Context-based virtual metrology

Peter Ebersbach, Adam M. Urbanowicz, Dmitriy Likhachev, Carsten Hartig, Michael Shifrin

Peter Ebersbach, Adam M. Urbanowicz, Dmitriy Likhachev, Carsten Hartig, Michael Shifrin, "Context-based virtual metrology," Proc. SPIE 10585, Metrology, Inspection, and Process Control for Microlithography XXXII, 1058514 (22 March 2018); doi: 10.1117/12.2302498

Event: SPIE Advanced Lithography, 2018, San Jose, California, United States

Context-based virtual metrology

Peter Ebersbach[a], Adam M. Urbanowicz[b], Dmitriy Likhachev[a], Carsten Hartig[a], Michael Shifrin[c]

[a]GLOBALFOUNDRIES Dresden Module One LLC & Co. KG, Wilschdorfer Landstr. 101, 01109 Dresden, Germany
[b] Nova Measuring Instruments GmbH, Moritzburger Weg 67, 01109 Dresden, Germany
[c] Nova Measuring Instruments, LTD, P.O. Box 266, Weizmann Science Park, Rehovot 76100, Israel

ABSTRACT

Hybrid and data feed forward methodologies are well established for advanced optical process control solutions in high-volume semiconductor manufacturing. Appropriate information from previous measurements, transferred into advanced optical model(s) at following step(s), provides enhanced accuracy and exactness of the measured topographic (thicknesses, critical dimensions, etc.) and material parameters. In some cases, hybrid or feed-forward data are missed or invalid for dies or for a whole wafer. We focus on approaches of virtual metrology to re-create hybrid or feed-forward data inputs in high-volume manufacturing. We discuss missing data inputs reconstruction which is based on various interpolation and extrapolation schemes and uses information about wafer's process history. Moreover, we demonstrate data reconstruction approach based on machine learning techniques utilizing optical model and measured spectra. And finally, we investigate metrics that allow one to assess error margin of virtual data input.

Keywords: hybrid metrology, optical modeling, virtual metrology, machine learning, process context, process commonality, process control

1. INTRODUCTION

Increased complexity in modeling of the optical structures due to shrinking design rules and ultra-thin film layers sparks the needs for complex optical modeling approaches. Typically, advanced thickness or optical-critical dimension (OCD) measurements require an additional data input due to mutual correlations and weak spectral sensitivities of the model parameters relevant to production control. In optical metrology there are two typical approaches called "data feed-forward" (or "injection") and "hybrid metrology". Hybrid metrology involving OCD models is widely explored in literature and implemented on factory floor.[1−2] Hybrid metrology for OCD is a common approach to combine information from several metrology sources, i.e., CD-SEMs, AFMs, etc., while data feed-forward is a measurement method of passing data input from previous metrology step(s) on the same tool's platform. Our work is relevant for both approaches. Furthermore, in some cases these previously extracted (or obtained from other metrology sources) data require additional assumptions based on process specifications. We call it a hybrid data feed-forward (HDFF) approach. Since implementation in high-volume manufacturing, hybrid and/or feed-forward based optical model configurations are improved further to meet growing metrology requirements for advanced technology nodes. One of the improvement directions is a methodology to recreate missing data feed-forward and hybrid inputs. This is especially important for the OCD measurements, providing automatic process control (APC) feedback, where a sampling ratio reaches 100%. We focus on re-creation of hybrid and/or data feed-forward inputs for the OCD models utilizing virtual metrology (VM). VM is the technology of prediction of post process metrology variables using process and wafer state information that

Metrology, Inspection, and Process Control for Microlithography XXXII, edited by Vladimir A. Ukraintsev, Ofer Adan, Proc. of SPIE Vol. 10585, 1058514
© 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2302498

could include metrology, sensor data or process context.[3] The concept of VM was already applied for several applications in microelectronic manufacturing to reduce an amount of real metrological measurements by adaptive sampling rules[4,5], to predict metrology values based on sensor data from various tools such as etch or thin film deposition equipment[6], or to implement a virtual reference for the OCD model creation.[7] Also, a whole fab-wide VM system architecture has been discussed in literature as well.[8] The essence of VM approach lies mainly in a regression model which maps accessible data, such as measured process variables, to metrology data. A large variety of linear and nonlinear regression models can be used to identify a VM model.[9] Moreover, we demonstrate data reconstruction approach based on machine learning techniques utilizing optical model, measured spectra and process context. In summary, in this study we focus on several complex models used in high-volume manufacturing conditions and present a few approaches for HDFF data reconstruction using VM, optical modeling, and machine learning.

2. RESULTS AND DISCUSSION

This Results and Discussion section contains several subsections. First, we focus on missing data classification in HDFF solutions (subsection 2.1). Next, we introduce definitions of VM and its links to manufacturing process flow (subsection 2.2). Following the definitions, we study the easiest case when the site-level data are missing while possible time-dependent process drift and process context do not play any role (subsection 2.3). We also discuss appropriate data reconstruction approach for given type of process signature. Following, this less complex scenario, we discuss the cases when whole wafer data are missing (subsection 2.4). We show example how different process context and data aggregation levels and process drift influence quality of the data reconstruction. We compare two different approaches to solve problem of missing data in HDFF metrology solutions. First approach involves missing hybrid data reconstruction using virtual metrology and process context. Second approach is based on machine learning. We replace hybrid configuration involving several metrology steps with one-step machine learning solution only when HDFF data are missing. We discuss effect of training size and usage of process context for machine learning on the quality data reconstruction. Finally, we compare all approaches.

2.1 Missing data classification in data feed forward and hybrid solutions

In our study VM must provide sufficiently accurate inputs to enable convergence of an advanced optical model to global minimum and, thereby, ensure an accurate and exact model output. Typical schemes with missing data inputs are depicted in Figure 1. We have two main scenarios for the missing HDFF inputs. The first scenario relates to the missing/invalid data for a whole wafer and the second scenario is associated with missing/invalid data for individual measurement sites on a wafer.

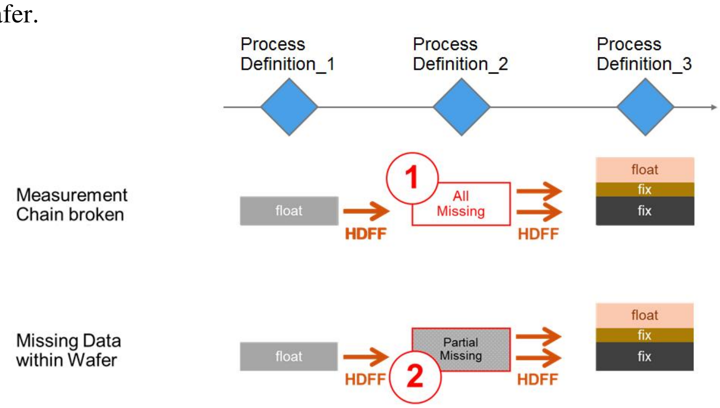

Figure 1 Examples of hybrid data feed-forward (HDFF) approach. Top picture shows HDFF scenario with incomplete data due to missing whole-wafer data. The picture at the bottom displays situation where the measurements of particular sites within the wafer are missed.

For complex solutions in optical metrology usually the ”one-to-one” site level HDFF used to cover wafer uniformity and minimize accuracy error in interpretation. Possible reasons for missing data can be the different sampling maps (static, dynamic) or filtered HDFF data by validation (i.e., due to bad fit). Furthermore, the measurement chain cannot be always secured in high volume manufacturing conditions. There are restrictions coming from sampling strategies (lot/wafer sampling by process step requirements) and/or capacity limitation. More complex situation occurs when a whole wafer data are missing. This could be due to possible skip in measurement, unavailability of a metrology tool ("tool down"), or other reasons such as factory host problems.

2.2 Data reconstruction based on virtual metrology

The virtual metrology approach can be schematically represented as shown in Figure 2. Virtual metrology approach accounts for wafer's process history and all calculations of virtual HDFF are based on that knowledge. Obviously, an accuracy of extracted virtual data depends on amount of available data with the same process history and data aggregation level. The error of calculated virtual data is expected to be the largest in case of interpolation based on the lot-level data and the smallest error would be from the die-level data from a set of wafers with the same process history. To demonstrate possible challenges of data reconstruction we analyze manufacturing process flow model.

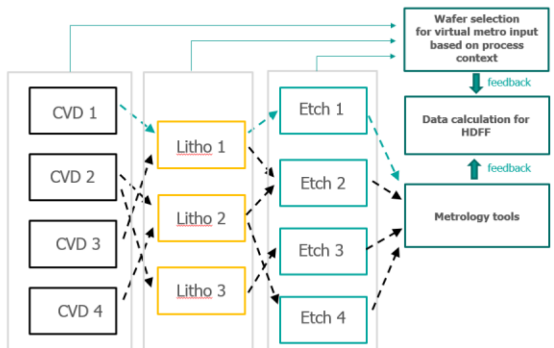

Figure 2 Schematic view of virtual metrology for reconstructing missing HDFF inputs based on wafer's process history: the dashed lines represent wafer dispatching events, while solid lines represent process information flows. The process context information is analyzed to be the same as for the wafer with missing HDFF inputs. For instance, if the OCD model misses the data from the wafer processed with CVD 1/Litho 1/Etch 1, the wafer with process history marked by blue dotted line will be considered for virtual HDFF input calculation.

Manufacturing process flow consists of many consecutive process and metrology steps as shown in Figure 3. depending on production control strategy, the different metrology steps can be defined such as OCD and CD-SEM in hybrid configuration utilizing metrology data from previous steps. There are many variables in process control flow that need to be specified such as sequence of metrology steps, within wafer measurement locations, overall sampling of the production volume. For instance, for APC applications in many cases a sampling needs to be 100 % while for typical production control within-lot sampling can be 20% of the whole volume. There could be differences between site sampling, wafer- and lot sampling due to consecutive metrology steps or product groups. For measurements that require HDFF all sampling differences need to be assumed. The differences in sampling between preceding metrology steps can occur due to missing site or wafer measurements which themselves can be a result of measurement failure, measurement skip or other reasons. Our focus is to ensure that the missing measurement input is somehow recreated. The data recreation can be done by process context knowledge and process history data. The process context includes assumptions of specific characteristic of a given tool. Furthermore, process context can assume FDC (Failure Detection and Classification) data, i.e., some specific signal which indicates that production tool's operation state is normal and given measurement data point can be included.

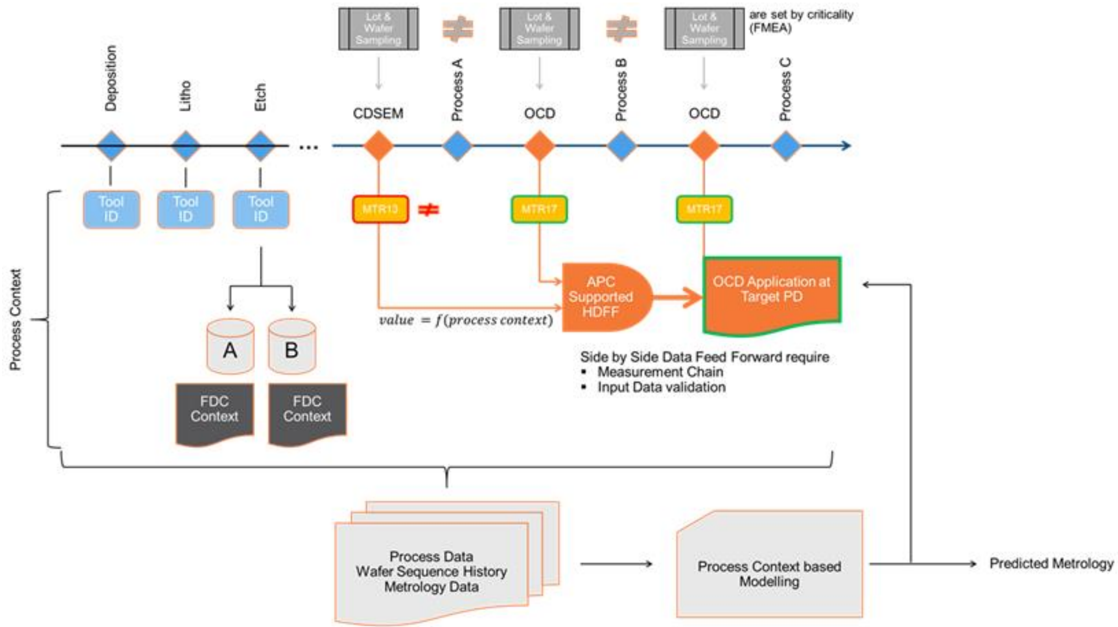

Figure 3 Schematic representation of information flow utilizing virtual metrology.

We specify the variables of the function of process context as follows:

- Every measured parameter has its unique coordinate that describe the location on the wafer wherefrom it was extracted. This we describe in the following formula.

  <parameter> =f (**XY**) (1)

- The run path in a process flow contains different process steps (*PD_ID*) with their tool platforms (*Tool_ID*) and chambers (*Chamber_ID*), called process context. The process context can be a combination of multiple *PD_ID*, *Tool_ID* and *Chamber_IDs*. Based on this information different aggregation level /commonalities to the parameter value can be build.

  <parameter> =f (XY) **(process context)** (2)

- Process characteristics plays a significant role to find other levels of commonalities to metrology results. Example of these are, Chuck temperature zones, MFC flows, chamber pressure, RF power and others found in sensor data in FDC monitoring.

  <parameter> =f (XY) (process context **; FDC**) (3)

- Time-based characteristics. This represent behavior in parameter drifts within *process context level.*

  <parameter> =f (XY ; **Time**) (process context ; FDC) (4)

2.3 Interpolation approaches for missing dies data within wafer

In this chapter we focus on the scenario when the data from just a few dies are missing from preceding metrology step used in HDFF. Figure 4 presents a sketch which represents a scenario with classification and potential solutions. We recognized two major cases which occur in manufacturing: wafer map shows radial or non-radial signature. Furthermore, we need to consider if the amount of measured dies is sufficient to perform interpolation. In the case when the data are insufficient, for instance, only a single die is measured on the wafer, the virtual metrology and machine learning approaches should be applied and they are discussed in detail in the next chapters. Here, we focus only on cases when we have sufficient measured dies data. Furthermore, we assume here that the time-dependent process drift do not play a role in this consideration.

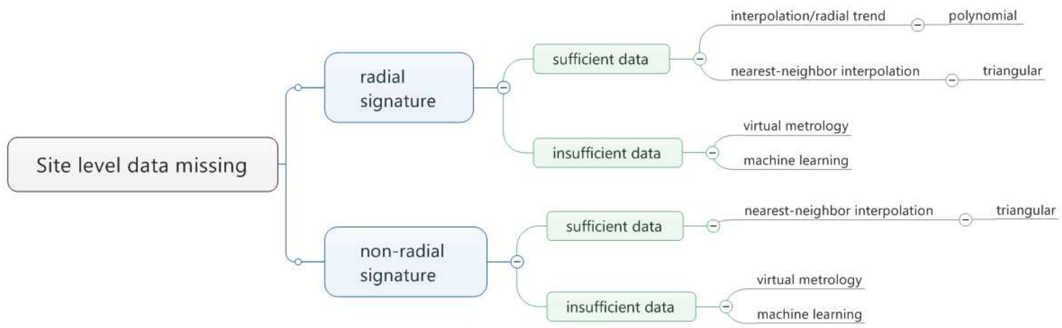

Figure 4 Schematic representation of missing data recreation scheme when site level data is missing

The scenario of missing measurement site(s) inputs can be addressed by different extrapolation and interpolation schemes.[10] However, selection of an extrapolation/interpolation scheme is strongly related to a manufacturing process signature. Typically, a radial process signature can be noticed in many process steps. When the process signature is close to radial or centrosymmetric, simple polynomial extrapolation/interpolation can be applied as shown in Figure 5a.

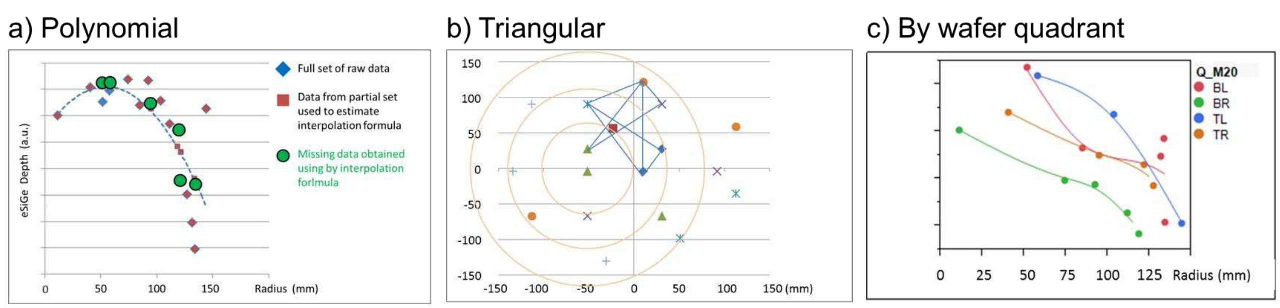

Figure 5. (a) Radial distribution interpolated by second-order polynomial function; (b) Triangular interpolation approach for missing data, and (c) comparison of triangular interpolation data per wafer quadrant vs wafer radius. The symbols in panels b) and c) denote respective wafer quadrants: TL (top left), TR (top right), BL (bottom left), and BR (bottom right).

For the case when single-die data are missing and there is statistically sufficient number of correctly measured dies the nearest-neighbor (NN) interpolation approach (for example, the triangular interpolation) works well. To demonstrate this, we present a case study for data feed-forward approach when the input data are missed from very beginning of the process chain as shown in Figure 6a. Since presented scheme contains a whole chain of passed parameters a possible error in the model should be amplified if the interpolated result is far from a real measurement. To appraise the effect of missing measurements we removed a few data points and then used the nearest-neighbor interpolation scheme to estimate the missing values, feed forwarded them into further analysis and compared the results with real measured output. The study shows that use of the values obtained by nearest-neighbor interpolation results in almost the same performance across all subsequent steps as shown in Figure 6b and subsequent error accumulation is negligible. In summary, for simple cases when a few dies data are missing within wafer, radial approach can be used if the process signature is radial and centrosymmetric. For the more advanced case when the process signature is non-centrosymmetric and amount of correctly measurement sites is statistically sufficient, the nearest-neighbor interpolation approach can be used to re-create missing data. We also should mention a less common, but an elegant and advantageous approach to describe arbitrary process signature on a wafer by means of a class of fairly easy to construct spline functions called B-splines (or Basis-splines).[11] B-splines have been proven to be highly useful for representation of complexly shaped curves and surfaces and currently became pretty popular in optical modeling (see, for instance, Refs.[12,13]). Implementation of this approach instead above-mentioned traditional interpolation/extrapolation schemes will be the subject of a future work.

The missing data problem becomes more significant when there is an insufficient number of correctly measured dies within the wafers or when whole-wafer data are missing. We focus on the latter cases in the next subsection.

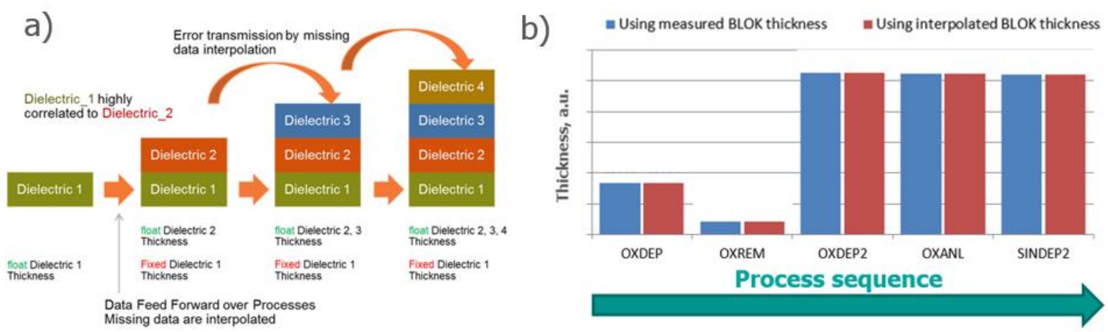

Figure 6 (a) Schematic representing missing data input and error propagation; (b) Comparison of the real vs virtual data input based on the nearest-neighbor interpolation scheme.

2.4 Context based virtual metrology and machine learning for missing data reconstruction

In this subsection we focus on the instances when the metrology data for a whole wafer or significant amount of them are missing and process signature is non-radial and non-centrosymmetric. In that case the data re-construction of missing wafer measurements requires more detailed process knowledge. We distinguished here some key cases illustrated in a sketch shown in Figure 7. In the first case there is a sufficient amount of the data for a short time interval and in the second case the data are insufficient to reconstruct the missing data input. In this case a longer data period needs to be examined and possible time-depended process drift needs to be taken into account.

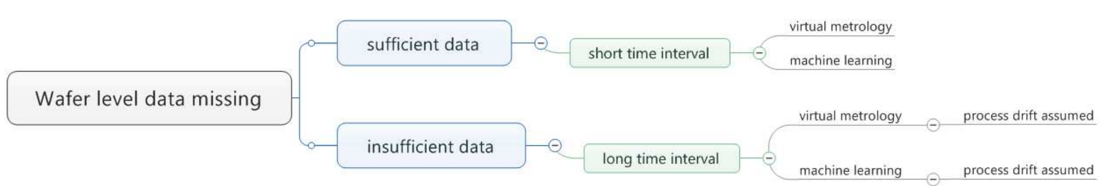

Figure 7 Schematic representation of missing data recreation scheme when full wafer data is missing

The both context-based VM and ML approaches are illustrated in Figure 8. In the process context based virtual metrology we need reconstruct the missing HDFF data. Data reconstruction process includes process context (discussed in detail in subsection 2.2) and process history data that can be averaged on different aggregation levels. Machine learning is another approach to solve missing HDFF data input challenge (Figure 8). The machine learning approach assumes a replacement of hybrid model when HDFF data are missed. An adequate training data set based on hybrid solution output and optical spectra need to be used. The key importance for machine learning approach is a training set that covers all expected variation. The advantage of mixing hybrid solution and machine learning solution (when hybrid input is missing) is higher confidence in the data. Data resulted from the machine learning can be compared with those obtained from ordinary way of using optical modeling in production. In the case of missing data for a whole wafer, the machine learning solution can be scored with process context knowledge. By this way there is higher confidence overall level for any excursion detection. Furthermore, the ML model can be updated (retrained) when corresponding confidence score will indicate such necessity.

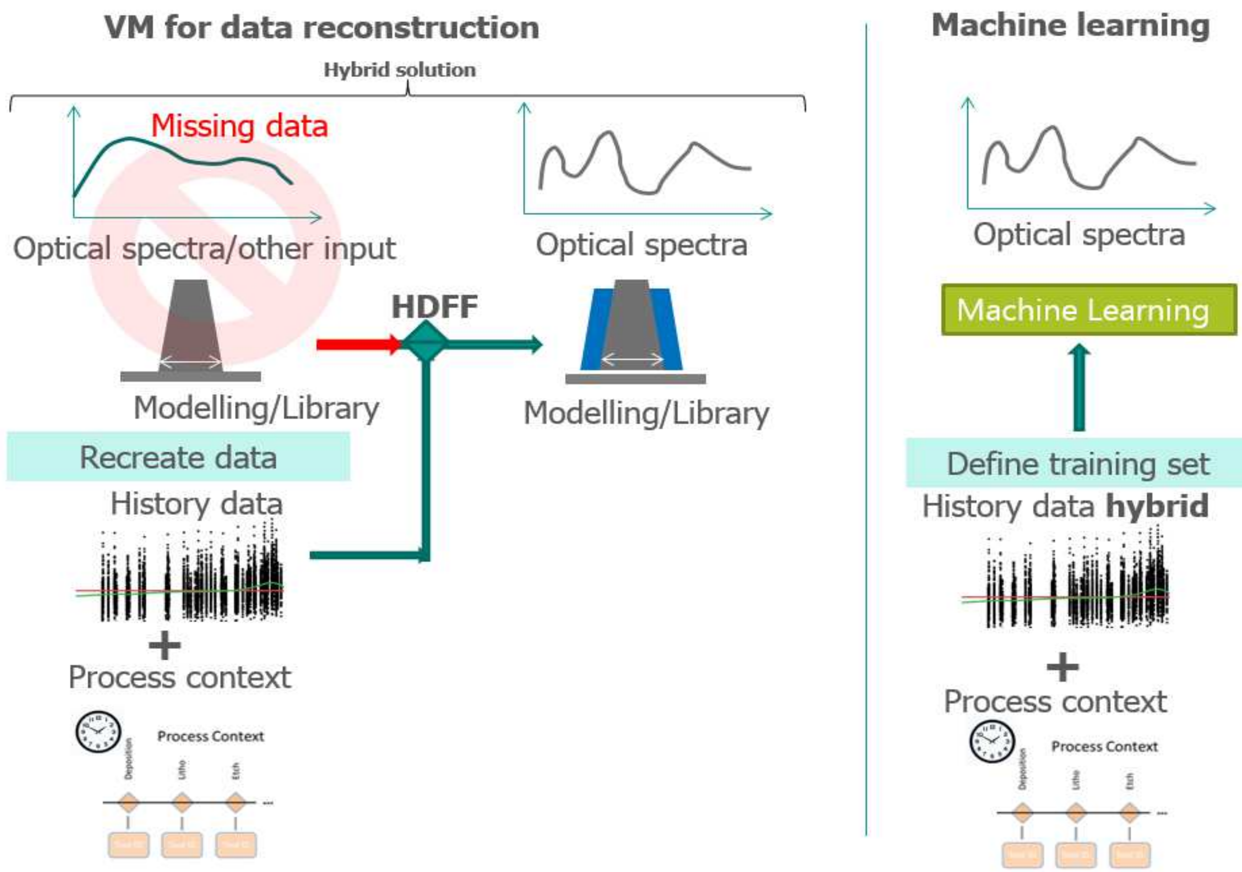

Figure 8 Concept of two approaches to solve the problem of missing HDFF data: virtual metrology and machine learning

To demonstrate how both proposed approaches work in practice, we use as example the results from HDFF solution for process steps and structures shown in Figure 9. Several parameters measured at the TJ-FI step (left picture in Figure 9) were passed to the following TJ-DEP metrology step (right picture in Figure 9). TJ-FI is defined by several process steps starting from gate layer depositions, gate patterning, spacer deposition and final cavity processing. At first, we focus on reconstruction of the cavity height parameter using process context-based virtual metrology. The cavity height parameter in the TJ-FI step depends mainly on the etch conditions (an influence from another process steps can be neglected in process context analysis).

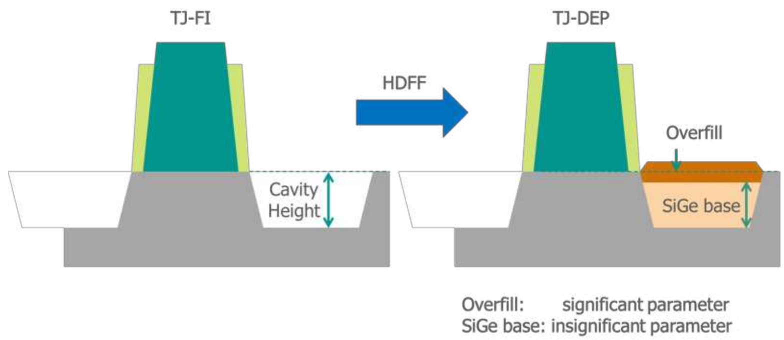

Figure 9. Schematic representation HDFF configuration. Cavity depth and other dimensions as TJ-FI are passed to TJ-DEP step.

Accuracy of extracted virtual data depends on amount of available data within the same process context and data aggregation level. The easiest way to reconstruct missing wafer level data is to estimate the average value of the whole data set. A statistic related to cavity height parameter is presented in Figure 10. Distribution for the whole data set can be presented on several levels. Figure 10a shows schematically multiple wafers and related site level data. The example shows three dies marked as A, B, and C at different locations. The lot-level distribution of the cavity height data set is shown in Figure10b where A, B, and C site level data are marked by circles. The median and average values of the lot-level data are very close and the whole set demonstrates a normal (Gaussian) distribution. Passing HDFF data input mean value as the whole set distribution would be a first approximation and it might lead to a significant error. To illustrate that we present the same data set but on die level in Figure 10c. It is clear that the average data value from the site level is a better approximation compare to the lot-level value. The edge site C average number differs the highest for center site A and middle site B. The site level averaging can work well when process drift is minimal (time interval is short). The site level averaging reflects significantly better the larger ranges in uniformity within wafer and/ or be independent on process signature (work for radial or non-radial wafer maps). One of the concerns for site level scenario is sufficient data set to achieve statistical significance of the die average value. On the other hand, when production volume is low the collecting more data can prolong time interval and therefore time dependent process drift need to be included.

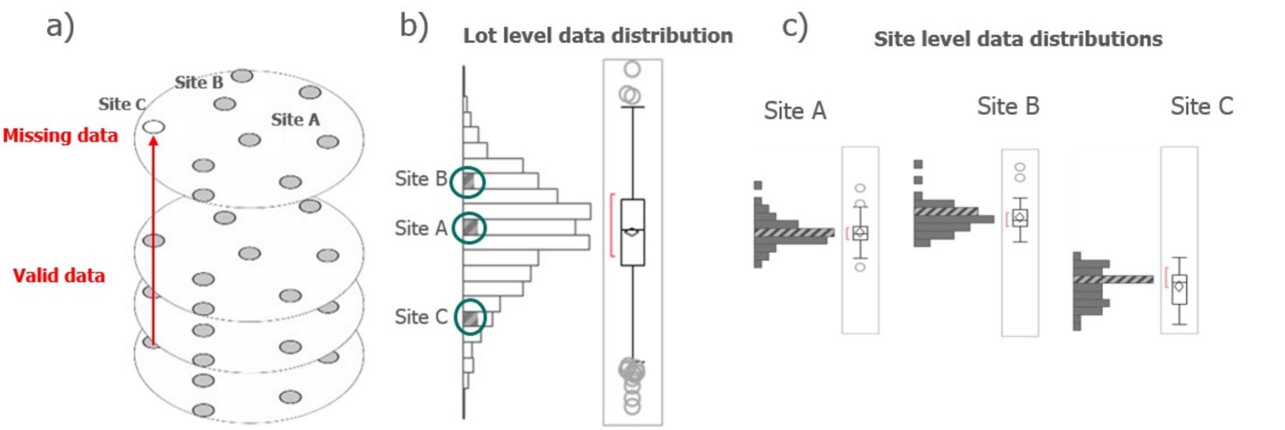

Figure 10 (a) schematic representation of site level data aggregation within large wafer set. (b) distribution of total data set with markups on the site level (c) the site level data distribution referenced in sketch (a) and histogram (b).

Despite purely statistical considerations the next level would be to include process context. In the case of TJ-FI the cavity height parameter is mainly defined by etch process (other process steps have negligible effect). In order to see the process chamber effect the data can be on process tool level as shown in Figure 11a. (the site A, B, C level data are also marked). It can be clearly seen that the tool's effect is negligible. However, more accurate picture can be achieved splitting data on chamber level as shown in Figure 11b. In this case it is clear that *Tool_A/Ch_A* is clear offset to other Chambers as *Tool_A/Ch_B*. In this case process context need to be used and it will result in better approximation of the missing data. The data reconstructed on four chosen levels are presented in Figure12. It is clear that etch tool level data in Figure12a show less variation that etch chamber level data (Figure 12b). Even more variation between averaged data is visible on the site level.

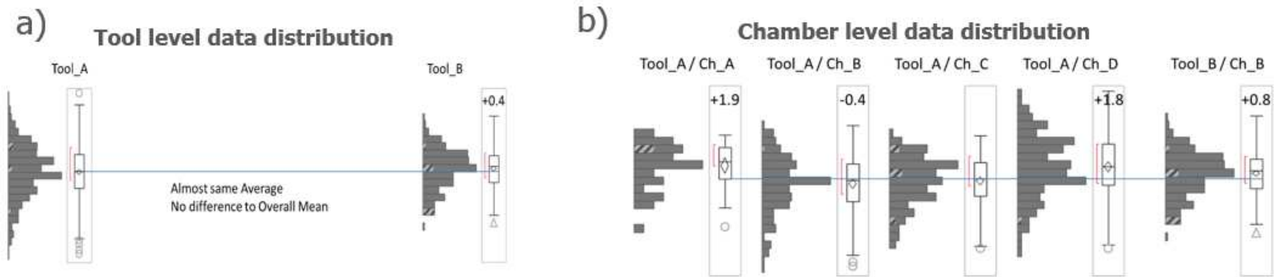

Figure 11 Statistical distribution for the same data set for two data aggregation levels: a) tool level, b) chamber level.

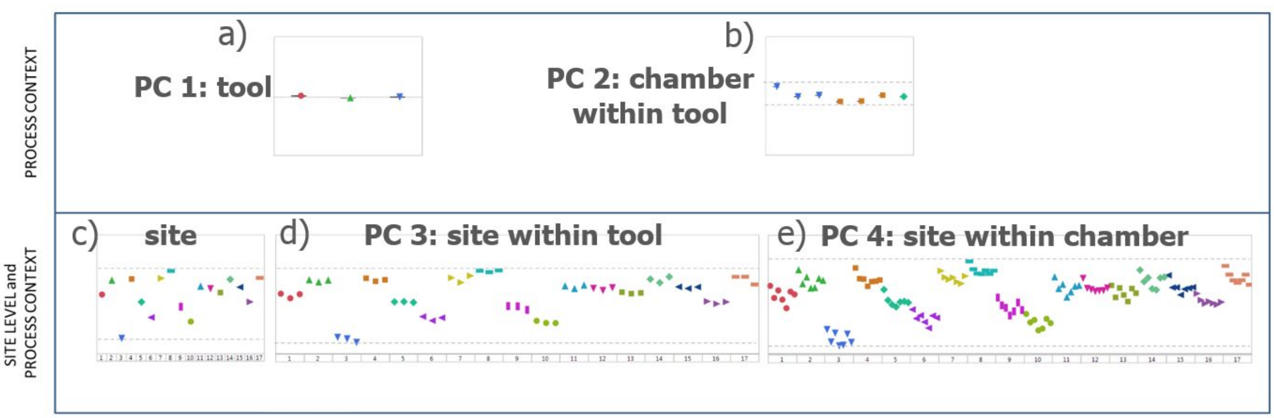

Figure 12 Data reconstruction using context-based virtual metrology: a) the etch-tool level, b) the etch-chamber level, c) the site level (no process context), d) the site level data on the tool level, and e) the site level data for the etch-chamber level.

Next we study the effect of reconstructed data as HDFF input on final output of hybrid TJ-DEP model. We have chosen two output parameters of interest. One of the parameters, namely, the overfill, parameter is highly influential (i.e., the spectral diffrences are singificant due to variations of this parameter) and another prameter is insignificant. We expect to see more effect on less important parameter since it is more prone to mutual correlations which are suppressed by HDFF solution. In order to perform this study, we create the folowing scenario. We used the data from dosens of production lots. For all selected data we had no missing HDFF data and model output was 100% according to process specificaitons. We call this dataset a "POR set". Next we correlated the POR results to the results obtained using the same dataset but without HDFF input. The whole hybrid interpretation was done assuming fixed/nominal value of passed parameters. Next, we correlated the POR results to those evaluated with HDFF fixed at nominal values for overfil and SiGe base parameters, as shown in Figure 13. In this case a linear coefficient of determination R^2 was 0.9595/and a slope s = 1.0006 for overfil, a very influential parameter. The effect due to missing HDFF input is visible more on insignificant parameter where R^2=0.7126 and s = 1.6119. We then study the effect of HDFF data recreation using four virtual metrology scenarions. For PC1 case the recreated data were just average of the datasets on etch tool level (see also Figure 12a). We see some improvement in the R^2 value for overfill parameter while small degradation of R^2 occurs for

SiGe base parameter. Similar situation is observer for PC2 case where reconstructed data are averaged out on chamber level (see also Figure 12b). Next we tested the same process context scenarios but we changed the data agrrevation level from wafer to site level. We obesrved an improvemement in R^2 for both overfill and SiGe base parameters. This shows that not only process context but data aggrevation level plays a role. Furthermore, the result agrees with our expectations about improvements in slope and R^2 with higher aggregation level (see Figure 10).

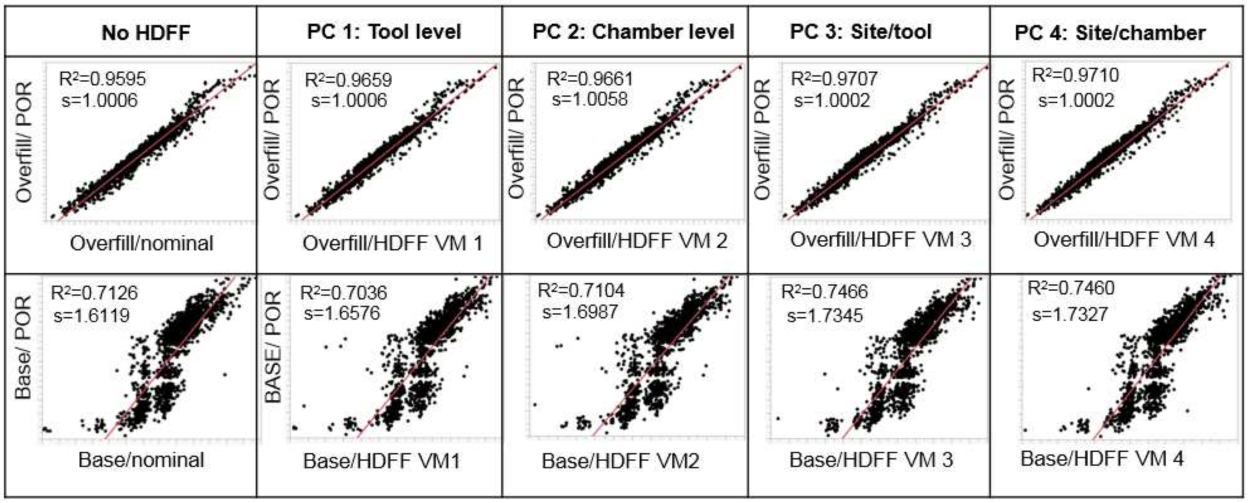

Figure 13 Correlation of the POR hybrid production data for overfill and base parameters of TJ-DEP application: (No DFF) the hybrid data were missing and interpretation was done with nominal values; (PC1) the data were reconstructed on the tool level; (PC2) the data were reconstructed on the chamber level, (PC3) the data were reconstructed on both site and tool levels, and, finally, (PC4) the data were reconstructed on the site/chamber level.

We look for futher improvement of virtual data input. For that purpose we need to assume time dependent process drift. In order to demostrate effect of process drift we plotted time-changing base height values for three etch tools at TJ-FI step. It is clear that the data averages change in time due to process drift and this fact contributes to an error of virtual metrology input estimation. The larger the time range of the sample dataset, the higher effect of the time-dependent process drift can be observed. Other errors can be due different product types running on the same route. Although the process sequence is the same, the pattern density and wafer locations of measured sites might differ across different products.

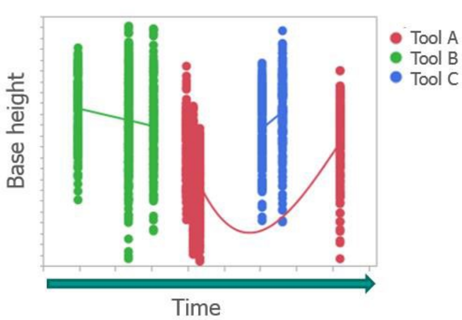

Figure 14 Tool data within studied set vs data acquisition time. Product-level data shows also different performace. The data reconstruction should be done on the product level.

After detailed study of the context-based VM for HDFF reconstrution, we focused on the machine learning approach. As shown in Figure 8, the ML approach is meant to replace HDFF, which typically requires two or more measurment steps for TJ-DEP, with a single ML measurment step. The previous consideration realted to VM and process context are also relevant for ML. Process context can be used to define optimal training set for the ML solution. Furthermore, since we propose HDFF solution when there is no missing HDFF data and ML when data are missing. The resulting production data set is mix of ML and HDFF solution data. By comparing the data set with process context knowledge the confidence metrics in training set can be build. Later this confidence metrics can be used to signalize the need for auto re-train of the ML model. In order to benchmark the efficiency of ML for the data reconstruction we use the same scheme and data set as we used for VM (see Figure 13 and Figure 15). Our benchmark data are the POR hybrid data. However, for ML solution the most important is the training size of final result of the data recreation. We are comparing four different training size on final output: ML1 to ML4. The ML4 is the training that involves more than ten lots of hybrid data. It is clear that the bigger training size will result in the better correlation for influential (overfill) as well as less important (SiGe base) parameters. We observe, as we had for the VM data reconstruciton, that correlation for insignificant SiGe base parameter gives worse results than in the case of overfill parameter. Other important observation is an effect of training size on the final data reconstruction quality. As expected, the lesser the training size, the lesser the correlation to the benchmark data. The correlation of ML1 is better than ML2 where only three lots are selected for traing. In the case of ML3+PC when only eight wafers were selected but process context was also used for selection. Here, still correlation is overal worst than M1 and ML2. And finally, for the research purpose we compare the training size for one of the wafers with 17 measured dies. Interestingly, we still have some correlation for influential overfill parameter but worse than in case of No HDFF. Too small size of the training set is very strongly appeared for insignificant SiGe base parameter. Beside the training size, it is worth to mention that the process context matters for data selection in the ML training. If we compare the slopes of the overfil parameter for ML2 and ML3, we wil note that the process context based training provides a better slope. And, finally, we compare VM (Figure 13) and ML (Figure 15) approaches. Both of them provide an improvement in the case when input data are missing. Furthermore, the ML approach shows better performance than the VM apprach except the ML1 case with single wafer training. This could be due to a fact that the ML training is performed on the die level and also assumes all discrete process context data. Obvioulsy, without process context knowledge training selection could be inappropriate and final result might be much worse.

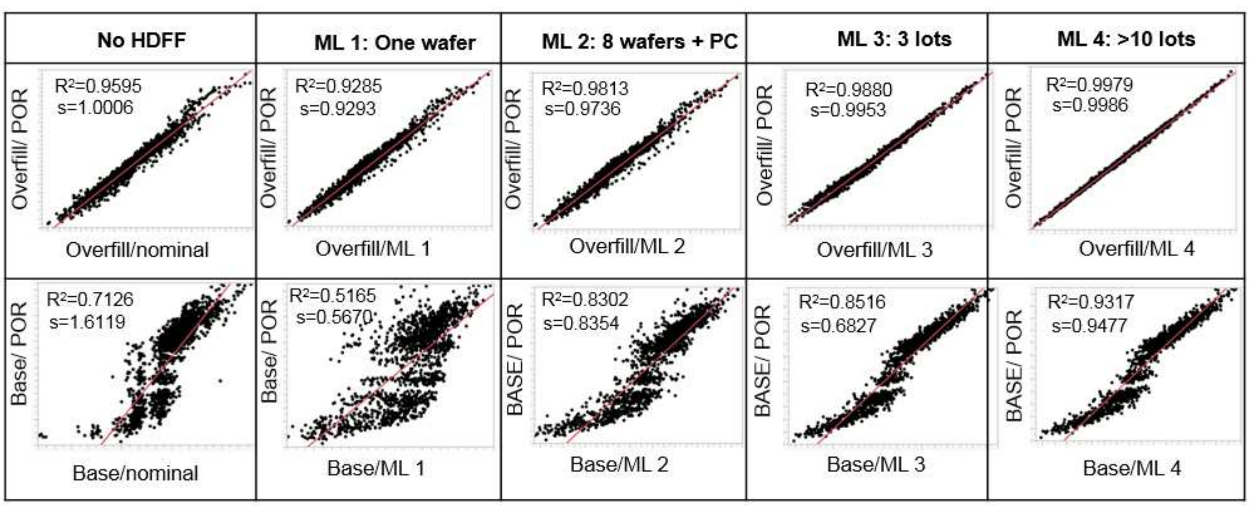

Figure 15 Correlation of POR hybrid production data for overfill and base parameters of TJ-DEP HDFF application: (No HDFF) the hybrid data were missing and interpretation was done with nominal values; (ML4) with training set larger than 10 lots; (ML 2) training with three wafer lots; (ML 3) training with eight wafers chosen based on process context knowledge, and, finally, (ML 1) training with single wafer.

3. Summary

We addressed the missing data problem in complex hybrid data feed-forward applications. We classified missing data in two major groups: missing dies data within the wafer and data missing for a whole wafer/lot. For scenarios when only the data from several dies are missing, we propose some interpolation schemes selected using information about process signature. For instance, in the case of radial process signature, the polynomial interpolation is sufficient and for the non-radial signatures various nearest-neighbor interpolations, such as, for example, triangular NN interpolation, need to be applied. We also described two approaches for more complex case where there is insufficient amount of measured dies within wafer or whole wafer/lot data are missing. First approach relays on the process context and virtual metrology and the second one uses the machine learning together with process context consideration to select an optimal training set. We described key variables of process context that are relevant for missing data reconstruction. We demonstrated that for virtual metrology approach despite process context data aggravation level plays an important role. The best result was achieved on die level data aggregation and assumption of relevant process context such as etch chamber level split which has impacted a key parameter. We addressed the time dependent process drift impact on the data reconstruction. As a second approach to solve missing data problem we use machine learning model. We replaced hybrid configuration involving several metrology steps with the single-step machine learning solution when HDFF data are missing. We demonstrated that a quality of the data reconstruction depends on training size and the process context information can be used to define more efficient training set. Finally, we compared the results of both virtual metrology and machine learning approaches. Both approaches improve the solution in the case when forwarded data are missing. The machine learning approach shows better results in comparison to virtual metrology approach. Possible reason is that the training set automatically assumes all process context and data aggravation on the die level. As an advantage of mixing the VM and ML approaches, one might consider the fact that the inline confidence metrics will indicate the ML model needs to be retrained.

REFERENCES

[1] A. Vaid, B. B. Yan, Y. T. Jiang, M. Kelling, C. Hartig, J. Allgair, P. Ebersbach, M. Sendelbach, N. Rana, A. Katnani, E. McLellan, C. Archie, C. Bozdog, H. Kim, M. Sendler, S. Ng, B. Sherman, B. Brill, I. Turovets, and R. Urensky, "A holistic metrology approach: hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM," Proc. of SPIE, 7971, 797103-797103 (2011).

[2] A. Vaid, A. Elia, G. Iddawela, C. Bozdog, M. Sendelbach, B. C. Kang, P. K. Isbester, and S. Wolfling, "Hybrid metrology: from the lab into the fab," Journal of Micro/Nanolithography, MEMS, and MOEMS, 13(4), 041410-041410 (2014).

[3] J. Moyne, and J. Iskandar, "Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing," Processes, 5(3), 39 (2017).

[4] Y.-C. Jonathan Chang, and C. Fan-Tien, "Application development of virtual metrology in semiconductor industry," Proc. IECON 2005., 6 pp. (2005).

[5] G. A. Susto, "A Dynamic Sampling Strategy based on Confidence Level of Virtual Metrology Predictions " Proc. of ASMC, (2017)

[6] H. Purwins, B. Barak, A. Nagi, R. Engel, U. Höckele, A. Kyek, S. Cherla, B. Lenz, G. Pfeifer, and K. Weinzierl, "Regression Methods for Virtual Metrology of Layer Thickness in Chemical Vapor Deposition," IEEE/ASME Transactions on Mechatronics, 19(1), 1-8 (2014).

[7] A. Vaid, G. Iddawela, J. Tsai, G. Wainreb, P. Isbester, B. C. Kang, M. Klots, Y. Katz, C. Bozdog, and M. Sendelbach, "Improved scatterometry time-to-solution using virtual reference," Proc. SPIE Advanced Lithography, 9424, 94240X (2015).

[8] F. T. Cheng, H. C. Huang, and C. A. Kao, "Developing an Automatic Virtual Metrology System," IEEE Transactions on Automation Science and Engineering, 9(1), 181-188 (2012).

[9] W. Jian, S. Pampuri, P. G. O. Hara, A. B. Johnston, and S. McLoone, "On regression methods for virtual metrology in semiconductor manufacturing," Proc. ISSC 2014/CIICT 2014, 380-385 (2014).

[10]J. S. Armstrong, “Forecasting by Extrapolation: Conclusions from 25 Years of Research,” Interfaces, 14(6), 52-66 (1984).
[11] C. de Boor, *A Practical Guide to Splines*, Revised Edition, Springer-Verlag, New York NY, 2001.
[12] B. Johs, J. S. Hale, “Dielectric function representation by B-splines”, Phys. Status Solidi A 205(4), 715-719 (2008).
[13] D. V. Likhachev, “Selecting the right number of knots for B-spline parameterization of the dielectric functions in spectroscopic ellipsometry data analysis”, Thin Solid Films 636, 519-526 (2017).