

Article

Open Access

Non-destructive thickness characterisation of 3D multilayer semiconductor devices using optical spectral measurements and machine learning

Hyunsoo Kwak¹, Sungyoon Ryu², Suil Cho², Junmo Kim³, Yusin Yang² and Jungwon Kim¹

Abstract

Three-dimensional (3D) semiconductor devices can address the limitations of traditional two-dimensional (2D) devices by expanding the integration space in the vertical direction. A 3D NOT-AND (NAND) flash memory device is presently the most commercially successful 3D semiconductor device. It vertically stacks more than 100 semiconductor material layers to provide more storage capacity and better energy efficiency than 2D NAND flash memory devices. In the manufacturing of 3D NAND, accurate characterisation of layer-by-layer thickness is critical to prevent the production of defective devices due to non-uniformly deposited layers. To date, electron microscopes have been used in production facilities to characterise multilayer semiconductor devices by imaging cross-sections of samples. However, this approach is not suitable for total inspection because of the wafer-cutting procedure. Here, we propose a non-destructive method for thickness characterisation of multilayer semiconductor devices using optical spectral measurements and machine learning. For > 200-layer oxide/nitride multilayer stacks, we show that each layer thickness can be non-destructively determined with an average of approximately 1.6 Å root-mean-square error. We also develop outlier detection models that can correctly classify normal and outlier devices. This is an important step towards the total inspection of ultra-high-density 3D NAND flash memory devices. It is expected to have a significant impact on the manufacturing of various multilayer and 3D devices.

Introduction

The increasing demand for data storage systems, ranging from data centres to various smart devices, has led to an increasing need for higher-capacity, more compact memory devices. As the cell-to-cell distance has decreased

to less than 10 nm, traditional two-dimensional (2D) scaling methods suffer from cell-to-cell interference and technical difficulties in fabrication processes^{1,2}. As an alternative approach, three-dimensional (3D) scaling has been proposed, and it has increased the number of transistors per area by overcoming the spatial limitations of traditional 2D devices³. Most notably, the storage capacity and energy efficiency of 3D NAND flash memory devices have been significantly improved by stacking memory cells vertically^{4,5}. Since the first launch of 3D NAND with 24 word line layers in 2013⁶, the number of layers has been rapidly increasing, and 3D NAND with approximately 100

Correspondence: Jungwon Kim (jungwon.kim@kaist.ac.kr)

¹Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

²Memory Metrology & Inspection Technology Team, Memory Manufacturing Technology Center, Samsung Electronics Co. Ltd., Gyeonggi-do 18448, South Korea

Full list of author information is available at the end of the article.

© The Author(s) 2021



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

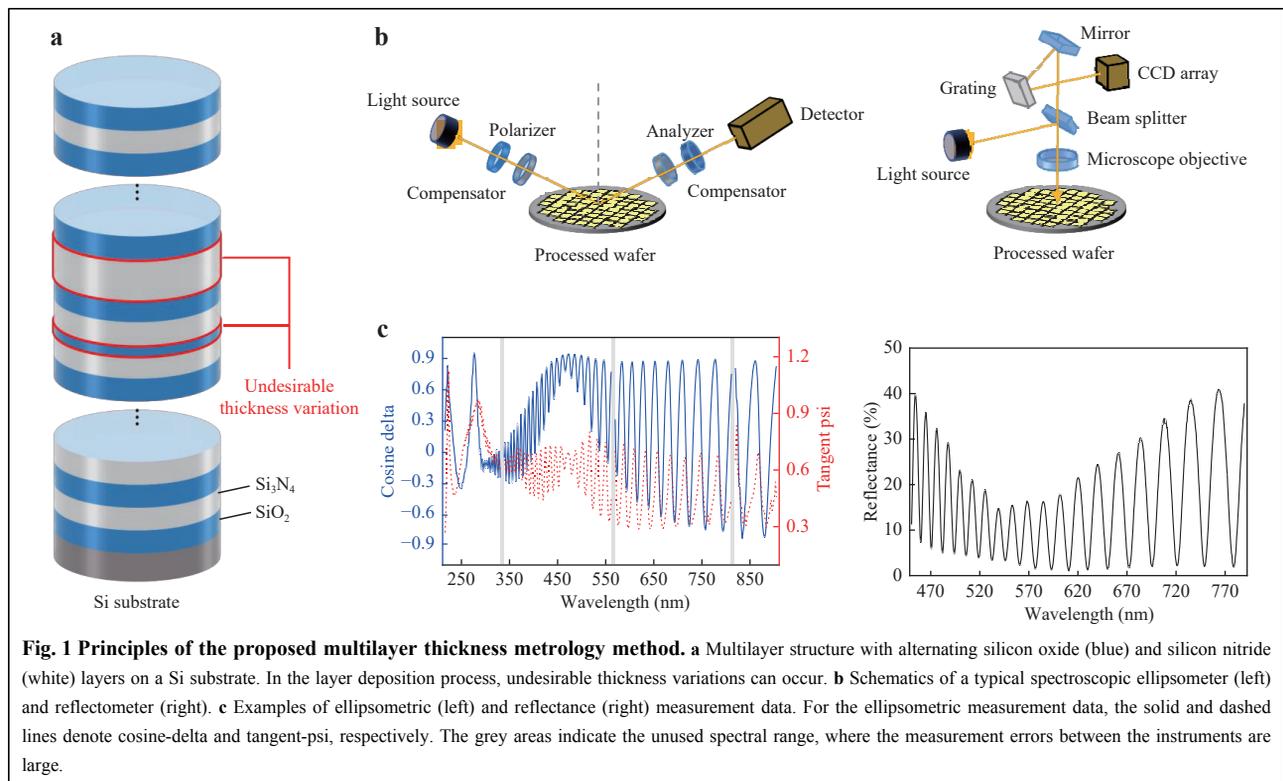
word line layers has recently been commercialised⁷. Recently developed 3D NAND flash memory has a storage density of 1 terabit per 180 mm² footprint⁸. Driven by demands for more massive storage devices, the market size for 3D NAND is expected to grow exponentially from \$9 billion in 2017 to \$100 billion by 2025⁹.

There are several different methods of fabricating 3D NAND^{10–13}. Nevertheless, building multilayer structures by alternating layers of semiconductor materials in the initial fabrication process (Fig. 1a) is the same for all approaches. In the multilayer deposition process, residual stresses can occur owing to the different thermal expansion coefficients between the layers¹⁴. This results in undesirable thickness variations after the process is complete. Even small thickness variations in each layer can affect the circuit performance of the final product^{15,16}. Therefore, it is highly desirable to accurately assess the thickness of the stacked semiconductor layers.

To date, various measurement methods have been used in semiconductor device fabrication facilities to measure the nanoscale features of semiconductor devices^{17,18}. In particular, transmission electron microscopy (TEM) has been used to measure the thickness of semiconductor multilayer stacks^{19,20}. TEM has the advantage of high resolution and high magnification. However, owing to the destructive nature of the required wafer-cutting process,

this technique cannot be used for total inspection. Another cross-sectional approach, for example, is to measure the cross-section of a multilayer by scatterometry with fast calculations using analytic approximation²¹. Interference microscopy can be used for the simultaneous characterisation of the multilayer thickness and the surface imaging²². Spectral ellipsometry, a non-destructive optical method, has been used for multilayer thickness characterisation^{23–26}. However, as the number of layers increases, accurate thickness characterisation using spectroscopic ellipsometry becomes more difficult on account of errors in the measurement instruments and changes in the material properties of each layer under different fabrication conditions. Meanwhile, because the number of layers in 3D NAND will increase well above 200 layers in the near future, machine learning can be more effective for thickness characterisation of multilayer structures as compared to fitting methods^{27,28}. This is because the machine learning algorithm effectively learns the correlations between spectroscopic data and multilayer thickness without physical interpretation. Although thickness characterisation by artificial neural networks (ANNs) has been previously reported^{29–31}, the characterisations were conducted only for a few (e.g. less than four) layers.

We herein demonstrate a non-destructive method for



thickness characterisation of each layer in the > 200-layer semiconductor multilayer stacks that are used in commercial 3D NAND devices. By exploiting the structural similarity between semiconductor multilayer stacks and dielectric multilayer mirrors^{32,33}, various spectroscopic methods, including ellipsometric and reflectance measurements^{34,35} (Fig. 1b), which are commonly used in dielectric mirror analysis, are employed. Based on the obtained spectroscopic data (Fig. 1c), machine learning is used to predict the thickness of each layer. From theoretical optical modelling (see ‘Materials and methods’ section), we exploit the well-known fact that the thickness of each layer affects the spectroscopic ellipsometric and reflectometric spectra. We can predict the thickness of each layer with an average root-mean-square error (RMSE) of approximately 1.6 Å (1.6×10^{-10} m, with ± 0.2 Å standard deviation) for > 200-layer 3D semiconductor devices. In addition, using a machine learning model trained with simulated data, it is possible to correctly classify normal and outlier devices (e.g. a multilayer structure having a layer with > 30 Å deviation from the targeted layer thickness).

Results

Accurate determination of layer-by-layer thickness for normal samples. The tested samples were multilayer semiconductor devices with alternating layers of oxide (SiO₂) and nitride (Si₃N₄) on a silicon substrate. The total number of layers was approximately 200, with a total thickness of approximately 5.5 μm. Most layers consisted of quasi-periodic oxide/nitride layers with a thickness of 200 Å–330 Å, except for several top and bottom layers with a thickness of 100 Å–1,600 Å. For multilayer thickness prediction, ellipsometric data of 148 normal samples were used. For an outlier detection test, reflectance data of 45 normal samples and three outlier samples were used. Commercial ellipsometers and reflectometers (Atlas XP+, Nanometrics, Inc.), which were installed in the production lines of the 3D devices characterised in this work, were used to obtain the spectroscopic data. The ellipsometric data (psi and delta) were measured at an incident angle of 65° for a spectral range of 216–905 nm (Fig. 1c). Psi and delta were measured at 991 different wavelengths. In total, each sample had 1982 (= 991 × 2) measured psi and delta values. The reflectance was measured at an incident angle of 0° for a spectral range of 450–790 nm (Fig. 1c). The reflectance was measured at 741 different wavelengths.

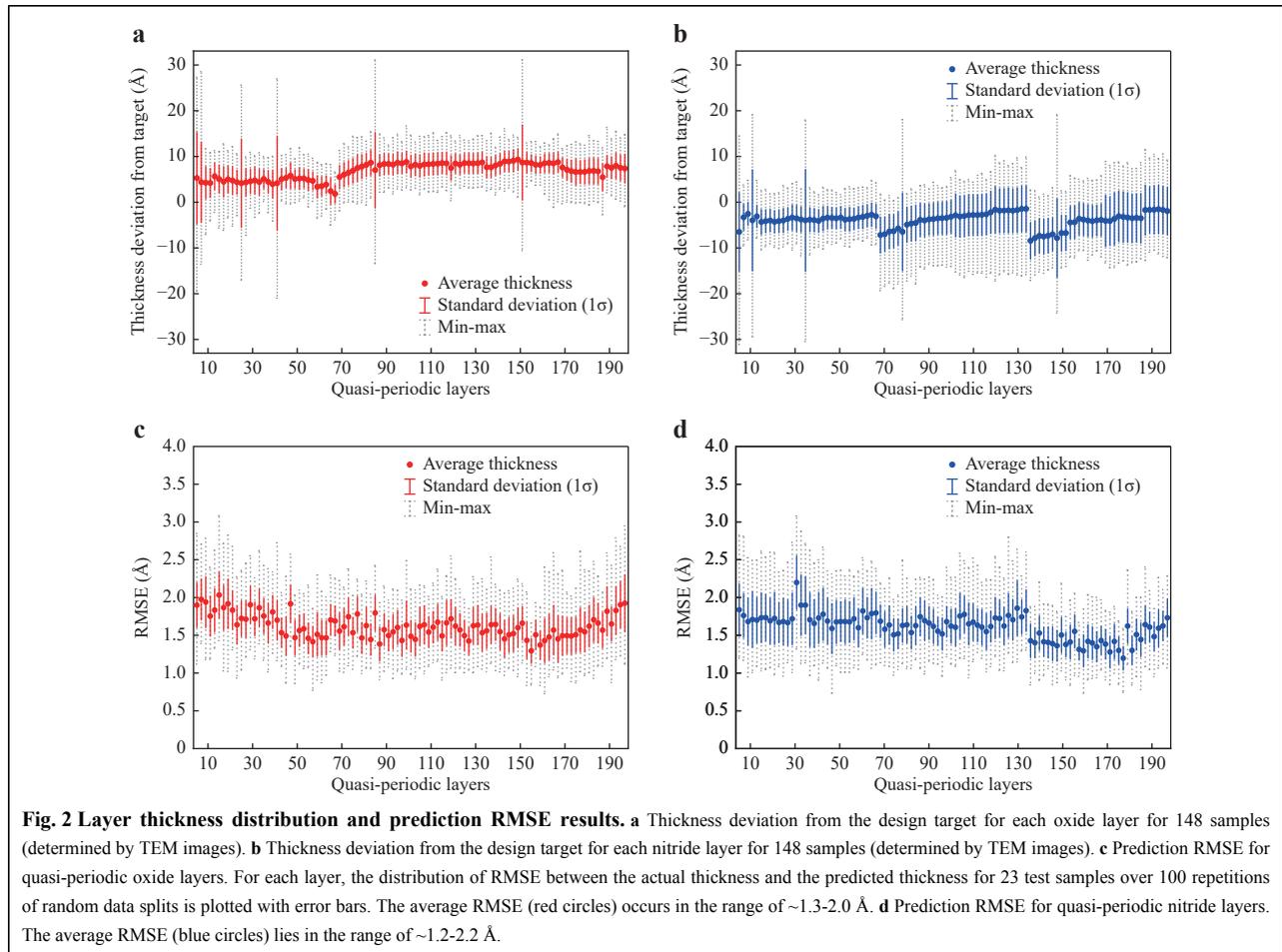
After the spectroscopic measurements were conducted, each wafer was cut, and its cross-section was imaged by TEM. The TEM images were used as a reference for evaluating the accuracy of the proposed method. From the

TEM images we determined that, even for the normal samples, the actual layer thickness could vary by up to approximately 20–30 Å from the target thickness, which corresponded to approximately 10–15% errors in the fabrication. The standard deviation of each layer thickness was in the range of approximately 3–11 Å (see Fig. 2a, b for the distributions of oxide and nitride layers, respectively).

To determine the thickness of each layer from the measured spectral data, machine learning was used. For the machine learning model, spectral data and layer thicknesses were used as inputs and outputs, respectively. A total of 148 normal samples were randomly split into a training set of 125 samples and a test set of 23 samples. Owing to the limited number of available samples with TEM data, the number of training samples was increased to 5,000 by data augmentation based on noise injection methods (see ‘Materials and methods’ section). We used various machine learning models, such as support vector regression, linear regression models, and artificial neural networks. To evaluate the models, we implemented a five-fold cross-validation test. The linear regression model showed the best performance. Because the initial random split of 148 samples could reflect a biased result, we randomly split the dataset into 100 different combinations of training and test sets and trained the linear regression model on each training set (see Fig. S1). Finally, we applied the trained model to each test set.

Fig. 2a, b show the actual thickness distributions of 195 quasi-periodic layers for 148 normal samples (determined by the TEM images). After the deposition process, the oxide layer thickness tended to increase by approximately 7 Å, and the nitride layer thickness tended to decrease by approximately 4 Å from the original design target. The peak-to-peak distribution (grey bars in Fig. 2a, b of each layer thickness ranged from 10 Å to 50 Å. The standard deviation of the actual thickness (red and blue bars in Fig. 2a, b) was approximately 3–5 Å in most layers, while deviations of up to approximately 11 Å were also observed in some layers. Fig. 2c, d show the distributions of the RMSE of each layer thickness (i.e. the RMSE between the predicted layer thickness and the actual layer thickness determined by spectral data-driven machine learning and TEM imaging, respectively) for 23 test samples over 100 repetitions of random data splits. This result shows that our spectral measurement-based machine learning method achieved an average prediction RMSE of approximately 1.6 Å for each layer (red and blue circles for oxide and nitride layers, respectively).

To demonstrate the effectiveness of the proposed method for prediction, Fig. 3 presents a comparison



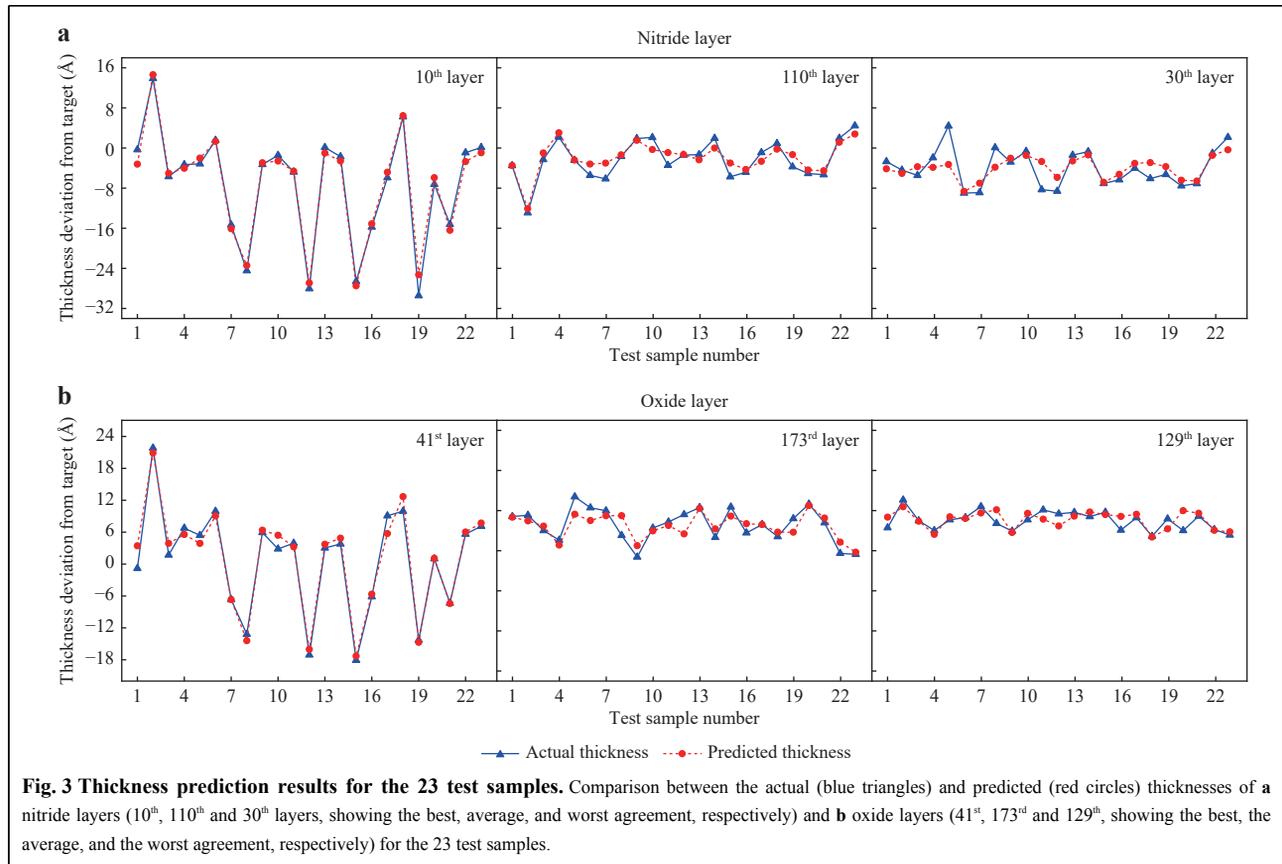
between the actual thickness (determined by the TEM images) and the predicted thickness of several nitride and oxide layers (selected from the bottom, middle, and top parts of the multilayer structure) in the 23 test samples. The predicted thickness aligns well with the actual thickness, regardless of the material or layer position used, with an average prediction RMSE of approximately 1.6 Å.

To evaluate the correlation between the predicted and actual layer thicknesses, the R-squared value was calculated for each layer. Fig. 4a, b show the distribution of the R-squared values for each of the oxide and nitride layers, respectively. As shown in Fig. 4a, the highest and lowest R-squared values are 0.97 (41^{st} layer, denoted by (i)) and 0.24 (35^{th} layer, denoted by (ii)), respectively. As shown in Fig. 4c, for both the 41^{st} and 35^{th} layers, the predicted thickness (denoted by circles) is consistent with the actual thickness (denoted by triangles) with an RMSE of approximately 1.6 Å. Thus, the resulting R-squared value for the 35^{th} layer is much lower than that of the 41^{st} layer because the actual thickness distribution is much narrower (2.1 Å and 10.3 Å RMS distributions for the 35^{th}

and 41^{st} layers, respectively). As shown in Fig. 4d, even though the actual total layer thickness is widely distributed from -400 to $+900$ Å from the design target, the predicted total thickness, which is the sum of all predicted layer thicknesses, has a high correlation (R-squared = 0.93) with the actual total thickness.

It is noteworthy that the number of training samples could be reduced at the expense of slightly degraded prediction performance. Figure S2 shows the average RMSE of each layer according to the number of training samples used for thickness characterisation of the > 200 -layer structure. For example, when using 25 and 75 training samples (instead of 125), the RMSEs for the test set remain ~ 2.1 Å and ~ 1.75 Å, respectively. In addition, to investigate the applicability of our method, we applied our method to different multilayer oxide/nitride structures with ~ 65 , ~ 110 , and ~ 130 layers. As summarised in Table S1, 1.4 – 1.7 Å level RMSEs (over 100 repetitions of random data splits) could be obtained for each case when > 60 samples for training were used.

Outlier device detection using simulated data. In

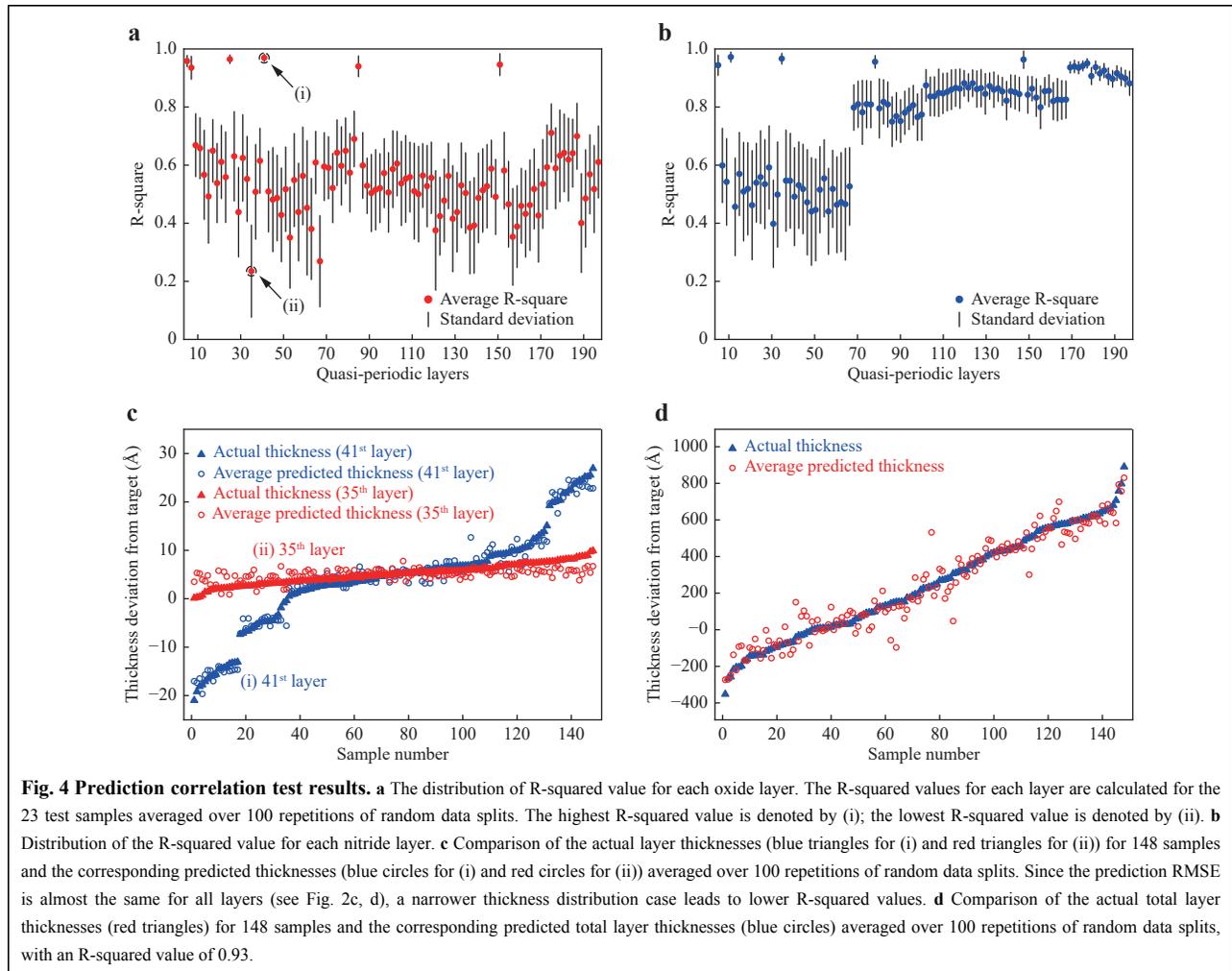


addition to the accurate determination of the multilayer thickness under normal fabrication conditions (as shown in Figs. 2-4), which is helpful for controlling etching and deposition processes, we developed another machine learning model that can detect outliers when layer thicknesses significantly vary (e.g. by more than 30 Å) from the design target. To distinguish outlier cases from normal cases, both normal and outlier samples are required to train the machine learning model. However, because it is impossible to fabricate all possible outlier samples for this training, we used a large number of simulated spectral data for more effective and economical training. The measured reflectance showed reasonable agreement with the simulated data. Therefore, reflectance data were used for the outlier detection models. We first generated 1,000 simulated data with a wide range of thickness distributions as outlier cases. We also generated 1,800 augmented data (as normal cases) by a noise injection method from 18 normal samples. A total of 2,800 training data points were used to train the linear regression model (see ‘Materials and methods’ section and Fig. S3).

To test the developed outlier detection method, three outlier samples were prepared by intentionally growing the 42nd layer thickness to be approximately 50 Å thinner than

the normal fabrication condition. As shown in Fig. 5a, the reflectance of the outlier sample (red circle line) is blue-shifted by approximately 5 nm compared to the normal sample (blue triangle line). For validation, 10 normal samples and one outlier sample were used. Finally, the test was performed for 17 normal samples and two outlier samples. Details are provided in the ‘Materials and methods’ section.

When defining an outlier case with a one-layer thickness exceeding 30 Å from the target, all the normal samples were classified as normal cases, and all the outlier samples were classified as outlier cases. When we modified the outlier threshold from 10 Å to 50 Å, we obtained a sensitivity-specificity graph, as shown in Fig. 5b. For the normal samples, the thicknesses of all layers were predicted to have an average RMSE of 7.4 Å from the target (Fig. 5c). For the outlier samples, the average thickness of the 42nd layer was predicted to have a -35 Å deviation from the target. The remaining layers were predicted to have an average RMSE of 8.6 Å from the target (Fig. 5d). Therefore, machine learning based on simulated data could successfully detect the outliers (faulty devices) and the exact erroneous layer location in the device.

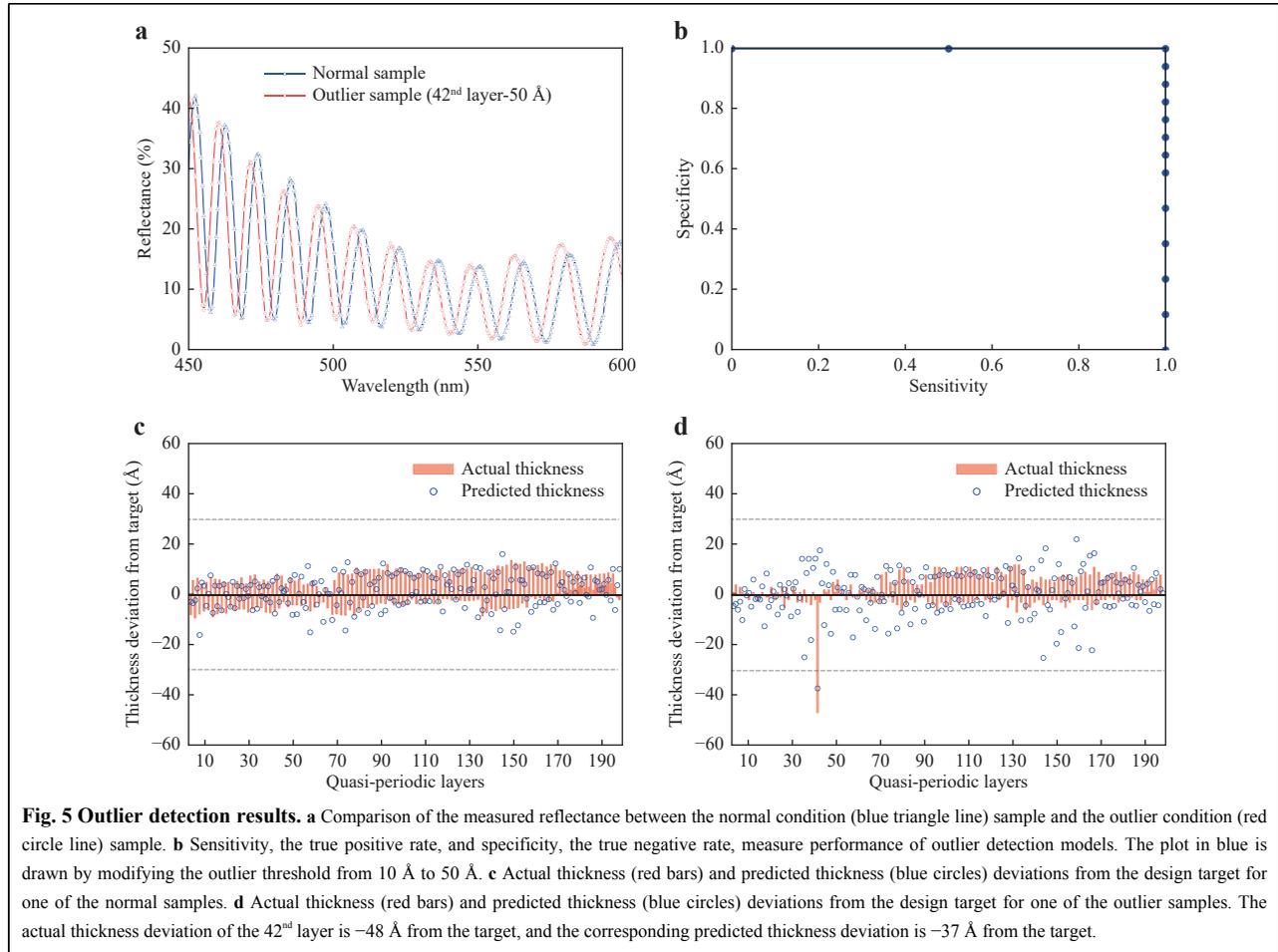


Discussion

In summary, we demonstrated a non-destructive method to accurately characterise each layer thickness and to detect outliers in ultra-high-density 3D semiconductor devices consisting of more than 200 layers. The machine learning approach uses a data-driven algorithm that considers only the correlation between spectral data and thickness information. We could thus eliminate many measurement-related issues, such as absolute accuracy errors and drift in measurement instruments, as well as in situ material properties that are not completely measurable (e.g. changes in the wavelength-dependent refractive indices of each layer under different fabrication conditions). When using noisy data as input to machine learning algorithms, the trained model is robust against various measurement errors. In addition, our outlier detection method can detect significant thickness defects by using a relatively small number of TEM measurements (e.g. 18 samples used as normal cases in this work) and massive simulated data (used as outlier cases). As a result, this method is highly

suitable for application in actual semiconductor manufacturing facilities. In our work, all the spectroscopic data were obtained in commercial 3D NAND manufacturing lines, and only tens to hundreds of TEM measurements were required for model training. It is noteworthy that the proposed approach is suitable for the thickness characterisation of multilayer systems composed of dielectric materials, whereas the thickness characterisation for multilayer systems composed of materials with high extinction coefficients, such as titanium nitride (TiN) or tungsten, is challenging owing to the relatively short penetration depth (of the order of tens of nanometres) of those materials.

Our demonstrated method can be readily applied for the total inspection of various 3D semiconductor devices as well as many other types of highly complex multilayer stacked devices, such as ultra-broadband dielectric mirrors for high field physics and ultrafast science^{36,37}, thin-film bio-sensors for biotechnology^{38–40}, and hyperbolic metamaterials^{41,42}.



Materials and methods

Theoretical model of spectroscopic data. In this section, we describe the process of deriving the theoretical values of reflectance, psi, and delta³². In a multilayer system, the tangential components of electric field E and magnetic field H are continuous at the boundary between each layer. The tangential components of the electric field and the magnetic field at the interface of each layer have the following relationship:

$$\begin{bmatrix} E_t \\ H_t \end{bmatrix} = \begin{bmatrix} \cos \delta & i \sin \delta / \eta_1 \\ i \eta_1 \sin \delta & \cos \delta \end{bmatrix} \begin{bmatrix} E_b \\ H_b \end{bmatrix} \quad (1)$$

where E_t and H_t are the fields at the top interface; E_b and H_b are the fields at the bottom interface. The phase thickness δ is expressed as $\frac{2\pi N d \cos \theta}{\lambda}$, where N is the complex refractive index of the layer; d denotes the layer thickness; θ denotes the incident angle of the light; λ represents the light wavelength, and η_1 is the optical admittance of the medium. In a multilayer system, Eq. 1 is extended for all layers and expressed by a matrix-cascaded system, as shown in Eq. 2:

$$\begin{bmatrix} B \\ C \end{bmatrix} = \left\{ \prod_{r=1}^q \begin{bmatrix} \cos \delta_r & i \sin \delta_r / \eta_r \\ i \eta_r \sin \delta_r & \cos \delta_r \end{bmatrix} \right\} \begin{bmatrix} 1 \\ \eta_m \end{bmatrix} \quad (2)$$

where δ_r is the phase thickness; η_r denotes the optical admittance of the r -th medium; η_m represents the optical admittance of the substrate; q is the number of layers, and B and C are the normalised electric and magnetic fields, respectively. Finally, we obtain the theoretical reflectance as follows:

$$R = r r^* = \left(\frac{\eta_0 B - C}{\eta_0 B + C} \right) \left(\frac{\eta_0 B - C}{\eta_0 B + C} \right)^* \quad (3)$$

where r is the reflection coefficient, and η_0 is the optical admittance of the incident medium. For an oblique incident angle, η_r is multiplied by $\frac{1}{\cos \theta}$ for p-polarisation and by $\cos \theta$ for s-polarisation. Psi (Ψ) and delta (Δ) are derived from Eq. 4 as follows:

$$\frac{r_p}{r_s} = \tan(\Psi) e^{i\Delta} \quad (4)$$

where r_p is the reflection coefficient for p-polarisation, and r_s is the reflection coefficient for s-polarisation. Varying

the layer thickness leads to a change in the phase thickness for each wavelength of incident light, which results in changes in interference patterns of reflectance, ψ , and Δ .

Data preparation. Semiconductor multilayer stacks that are used for commercial 3D NAND devices were obtained at different locations on each wafer. For multilayer thickness prediction, 148 normal samples were obtained from 10 different wafers (10 to 17 different locations on each wafer). A spectroscopic ellipsometer installed in the production lines was used to measure 991 ψ - Δ pairs for each sample. For outlier detection, 45 normal samples were obtained from four different wafers, and three outlier samples were obtained from one wafer. A total of 741 reflectances were measured for each sample using a spectroscopic reflectometer in the production lines. High-resolution cross-sectional images of the samples were obtained using TEM.

Noise injection method. Data augmentation is widely used for a relatively small amount of data in many applications^{43–45}. Because our objective was to access only a small number of normal-condition samples (in commercial device production lines), we augmented the training samples by employing a noise-injection method. For multilayer metrology of normal conditions, 125 training samples were augmented by injecting noise, resulting in a total of 5,000 augmented data points (40 augmented data points per training sample). Note that 40 augmented data points for each training sample shared the same thickness profile. For each augmented data point, the spectral data could be shifted to the left or right (in wavelength) or shifted up or down as a whole from the original positions.

As shown in Fig. S4, α is added to all spectral data to inject vertical offset noise. To inject lateral offset noise by β , shifted spectral data at $(216 + \beta)$ nm to $(905 + \beta)$ nm should be obtained. However, since ψ and Δ were measured for wavelengths of 216–905 nm, we interpolated the original spectral data into shifted spectral ranges to obtain the shifted spectral data. When shifting spectral data to the left or right, redundant values can be generated because the shifted spectral data deviate from the actual measured range. These redundant values are truncated at both ends. A total of 12 redundant values were removed. Thus, 1,970 dimensional inputs were used.

Considering various possible noise sources during measurement (such as drift errors, wavelength errors, and refractive index changes), we added different amounts of noise under various conditions to the original spectral data. As shown in Fig. S5, injection of vertical offset noise uniformly distributed from -0.04 to $+0.04$ and lateral offset noise uniformly distributed from -6 to $+6$ nm was the best condition for thickness prediction performance for the

validation set. For an outlier detection test, we also augmented the training samples by employing the noise injection method. Eighteen training samples representing normal conditions were increased to 1,800 augmented data points (100 augmented data per training sample). For lateral noise injection, three redundant values were truncated at each end such that the shifted reflectance did not contain the redundant values. A total of 735 dimensional inputs were used. As shown in Fig. S5, injection of lateral offset noise uniformly distributed from -4 to $+4$ nm without using vertical noise injection was the best condition for thickness prediction performance for the validation set.

Performance comparison with ψ and Δ combinations. For multilayer metrology under normal conditions, we compared the RMSE of the validation set using a linear model to determine which combination of spectral data (ψ and Δ) should be used as the input to the model. As shown in Table S2, when both ψ and Δ are used, the RMSE of the validation set has the lowest value of 2.75 Å. Because machine learning learns the correlation between the input (spectroscopic measurements) and the output (layer thicknesses) rather than interpreting the physical meaning of the input data, there is no significant difference in the prediction RMSE, regardless of whether ψ or Δ is used. From these results, we find that the thickness prediction model performs best when using all spectroscopic data as inputs.

Evaluation of machine learning models For multilayer metrology under normal conditions, we first randomly split 148 samples into 125 training samples and 23 test samples. The 125 training samples were divided into five folds (25 samples per fold). One hundred samples (four of five folds) were converted to 4,000 augmented data using the noise injection method to train the model, and the remaining 25 samples (one of five folds) were used as the validation set to evaluate the model. Each fold was used as a validation set, and five validation results were averaged to measure the model performance. This method, called K-fold cross-validation^{46,47} (in this case, five-fold cross-validation), is widely used to identify the best model. For the model evaluation, the RMSE between the predicted thicknesses and the actual thicknesses of the validation set was used. As shown in Table S3, the RMSE of the validation set was found to be the lowest for the linear model. We evaluated the performance of the linear model according to the training data size, which is shown as the learning curve in Fig. S6. The RMSE was calculated by increasing the training data size from 40 to 4,000 (with 40 intervals).

As the number of training data increased, the RMSE of the training set increased because it became more difficult

to perfectly fit the training data. Meanwhile, the RMSE of the validation set decreased as the model became better fitted to unseen data. Because the RMSE of the validation set approached the RMSE of the training set until settling, the 4,000 training data were sufficient to train the linear model without overfitting.

For the implementation, we used a Titan X graphical processing unit (GPU). Generating 5,000 augmented data from 125 training samples by the noise-injection method required approximately 2 s. The complete training for the linear model required approximately 116 s. With the trained model, the prediction time for the test samples was less than 0.01 s. The most time-consuming process in this study was the model validation process because all the models were evaluated with the five-fold cross-validation technique. The linear model, support vector regression (SVR⁴⁸), and deep neural network (DNN) required ~550 s, ~9 h, and ~2 h for the cross-validation, respectively. Furthermore, to evaluate the model by modifying the hyperparameters (such as the number of hidden neurons or level of regularisation), the validation time for each model was multiplied by the number of hyperparameter sets used. However, since we found that the linear model performed the best in this study, the actual model validation time was relatively short.

Implementation details of machine learning models.

For multilayer metrology under normal conditions, we used three different machine learning models: SVR, linear regression⁴⁶, and an artificial neural network (ANN^{49,50}). All these models are regression models that predict continuous values (layer thicknesses). For all spectral data, feature standardisation is applied; thus, each feature vector has zero mean and unit variance. We compared the RMSE of the validation set with various conditions for each model. As shown in Fig. S7, when using the DNN model with a large number of hidden neurons, the RMSE of the training set converges to zero; however, the RMSE of the validation set does not decrease on account of overfitting during training. For the linear model, we applied L2 regularisation⁴⁶ to avoid overfitting and used a conjugate gradient function⁵¹ with 1,000 iterations to minimise cost function J as follows:

$$J = \frac{1}{2N} \left[\sum_{i=1}^N (p_i - y_i)^2 + \lambda \|\mathbf{w}\|^2 \right] \quad (5)$$

where N is the number of training samples; p_i denotes the predicted thickness; y_i is the actual thickness; λ represents a parameter that controls the level of L2 regularisation, and \mathbf{w} is the weight vector. A step-by-step algorithm operation process for the linear model is provided as a flow chart shown in Fig. S9. For the SVR model, we used a radial

basis function as the kernel function. Scikit-learn⁵² was used for implementing the SVR with cost function J as follows:

$$J = C \sum_{i=1}^N L_\varepsilon(p_i - y_i) + \frac{1}{2} \|\mathbf{w}\|^2 \quad (6)$$

where C is a regularisation parameter. L_ε is an ε -insensitive loss function given by

$$L_\varepsilon(p_i - y_i) = \begin{cases} 0, & \text{if } |p_i - y_i| < \varepsilon; \\ |p_i - y_i| - \varepsilon, & \text{otherwise} \end{cases} \quad (7)$$

Here, ε is the margin of tolerance, where no penalty is given to errors. ANNs with different architectures were implemented using Tensorflow⁵³. For the ANN models, batch normalisation⁵⁴, a ReLU activation function⁵⁵, and dropout⁵⁶ were applied. Batch normalisation and ReLU were applied to all hidden layers, while dropout was applied to the last hidden layer. A linear activation function was used for the output layer. As the best result for each ANN model, for the two-layer neural network (NN), eight neurons were used for the hidden layer without a dropout layer. For the three-layer DNN, 512 neurons were used for each hidden layer with a dropout layer (50% drop probability). For the four-layer DNN, 512 neurons were used for each hidden layer with a dropout layer (50% drop probability). The batch size was 128 in all cases. We used the Adam optimiser⁵⁷ with 10,000 epochs. It should be noted that an epoch denotes one full training iteration for each training data. The learning rate was 0.003.

Outlier detection methods To detect outliers, we used simulated spectroscopic data for model training. The matrix method³² was used to obtain the theoretical values of reflectance (see the ‘Theoretical model of spectroscopic data’ section in ‘Methods’). To simulate the spectroscopic data, the thickness of each layer and the refractive index of each medium were required. We used the measured refractive index obtained by a single layer measurement of each material with an ellipsometer (Fig. S8) as the refractive index of each material (oxide, nitride, and Si substrate) in the modelling. Because the outlier detection method focuses on detecting relatively large thickness changes, precise optical modelling by accurate refractive index characterisation was not required. We assumed that all the oxide and nitride layers shared the same oxide and nitride refractive indices, respectively. In addition, we assumed that there were no surface roughness or interface layers in the multilayer structures, which was also confirmed by the TEM measurement results.

Instead of using one model to predict the thickness of all layers, multiple models (i.e. one model per layer) were used to predict the multilayer thickness. The reasons were (a) to avoid overfitting the model to a large amount of

simulated data generated for potential outlier cases, and (b) to magnify the sensitivity to the critical thickness changes of a single layer.

To train and test the outlier detection models, we prepared 45 normal samples and 3 outlier samples. As shown in Fig. S3, the 45 normal samples were first randomly split into 18 training, 10 validation, and 17 test samples. The three outlier samples were randomly split into one validation and two test samples. Eighteen normal samples were increased to 1,800 augmented data by the noise injection method, and 1,000 simulated data, which were designed with a relatively large thickness variation in each layer, were generated. When designing the simulated outlier case data, the thickness of the outlier layer was uniformly distributed within a $\pm 20\%$ variation with respect to the reference thickness, and the thicknesses of the other layers were uniformly distributed within $\pm 4\%$ of the reference thickness. Here, the reference thickness denotes the average thickness of each layer for the 18 normal samples used for model training. A total of 2,800 training data (1,800 for normal cases and 1,000 for outlier cases) were used to train each outlier detection model. We used the linear model (with the L2 regularisation parameter of 100) for the outlier detection model because we found that the linear model performed the best in thickness characterisation of the multilayer (Fig. S7).

To determine the best noise-injection condition for the 18 training samples, 10 normal samples and one outlier sample were used as the validation set. We found that the lowest RMSE of the validation set was 4.78 Å when applying lateral offset noise uniformly distributed from -4 to $+4$ nm without vertical noise injection. For model testing, 17 normal samples and 2 outlier samples were put into each model to predict the thickness of each layer. Although a sample with a single-layer defect is used in this study, we anticipated that defects in multiple layers could be detected because our scheme is based on multiple models (i.e. one model per layer) for outlier detection.

Acknowledgements

This research was supported by the Industry–Academia Cooperation Program of Samsung Electronics Co., Ltd.

Author details

¹Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea. ²Memory Metrology & Inspection Technology Team, Memory Manufacturing Technology Center, Samsung Electronics Co. Ltd., Gyeonggi-do 18448, South Korea. ³School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

Author contributions

S.R., Y.Y. and Jungwon K. initiated and managed the project. Jungwon K. and H.K. conceived the main ideas and designed the experiments and

simulations. S.R. and S.C. prepared the devices under testing and obtained the spectral and TEM data. H.K. performed the modelling and simulations. Junmo K. advised on the machine learning models. All authors participated in the analysis of the data. H.K. and Jungwon K. wrote the manuscript with input from all authors.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Conflict of interest

H.K., S.R., S.C., Y.Y., and Jungwon K. are inventors with a patent application related to this work filed by Samsung Electronics Co., Ltd. and Korea Advanced Institute of Science and Technology (Korean patent application number 10-2018-0160334 filed on December 12, 2018 and US patent application number 16/678755 filed on November 8, 2019). The authors declare that they have no conflict of interest.

Supplementary information is available for this paper at <https://doi.org/10.37188/lam.2021.001>.

Received: 26 July 2020 Revised: 23 October 2020 Accepted: 06 November 2020

Published online: 12 January 2021

References

- Park, Y. et al. Scaling and reliability of NAND flash devices. Proceedings of 2014 IEEE International Reliability Physics Symposium. Waikoloa, HI, USA: IEEE, 2014, 2E-1.
- Li, Y. & Quader, K. N. NAND flash memory: challenges and opportunities. *Computer* **46**, 23-29 (2013).
- Nitayama, A. & Aochi, H. Vertical 3D NAND flash memory technology. *ECS Transactions* **41**, 15-25 (2011).
- Micheloni, R., Aritome, S. & Crippa, L. Array architectures for 3-D NAND flash memories. *Proceedings of the IEEE* **105**, 1634-1649 (2017).
- Kim, H. et al. Evolution of NAND flash memory: from 2D to 3D as a storage market leader. Proceedings of 2017 IEEE International Memory Workshop (IMW). Monterey, CA, USA: IEEE, 2017.
- Park, K. T. et al. Three-dimensional 128 GB MLC vertical NAND flash memory with 24-WL stacked layers and 50 MB/s high-speed programming. *IEEE Journal of Solid-State Circuits* **50**, 204-213 (2015).
- Maejima, H. et al. A 512Gb 3b/Cell 3D flash memory on a 96-word-line-layer technology. Proceedings of 2008 IEEE International Solid-State Circuits Conference (ISSCC). San Francisco, CA, USA: IEEE, 2018, 336-337.
- Lee, S. et al. A 1Tb 4b/cell 64-stacked-WL 3D NAND flash memory with 12MB/s program throughput. Proceedings of 2008 IEEE International Solid-State Circuits Conference (ISSCC). San Francisco, CA, USA: IEEE, 2018, 340-342.
- Kumar, R. & Tewari, D. Global 3D NAND Flash Memory Market to Reach \$99,769.0 Million by 2025 (2018). at <https://www.alliedmarketresearch.com/press-release/3D-NAND-flash-memory-market.html>.
- Tanaka, H. et al. Bit cost scalable technology with punch and plug process for ultra high density flash memory. Proceedings of 2007 IEEE Symposium on VLSI Technology. Kyoto, Japan: IEEE, 2007, 14-15.
- Parat, K. & Dennison, C. A floating gate based 3D NAND technology with CMOS under array. Proceedings of 2015 IEEE International Electron Devices Meeting (IEDM). Washington, DC, USA: IEEE, 2015, 48-51.
- Whang, S. et al. Novel 3-dimensional dual control-gate with surrounding floating-gate (DC-SF) NAND flash cell for 1Tb File Storage Application. Proceedings of 2010 IEEE International Electron Devices

- Meeting. San Francisco, CA, USA: IEEE, 2010, 668-671.
13. Jang, J. et al. Vertical cell array using TCAT (Terabit Cell Array Transistor) technology for ultra high density NAND flash memory. Proceedings of 2009 Symposium on VLSI Technology. Honolulu, HI, USA: IEEE, 2009, 192-193.
 14. Sinha, A. K., Levinstein, H. J. & Smith, T. E. Thermal stresses and cracking resistance of dielectric films (SiN, Si₃N₄, and SiO₂) on Si substrates. *Journal of Applied Physics* **49**, 2423-2426 (1978).
 15. Singh, H. Overcoming challenges in 3D NAND volume manufacturing. *Solid State Technology* **60**, 18-21 (2017).
 16. Miyaji, K. et al. Control gate length, spacing, channel hole diameter, and stacked layer number design for bit-cost scalable-type three-dimensional stackable NAND flash memory. *Japanese Journal of Applied Physics* **53**, 024201 (2014).
 17. Orji, N. G. et al. Metrology for the next generation of semiconductor devices. *Nature Electronics* **1**, 532-547 (2018).
 18. Brown, K. A. et al. Machine learning in nanoscience: big data at small scales. *Nano Letters* **20**, 2-10 (2020).
 19. Kang, K. et al. Layer-by-layer assembly of two-dimensional materials into wafer-scale heterostructures. *Nature* **550**, 229-233 (2017).
 20. Ohashi, T. et al. Precise measurement of thin-film thickness in 3D-NAND device with CD-SEM. *Journal of Micro/Nanolithography, MEMS, and MOEMS* **17**, 024002 (2018).
 21. Abdulhalim, I. Simplified optical scatterometry for periodic nanoarrays in the near-quasi-static limit. *Applied Optics* **46**, 2219-2228 (2007).
 22. Abdulhalim, I. Spectroscopic interference microscopy technique for measurement of layer parameters. *Measurement Science and Technology* **12**, 1996-2001 (2001).
 23. Likhachev, D. V. Efficient thin-film stack characterization using parametric sensitivity analysis for spectroscopic ellipsometry in semiconductor device fabrication. *Thin Solid Films* **589**, 258-263 (2015).
 24. Hilfiker, J. N. et al. Spectroscopic ellipsometry characterization of multilayer optical coatings. *Surface and Coatings Technology* **357**, 114-121 (2019).
 25. Hilfiker, J. N. et al. Survey of methods to characterize thin absorbing films with spectroscopic ellipsometry. *Thin Solid Films* **516**, 7979-7989 (2008).
 26. Nazarov, A., Ney, M. & Abdulhalim, I. Parallel spectroscopic ellipsometry for ultra-fast thin film characterization. *Optics Express* **28**, 9288-9309 (2020).
 27. McGahan, W. A., Johs, B. & Woollam J. A. Techniques for ellipsometric measurement of the thickness and optical constants of thin absorbing films. *Thin Solid Films* **234**, 443-446 (1993).
 28. Polgár, O. et al. Comparison of algorithms used for evaluation of ellipsometric measurements random search, genetic algorithms, simulated annealing and hill climbing graph-searches. *Surface Science* **457**, 157-177 (2000).
 29. Fried, M. & Masa, P. Backpropagation (neural) networks for fast pre-evaluation of spectroscopic ellipsometric measurements. *Journal of Applied Physics* **75**, 2194-2201 (1994).
 30. Rédei, L. et al. A modified learning strategy for neural networks to support spectroscopic ellipsometric data evaluation. *Thin Solid Films* **313-314**, 149-155 (1998).
 31. Battie, Y. et al. Demonstration of the feasibility of a complete ellipsometric characterization method based on an artificial neural network. *Applied Optics* **48**, 5318-5323 (2009).
 32. Macleod, H. A. *Thin-Film Optical Filters*. (New York: Elsevier, 1969).
 33. Lissberger, P. H. Optical applications of dielectric thin films. *Reports on Progress in Physics* **33**, 197-268 (1970).
 34. Bhattacharyya, D. et al. Spectroscopic ellipsometry of multilayer dielectric coatings. *Vacuum* **60**, 419-424 (2001).
 35. Tikhonravov, A. V. et al. Optical characterization and reverse engineering based on multiangle spectroscopy. *Applied Optics* **51**, 245-254 (2012).
 36. Pervak, V. et al. 1.5-octave chirped mirror for pulse compression down to sub-3 fs. *Applied Physics B* **87**, 5-12 (2007).
 37. Pervak, V. et al. Dispersive mirror technology for ultrafast lasers in the range 220-4500 nm. *Advanced Optical Technologies* **3**, 55-63 (2014).
 38. Siqueira, J. R. Jr. et al. Penicillin biosensor based on a capacitive field-effect structure functionalized with a dendrimer/carbon nanotube multilayer. *Biosensors and Bioelectronics* **25**, 497-501 (2009).
 39. Ferreira, M. et al. Enzyme-mediated amperometric biosensors prepared with the Layer-by-Layer (LbL) adsorption technique. *Biosensors and Bioelectronics* **19**, 1611-1615 (2004).
 40. Morais, P. V. et al. Nanofilm of ZnO nanocrystals/carbon nanotubes as biocompatible layer for enzymatic biosensors in capacitive field-effect devices. *Journal of Materials Science* **52**, 12314-12325 (2017).
 41. Poddubny, A. et al. Hyperbolic metamaterials. *Nature Photonics* **7**, 948-957 (2013).
 42. Maas, R., van de Groep, J. & Polman, A. Planar metal/dielectric single-periodic multilayer ultraviolet flat lens. *Optica* **3**, 592-596 (2016).
 43. Novak, R. et al. Sensitivity and generalization in neural networks: an empirical study (2018). at <https://arxiv.org/abs/1802.08760>.
 44. Jiang, Y. L. et al. A study of the effect of noise injection on the training of artificial neural networks. Proceedings of 2009 IEEE International Joint Conference on Neural Networks. Atlanta, GA, USA: IEEE, 2009, 1428-1432.
 45. Oviedo, F. et al. Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. *npj Computational Materials* **5**, 60 (2019).
 46. Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. (Cambridge: MIT Press, 2012).
 47. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence. Montréal Québec, Canada: ACM, 1995, 1137-1143.
 48. Smola, A. J. & Scholkopf, B. A tutorial on support vector regression. *Statistics and Computing* **14**, 199-222 (2004).
 49. Cheng, B. & Titterton, D. M. Neural networks: a review from a statistical perspective. *Statistical Science* **9**, 2-30 (1994).
 50. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).
 51. Le, Q. V. et al. On optimization methods for deep learning. Proceedings of the 28th International Conference on Machine Learning. Bellevue, Washington, USA: ACM, 2011, 265-272.
 52. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830 (2011).
 53. Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems (2015). at <https://arxiv.org/abs/1603.04467>.
 54. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. Lille, France: ACM, 2015, 448-456.
 55. Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2011, 315-323.
 56. Srivastava, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15**, 1929-1958 (2014).
 57. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA, 2015.