PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Full wafer OCD metrology: increasing the sampling rate without the cost of ownership penalty

Daniel Doutt, Ping-ju Chen, Bhargava Ravoori, Tuyen Tran, Eitan Rothstein, et al.

Daniel Doutt, Ping-ju Chen, Bhargava Ravoori, Tuyen K. Tran, Eitan Rothstein, Nir Kampel, Lilach Tamam, Effi Aboody, Avron Ger, Harindra Vedala, "Full wafer OCD metrology: increasing the sampling rate without the cost of ownership penalty," Proc. SPIE 12496, Metrology, Inspection, and Process Control XXXVII, 124960B (27 April 2023); doi: 10.1117/12.2657471

SPIE.

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

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Full wafer OCD metrology: Increasing sampling rate without the cost of ownership penalty

Daniel Doutt*[a], Ping-ju Chen[a], Bhargava Ravoori[a], Tuyen K. Tran[a], Eitan Rothstein[b], Nir Kampel[b], Lilach Tamam[b], Effi Aboody[b], Avron Ger[b], Harindra Vedala[c]
[a]Logic Technology Development, Intel Corp, 2501 NE Century Blvd, Hillsboro, OR 97124;
[b]Nova LTD, 5 David Fikes St, Rehovot, Israel 761020;
[c]Nova Measuring Instruments Inc, 3342 Gateway Blvd, Fremont, CA, USA 93117

ABSTRACT

Optical Critical Dimension (OCD) spectroscopy is a reliable, non-destructive, and high-throughput measurement technique for metrology and process control that is widely used in semiconductor fabrication facilities (fabs). Wafers are sampled sparsely in-line, and measured at about 10-20 predetermined locations, to extract geometrical parameters of interest. Traditionally, these parameters were deduced by solving Maxwell's equations for the specific film stack geometry. Recently advanced machine learning (ML) models, or combinations of ML and geometric models, has become increasingly attractive due to the several advantages of this approach.

Advanced node processes can benefit from more extensive data sampling, but this conflicts with measurement cycle time goals and overall metrology tool costs, which cause fabs to use sparse sampling schemes. In this paper, we introduce a novel methodology that allows wafers to be sampled sparsely but provides the parameters of interest as if they were densely measured. We show how such a methodology allows us to increase data output with no impact on overall measurement time, while maintaining high accuracy and robustness. Such a capability has potentially far-reaching implications for improved process control and faster yield learning in semiconductor process development.

Keywords: OCD metrology, machine learning, WiW, wafer map, sampling scheme.

*daniel.r.doutt@intel.com; phone +1 (503) 840-2565

1. INTRODUCTION

A fundamentally critical challenge in the development of advanced node semiconductor processes is the reliable and repeatable production of smaller and increasingly intricate geometric structures, typically a few tens of nanometers in size. Achieving this requires extensive measurements of the critical dimensions of these structures which generally are performed through either imaging, in particular Critical Dimension Scanning Electron Microscopy (CD-SEM), or optical scatterometry. While CD-SEM measurements offer higher spatial resolution, they are speed constrained and usually are unable to make depth/height measurements. In contrast, Optical Critical Dimension (OCD) metrology offers a balance of speed and resolution and consequently is implemented as an integral part of both R&D and high-volume manufacturing (HVM) production lines. The ability to collect data quickly for real-time adjustments and tool monitoring is essential for process quality control. Even while advanced OCD metrology equipment provides fast and non-destructive measurements, fabs generally sample only a subset of the dies on the wafer (typically ~25% coverage) to meet cycle time goals.[1] For example, Nova's latest multi-channel PRISM® tool can measure dozens of dies per minute. While wafer sampling plans are carefully designed to represent all wafer quadrants and known radii of interest, inevitably, useful data is missed. These details can include process specific within-wafer (WiW) profiles, topography variation within die regions of interest, or process-chamber specific WiW profiles. Furthermore, while supporting process changes, qualifying new equipment, and for excursion detection and recovery, it is often necessary to oversample the wafers.

Metrology, Inspection, and Process Control XXXVII, edited by John C. Robinson,
Matthew J. Sendelbach, Proc. of SPIE Vol. 12496, 124960B
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2657471

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

In recent years, advanced machine learning methods have been successfully applied to enhance traditional Rigorous Coupled Wave Analysis (RCWA) solutions[2] by improving time to solution, precision, and by minimizing non-linear correlations.[3] Other recent applications of ML methods include measuring aperiodic structures (such as logic or graphics) and predicting e-test data based on optical spectroscopic measurements.[4] In this paper, we present a novel approach, MapPro, that extends beyond previously demonstrated techniques by leveraging advanced ML algorithms to deliver a full wafer map output using only sparsely sampled inputs with almost no impact on overall measurement times. Section 2 discusses the motivation for this innovative approach. In Section 3, we explain the methodology and applications of MapPro. Examples of full wafer map output results are presented in Section 4. The robustness and sensitivity to process variability is addressed in Section 5. Lastly, in Section 6, we discuss our conclusions and future work.

2. MAPPRO: MOTIVATION FOR ML-ENABLED DENSE SAMPLING SCHEME

Spatial sampling schemes are motivated by a desire to capture all process variations across the wafer. In general, CD measurement data is consolidated into radial profiles, contour maps, and various statistical measures such as wafer mean and WiW standard deviation. It is important to engineer sampling schemes to adequately capture all radial variations and spatial variability patterns to ensure the metrics are truly representative of the process. A further complication is that process tools have dissimilar spatial variability signatures, which makes it challenging to obtain a uniform sampling scheme that applies to all process steps. However, in order to support data analysis and parameter tracking needs, fabs are obliged to compromise by accommodating a one-size-fits-all sampling scheme rather than process specific ones. To illustrate this, a common challenge is that electrical test (e-test) measurements are collected during various process steps and at different wafer locations, making OCD correlations difficult. Moreover, it takes valuable time for wafers to reach end of line sort diagnostics, where the impact of process defects may be exposed and material discarded. Such challenges suggest that the fab environment would considerably benefit from high-resolution sampling and the ability to detect excursions earlier than current capabilities allow. Full wafer sampling can address these challenges by allowing modules to employ process-tailored data and enhanced visibility, leading to faster yield learning and improved process control. In this section, we present the characterization benefits of MapPro full wafer measurements and discuss how this approach can be implemented without incurring the burden of increased physical wafer measurements.

To illustrate the range of information delivered by different sampling schemes, Figure 1 shows the contour maps obtained by a sparse high throughput in-line scheme, dense low throughput full wafer map (FWM), and in-line sparse high throughput sampling enhanced by MapPro. In the FWMs one can see that the parameter behaves in a characteristic sombrero-like distribution, which is due to symmetries in some of the process tools but is not effectively captured by the sparse sampling scheme.

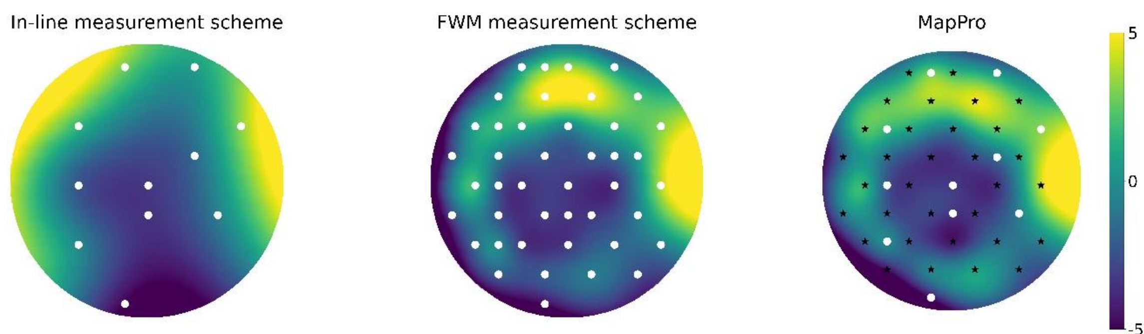


Figure 1. Parameter wafer contour plots (in arbitrary units). Left: parameter wafer contour map obtained using the sparse in-line measurement where the measurement points are denoted by the white dots. Middle: map using a FWM sampling plan. Right: map using in-line sparse measurements (white dots) with extrapolated dies obtained by MapPro (black stars). For interpolation between the measured and predicted points we used a simple radial basis function with a cubic spline interpolation.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The idea of MapPro takes inspiration from the recent achievements of deep learning techniques[5], in particular those that map low-resolution input to a high-resolution output[6]. A notable example is the super resolution convolutional neural network developed by Dong et. al.[7], which is used to increase the resolution of image files. The concept involves generating a training set of high-resolution images, whereby you proceed to lower the resolution for the model input features while keeping the high-resolution ones as the target output. This technique requires no labeled data; only a sufficient number of high-resolution images is needed.

Another interesting use-case comes from the field of weather forecasting. While this industry is very different from that of semiconductor manufacturing, from a modeling perspective, there are in fact many similarities. The weather forecast can be predicted using complex physical models that requires an advanced computing infrastructure. As such, delivering an accurate and high spatial resolution forecast is both expensive and time consuming. Rodrigues et. al.[8] have shown that training weather models on archived weather observations can bypass the need for human categorizing or additional computational resources. They demonstrated that, with their Convolutional Neural Network based model they were able to interpolate between the locations where weather forecasts take place and predict weather patterns with high accuracy and dense spatial resolution. It is with these considerations and techniques in mind that we developed MapPro.

3. METHODOLOGY AND APPLICATIONS

As previously indicated, MapPro allows the prediction of FWM output by directly mapping the spectra from a sparse in-line measurement scheme to a dense high-resolution output. For the offline training stage (see the training flow in Figure 2), the required data consists of several dozen wafers that were measured with a dense sampling plan, which includes the dies measured during the in-line process (in-line dies) and the dies for which the model outputs a parameter prediction (extrapolated dies) without them actually being measured.

Usually, in order to train a supervised machine learning algorithm, one requires to obtain labeling data, for instance using CD-SEM, AFM (atomic force microscopy), or TEM (transmission electron microscopy), which are slow, expensive, and potentially destructive. In our approach, we circumvent the need of any such externally labeled datasets by effectively utilizing the measured spectra and the OCD model generated parameters.

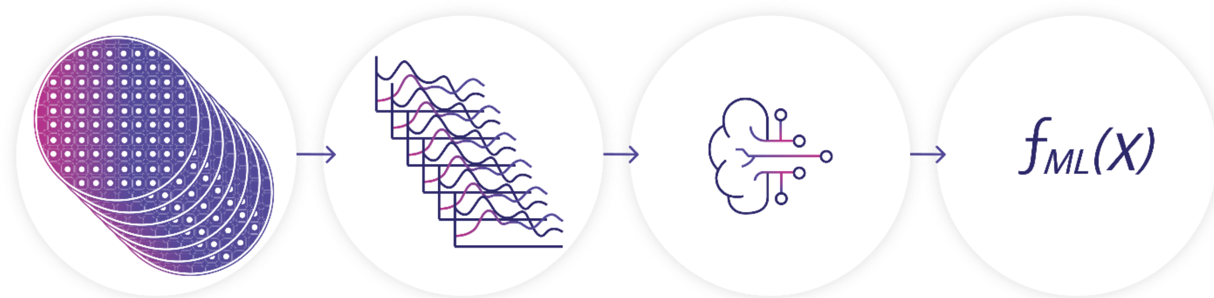


Figure 2. The MapPro training flow. We start with several dozen densely measured wafers. This data is used for training the ML algorithm where for the labeling we use the qualified OCD solution. After the training stage we obtain an ML model that can be run on the measuring tool and publish the results to the host.

During the training stage we execute an automatic sequence that finds the optimal model parameters and hyperparameters to optimizes a function $f_{ML}(X)$ that directly maps between the process of record (POR) spectra (denoted by X) and the FWM output for the desired parameter (Figure 2). This trained model is then run on the OCD metrology tool and after each wafer is measured the model outputs to the fab host the parameter values for both the in-line and extrapolated dies (Figure 3). During prediction the model also outputs a prediction confidence measure, MapPro Trust, to quantify the reliability of the prediction. The Trust threshold value for a reliable prediction can be tuned during the training stage. For

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

the results reported here, we tuned this value to be 0.9. Therefore, if we achieve $Trust > 0.9$ the predictions are reliable. If, for some reason, the predictions are no longer reliable, for example due to a process change that alters the process stack, the Trust score will be lower than 0.9; in such cases the user needs to analyze the data and, occasionally, a model retrain is needed to accommodate the new stack.

The reason we are able to train a model that optimizes a function $f_{ML}(X)$ that directly maps between low-resolution and high-resolution parameter outputs for a wafer is that the fabrication process for each die is statistically similar to that for other dies within the wafer. They all pass through the same process tools; however, in some of the tools the wafer rotates during the process, which induces parameter rotational symmetry. Moreover, the wafer die parameters after each process step are not isolated from the parameters in previous steps. In other words, the die parameters are also a function of the previous process steps that the wafer went through, and this information is encapsulated in the spectra.

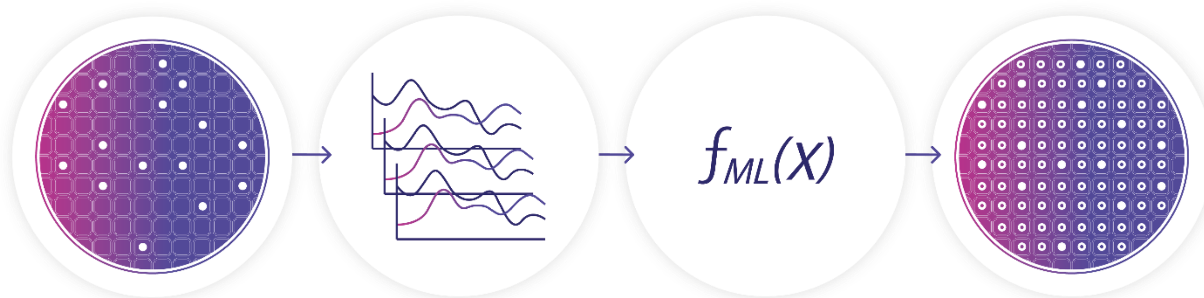


Figure 3. MapPro runtime flow. The spectra of sparse measurements are used as the input for the trained ML algorithm, which outputs high-resolution, densely located, parameter values and publishes the results to the host in real time during the in-line process.

4. RESULTS

In this section, we discuss the data sets from in-line sampling, measured FWM, and extrapolated results using MapPro for multiple in-line process steps and target structures. We highlight the improved visibility that full wafer metrology facilitates and show that these same details can be predicted reliably using MapPro. For clarity, the results section focuses on examples of representative Etch, chemical-mechanical planarization (CMP) and film deposition process steps. POI-1 and POI-2, in each example, represent the two primary parameters of interest for process and line control at that particular step. MapPro models were developed for each process step, for both scribe-line metrology pads (SCRIBE) and in-die full active structures (SRAM).

4.1 ETCH STEP 1

In the Etch Step 1 process, FWM sampling was performed on two different structures over several hundred in-line wafers. Figure 4 (a1) - (a4) show the complementary set of contour and radial plots for the in-line, FWM OCD, and in-line + MapPro sampling for POI-1 and POI-2, collected on the SRAM structure.

In Figure 4 (a1) POI-1, the in-line scheme contour plot shows very little WiW parameter variability, with features that become washed out due to the low-resolution data collected. In the FWM scheme contour plot we see a distinctive ring feature in the mid-radial band and a clear spoke-like symmetric feature that originates at the wafer center. These features are accurately predicted and clearly visible in the contour figure for the MapPro in-line with extrapolated data. Similarly, in Figure 4 (a3) POI-2, the in-line scheme contour simply shows a broad ridge-like feature extending across the diameter of the wafer and falling off on either side. Conversely, the FWM OCD and MapPro extrapolated contour plots clearly show the radial symmetry of this feature as well as numerous zones of elevated POI-2 peaks. The results of both POI-1 and POI-2 of ETCH STEP 1 highlight the limited data delivered by the in-line sampling schemes and the increased visibility to features or potential zones of interest in the full wafer data. Figure 4 (a2) and (a4) show the radial profiles of the in-line,

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

FWM, and MapPro in-line with extrapolated data, plotted together with a standard spline (lambda = 0.1) fitting. Note the similarities between the FWM and MapPro in-line with extrapolated radial profiles, which is indicative of the algorithm's accuracy and robustness. Also note the similarity between the in-line radial profiles for the FWM and extrapolated lines. Each of the radial plots deliver similar WiW profiles with POI starting high at the center of the wafer, dipping down in the mid-radius bands, increasing again towards the outer radii, and then dropping off near the edge. While radial plots are commonly used to monitor WiW variability, it is important to realize that they are simply another low-resolution output for POI. In fact, the radial profiles consistently underreport much of the detail of the in-line scheme and information delivered by the full wafer contours.

4.2 CMP STEP

Figure 4 (b1) - (b4) show the complementary set of contour and radial plots for the in-line, FWM OCD, and MapPro predicted results for POI-1 and POI-2, collected from the SCRIBE structure for the CMP process step. Similar to the previous figures, in Figure 4 (b1) POI-1 it is evident that the in-line scheme contour plot shows very little WiW parameter variability and diffuse low-resolution detail of the wafer map. The FWM and MapPro in-line with extrapolated contours show improved WiW detail. However, due to nature of the CMP process, the WiW variability is low and both contour and radial plots are without strong features. But closer examination of the radial profiles reveals two lobes of increasing POI-1 at the ~30mm and ~120mm radial bands, that are only captured with full wafer sampling. Notably, the overall wafer flatness is a key deliverable of the CMP process. Likewise, the impact of the CMP process on underlying structures, like those represented by POI-2, is of equal importance. The POI-2 results for FWM OCD and MapPro prediction (shown in Figure 4 (b3)) reveal well defined features at the wafer center and edge which are regions of general interest to the CMP process modules. The high-resolution data provided by the MapPro prediction reveals enhanced visibility of POI variability that could help to provide early detection for process and hardware performance and contribute to successful fulfillment of key deliverables.

4.3 ETCH STEP 2

In Figure 4 (c1) - (c4) we present the complementary set of contour and radial plots of the in-line, FWM OCD, and MapPro predicted results for POI-1 and POI-2 collected from a SCRIBE geometric structure. Similar to the CMP STEP, in Figure 4 (a1) POI-1, each sampling scheme shows little within-wafer variability due to the uniformity and characteristics of this process step. Still, the FWM contour plots again clearly show more detail of the WiW variability when compared to in-line sampling. Much like the features highlighted in ETCH STEP 1, in Figure 4 (c3) we again notice a distinct ring feature in the mid-radial band. These features once again accurately predicted and clearly visible within the MapPro extrapolated data and entirely absent from the in-line sampling results. As before, we can see more distinct zones of interest within the radial band feature that characterizes the WiW profile. And much like the previous process step examples, while the radial profiles of both POI-1 and POI-2 are well matched across sampling schemes, the full wafer contour plots clearly show information that is not delivered by the low-resolution data sets. And, like most process modules, these features and zones of interest may be highly dependent on etch tool performance or modulate from product-to-product according to the density of the geometric structure or die layout definitions.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

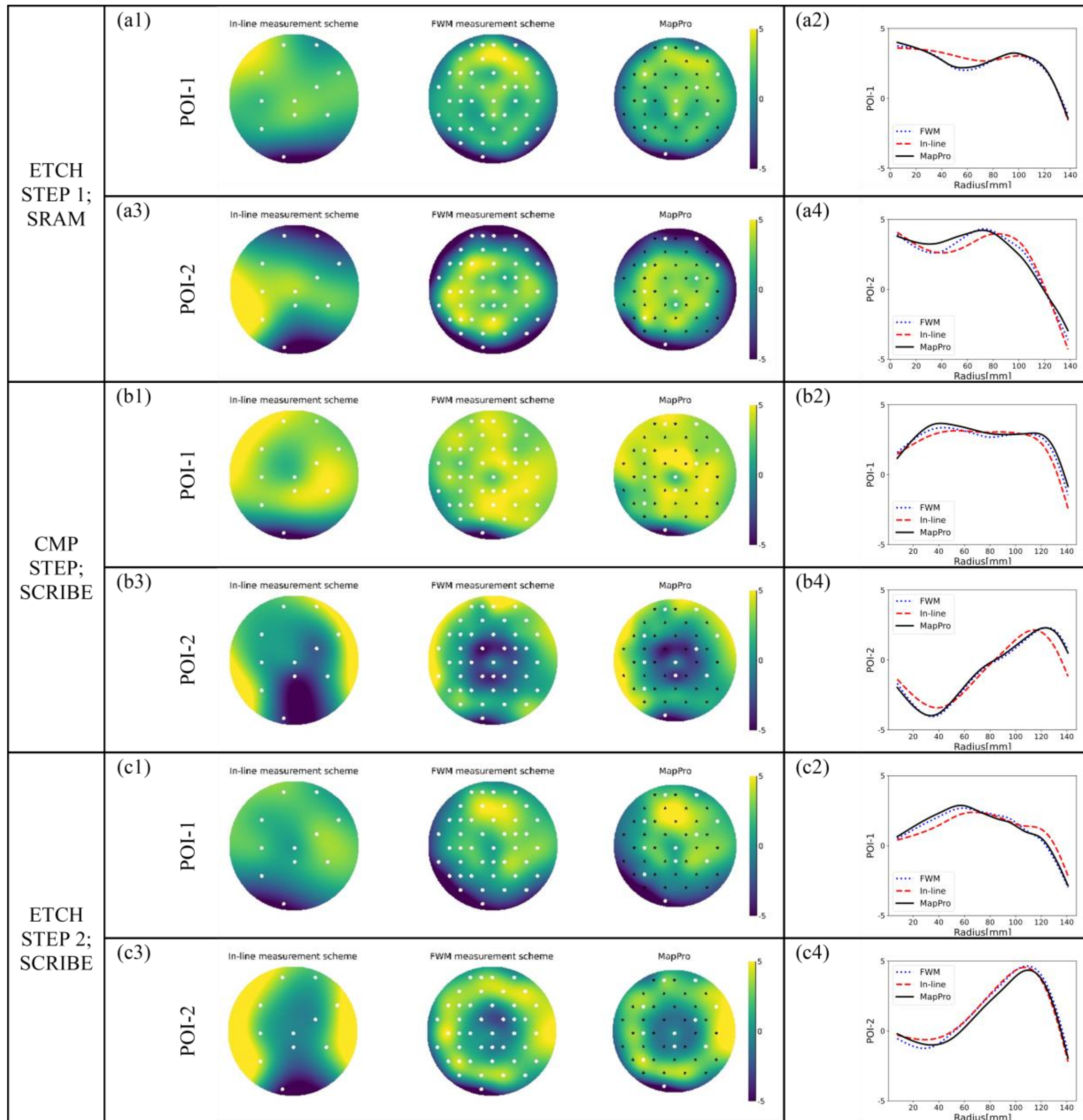
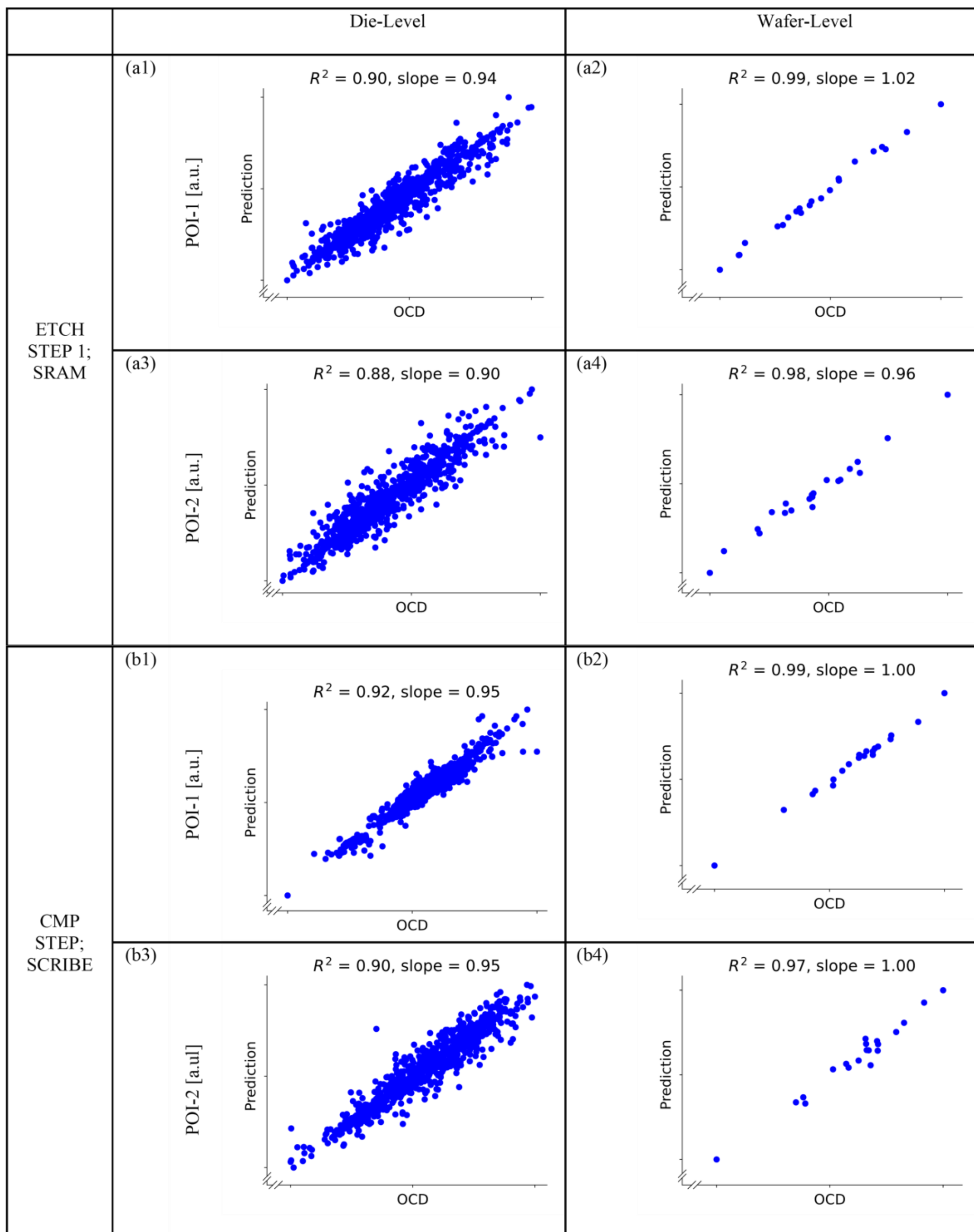


Figure 4: Contour plots and radial profiles of parameter of interest (POI) presented with arbitrary units [a.u.]. Subfigures (a1) - (a4) show the results from ETCH STEP 1, (b1) - (b4) show the results from the CMP STEP, and (c1) - (c4) show the results from the ETCH STEP 2. Subfigures (a1), (b1), and (c1) show, from left to right, respectively, the contour plots from in-line measurement scheme, full-wafer map (FWM) measurement scheme, and in-line with MapPro prediction scheme of POI-1. Subfigures (a3), (b3), and (c3) show, from left to right, respectively, the contour plots from in-line measurement scheme, full-wafer map (FWM) measurement scheme, and in-line with MapPro prediction scheme of POI-2. Subfigures (a2), (b2), and (c2) show the radial profile overlay radius (mm) for POI-1. Subfigure (a4), (b4), and (c4) shows the radial profile overlay radius (mm) for POI-2. Note that the color scale in the contour maps and vertical scale in radial plots are in arbitrary units; measurement locations also are arbitrary.

In Figure 5 we show the correlation statistics of the FWM OCD results versus the MapPro FWM blind test results for each of the ETCH STEP 1, CMP STEP, and ETCH STEP 2. All process steps show a n R^2 linear fit of greater than 0.88 with

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

slopes close to unity for both both die-level and wafer-level POIs. Here we can see that both the die-level and wafer-level distribution of data across all the many wafers are well correlated for both POI-1 and POI-2.



Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

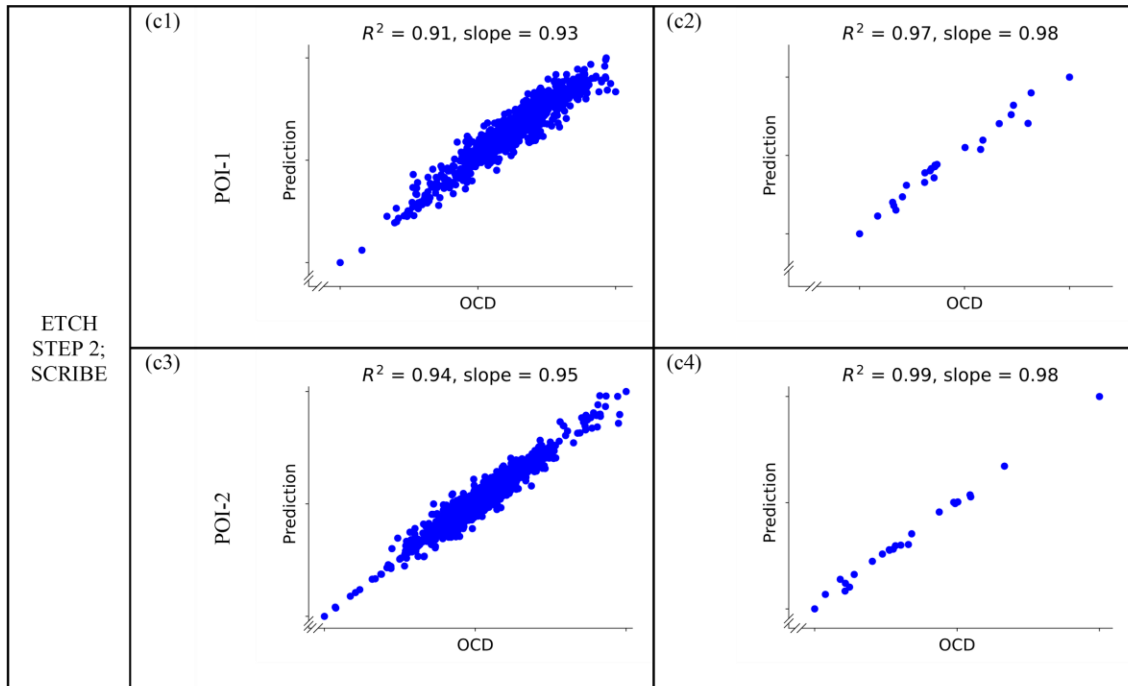


Figure 5: Correlation plots of the parameter of interest (POI). Subfigures (a1) - (a4) show the results from the ETCH STEP 1; (b1) - (b4) show the results from the CMP STEP; (c1) - (c4) show the results from the ETCH STEP 2. Subfigures (a1), (b1), (c1) and (a3), (b3), (c3) shows the die-level correlation plots for the OCD full-wafer map (FWM) measurement scheme versus MapPro prediction scheme of POI-1 and POI-2, respectively. Subfigures (a2), (b2), (c2) and (a4), (b4), (c4) show the die-level correlation plots for the OCD full-wafer map (FWM) measurement scheme versus MapPro prediction scheme of POI-1 and POI-2, respectively. Note that the correlation plots are presented in arbitrary units [a.u.])

			Die-Level		**Wafer-Level**	
Step	**Parameter**	**Target**	**r2**	**slope**	**r2**	**slope**
Etch 1	POI-1	SCRIBE	0.93	0.95	0.99	1.00
	POI-2	SCRIBE	0.88	0.89	0.97	0.90
	POI-1	SRAM	0.90	0.94	0.99	1.02
	POI-2	SRAM	0.88	0.90	0.98	0.96
CMP	POI-1	SCRIBE	0.92	0.95	0.99	1.00
	POI-2	SCRIBE	0.90	0.95	0.97	1.00
	POI-1	SRAM	0.90	0.97	0.99	1.00
	POI-2	SRAM	0.78	0.86	0.97	0.91
Etch 2	POI-1	SCRIBE	0.91	0.93	0.97	0.98
	POI-2	SCRIBE	0.94	0.95	0.99	0.98
	POI-1	SRAM	0.93	0.93	0.99	0.95
	POI-2	SRAM	0.96	0.97	0.99	1.00
Deposition	POI-1	SCRIBE	0.87	0.90	0.96	0.94
	POI-2	SCRIBE	0.94	0.97	1.00	1.00
	POI-1	SRAM	0.92	0.96	0.98	0.99
	POI-2	SRAM	0.96	0.97	1.00	0.99

Table 1: Summary of correlation coefficients from linear regression for various layers. Each layer shows the correlation between the FWM measurements and MapPro predictions per die (labelled with die-level type) and for the wafer mean (labelled with wafer-level type) for two target locations, SCRIBE and SRAM.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

In Table 1 we present the correlation statistics of the OCD-collected FWM results versus the full wafer output results of the blind test evaluation for four unique process steps: ETCH STEP 1, ETCH STEP 2, CMP STEP 1, and DEPOSITION STEP. The table details the die-level and mean Wafer-Level R^2 fit quality and slope for each parameter of interest for both SCRIBE and SRAM structures at each process step. In general, the die-Level POIs show an R^2 of greater than 0.90 and slopes of near unity for every process step. The CMP Step POI-2 on the SRAM structure is an exception with R^2 ~ 0.78 and slope ~ 0.86. While these correlations are still strong, this indicates that a more detailed training set might be needed. Typically, these correlations can be driven by edge die where processing variability may be more common. Similarly, the DEPOSITION STEP POI-1 SCRIBE die-level R^2 and slope are below 0.90. This may be due the thickness and WiW variability of this particular process step being very small. Comparably, all mean wafer-level POIs show an R^2 of greater than 0.95 and slopes of near unity for all process steps which demonstrates the robustness of the ML algorithm ability to predict wafer-level statistics. Table 1 clearly shows that the prediction is not limited to simple structures or materials. Indeed, the prediction's benefits can extend to many varieties of structures and process conditions. Similar statistics showing good R^2, and slope correlations are also given for the mean wafer-level data.

Collectively, these results reveal the limitations of a "one-size-fits-all" sparse sampling scheme. In fact, variation from different process steps and tooling steps can have very distinct WiW signatures originating from a variety of sources. While process modules may share a goal of consistent, uniform, and flat WiW POI distributions the sources of variation and the challenges to overcome them are tool and/or process specific. As such, process modules have dramatically different zones of interest. For example, the primary ETCH process hurdle may be uniform etching across the mid-radius bands of the wafer shown in ETCH STEP 1. While for CMP, the dramatic edge roll-off shown in CMP STEP may pose the most immediate challenge for improving yield. When in-line problems present themselves, these unique considerations lead to specific sampling needs for root cause diagnosis. These actions are typically in response to downstream and time-delayed data. Currently, process developers must rely on POR sparse sampling and downstream data correlations for excursion detection. Once an excursion signal is detected integrators may then respond by inserting dense sampling across upstream process steps and patiently awaiting downstream validation data. Ultimately, modules and process developers alike would benefit from early prediction via dense sampling. MapPro predicted sampling empowers process modules to do just that by enabling users to choose data that is most relevant to each process step and improve control and excursion detection.

5. ROBUSTNESS OF ML-BASED SOLUTION

The MapPro Trust metric was developed for evaluating machine learning data and its robustness against the trained model feature vectors. With the MapPro Trust metric, we can verify that incoming data falls within the expected range of training feature vectors to show confidence in the predicted results. Figure 6 shows the MapPro Trust scores for a compilation of ML data collected at the ETCH STEP 1. This data presents results collected on in-line POR wafers (PROCESS) compared with results collected on experimental wafers that have undergone stack changes (EXPERIMENTAL). The in-line process data in Figure 6 show Trust scores of well over 0.90 for a most sites and for both targets. The few points that fall below 0.90 are confirmed edge-die locations that exhibit larger stack variability than those contained in the training set. The experimental data in Figure 6 show Trust scores predominantly below 0.90, indicating that the process stacks of these wafers are measurably different than those that produced the training set. These results indicate that the MapPro predicted solutions are sensitive to in-line process perturbations. As such, the Trust quality metric is capable of alerting users to excursionary wafers and provides an indicator to inform users when additional ML training data is required.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

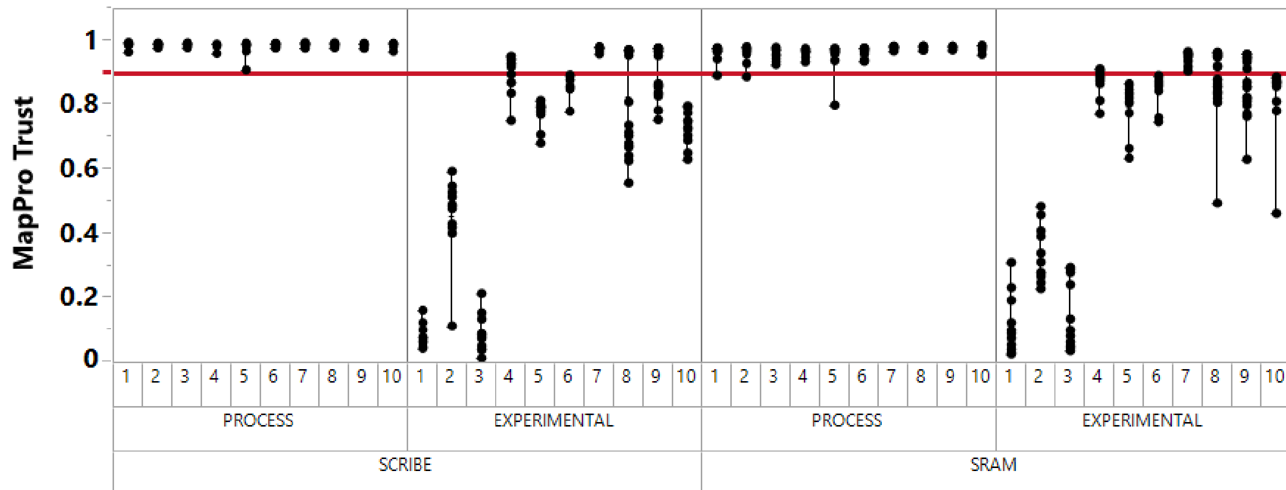


Figure 6: MapPro Trust values of ETCH Layer 1 for Target SCRIBE and Target SRAM under different process conditions. Here we see that the experimental wafers have a low MapPro Trust value, which indicates that the training set distribution does not cover these wafers, as expected.

6. SUMMARY AND FUTURE DEVELOPMENT

In this paper, we described a novel machine learning based approach (MapPro) to provide full wafer metrology data based only on sparse wafer sampling without incurring additional measurement time. This technique is shown to be generally applicable to various process steps and geometries with data from multiple advanced nodes for etch, CMP, and deposition steps. While the results presented here use spectra collected from the Nova PRISM® platform as the feature vector, MapPro is readily extensible to assimilate data from disparate measurement sources. The implications of this technique extend beyond providing densely sampled data. One can readily envision its applications to engineer more efficient dynamic sampling schemes, where downstream measurement locations are chosen based on the predicted full-wafer data with a potential to further speed up yield learning. The MapPro Trust metric provides additional value by highlighting potential excursions as it is inherently based on identifying spectral deviations from a reference training set. In conclusion, we demonstrated the value of ML based techniques to enhance traditional OCD measurements.

REFERENCES

[1] Park C., Lee H., Lee D., Choi A., Buhl S., Kim W.-S., Groeger P., Guhlemann S., Kim S., and Kim M., “CD and OCD sampling scheme optimization for HVM environment,” Proc. SPIE 10959 (2019).

[2] Moharam M. G. and Gaylord T. K., “Rigorous coupled-wave analysis of planar-grating diffraction,” J. Opt. Soc. Am. 71, 811–818 (1981).

[3] Bringoltz B., Rothstein E., Rubinovich I., Kim Y. H., Tal N., Cohen O., Yogev S., Broitman A., Rabinovich E., and Zaharoni T., "Machine Learning and Big Data in optical CD metrology for process control," 2018 e-Manufacturing & Design Collaboration Symposium (eMDC), Hsinchu, Taiwan, pp. 1-4, (2018).

[4] Breton M., Chao R., Muthinti G., Peña, A., Simon J., Cepler, A., Sendelbach M., Gaudiell, J., Tang H., Emans S., Shifrin M., Etzioni Y., Urenski R., and Lee W., “Electrical test prediction using hybrid metrology and machine learning”, Proc. SPIE 10145 (2017).

[5] LeCun Y., Bengio, Y., and Hinton, G. “Deep learning”. Nature 521, 436–444 (2015).

[6] Wang Z., Chen J., and Hoi S. C. H., "Deep Learning for Image Super-Resolution: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 10, 3365-3387 (2021).

[7] Dong C., Loy C. C., He K., and Tang X., “Learning a deep convolutional network for image super-resolution,” in European Conference on Computer Vision. Springer, 184–199 (2014).

[8] Rodrigues E. R., Oliveira I., Cunha R., and Netto M., "DeepDownscale: A Deep Learning Strategy for High-Resolution Weather Forecast," in 2018 IEEE 14th International Conference on e-Science (e-Science), pp. 415-422 (2018).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 17 May 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use