PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Characterization of asymmetries in 3D NAND memory devices

Jie Li, Shashank Srivastava, Joyce Li, Zhuo Chen, Petar Žuvela, et al.

Jie Li, Shashank Srivastava, Joyce Li, Zhuo Chen, Petar Žuvela, Boyang Chor, Jinyu Deng, Haodong Qiu, YaChing Chang, Sadao Takabayashi, Xadric Yiew, Bo Hui Ng, Rohit Kothari, Dan Engelhard, Han Yang Tan, "Characterization of asymmetries in 3D NAND memory devices," Proc. SPIE 12496, Metrology, Inspection, and Process Control XXXVII, 124960F (27 April 2023); doi: 10.1117/12.2658148

SPIE.

Event: SPIE Advanced Lithography + Patterning, 2023, San Jose, California, United States

Characterization of Asymmetries in 3D NAND Memory Devices

Jie Li[1*], Shashank Srivastava[1], Joyce Li[1], Zhuo Chen[1], Petar Žuvela[1], Boyang Chor[1], Jinyu Deng[1], Haodong Qiu[1], and YaChing Chang[1]

Sadao Takabayashi[2], Xadric Yiew[2], Bo Hui Ng[2], Rohit Kothari[2], Dan Engelhard[2], and Han Yang Tan[2]

[1]Onto Innovation, Inc., 1550 Buckeye Dr., Milpitas, CA, USA 95035-7418; [2]Micron Technology, Inc., 8000 South Federal Way, Post Office Box 6, Boise, ID, USA 83707-0006

ABSTRACT

The adoption of tier stacking (dual deck) leads to increasingly high aspect ratios and poses challenges in controlling overlay, tilt, and misalignment in the manufacturing processes for next generation 3D NAND devices. In this work we address metrology challenges such as tilt and overlay separation, measurement robustness influenced by process variation, and nonlinearity of spectral response to asymmetries. We show that Mueller measurement can separate overlay and tilt signals through distinct spectral response analyzed by a machine learning method with reference data. To reduce asymmetry measurement errors caused by process variation such as critical dimension (CD) and thickness changes, we propose and demonstrate improvement of tilt measurements on blind test wafers by feeding forward CD measurement results to the analysis of tilt signal. We also investigate nonlinear regression and show its capability to extend overlay measurement limit from linear response range, ±0.25pitch, to ±0.43pitch. In addition, for small structural asymmetries introduced by channel hole tilt, test RMSE is reduced by 20–40% from nonlinear regression alone or combined with CD feed-forward. We demonstrate that spectroscopic Mueller matrix measurements, paired with advanced machine learning analysis, provide nondestructive and accurate measurement of tilt, overlay, and misalignment for 3D NAND devices with high throughput and fast recipe creation.

Keywords: 3D NAND memory, asymmetry, ellipsometry, machine learning, Mueller matrix, overlay, tilt

1. INTRODUCTION

Continuous scaling of 3D NAND flash memory capacity requires stacking memory cells vertically in a direction perpendicular to the substrate. In these memory structures, cell density increases with the number of layers in the stack. As more layers are added, the high aspect ratio (HAR) etch process becomes more difficult. A dual deck approach has been invented and implemented beyond 128-layer devices [1]. The HAR etch process challenges are addressed by splitting the massive stack into two decks, that is, replacing one single etch step with two more manageable etch steps. Adoption of tier stacking (dual deck) leads to increasingly high aspect ratios and poses metrology challenges [2] in controlling overlay, tilt, and misalignment in the manufacturing processes for next generation 3D NAND devices.

For dual-deck 3D NAND devices, the joint between the two decks needs to be well aligned to ensure proper downstream processing and robust electrical contact. Fig. 1 shows the two major sources of errors, overlay and tilt, which contribute to the misalignment of the bottom of the top channel hole (Deck 2) and the top of the bottom channel hole (Deck 1). The overlay registration error between the two decks is controlled by the lithographic process when Deck 2 is patterned. The tilt of the channel holes is introduced in the HAR etch process and is often a bigger contributor to the misalignment at the joint of the top and bottom decks if not controlled well. A nondestructive, fast, and robust metrology solution is required to measure and monitor overlay and tilt in a manufacturing environment to minimize misalignment yield loss.

* jie.li@ontoinnovation.com; phone 1 (408) 545-6225

Metrology, Inspection, and Process Control XXXVII, edited by John C. Robinson, Matthew J. Sendelbach, Proc. of SPIE Vol. 12496, 124960F
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2658148

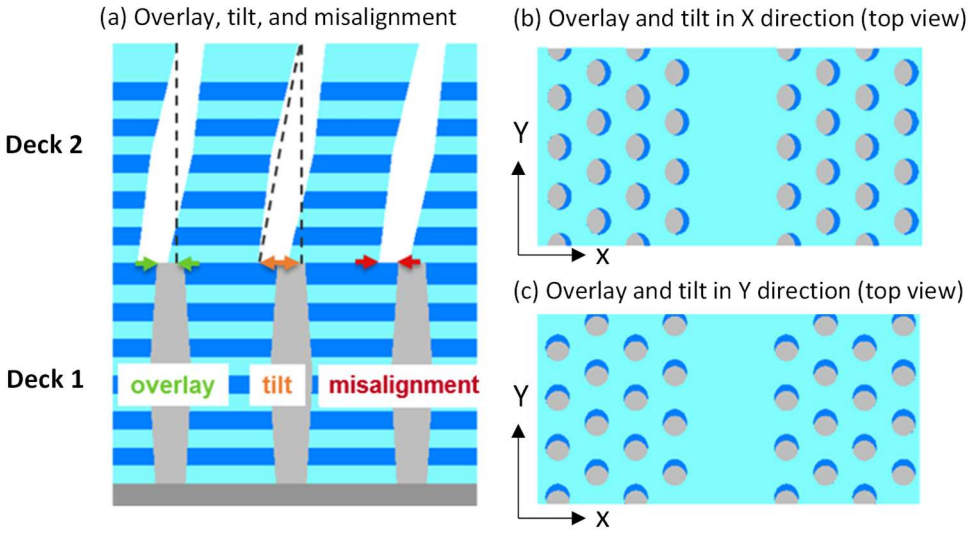


Figure 1. Overlay, tilt, and misalignment for a 3D NAND dual deck structure.

In this work, we apply spectroscopic Mueller matrix measurements to characterize structural asymmetries introduced by overlay and tilt in 3D NAND memory devices. We address metrology challenges such as tilt and overlay separation, measurement robustness influenced by process variation, and nonlinearity of spectral response to asymmetries. In Section two, we introduce experimental setup and data analysis methodology. In Section three, we present theoretical simulation and experimental results to demonstrate that broadband optical full Mueller matrix measurement is capable of differentiating overlay and tilt in a 3D NAND dual deck device. We illustrate that machine learning data analysis can produce overlay and tilt predictions of high accuracy. In Section four, we discuss how asymmetry measurement could be compromised in presence of process variations and propose a data feed-forward approach to mitigate the problem. In Section five, we study and apply nonlinear models to analyze Mueller signals. We show benefits of nonlinear regressors for three cases. In Section six, we summarize our work and discuss the capability of spectroscopic Mueller matrix measurements as an in-line metrology solution to monitor channel hole tilt and overlay for 3D NAND memory device wafer fabrication.

2. METHODOLOGY

2.1 Experimental setup

The experimental apparatus used in this study is the Onto Innovation Atlas® platform that consists of a focused beam rotating compensator spectroscopic ellipsometer (J. A. Woollam Company Model RC2). With a dual rotating compensator configuration, we can measure all 16 Mueller Matrix elements with available wavelengths in 195–1700nm range. Data collection is performed in the specular mode at 65° angle of incidence, and the plane of incidence (POI) can be adjusted from 0° to 360° for optimal spectral sensitivity.

Under the Stokes-Mueller formulism, the effect of an optical component or a sample can be described by a 4x4 Mueller matrix:

$$MM = \begin{pmatrix} M_1^{(2x2)} & M_2^{(2x2)} \\ M_3^{(2x2)} & M_4^{(2x2)} \end{pmatrix} = \left(\begin{array}{cc:cc} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ \hdashline m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{array}\right) \quad (1)$$

In the absence of depolarization, as it is the case with our samples, the sample-light interaction can also be described by the Jones reflection matrix:

$$J = \begin{pmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix} \quad (2)$$

The diagonal elements describe the complex reflectance (amplitude and phase) for polarization orthogonal (r_{pp}) and parallel (r_{ss}) to POI. The off-diagonal terms r_{ps} and r_{sp} are related to polarization conversion between s and p polarization states in the presence of sample anisotropy. The 4x4 Mueller matrix is related to the Jones matrix by Eq. 3:

$$MM = \begin{bmatrix} (r_{ss}^* r_{ss} + r_{pp}^* r_{pp} + r_{sp}^* r_{sp} + r_{ps}^* r_{ps})/2 & (-r_{ss}^* r_{ss} + r_{pp}^* r_{pp} + r_{sp}^* r_{sp} - r_{ps}^* r_{ps})/2 & Re(r_{pp}^* r_{ps} + r_{sp}^* r_{ss}) & -Im(r_{pp}^* r_{ps} + r_{sp}^* r_{ss}) \\ (-r_{ss}^* r_{ss} + r_{pp}^* r_{pp} - r_{sp}^* r_{sp} + r_{ps}^* r_{ps})/2 & (r_{ss}^* r_{ss} + r_{pp}^* r_{pp} - r_{sp}^* r_{sp} - r_{ps}^* r_{ps})/2 & Re(r_{pp}^* r_{ps} - r_{sp}^* r_{ss}) & Im(-r_{pp}^* r_{ps} + r_{sp}^* r_{ss}) \\ Re(r_{pp}^* r_{sp} + r_{ps}^* r_{ss}) & Re(r_{pp}^* r_{sp} - r_{ps}^* r_{ss}) & Re(r_{pp}^* r_{ss} + r_{ps}^* r_{sp}) & Im(-r_{pp}^* r_{ss} + r_{ps}^* r_{sp}) \\ Im(r_{pp}^* r_{sp} + r_{ps}^* r_{ss}) & Im(r_{pp}^* r_{sp} - r_{ps}^* r_{ss}) & Im(r_{pp}^* r_{ss} + r_{ps}^* r_{sp}) & Re(r_{pp}^* r_{ss} - r_{ps}^* r_{sp}) \end{bmatrix} \quad (3)$$

The Mueller elements are sensitive to the profile details of the structures as well as any asymmetries present. In the most general configuration where POI is chosen at arbitrary angles, the Mueller matrix is fully dense. The off-diagonal blocks $M_2^{(2x2)}$ and $M_3^{(2x2)}$ are nonzero due to the intentionally introduced asymmetry from nonsymmetric hardware-sample configuration at arbitrary azimuth angles. For intrinsically symmetric structures, however, the cross-reflection coefficients are anti-symmetric, that is, $r_{sp} = -r_{ps}$ [3]. From Eq. 3, it is seen that the following relationships in Eq. 4 are valid for m_{13}, m_{23}, m_{14}, m_{24}, and their corresponding transposed elements:

$$m_{13} + m_{31} = 0,\ m_{23} + m_{32} = 0,\ m_{14} - m_{41} = 0,\ m_{24} - m_{42} = 0 \quad (4)$$

When the structural symmetry is broken, e.g. due to tilt and overlay (Fig. 1), we have $r_{sp} \neq -r_{ps}$, and the relationships in Eq. 4 are violated. For a small amplitude of structural asymmetry, we expect that the composite signals from the Mueller off-diagonal 2 x 2 blocks, that is, $m_{13} + m_{31}$, $m_{23} + m_{32}$, $m_{14} - m_{41}$ and $m_{24} - m_{42}$ respond to asymmetry linearly:

$$M_c = S * A + \text{noise} \quad (5)$$

where M_c represents the four composite signals from Mueller off-diagonal blocks, S is the pre-factor or the slope of the linear response, and A represents asymmetry such as tilt and overlay. Because the linear combination of corresponding Mueller elements from off-diagonal blocks $M_2^{(2x2)}$ and $M_3^{(2x2)}$ cancels the cross-polarization signals due to nonsymmetric hardware-sample configuration and only responds to the intrinsic structural asymmetries, we will always use composite Mueller signals in our data analysis.

2.2 Data analysis

Machine learning (ML) has been considered and adopted in optical metrology for monitoring critical dimensions, thickness, trench depth, and so on for quality control and yield improvement in semiconductor manufacturing processes [4], [5]. The major benefit of ML is to bypass electromagnetic simulation required by the conventional physical modeling approach [6], so recipe creation time is greatly reduced. ML has limitations though. It typically requires a larger number of reference data than conventional physical modeling due to lack of physics constraints [5]. Otherwise, it could suffer from overfitting with degraded measurement accuracy on unseen wafers. Mueller matrix has a unique response to structural asymmetries such as tilt and overlay (Eq. 5). Additional overlay references can be obtained from image-based overlay (IBO) and high-density tilt references can be acquired with destructive imaging techniques. These make ML an ideal candidate to analyze Mueller matrix spectra and make tilt and overlay predictions. In this study, we perform ML data analysis with linear or nonlinear regressors using all four Mueller composite signals as model inputs. Cross validation is applied to optimize hyper parameters such as number of features, regularization, etc. based on validation errors. We leave out some labeled data used as blind test data that are not used in any sense during ML model training. The blind test data measurement performance is an indication of how the ML model can be generalized to predict unseen targets.

3. TILT AND OVERLAY SEPARATION

3.1 Simulation results

In 3D NAND channel hole dual deck structures, both overlay and tilt (Fig. 1) contribute to structural asymmetries. One basic question is whether Mueller matrix experiments can differentiate overlay and tilt in a single measurement. To address this, we performed electromagnetic simulations with Rigorous coupled-wave analysis (RCWA) formulation using Onto Innovation's Ai Diffract™ modeling software on a dual-deck 3D NAND device with ~ 200 silicon dioxide/silicon nitride (ON) pairs. For X tilt and X overlay, plane of incidence (POI) is along y-axis, and for Y tilt and Y overlay, POI is along x-axis (Fig. 1). With this configuration, Mueller off-diagonal blocks respond to the asymmetry in one direction only and have zero response to the asymmetries along the orthogonal direction. In the simulation, overlay and tilt are changed by ±5nm in x- and y-directions, one at a time, and the sensitivity spectra is calculated as the spectral difference between the spectra with no asymmetry (that is, X_Tilt = Y_Tilt = X_Overlay = Y_Overlay = 0) and the spectra with asymmetry of 5nm amplitude. Fig. 2 displays sensitivity spectra of four Mueller off-diagonal elements m_{13}, m_{14}, m_{23}, and m_{24} that correspond to ±5nm in X_Overlay, Y_Overlay, X_Tilt, and Y_Tilt, respectively. The asymmetry signals spread across the full spectrum from UV to IR. For a given asymmetry parameter, 5nm signal is identical to –5nm signal except the sign is flipped. Previous studies showed that these Mueller elements are identically zero when there are no structural asymmetries and have linear response for small amplitude of asymmetries such as tilt and overlay [7]. This unique property makes Mueller matrix a good candidate for asymmetry measurement with simplified data analysis using machine learning. A closer look at sensitivity spectra of tilt and overlay in x- or y-directions shows that they have distinct spectra features. The spectral correlation of tilt and overlay signals is very low, as shown in Fig. 3. The position of each data point on the scatter plots indicates values of tilt and overlay spectral response at the same wavelength. If tilt and overlay spectra response are similar, the data points will closely follow a straight line. For both x- and y-directions, the data points do not follow any clear direction and the shape is almost round indicating low correlation between tilt and overlay signals. Wavelength by wavelength spectral correlation R^2 are 0.015 and 0.005 respectively for x- and y-directions. Although both tilt and overlay contribute to structural asymmetries in 3D NAND channel holes, the broadband spectral features are different from each other for each Mueller element. The simulation results show the great potential of broadband Mueller matrix technology to separate and measure tilt and overlay.

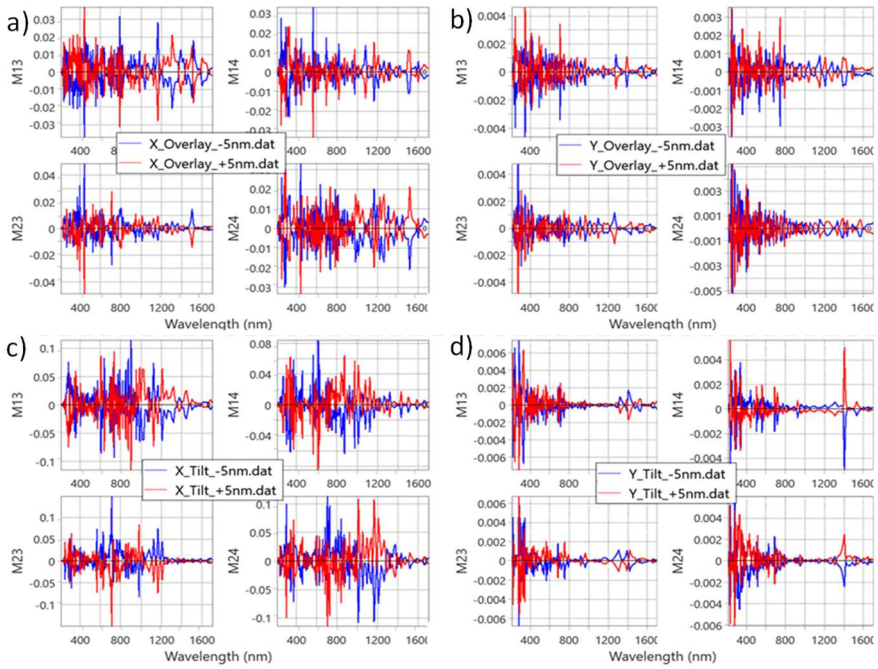


Figure 2. Sensitivity spectra of ±5nm tilt/overlay in x- and y-directions for Mueller elements, M_{13}, M_{14}, M_{23}, and M_{24}.

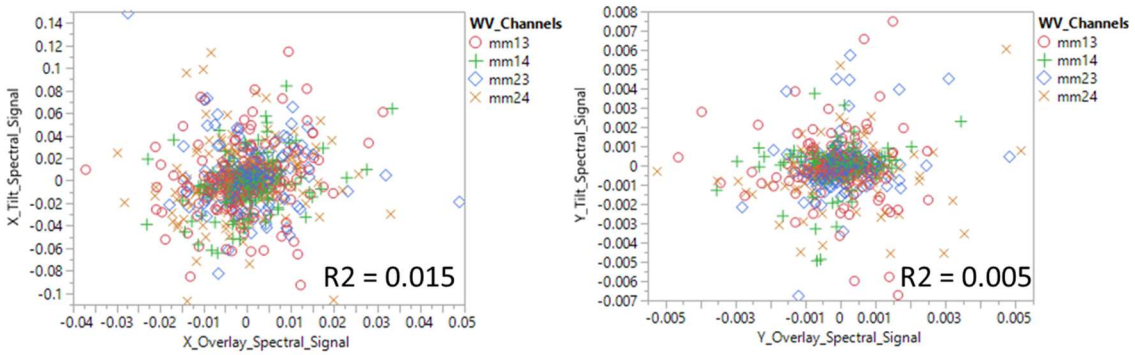


Figure 3. Correlation plots of tilt signal versus overlay signal for x- and y-directions with each data type labeled by different markers. The correlation coefficient R^2 is calculated using all four data types.

3.2 Experimental results

We measured tilt and overlay from four 3D NAND dual deck channel hole wafers using the spectroscopic Mueller matrix technique. Three wafers have skew on channel hole tilt through etch package changes. The intended skew in tilt is expected to mainly occur on the edge of the wafers. One wafer has a field-based overlay skew performed at Deck 2 during the lithographic process step. The programmed overlay was applied on 20 selected fields with four different design types with a magnitude from –30nm to 30nm (Fig. 4). The overlay skew patterns are indicated by the vectors in each field. For inward and outward skew patterns, overlay is designed to be 0nm at the center of the field and increases linearly along x-direction for X overlay and increases linearly along y-direction for Y overlay. At the four corners, overlay reaches its maximum value. For shift down pattern, Y overlay is designed to be –20nm, and X overlay is fixed to 0 at all locations. For shift right pattern, X overlay is designed to be +20nm, and Y overlay is fixed to 0 at all locations.

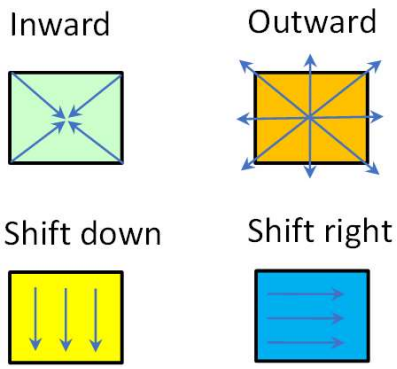


Figure 4. Four overlay skew design patterns within a field.

Mueller matrix spectra were collected on device using the Atlas platform with four plane of incidence (POI) angles, 0º, 90º, 180º, and 270º. Tool-induced shift (TIS) correction was performed using 0º and 180º spectra for Y tilt and Y overlay, and 90º and 270º spectra for X tilt and X overlay. IBO reference data were measured on designed overlay marks from scribe lines and used as reference. In each field, spectra were collected from 15 different locations. The locations were selected to be on device, but close to the IBO targets in the scribe lines. There are totally about 1,200 spectra and IBO references collected from the overlay skewed wafer. X overlay and Y overlay are trained separately with linear models, with one third of the data used as training samples and the other two thirds used as blind test samples. After ML hyperparameter optimization, good correlation (R^2 and slope) between ML predictions and IBO reference are achieved for X overlay and Y overlay for both training and blind test samples (Fig. 5).

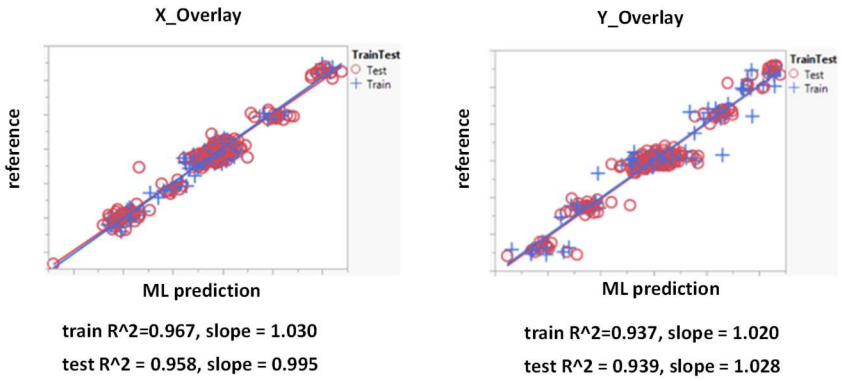


Figure 5. Correlation of machine learning prediction and IBO reference for X_Overlay and Y_Overlay measured from the overlay skew wafer.

For tilt measurement, linear machine learning models are trained with Mueller spectra and destructive imaging reference collected from the overlay skew wafer with 1:1 train/test ratio. Fig. 6 shows the correlation plots between ML predictions and tilt reference. Both training and test results show high correlation and consistency to reference data. The ML predicted tilt wafer maps and radial plots for all four wafers are shown in Fig. 7. The wafer in the green square is used to train the tilt ML models. It is seen that the predicted tilt radial profiles for all four wafers are very similar to each other, with the only difference showing up at the extreme edge. The tilt wafer maps are also very similar to each other. The results from Figs. 5–7 confirm that the Mueller matrix technique can differentiate overlay and tilt signals and make accurate measurements. The broadband Mueller matrix spectra carry rich information about tilt and overlay with distinguishable spectral features (Figs. 2 and 3). With appropriate reference data provided, the ML model is trained to learn the underlying spectral patterns and make correct predictions for overlay or tilt. High quality reference data are required for machine learning offline training. After validation from blind test samples, the ML model can be deployed for inline measurements with no reference data needed.

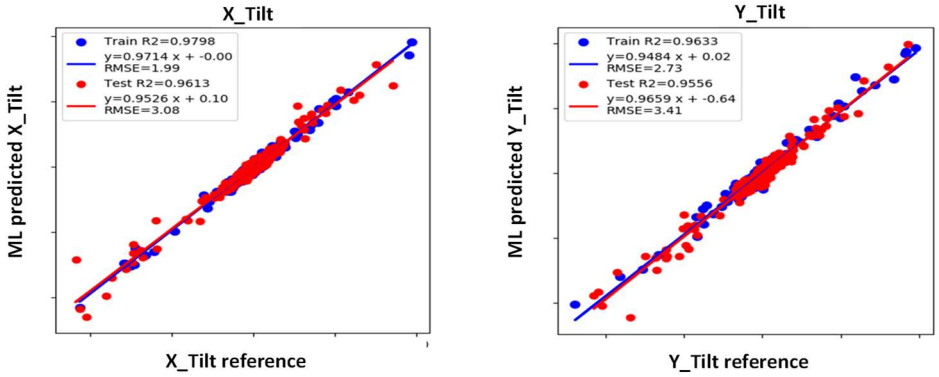
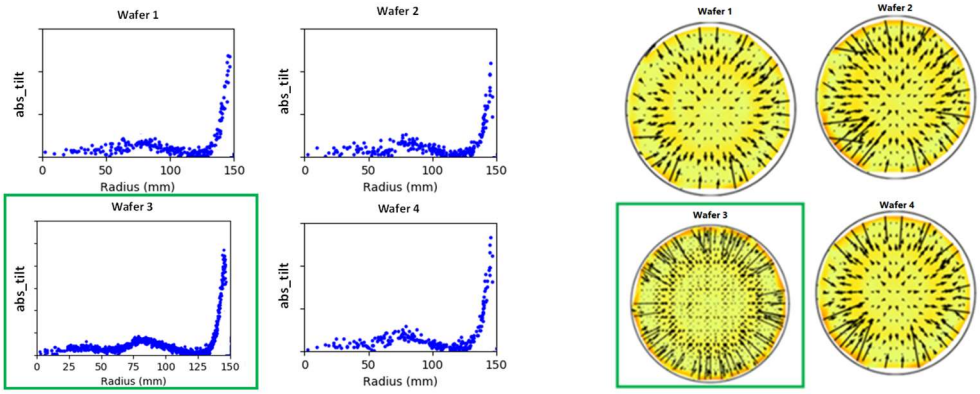


Figure 6. Correlation of machine learning prediction and destructive imaging reference for X_Tilt and Y_Tilt.



(a) (b)

Figure 7. a) Tilt radial maps and b) tilt quiver wafer maps. The wafer with an overlay skew is highlighted in a green box and used to train tilt ML models. The other three tilt wafers are used as blind test samples.

4. ECOSYSTEM DATA FEED-FORWARD

4.1 Motivation

In the previous section, we showed that spectroscopic ellipsometry with Mueller matrix measurements in UV to near IR spectral range with a recipe trained with simple linear machine learning models can predict structural asymmetries such as tilt and overlay with reasonable accuracy. However, the asymmetry prediction is subject to larger errors if test wafers have different process conditions that are not covered by training samples because the pre-factor S (Eq. 5) is a function of structural parameters such as channel hole profile, critical dimension, and thickness. It can no longer be considered as constant if test samples have significant different process conditions from training samples, so measurement accuracy for test samples will suffer.

Recently, mid-IR optical metrology [8] with Onto Innovation's Aspect® platform has shown success and outperforms the Atlas (UV-VIS) platform for CD profile measurements for high aspect ratio channel holes in 3D NAND memory manufacturing. Here we propose an Aspect and Atlas platform ecosystem to improve tilt and overlay measurement robustness by feeding forward CD measurement results from the Aspect system to the Atlas system's tilt and overlay measurements. The CDs are used as machine learning model input features to compensate the impact of process variations and reduce the prediction errors for asymmetry.

4.2 Simulation results

To study the impact of process condition changes on tilt measurement robustness, we perform theoretical simulation on a 3D NAND channel hole single deck structure. 4,500 Mueller matrix spectra in the wavelength of 200nm to 1700nm are calculated using RCWA electromagnetic simulation with varied CDs, thin film thickness, and tilt within predefined ranges. Fig. 8 shows the structure and floating parameters used in the simulation. For the X tilt study, an azimuth angle of 90º is applied so that Mueller off-diagonal elements respond to tilt in the x-direction only. In the simulation test, we focus on the influence of channel hole CD profile change on X tilt measurement accuracy. The observation and conclusion should also apply to Y tilt.

Linear machine learning models are trained on calculated Mueller off-diagonal signals to predict X tilt parameter. The X tilt value applied to calculate each spectrum is used as reference to train the ML model. We intentionally split the training and blind test samples based on channel hole CDs. Simulated spectra with average CD in the range of 90nm to 115nm are used as training samples, and spectra with average CD in the range of 80nm to 125nm are used as blind test samples. For a linear model baseline, Mueller composite signal $m_{13} + m_{31}$, $m_{23} + m_{32}$, $m_{14} - m_{41}$ and $m_{24} - m_{42}$ are used as ML input features. To compensate for impact of CD variations on tilt results, another linear model is trained with four Mueller composite signals and twelve tier CDs as input features. The optimized results for both ML models are displayed in Fig. 9.

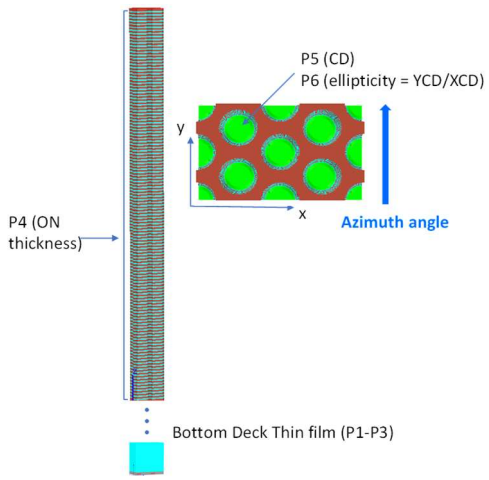


Figure 8. 3D NAND channel hole structure and floating parameters used to generate calculated spectra and perform simulation on X_Tilt measurement.

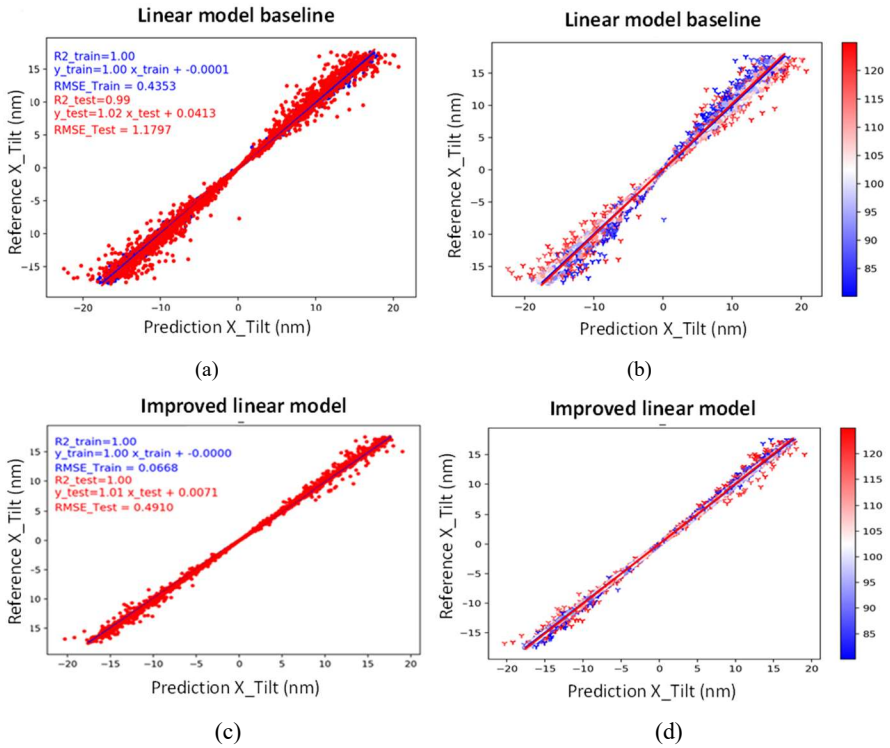


Figure 9. Correlation plots for X-tilt between ML predictions and reference using simulated spectra. a) Train and test correlation for linear model with spectra as ML inputs, b) same results as a) but with samples colored based on CD values, c) train and test correlation for linear model with spectra and twelve CDs as ML inputs, and d) same results as (c) but with samples colored based on CD values.

The optimized baseline results for train and test samples are shown in Fig. 9a. Although the test sample CD range is outside the training sample range, X tilt test R^2 is very reasonable, 0.99 because simulation range of X tilt is relatively large, and the simulated spectra and labels are perfect with no noises. However, test root mean square error (RMSE) is 1.1797nm, 2.7 times worse than training RMSE (0.4353nm). Test sample points are much more scattered than training samples on the correlation plot (Fig. 9a). The deviation of predicted results from actual value (reference) is worse for samples of larger tilt magnitudes. This is expected since the slope S from Eq. 5 varies with the volume of the features including channel hole

CDs. For simulation data, set noise = 0 in Eq. 5 and taking the derivative with respect to slope S and asymmetry A leads to the following equation:

$$dA = -(dS/S) * A \quad (6)$$

From Eq. 6, the asymmetry error dA introduced by the error from slope dS is proportional to the amplitude of asymmetry, that is, the error bar on the linear fit is zero at origin and increases with growing asymmetry amplitude toward both ends of the linear line. The observed behavior in the correlation plot is consistent with what is indicated by Eq. 6. Fig. 9b shows the same results as Fig. 9a with data points colored based on CD values. The samples with darker blue and darker red colors are test samples with CDs <90nm or >115nm, that is, outside training sample CD range. They have larger prediction errors than the test samples colored as light blue and light red with CDs within train sample range. Fig. 9c shows the optimized results with improved ML model by adding CDs as input features. ML prediction for training samples is greatly improved to nearly perfect correction to reference with $R^2 = 1.00$ and RMSE = 0.0668nm, compared with baseline training RMSE of 0.4353nm. Test R^2 is improved from 0.99 (baseline) to 1.0 and test RMSE is improved from 1.1797nm (baseline) to 0.4910nm. Comparing Figs. 9b and 9d, the improved ML model mostly reduce big prediction errors for samples with CDs that largely deviate from the center of the CD range. This confirms that tilt measurement errors due to CD variations from process change can be at least partially reduced by the proposed CD feed-forward method. The results could be further improved by using more complex ML models with carefully designed feature engineering to select and transform input variables for better predictive power.

4.3 Experimental results

We performed CD and tilt measurements for seven 3D NAND dual deck channel hole wafers immediately after etching and cleaning processes. Table I displays the wafer process conditions and train/test split. To check if CD feed-forward can compensate the pre-factor change (Eq. 5) in the presence of process condition variations, we carefully design how training and test wafers are split based on process conditions (Table I). We use three CD process of record (POR) wafers from etch package 1 as training wafers to optimize ML models for tilt prediction. The training wafers have intentionally introduced tilt skew so that the ML model can cover a larger range of tilt with better generalizability for blind test wafers. Four wafers processed from a different etch package with CD skews, POR, POR+ and POR–, are used as test wafers. The test wafers are expected to have different CD profiles from training wafers so that tilt accuracy will be degraded due to profile change.

Table 1. Wafer process conditions and train/test split for tilt measurement with CD feed-forward. The wafers were processed from two different etch packages and have three conditions, process of record (POR), POR+ and POR– for CD and tilt. The test wafers are chosen with process conditions different from train wafers for data feed-forward study.

Scribe	CD Skew	Tilt Skew	Etch Package	Number of tilt measurements	Train/test condition
Wafer 1	POR	POR	Etch1	1933	train
Wafer 2	POR+	POR	Etch1	1919	test
Wafer 3	POR-	POR	Etch1	1923	test
Wafer 4	POR	+	Etch1	1928	train
Wafer 5	POR	–	Etch1	1919	train
Wafer 6	POR+	POR	Etch2	1916	test
Wafer 7	POR-	POR	Etch2	1935	test

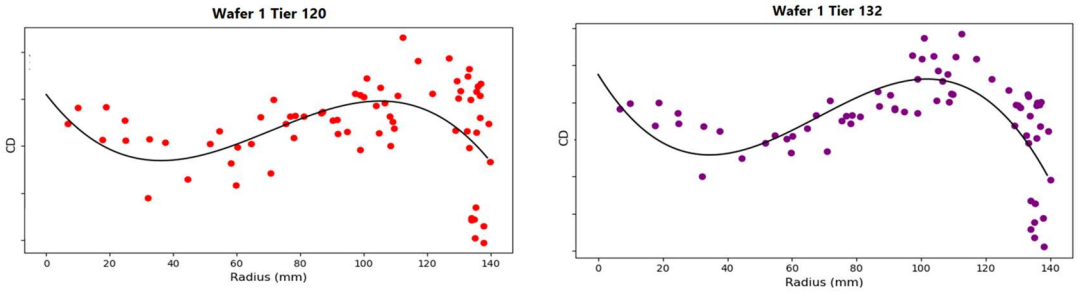


Figure 10. Polynomial interpolation to CD results measured with mid-IR Mueller matrix.

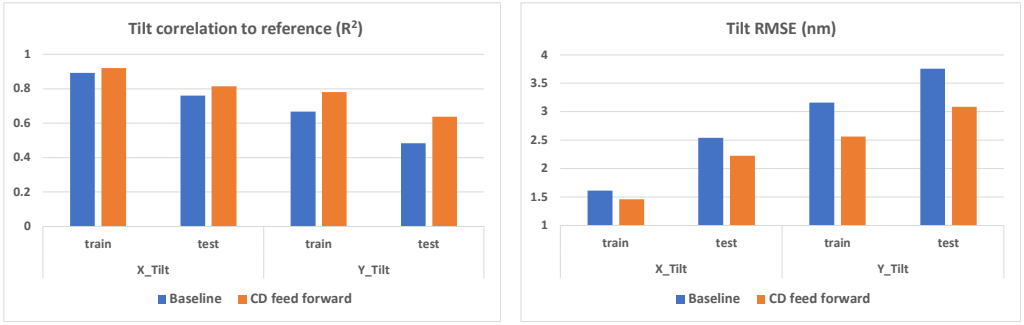


Figure 11. Correlation of X tilt and Y tilt ML predictions to reference for baseline and CD feed-forward.

We first measure channel hole CDs using the Aspect platform with standard optical critical dimension (OCD) physical modeling data analysis. Twelve CDs are measured from top to bottom of the channel hole for 80 dies across each wafer. Channel hole tilts in x- and y-directions are measured on the Atlas platform with much denser sampling, approximately 1,900 points per wafer. We perform radial based interpolation on Aspect system CD results to provide data sources for data feed-forward to Atlas system tilt measurement. Fig. 10 shows two examples of tier CDs from wafer 1 as function of measurement location in term of radius from center of wafer. A third order polynomial fit is performed to capture the trend of CDs as function of radius. Interpolated CD values are used to match the measurement location for tilt. For ML baseline, only spectra are used as ML inputs to train a linear model using destructive imaging reference. The measurement locations from the Atlas platform are on device and selected to be adjacent to the destructive imaging measurement locations. To improve tilt measurement results, the Aspect system CD values after interpolation to match tilt measurement locations are added as model input features for ML training and prediction.

The tilt measurement results from linear model ML prediction with and without CD feed-forward are displayed in Fig 11. The ML prediction for tilt has better correlation to reference for both training and test wafers, with improved R^2 and reduced RMSE. The improvement with data feed-forward from the experimental study is not as significant as the simulation results, due to several reasons: a) the structure used in simulation is a single deck structure while the measured wafers are from dual deck channel hole. The structural difference can impact how sensitive tilt signal can be as CD profiles change; b) The simulated CD range for training and test samples is different from the actual CD changes in the experimental training and test wafers; c) The simulated spectra and the labels for tilt are noise free. For experimental study, other error sources including but not limited to spectrometer noise, reference uncertainties, unknown process changes are not addressed by CD feed-forward. The simulated results provide theoretical understanding on how channel hole CD profile can impact tilt measurement and whether CD feed-forward can help. The conclusion should be taken qualitatively rather than quantitively. The experimental results further confirm the benefit of applying CD feed-forward, although the quantitative improvement depends on multiple factors that are practically difficult to design and study systematically.

5. NONLINEAR MODEL FOR ASYMMETRY MEASUREMENT

The Mueller matrix elements from off-diagonal blocks respond linearly to small structural deviations from symmetric geometries. With the features extracted from these Mueller elements measured over broad range of wavelengths (UV-VIS-

IR), linear machine learning models with optimized hyperparameters work well for a small range of tilt and overlay. However, under some conditions nonlinear regressors may be required to improve measurement accuracy. When the structural asymmetry is sufficiently large so that the spectral response starts deviating from linear relationship, nonlinear regression model is required to extend the measurement range beyond what the conventional linear ML regression can handle. Furthermore, even within a constrained range, the slope of the linear response (Eq. 5) varies with structural parameters such as CDs and thickness. A more complex nonlinear model could reduce the errors caused by nonlinearity and structural variations. In this section, we demonstrate benefits of applying nonlinear regressors to asymmetry characterization for three 3D NAND use cases.

5.1 Extend overlay measurement limit

We measured X and Y overlay between the bit-line contact hole and the channel hole plug from four 3D NAND wafers, two POR and two overlay skew wafers. Figs. 12a and 12b show the structure under study and the overlay key parameters.

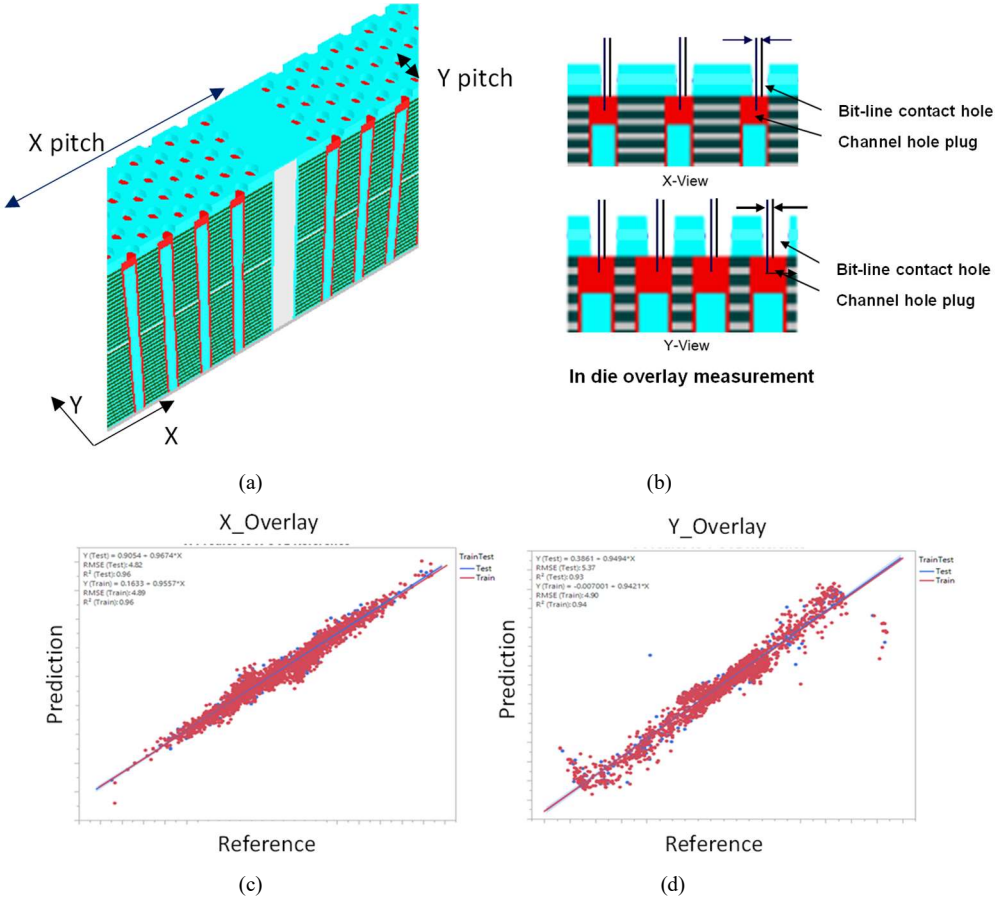
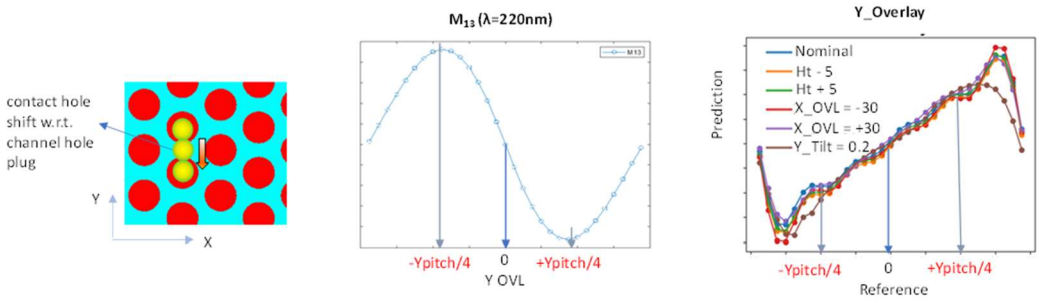


Figure 12. a) Structure under study; b) overlay key parameters for 3D NAND bit-line contact hole to channel hole polysilicon plug, c) correlation of Mueller measurements to reference for X_Overlay; d) correlation of Mueller measurement to reference for Y_Overlay.

The measurement results from Mueller matrix spectra analyzed by linear ML model are shown in Fig. 12c and 12d for X overlay and Y overlay, respectively. The reference data are provided by destructive imaging. There are some outliers observed on Y overlay ML predictions for samples with large overlay errors on both ends of linear fit from the correlation plot as shown in Fig. 12d. These outliers come from Y overlay ++ skew fields on the two skew wafers with overlay magnitude close to or greater than 1/3 of Y pitch. It is known that the Mueller off-diagonal block elements respond linearly to structural asymmetries within a small range and start to distort nonlinearly as magnitude of asymmetries increases approaching half of the pitch [7]. The Y overlay outliers are expected since a linear ML model is applied to infer measurement results while the spectral response is out of linear region. Because X Pitch is nearly 10 times larger than Y

pitch, the range of X overlay values is a very small faction of X pitch. Therefore, the spectral response to X overlay still follows linear relationship and no outliers are observed for X overlay measurement results.

To illustrate the nonlinear response of Mueller matrix, we perform simulation studies on a simplified structure. Fig. 13a shows the top view of the structure with Y overlay as key parameter. POI is chosen to be along x-direction so that the calculated Mueller elements are sensitive to asymmetry along y-direction only. We calculate 31 Mueller matrix spectra with Y overlay varied from –Y_Pitch/2 to +Y_Pitch/2 with all other parameters fixed at nominal values, that is, Y overlay is the only varied parameter and X overlay and tilt are fixed at 0. Fig. 13b plots calculated Mueller element m_{13} as a function of Y overlay over a range of full pitch at wavelength 220nm. The response follows a linear relationship for Y overlay close to 0. The response shows sinusoidal behavior over the full pitch range and nonlinearity becomes obvious when overlay is approaching quarter pitch. All Mueller off-diagonal block elements at different wavelengths show similar behavior. In another study, we train a linear ML model with the 31 simulated spectra using the Y overlay labels as reference. The predicted Y overlay plotted against the reference labels are shown in Fig. 13c with the legend 'Nominal' indicating all other parameters fixed at nominal values. The other five curves in Fig. 13c are from five additional runs with spectra generated by fixing a parameter to a new value. The simulation conditions for the five runs are contact hole height (nominal value ±5nm), X overlay (–30nm and +30nm) and Y tilt (0.2 deg). For the results shown in Fig. 13c, independent linear ML models are trained for the six simulation conditions. Within ± quarter pitch, ML predictions reasonably follow the reference values, and the variation of other parameters has minor influence for Y overlay prediction within the linear range. Beyond quarter pitch, ML predictions from linear models have much worse accuracy. A nonlinear regressor will be required to extend the overlay measurement limit beyond quarter pitch. It is also noticed that Y tilt has much more impact on Y overlay predictions than height and X overlay in the nonlinear region. X overlay contributes to structural asymmetry as well but in an orthogonal direction, so it has minor impact on Y overlay measurement. Out of all the simulated parameters, only Y tilt contributes to structural asymmetry in the same direction as Y overlay, so it is expected to have the largest impact on Y overlay measurement.



(a) (b) (c)

Figure 13. a) Simulation structure with key parameter Y overlay representing contact hole shift with respect to channel hole plug; b) simulated Mueller element m_{13} response to Y overlay over a range of full pitch at wavelength 220nm; c) ML prediction from a linear model versus reference with the ML model trained with Y overlay labels using simulated spectra calculated with varied parameters values for thickness, X/Y overlay, and Y tilt.

To extend overlay measurement limit beyond quarter pitch, we explore the potential benefits of several nonlinear machine learning models including polynomial, support vector machine (SVM) with radial basis kernel function (RBF) and gradient boosted regression (GBR). We first analyze Mueller spectra from one skew wafer with the three nonlinear models and optimize hypermeters associated with each model. Training and test samples are split with a 70:30 ratio. Special care is taken to find out appropriate model complexity for balanced training and test performance with controlled overfitting risk. The optimized results are displayed in Fig. 14 and compared with the linear model. All ML models have similar performance for X overlay. For Y overlay, nonlinear models improve training R^2, while improvement on test R^2 are moderate. GBR shows the greatest potential to reverse the nonlinear spectral response to overlay of large amplitude and reproduce the correct overlay values, so it is further optimized for Y overlay with the data from all four wafers. The optimized GBR results are compared with the linear model in Fig. 15. Both training and test R^2 are noticeably improved by GBR. The outliers from linear regression on both ends of the fit are fixed in the GBR model. Other nonlinear models could also be studied to further improve overlay measurement results over a larger dynamic range.

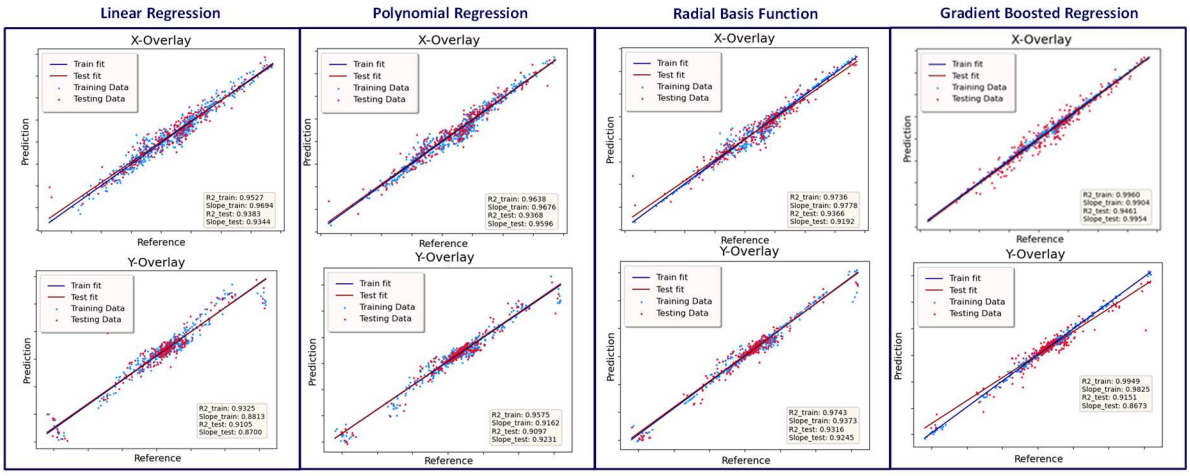


Figure 14. Overlay results of linear model compared with three nonlinear regression models: polynomial regression, SVM with RBF kernel function and gradient boosted regression. All four ML models are trained and optimized using data from one overlay skew wafer.

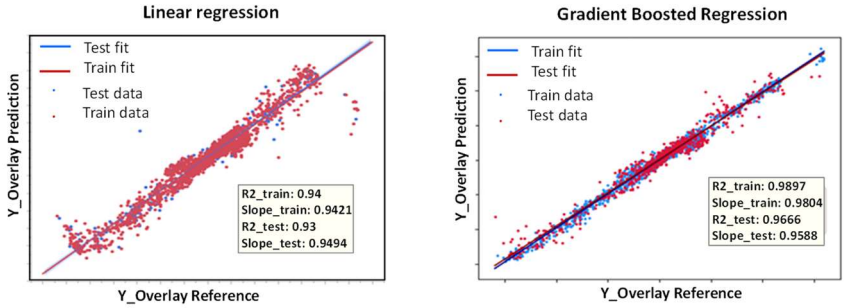


Figure 15. Overlay results of linear regression and gradient boosted regression. Both models are trained and optimized using 70% of the data from two POR and two overlay skew wafers. The remaining 30% are used as test samples.

5.2 Nonlinear regression for tilt measurements

A set of 3D NAND wafers including three POR wafers and two CD skew wafers (POR ±4nm) were measured with ~1500 sites per wafer to test nonlinear regression for tilt measurement. Mueller matrix spectra were measured on channel hole arrays in close proximity to the destructive imaging reference measurement sites. Seventy percent of the sites are randomly selected from five wafers and used as training samples. The rest are used as test samples. Baseline performance is benchmarked using feature selection followed by linear regression. The nonlinear models under test here are fully connected multi-layer neural networks, referred to as Multilayer Perceptron (MLP). We optimize shallow MLP with one hidden layer and deep MLP with three hidden layers for tilt prediction. Hyperparameters such as number of nodes in hidden layers, regularization, learning rate, etc. are optimized to minimize train and test errors. Fig. 16 shows the comparison of MLP and baseline results for X tilt and Y tilt. For both training and test samples, MLP models improved R^2 and reduced RMSE. While deep MLP can improve train errors more than shallow MLP, it does not necessarily improve test errors if there is overfitting.

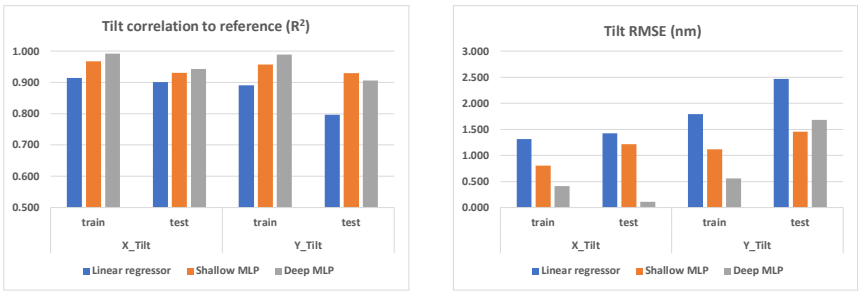
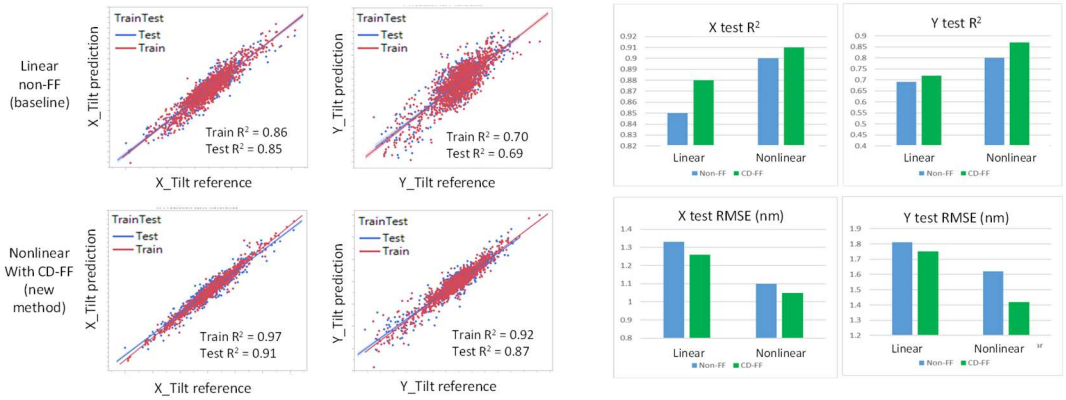


Figure 16. Correlation of tilt prediction and reference for three ML models: linear, shallow MLP and deep MLP. Data are shown as R^2 and RMSE (nm).

5.3 Nonlinear regression with data feed-forward

For 3D NAND HAR etches, tilt is often a bigger contributor to upper and lower deck misalignment than overlay registration error. In previous sections we demonstrated improved tilt measurements by applying CD feed-forward or nonlinear ML models. In this section, we demonstrate tilt performance enhancement by combining these two approaches.

We collected Mueller matrix spectra from five 3D NAND channel hole dual deck POR wafers using the Atlas platform for tilt measurement. The Aspect platform spectra were collected on the same set of wafers and analyzed with physical modeling to provide CD results for data feed-forward. The channel hole tilt for upper deck is predicted by training ML models with reference data provided from destructive imaging metrology. Training samples are randomly selected from 70% sites of all five wafers and the rest are used as test samples. Four algorithms are applied including linear model without CD feed-forward (baseline), linear model with CD feed-forward, nonlinear model without CD feed-forward and nonlinear model with CD feed-forward. Fig. 17a shows correlation plots of ML prediction and reference for X tilt and Y tilt for baseline and proposed new algorithm combining nonlinear model and CD feed-forward. For both training and test samples, R^2 is significantly improved. The test results are compared in Fig. 17b for four models. Test R^2 and test RMSE of both X tilt and Y tilt are consistently improved from baseline to the new method as nonlinear regression and CD feed-forward are introduced in the data analysis. It is noticed that X tilt performance is typically better than Y tilt for this use case and the several use cases studied in previous sections. In 3D NAND memory devices, the slit pitch along x-direction is typically 10 times larger than Y pitch. Mueller matrix spectra respond to asymmetries in x- and y- directions differently because of the difference in X and Y pitches. From the sensitivity spectra shown in Fig. 2, X overlay and X tilt spectral signals are much stronger than y-direction for the same amount of asymmetry. Because of the stronger spectral sensitivity, asymmetry measurement for x-direction typically outperforms y-direction.



(a) (b)

Figure 17. Correlation of upper deck tilt prediction and reference for channel hole dual deck wafers. a) correlation plots of ML prediction and reference for X tilt and Y tilt for linear model without CD feed-forward (baseline) and nonlinear model with CD feed-forward (proposed new algorithm) and b) X tilt and Y tilt test R^2 and test RMSE for four models: linear model without CD feed-forward, linear model with CD feed-forward, nonlinear model without CD feed-forward and nonlinear model with CD feed-forward.

6. DISCUSSION AND CONCLUSION

Mueller matrix off-diagonal blocks have a unique response to structural asymmetry, that is, the response is zero in absence of asymmetry and responds linearly within a small range of asymmetry. They provide direction indications for tilt and overlay in 3D NAND memory structures. We show that these Mueller matrix elements respond to tilt and overlay with distinct spectra features in UV to near IR range. The wavelength dependent Mueller matrix spectra carry sufficiently rich information to separate tilt and overlay when analyzed by simple linear ML models. To further improve tilt and overlay measurement robustness, we propose and demonstrate benefits using CD feed-forward and nonlinear regression and a combination of both. CD feed-forward helps compensate for the impact of process variation to tilt and overlay measurement. Nonlinear regression can extend the overlay measurement limit beyond the linear response range (–pitch/4, +pitch/4); it can also reduce measurement errors caused by imperfect linearity and structural variations for asymmetry within a quarter pitch range. We demonstrated tilt and overlay measurement capability using the broadband Mueller matrix technique as a nondestructive inline metrology solution with high throughput. The technology can be applied for on device measurement with fast recipe creation enabled by machine learning data analysis.

ACKNOWLEDGEMENT

The authors thank Wenmei Ming, Franklin Wong, and Hepeng Ding for performing sensitivity analysis of tilt and overlay and analyzing spectra and optimizing machine learning models for tilt/overlay separation. The authors thank Wenmei Ming for performing simulation for tilt measurements with CD feed-forward.

REFERENCES

[1] L. Liu, "3D NAND flash status and trends," in *Proc. IEEE Int. Memory Workshop (IMW),* May 2022, pp. 1–4.
[2] W. Zhang, J. Xu, S. Wang, Y. Zhou, and J. Mi, "Metrology challenges in 3D NAND flash technical development and manufacturing," *J. Microelectronic Manufacturing,* 3(1):1–8, 2019.
[3] L. Li., "Symmetries of cross-polarization diffraction coefficients of gratings," *J. Opt. Soc. Am. A 17,* pp. 881–887, 2000.
[4] N. Rana, Y. Zhang, T. Kagalwala, and T. Bailey, "Leveraging advanced data analytics, machine learning, and metrology models to enable critical dimension metrology solutions for advanced integrated circuit nodes," *J. Micro/Nanolithography, MEMS, and MOEMS,* 13(4), 041415, 2014.
[5] F. J. Wong, Y. Hao, W. Ming, P. Žuvela, P. Teh, J. Shi, and J. Li, "Methods to overcome limited labeled data sets in machine learning-based optical critical dimension metrology," *SPIE Advanced Microlithography, Proc. SPIE,* vol. 11611, 2021.
[6] A. C. Diebold, G. A. Antonelli, and N. Keller, "Perspective: Optical measurement of feature dimensions and shapes by scatterometry," *APL Materials,* 6 058201, 2018.
[7] J. Li, J. J. Hwu, Y. Liu, S. Rabello, Z. Liu, and J. Hu, "Scatterometry measurement of asymmetric gratings," in *Proc. SPIE,* vol. 7520, Lithography Asia 2009, 75201B, 2009.
[8] G. A. Antonelli, N. Keller, T. Ribaudo, F. Wong, W. Ming, H. Ding, Z. Chen, R. Grynko et al., "Ellipsometric critical dimension metrology employing mid-infrared wavelengths for high-aspect-ratio channel hole module etch processes," *SPIE Advanced Microlithography,* in *Proc. SPIE,* vol. 11611, 2021.