. There are a large amount of variables measured during the production process so that the dimension of the model inputs is sincerely high. Consequently, the precision of the prediction would be impacted. Also, the individual wafer polishing is not independent. The previous polishing performance and the usage of replaceable components in the machine bring dynamic influence on the MRR prediction. Thus, these unmet needs on advanced feature extraction and selection along with enhanced modeling techniques are looking for further study.

This paper proposes an enhanced VM approach for the prediction of the MRR based on an integrated data-driven model. In feature extraction, it not only considers the physical mechanism of the CMP process but also takes advantages of nearest wafers' polishing results. In modeling, the prediction model is enhanced as well by combining outputs from each individual regression approach. The presented approach was evaluated by a public data set from the Prognostics and Health Management (PHM) Society 2016 Data Challenge and it showed promise since the result earned the top score. The remaining of the paper is organized as follows: Section 2 explains the proposed approach for the MRR prediction. Section 3 goes through an implementation of the method on the data set provided by the competition committee. Following the evaluation results and discussions in Section 4, conclusions are included in Section 5.

2. TECHNICAL APPROACH

This section mainly discusses the proposed integrated prediction approach. An overview of the proposed method is introduced in Section 2.1. Following the overview, the details for each step are then presented.

2.1. Methodology Overview

There are four major steps in the proposed methodology, which is summarized in Figure 1. In step 1, features, which are highly correlated to the MRR, are extracted from the sensor measurements. A subset of the features is then selected in step 2 to reduce the dimension and the redundancy in the feature space. After constructing multiple regression models utilizing the feature subset in training data (step 3), a cross-validation (CV) is employed, and the weights for multi-model integration are calculated. For testing data sets, weighted averaging results will be generated to predict the MRR.

2.2. Feature Extraction

The feature extraction strategy illustrates that various features will be extracted in an exhaustive way from the measurements firstly, and then a subset of them is selected. Nevertheless, there needs a pre-processing step in advance due to the excessive sensor environment in semiconductor field.

Like other tools in semiconductor manufacturing, there are hundreds of built-in sensors in the machine in CMP process. However, not all sensor measurements are able to be utilized in MRR prediction since they bring little impacts to the polishing. Thus, specific sensor variables, which are correlated to the mechanisms of the CMP treatment, need to be selected. Analysis of the physical models can provide useful insights into this physical process and can also reveal underlying physical features for finer modelling.
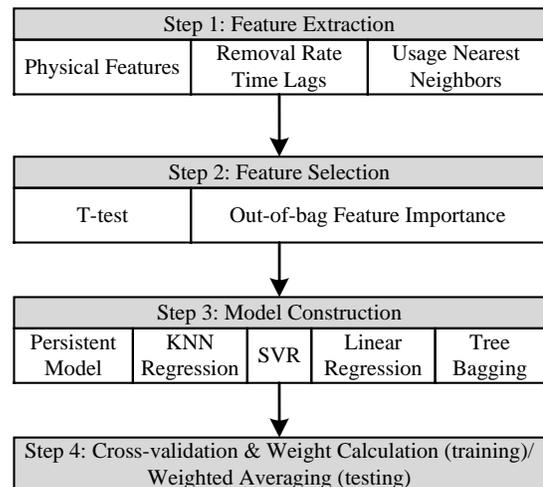

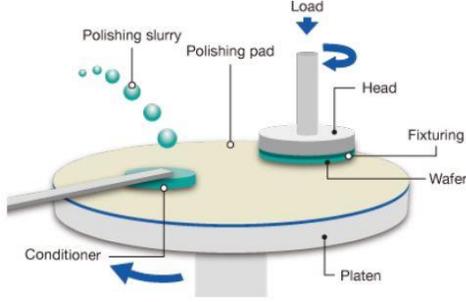Figure 1. Flow chart of the integrated prediction approach

Figure 2. CMP description

The CMP machine tool and the operation are depicted in Figure 2. The tool contains a polishing pad which is attached to a rotating table, a rotating and translating wafer carrier, a rotating and translating dresser and a slurry dispenser (Jebri et al., 2016). The polishing is employed by both the relative rotation between the wafer and the pad, and the slurry containing the abrasives and the chemicals. Also, the pad needs conditioning by the dresser in order to keep its polishing property. During the polishing process, the performance of both the pad and the dresser is degrading and they will be replaced if needed.

The most frequently used physical model for removal rate prediction is an empirical model based on Preston's equation (Luo & Dornfeld, 2001), which is denoted in Eq. (1):

$$ARR = K \cdot P \cdot V \tag{1}$$

Where ARR represents the average removal rate (ARR); P is the mean interface pressure; V is the relative velocity between the wafer and the pad table; K is the Preston constant. All other physical variables are combined into the constant K. However, the ARR will decrease dramatically without conditioning. Thus, the pad condition will be investigated first.

The performance of the pad conditioning can be represented by Dressing Rate (Tso & Ho, 2007), which is denoted as Eq. (2):

$$Dressing\ Rate = K_D \frac{V_D}{RA} \lambda d_0 (\frac{P}{H_p})^{1.5} \tag{2}$$

Where K_D is the dressing coefficient; V_D is the dressing speed; R, A, λ and d_0 are geometric and material related information of the dresser; P is the applied load; H_P is the hardness of the pad. Among these, the usage of the dresser is correlated to the geometric and the material related parameters. Thus, assuming that the dressing speed, the load and the hardness do not variate so much for each run, the Dressing Rate could be simplified as Eq. (3):

$$Dressing\ Rate = K_D U_D \tag{3}$$

Where U_D is the usage of the dresser. Further, the influence of the Dressing Rate could be integrated in the MRR model.

In this study, the particle-scale model (Luo & Dornfeld, 2001) was selected, which can be described as Eq. (4):

$$MRR = f(F_s, P, V, U_P, U_D) \tag{4}$$

Where F_s is the slurry related feature; P is the pressure; V is the relative speed between the pad and the wafer; U_P is the usage of the pad; U_D is the usage of the dresser. On the basis of the physical findings in Eq. (4), various machine learning algorithms are adopted to model physical relationship between the MRR and relevant features. In this study, extracted physical features include the statistics (mean, standard deviation, range and area under curve) of each relevant physical variable.

Besides the physical features, there are dynamic feature representations that are extracted. On one hand, since the CMP treatment is a continuous process, the feature of time lags treats the MRR as time series and it selects the MRR in the most recent past as features. In this paper, we use r_{t-i} denoting the i-th MRR in the recent past. On the other hand, under the same recipe settings, all sensor measurements are expected to be steady regarding to different wafers except the usages of the consuming components such as the pad and the dresser. The wafers which share the same usage state might also own similar MRRs. Therefore, the usage nearest neighbor features are selected based on the proximity of consuming components' usages. The Euclidian distance is implemented to determine the k nearest neighbors (KNN) for each training sample, and the MRR of the selected neighbors are used as input features (Jia, Jin, Buzza, Wang, & Lee, 2016).

2.3. Feature Selection

Even though the number of sensor variables has been reduced by investigating the polishing mechanism, the dimension of the feature space is sincerely high. To reduce the redundancy in the features and mitigate overfitting, the feature selection is employed. In the feature selection step, the Student's t-test (Walpole, Myers, Myers, & Ye, 1993) and the Out-Of-Bag (OOB) feature importance (Breiman, 1996) are mainly used as criteria to select useful features. Student's t-test is also called the significant test for linear regression, it assumes the prediction error of the linear model is independent of the input features and is normally distributed (Walpole et al., 1993). In this scenario, the significant relationship between the input features and the target variable can be evaluated through Student's t-test. However, the assumption of normal distribution is not always valid in real applications. Therefore, we proposed to employ OOB feature importance as a supplemental criterion. The OOB feature importance is a tree bagging based method. It is obtained by permuting the value of each feature across all the training observations and then evaluating how much worse the prediction error becomes after the permutation. In this study, the Student's t test for linear regression (LR) and OOB feature importance

for tree bagging are considered collectively to vote for the important features.

2.4. Regression Model Construction and Integration

The prediction models employed in this study involve the persistent model, KNN regression (Kramer, 2011), LR, tree bagging (Breiman, 1996), and Support Vector Regression (Basak, Pal, & Patranabis, 2007; Smola & Schölkopf, 2004). The persistent model treats the MRR as time series and predicts r_t as $r_t = r_{t-1}$. The KNN regression averages the MRR of k nearest neighbors and these neighbors are determined by the usage neighbor features. On the other hand, the LR, SVR and tree bagging methods employ the selected feature subset in step 2 as inputs to train the models and then predict the MRR for testing observations.

In the final step, the prediction results given by these models are averaged following the model integration strategy in (Parks, Wan, Wiener, & Liu, 2011). The weight for individual predictor is defined by the prediction error obtained in the CV. In this study, the Monte Carlo CV is applied, which means that the training dataset is randomly split into two sets – the training set and the validation set, for each CV test. The Monte Carlo method is preferred since it can evaluate the fluctuation of the prediction error by repeating the CV tests for many times. After the CV is finished, the prediction error for individual predictor is evaluated by Eq. (5):

$$e = \text{mean}(\boldsymbol{\epsilon}) + 3 * \text{std}(\boldsymbol{\epsilon}) \tag{5}$$

Where $\boldsymbol{\epsilon} \in \mathbb{R}^N$ represents the vector of error obtained in the validation tests, $\boldsymbol{N}$ represents the times of CV tests and e denotes the upper bound of the prediction error. After CV, the weight for each model is calculated. In averaging, the model output which obtains smaller prediction error will be assigned higher weight. Therefore, the weight is derived in Eq. (6):

$$\mathbf{w} = \frac{1/\mathbf{e}^3}{\text{sum}(1/\mathbf{e}^3)} \tag{6}$$

Where $\mathbf{e} = [e_{persis}, e_{LR}, e_{SVR}, e_{KNN}, e_{treebagger}]$ represents the vector of prediction error and $\mathbf{w}$ denotes the weighting vector for the individual prediction model.

3. CASE STUDY

In this research, the proposed method was evaluated by a CMP public data set provided by the PHM Data Challenge 2016 (PHM Society, 2016). In this Data Challenge, attendees competed with others predicting the MRR of the wafer through the data collected from a CMP machine. The data were briefly introduced in Section 3.1 and the approach implementation was employed in Section 3.2 respectively.

3.1. Data Description

The Data Challenge provided three different data sets: training data, testing data and validation data. All three data sets included operating parameters and components usage conditions measured from the machine during the CMP process. As shown in Table 1, each row of the table represents a variable of at any given time. Besides the measurements, the metrology information was given as an average MRR for each wafer in the training data sets. Each MRR had its corresponding wafer identification number and the stage type. On the contrary, the MRR was blinded in testing and validation data sets.

There were 1981 MRR records in the training data sets while 424 wafers were required to predict their metrology variables in the testing and validation data sets respectively. The evaluation of proposed algorithms was based on the mean squared error (MSE) accuracy and the understanding of the CMP mechanism.

3.2. Approach Implementation

Based on the provided datasets in the Data Challenge, the method in Figure 1 was implemented on 3 different recipes, which were identified by the Stage and the Chamber. The MRR of the training data is plotted in Figure 3, which indicates that the MRR for these 3 recipes, named Cond1, Cond2 and Cond3, falls in different ranges. Therefore, we proposed to predict the MRR separately. The implementation details based on multiple recipes are summarized in Table 2.

As discussed above, the features for each recipe were extracted and selected separately. It was also noted that most of the wafers went through three chambers in the whole CMP process. Preliminary investigation for each recipe revealed that the processing time of Chamber 4 or 1 was approximately equal to the time of Chamber 5&6 or 2&3 for Cond1 and Cond2 & 3 respectively. Therefore, the physical features from different chambers were extracted separately as listed in Table 3. The extracted features were then selected on the basis of Student's t-test and OOB criteria. Taking the selection results for Cond3, as shown in Figure 4, the features with OOB value and t-stat value beyond the specified threshold were identified as important features. It was also worth noting that the selection result in Figure 4 was obtained from one CV test and the selected features could be different if the training data were split differently. Therefore, in order to exclude the bias introduced by random splits of the training dataset, the feature selection strategy was implemented repeatedly for all CV tests. Only the highly voted features were preserved to predict the testing samples. Regarding the CV, the prediction error e obtained after 20 times CV tests was employed to evaluate the performance of individual predictor, since we found that the prediction error tended to be stable after 20 times tests.

Variable Name	Description
Timestamp	Seconds
Wafer_ID	Number representing ID of wafer
Stage	A or B representing a different type of processing stage
Chamber	Chamber in machine for wafer processing
Usage_Of_Backing_Film	A usage measure of polish-pad backing film
Usage_Of_Dresser	A usage measure of dresser
Usage_Of_Polishing_Table	A usage measure of polishing table
Usage_Of_Dresser_Table	A usage measure of dresser table
Pressurized_Chamber_Pressure	Chamber pressure
Main_Outer_Air_Bag_Pressure	Pressure related to wafer placement
Center_Air_Bag_Pressure	Pressure related to wafer placement
Retrainer_Ring_Pressure	Pressure related to wafer placement
Ripple_Air_Bag_Pressure	Pressure related to wafer placement
Usage_Of_Membrane	A usage measure of polishing membrane
Usage_Of_Pressurized_Sheet	A usage measure of wafer carrier flexible sheet
Slurry_Flow_Line_A	Flow rate of slurry type A
Slurry_Flow_Line_B	Flow rate of slurry type B
Slurry_Flow_Line_C	Flow rate of slurry type C
Wafer_Rotation	Rotation rate of wafer
Stage_Rotation	Rotation rate of stage
Head_Rotation	Rotation rate of head
Dressing_Water_Status	Status of dressing water
Edge_Air_Bag_Pressure	Pressure of bag on edge

Table 1. Measurements description

The CV results for the proposed methods are tabulated in Table 4. The mean and standard deviation of prediction MSE demonstrates that the integrated model tends to outperform individual predictors for all the recipes. It obtains the lowest mean value of MSE compared with other approaches while the standard deviation shows slightly less performance than tree bagging method. By weighted averaging, the integrated model improves the bias issue of the prediction. Also, it demonstrates the robustness of the integrated model in respect of various recipes.

Among the five prediction models adopted, LR is able to give rather satisfactory prediction results if the model is tuned properly. However, finding important features for the LR is not trivial since the performance of the linear model can be seriously affected by the input features. In comparison, the tree bagging method is more robust than the linear model and

it is less affected by the input features. The major disadvantage of tree bagging method involves that it may require more computing resources and the training of the bagged trees can be very slow if the model parameters are not tuned properly. SVR is a very good candidate to overcome the shortcomings of the linear models and the tree bagging method (Murphy, 2012). It is robust and efficient, and more importantly, it can give rather satisfactory results for all the recipes.
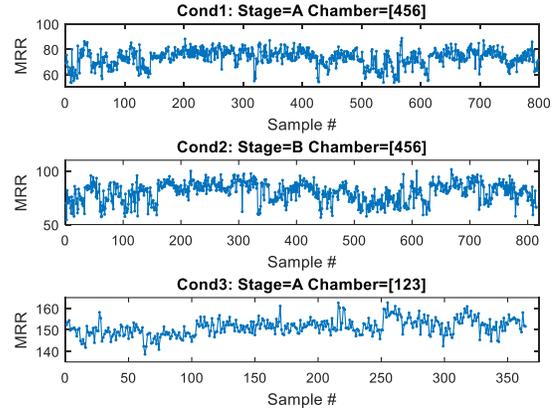

Figure 3. MRR for training data

In contrast, the persistent model and KNN regression are less accurate. Even though the prediction accuracy of these two models strongly indicates that the removal rate has the apparent property of time series, it can be also greatly affected by the physical variables. The comparison of the persistent model and KNN regression also implies that the time series property of removal rate is stronger than the affections of usage variables, since the persistent model tends to outperform the KNN regression for all the recipes.

4. RESULTS AND DISCUSSIONS

In this section, the proposed approach was implemented on the testing data sets. The MRR prediction results were obtained from each individual regression model firstly. And the final prediction was employed by weighted averaging, in which the weights were well determined through the CV on the training data.

Step	Description
Step1:	-Train the prediction models for Cond1; -Extract testing features for Cond1; -Predict the testing MRR for Cond1 following the procedure in Figure 1;
Step 2:	-Train the prediction models for Cond2; -Extract testing features for Cond2; -Predict the testing MRR for Cond2;
Step 3:	-Train the prediction models for Cond3; -Extract testing features for Cond3; -Predict the testing MRR for Cond3;
Step 4:	-Sort the predicted MRR and output the results

Table 2. Pseudo-code for the method implementation

List of features		Feature Id
Removal Rate Time Lags:	11 time lags	1~11
Removal Rate from neighbors:	10 nearest neighbors	12~21
Polishing time:	Chamber = 4 or 1	22
	Chamber = {5,6} or {2,3}	23
Mean, std, decreasing rate of usage variables:	Chamber = 4 or 1	24~29, 36~41, 48~53
	Chamber = {5,6} or {2,3}	30~35, 42~47, 54~59
Mean, std, area under curve of pressure variables	Chamber = 4 or 1	60~65, 72~77, 84~89
	Chamber = {5,6} or {2,3}	66~71, 78~83, 90~95
Mean, std, area under curve of flow rate variables	Chamber = 4 or 1	96~99, 104~107, 112~115
	Chamber = {5,6} or {2,3}	100~103, 108~111, 116~119
Mean of rotating speed variables	Chamber = 4 or 1	120~122
	Chamber = {5,6} or {2,3}	123~125

Table 3. List of extracted features

A comparison between the predicted results and the ground truth for the testing data sets is depicted in Figure 5. The predicted results are denoted as circle while the real MRRs are represented as cross. It is revealed that both the predicted results and the ground truth are close to each other for all 3 recipes, which validate the prediction capability of the proposed approach. Also, compared with Cond3, Cond1 and Cond2 represent higher prediction accuracy.

To quantify the prediction performance, a summary of MSE for both the integrated model and each individual model is shown in Table 5. Since the approaches applied in individual models have been studied for years achieving the VM in semiconductor manufacturing, we can also regard them as baseline methods and benchmark with the integrated model. It is validated that the integrated model also obtains the best prediction performance for the testing data sets. The integration is able to achieve higher prediction accuracy than other conventional approaches. In addition, compared with the MSE from the CV for the training data sets, the prediction capability for each model does not vary significantly. This consistency not only reveals that the proposed method keeps good generalization but also indicates that the Monte Carlo CV can well estimate the performance of the model on the testing data sets which have not been used. Further, the proposed method was compared with that applied deep belief networks (Wang, Gao & Yan, 2017). Even going through deep learning architecture, the model performs less than the proposed approach.

The MSE on testing data sets were also compared and ranked for all the Data Challenge competitors. There were 24 teams

that attended the competition and the MSE results for top 5 teams are described in Table 6. The proposed approach is verified to outperform others in the Data Challenge as well.
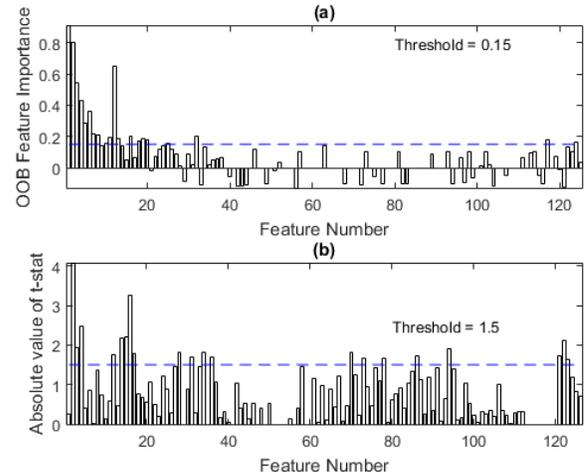

Figure 4. Feature selection for Cond3 on the basis of OOB feature importance (a) and Student's t test (b)

5. CONCLUSIONS

This paper proposed an enhanced VM approach on CMP process. The influence of nearest neighbors, along with the impacts learned by the physical model, were utilized as features which were correlated to the MRR. And the weighted averaging of all outputs from multiple regression models was derived as the final predictions. The algorithm resulted in the most accurate prediction result for the PHM 2016 Data Challenge, which highlighted the effective application for VM. The study advances the VM in semiconductor manufacturing by two contributions: 1 In feature extraction, nearest neighbor related features are added so that it is able to capture detailed dynamic behaviors during the process. 2 In modeling, the weighted averaging takes advantages of all individual regression model hence improves the prediction accuracy.
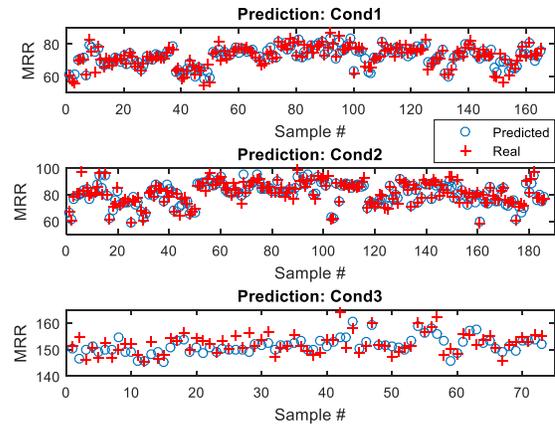

Figure 5. MRR prediction on testing data

	Persistent Model		KNN Regression		SVR		Linear Regression		Tree Bagging		Integrated Model	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Cond1	7.55	1.27	8.78	5.43	5.54	1.30	5.77	1.10	5.65	0.33	5.43	0.84
Cond2	11.4	1.36	15.22	2.03	7.43	1.12	10.95	1.15	8.32	1.17	7.33	0.94
Cond3	5.62	1.06	7.55	0.93	5.05	0.81	5.00	0.92	5.32	0.88	4.91	0.83
Overall	8.78	0.92	11.21	1.29	6.23	0.85	7.76	0.91	6.48	0.68	6.18	0.77

Table 4. MSE for different prediction models after 20 CV tests (Cond1: Stage = A, Chamber = {4,5,6}; Cond2: Stage = B, Chamber = {4,5,6}; Cond3: Stage = A, Chamber = {1,2,3})

Model	MSE
Integrated Model	7.07
Persistent Model	8.23
KNN Regression	9.60
SVR	7.44
Linear Regression	7.32
Tree Bagging	7.22
Deep Belief Networks	7.29

Table 5. MSE from each approach on testing data

Team ranking	MSE
NO. 1	7.07
NO. 2	7.4
NO. 2	7.4
NO. 2	7.4
NO. 5	7.5

Table 6. MSE for top 5 teams on testing data

There are several refinements of the method that can be employed in the future. In feature extraction step, only a summary of statistics is extracted from the sensor measurements, which necessitates advanced techniques mining more accurate patterns in the data and capturing unique characteristics for each variable. Additionally, the trained models need updates periodically including the parameters of the model and the corresponding weights in order to be adapt to the changing settings and recipes in online monitoring and prediction. Also, there exists space enhancing the approach by fusing the physical representation and the data-driven model. Pillai et al. (2016) have proposed a hybrid method predicting the turbine blade failure. Finally, the proposed method is expected to obtain further validations by other data sets.

ACKNOWLEDGEMENT

REFERENCES

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews, 11*(10), 203-224.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, *185*, 1-17.

Hirai, T., Hazama, K., & Kano, M. (2014, October). Application of locally weighted partial least squares to design of semiconductor virtual metrology. In *2014 IEEE Conference on Control Applications (CCA)* (pp. 1771-1776). IEEE.

Hirai, T., & Kano, M. (2015). Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares. *IEEE Transactions on Semiconductor Manufacturing*, *28*(2), 137-144.

Hocheng, H., Tsai, H. Y., & Chen, L. J. (1997). A kinematic analysis of CMP based on velocity model. *Proc. 2nd CMP-VMIC*, 277-280.

Jebri, M. A., Graton, G., El Adel, E. M., Ouladsine, M., & Pinaton, J. (2016, May). Virtual metrology on Chemical Mechanical Planarization process based on Just-In-Time Learning. In *2016 5th International Conference on Systems and Control (ICSC)* (pp. 169-174). IEEE.

Jia, X., Jin, C., Buzza, M., Wang, W., & Lee, J. (2016). Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves. *Renewable Energy, 99*, 1191-1201.

Kang, P., Kim, D., & Cho, S. (2016). Semi-supervised support vector regression based on self-training with label uncertainty: An application to virtual metrology in semiconductor manufacturing. *Expert Systems with Applications*, *51*, 85-106.

Kang, P., Lee, H. J., Cho, S., Kim, D., Park, J., Park, C. K., & Doh, S. (2009). A virtual metrology system for semiconductor manufacturing. *Expert Systems with Applications*, *36*(10), 12554-12561.

Kramer, O. (2011). *Dimensionality reduction by unsupervised k-nearest neighbor regression.* Paper presented at the Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on.

Lenz, B., & Barak, B. (2013, January). Data mining and support vector regression machine learning in

semiconductor manufacturing to improve virtual metrology. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 3447-3456). IEEE.

Luo, J., & Dornfeld, D. A. (2001). Material removal mechanism in chemical mechanical polishing: theory and modeling. *IEEE transactions on semiconductor manufacturing*, *14*(2), 112-133.

Luo, J., & Dornfeld, D. A. (2003). Review of chemical-mechanical planarization modeling for integrated circuit fabrication: From particle scale to die and wafer scales. *Laboratory for Manufacturing and Sustainability*.

Moyne, J., Del Castillo, E., & Hurwitz, A. M. (Eds.). (2001). *Run-to-run control in semiconductor manufacturing* (Vol. 200). Chichester: CRC press.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*: MIT press.

Park, C., & Kim, S. B. (2016). Virtual metrology modeling of time-dependent spectroscopic signals by a fused lasso algorithm. *Journal of Process Control*, *42*, 51-58.

Parks, K., Wan, Y.-H., Wiener, G., & Liu, Y. (2011). Wind energy forecasting: a collaboration of the national center for atmospheric research (NCAR) and xcel energy. *Contract, 303*, 275-3000.

PHM Society 2016 Data Challenge Competition, (2016). [http://www.phmsociety.org/events/conference/phm/16/data-challenge]

Pillai, P., Kaushik, A., Bhavikatti, S., Roy, A., & Kumar, V. (2016). A Hybrid Approach for Fusing Physics and Data for Failure Prediction. *International Journal of Prognostics and Health Management, 7*(025), 1-12

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing, 14*(3), 199-222.

Steigerwald, J. M., Murarka, S. P., & Gutmann, R. J. (2008). *Chemical mechanical planarization of microelectronic materials*. John Wiley & Sons.

Stine, B. E., Ouma, D. O., Divecha, R. R., Boning, D. S., Chung, J. E., Hetherington, D. L., ... & Oh, S. Y. (1998). Rapid characterization and modeling of pattern-dependent variation in chemical-mechanical polishing. *IEEE Transactions on Semiconductor manufacturing*, *11*(1), 129-140.

Su, Y. C., Lin, T. H., Cheng, F. T., & Wu, W. M. (2008). Accuracy and real-time considerations for implementing various virtual metrology algorithms. *IEEE Transactions on Semiconductor Manufacturing*, *21*(3), 426-434.

Tso, P. L., & Ho, S. Y. (2007). Factors influencing the dressing rate of chemical mechanical polishing pad conditioning. *The International Journal of Advanced Manufacturing Technology*, *33*(7-8), 720-724.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). *Probability and statistics for engineers and scientists* (Vol. 5): Macmillan New York.

Wang, P., Gao, R. X., & Yan, R. (2017). A deep learning-based approach to material removal rate prediction in polishing. *CIRP Annals-Manufacturing Technology*.

Yeh, H. M., & Chen, K. S. (2010). Development of a pad conditioning simulation module with a diamond dresser for CMP applications. *The International Journal of Advanced Manufacturing Technology, 50(1-4)*, 1-12.

BIOGRAPHIES

Yuan Di is currently a Ph.D researcher at Mechanical Engineering department at University of Cincinnati. He received his BS and MS from East China University of Science and Technology in 2010 and 2013 respectively. His research interests include prognostics and health management, data mining and machine learning.

Xiaodong Jia is currently a Ph.D researcher at Mechanical Engineering department at University of Cincinnati. He received his BS and MS from Central South University in 2008 and Shanghai Jiaotong Unviersity in 2014 respectively. His research interests include prognostics and health management, data mining and machine learning.

Jay Lee, PHD is Ohio Eminent Scholar, L.W. Scott Alter Chair Professor, and Distinguished University Professor at the University of Cincinnati and is founding director of National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) on Intelligent Maintenance Systems (IMS, www.imscenter.net) which is a multi-campus NSF Industry/University Cooperative Research Center which consists of the University of Cincinnati (lead institution), the University of Michigan, Missouri Univ. of S&T, and the University of Texas-Austin. Since its inception in 2001, the Center has been supported by over 80 global companies. His current research focuses on Cyber Physical Systems and industrial Big Data analytics.

He has authored/co-authored numerous highly influential articles and technical papers in the areas of machinery monitoring and prognostics, E-manufacturing, and intelligent maintenance systems. He has over 20 patents and trademarks. He is a Fellow of ASME, SME, as well as a founding fellow of International Society of Engineering Asset Management (ISEAM) and has received a number of awards including the Alexander Schwarzkopf Prize for Technical Innovation by NSF in 2014. He is also the catalyst in a new start-up company Predictronics through a NSF I-Corps Award in 2012.