Methods to Overcome Limited Labeled Data Sets in Machine Learning-Based Optical Critical Dimension Metrology

Franklin J. Wong*, Yudong Hao, Wenmei Ming, Petar Žuvela, Peifen Teh, Jingsheng Shi, and Jie Li
Onto Innovation Inc., 16 Jonspin Road, Wilmington, Massachusetts 01887
*franklin.wong@ontoinnovation.com; phone 1 408 515-1231

ABSTRACT

With the aggressive scaling of semiconductor devices, the increasing complexity of device structure coupled with tighter metrology error budget has driven up Optical Critical Dimension (OCD) time to solution to a critical point. Machine Learning (ML), thanks to its extremely fast turnaround, has been successfully applied in OCD metrology as an alternative solution to the conventional physical modeling. However, expensive and limited reference data or labeled data set necessary for ML to learn from often leads to under- or overlearning, limiting its wide adoption. In this paper, we explore techniques that utilize process information to supplement reference data and synergizing physical modeling with ML to prevent under- or overlearning. These techniques have been demonstrated to help overcome the constraint of limited reference data with use cases in challenging OCD metrology for advanced semiconductor nodes.

Keywords: OCD, machine learning, semiconductor, time to solution, robustness, metrology

1. INTRODUCTION AND PROBLEM STATEMENT

Machine learning (ML) and artificial intelligence have exhibited an increased demand in semiconductor fabs, and their presence in semiconductor metrology is growing [1-4]. In particular, applications of ML in optical metrology has exhibited a notable increase wherein the workhorse has been physical electromagnetic (EM) simulations followed by subsequent regression to match calculated to experimental spectra in order to extract fitted parameters. Specifically, in the case of optical critical dimension (OCD) metrology, the key fitted parameters are physical dimensions. Currently, physical modeling is still the incumbent solution in semiconductor fabs due to robustness challenges and general reliance on large sets of costly reference data most standard ML solutions face. Nevertheless, there are two key reason to pursue ML as an alternative OCD solution:

a. **Reduce Time to Solution**: Even the most efficient electromagnetic (EM) calculations of periodic structures, i.e. Rigorous Coupled Wave Analysis (RCWA) [5], in advanced devices are becoming increasingly computationally expensive, making the investment of computing resources high and time-to-working-metrology-recipe long. Despite the investment in computing resources, often the time to metrology solution does not meet the requirements to keep up with fast-paced environments that are typical in semiconductor research and development phases, that is, prior to high-volume manufacturing.

b. **Reduce Measurement Error**: As device dimensions scale downward, the metrology error tolerance likewise tightens. Novel data analysis packages for indirect metrology techniques are aimed at improving accuracy and increasing recipe robustness.

The metrology problem has unique features that need to be understood prior to application of ML techniques.

I. Metrology is a regression problem that requires high-precision analog output.

II. ML inputs for OCD metrology are comprised of optical spectra (i.e. signal vs. wavelength) and labeled reference data of physical dimensions corresponding to a small subset of the spectra.

III. In general, for ML-based OCD solutions spectroscopic data is readily available as such tools offer fast spectral acquisition. On the other hand, reference data is typically acquired through destructive imaging, and hence not as readily available. For a typical case, we can expect a total of 10 to 20 points of reference data acquired using a destructive technique.

Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV,
edited by Ofer Adan, John C. Robinson, Proc. of SPIE Vol. 11611, 116111P
 · doi: 10.1117/12.2583774

IV. In wafer preparation for metrology development, process engineers can design experiments to deliberately introduce process variations. These types of wafers are commonly referred to as Design of Experiment (DOE) wafers, also known as skewed wafers.

V. Physics-based OCD models are a reliable and established baseline for inline metrology. They required EM simulations of device structures, and hence more time-consuming compared to ML operations. These solutions can be developed with 10 to 20 points of reference data.

Therefore, the primary goal of metrology ML-based algorithm development is to design within the constraints in (I-V), while providing one or more of the values in (a-b).

2. CONCEPTUAL OVERVIEW OF THE PROPOSED SOLUTIONS

2.1 Pure ML Customized to the Metrology Problem Statement

ML of measured signal for metrology application is well-suited to meet the needs of shorter time to solution because computationally expensive EM simulations are replaced by cheaper pure mathematical. Standard ML requires (1) input data and (2) reference data or labels corresponding to a subset of the input data. In the case of optical metrology, measured spectra are the input data and reference data are numerical physical structural parameters corresponding to a subset of the locations of the measured spectra (Figure 1). In order to develop a high performing standard ML solution, one measures spectra from multiple wafers, and acquire as much reference data as possible. It is well-known that the quality of solution in this standard approach relies on a sufficient quantity of, as well as high-quality, reference data; in other words, often with limited reference data, the standard ML can be inaccurate.

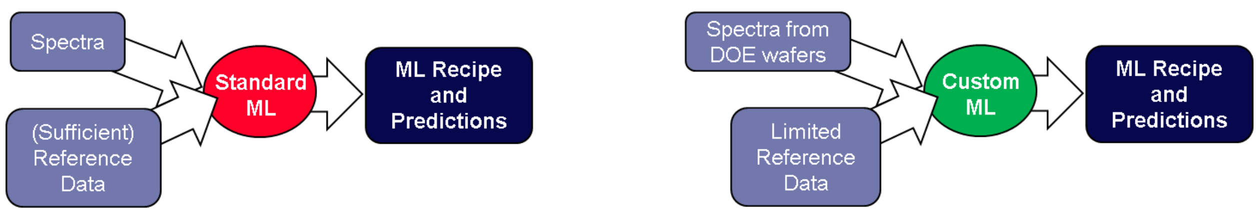

Figure 1. Conceptual difference between standard ML and ML customized for the metrology use case.

In order to overcome this challenge relating to standard ML, we developed custom ML techniques, algorithms, and software solutions that make use of the flexibility and constraints outlined in Section 1 to deliver robust, high-accuracy ML solutions. To build a good custom ML solution, we outline a few key recommended inputs:

i. Three DOE wafers containing per-wafer skews target at the key parameter
ii. 50-100 points of spectra measured per DOE wafer
iii. 3-5 reference data points per DOE wafer that are evenly spread in radius

Detailed methodology will be outlined in Section 3A. Such wafer and reference data requirements are comparable to what is needed to develop a traditional physical model-based solution, so no extra inputs are required, while the time to solution will drastically shortened. The requirement of DOE wafers skewing the key parameter of interest is readily accessible in semiconductor fabs, and it remedies the reference data challenge in ML immensely. It is a recurring theme in ML that careful design of wafer skews can dramatically improve the quality of solution. Contrary to many ML applications leveraging massive amounts of information employing data mining methods, in physical metrology wherein the hardware signal comes directly from the area of interest, careful design of wafers and spectral inputs is critical and requires subject-matter expertise.

2.2 Synergized Physical Modeling and ML Hybrid for Best Performance

Physical modeling has been a proven standard for many indirect metrology techniques in semiconductor manufacturing, especially in OCD [6,7]. When the development of a process node matures, metrology error budgets will tighten, process revisions decrease in frequency, and therefore the focus shifts from short time to solution to improved measurement accuracy. If the time allows for a physics-based OCD models to be developed, ML can be augmented to enhance the accuracy of final solution; we believe a hybrid approach offers the best overall performance, and should be used as the process development stage wherein accuracy and stability are the most critical metrics. Because the hybrid solution has a backbone of the physical model, the wafer design and data requirements are not as strict as in Section 2.1 for the pure ML approach. Figure 2 illustrates the inputs required for the solution. In addition, the reference data requirements are no greater than for traditional RCWA OCD modeling.

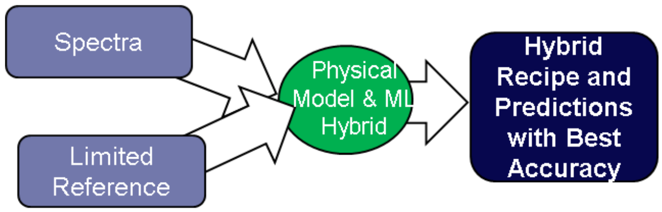

Figure 2. Hybrid OCD modeling and ML approach for improved accuracy.

3. CUSTOM MACHINE LEARNING

3.1 Proof of Principle Methodology: DRAM Fin Formation Example, A Buried Key Parameter

We will focus on a DRAM etch application. In this example, we demonstrate that ML techniques can achieve a solution comparable for a physics-based RCWA modeling solution, but can be developed inherently faster without the need to perform EM computations. The structure is shown in Figure 3. It is the process step that defines the FinFET height of the DRAM transistor. Physically, the key parameter is the amount of oxide recessed with respect to the etched silicon. This fin height is key to determining device performance, and metrology at this step has been critical for every device node in the $6F^2$ DRAM design. Because the key parameter is buried rather than on near the top and hence low sensitivity, it is quite challenging for OCD metrology. We chose this example because while for high optical sensitivity parameters, many types of ML approach can give decent results due to the inherent rich information content in optical spectra, lower sensitivity parameters pose extra challenges and require a more tailored approach. A challenging low sensitivity example better highlights the benefits of the methods we developed.

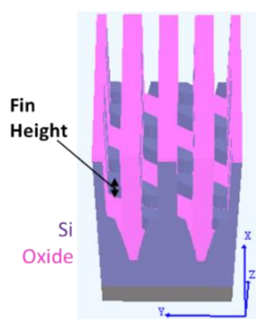

Figure 3. Schematic of a DRAM structure on the fin formation process step, with key parameter fin height labeled.

To demonstrate the proof of principle we acquired 70 spectra from three DOE wafers (POR^+, POR, and POR^-) whereby the fin height parameter was skewed. We will consider this set of wafers as **Group 1**. Seventy spectra were acquired from three additional validation wafers that underwent the standard process of fin height, but another process parameter was skewed – controlled independently from the fin height. We will consider this set of wafers as **Group 2**.

An established RCWA physical model-based recipe was used as a benchmark for evaluating the quality of the ML solution. Such a comparison is carried out due to limited true reference data for virtually all applications. It is uncommon that a sufficient number of reference points are available to reliably validate the quality of the results of the ML-based metrology solutions. This limitation leads to overly optimistic projections based on training results of limited reference that may not be robust. Therefore, the availability of a large number of model-based outputs allows for a better assessment of the proposed custom ML metrology solution's efficacy in extracting hardware signal and accurately mapping to physical parameters of interest.

As ML inputs we use **Group 1** spectra and their wafer skew information, and consider **Group 1** as the **validation set**. We have RCWA outputs for all of the wafers, 70 points per wafer, and perform different iterations using 3, 4, 5, 7, and 10 reference RCWA outputs per wafer to assess the quality of our solution for each scenario. The points that we explicitly use the RCWA outputs are considered as the **training set**. **Group 2** wafers and spectra completely left out from ML recipe creation, and we consider them as true **blind test**. To evaluate the quality of solution, we do not consider the **training set** performance with RCWA outputs explicitly used for ML recipe creation at level. Instead, we evaluate based only on the points *not* used. We show the performance in the **Group 1 validation set** and **Group 2 blind test** wafers separately. We compare the performance of the custom ML to a standard ML approach to highlight the benefits of the custom method.

3.2. Custom ML Performance Enhancements

Our custom ML has the following characteristics that are essential inline semiconductor metrology:

A. *Higher accuracy* than standard ML approaches as judged from the **validation** and **blind test** performance. This enables inline metrology for accuracy comparable to physical modeling.
B. *Improved ML error convergence* of the **training** versus **blind test** compared to standard ML. This allows practitioners to better predict recipe inline performance and hence improved usability.
C. *Robustness against reference noise*. This frees the dependence on perfect reference data, which can never be truly achieved.
D. *Precision at the Å level*. This enables high-precision inline metrology.

3.2.A. Higher Accuracy

The performance of the optimal solution of the Group 1 **validation** and Group 2 **blind tests** for different number of reference data inputs is shown on Figure 4. As expected, the error reduces upon inclusion of more reference data. Typically, in a real fab scenario, 3-5 reference points per wafer are accessible, but higher quantities become increasingly unlikely. The error of the custom ML is always lower than the standard ML, particularly in the regime of small number of reference points; the advantage is also more evident in the **blind test** wafers – more indicative of the inline performance.

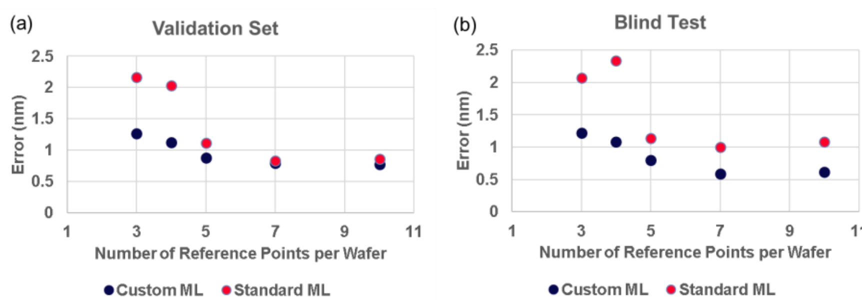

Figure 4. Custom ML and standard ML errors for (a) **Group 1** skewed wafer (**validation set**) and (b) **Group 2** wafers with standard fin height but skewed in other parameters (**blind test**) as they depend on the number of reference data points used for training.

3.2.B. Improved ML Error Convergence

The custom ML technique also drastically improves the convergence between the **training set** and the **blind test**, as depicted in Figure 5. In a real fab scenario, the **blind test** data set corresponds to inline sampling, for which there will be not viable method to validate often. The ideal ML solution not only should yield low training error, but also provide an estimate of **blind test** inline performance. A perennial challenge of ML applications in semiconductor metrology is with limited reference, extremely low training errors can be easily achieved even in the simplest regression algorithms due to overfitting; however, this error is by no means representative of the predictive performance inline in production environments. This can clearly be visible from Figure 5b, wherein the **training set** error is extremely small in all cases. The issue is worsened by the application of more complex ML regression algorithms; solutions with close to zero training error are readily accessible, but they are not unique and therefore will be unpredictable in the production environment. In contrast, Figure 5a shows a notably smaller gap between **training** and **blind test** and a well-behaved convergence trend noticeable even at 3-5 reference points.

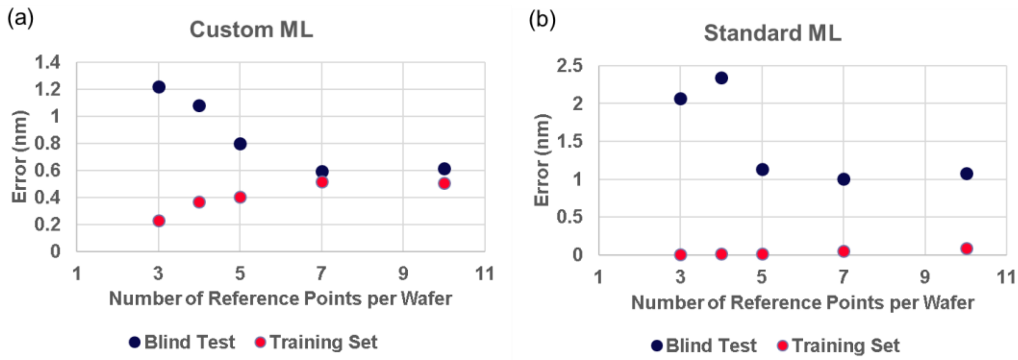

Figure 5. **Blind test** and **training set** error convergence for custom and standard ML.

3.2.C. Robustness against Reference Data Noise

Techniques to acquire experimental metrology reference data contain error. Naturally, the reference noise can propagate to the ML-based OCD recipe and increase its error. Figure 6 shows the impact of changing the noise level in the reference data from 0.2nm to 2.0nm for the case of 5 reference points per wafer [8]. This was done by adding a normally distributed random noise to each reference data point in the **training set**. It can be observed that the custom ML method is far more robust against reference noise as compared to the standard ML approaches. Figures 6(a1)-6(a4) display minimal degradation despite a substantial amount of noise added. On the other hand, while in the low noise case of 0.2nm, standard ML yields reasonable predictions on both the **validation** and **blind test** sets, the solution rapidly degrades as the reference noise increases. In a real case scenario, reference noise cannot be separated from the data itself and is difficult to estimate; therefore, the ideal ML solution should be robust against noise as in the custom method. Unlike the custom ML method, in the standard ML approach, semiconductor metrology-specific subject matter expertise was not considered, and one potential outcome is overtraining to noisy reference data as exhibited in Figures 6(b1)-(b4) which leads to poor inline performance.

3.2.D. Precision at the Å-level

Figure 7 shows the dynamic precision pooled over multiple sites, wafers, and cycles expressed as three times of the standard deviation. The baseline RCWA model yielded a precision (3σ) value of 0.95Å. Therefore, we can conclude that ML can yield comparable precision values. The number of reference data points has minimal impact in the precision of the custom ML solutions.

4. SYNERGIZED PHYSICAL MODELING AND MACHINE LEARNING HYBRID

In the previous section, we showed that a custom ML solution can (1) reproduce results comparable to physical modeling and (2) outperforms standard ML approaches on validation set and blind test. In situations wherein a physical model has already been developed for the process step of interest, we believe a hybrid physical modeling and ML combination will produce the highest accuracy solution. With a RCWA model as the backbone of the solution, this approach naturally needs no more reference data than the physical modeling approach; in combination with ML, the hybrid approach can extract and isolate more information from the measured signal than can be discerned by the physical model alone. In addition, there is not a requirement for exact wafer preparation as in custom ML (Section 2.1). Figure 8 shows a comparison between the RCWA versus the hybrid performance on the same DRAM fin structure. Here, the reference is from destructive imaging. The wafer set contains baseline POR wafers as well as wafers skewed on various structural parameters. It is shown that the correlation to the reference is stronger in the hybrid model with a close to unity sensitivity response.

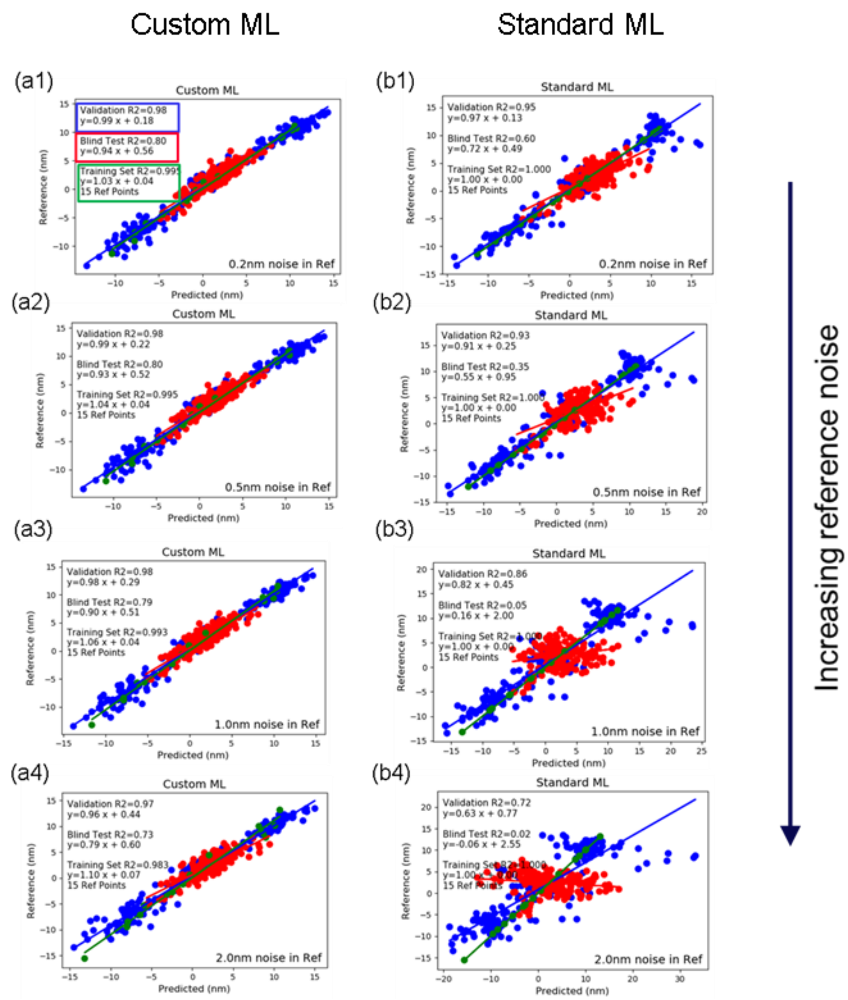

Figure 6. Custom ML versus standard ML dependence on reference data noise in the case of 5 reference points per wafer. The training, validation, and blind test points are highlighted by color.

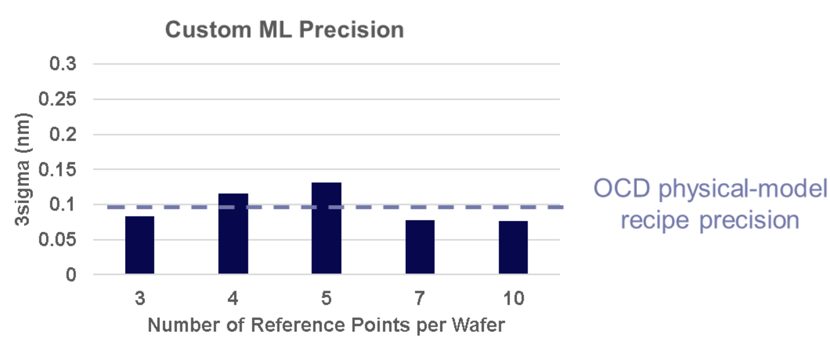

Figure 7. Precision of the custom ML method for varying numbers of reference points per wafer.

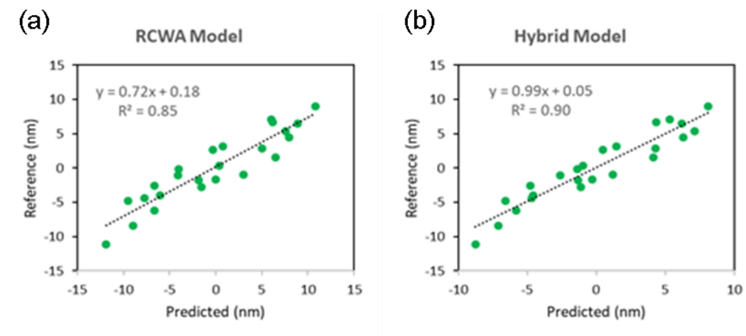

Figure 8. Performance of (a)RCWA model compared to (b) hybrid approach for DRAM buried fin etch.

Lastly, standard ML for multiple key parameters has always been challenging, particularly if they are produced by the same process and hence have physical, in addition to numerical, correlation to one another. In such a scenario, modeling of the exact key physical parameters and their differences can have a lot of benefit. Our hybrid approach can maintain the advantages of RCWA modeling as well as improve on its accuracy. Figure 9 shows an example of a three-key-parameter etch step, and a rough schematic illustrating the nature of the applications is shown on Figure 9c. One can see that the hybrid approach improves the accuracy of all three key parameters. Because there is always some unknown error in destructive imaging reference techniques, despite the small data set, the correlation of any robust model to the given reference data should never be perfect. Typically, that would be a sign of overtraining. We believe the hybrid yields solutions that maximize the information content of the signal, but does not overtrain to the errors of the limited labeled reference.

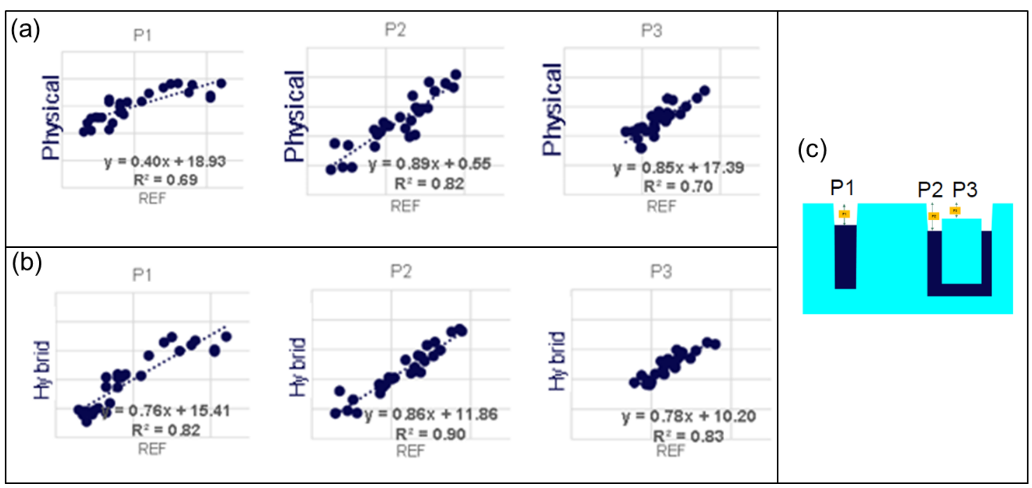

Figure 9. Multi-parameter performance of (a) RCWA model compared to (b) hybrid approach for a FEOL etch step in a logic structure. (c) The schematic is a rough illustration of the type of application.

5. CONCLUSION

In this work, we demonstrate the capability of custom ML with limited reference data on a challenging, comparatively low sensitivity buried parameter in **Section 3**. With such a technique, essentially metrology engineers can achieve with ML a solution comparable to physical modeling, but at a significantly faster time to solution. The custom ML should be utilized by engineers with metrology subject-matter expertise to place value-added solutions in various stages of the device development and manufacturing product lifecycles, particularly when time to solution is critical. This analysis module has been implemented in the AutoML Studio™ and SpectraProbe™ software products. In addition, we show in **Section 4** applications of hybrid physics-based modeling in conjunction with ML techniques applied to yield solutions with highest accuracy. Fab engineers can make use of this approach when the process control needs are strict and the highest performance metrology solution is required. The hybrid approach is available in the NanoDiffract™ analysis suite.

REFERENCES

[1] Rana, N., Zhang, Y., Kagalwala, T., and Bailey T. "Leveraging advanced data analytics, machine learning, and metrology models to enable critical dimension metrology solutions for advanced integrated circuit nodes," Journal of Micro/Nanolithography, MEMS, and MOEMS 13(4), 041415 (2014).

[2] Henn, M.A., Zhou, H., Richard M. Silver, R.M., and Barnes, B.M. "Applications of machine learning at the limits of form-dependent scattering for defect metrology", Proc. SPIE 10959, Metrology, Inspection, and Process Control for Microlithography XXXIII, 109590Z (2019);

[3] Timoney, P., Kagalwala, T., Reis, E., Lazkani, H., Hurley, J., Liu, H., Kang, C., Isbester, P., Yellai, N., Shifrin, M., and Etzioni Y. "Implementation of machine learning for high-volume manufacturing metrology challenges", Proc. SPIE 10585, Metrology, Inspection, and Process Control for Microlithography XXXII, 105850X (2018).

[4] Lee, H., Han, S., Hong, M., Kim, S., Lee, J., Lee, D., Oh, E., Choi, A., Park, H., Liang, W., Choi, D., Kim, N., Lee, J., Pandev, S., Jeon, S., and Robinson, J.C. "In-cell overlay metrology by using optical metrology tool", Proc. SPIE 10585, Metrology, Inspection, and Process Control for Microlithography XXXII, 105851D (2018)

[5] Moharam, M.G., and Gaylord, T.K, "Rigorous coupled-wave analysis of planar-grating diffraction," Journal of the Optical Society of America, **71**, 811 (1981).

[6] Yang, W., Lowe-Webb, R., Korlahalli, R., Zhuang, V., Sasano, H., Liu, W., and Mui, D., "Line-profile and critical dimension measurements using a normal incidence optical metrology system," 13th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference. Advancing the Science and Technology of Semiconductor Manufacturing. ASMC 2002 (Cat. No.02CH37259), 119-124 (2002).

[7] Diebold, A. C., Antonelli, G. A., and Keller, N. " Perspective: Optical measurement of feature dimensions and shapes by scatterometry," APL Materials, 6 058201 (2018).

[8] 0.2nm is small for some reference metrology techniques. 2.0nm is very large, particularly for the small feature sizes in advanced DRAM and logic. Typically, we expect random errors in the order of 0.5 to 1.0nm from cross-sectional reference techniques.