Deep Neural Network Modeling of Multiple Oxide/Nitride Deposited Dielectric Films for 3D-NAND Flash

Received October 8, 2020; revised November 16, 2020; accepted November 16, 2020

Jeong Eun Choi^a, Jungho Song^b, Yong Ho Lee^a, and Sang Jeen Hong^{a,*}

^aDepartment of Electronic Engineering, Myongji University, Yongin 17058, Republic of Korea

^bPlasma Technology Research Center, NFRI, Gunsan 54004, Republic of Korea

*Corresponding author E-mail: samhong@mju.ac.kr

ABSTRACT

The fast and accurate measurement of the thickness of multiple oxide/nitride layer deposition (MOLD) films is desirable to improve the quality of the plasma deposition process and potentially simplify the metrology in 3D NAND flash memory devices. In this study, we performed deep neural network modeling of the reflectance spectrum data of two pairs of oxide/nitride films on a silicon substrate. We designed a deep neural network model to estimate the thickness of four stacked thin-film layers and the MOLD film. Principle component analysis of this model was performed to develop another model with 27 features. Finally, a combined model was designed by fine tuning both the models and applying an ensemble algorithm to both. The mean absolute error of the combined predictive model was lower than that of the individual models. We verified the performance of the proposed model by considering the thin-film deposition mechanism with respect to the infrared reflectance metrology of MOLD thin films. This study demonstrates the potential of machine learning for predicting the thickness of multiple layered films, addressing the limitations of optical metrology.

Keywords: Deep neural network, Ensemble, Prediction, Machine learning, Thickness metrology

1. Introduction

In the area scaling technology, three dimensional (3D)-NAND flash memory has replaced conventional planar NAND devices owing to smaller size of NAND flash memory devices [1]. A high aspect ratio etching process and multiple dielectric deposition for charge trapping and isolation are required to perform area scaling in 3D-NAND flash memory; this increases the complexity of device manufacturing. The variation in the thickness of the vertically stacked films in the dielectric multiple oxide/nitride layer deposition (MOLD) process is similar to the variation in the gate critical dimension in planar NAND. This affects the amount of charge stored in NAND devices and causes bit error in storage devices. The main challenge in 3D-NAND manufacturing is to form the required MOLD dielectric films exactly over the substrate. It has been reported that the thickness of actual deposited films tends to gradually increase in plasma enhanced chemical vapor deposition (PECVD) even though the films are deposited using the same process [2]. Hence, the fast and accurate measurement of the thickness of MOLD films has become increasingly important to improve the quality of the plasma deposition process.

Various techniques are used to measure film thickness, including stylus profilometry and interferometry. However, optical metrology techniques, such as ellipsometry and reflectometry, is preferred because it is nondestructive [3]. Ellipsometry measures the change in the polarization state of reflected light by irradiating light in the polarized

state onto the surface of a sample. Several desirable optical properties can be derived by measuring multiple wavelengths with a spectroscopic combination [4]. Ellipsometry parameters, such as the phase difference (Δ) and amplitude ratio (Ψ), as a function of ultraviolet to infrared wavelengths can be determined with high accuracy in a few seconds. Such data can be processed to provide the following information: (i) the depth profiles of interfaces, thin films, and multilayer structures with almost atomic resolution; (ii) the composition of any layers (bulk, interface, or surface) that are composites or alloys; (iii) the micro-roughness of surface layers [5]. Unlike a single layer, the reflections from multiple interfaces that occur in multiple layers form several peaks and valleys in the ellipsometry spectrum [6]. Moreover, measurement requires considerable time because numerous steps must be performed to measure the changes in the polarization state of reflected light by changing the incident angle and wavelength. Reflectometry can be a suitable method for overcoming these drawbacks [7].

Reflectometry detects or characterizes an object using the reflection of waves from a surface. This approach can rapidly analyze data from complex thin-film stacks to obtain thin-film thickness data, and it measures multi-angular reflectance for various polarization states in real time [8]. It is challenging to reduce the size of 3D-NAND devices because their metrology becomes more complex with decreasing size. Unlike planar NAND devices, 3D-NAND devices have vertical structures with a high aspect ratio. It causes problems in the metrology, such as multilayer structures, and opaque film measurement. Destructive

Figure 1. Principle of thin-film thickness measurement.

methods can be used for technological development; however, only nondestructive methods can be used for manufacturing [9]. Therefore, multilayer thickness measurement data must be obtained nondestructively using a reflectometer.

To resolve the issues of thin-film metrology in 3D-NAND devices, advanced data mining and the physical understanding of material forming are required for accurate and fast measurement. Spectroscopic analysis based on first principles is commonly used for film thickness metrology. The combination of the conventional Levenberg–Marquardt algorithm and the artificial neural network (ANN) modeling of the spectroscopic reflectometry data of transparent silicon and polysilicon on an oxide stack was suggested [10]. ANNs were used to obtain accurate initial estimates of film thickness and dispersion model parameters to determine the starting point of the Levenberg–Marquardt algorithm to improve the iterative computational efficiency. Recently, a machine learning methodology based on the measured thin-film thickness for virtual metrology was proposed. Machine learning with three types of algorithms that employed ANNs, gradient boosting regression, and random forest regression was investigated [11]. An improved film thickness measurement method that utilized a convolutional neural network was developed to prevent the ambiguity caused by the nonlinear fitting of spectroscopic data, which may fall into a local minimum problem [12]. The method consisted of two sets of convolution and pooling layers for feature extraction and one fully connected layer for classification. First, a 300×1 initial featured vector was pooled into $32 \times 150 \times 1$ featured vectors, which were pooled into $64 \times 75 \times 1$ featured vectors. Finally, a 1×2 output vector was obtained via flattening. Previous studies demonstrated the potential of machine learning for thin-film metrology and prediction, but prediction for a single thin film or two thin films. It is worthwhile to investigate multiple-layer thin-film prediction for the need for multilayer thin film deposition in 3D devices.

In this study, we established a machine learning model for predicting the thickness of multiple oxide/nitride thin films using optical spectroscopic reflectometry data by employing deep neural networks (DNNs). The model can predict the thickness of a four-layer PECVD thin film using 226 wavelength features. We preprocessed the data using principal component analysis (PCA), which is a dimensionality reduction method. Furthermore, we applied an ensemble algorithm to two DNN models to increase learning speed and decrease prediction error.

2. Modeling

Optical spectrum analysis is widely used to measure film thickness [13]. We performed the modeling of the spectroscopic reflectometry data, thickness, and reflectance spectrum acquired from the PECVD MOLD process provided by the DACON data science competition [14]. Reflectance is the ratio of the intensities of reflected and incident light, and it varies with the wavelength of light, as illustrated in Fig. 1. The reflectance spectrum shows the distribution of reflectance according to wavelength. The analyzed device consisted of five layers, i.e., silicon nitride (Layer 1), silicon dioxide (Layer 2), silicon nitride (Layer 3), silicon dioxide (Layer 4), and silicon (Substrate). The thicknesses of the four layers were predicted using 226 wavelength features. The DNN model was trained with 810,000 data points and evaluated using 10,000 test data points. The model prediction error was calculated as the mean absolute error (MAE).

The first model used all 226 wavelength features as input features. In order to facilitate learning, the thicknesses of the four layers, which comprised output data, were scaled using MinMaxScaler. The prediction error of the model was minimized by constructing deeper hidden layers and performing intensive training. The DNN model with 13 hidden layers was trained with 600 epochs. Batch normalization was used to prevent gradient vanishing or exploding in the learning process of the DNN model. The batch size for training with 810,000 data points was set as 1024. The exponential linear unit was used as the activation function, and it was optimized using the adaptive moment estimation optimizer [15]. The structural diagram of the DNN model is shown in Fig. 2.

The DNN model with all features required considerable time to learn the training data. In the second model, PCA was applied to the input features to achieve a performance similar to that of the first model and to increase the learning rate. Figure 3 shows an example of reducing two dimensions to one dimension to help understand dimension reduction using PCA. The PCA-DNN model was optimized using the same weight update method as that for the DNN model. However, the difference was that the dimensions of the input data were reduced through PCA. The PCA data variance was set as 0.99 to retain 99% of the data. As a result, the number of wavelength features was reduced from 226 to 27. The second model was implemented using only the 27 features as input data. In this model, 8 hidden layers were configured, and the remaining model configuration was the same as that for the first model.

The third model was an ensemble DNN model that simultaneously

Figure 2. Structural diagram of DNN model.

Figure 3. Example of PCA.

Figure 4. Schematic of DNN ensemble model.

used the first and second models, as shown in Fig. 4. In this model, the previously established DNN and PCA-DNN models were refined using an ensemble algorithm. In machine learning, the ensemble algorithm can be employed to combine less accurate models into more accurate models by voting, bagging (bootstrap aggregating), boosting, and stacking [16]. Bagging increases the accuracy and stability of a machine learning algorithm, helps decrease variance, and addresses overfitting. We used bagging to obtain the ensemble of the first two models. The pseudo code for the ensemble DNN model is shown in Fig. 5.

Figure 5. Pseudo code for DNN ensemble model.

Algorithm 1 Pseudo code for DNN ensemble

1: **Input:** Data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$

2: **Procedure** Ensemble

3: Load the DNN model to L_1

4: Load the PCA_DNN model to L_2

5: $h_t = L_t(D)$ ($t = 1, \dots, T$) # In this study, $T = 2$

6: $H(x) = \sum_{i=1}^T h_t(x)/T$

7: end.

8: **Output:** Predicted set $H(x) = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m\}$

3. Results and discussion

Deeper hidden layers were implemented to increase model accuracy for training the DNN model. The DNN model with 226 input features contained 13 hidden layers. Batch normalization helped the model learn despite the number of deep hidden layers. The number of training epochs was increased to 600 to reduce the error rate of the DNN model. The MAE of the DNN model trained 600 times was low at 0.74. The MAE was obtained by averaging the absolute value of the deviation between actual and predicted values. In our study, it was intuitively considered as the prediction error of thin-film thickness.

Even though the error rate of the DNN model was low, considerable time was required to train the model. Data are not appropriately learned when dimensionality is high; this is referred to as the curse of dimensionality [17]. The learning time of the DNN model was high even though the dimensionality of the data was not sufficiently high to correspond to the curse of dimensionality. Thus, PCA was used to reduce dimensionality and hence learning time. PCA has been used in numerous studies for reducing the number of input features to decrease learning time and increase accuracy [18]. When the variance of PCA was 0.99, the number of input features was reduced from 226 to 27. Owing to the smaller number of input features, the number of hidden layers in the PCA-DNN model was decreased to eight. Thus, the number of parameters to be learned was reduced. The learning time of the PCA-DNN model was approximately half of that of the DNN model. However, the MAE of the PCA-DNN model (0.87) was slightly higher than that of the DNN model (0.74). This difference in MAE (0.13) was extremely small. Furthermore, the PCA-DNN model provided better results in terms of model complexity and learning time.

There are several limitations in implementing a new model by tuning its hyperparameters to reduce its error rate, e.g., an increase in training time. Thus, instead of implementing a new model, we fine tuned the hyperparameters, such as the batch size and learning rate, of the DNN and PCA-DNN models that were previously trained 600 times. Then, an ensemble algorithm was applied to the fine-tuned DNN and PCA-DNN models to create an ensemble model with a lower error rate. The prediction results of the ensemble DNN model are shown in Fig. 6. The test MAE of the model was 0.36, which was the lowest among the three models. The disadvantage of the ensemble model was that several models had to be trained to obtain it. As shown in Fig. 6, the error was large at large and small thin-film thicknesses. The thin-film thickness varied from 10 to 300 nm in the training data, and the error was the largest at 10 and 300 nm. This is presumed to be due to the scaling of the values from 10–300 nm to 0–1 by MinMax

Figure 6. Prediction results of the DNN ensemble model.

Figure 7. Change in thickness during deposition of pair of oxide/nitride layers [18].

Scaler. We consider that this problem can be resolved by utilizing StandardScaler instead of MinMaxScaler and using thin films of various thicknesses rather than limited data.

We examined the material properties of the PECVD oxide/nitride thin film to enhance the understanding of MOLD thin films for optical metrology. It should be noted that the silicon nitride in Layers 1 & 3 was deposited on silicon oxide; the silicon oxide in Layer 2 was deposited on silicon nitride, and the silicon nitride in Layer 4 was deposited on the silicon substrate. PECVD is a well-known thin-film deposition process. It forms a thin-film precursor in the plasma to be absorbed and agglomerated on an exposed surface. In conventional oxide/nitride thin-film dielectric deposition using PECVD, individual

deposition process chambers are utilized for different materials to prevent the cross contamination of two or more materials. However, the MOLD for 3D-NAND employs a single chamber for depositing relatively thinner oxide/nitride films. The thickness of individual layers is at most 150 Å, and it is easily affected by the chamber environment. The thickness of an individual layer is a function of the surface condition of the underlying layer, the chamber condition, and the accumulated temperature with respect to each layer. In one study, as shown in Fig. 7, layers were deposited in the same chamber using the same method. However, the thickness of the layers gradually increased owing to the aforementioned reasons [19].

4. Conclusions

We implemented a deep learning algorithm to simultaneously predict the thickness of four semiconductor layers based on the reflectance spectrum. We developed a DNN model, a PCA-DNN model by applying PCA to the DNN model, and an ensemble DNN model by applying an ensemble algorithm to the first two models. The accuracy of the PCA-DNN model was slightly lower than that of the DNN model. However, it required less learning time because the number of parameters to be trained was small. The ensemble DNN model exhibited the highest accuracy. This study utilized the data of two pairs of oxide/nitride films. The accuracy of the deep learning algorithm can be improved if the data for more MOLD layers can be obtained via experiments. Furthermore, this study used only the thin-film thickness data. However, as the roughness of the interface is an extremely important factor in multilayer thin-film systems, a model that predicts various thin-film properties, including roughness, can be implemented in future studies.

Acknowledgements

This work was supported by the Korea Institute of Advanced of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008458, The Competency Development Program for Specialist) and the MOTIE (Ministry of Trade, Industry & Energy (project number 20006499) and KSRC(Korea Semiconductor Research Consortium) support program for the development of the future semiconductor device.

References

- [1] J. Lee, J. Jang, J. Lim, Y. G. Shin, K. Lee, and E. Jung, *Proceedings of the IEEE International Electron Devices Meeting* (San Francisco, CA, USA, December 3-7, 2016), pp. 11.2.1-11.2.4.
- [2] D. B. Jang and S. J. Hong, *Trans. Electr. Electron. Mater.* 19, 21 (2018).
- [3] A. Piegari and E. Masetti, *Thin Solid Films* 124, 249 (1985).
- [4] K. H. Kim, *J. Korean Solar Energy* 39, 11 (2019).
- [5] K. Vedam, *Thin Solid Films* 313, 1 (1998).
- [6] M. Hasan, K. Lyon, L. Trombley, C. Smith, and A. Zakhidov, *AIP Adv.* 9, 125107 (2019).
- [7] Y. S. Ghim and H. G. Rhee, *Opt. Lett.* 44, 5418 (2019).
- [8] H. T. Huang and F. L. Terry Jr., *Thin Solid Films* 455, 828 (2004).
- [9] W. Zhang, J. Xu, S. Wang, Y. Zhou, and J. Mi, *J. Microelectron. Manuf.* 3, 20030102 (2020).
- [10] M. F. Tabet and W. A. McGahan, *Thin Solid Films* 370, 122 (2000).

- [11] D. H. Kim, J. E. Choi, T. M. Ha, and S. J. Hong, *J. Semicon. Display Technol.* 18, 48 (2019).
- [12] M. G. Kim, *Int. J. Precis. Eng. Manuf.* 21, 219 (2020).
- [13] J. Park, J. Kim, H. Ahn, J. Abe, and J. Jin, *Int. J. Precis. Eng. Manuf.* 20, 463 (2019).
- [14] DAICON <https://dacon.io/competitions/official/235554/data/> (in Korean) (accessed Jan. 1, 2020).
- [15] S. Ruder, *arXiv preprint* arXiv:1609.04747 (2016).
- [16] A. Dey, *Int. J. Comput. Sci. Inf. Technol.* 7, 1174 (2016).
- [17] L. Van Der Maaten, E. Postma, and J. Van den Herik, *J. Mach. Learn. Res.* 10, 66 (2009).
- [18] A. M. Sharifi, S. A. Kasmani, and A. Pourebrahimi, *J. Converg. Inf. Technol.* 10, 42 (2015).
- [19] D. B. Jang and S. J. Hong, *Trans. Electr. Electron. Mater.* 19, 21 (2018).