

# Assessment of Physics-Based and Data-Driven Models for Material Removal Rate Prediction in Chemical Mechanical Polishing

Xuan Li<sup>1</sup>, Cheng Wang<sup>1</sup>, Li Zhang<sup>1,\*</sup>, Xinnong Mo<sup>1</sup>, Dewen Zhao<sup>2</sup> and Changkun Li<sup>2</sup>

<sup>1</sup>School of Software, Tsinghua University, Beijing, 100084, China

<sup>2</sup>Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, China

\*Corresponding author

**Abstract**—Material removal rate (MRR) during the chemical mechanical polishing (CMP) process affects the control of product quality. Complexity of various parameters makes it challenging to predict MRR accurately. We addressed this challenge by integrating physics-based modeling with data-driven statistics. First, we analyzed the raw data using data profiling techniques. Then, we extracted features from a physical point of view. Finally, we constructed two Random Forest models respectively based on the feature selection results via the Generic Algorithm. Experiments show that the features we extracted embody key information of each process. The final score predicted by this approach ranked in the second place in a Data Challenge Competition.

**Keywords**—process control; material removal rate; chemical mechanical polishing

## I. INTRODUCTION

With the development of manufacturing technology, effective process control plays an increasingly essential role in today's manufacturing system. It can ensure the quality of product. Chemical Mechanical Polishing (CMP) is a highly effective technique for planarizing wafer surfaces. During the CMP process, the surface of the wafer is polished both by mechanical abrasion and by chemical erosion to achieve a global planarization [1]. The material removal rate (MRR) affects the polished product quality. Due to the complexity of various parameters, predicting MRR accurately remains at an empirical stage.

This paper approach this challenge using assessment of physics-based and data-driven models. Combined with the prior CMP system knowledge, we first analyzed the raw data using data profiling. Then multiple machine learning techniques were used to build the prediction model, and finally, the Random Forest method was chosen to make the best prediction. To detail the analysis results and techniques, this paper is organized as follows: Section 2 provides an overview of the problem and the raw data. Section 3 presents the techniques used to analyze the raw data and conclusions drew from the analysis results. Section 4 details how the raw data would be best represented by extracting features and condition indicators. Section 5 explains the model selection strategy and techniques used to build the final model. It also analyzes the

effect of each parameter based on experiment results. Lastly, conclusions are discussed in Section 6.

## II. PROBLEM STATEMENT AND DATASETS

The challenge is to predict removal rate of material during the CMP process. This section first of all gives the basic information of the system under investigation. Then, the raw data we used are described in detail. Finally, the scoring metric is defined to evaluate the prediction performance of the test instances.

### A. Background

Chemical Mechanical Polishing (CMP) is an effective method in the semiconductor industry to planarize material from the surface of the wafer [2]. During the CMP process, the surface of the wafer is polished both by mechanical abrasion and by chemical erosion to achieve a global planarization. Figure 1 shows a schematic view of the CMP system. It is composed of a rotating table, a replaceable polishing pad, a rotating wafer carrier, a slurry dispenser and a translating dresser. A wafer is held upside down in a wafer carrier and pressed into contact with a slurry film flowing over a polishing pad. When the process starts, both the wafer carrier and polishing pad are rotated in the same direction. The slurry, which is composed of various chemicals and abrasive materials, are dispensed onto the polishing pad, thereby removing material from the surface of the wafer. After finishing polishing, the polishing pad may be conditioned by the dresser to prepare for the future polishing operation. Over time, the polishing pad and the dresser degrade and therefore must be replaced with new ones.

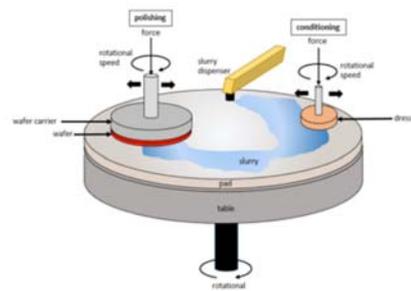


FIGURE 1. SCHEMATIC DIAGRAM OF CMP SYSTEM

TABLE I. TIME SERIES DATA DESCRIPTION

Symbol	Column Name
X1 - X6	MACHINE ID   MACHINE DATA   TIMESTAMP WAFER ID   STAGE   CHAMBER
X7 - X10, X16-X17	USAGE OF - BACKING FILM   DRESSER   POLISHING TABLE   DERSER TABLE   MEMBRANE   PRESSURIZED SHEET
X11 - X15, X25	PRESSURE - PRESSURIZED CHAMBER   MAIN OUTER AIR BAG   CENTER AIR BAG   RETAINER RING   RIPPLE AIR BAG   EDGE AIR BAG
X18 - X20	SLURRY FLOW LINE - A   B   C
X21 - X23	ROTATION - WAFER   STAGE   HEAD
X24	DRESSING WATER STATUS

As we can see from the CMP process, various parameters affect MRR, such as material properties of the wafer, the pressure and rotation of some components, the slurry flow line and the usage of table and sheet. The material removal mechanism has been the subject of research all long. Most of the earlier research has concentrated on predicting MRR based on physical view [3]. Due to the complexity of the physics during the CMP process, only a subset of process parameters was considered from aspects of chemical reaction and mechanical abrasion [4-10]. Data-driven approaches make use of large amount of data to discover underlying patterns. Complex relationship between MRR and various process parameters can be discovered using machine learning techniques. However, effective feature extraction often requires physical knowledge. This paper combines this two together to get a better model for MRR prediction during the CMP process.

### B. Data Description

We used the 2016 PHM Data Challenge [11] Competition Datasets as the raw data. The data provided can be split into training sets, test sets and validation sets. Each type of datasets included time series data and measurement data. Time series data was collected during various runs of the CMP system for the specified wafers. As is shown in Table 1, the raw data include run properties, pressure, usage, rotation, slurry flow line and status. Measurement data represented the average rate of removal, which was determined from the measurements of the thickness of the material before and after the CMP polishing. Only the training sets had the value of removal rate for each wafer identifier and stage. The test sets and validation sets were all missing the removal rate value and need to be predicted.

### C. Evaluation

The performance of the prediction model can be evaluated using mean square error (MSE). MSE was calculated by the following equation, where  $N$  is the number of runs,  $T_n$  is the actual average removal rate and  $P_n$  is the prediction result:

$$MSE = \frac{\sum_{n=1}^N (P_n - T_n)^2}{N} \quad (1)$$

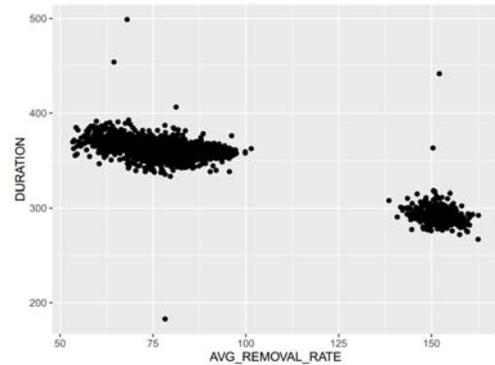


FIGURE II. CLUSTERING RESULT OF THE TRAINING SETS

## III. PHYSICAL-BASED DATA PROFILING

To get a better understanding of the raw data, data profiling is an effective tool. Data profiling is the set of activities and processes to determine the metadata about a given dataset. It can be used to collect statistics and information about the data. Traditional data profiling can be classified to single column profiling, multiple columns profiling and dependencies profiling. Single column profiling techniques can be used to get value distributions and patterns while multiple columns profiling methods achieve more complex information that reflects the relationships between different columns.

Combined with the prior CMP system knowledge, we analyzed the training sets by clustering, outlier detection and visualization. We drew some important conclusions, which provided an important basis for the following steps:

### A. Classification

In conjunction with the average removal rate, the training sets can be classified into two sets: the first set contained runs going through chamber 4-5-6 and the average removal rate was between 50 and 100; the second set contained runs going through chamber 1-2-3 and the average removal rate was around 100. Figure 2 shows the clustering result of the training sets on the average removal rate and the duration by calculating the difference between the start timestamp and the end timestamp. Based on the prior system knowledge, we inferred that there existed two different types of the polishing process: rough polishing and fine polishing. Each polishing run in the same class had the similar average removal rate.

In addition, we gave metrics on data quality by data profiling and found that the monitoring data of the second class (chamber 1-2-3) was missing seriously. Different prediction models should be constructed due to different data quality.

### B. Division

By means of the visualization technique, we divided the polishing process into four phases: the preparation phase, the main polishing phase, the ending phase and the post-CMP

cleaning phase. The former three phases were all in the first chamber and the division among them was based on the observation value of pressure, slurry and rotation columns. The fourth phase went through the second and third chamber.

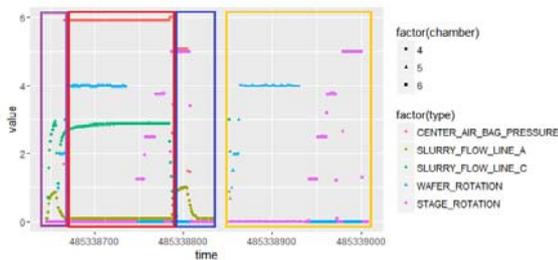


FIGURE III. AN EXAMPLE OF DIVISION

Figure 3 shows an example of division. The part in the purple box represents the preparation phase. It begins from the start time of each run and ends when the second phase starts. The phase in the red box is the main polishing phase and has the stable value on pressure and slurry. The third phase begins when the slurry A flow increases which is signed with the blue box. The fourth phase in the yellow box is called the post-CMP cleaning phase. It goes through the remaining two chambers. We found that all data in this phase was a copy of the polishing data that generated by the former wafer.

As the phases were divided according to the monitoring data, the division method described below only applied to the first type runs. Only the post-CMP cleaning phase can be recognized based on the chamber information for the first type runs. The former three phases were merged into one polishing phase roughly. That is, each polishing run in the first class was divided into two phases: the polishing phase and the post-CMP cleaning phase.

Each phase imposed different effects on the wafer. For the first type, the main polishing phase was most decisive to the average removal rate. Similarly, the polishing phase was the most decisive one for the second type. The preparation phase and the end phase may have some effects on the wafer due to the slurry was dispensed and the polishing pad was conditioned respectively. However, the post-CMP cleaning phase of both

types had little effect on the removal rate as it had no operation on the polishing pad.

### C. Continus Polishing Process

In addition to the visualization of a single run, we analyzed all 2929 runs and found that there existed 1267 Continuous Polishing Processes (CPPs). CPPs were determined by time interval. That is time interval between different CPPs should be larger than 500. Each CPP consisted of several runs and had the following two important characteristics:

- Runs in the same CPP had consistent tool settings, similar health states of various components and similar removal rates.
- The average removal rate of the first run was relatively high.

Figure 4 shows an example of a CPP. It is obvious that this CPP has similar value variety and the first run's removal rate is the greatest. All of these are corresponded with the above characteristics.

By combining with the physical model, we can speculate a reason for the second characteristic: Before a CPP started, the polishing pad may be conditioned by the offline dressing. The polishing pad's condition was the best, so that the first run had a higher removal rate.

### D. Summary of Data Profiling

We used data profiling techniques to analyze the raw data and drew the following conclusions:

The polishing runs can be classified into two types according to the combination of the chambers. Prediction models should be constructed respectively.

The polishing process can be divided into different phases. Each phase had different effect on the wafer and there existed a decisive phase.

Runs in the same CPP were similar and the average removal rate of the first run was relatively high due to the offline dressing.

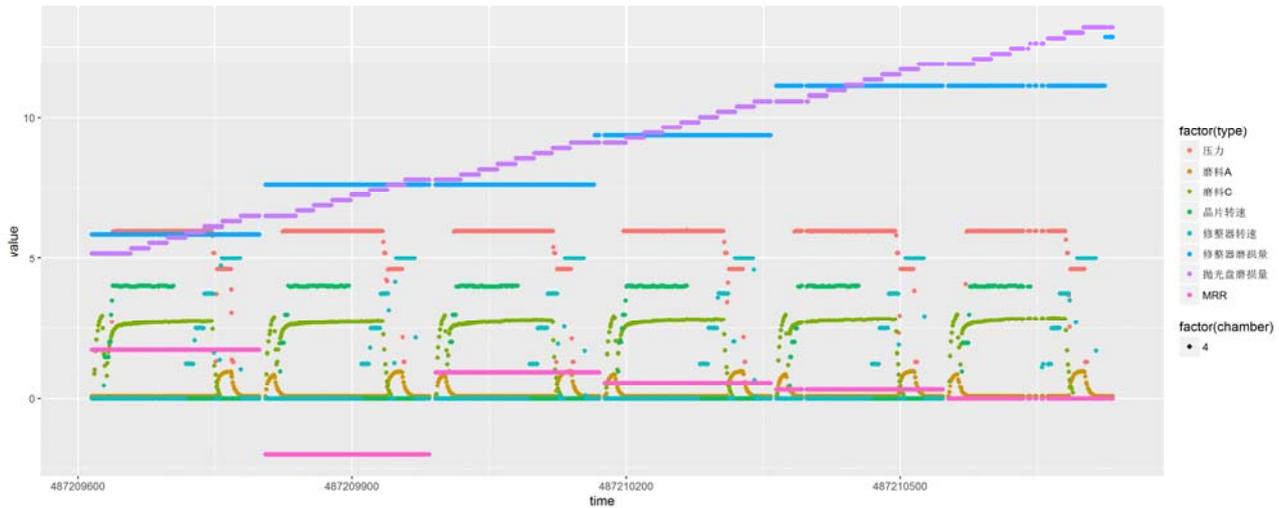


FIGURE IV. EXAMPLE OF A TWO-COLUMN FIGURE CAPTION: (A) THIS IS THE FORMAT FOR REFERENCING PARTS OF A FIGURE.

#### IV. PHYSICAL-BASED DATA REPRESENTATION

The direct method of MRR prediction is to predict at each time step independently and finally aggregated all to get the average removal rate. But this method lacked timestamp-based training data. An alternative way is to predict based on process. It also brought a challenge as to how the time series data would be best represented. This section details how to extract features for the different type runs respectively.

##### A. Physical Evidence

To extract more useful features, we speculated on all possible input of the prediction model from a physical point of view. Conventionally, the Preston equation [9] is widely adopted to predict the material removal rate. It is given by:

$$MRR = K_p PV \quad (2)$$

where MRR is the material removal rate, P is the down force, V is the relative velocity and  $K_p$  is the coefficient, which represents the combined effects of other parameters such as the wafer, polishing pad and slurry. The Preston equation can only provide a linear prediction and many revised versions were proposed, e.g. Wrschka et.al expressed the nonlinear dependency [10].

We inferred the average removal rate was calculated by:

$$\overline{MRR} = \frac{\int_0^T MRR_t dt}{T} \quad (3)$$

where  $\overline{MRR}$  is the average removal rate and  $T$  is the effective polishing duration.

Therefore, the average removal rate of the CMP process was related to the pressure, relative velocity and effective duration. Meanwhile, the material properties of the wafer,

condition of the polishing pad and slurry also had a significant influence on the final polishing results. Features should be extracted to describe the CMP process based on the above physical evidence.

##### B. Feature Extraction

As the prediction model was based on process, it was impossible to directly use the time series data as the input due to variation in quantity. Features representing the key information should be extracted from the raw data. Because of the difference in data quality, we extracted features for the two types of the CMP runs respectively.

For the first type (chamber 4-5-6), the monitoring data was complete. We extracted overall 47 features from the original 25-columns dataset.

The direct extraction methods included: computing the statistics (e.g. mean, media) of the pressure, speed, slurry variables in the main-polishing phase using the wafer identifiers as the material properties of the wafer using initial value to evaluate health states.

The indirect extraction methods included: calculating the effective polishing duration calculating the integration of pressure and speed on the effective polishing duration identifying the order of CPPs based on time intervals identifying the first run of each CPP.

For the second type (chamber 1-2-3), the monitoring data was missing. Therefore, only 12 features were extracted by the same direct and indirect extraction methods: using the wafer identifiers and initial value using initial value to evaluate health states calculating the effective polishing duration identifying the order of CPPs based on time intervals.

#### V. MODEL CONSTRUCTION

Once finishing the feature extraction, the next step was to construct a model to make predictions for the CMP processes. The model should summarize patterns from the underlying data

and correlate features to the behavior of the CMP system. This section details how to construct the final prediction model.

#### A. Model Selection Strategy

Machine Learning and Statistics provide a large number of algorithms to build the model. These algorithms automatically learn to recognize complex patterns and generate the most likely outcome for a new case. Some of the commonly used approaches include: Clustering, Decision Tree, Neural Networks, Support Vector Machines (SVMs), K Nearest Neighbor (KNN) and Bayesian Networks. By removing redundant or irrelevant features, feature selection methods can be used to reduce the complexity and enhance the generalization of models. Also, since the training data is usually limited, various techniques have been used to train a group of models or pick the best model. Some of the widely used techniques include: Ensemble Learning and Cross Validation.

Based on the Genetic Algorithm [12], a flow has been designed to select the best prediction model for the CMP system: At the beginning, all extracted features were taken as the input. Then we evaluated the fitness of each individual based on the MSE score of each candidate machine learning model. In the end, the best model with the optimal feature subset was chosen.

TABLE II. AVERAGE MSE SCORE OF ENSEMBLE MODELS

<i>Model</i>	<i>Self test MSE</i>	<i>Real test MSE</i>
Random Forest	7.4	7.6
Ensemble Neural Network	7.5	8.2
Random Forest (REVISED)	7.2	7.4

#### B. Final Model

We chose Decision Tree, Neural Network, KNN and SVMs as the candidate models. Among these four models, the Neural Network performed best. To improve the accuracy and robustness of individual models, we used the Ensemble Learning methods such as boosting and bagging to get the ensemble model for neural network [13]. Inspired by the improvement, we tried the Random Forest [14] and found that it had a better performance. Random Forest is an ensemble model that consists of many decision trees.

In addition, the revision method based on the CPPs was used to improve the accuracy further according to the second conclusion in Section 3:

$$P_n = O_n - \overline{E}_i \quad (4)$$

$$E_i = \sum_{m=1}^{M_i} O_m - T_m \quad (5)$$

where  $M$  is the number of training runs in  $i$ -th CPP,  $O_m$  is the output prediction result of the training CMP run,  $T_m$  is the

actual removal rate and  $O_n$  is the output prediction result of the test CMP run.

Table 2 shows the comparison results of the average MSE score on self-test sets and real test sets. It shows that the evaluation based on self-test sets is effective and the revision method based on the CPPs also improves the accuracy.

Considering the prediction performance on both self-test sets and real test sets, we finally chose the Random Forest as our modeling method. As the extracted features were different, we constructed two Random Forest models for the two type runs respectively. The results of feature selection can be seen in Table 3.

The final score predicted by this approach ranked in the second place in the 2016 PHM Data Challenge Competition.

TABLE III. FEATURE SELECTION RESULTS OF FINAL MODELS

<i>Model for rough polishing</i>	<i>Model for fine polishing</i>
wafer identifier, stage, usage of dresser, usage of dresser table, usage of polishing table, usage of membrane, effective polishing duration, integration of pressure, first run identifier, start time, CPP identifier	wafer identifier, usage of dresser, usage of dresser table, usage of polishing table, usage of membrane, CPP identifier

## VI. CONCLUSION

This paper introduces a physics-based and data-driven modeling solution for addressing the MRR prediction problem in CMP System. Essential steps included data profiling, feature extraction and model construction. Physical knowledge provided the solid foundation for better understanding the raw data and data profiling was an effective tool to collect statistics and information. Based on the data analysis results, features were extracted to best represent the process. Meanwhile, physical evidence was also taken into consideration and achieved the best extraction results. Finally, a large number of machine learning methods were used to construct the model. The model selection strategy and a revision method were proposed to select the best model and improved the final model further.

## ACKNOWLEDGMENT

This work is supported by one project (No. 2017YFC0804307) and two projects of National Science and Technology Supporting Plan (No. 2015BAF21B00 and No. 2015BAF32B01).

## REFERENCES

- [1] Runnels, Scott R, and L. M. Eyman. "Tribology analysis of chemical-mechanical polishing." *Journal of the Electrochemical Society* 141.6(1994):1698-1701.
- [2] Steigerwald, Joseph M., S. P. Murarka, and R. J. Gutmann. *Chemical Mechanical Planarization of Microelectronic Materials*. 2007.
- [3] Luo, Jianfeng, and D. A. Dornfeld. "Material removal mechanism in chemical mechanical polishing: theory and modeling." *Semiconductor Manufacturing IEEE Transactions on* 14.2(2001):112-133.

- [4] Tseng, Wei Tsu. "Re-examination of Pressure and Speed Dependences of Removal Rate during Chemical-Mechanical Polishing Processes." *Journal of the Electrochemical Society* 144.2(1997):L15-L17.
- [5] Wang, Y. G., L. C. Zhang, and A. Biddut. "Chemical effect on the material removal rate in the CMP of silicon wafers." *Advanced Materials Research* 126-128.3-4(2011):511-514.
- [6] Huang, Pay Yau. "A material removal rate model considering interfacial micro-contact wear behavior for chemical mechanical polishing." *Journal of Tribology* 127.1(2005):190-197.
- [7] Park, K. H., et al. "Effects of pad properties on material removal in chemical mechanical polishing." *Journal of Materials Processing Technology* 187-188.4(2007):73-76.
- [8] Zhao, Yongwu, and L. Chang. "A micro-contact and wear model for chemical-mechanical polishing of silicon wafers." *Wear* 252.3-4(2002):220-226.
- [9] Preston, F. W. "The Theory and Design of Plate Glass Polishing Machines." *J.soc.glass Tech* 11(1927).
- [10] Wrschka, P., et al. "Polishing Parameter Dependencies and Surface Oxidation of Chemical Mechanical Polishing of Al Thin Films." *Journal of the Electrochemical Society* 146.7(1999):2689-2696.
- [11] Data Challenge PHM (2016) Society of Prognostic and Health Management.<http://www.phmsociety.org/events/conference/phm/16/data-challenge>
- [12] Houck, Christopher R., J. A. Joines, and M. G. Kay. "A Genetic Algorithm for Function Optimization: A MATLAB implementation." *Ncsu* 22(2008).
- [13] Zhou, Zhi Hua, J. Wu, and W. Tang. "Ensembling Neural Networks: Many Could Be Better Than All." *ARTIFICIAL INTELLIGENCE* 2002.
- [14] Liaw, Andy, and M. Wiener. "Classification and Regression by randomForest." *R News* 23.23(2002).