Mathematical Problems of Computer Science 53, 39--48, 2020

UDC 519.6

Application of Machine Learning-Based Electrochemical Deposition Models to CMP Modeling

Ruben G. Ghulghazaryan[1], Davit G. Piliposyan[1], Misak T. Shoyan[1] and Hayk V. Nersisyan[2]

[1] Mentor, a Siemens Business, Yerevan, Armenia
[2] American University of Armenia, Yerevan, Armenia
e-mail: ruben_ghulghazaryan@mentor.com

Abstract

Chemical mechanical polishing/planarization (CMP) is the primary process used for modern integrated circuits (IC) manufacturing. Modeling of the post-CMP surface profile is critical for detecting planarity hotspots prior to manufacturing and avoiding fatal failures of chips. Electrochemical deposition (ECD) is a key process for the void-free filling of interconnection wires and vias in modern chips. Large surface topography variations generated after ECD affect the post-CMP surface profile. In this paper, several machine learning approaches are used to model surface profiles after ECD that are used as input to CMP models. Different combinations of deep neural networks, long-short-term-memory (LSTM) recurrent networks, convolutional neural networks and XGBoost algorithms are investigated and compared. The model based on the XGBoost library showed superior performance and accuracy.

Keywords: CMP, ECD, Machine learning, Neural networks, LSTM, XGBoost.

1. Introduction

Chemical mechanical planarization/polishing (CMP) is a crucial technology used during the manufacturing of multi-level interconnections of semiconductor chips and electronic devices [1, 2]. Many process steps used for chip manufacturing require planar surfaces for correct pattern printing during lithography to generate structures for the next layer. CMP is the key process used in the chip production flow for achieving surface planarity required for depth of focus (DOF), lithography requirements, further etch steps for construction of multi-level interconnection wires, high-k replacement metal gate transistors, 3D stacked chips, 3D NAND memory cells, etc.
During CMP, a polishing pad is pressed against the rotating wafer. A chemical slurry containing abrasive particles and chemical agents is added between the pad and the wafer. Combined mechanical and chemical interactions that take place simultaneously at the wafer pad contact area

result in material removal from the wafer surface, leading to wafer surface profile planarization [1, 2]. The post-CMP surface profile depends on the applied pressure, slurry chemistry, and the pattern printed on the wafer. An inappropriate combination of these parameters can lead to non-uniform material removal, which can lead to defects like dishing of metal lines and erosion of dielectrics, causing hotspots such as open contacts, circuit shorting, timing delays, and RC violations.

In modern chips, copper interconnection lines are built using dual damascene processes. During the copper dual damascene process, both wires and vias are filled with copper simultaneously. The electrochemical deposition (ECD) process is then used to fill the wires and vias with copper [3]. After ECD, copper is not only deposited over trenches, but also between them. CMP is used after ECD to remove the excess copper over oxides and isolate wires to avoid fatal electrical shorts. Large surface topography variation created after ECD introduces additional challenges for Cu CMP (Fig. 1), making modeling of the ECD surface profile critical for high-accuracy CMP modeling.

Modeling of the post-CMP surface profile enables the detection of possible CMP hotspots prior to manufacturing [4, 5]. On the other hand, modeling of deposition surfaces prior to CMP is crucial for correct CMP simulation, since the post-deposition surface profile is used as input for CMP simulation, and may affect the final surface profile. Even with advanced deposition processes, the pre-CMP profile on a patterned wafer is non-uniform and affects the surface planarity after CMP. A fully connected neural network (NN)-based full-chip deposition model for predicting post-deposition surface profile for shallow trenches isolation (STI) and inter-level dielectric (ILD) deposition steps has been developed [6]. However, the post-ECD surface profile shows even complicated topography variation, depending on the underlying layer pattern geometries [3].

To predict these complicated topography variations, the use of more advanced machine learning (ML) modeling techniques should be investigated. ML is a sub-area of artificial intelligence (AI) that enables computer systems to recognize patterns using algorithms and data sets, and use that information to develop solutions. ML is used in a wide variety of applications, including fraud detection, computer vision, transportation, bioinformatics, etc. [7, 8]. The ML algorithms make predictions based on initially gathered (training) data without being explicitly programmed for that task. Typical ML applications use neural networks (NNs), which consist of layers of artificial "neuron" data processing elements with an input layer, several hidden processing layers, and an output layer, with weighted connections between the layers. Training an NN means finding values for the weights of connections (based on minimization of a cost function) that best fit the training and validation data. There are multiple types of NN architectures, including deep neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) [7, 8].

A gradient boosted decision tree (GBDT) is another class of prediction technique widely used in ML modeling [9]. A GBDT model consists of an ensemble of decision trees the predictions of which are combined and optimized based on the prediction error on training data to reduce the error of the overall model. Extreme gradient boosting (XGBoost) is one of the implementations of GBDT algorithms, popular for its ability to learn non-linear decision relations and used in both industry and academia alike [9].

In this paper, we apply ML methods such as CNN, long short-term memory (LSTM), RNN, and XGBoost to model surface profiles after ECD for CMP modeling.

2. Electrochemical Deposition Process

Dual damascene is a state-of-the-art technique containing copper deposition and CMP steps used in the manufacturing of back-end-of-line (BEOL) wiring. To date, ECD is the most suitable copper deposition method for the copper deposition step. Using ECD, the vias and trenches can be filled simultaneously. To avoid air gaps, voids or defects during trench filling, chemical additives are used during the ECD process. These additives strongly affect the local deposition rate, leading to superfill, where the copper deposition rate at the trench bottom is higher than on trench sidewalls. These chemical additives are mainly categorized into two types: suppressors and accelerators. Accelerators consist of surfactant molecules that are absorbed on the surface and have the feature of forcing out more weakly-absorbed additives. For a given voltage, they enhance the current value, resulting in a higher copper plating rate. Suppressors consist of polymer-like molecules. In contrast to accelerators, they reduce current value and passivate the copper plating rate. Levelers are a subordinate class of suppressor additives that polarize the areas with high current densities and even out current distribution. They are commonly used to reduce overburden thickness or bumps. During the ECD process, the trench bottom surface area shrinks, causing the accelerator molecules to displace more weakly-absorbed suppressor molecules. This leads to a higher concentration of accelerator molecules at the bottom of the trenches than on trench sidewalls and non-trench areas, resulting in the desirable void-free superfill effect. The side effect of this process is that the sites with narrow trenches fill up faster and continue to grow at a higher rate than sites with wider trenches. This causes a formation of bumps. When narrow trenches are located close to each other, these bumps accumulate into a single large bump. These bumps lead to a large profile surface variation (Fig. 1) and require excessive polishing, introducing additional challenges for the CMP. The effect of bump formation can be weakened by adding levelers; however, in general, bump formation cannot be avoided.

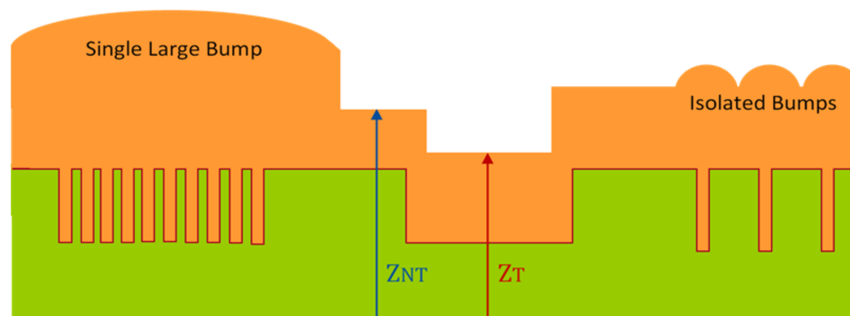

Fig. 1. Schematic plot of post-ECD surface profile.

The ECD process is complicated and continuously evolving. It is hard to follow all the effects of chemistry and physics in physics-based models. Meanwhile, even minimalistic chemistry and physics-based models should keep track of deposition rate changes during ECD to properly model the surface profile. This tracking requires a huge number of computations for solving a large system of differential equations needed to track surface profile evolution during ECD. The runtime of chemistry- and physics-based ECD models may take several minutes up to hours for a layer, depending on the chip size.

To model the ECD process, the design is first divided into fixed-size tiles and for each tile, average values of geometrical characteristics like width, space, pattern density, and perimeter are

extracted. Using these characteristics, the surface dynamic transformations are modeled using the "effective trench" (ET) approximation for each tile.

In ET approximation, the tile surface profile is characterized by previously extracted characteristics together with parameters describing the height of the profile surface material inside (ZT) and outside (ZNT) of the trenches. All these parameters are determined dynamically and passed to the ECD model for simulation. During the simulation, ZT, ZNT, and the geometric data are updated for each tile. After the ECD simulation, the surface profile data for each tile are used as input for the CMP model.

3. Training Data Generation for Machine Learning

The ML training data should include topography height and geometry data of patterns collected from design layouts. When collecting geometry data for ML model input, both specially designed test chips and production design layouts were used to provide sufficient coverage of possible geometry patterns supported by the technology. For training an ML model, large amounts of data are required for separation into training, validation, and test data. We carried out a series of experiments using Calibre™ CMP ModelBuilder and CMPAnalyzer tools. First, we created physics-based models using data collected at the factories. We then used these physics-based models to generate the training, validation, and test data for ML model building.

To generate input and output data, we extracted six grids with geometry characteristics of patterns (width, space, pattern density, perimeter, etc.) for input, and surface profile grids ZT and ZNT for output, from different test chip and production design layouts.

4. Results and Discussion

We experimented with different architectures of DNN, CNN, RNNs and GBDT methods and their combinations for modeling post-ECD surface profiles (Fig. 2). In this section, we review the architectures and models that provided the best combination of running time and accuracy.

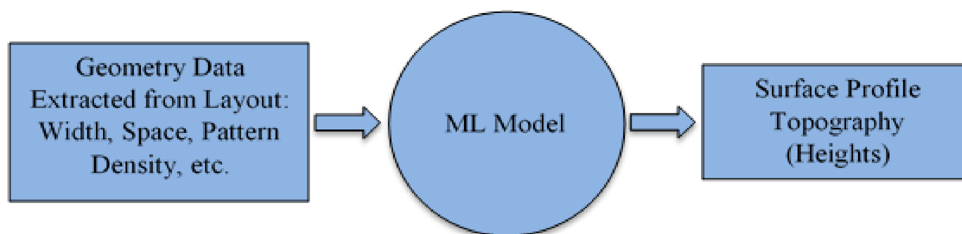

Fig. 2. NN input and output data.

5. RNN-Based Modeling

For the first model, we considered a model based on the combination of DNN and LSTM RNN networks. The DNN part consists of a feed-forward neural network with three hidden layers and 10 neurons per layer. Six-dimensional geometry data (width, space, pattern density, etc.) are used

as input to the network. Two-dimensional output is used for modeling post-ECD profiles. A sigmoid activation function is used for all layers. The model is trained using a stochastic gradient descent optimization method for reverse Huber (berHu) loss function [1], which has the form

$$B(x) = \begin{cases} |x| & |x| \leq c, \\ \dfrac{x^2 + c^2}{2c} & |x| > c. \end{cases}$$

Here, c is a scalar related to 10% of the maximum error on the batch. This method has been proved to have several advantages over other standard loss functions usually used for regression models, such as the mean square error function, etc. [1].

The output of the DNN model is then used as input to a bidirectional LSTM RNN. This input is extended to include several neighbor sites as a new input to LSTM RNN, to take into account weak long effects inherent to ECD process (i.e., the effects of neighbor site patterns on the local surface profile heights after deposition). Using the DNN output values of neighbors, the new input dimension becomes 7*7*2. For each site, we use as input an extended block containing three neighbor sites, from left, right, top and bottom, or radius 3 neighbors of a given site. The goal of this block of data is to improve the predictions of the DNN results by taking into account the effects of neighbors. In Fig. 3, scatter plots of height data of predictions vs. targets are presented after the application of LSTM RNN.

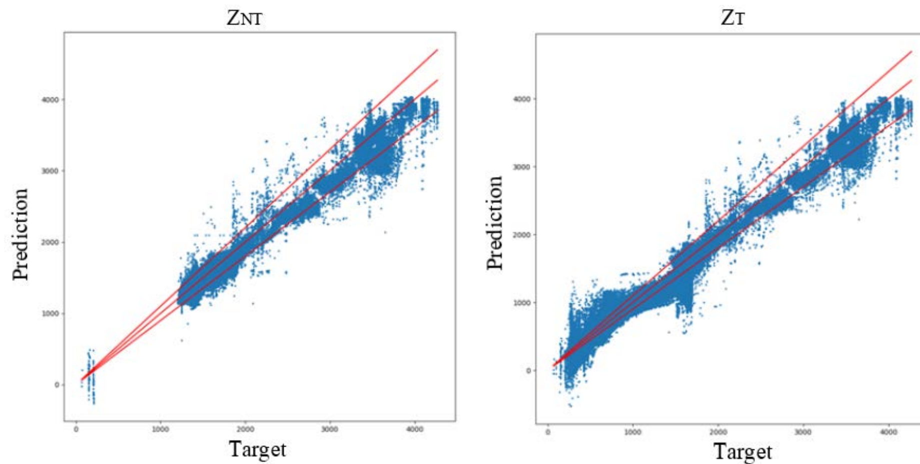

Fig. 3. Scatter plots of combined DNN and LSTM RNN model predictions vs. targets for ZNT and ZT height data.

6. CNN-Based Modeling

The next model was based on CNN [7-8]. The architecture we used consists of two blocks, as shown in Fig. 4. The first block is a separate feature extraction for each input parameter through CNN. The collection of all outputs from the first phase serves as an input to the second block of the network. The second block consists of a DNN with three hidden layers and 10 neurons per layer. Again, the berHu function is used for the loss function.

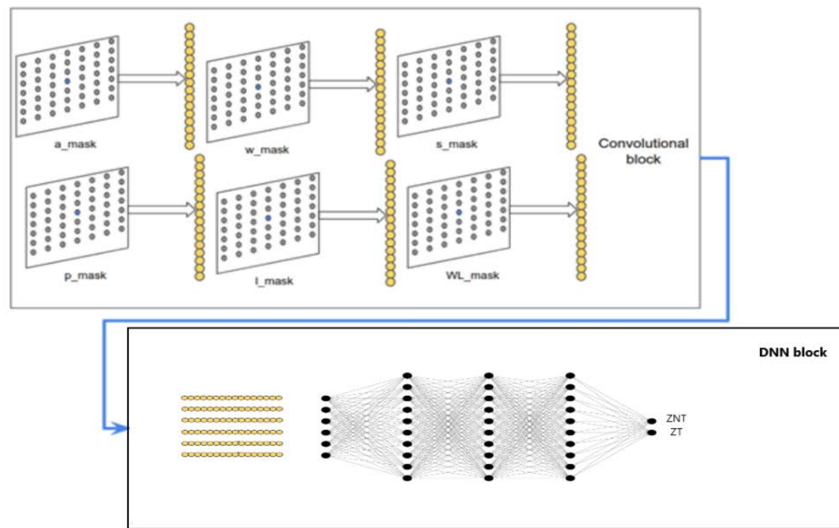

Fig. 4. Schematic plot of the combined CNN-DNN model.

The accuracy of the model predictions on the test set is analyzed using scatter plots of prediction vs. target data for surface heights, as shown in Fig 5. Red lines show 10% deviations from the target.

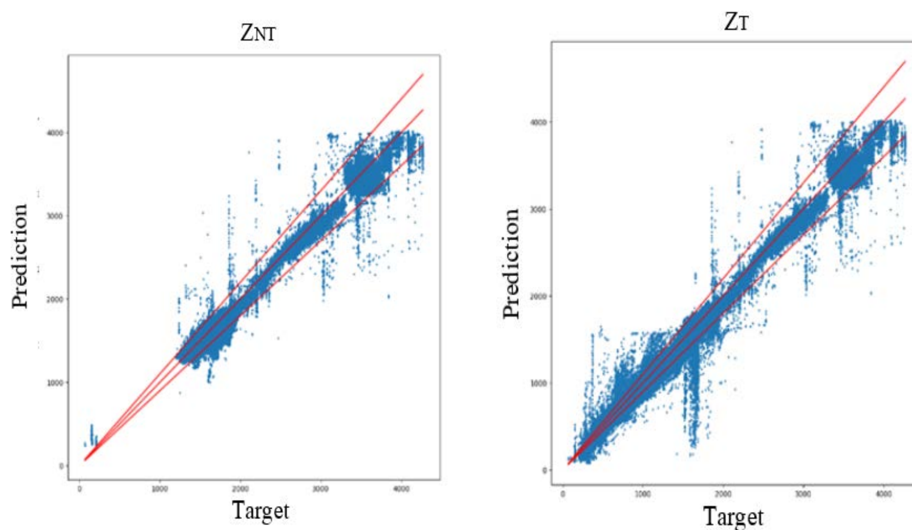

Fig. 5. Scatter plots of predictions vs targets for ZNT and ZT using NN model.

7. XGBoost Method for Modeling ECD Surface Profiles

Finally, algorithms of the XGBoost library [9] were used for ECD surface profile modeling. The motivation for using XGBoost is that it uses fewer resources compared to DNN and CNN approaches. XGBoost is a form of a gradient tree boosting or a gradient boosting machine. It is implemented with a high-level optimization strategy for fast execution speed and high

performance [9]. A gradient boosting algorithm is a sequential combination of several predictors, where each predictor is an improvement on its predecessor (e.g., by fitting to the errors of its predecessor). This algorithm uses a number of hyper-parameters. We tune the parameters using a greedy search. We use a linear regression model as the objective function, and choose the best results by minimizing the root mean squared error. We use the same training, validation, and testing datasets for these experiments as for the previous models.

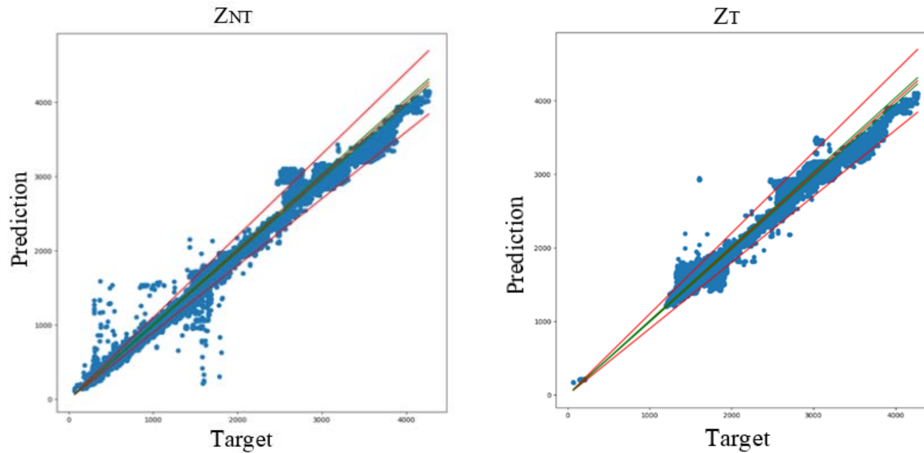

Fig. 6. Scatter plots of predictions vs targets for ZNT and ZT using XGBoost.

A significant improvement in prediction accuracy is obtained with the XGBoost algorithm, as shown in scatter plots in Fig 6-7. Moreover, we noticed a significant reduction in model training time with XGBoost compared to the other approaches we tried.

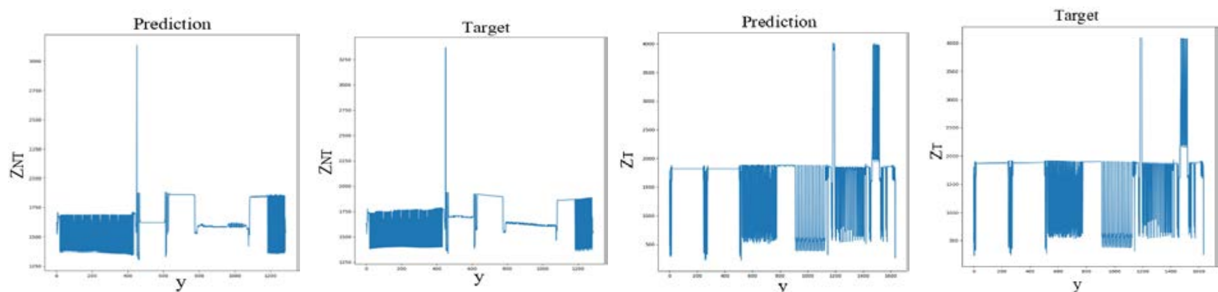

Fig. 7. Linescan (for fixed x value) predictions vs targets plots for ZNT and ZT using XGBoost.

8. Conclusion

In this paper, different ML methods were used to model surface profiles after electrochemical deposition of copper over patterned wafers for the creation of interconnection wires for chips. Weak long-range interactions of patterns on the design are inherent for post-ECD surface profiles, which means that the surface height above the given pattern is not defined solely by the pattern itself, but is affected by neighbor patterns. Using physics and chemistry-based electrochemical deposition simulation, training, validation, and test data was generated for the ML model building.

Several ML models were used for modeling, including combined DNN-LSTM RNN, combined CNN-DNN, and XGBoost-based models. It was found that the XGBoost-based model provided the best accuracy in surface height prediction, as well as correct data trends and high correlation with simulated linescans. It also has a much shorter training time (a couple of hours) compared to other methods (several hours or several days).

Acknowledgements

The authors would like to express their appreciation to Shelly Stalnaker for her editorial assistance in the preparation of this paper.

References

[1] D. Zhao and X. Lu, "Chemical mechanical polishing: Theory and experiment," Friction, vol. 1, no. 4, pp. 306-326, 2013.

[2] G. Banerjee and R. L. Rhoades, "Chemical mechanical planarization historical review and future direction," *ECS Transactions*, vol. 13, no. 4, pp. 1-19, 2008.

[3] J. Jhothiraman and R. Balachandran. "Electroplating: Applications in the Semiconductor Industry." Advances in Chemical Engineering and Science, vol. 9, pp. 239-261, 2019.

[4] R. Ghulghazaryan, J. Wilson and N. Takeshita, "CMP Model building and hotspot detection by simulation", *Proceedings of 158th Meeting of Planarization CMP Committee*, Nagoya, Japan, vol. 55 pp. 55-59, 2017.

[5] R. Ghulghazaryan, J. Wilson and N. Takeshita, "Building CMP Models for CMP Simulation and Hotspot Detection", *Mentor, a Siemens Business*, mentor.com (whitepaper), 2017.

[6] R. Ghulghazaryan, D. Piliposyan and J. Wilson, "Application of neural network-based oxide deposition models to CMP modeling", *ECS Journal of Solid State Science and Technology*, vol.8, no. 5, pp 3154-3162, 2019.

[7] I. Laina, Ch. Rupprecht, V. Belagiannis, F. Tombari and N. Navab, "Deeper depth prediction with fully convolutional residual networks", CoRR, abs/1606.00373, 2016.

[8] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning,* MIT press, 2016.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in KDD '16: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

Submitted 18.12.2019, accepted 22.04.2020

Մեքենայական ուսուցման միջոցով ստեղծված էլեկտրաքիմիական նստեցման մոդելների կիրառումը քիմիա-մեխանիկական փայլեցման մոդելավորման համար

Ռուբեն Գ. Ղուլղազարյան[1], Դավիթ Գ. Փիլիպոսյան[1],
Միսակ Տ. Սհոյան[1] և Հայկ Վ. Ներսիսյան[2]

[1] Mentor, a Siemens Business, Երևան, Հայաստան
[2]Հայաստանի Ամերիկյան Համալսարան Երևան, Հայաստան
e-mail: ruben_ghulghazaryan@mentor.com

Ամփոփում

Քիմիա-մեխանիկական փայլեցումը/պլանարիզացիան (CMP) հանդիսանում է ժամանակակից ինտեգրալ միկրոսխեմաների արտադրության հիմնական գործընթացը: CMP-ից առաջացած մակերևույթի պրոֆիլի մոդելավորումն ունի վճռորոշ նշանակություն արտադրությունից առաջ չիպերի պլանարության դեֆեկտների, «տաք կետերի» (hotspots) ստացման և կրիտիկական խափանումերը բացահայտելու համար: Էլեկտրաքիմիական նստեցումը (ECD) ժամանակակից միկրոսխեմաների միջմիացումների և միջանցիկ միացումների (vias) անթերի լցոնման կարևորագույն պրոցես է: ECD-ից հետո առաջանում են մակերևույթի տոպոգրաֆիայի զգալի փոփոխություններ, որոնք ազդում են CMP-ից հետո առաջացած մակերևույթի պրոֆիլի վրա:

Այս հոդվածում ուսումնասիրված են մեքենայական ուսուցման մի քանի մոտեցումներ՝ ECD-ից ստացված մակերևույթի պրոֆիլը մոդելավորելու համար, որն օգտագործվելու է որպես մուտքային տվյալ CMP մոդելավորման համար: Հետազոտված են խորը նեյրոնային ցանցերի, երկար կարճաժամկետ հիշողության (LSTM) ռեկուրրենտ ցանցերի, փաթույթային նեյրոնային ցանցերի և XGBoost գրադիենտային բուսթինգի ալգորիթմների տարբեր համակցություններ: Այլ մոդելների նկատմամբ XGBoost գրադարանի վրա հիմնված մոդելը ցույց տվեց ամենաբարձր արտադրողականություն և ճշտություն:

Բանալի բառեր՝ CMP, ECD, մեքենայական ուսուցում, նեյրոնային ցանցեր, LSTM, XGBoost.

Использование моделей электрохимического осаждения основанных на машинном обучении для моделирования химико-механического полирования

Рубен Г. Гулгазарян[1], Давит Г. Пилипосян[1], Мисак Т. Сгоян[1] и Айк В. Нерсисян[2]

[1] Mentor, a Siemens Business, Ереван, Армения
[2] Американский университет Армении, Ереван, Армения
e-mail: ruben_ghulghazaryan@mentor.com

Аннотация

Химико-механическое полирование/планаризация (CMP) - основной процесс, используемый в производстве современных интегральных микросхем. Моделирование профиля поверхности после CMP имеет решающее значение для дефектов планарности, «горячих точек» (hotspots) перед началом производства чипов и выявления их фатальных отказов. Электрохимическое осаждение (ECD) - это ключевой процесс бездефектного заполнения межсоединений и сквозных соединений (vias) в современных микросхемах. После ECD возникают значительные изменения топографии поверхности, которые влияют на профиль поверхности после CMP. В данной статье, для моделирования профилей поверхности после ECD, используемых в качестве входных данных для CMP моделирования, рассмотрены несколько подходов машинного обучения. Исследуются различные комбинации глубоких нейронных сетей, рекуррентных нейронных сетей с долгой краткосрочной памятью (LSTM), сверточных нейронных сетей и алгоритмов градиентного бустинга XGBoost. По сравнению с другими моделями, модель на основе библиотеки XGBoost проявила наивысшую производительность и точность.

Ключевые слова: CMP, ECD, машинное обучение, нейронные сети, LSTM, XGBoost.