IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 17, 2021, accepted April 20, 2021, date of publication April 28, 2021, date of current version May 6, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3076193*

# A Realizable Overlay Virtual Metrology System in Semiconductor Manufacturing: Proposal, Challenges and Future Perspective

**TZE CHIANG TIN[1], SAW CHIN TAN[1], (Senior Member, IEEE), HING YONG[2], JIMMY OOK HYUN KIM[2], ERIC KEN YONG TEO[2], CHING KWANG LEE[3], (Senior Member, IEEE), PETER THAN[2], ANGELA PEI SAN TAN[2], AND SIEW CHEE PHANG[2]**

[1]Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63100, Malaysia
[2]X-FAB Sarawak Sdn. Bhd., Kuching 93350, Malaysia
[3]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia

Corresponding author: Tze Chiang Tin (alvin.tin.i@gmail.com)

This article was supported by the Multimedia University, Cyberjaya, Malaysia

**ABSTRACT** Integrated circuits (IC) are fabricated on a wafer through stacked layers of circuit patterns. To ensure proper functionality, the overlay of each pattern layer must be within the tolerance. Inspecting each wafer's overlay is unrealistic and impractical. Hence, wafers are selectively inspected at metrology stations through sampling strategies. With virtual metrology (VM), the metrology quality of the uninspected wafers can be estimated. Motivated by a real-world production environment of a 200mm semiconductor manufacturing plant (fab), a VM to estimate the overlay of the photolithography process is envisioned. Past researches on overlay VM leveraged fault detection and classification (FDC) data to estimate the overlay errors. As such, for fabs in the progress of completing their FDC development for photolithography equipment, a different modeling approach is required to realize an overlay VM that sustains the production line until FDC data can be leveraged for VM. With practical gaps that must be addressed in real fabs, this paper focuses on realizing an overlay VM for the photolithography process without leveraging FDC data. Therefore, the objectives of this paper are two folds: First, to identify the research challenges towards realizing the overlay VM. Second, to propose the future research perspectives of the envisioned overlay VM. Based on the future research perspectives, a two-steps overlay VM modeling approach utilizing data mining techniques is proposed toward realizing the envisioned overlay VM system. The proposed approach first classifies the process stability at the wafer lot level, and subsequently, performs overlay error estimations for wafers in the wafer lots classified with stable process. Linear regression models are proposed to perform overlay error estimations in this work to augment the interpretability of the overlay VM.

**INDEX TERMS** Virtual metrology, photolithography, overlay, classification, regression.

## I. INTRODUCTION

Semiconductor manufacturing is a composite process that transforms raw wafers into computer chips. The entire process manufacturing process consists of four distinct stages of operations. The first stage is the wafer fabrication stage. In this stage, bare silicon wafers first enter the wafer fabrication step to produce integrated circuits (IC) in the form of dies on their surface through a repetitive process of a series of sequential processing steps. The fabricated wafers are then sent to the second stage of the manufacturing process, called wafer test, to inspect the electrical properties of the wafer dies. Wafer dies that past these electrical tests are then sent for assembly to be packed as computer chips, which is the third stage of the manufacturing process. At the last stage, the final test on the functionality of the chips is performed to ensure that only chips that pass the tests are shipped to the customers.

The required IC design is translated onto the surface of the wafer layer by layer through sequential steps. The processing steps involved to translate a layer of IC design onto the wafer surface can be categorized into seven steps: lithography,

The associate editor coordinating the review of this manuscript and approving it for publication was Okyay Kaynak.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

etching, deposition, chemical mechanical planarization (CMP), oxidation, ion implantation, and diffusion. These steps are categorized as process steps, as opposed to metrology or inspection steps that examine the qualities of the wafers. The following describes these major fabrication steps.

These fabrication steps are presented through the following sequence [1], [2]: lithography, etching, deposition, chemical mechanical planarization (CMP), oxidation, ion implantation, and diffusion.

The lithography process step, also known as the photolithography process step, typically marks the beginning of a new layer. In this process step, the pattern of the IC is imprinted onto the surface of the wafer through a mask (or reticle). The wafer surface is first coated with photoresist material. Then, a mask that contains the IC design pattern of the layer is placed on top of the wafer in close proximity in a machine called the stepper for the pattern imprinting on the wafer's surface. The pattern imprinting is accomplished by exposing the mask and the wafer under ultra-violet light. The exposed resist will be removed in the subsequent step, leaving the protected area to represent the pattern imprinted.

At the Etching process, IC patterns are created on the surface of the wafer by selectively removing the deposited material. An etching mask is applied on the wafer surface to protect the area that should remain. The material in the unmasked area is then removed either through dry (physical) etching or wet (chemical) etching. The etching processing occurs in a localized environment called chambers of the etching machine.

At the deposition process, thin organic and inorganic films are deposited into the wafer to either create interconnectors of the IC or as intermediate layers for specific processing step and subsequently removed after the processing step is completed. Deposition methods include physical vapor deposition (PVD) and chemical vapor deposition (CVD), with the former utilizes the sputtering of accelerated ion gas, while the latter utilizes the chemical reaction from a mixture of gasses at high temperatures.

At the chemical mechanical planarization (CMP) process step, a planarization process is applied to the wafer to flatten its surface to create a non-planar surface. This surface conditioning step is crucial for the lithography process for the correct transfer of the IC pattern. CMP process performs the planarization process with the help of designated chemical slurry.

At the oxidation process, silicon on the wafer surface is converted to silicon dioxide, which is a layer necessary for the ion implantation process later. In the modern fabrication process, thermal oxidation is commonly used for the conversion process. In thermal oxidation, a wafer is exposed to oxygen or water vapor in a furnace environment for elevated temperatures for effective conversion of the silicon oxide.

At the ion implantation process, dopant impurities are introduced into the wafer to modify its electrical properties. Boron, arsenic, phosphorus, and antimony are examples of impurities used. Ionized dopants are energized through electrical field acceleration so that they can penetrate the wafer to the desired depth. At the diffusion process, implanted dopants are shifted across wafer sites at an elevated temperature. Dopants shifting is necessary to adjust their concentration across the wafer's site. The diffusion process is used to introduce dopant into wafer from dopant gas. The diffusion time and temperature determine the implantation depth. The diffusion process is also used to perform thermal oxidation in order to obtain the silicon oxide layer.

The wafer fabrication process to produce IC required for today's technology may require 350 process steps or more. With such a long processing sequence, and each process step having its own process characteristics, wafer quality has to be ensured throughout the various stages of the fabrication process. Metrology steps are therefore placed in between designated process steps to perform the quality checks. Wafer quality can be determined by measuring critical parameters to ensure that these parameters are within the product specification limits and inspecting the surface of the wafer to ensure no physical damage, defects, or unwanted particles are found on the wafer. A failure in the metrology steps indicates the presence of abnormalities such as equipment parts failure or process performance drifts. Metrology steps are therefore crucial to ensure wafer quality and maintain yield.

Although metrology steps are crucial to ensure wafer quality, they are considered non-value-added steps to the production. In a high-throughput wafer fabrication foundry (fab) where thousands of wafers are being processed each day, it is both costly and inefficient in terms of time and manpower to carry out metrology on every wafer. Hence, a sampling approach is conventionally used. In a fab, a wafer lot is a single unit quantity of wafer transportation cassette that typically contains 25 wafers. Using the sampling approach, metrology is performed after a certain number of lots are being processed for a particular recipe-equipment pair, and for the lot sent for metrology, a subset of the wafers are selected for metrology. The metrology results obtained from these wafers are used to represent the quality of the entire lot. In terms of cycle-time (CT) of a fab, the sampling approach is beneficial as it shortens the CT of the manufacturing process. However, in terms of wafer quality assurance, the sampling approach is undesirable. To address this gap, virtual metrology (VM) has emerged as the means to examine the un-sampled wafers in a lot without incurring additional cost to cycle-time.

VM is defined in the literature as mathematical models that conjecture, estimate, or predicts the target metrology variables of interest by utilizing historically sampled metrology measurements, process, and equipment state information [3], [4]. The realization of VM as a low-cost solution that compliments the physical metrology to bridge the aforementioned gap is made possible owing to the availability of substantial computing power and storage at low cost.

VM has been actively studied by researchers for the past decades for various fabrication process steps by leveraging FDC data to provide real-time depiction of fabrication process characteristics. FDC is still actively researched in recent

years for various improvements to augment its capabilities to handle various complex process characteristics in wafer fabrication. Examples of these works can be found in [3]–[7], and [8]. As such, active FDC development works are necessary for real fabs to apply the various new approaches for effective fault detection and classification.

In the event that the FDC development is in progress or undergoes various improvement activities that render its data unavailable until it is qualified for production use, FDC data could not be leveraged to realize VM similar to past researches. Hence, a different modeling paradigm is required to realize an overlay VM to sustain the production line until FDC is ready for production use and its data can be leveraged by VM. Motivated by a real fab production environment of a 200mm semiconductor manufacturing plant (fab), this research focuses on realizing an overlay VM for the photolithography process without leveraging FDC data to cater for such an event. The realization of the envisioned VM that does not leverage FDC data is non-trivial with the presence of practical challenges that must be addressed. Therefore, the objectives of this article are two folds: First, to identify the research challenges towards the realization of the envisioned VM. Second, to propose the future research perspectives towards a production overlay VM that does not leverage FDC data for the real-world production settings.

The presentation of this paper is therefore organized as follow: Section II presents the literature review of the related research works, Section III presents the research analysis and challenges, Section IV presents the future research direction and the proposed VM model, and finally, Section V presents the conclusion of this work.

## II. RELATED WORKS

This section presents the literature review of related virtual metrology (VM) works for the past decades.

In [9], the authors presented a faulty wafers detection method for batch processes using k-Nearest Neighbor (kNN). According to the authors, statistical process control (SPC) has traditionally been used to identify faulty wafers in semiconductor manufacturing. However, SPC is incapable of multivariate detection approaches. On the other hand, the widely explored principle component analysis (PCA)-based and partial least squares (PLS)-based detection for multivariate fault detection were inefficient for batch processes as a result of unique process characteristics involving nonlinearity, multimodal trajectories as a result of product mix, and process time variation. PCA or PLS-based methods require the unfolding of data into 2 dimensional (2-D) data array, thus requiring a significant amount of data to build a reliable detection model and restricting automated solution for online applications. Hence, employing FDC and kNN, the authors proposed kNN based fault detection to overcome the PCA-based detection method. As a nonlinear classifier, kNN is capable of handling data with nonlinear characteristics. With its simplicity and flexibility, kNN is also suitable for practical use in the production system. Differing from fault classification, faults in fault detection cannot be defined and thus, labeled upfront to train the model. Hence, the authors adapted the conventional rule from kNN by considering the distinction between the trajectories of the samples: should an incoming trajectory is faulty, it will exhibit deviations from those of the normal ones. Hence, by considering only the trajectories of the normal samples and distance thresholds of the training samples, the proposed model can achieve fault detection for batch processes. Employing an industrial example from the etching process, the proposed approach performed better than PCA-based approaches in certain cases.

In [10], the authors presented a VM model that can capture the process variations in multiple inputs and multiple outputs (MIMO)-based processes. The predicted output is then utilized by the authors to create a wafer-level run-to-run (R2R) feedback process control scheme. Prior to being able to integrate into a process control scheme, the VM model must be able to capture the process variations in a MIMO process. Therefore, the authors first presented a detailed formulation of the VM modeling using Partial least squares (PLS) that take into consideration metrology delays and two types of process drifts: consistently and suddenly. Various statistically summarized variables obtained from fault detection (FD) systems are used for the model's input. After the VM model is developed, the quality of each predicted measurement needs to be assessed quantitatively so that only good predictions are fed into the process control scheme. Hence, in this work, the authors also proposed quality metrics to perform the required quality assessment. The double exponential weighted moving average (dEWMA) was used as the baseline to evaluate the efficiency of the proposed control mechanism using VM. By using data from the simulated MIMO process, the experimental results showed that R2R controlling scheme utilizing VM data achieved superior results compared to R2R controlling scheme without the VM data.

In [11], the authors attempted to identify faulty wafers in semiconductor manufacturing through VM. The motivation was stemmed from the fact that early detection of faulty wafers can avoid unnecessary resource consumption and in some events, avoid the wafer being scrapped. However, applying VM in such detection has two challenges. Firstly, an optimum number of model inputs need to be selected from high-dimensional datasets containing measurements from both machine sensors and physical metrology. Secondly, wafer metrology data samples required to build the VM model are usually small as physical metrology only measures a minimum number of wafers per lot. In the presence of such adversarial data characteristics, the authors employed data mining techniques to develop a VM model with high accuracy. In this work, two VM models were built using the same techniques to address the etching process of two etching equipment that have been suspected to cause wafer scrap in a real production environment. VM1 denoted the VM of the first etching equipment while VM2 denoted the VM of the second etching equipment. The data acquisition of the first etching equipment involved 48 sensors while the second

equipment involved 56 sensors. Each etching process has 8 unique steps with all the involved sensors operating during the process. Deriving 4 summary statistics for each sensor's FDC data, a total of 1536 input variables were available for VM1 while 1792 input variables were available for VM2. In real-world practice, fabs only measure one wafer out of 25 wafers for each wafer lot at the physical metrology station. Hence, the data available to develop the two VM models were significantly low with only 118 wafers' data available for VM1 while 241 for VM2. To develop the VMs, the authors first applied data pre-processing steps by converting them to the relevant data structure, followed by dimension reduction. The VM models were then trained using the input variables obtained from the dimension reduction schemes. The large numbers of available input variables were reduced significantly to obtain the crucial ones for both VMs. With only a small data sample size, the VM models built by the authors were still able to perform abnormal wafers detection very well with no misclassification on the normal wafers. The regression models that achieved the best results were linear regression and support vector regression (SVR).

In [12], the authors proposed the use of stepwise selection (SS) for an artificial neural network (ANN) to achieve optimum variable selection for VM models that employ ANN as its prediction algorithm. According to the authors, the widely used Multi-regression-based (MR-based) SS has three limitations that lead to lower predictions for the ANN-based VM model. First, the features obtained from MR-based can be suboptimum. Second, the use of MR-based SS often requires subjective judgment based on engineering and domain knowledge when high-dimensional variables are present. Third, the variables selected as features using MR-based SS may not be suitable for ANN-based VM. In this work, the etching process considered has 12 unique steps with 36 equipment sensors involved to monitor these process steps. Among the input variables, 66 important ones were identified by the process experts. These variables formed the expert-recommended (ER-based) feature set and served as inputs to the MR-based SS. The dataset consisted of 248 wafer lots with 25 wafers or less. 247 sets were used to construct the proposed VM while the last set was reserved as the prediction set. For the 247 sets, the normal wafer sampling procedure was applied to measure only one wafer per wafer lot, while the prediction set has all 25 wafers measured to assess the capability of the proposed VM in predicting the metrology values of all 25 wafers. Using the proposed method, a preliminary list of input variables is first obtained using MR-based SS from the 66 input variables identified by the process experts. Then, forward and backward elimination processes are carried out repeatedly by the ANN algorithm selected to perform prediction on wafers' critical dimension (CD) from the etching process. One-hidden-layer backward propagation neural network (BPNN-1), simple recurrent neural network (SRNN), and generalized regression neural network (GRNN) were employed as the algorithms to evaluate both ANN-based SS and VM's conjecture models. The results from the experiment demonstrated the highest prediction accuracy was obtained by the proposed ANN-based SS, and all three ANN models achieved similar prediction performance.

In [13], the authors presented a VM model that predicts the etch bias (EB), which is the critical dimension (CD) difference between the two patterns etched by plasma etch equipment. The plasma etch equipment considered in this study has dozens of sensors sampling the process characteristics at the frequency of 1-2Hz for each wafer. With fabrication's duration ranges from several minutes to an hour, thousands of sensor readings can be collected for each fabricated wafer to depict the process characteristics. Such datasets collected from the equipment typically contain characteristics of high-dimensionality, varying data structures, collinearity, and non-linearity interactions between the variables, to name a few. Variable selection and outlier removal were the two steps deemed crucial by the authors in order to build a VM model that is both robust and reliable in the presence of such challenging data characteristics. According to the previous studies, stepwise selection, random modeling and genetic partial least squares (PLS) methods have shown descent results in variable selections. Principal component analysis (PCA) was employed by authors for outlier removal. Three VM models were then built using linear PLS, stepwise regression, and BPNN to determine the best combination of the aforementioned variables selection and outlier removal techniques that is capable of the highest prediction accuracy. The best experimental results were obtained using stepwise variable selection and the BPNN prediction algorithm.

In [14], the authors presented a VM using the data mining approach to predict the overlay metrology qualities of the photolithography process of a real fab. Two chucks of the photolithography equipment were studied in this work with a total of 37 equipment sensors involved to monitor the fabrication process of each wafer. 1612 wafers' sensor and metrology data were collected for chuck 1 and 1563 wafers for chuck 2 over a period of 8 months. Deriving 4 summary statistics of minimum, maximum, mean, and variance, a total of 148 statistical sensor parameters were made available to construct the proposed VM that predicts 8 overlay variables. Employing various dimension reduction schemes, the statistical sensor parameters were reduced to only the ones deemed crucial by the dimension reduction schemes. The authors also analyzed the effects of the data collection period on VM's prediction performance. By utilizing the moving window method, the authors discovered that when process drift was absent in the data collection period, a substantial training period on the VM model was preferably in order to gain prediction accuracy. Among the regression model, kNN gave the best prediction performance. The authors also developed an R2R process control system embedded with VM by utilizing EMWA in Monte Carlo simulations. The empirical results obtained through a large number of simulations carried out demonstrated that the proposed R2R control system was able to perform process recipe adjustment correctly in order to

correct the overlay metrology measurement that has drifted far from the target values defined.

In [15], the authors presented a machine learning-based faulty wafer detection method using novelty detection instead of the conventional binary classification method. In this work, the authors defined a faulty wafer as a wafer with large deviations in its metrology values. According to the authors, FDC data are conventionally used by SPC for fault detection in a fabrication process because of the immediacy of these methods. However, SPC methods for fault detection have various limitations. Firstly, SPC only inspects and controls a subset of the variables that have been known to have a high impact on wafer quality. Secondly, SPC assumes independence between variables while in actuality, the interaction between multiple variables affects the wafer's quality. Employing Principle Component Analysis (PCA) in SPC may address the problem, but at the cost of information lost. Thirdly, SPC-based methods assume linearity and unimodality in the data, but in reality, the opposite is true. Lastly, FDC data, which are direct observations of process conditions sampled by the equipment's sensor during the fabrication process, are not a direct representation of the wafer's quality; wafer quality has to be derived from the sampled process conditions. An alternative way to detect faulty wafers is through physical metrology inspection. However, it is both impractical and unrealistic to employ a metrology step after each process step and measure all wafers processed because doing so will induce great costs in finances, human resources, and production cycle time. Virtual metrology (VM), as opposed to physical metrology, allows the conjectures of metrology values through process and equipment sensor data. Regression models are used to perform numerical estimation on its predictors, which are the targeted metrology variables' values. Although VM has conjecture capability, regression models are not sensitive to deviations of the predictors. Therefore, the authors proposed to detect faulty wafers using novelty detection using machine learning. Focusing on photolithography equipment with two chucks in this work, 2583 and 2509 wafers' sensor and metrology variables were collected for chuck 1 and chuck 2, respectively. A total of 148 statistical sensor parameters were derived as inputs for the dimension reduction schemes to identify the crucial ones for the prediction task. Both cross-validation and moving windows methods were employed to measure the accuracy of the models evaluated by the authors. From the experimental assessment conducted, One-Class Support Vector Machine (1-SVM) achieved the highest detection accuracy through the moving windows method. The obtained results were significantly high enough for this work to be further researched in the future for practical use in the production environment.

In [16], the authors demonstrated the use of a VM-based control scheme to perform estimation of plasma electron density and plasma etch rate. According to the authors, the prediction of these metrology variables in the plasma etch process is non-trivial due to its process characteristics that are time-varying, process drifts, and sudden process shift owing to maintenance activities. Traditionally, SPC has been used to manage the plasma etch process. However, the SPC approach introduces metrology delays that could result in a large number of wafers being processed erroneously before the first faulty wafer's metrology result is detected by SPC for process issues. Advanced process control (APC) and VM have both availed as the solution to the control issues. However, the VM approach is preferable to APC because the APC system could not be utilized efficiently for wafer-to-wafer control due to infrequent measurement and metrology delays. In this work, the authors investigated the use of plasma independence (PIM) data to realize the proposed real-time control scheme. Two PIM sensors were involved to provide the process condition data at the sampling frequency of 13.56MHz. A maximum of 52 harmonics of this frequency can be recorded to depict the process condition. The control method was implemented using the predictive functional control (PFC) model. Employing multiple linear regression (MLR), ANN, and Gaussian process regression (GRP) as the prediction algorithms to evaluate the proposed method, the best prediction results were obtained by ANN. Although the obtained experimental results were promising, migrating the solution to the production environment requires further investigation on the limitations encountered in the experiment.

In [17], the authors studied the efficacy of a two modeling approach to develop an accurate VM model for the plasma etch process in semiconductor manufacturing. According to the authors, the plasma etch process is still one of the most challenging processes to model for accurate metrology quality prediction. A two modeling approach consisted of global modeling and local modeling were proposed by the authors in an attempt to model this process. The authors defined global modeling as the use of all available training data to model the behavior of the process, while local modeling referred to the use of a subset of the available training data to perform the modeling. The selection criteria for the subsets depend on the context information of interest. Hence, local modeling can produce VM models with higher prediction accuracy over certain operational behavior, while global modeling is more suitable when a general prediction spanning across the operational space is required. Partial least squares (PLS), artificial neural networks (ANNs), and Gaussian process regression (GPR) were the algorithms selected to compare the efficacy of the two modeling approaches to predict the etch rate of the plasma etching operation. In global modeling, the data are prepared using two approaches: chronologically and interleaving with reference to known events. In local modeling, three preparation methods are used: partitioning by wafer position during preventive maintenance (PM) cycle, clustering through PM cycle, and time windowing method. Two datasets of 12133 wafers and 18513 wafers were formed to evaluate the efficacy of the proposed modeling. A total of 103 process variables were collected through equipment built-in sensors. An additional 159 process variables were collected through PIM sensors. The best prediction result was

obtained through the use of the localized GPR model using the time-windowing method.

In [18], the authors presented a data fusion approach to develop a VM model for the Cu-CMP process when only a small data sample size is available for a product mixed production environment. According to the authors, conventional VM models are built when the number of observations is sufficient to form a stratified matrix between processing equipment and the products processed. Hence, an accurate VM model could not be developed using the conventional approach when mass production is still in its early stage. Therefore, the authors proposed to fuse the data from different equipment that perform the Cu-CMP process. There are two main parts in the proposed method: i) fusion method utilizing Markov chain Monte Carlo (MCMC) for identifying the significant parameters and ii) the derivation of a hierarchical Bayesian model using the parameters from the MCMC. The proposed method was compared with a conventional VM-APC (advance process control) model through a simulation experiment to evaluate their prediction accuracy. The best result was obtained by the former model.

In [19], the authors presented a VM approach that enables early-stage metrology outcome prediction and process control through the generated feedforward control signal. According to the authors, several process steps may influence the variability of target metrology and not just the immediate process step before the target metrology. The scenario studied by the authors consists of a sequence of four process steps, with the target metrology as the immediate metrology step after the fourth process step. The process' data that are taken into consideration for prediction was denoted as the observable portion of the VM scheme, which consisted of the first two process steps, while the unobservable portion was made up of the process step thereafter. Hence, the authors aimed to achieve early-stage detection by estimating the target metrology's result at the observable point and subsequently initiates process control on the processes in the unobservable portion to reduce the metrology variability. The advantages of the proposed methods are four folds: 1) provides the ability to comprehend the effects of early-stage processes on the target metrology; 2) provides lower bound observable processes on overall metrology variability; 3) early-stage detection of process drifts or abnormality, thereby potentially used for predictive maintenance purposes; 4) enhance process control on the unobservable process steps through its feedforward control signal. Elastic Net was employed as the prediction model of the proposed VM. According to the authors, this work is the first to apply Elastic Net in VM research. By using a real-world dataset that consisted of 870 wafers and 327 process variables collected through a period of 5 months, the experimental assessment showed that Elastic Net obtained the lowest prediction error scores in comparison with other regression models.

In [20], the authors presented a feasibility evaluation of the VM developed to predict the etch depth of two different recipes of the plasma etch process in a high product-mixed scenario. According to the authors, although various VM algorithms have been proposed as a result of active research in this area, applying VM for accurate prediction in a real complex production environment remains the key challenge. The authors conducted the VM development according to the phases in Cross-Industry Standard Process for Data-Mining (CRISP-DM) model. The variable selection phase is typically performed separately from the prediction model development phase by using different algorithms. However, in this work, these two phases are embedded by using the stochastic gradient tree boosting algorithm. Two model update approaches are evaluated in this work. First, the model is updated on monthly basis. Second, the model is updated whenever metrology data are available. Using data collected over a period of 6 months in a real fab, the data used in this work consisted of 64000 wafers' process data, with 2900 wafers having their metrology data recorded. A total of 120 potential process predictors were derived from the collected data. The experimental results showed that by jointly optimizing the data pre-processing and the parameters of the model, coupled with model updates, it is possible to achieve an accurate VM that is feasible to be used in a complex production environment.

In [21], the authors presented a VM model for the plasma-enhanced chemical vapor deposition (PECVD) fabrication process. The metrology quality of interest is the average thickness of the silicon nitride layer on the surface of the wafer. FDC data for the PECVD process is first filtered through various statistical methods to generate a final list of available predictor variables. Then, three variable sets were prepared. The first variable sets contain a full list of variables. The second variable set contains a subset of the full list selected based on expert knowledge. The last variable set is further filtered from the second variable set to only contain the variables deemed most important by the FDC experts. Using real fab's data collected over a period of 9 months, the data used in this work consisted of 450 wafers' FDC data with more than 150 FDC sensor variables available as potential predictors. The regression methods employed in this work were multiple linear regression (MLR), simple linear regression (SLR), ridge linear regression (RLR), partial least square (PLS), and support vector regression (SVR). Experimental results showed that SVR utilizing the expert knowledge variable set outperformed the other regression methods used with its capability to generalized unseen conditions better than then its counterparts.

In [22], the authors presented the use of the relevance vector machine (RVM) to develop a robust regression VM model with variation inference. According to the authors, neural network (NN) has been the most employed prediction algorithm for VM. Although widely used, NN is known for three limitations. First, NN suffers from the over-fitting problem. Second, NN is not robust in the presence of potentially outlier data. Third, NN lacks inferential information to measure the performance of the NN statistically. The authors are not aware of any previous works that have attempted to solve all three

weaknesses in a single VM solution. The proposed model was termed RVM-VI. Using actual plasma etch process data from the industry collected over a period of 5 months, 76 wafers' process and metrology data were available to conduct the experiment, with 40 equipment sensors involved for process condition monitoring. The process condition was sampled by the sensors at the rate of 1Hz, i.e. every second. 2 summary statistics were derived for each equipment sensor. Hence, a total of 80 potential features were available for selection. Utilizing stepwise selection procedure for dimension reduction and feature selection, the RVM-VI showed better prediction accuracy compared to other models in the experimental evaluation conducted.

In [23], the authors presented a VM model using locally weighted partial least square (LW-PLS). According to the authors, the PLS regression method is a widely used VM model due to its capability to handle the collinearity present in the variables. However, the prediction accuracy of PLS degrades with the characteristics of the modeled process changes. An example of an event that can alter process characteristics is the equipment maintenance activities. Hence, in this work, the authors proposed LW-PLS as an adaptive VM model to cater to process characteristics changes. Using the proposed method, a VM model was developed for the dry etching process to predict the etching conversion difference by utilizing variable importance in the projection (VIP) to select the relevant mode inputs. The proposed VM model was compared with the conventional method called sequential update model (SUM) and ANN model for prediction accuracy using real-world dry etching equipment data. The real-world data consisted of both process data from the equipment engineering system (EES) and the optical emission spectroscopy (OES) signals sampled at 100 milliseconds. For EES data, a total of 400 types of signals were stored to depict the process condition. From these process data, 9 statistical representatives were derived by the authors: maximum, minimum, the range between maximum and minimum, median, average, standard deviation, integral, differential and count. The etching process considered in this work has 16 unique process steps. Hence, a total of 57600 (i.e. 400 process signals x 16 process steps x 9 statistical representatives) features were derived for one etched wafer. The results from the experiment conducted showed that the proposed VM model achieved higher prediction accuracy than the comparison models. In addition, the proposed model was resilient against equipment maintenance activities with lower variation in its prediction performance pre and post-maintenance activities than SUM and ANN models.

In [24], the author presented a novel strategy to VM called multi-step VM. According to the author, the classical VM approach only takes into consideration only the last process step before the metrology step of interest as input variables when developing the VM model. However, since the semiconductor process is a sequential process step, it was a reasonable assumption to the authors that the metrology quality of the wafers not only depend on the process step immediately prior to the metrology step, but also on the previous process steps. Hence, a multi-step VM was proposed in this work utilizing regularized machine learning methodologies. The process steps considered in this work start from Chemical Vapor Deposition (CVD) step, followed by the photolithography step, and ends at the etching step. Using real-world data that consisted of 583 wafer samples, various combinations of these process steps were evaluated using two regression models to find the combination that gives the best prediction result. The two regression models employed were ridge regression (RR) and least absolute shrinkage and selection operator (LASSO). The proposed multi-steps approach showed improved prediction results over the classical approach. The acceptancy of prediction accuracy from the experimental results is subject to each semiconductor company. If the results are considered accurate enough, this strategy may be used in a production environment in place of the real metrology step.

In [25], the authors developed a VM model for the plasma-assisted oxide etching process. According to the authors, the prediction accuracy of VM is vital for the model to be reliably used in various monitoring systems in a fab. Due to the fact that the etch rate could not be measured in-situ to the process, and the challenging characteristics of this process, VM models utilizing to perform the etch rate estimation in real-time. However, these models that are typical driven by nonlinear statistical methods, such as the principal component regression (PCR), have low reliability as they tend to perform well with training datasets but not on the validation dataset. In this work, the authors showed that by introducing information that better represents the process states, the reliability and hence their prediction accuracy can be augmented. With reference to the plasma-assisted oxide etching process, the new parameters were selected by referring to the reaction mechanism of the plasma. These parameters were obtained by analyzing the data from EES and OES. From a total of 1670 parameters, 79 of them were selected through the sensitivity ranking test (SRT) that assesses the sensitivity of these parameters during the etching process. By using these parameters to form new principal components (PC) through PCA, an improvement in the etch rate prediction accuracy was demonstrated through the use of PI parameters in PCR-based VM.

In [26], the authors presented the use of extreme machine learning as the conjecture model in VM for the plasma etching process's etching rate (ER). According to the authors, controlling the plasma etching process is known to be challenging due to the process characteristics that are difficult to model. As such, estimating the etch rate is a non-trivial task. With the advent of VM, it is now possible to perform in-situ etch rate prediction. However, obtaining robust and reliable VM schemes is challenging in actual an production environment due to restricting factors such as low metrology samples to develop the model, process drifting across time, and effects from both periodic and ad-hoc maintenance activities on the equipment. Nonlinear modeling techniques are therefore

needed for such processes. Among the nonlinear models, support vector machines (SVM) and neural network (NN) has been widely applied in nonlinear modeling. However, they are not easy to train, interpret and scale efficiently with problem dimension. Hence, the authors proposed to use extreme learning machines (ELM) as the nonlinear model for VM which is capable of addressing the shortcomings of SVM and NN. By using real-world OES data of 2194 wafers with their measured ER to compare the performance of both linear and non-linear models in the experimental setup, the authors showed that the proposed model has higher effectiveness than the VM model to predict the ER.

In [27], the authors presented a feature selection method that incorporates randomization to improve feature search efficiency. VM development commonly involves feature selection over high-dimensional datasets to obtain input variables that best represent the characteristics of the process being modeled. Feature selections commonly used in VM modeling are heuristic-based which can be classified into sequential and stochastic methods. The former class is prone to suboptimal solutions while the latter, over-fitted solution. The computational cost of these methods is also sufficiently high when the datasets are dimensionally high. To address these issues, the authors proposed a feature selection method using the random forward search. By introducing randomization into feature selection, the proposed method performs a sequential search over subsets of data that are randomly joined. Two datasets of a plasma etching process from real fab were used to conduct the experimental evaluation. The first dataset consisted of 839 features with 3 target metrology variables for 118 wafers, while the second dataset consisted of 1224 features with 1 target metrology variable for 241 wafers. The evaluation results showed that the proposed feature selection method was computationally less intensive when dealing with high-dimensional datasets and at the same time, prevents suboptimal and over-fitted solutions. With PLS used as the learning algorithm, the proposed method not only reduced computational cost but also achieved higher prediction accuracy compared to other methods in the experiment.

In [28], the authors presented a VM model for the plasma etching process through the use of OES data. The enormous amount and the vast number of highly correlated features in spectroscopic signals data are the challenges that authors intents to tackle. The contribution of this work is two folds. First, the fused lasso model is used for feature selection from a high-dimensional dataset. According to the authors, this work was the first to evaluate the performance of the fused-lasso algorithm in VM modeling. Previous studies mainly used Principle Component Regression (PCR) and Partial Least Square (PLS) to extract important features from high-dimensional datasets by performing a transformation on the dataset. Although these methods have been proven to be robust in high-dimensional dataset reduction, these methods do not provide a clear interpretation between the transformed features and the original features in the dataset. The lack of interpretability prevents the engineers operating the semiconductor equipment to utilize the transform features for troubleshooting and performance improvement plans. Second, the feature selection procedure in this work handles both the wavelength and the time factor in the dataset. Due to the dynamicity of plasma processing, preserving time information in the reduced datasets is crucial. According to the authors, the existing studies have focused on reduced datasets already summarized statistically. Statistically summarized datasets not only lacks interpretability but also fail to preserve the time information of the plasma process. These limitations hinder information tracing between the wavelength measured and the time of the measurement. Using OES data from real fab sampled, 61 samples were collected from 3 wafer lots. Each of the samples was comprised of 2045 wavelength channels sampled 160 times. Hence, the dataset contained a total of 327200 process variables to be evaluated. The experimental results evaluated using the dataset showed that the performance of the proposed VM model surpassed the performance of other VM models compared.

In [29], the authors presented a VM using the support vector machine (SVM) to predict if a wafer will exceed defect threshold counts. According to the authors, as wafers proceed in a series of loops over various major fabrication process steps involving diverse process recipes, defects in terms of cracks and particles can potentially be deposited onto the surface of the wafer, resulting in improper functional behavior and thus, lead to yield loss. Defect threshold in terms of numbering counts are defined to decide if a wafer can proceed for further fabrication, a rework is required or the wafer should be scrapped completely. To determine the defect counts, regular physical metrology is required. While ensuring wafer quality, physical metrology could not be performed on all wafers in production as such actions will be too expensive in terms of equipment costs, manpower, and production time. Conventional practice adopts various sampling strategies at metrology steps by measuring a subset of the wafers to determine to overall quality level. However, such an approach potentially missed out on problematic wafers. Hence, the authors proposed the development of a VM that estimates if a wafer is good enough to proceed with further fabrication. Support vector machine (SVM) was employed as the prediction model to address the challenges in the formulated classification task. Sensor readings from high sampling rate sensors of the equipment involved were utilized to first derive a set of informative features, followed features reduction to obtain the critical ones, and lastly to train the SVM using the available data. The experiment was conducted in a real fab over many months with two fabrication equipment over hundreds of wafers. With each fabrication equipment having more than 100 sensors, more than 1500 features were derived for each wafer. The experimental results showed that the classifier was capable to achieve more than 90% prediction accuracy even when the training data was limited. The experiment also revealed that in order to obtain higher prediction accuracy, VM needs to be developed separately for each machine instead of developing a single VM for all

the machines. It is possible for VM to obtain better results should chamber-specific VM is developed. However, the lack of training data prevented this direction.

In [30], the authors presented an in-situ particle monitoring system using VM to measure particle contamination in the plasma etch process. According to the authors, the plasma etch process is a complex nonlinear process and sensitive to particle disturbances. Expensive metrology steps are often necessary to measure the fabrication process's etching rate. Particles are generally contributed from three sources: clean room, wafer handling, and from the equipment itself due to product-mixed run. While the first two contributions can be handled through various protocols, practices, and guidelines, mitigating particle from the last source is the most challenging. Hence, the VM model is employed to provide an in-situ particle monitoring system for the etching process. The method proposed by the authors performs early particle detection during the oxide etch process. 6 months of real-world data from SPC, APC, and the plasma process monitor system were acquired to form the necessary datasets for 130 wafer samples. With a sampling rate of 1Hz from the plasma etch sensors, 212 data points were used as inputs to the VM model for particle count prediction. Multilayer Perceptron Network (MLP) was used as the prediction algorithm, and the comparison was made between two learning algorithms: Levenberg Marquardt algorithm and resilient back-propagation algorithm. The experimental evaluation showed that the best prediction results were obtained by the former learning algorithm.

In [31], the authors addressed the reliability of a VM model over time in the semiconductor production environment. According to the author, the reliability of the VM degrades over time as data characteristics changes due to various scheduled activities (such as equipment maintenance activities) and non-scheduled events (such as faulty parts replacement in the equipment) that occur in the semiconductor equipment. Hence, the VM model must be updated to include the new data characteristics. However, frequent updates on the VM model can incur higher costs as the VM model needs to be re-trained for each update. In addition, not all updates performed will enhance the model's performance since not all disturbance at the equipment causes a change in the data characteristics. Hence, the authors proposed an intelligent VM model using the ensemble artificial neural network (ANN) with an adaptive update. In the proposed approach, ensemble ANN was used as the prediction algorithm, and the prediction accuracy variance was used to gauge the reliability of the VM. Two sets of four months data corresponded to two photolithography equipment of a real fab were collected to conduct the experimental evaluation. Each dataset contained data from 2301 lots, 133 process variables recorded from the equipment's sensors for each lot, and the measurement outputs of the 6 corresponding metrology variables. The experimental evaluation demonstrated that the required performance can be obtained by the proposed method at a lower cost compared to other models evaluated. The proposed VM model also capable of performing anomaly process events detection, hence, allowing it to perform wafer quality monitoring.

In [32], the authors introduced an adaptive VM methodology using the group method of data handling (GMDH) type polynomial neural networks (NN) to automate feature selection and NN topology specifications. The motivation of this work stems from the authors' observation that the previous VM studies lack the discussion on both feature and model selections. The development of a VM model commonly encounters high dimensional inputs of process variables collected from the semiconductor processes. The reduction of high-dimensional inputs coupled with appropriate model complexity is crucial to develop a VM model with high prediction accuracy. The novelty of this work is four-folds. First, GMDH is proposed to tackle the challenges. Second, two novel features were proposed by the authors to augment the prediction accuracy. Third, an enhancement to the Material Removal Rate (MRR) estimation of the CMP process utilizing the proposed method was presented, and lastly, the ease of adaptability of the proposed method to other semiconductor processes was discussed. The dataset from the prognostics and health management (PHM) data challenge in 2016 was used in the experimental evaluation of this work. The authors showed that better prediction accuracy and enhanced scalability were achieved by the proposed method compared to the data challenge champion method of the same year.

In [33], the authors presented a VM model utilizing multitask learning. Developing a VM model for equipment with various process chambers is challenging. Firstly, even though the chambers of the same equipment perform the same chemical process, the chambers' conditions will not be identical due to process variations over time. The occurrence of these variations is frequent enough that dedicated VM models for each chamber are justifiable. However, the lack of the number of observations per chamber necessary to develop a reliable VM model restricts this approach. In addition, this approach requires high resources in both time and manpower to perform the measurement at the required metrology step in order to obtain the relevant metrology data. Hence, a global model using grouped observations of each chamber is more feasible. However, the global model is incapable of capturing localized chamber process variances, causing the model to have inaccurate predictions. Multitask learning is therefore proposed to compensate for such sparse and diversified chamber information. The advantages of multitask learning are two folds. First, multitask learning performs information learning simultaneously, allowing the model to explore shared information among the chambers. Second, simultaneous learning increases the number of observations in each learning. This capability is beneficial especially for chambers that have relationships but the number of observations available is small. In this work, the authors proposed the use of non-linear and ensemble multitask methods, utilizing Multitask Adaboost and Multiboost model, to develop the VM model.

The authors were unaware of any previous attempt to use such an approach in semiconductor VM modeling. The experimental results using real-world FDC data demonstrated showed superior results over the single-task learning model while compared with the other multitask learning models, the proposed model can minimize wrong information sharing conditions, thus minimizing prediction inaccuracy.

In [34], the author investigated the efficacies of applying transfer learning for the neural network-based VM model. VM model is a prediction model that estimates the metrology measurement outcome as a function of the metrology variables with their corresponding process variables. Data-driven VM models use various mathematical models to derive the function that best represents the relationship from the historical data collected for two groups of variables. The need for sufficient historical data to derive an accurate function poses challenges to develop a VM model for new equipment. Instead of performing data collection afresh, the authors proposed to use transfer learning utilizing the existing VM models from similar sets of equipment to develop the VM model for the new equipment. Two transfer learning strategies were investigated which are model weights transfer and feature representation transfer. VM models built using the two strategies were then compared with VM models built using independent learning. Two real-world photolithography process datasets with the same target metrology quality were used to conduct the experimental evaluation, with the first dataset consisted of 1954 wafers and 1952 wafers for the second dataset. For each of these wafers, 133 process variables and 3 metrology variables were recorded. The models in the experiment were required to estimate the values of the 3 metrology variables. The experimental results showed that transfer learning is capable of developing a VM model with sufficient prediction accuracies when there is a shortage of data to conduct independent learning.

In [35], the authors presented a VM development model to cope with the equipment condition change for accurate wafer critical dimension (CD) prediction. According to the author, APC systems are conventionally used to control CD by adjusting process recipe parameters run-to-run. However, APC systems encounter control delays due to the delay in metrology, in addition to incurring high metrology costs. Incorporating VM into APC systems can solve the problem by using VM to predict the CD and feed the prediction into APC systems for process control. Hence, accurate CD prediction is crucial. Equipment conditions can change due to various conditions. These changes affect the accuracy of VM. Differing from the previous works, the authors approached the modeling by considering the correlation between the results obtained from the wafer's CD measurement and the condition change of etching equipment. Both global and local model approaches were used by the authors to construct the VM. In addition, the authors proposed an APC system that contains multiple VM models with the capability to select the optimum one based on the highest similarity to the current equipment condition. A simulation dataset was generated to conduct the experimental evaluation. The simulation data consisted of 2976 wafers with 2 equipment sensors' data. The experimental results showed that the proposed method obtained better prediction and error reduction compared to the conventional APC systems.

In [36], the author evaluated a joint modeling approach to augment faulty wafer detection in semiconductor manufacturing. According to the author, faults in wafer processing are inevitable due to various internal and external factors. A faulty wafer or wafer estimated to be faulty should be held for further inspections to determine if the wafer can be reworked or have to be scrapped. Predictive modeling has been applied actively in semiconductor manufacturing in recent decades to detect potential wafer faultiness and enhance production yield. Conventionally, a predictive modeling task is formulated as a classification task if the targeted metrology variables are categorical value, and formulated as a regression task if the targeted metrology variables are continuous values. Each task has its own sets of prediction algorithms, and there is no single setting that will consistently be the best for all problems. Hence, this work proposed joint modeling of both tasks to augment the performance of faulty wafer detection. First, a regression task is performed to predict the numerical values of the targeted metrology variables. Then, the predicted numerical outputs are used to classify the wafer for fault detection. Two datasets from a real-world fab were collected over a period of 7.5 months. The two datasets corresponded to the photolithography process of two different equipment in the fab, with the first dataset recorded data for 2583 wafers and 2509 for the second. For each of the wafer, 102 process variables and 4 target metrology variables were recorded. The measured values for each of the 4 target metrology variables were recorded in both numeric and binary forms to meet the requirement of this work. Averagely, only less than 1% of faulty wafers were present in the datasets. The experimental results showed that with such highly imbalanced datasets, the proposed modeling was able to achieve superior prediction performance over the comparison model. The experimental results also showed that the ANN model performed the best in both tasks compared with other models evaluated.

In [37], the authors presented a feature-based VM framework (FVM) to develop a VM model for wafers processed in batches of lots. The FVM framework is built upon the statistics pattern analysis (SPA) framework previously developed by the authors for process monitoring purposes. According to the authors, feature development in the current VM approaches correlates the process variables to the metrology measurements. In the FVM approach, batch process features are first derived and then correlates to the metrology measurements. The SPA quantifies process characteristics using various statistical measurements instead of the process variables themselves. Therefore, the FVM utilizes the statistics from SPA, coupled with other features such as process knowledge-based landmark features, profile-driven features, and geometry-based features to enrich the features available

for VM. The proposed method was first used to develop a dedicated VM model for two process characteristics of the CMP process based on simulated data. These two process characteristics are the material removal rate (MRR) and the within-wafer non-uniformity (WWNU), respectively. Then, comparisons were made between the proposed method and other VM methods by predicting the end-of-batch sheet resistance of a plasma etch system using real-world OES signal data. 18 process variables of the OES signals were recorded for 1121 wafers at the rate of 0.1 seconds. 6 features were derived for each of the 18 process variables, resulting in a total of 108 features available to construct the VM. The VM model developed using the proposed approach demonstrated performances that surpassed other VM models in both simulation study and industrial case study.

In [38], the authors proposed the use deep learning approach to develop a VM model that is robust against chamber condition variation of plasma etching equipment. According to the authors, optical emission spectroscopy (OES) data has found frequent usage in developing plasma etch equipment's VM model because it contains a large amount of process quality information. On the other hand, deep learning such as convolutional neural network (CNN) has proven its success in the field of computer vision and image processing. Applying deep learning (DL) to develop a VM model using OES data is not straightforward as OES data cannot be treated as an image as this approach will cause much significant loss to the process information. It is therefore necessary to modify the network configuration of the DL in order for DL to process OES data successfully. Two DL configuration changes were carried out by the authors in the work. The first configuration change is at the convolution calculation in order to cater to the time series data in the OES, while the second configuration is at the normalization method of the DL in order to preserve signal intensity information in the OES data. The proposed DL model is termed OESNet. To assess the performance of the proposed DL model, various DL models from the ImageNet Large-scale Visual Recognition Challenge (ILSVRC) were employed for comparison candidates. The experimental results showed that the proposed DL model achieved better performance than ILSVRC models in aspects of generalization capabilities, prediction accuracies, and inference time when processing OES data. In addition, the proposed DL model is robust against chamber sparsity and chamber condition variations.

In [39], the authors presented the use of deep autoencoders (AE) utilizing clipping fusion regularization to perform feature extraction for VM model development. According to the authors, a single wafer fabrication process can contain multiple sub-processes defined according to the product recipe. Each product recipe also has a different setup. A case in point is the etching process. With each sub-process monitored by various sensors, the signals captured reflect such heterogeneity and transient characteristics in the recorded data across sub-processes. However, current feature extraction methods do not take into these characteristics. Hence, in this work, the authors aimed at performing feature extraction over such data by into consideration these signals characteristics. This was achieved by applying clipping fusion regularization into the AE. Real-world plasma etching data with 298 observations involving 5 equipment sensors were used to evaluate the performance of the proposed model. 6 summary statistics were derived for each of the signals: process duration, maximum, minimum, the range between maximum and minimum, average, variance, skewness, and kurtosis. A total of 1740 features were derived to cater to all sub-process steps of the main process step. The proposed model was compared with conventional feature extraction methods using the dataset. The features extracted were then evaluated through various prediction models for wafer critical dimension (CD) prediction. The experimental evaluation showed that the proposed feature model successfully reduced the prediction errors of the models tested in the authors' work.

In [40], the authors presented a data-driven framework for VM modeling that emphasizes not only model prediction accuracy, but also model interpretability. According to the authors, the latter criteria have been missing from the study of VM modeling. The existing framework of a data-driven VM model typically focuses on regression model comparisons for the highest prediction accuracy using sets of features selected through automated feature selection methods. The invaluable knowledge of the subject-matter experts (SME) has not been taken into consideration as an important source of information to be integrated into the VM modeling framework. The development of a VM modeling framework that integrates the domain knowledge of the SME is, therefore, the focus of this work. In this work, the CMP process was selected as the subject domain and Gaussian Bayesian Network (GBN) was selected as the learning model of the proposed framework. The proposed framework can be divided into four phases. In the first stage, data pre-processing is carried out on FDC and metrology data to construct training and testing datasets. In the second stage, blocking rules are generated based on SME knowledge to govern the GBN's learning procedure. The third phase performs the structure learning for the GBN using the output from the first and second phase, and at the fourth phase, prediction accuracy is examined. Real-world CMP process data were used to conduct the experimental evaluation. After data pre-processing was applied to the dataset, the finalized dataset consisted of 545 wafers with 129 process variables available for each wafer. The results obtained from the experiments conducted showed that the prediction accuracy of GBN is on par with other regression methods while using only lesser features. The inclusion of the SME blocking rules did not reduce the prediction error. However, with the inclusion of the governing rules, root-cause analysis is made possible through the VM model as the cause and effect relationship is made clearer for interpretation through the GBN.

In [41], the authors presented the use of deep learning (DL) in VM for feature extraction. According to the authors, although VM has been widely studied, wide-scale

implementation of this enabling technology in the production environment has yet to be successful. This is mainly due to the limitations of the current feature extraction methods in handling semiconductor process data that are both large and complex, two-dimensional data. The case in point is the OES data of the plasma etch process. Manual feature extraction is not feasible in terms of time and scalability while automated feature extraction potentially missed out on crucial information. Hence, the authors explored the use of Convolutional Autoencoders to perform feature extraction on the OES data, of which such approach has not been explored in the existing studies. The extracted features were then fed into a regression model for plasma etch rate prediction. The proposed method was experimented with using other types of autoencoders as well to construct a variety of DeepVM-based models. These models were compared with other non-DL VM models recently developed for plasma etch rate prediction using OES data. Real-world OES and etch rate data were collected for 1554 wafers to conduct the experimental evaluation. 6 summary statistics were derived to create features for the OES signals: maximum, minimum, average, variance, skewness, and kurtosis. The results from the experiment conducted revealed that the proposed approach outperforms conventional regression VM models.

In [42], the authors presented a VM model to predict to perform wafer die inspection using ANN with a multi-task learning scheme. Wafers that completed the fabrication process are first tested at the wafer test phase to ensure the wafers meet the required electrical properties at the die level. At the final test phase, the functionality of the chips is inspected to filter out the faulty ones. It is possible for a wafer die to fail the final test even though it first passed the wafer test, resulting in yield loss. Therefore, the capability to predict the final test failure before the dies reached the final test is much desired in order to minimize various costs. In order to predict die failure at the final test, VM was first constructed at the wafer test phase to estimate the results of the non-sampled dies. Then, a joint model approach was constructed, containing both VM and the final test failure prediction. A real-world dataset was used to conduct the experiment. The dataset was collected over a period of 1 week. The sample size for each day consisted of 500000 wafer dies. 54 variables were recorded for each wafer die, with another 5 variables derived from the wafer map. The proposed model obtained the highest prediction results in comparison with various baseline models.

In [43], the authors proposed a tree-based ensemble VM model for the PVD semiconductor process. According to the authors, the existing VM models have low competency in handling data that are stochastic and nonlinear in nature. Certain models are also liable to the over-fitting issue. In addition, a large amount of data are usually required to successfully model the semiconductor process of interest, which may be costly to collect. Hence, in this work, the authors proposed a VM model that overcomes these limitations, focusing on predicting the wafer resistivity of the physical vapor deposition (PVD) process. The proposed model generally has two parts. First, preliminary prediction results are obtained from the ensemble model. The prediction results then serve as features input to another model to obtain the final prediction. Sequential model-based optimization (SMBO) was employed as the optimization algorithm for the proposed model. A dataset consisted of 22327 wafers with 70 equipment parameters was used to conduct the experiment. In the experiment, the authors demonstrated that superior performance is obtained by the VM model utilizing SMBO in comparison to the VM model utilizing random search for the same prediction algorithm. The authors then made a comparison between the proposed SMBO-based VM model and other existing VM models for prediction accuracy assessment. The comparison showed that the proposed model achieved better prediction accuracy by gaining more robustness against noises present in the data.

In [44], the authors proposed a VM model for the CMP process to predict the material removal rate (MRR). Differing from the previous studies in that attempts MRR prediction for CMP, this work introduced a dynamic prediction approach as opposed to the static approach used in the previous studies. The dynamic prediction approach is deemed more suitable as MRR changes over time due to process and machine performance variations. In the proposed method, K-Nearest Neighbor (KNN) is first used to select MRR samples from the historical datasets. These samples serve as past references to the MRR in order to model its behavior changes over time. Gaussian process regression (GPR) model is then used to join the selected samples. Lastly, the MRR's prediction and its uncertainty are obtained through the multi-task Gaussian process (MTGP). The proposed method outperformed its counterpart with lower prediction errors in the experiments conducted and achieved satisfactory results in comparison with both ensemble and deep learning models by leveraging information from past references of the MRR. In addition, the proposed model is capable of demonstrating the MRR's behavioral changes over time and provides prediction uncertainties along the prediction timeline.

In [45], the authors proposed a deep learning VM model with prediction uncertainty for the CVD fabrication process. The target metrology quality is the electrical property of the CVD fabrication process. According to the authors, although the data-driven VM approach has been widely studied, there are still limitations that warrant research attention. First, wafer fabrication usually spans multiple processing stages before a physical metrology measurement operation is performed, but the studies of the multi-stage VM approach are lacking. Second, the existing multi-stage VM studies do not include time-series dimension in their modeling. Third, the conventional VM approach performs feature extraction explicitly from prediction model development. As such, the end solution may not be properly optimized for the two operations. Fourth, conventional feature extraction is performed over statistically summarized data instead of the raw data. The former data are known to provide more descriptive information to the learning algorithm at the cost of

information loss, potentially resulting in less accurate prediction. The aforementioned challenges can be addressed by using a multi-stage convolutional neural network (CNN) VM model. CNN allows information extraction over high dimensional raw data implicitly, resulting in a low dimensional feature set, and subsequently, performs scalar value metrology variables prediction using the extracted features. However, the DL approach is liable to the over-fitting problem. The use of the Gaussian process model (GPR) to quantitatively measure prediction uncertainty can aid in solving the over-fitting problem, with the limitation that it could not be applied on high dimensional data without the risk of model instability. Hence, by utilizing the strength of CNN and GPR, the authors proposed the fusion of CNN and GPR for a multi-stage VM model, termed CNN-GPR. In addition, instead of utilizing backpropagation learning for the model, the authors trained the proposed model using posterior density distribution maximization. The VM model was experimented with using data from a real semiconductor plant to predict the wafer electrical property prediction of the CVD process. The CVD process consisted of 4 stages. A total of 27, 27, 27, and 20 equipment sensors for each of the 4 stages were involved in process data collection. Data for a total of 170 wafers were collected over a period of approximately 2 months. 4 descriptive statistics were derived for each of the process variables collected through the sensors. These 4 descriptive statistics were maximum, minimum, variance, and average. The experimental results showed lower prediction error in comparison with other regression models and at the same time, able to quantify the confidence level of the prediction.

In this section, past decades of VM researches were reviewed. Table 1 presents the comparison of these researches in terms of the fabrication process step studied, the targeted metrology quality, and the research contributions. In the next section, research analysis and the research challenges to realize the envisioned VM will be presented.

## III. RESEARCH ANALYSIS AND CHALLENGES OF OVERLAY VIRTUAL METROLOGY

This section first presents the analysis of the related VM research works presented in section II to identify the key criteria for developing a successful VM, followed by the current research challenges towards realizing an overlay VM for potential future researches.

### A. RESEARCH ANALYSIS

#### 1) VM MODELLING

In the literature, VM is identified as a data-driven model with soft sensors capable of virtually sensing the quality of an un-sampled wafer based on process and equipment data [23], [26]. The virtual sensing capability of the VM, also known as its prediction capability, is achieved through various data mining algorithms. Examples of these algorithms are SVM, PLS, and GPR, to name a few from the literature. These algorithms attempt to derive mathematical models in the historical data to map the key process representative variables to the targeted metrology [27]. According to [23], data preprocessing is crucial for realizing a high-performance VM. The data preprocessing steps are made up of 1) derivation of statistical process representative variables from process and equipment state information for a list of candidature input variables, 2) exclusion of outliers to minimize prediction error, and 3) shortlisting of high influential input variables through selection algorithms. Step 3) is commonly known as the feature selection step. Accurate feature selection is crucial for prediction accuracy enhancement and computational load reduction [23], [27]. Hence, in summary, VM modeling involves building a prediction model by using appropriately derived input and output representatives from the historical data for the data mining algorithms to model the relationship between the two data entities. A successfully derivative of this relationship will result in high prediction accuracy by the data mining algorithms.

#### 2) HIGH SAMPLING PROCESS DATA

Fabrication processes are sampled at high frequency by various physical sensors on equipment to preserve their characteristics for analysis. According to [29], process data sampled by sensors with high frequency are, in most cases, capable of depicting the fabrication process characteristics accurately. From the prior works reviewed, the number of sensors reported ranged tens to hundreds, and their sampling rates ranged from hertz (Hz) to megahertz (MHz) [13], [16], [22], [23], [29], [37]. These sampled data are retrieved from fabrication equipment and stored by advanced data acquisition systems, such as the FDC and the EES, in near real-time and made retrievable through various data communication channels provided by these systems. The immediate availability of these data enables various near real-time automated fault detections to be implemented. Leveraging these data for VM, a wafer's metrology variables can also be estimated in near real-time to gauge its metrology quality.

#### 3) DATA CHARACTERISTICS VARIATIONS

Fabrication process characteristics, and hence its data characteristics, may change over time owing to various internal and external disturbances, such as maintenance events, load change, equipment condition change, and equipment part replacements [23], [31], [35]. As process characteristics are preserved in the data sampled by the equipment sensors, changes in the process characteristics are directly reflected in the sampled data. According to [23], equipment part replacement strongly affects the process characteristics. Reference [31] further stated that data characteristics changes will not always correspond to the disturbances that occurred.

Data characteristics will gradually change over time regardless. Hence, it is important for VM modeling to take into consideration this factor in order to develop a reliable VM for production use. Various techniques were used in the

**TABLE 1. Comparison of the fabrication process step, the targeted metrology and the contribution of VM researches.**

| Research Work | Process Step | Target Metrology | Contribution |
|---|---|---|---|
| [9] | Etching | Faulty batch process detection | • A faulty batch process detection approach by adapting the traditional kNN rule to construct the detection model using only normal samples. |
| [10] | - | N/A | • A VM model capable of capturing process variations in multiple input and multiple output (MIMO) fabrication process.<br>• A wafer level run-to-run (R2R) feedback process scheme utilizing output from the proposed VM. |
| [11] | Etching | Faulty wafer | • A VM model that identifies faulty wafer in the presence of high-dimensional and small dataset. |
| [12] | Etching | Critical dimension | • An input variables selection approach utilizing NN for NN-based VM. |
| [13] | Etching | Etch bias | • Evaluated the best combinations of variable selection method, outlier removal method and prediction algorithm to achieve VM model with highest prediction accuracy for etch bias (EB). |
| [14] | Photolithography | Overlay | • A VM model to predict the variables of overlay metrology.<br>• A VM-embedded R2R control system that performs process recipes adjustment in order to correct the drifted metrology measurements. |
| [15] | Photolithography | Faulty wafer | • A novelty detection method using machine learning for faulty wafer detection, instead of the conventional binary classification approach. |
| [16] | Etching | Plasma electron density, etch rate | • A VM model utilizing PIM data for real-time electron density and etch rate control in plasma etch process. |
| [17] | Etching | Etch rate | • A two modelling approach to develop VM, termed global modelling and local modelling by selectively preparing the datasets according to the context information of interest. |
| [18] | CMP | Polishing rate | • A data fusion approach to develop VM model for Cu-CMP process in the presence of small data sample and product-mixed production environment. |
| [19] | Etch | The final width of a key component | • A VM approach that enables early stage metrology outcome prediction and a process control through the generated feedforward control signal |
| [20] | Etching | Etch depth | • A VM model developed using Cross-Industry Standard Process for Data-Mining (CRISP-DM) model.<br>• Feature selection and prediction steps are embedded as oppose to conventional approach that performs these two steps explicitly. |
| [21] | CVD | Silicon nitride layer thickness | • A study of the VM model development for plasma-enhanced CVD process.<br>• VM performance evaluation by varying the input variable sets utilizing experts' knowledge. |
| [22] | Etching | Critical dimension | • First VM that utilizes RVM<br>• A VM model that addresses that weakness of conventional NN-based VM models. |
| [23] | Etching | Etching conversion difference | • An improved partial least squares (PLS) termed locally weighted partial least squares (LW-PLS) that is more robust to process variations. |
| [24] | Photolithography | Critical Dimension | • A multi-process-steps VM model |
| [25] | Etching | Etch rate | • A prediction accuracy improvement for PCR-based VM etch rate prediction by introducing plasma information parameters. |
| [26] | Etching | Etch rate | • A VM model utilizing extreme machine learning to predict the etch rate of the plasma etching process. |
| [27] | - | N/A | • A feature selection method that incorporates randomization to improve its search efficiency over high-dimensional datasets. |
| [28] | Etching | Plasma intensities | • A feature selection model over spectroscopic signals data that are enormous in amount and highly correlated between variables. |
| [29] | - | Defect counts | • A VM utilizing VM to predict the defect counts of a wafer. |
| [30] | Etching | Particle counts | • An in-situ particle monitoring system utilizing VM to measure the particle contamination of etch process utilizing NN. |
| [31] | Photolithography | N/A | • A VM model using ensemble ANN with adaptive update model to maintain the reliability of the VM model over time. Reliance index is calculated to determine if the model requires an update is required. |
| [32] | CMP | Material removal rate | • A VM model built using group method of data handling (GMDH) type polynomial neural networks (NN) that automates feature selection and NN topology specifications. |
| [33] | Etching | Wafer thickness | • A VM model using multitask learning for etching equipment with various process chambers. |

**TABLE 1.** *(Continued.)* **Comparison of the fabrication process step, the targeted metrology and the contribution of VM researches.**

| | | | |
|---|---|---|---|
| [34] | Photolithography | N/A | • A transfer learning approach for NN-based VM model in the presence of insufficient historical data. |
| [35] | Etching | Critical dimension | • A global and local modeling approach to develop VM capable of coping with equipment condition change by considering the correlation between the metrology variables and the equipment conditions. |
| [36] | Photolithography | Faulty wafer | • A joint modelling approach using NN for faulty wafer prediction. |
| [37] | Etching | Plasma etch end-of-batch sheet resistance | • A feature based framework to develop VM model by considering wafers processed in batches of lots. |
| [38] | Etching | Chamber condition | • A deep learning approach using CNN to develop VM model that is robust against chamber condition variation in plasma etch equipment, utilizing optical emission spectroscopy (OES) data. |
| [39] | Etching | Critical dimension | • An AE feature extraction method that takes into consideration the signals from various sensors of a sub-processes and signals heterogeneity across sub-processes of etching recipe from diverse products. |
| [40] | CMP | Wafer thickness | • A data-driven framework that incorporates subject matter expert (SME) into the modelling process using Gaussian Bayesian Network (GBN). |
| [41] | Etching | Etch rate | • A feature extraction method using deep learning. |
| [42] | Final test | Wafer die failure | • A die failure prediction model utilizing VM in the joint model. |
| [43] | PVD | Electrical parameters of wafer resistivity | • A tree-ensemble VM model to augment the competency of the model in handling stochastic and nonlinear data and the reliability of the model in the presence of small datasets. |
| [44] | CMP | Material removal rate | • A VM model with dynamic prediction as opposed to the static prediction in the previous works, and provides prediction uncertainty along the prediction timeline. |
| [45] | CVD | Wafer electrical property | • A VM model with prediction uncertainty that takes into account:<br>o multiple processing steps that precedes the metrology step<br>o the time series dimensionality<br>o an implicit approach to perform both feature extraction and prediction model construction<br>o Feature extraction over raw data instead of summarized statistically features used conventionally |

literature to cater to this variation. Reference [23] presented the use of the Just-In-Time (JIT) model to handle both characteristics variations and nonlinearity of the fabrication process. Reference [31] presented the use of the ensemble artificial neural network as the model for both reliable estimation and adaptive model update. References [17] and [35] adopted global and local modeling approaches to cope with disturbances and correlation changes. References [43] and [44] employed GRP to provide the reliability index for the predicted metrology outcome.

#### 4) HIGH-DIMENSIONAL FEATURE SET

Existing VM works highly utilized FDC data to perform metrology quality prediction. The FDC data, which typically contained sensor readings in their raw forms [15], are derived into various statistical process representatives for meaningful depictions of the process characteristics. The wafer fabrication process requires the repetitive performance of major fabrication processes as listed in Section I. Metrology qualities of a wafer are potentially affected not only by the current process step but also the steps before it [27]. In addition, a single fabrication process step may consist of various sub-process steps. As such, the number of statistical representatives derived from the raw data to create a list of potential features as inputs to VM can range from hundreds to thousands [23], resulting in high-dimensional characteristic input data.

#### 5) PROCESS REPRESENTATIVES SELECTION

The availability of a large number of raw sensor readings enables a large number of process representatives to be derived as inputs to VM. To obtain high accuracy prediction in VM, it is necessary to only select the most relevant process representatives from this large number of potential process representatives [27]. This selection is performed through feature selection or feature extraction methods. Reference [27] distinguished between feature selections from feature extraction where the latter is defined as the extraction of new variables from the combinations of the original variables. Sequential forward selection is an exemplar of the feature selection method while PCA is an exemplar of the feature extraction method. Various algorithms were also proposed in the literature to perform this feature filtering process. For example, reference [27] proposed a feature selection algorithm that incorporates randomization for search efficiency, reference [28] applied fused LASSO algorithm to address this issue, while references [39], [41], and [45] employed deep learning models to deal with this issue.

With these research analyses defined, research challenges that need to be addressed in order to realize the desired VM in a production environment of semiconductor manufacturing can be identified.

### *B. RESEARCH CHALLENGES*

#### 1) A SHORTAGE OF RESEARCH IN OVERLAY VM

As the demand for device miniature continues to increase, the photolithography process remains the most critical wafer fabrication process step in order to shrink the feature sizes and reduce circuits' linewidth [46]. As ICs are fabricated on a wafer through a multilayer wiring process achieved through the major fabrication process steps, each patterned layer must overlay each other within the permitted range defined in their design specification to ensure proper functionality and thus, the yield of the products. Misalignment results in bad dies that eventually fail the final test and thus, yield loss [47]. Overlay error is therefore defined as the displacement between the present exposure layer relative to the preceding exposure layer [46], [47]. From Table 1, it can be seen that the majority of the VM research works focused on the metrology qualities of the etching process. Research works related to VM for overlay, on the other hand, were not actively researched. It can be seen that only works from [14], [31], [34], and [36] focused on photolithography with only [14] explicitly stated their targeted metrology was the overlay while the [31], [34], and [36] did not. In addition, the datasets used were real-world data proprietary to the specific semiconductor manufacturer. Implementing the same modeling approach in the production environment of another fab may yield different results owing to various varying production practices and product mixtures. Hence, additional research with reference to the prior works is necessary to realize a VM model in a different production environment. With the lack of research in overlay VM, it is also necessary to derive knowledge from VM research conducted for other process steps as well to realize an envisioned VM.

#### 2) A LACK OF NON-FDC VM MODELLING APPROACH

With high-frequency sensing technology, process and equipment data during a fabrication process can be sampled at very short time intervals [29]. These data sampled at high-frequency rates are, in most cases, sufficient to accurately depict the dynamicity of the fabrication process of the equipment for each wafer [29]. Utilizing an FDC system, these raw sensor readings can be retrieved from fabrication equipment and stored in the FDC system in near real-time. Besides sensor data retrieval and storage, the FDC system also performs process and equipment fault detections. The complexities of fault detections in fabrication processes have rendered active research in FDC systems for accurate fault detections, as showed in the works by [3]–[7], and [8]. From the literature, FDC data have been leveraged as the primary data source to derive various process representatives for VM modeling. Hence, in the event FDC data are unavailable owing to FDC system development and implementation works, a VM that is modeled independently of FDC data is required to sustain the production environment until the FDC system resumes production status again.

#### 3) VM MODELLING USING LOW PROCESS CHARACTERISTICS DEPICTION DATA

From observations in a real production environment, real-time sensor data retrieval without an FDC system will incur high computational cost on the equipment, leading to degraded equipment performance that affects the wafer fabrication duration. In order to preserve the computational power of the equipment for fabrication operations, only a single summarized reading for each sensor of a fabricated wafer is obtainable. That is, only an averaged reading per sensor is available for the entire fabrication duration of a wafer, instead of short time interval readings per sensor. Comparing the characteristics of these two types of data, the formal resembles data sampled at an extremely low-frequency rate. Therefore, its process characteristics depiction capability at the wafer level is also extremely low. Realizing a VM for a production environment using such data characteristics has not been attempted by prior works. Hence, it is a research challenge of this work to realize an overlay VM utilizing data with low process characteristics depiction. Table 2 summarizes the research challenges in the realization of the envisioned VM in real-world production settings.

In this section, research analysis for successful realization of VM and research challenges of this work were presented. With the research challenges identified, the next section presents the research perspective towards realizing the envisioned overlay VM.

## IV. RESEARCH PERSPECTIVE

With the research analysis and challenges defined in the previous section, this section presents the research perspective of this work that envisioned the overlay VM capable of addressing the aforementioned research challenges.

### *A. A LOT-LEVEL VM MODELLING*

Prior works have largely focused on leveraging FDC data to derive process representatives in VM modeling. As FDC data are raw sensor readings sampled at high frequency during the fabrication process of a wafer [15], VM modeling using FDC data is wafer-level modeling. To cater to the events that FDC data are rendered inaccessible, a different modeling paradigm is required. Without FDC data, deriving process representatives at the wafer level is not viable owing to the single averaged sensor readings that render low process characteristics depiction quality. As more process characteristics depictions are necessary to increase the accuracy of the metrology prediction, this work proposed to derive process characteristics at lot-level. As a wafer lot is a batch of wafers with the same fabrication steps sequence stored in a wafer cassette, wafers in the same lot will be processed at the same fabrication equipment. Hence, utilizing the averaged sensor readings from all 25 wafers in a single wafer lot can provide

**TABLE 2.** Research challenges in realization of the envisioned overlay VM in the real-world production settings.

| Research Challenge | Description |
| --- | --- |
| A Shortage of Research in Overlay VM | VM for overlay has not been widely studied. As each real fab has different production environment, implementing a VM in real production environment requires additional research effort with reference to existing work. Deriving VM knowledge from prior works on other process steps' metrology is necessary towards realizing the envisioned overlay VM. |
| A Lack of Non-FDC VM Modelling Approach | FDC data has been leveraged as the primary data source for VM modelling. In the event that FDC data are unavailable owing to development activities, a VM that is modelled independent of FDC data are required to sustain the product environment until FDC system resumes production status again. |
| VM Modelling using Low Process Characteristics Depiction Data | Retrieving high frequency sampled sensor readings is inefficient without FDC system. Only averaged readings per sensor can be obtained for each fabricated wafer. These data has extremely low depiction of the process characteristics. Realizing a VM for production environment using these data is a research challenge. |

higher visibility into the overall quality of the fabrication process of the equipment. Let $x$ denotes a single process characteristic derived and each of the ovals denotes a wafer, Figure 1 illustrates the difference between deriving a process characteristic between the wafer-level and lot-level. The proposed lot-level modeling paradigm resembles to approach taken by [9] and [37] that considered the process characteristics of a batch of lots for process fault detection, but with a distinctive difference in the time dimension. In [9] and [37], all wafer lots in a single batch are fabricated at the same time, as opposed to fabricating each wafer in a lot separately. The photolithography fabrication characteristic is represented by the latter. With the difference in time-dimension of each wafer in a wafer lot for a photolithography process, this work applies the process capability index, $C_{pk}$, to determine if a photolithography process can be qualified for overlay error estimation using the lot-level VM proposed by this work. $C_{pk}$ has been conventionally used by the industry to gauge the performance of a process with reference to its specification limits, with $C_{pk} \geq 1.33$ adopted as the standard measurement to indicate that a process is well capable of performing within its specification limits [48]. Let $USL$ denotes the upper specification limit, $LSL$ denotes the lower specification limit, $C_{pku}$ and $C_{pkl}$ denotes the process capability approximating the upper and lower specification limit, respectively. Let $\mu$ denotes the mean and $\sigma$ denotes the standard deviation measured from a sample, $C_{pku}$, $C_{pkl}$ and $C_{pk}$ of a process are given by equations (1), (2} and (3), respectively.

$$C_{pku} = \min\left\{\frac{USL-\mu}{3\sigma}\right\} \tag{1}$$

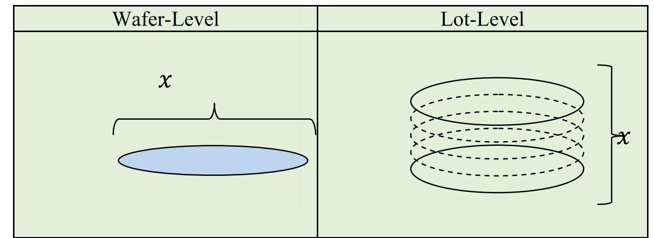


**FIGURE 1.** Difference between deriving process characteristics at wafer-level and lot-level.

**TABLE 3.** Comparison of the joint modelling steps sequence in [36] and the two-steps modelling sequence of this work.

| Research Work | First Task | Second Task |
| --- | --- | --- |
| [29] | Regression | Classification |
| This Work | Classification | Regression |

$$C_{pkl} = \min\left\{\frac{\mu-LSL}{3\sigma}\right\} \tag{2}$$

$$C_{pk} = \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right\} \tag{3}$$

$C_{pk}$ is selected as the smallest index among the upper and lower process capability to exercise precaution from over judging the capability of a process. Hence, the higher the value of $C_{pk}$, the smaller the fabrication process characteristic variation between subsequent wafer in a lot. Exploiting this characteristic, lot-level features can be derived to provide higher visibility into the stability of the photolithography process.

With the research perspective defined to envision an overlay VM capable of addressing the aforementioned research challenges, an overlay VM model can be proposed to realize the envisioned overlay VM. Hence, the next section proceeds to present the overlay VM model proposed by this work.

## V. THE PROPOSED OVERLAY VM MODEL

Drawing from the research insights from the invaluable prior VM works, this section presents the proposed VM model of this work to realize the envisioned overlay VM presented in the previous section. The proposed overlay VM model in this section also sets the future research endeavors of this work. The proposed model will be presented in four parts. Each part illustrates a characteristic of the proposed model. The first part pertains to the descriptive statistics for the process representatives. The second part pertains to prediction modeling. The third part pertains to the prediction algorithms. The last part pertains to the workflow of the proposed model.

### A. DESCRIPTIVE STATISTICS FOR PROCESS REPRESENTATIVES

The derivation of statistical process representatives is crucial to provide meaningful depictions of the process characteristics to the prediction algorithms. Differing from prior works that derived statistical process representatives at the wafer level, this work derives the statistical process representatives

**TABLE 4.** Statistical sensory features that will be derived at lot-level modelling of this work.

| Statistical Sensory Feature | Applied in Photolithography VM (Section II) |
|---|---|
| Mean | Yes |
| Maximum | Yes |
| Minimum | Yes |
| Variance | Yes |
| Standard Deviation | No |
| Skewness | No |
| Kurtosis | No |

at the lot level. Referring to the work by [14] that modeled an overlay VM, 4 descriptive statistics were derived at the wafer level: minimum, maximum, mean, and variance. In view that this work derives statistical process representatives from a single wafer lot that consists of multiple wafers, additional descriptive statistics will be derived to augment the depiction of the process characteristic. These statistical features were also found in VM research works that derive statistical sensory features from FDC data, such as [23]. [39] and [41]. Further referencing [49], Table 4 presents statistical sensory features that will be derived in this work. The use of these statistical sensory features in the prior photolithography VM works is also noted.

### B. A TWO-STEPS PREDICTION MODEL

The prediction model of the proposed overlay VM is a two-steps prediction model with each step represents a prediction task. The first prediction task is a classification task and the second is a regression. The proposal of this prediction model was based on the knowledge derived from the work by [15] and [36]. According to the findings of [15], modeling faulty wafer detection as a classification task would yield higher detection accuracy than a regression task. In [36], the author employed joint-modeling of both regression and classification tasks to detect faulty wafers. The regression task was performed first to predict the wafer's metrology variables' values. The predicted values were then used as input to the classification task for faulty wafer detection. Deriving knowledge from these two works, this work presents a two-steps prediction modeling that first performs the classification task, followed by the regression task – a reverse of the prediction task sequence compared to the sequence in [36]. Table 3 illustrates the difference between the task sequence taken by [36] and the proposed sequence of this work.

### C. PREDICTION ALGORITHMS

From the review of the related works in Section II, the following prediction algorithms are identified for each of the estimation tasks in the proposed VM modeling. For each of the tasks, two prediction algorithms are selected with the first algorithm as the proposed algorithm while the second as the comparison algorithm. The advantages of each of the selected prediction algorithms are briefly described as they are listed.

#### 1) CLASSIFICATION TASK

##### a: k-NEAREST NEIGHBOR (kNN)

The k-Nearest Neighbor (kNN) algorithm is selected as the proposed prediction algorithm for this work due to its simplicity, flexibility and its capability to be modified to perform novelty detection, as demonstrated by work in [9]. KNN is also inherently capable of nonlinear classification.

##### b: ONE-CLASS SUPPORT VECTOR MACHINE (1-SVM)

In [15], the authors examined the accuracy of various novelty detection algorithms in detecting faulty wafers for the photolithography fabrication process and concluded one-class SVM (1-SVM) as the best algorithm. As the work in [15] has the closest resemblance to the classification task of this work, 1-SVM is selected as the comparison algorithm to evaluate the performance of KNN in novelty detection of wafers with faulty overlay.

#### 2) REGRESSION TASK

##### a: ELASTIC NET

Elastic Net is an enhanced model deriving from the best features of both ridge regression and least absolute shrinkage and selection operator (LASSO) model. In ridge regression, the final model contains all the predictors while in LASSO, only a subset of the predictor enters the final model, hence achieving a sparse solution and simplifies interpretation compare to ridge regression. However, when there exist highly correlated predictors in a group setting, LASSO randomly selects only one predictor from each group, thereby discarding significant predictors in its modeling and hence lower prediction accuracy. With Elastic Net, this weakness can be overcome, resulting in higher prediction accuracy without much cost to the sparsity of the model [19]. This capability is crucial to this work for accurate selection of the equipment parameters most influential to the overlay error for a given photolithography process that took place not only contributes to accurate prediction in VM but also prepares for the extension of this work towards causality analysis of the overlay error. Elastic Net is selected as the proposed model of this task.

##### b: k-NEAREST NEIGHBOR (kNN)

In [14], the authors assessed the prediction accuracy of various linear and nonlinear regression algorithms for photolithography process VM. The evaluation results showed that kNN was the best regression algorithm. As the work in [14] has the closest resemblance to the regression task of this work, kNN is selected as the comparison algorithm to assess the performance of Elastic Net.

### D. VM SCHEME

The VM scheme based on the proposed model of this work can be figuratively depicted in Figure 2. The VM scheme involves five steps. At the first step, a unit of lot completes the photolithography process. The lot then moves into the next

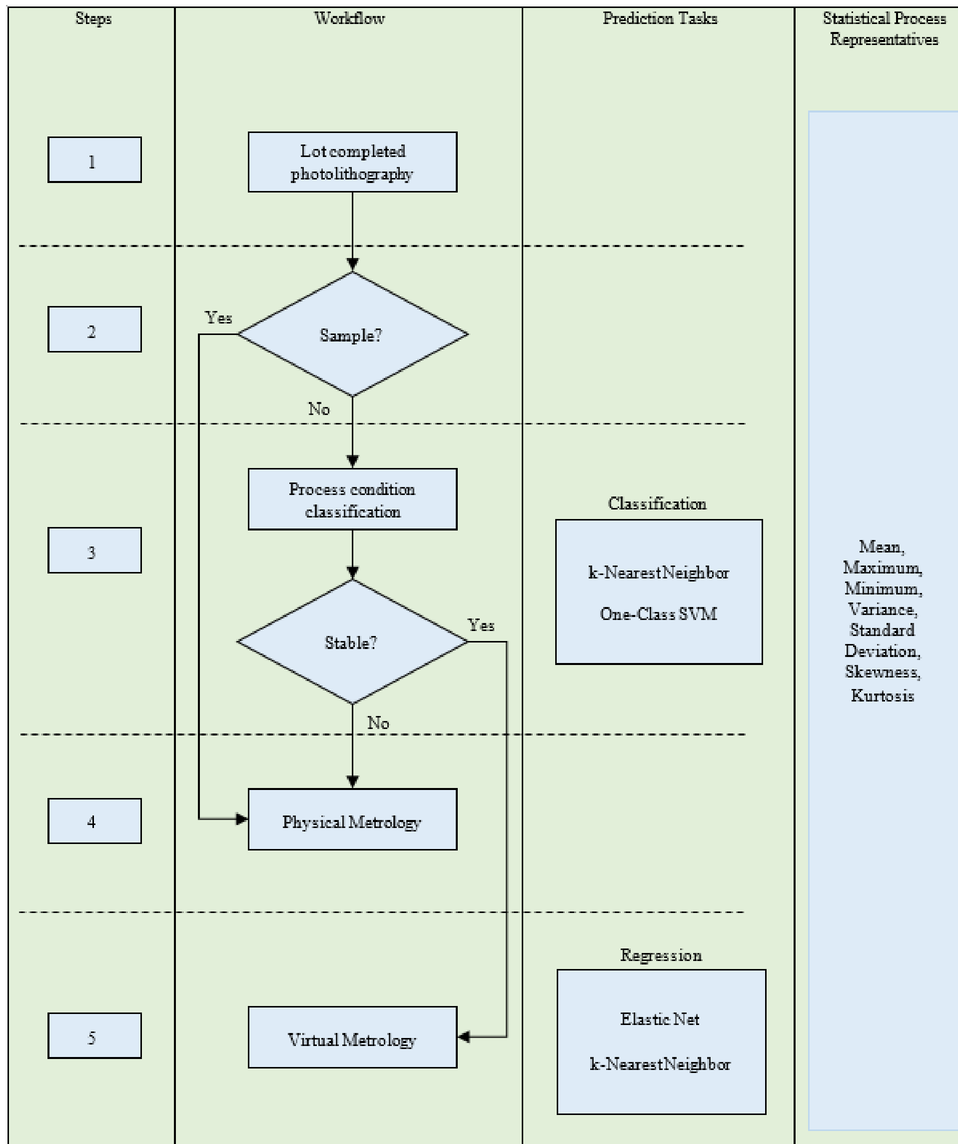


**FIGURE 2.** **Workflow of the proposed VM scheme.**

step to determine if it will be sent for physical metrology by the sampling strategy currently in place. If the lot is selected for wafer sampling at physical metrology, the lot will move to step 4. Otherwise, the lot will move the step 3 where its process condition is classified. This is the first prediction task of the proposed VM model, which is the classification task by means of the novelty detection method. If the process condition of the lot is classified as unstable, the lot will be sent for physical metrology where physical inspection will be performed. This is denoted as the third step in the VM scheme. If the process condition of the lot is classified as stable, virtual metrology will be performed on the wafers of the lot. This is the second prediction task of the proposed VM model, which is the regression task. Otherwise, the lot will be sent for physical metrology. The proposed prediction algorithms for each of the two tasks are listed in the third column of Figure 2, while the descriptive statistics to derive process representatives for both the tasks are listed in the fourth column.

In this section, the proposed overlay VM model of this work was presented. With the given real-world production settings that differ from those in the literature, a two-steps modeling approach was derived by drawing insights and design principles from the prior VM works. With the proposed overlay VM model in place, the next section concludes this study in preparation for future research endeavors to realize the envisioned overlay VM.

## VI. FINAL COMMENTS

Prior works have mainly leveraged FDC data in VM modelling. Sampled at high frequency, FDC data are capable

**TABLE 5. Comparison of the nominated data mining models for the proposed VM modelling.**

| Tasks | Data mining models | Advantages | Disadvantages |
|---|---|---|---|
| Classification | k-Nearest Neighbor | • Simple and flexibility<br>• Capable of handling nonlinearity in data | • For novelty detection, it requires explicit determination of the number of nearest data points to vote for novelty |
| | One-Class Support Vector Machine | • Novelty detection based model<br>• Capable of handling nonlinearity in data | • Higher complexity model, hence the obtained results lacks transparency and interpretability |
| Regression | Elastic Net | • Developed based on the best features of Ridge Regression and LASSO.<br>• Takes into account grouping tendency to select the correlated features | • Higher risk of over-fitting<br>• Feature selection requires sufficient number of observations in the data |
| | k-Nearest Neighbor | • Simple and flexibility<br>• Capable of handling nonlinearity in data. | • Requires explicit determination of the number of nearest data points to achieve higher accuracy for the numerical estimation |

to provide high depiction of the process characteristics. However, in the event that FDC data are unavailable owing to various FDC system development activities, only averaged reading per equipment sensor can be collected from a real photolithography equipment to depict the entire photolithography process of a wafer. These averaged data have extremely low process characteristics depiction capability. With the need to sustain the overlay metrology of a production environment until FDC data are available again, this work proposed an overlay VM modelling approach that utilizes these averaged data from the real photolithography equipment. A lot-level modelling paradigm is proposed in this work, which differs from most of the prior works that utilized wafer-level modelling. The lot-level concept has similarity to the batch-level concept presented in the prior works of [9] and [37], with the difference in the treatment of time-dimension. In [9] and [37], batches of wafers step through the process step at the same time, while in photolithography process, wafers of a wafer lot are processed sequentially. Hence, $C_{pk}$ index is necessary to select only stable photolithography process for lot-level modelling.

With wafer-level modelling, the process condition of a wafer is assessed after it completes fabrication. With lot-level modelling, the process condition of a wafer is assessed after all wafers in a lot completes photolithography. Hence, the formal exhibits wafer-to-wafer level control while the latter exhibits lot-to-lot control. To realize the lot-to-lot control for overlay metrology in production environment, a two-steps prediction model is proposed to define a VM scheme for the production environment. The VM scheme, as depicted in Fig. 2, takes into consideration the existing sampling strategy and complements it with the proposed overlay VM. Each of the step in the proposed VM model represents a prediction task. The first prediction task is a classification task that estimate if a wafer is potentially faulty and should be routed for overlay metrology station for physical inspection. If a wafer is classified as faultless, the second prediction task, which is a regression task, will attempt to estimate the numerical values of its overlay errors.

With the proposed overlay VM model, the overlay quality of the fabricated wafers can continue to be monitored until FDC data is available again and leveraged for VM using the approach proposed by the prior works. As such, it is the also the aim of this work that the proposed modelling approach can further contributes towards the research of VM in catering various scenario of a production environment.

## VII. CONCLUSION

The semiconductor technology advancement as predicted by Moore's law has become the benchmark to drive the performance of the fab with shorter cycle-time to produce high-quality semiconductor end products. Such challenging demands brought about the emergence of VM and the proliferation of its research. As ICs are fabricated on a wafer through a multilayer wiring process achieved through the major fabrication process steps, each patterned layer must overlay each other within the permitted range defined in their design specification to ensure proper functionality and thus, the yield of the products. The displacement between the present exposure layer relative to the preceding exposure layer is defined as overlay error. With feature size shrinkage and linewidth reduction in IC for device miniature, photolithography continues to be the most crucial wafer fabrication process step to minimize overlay error. Motivated by a real-world fab production environment, this research aims to realize an overlay VM in the event that the availability of FDC data of a real- photolithography process is interrupted due to various development activities of the FDC system. Realization of the envisioned VM is a non-trivial task owing to practical challenges that must be addressed. Hence, in this paper, a thorough research analysis was carried out on the related research works to identify the relevant research challenges. Based on the research challenges identified, the research perspective towards the envisioned overlay VM was defined. Next, the proposed VM model to realize the envisioned overlay VM was presented. Through the proposed VM model, the required future research endeavors of this work were also defined and the execution of these works are underway.

## ACKNOWLEDGMENT

The authors would like to thank X-FAB Sarawak Sdn. Bhd. for their support in this research by providing the environment and resources to extract the relevant data.

## REFERENCES

[1] G. S. May and C. J. Spanos, *Fundamentals of Semiconductor Manufacturing and Process Control*. New York, NY, USA: Wiley, 2006.

[2] M. Quirk and J. Serda, *Semiconductor Manufacturing Technology*. 1st ed. London, U.K.: Pearson, 2001.

[3] Q. P. He and J. Wang, ''Large-scale semiconductor process fault detection using a fast pattern recognition-based method,'' *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 2, pp. 194–200, May 2010.

[4] G. Verdier and A. Ferreira, ''Adaptive mahalanobis distance and $k$-nearest neighbor rule for fault detection in semiconductor manufacturing,'' *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 59–68, Feb. 2011.

[5] J. Yu, ''Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes,'' *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 432–444, Aug. 2011.

[6] Z. Zhou, C. Wen, and C. Yang, ''Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes,'' *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 70–79, Feb. 2015.

[7] K. B. Lee, S. Cheon, and C. O. Kim, ''A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes,'' *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[8] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, ''Convolutional neural network for wafer surface defect classification and the detection of unknown defect class,'' *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 163–170, May 2019.

[9] Q. P. He and J. Wang, ''Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes,'' *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.

[10] A. A. Khan, J. R. Moyne, and D. M. Tilbury, ''Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares,'' *J. Process Control*, vol. 18, no. 10, pp. 961–974, Dec. 2008.

[11] P. Kang, H.-J. Lee, S. Cho, D. Kim, J. Park, C.-K. Park, and S. Doh, ''A virtual metrology system for semiconductor manufacturing,'' *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12554–12561, Dec. 2009.

[12] T.-H. Lin, F.-T. Cheng, W.-M. Wu, C.-A. Kao, A.-J. Ye, and F.-C. Chang, ''NN-based key-variable selection method for enhancing virtual metrology accuracy,'' *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 1, pp. 204–211, Feb. 2009.

[13] D. Zeng and C. J. Spanos, ''Virtual metrology modeling for plasma etch operations,'' *IEEE Trans. Semicond. Manuf.*, vol. 22, no. 4, pp. 419–431, Nov. 2009.

[14] P. Kang, D. Kim, H.-J. Lee, S. Doh, and S. Cho, ''Virtual metrology for run-to-run control in semiconductor manufacturing,'' *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2508–2522, Mar. 2011.

[15] D. Kim, P. Kang, S. Cho, H.-J. Lee, and S. Doh, ''Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing,'' *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4075–4083, Mar. 2012.

[16] S. A. Lynn, N. MacGearailt, and J. V. Ringwood, ''Real-time virtual metrology and control for plasma etch,'' *J. Process Control*, vol. 22, no. 4, pp. 666–676, Apr. 2012.

[17] S. A. Lynn, J. Ringwood, and N. MacGearailt, ''Global and local virtual metrology models for a plasma etch process,'' *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 1, pp. 94–103, Feb. 2012.

[18] K. Tamaki and S. Kaneko, ''Multiparametric virtual metrology model building by job-shop data fusion using a Markov chain Monte Carlo method,'' *IEEE Trans. Semicond. Manuf.*, vol. 26, no. 3, pp. 319–327, Aug. 2013.

[19] G. A. Susto, A. B. Johnston, P. G. O'Hara, and S. McLoone, ''Virtual metrology enabled early stage prediction for enhanced control of multi-stage fabrication processes,'' in *Proc. IEEE Int. Conf. Automat. Sci. Eng. (CASE)*, Madison, WI, USA, Aug. 2013, pp. 201–206.

[20] G. Roeder, S. Winzer, M. Schellenberger, S. Jank, and L. Pfitzner, ''Feasibility evaluation of virtual metrology for the example of a trench etch process,'' *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 3, pp. 327–334, Aug. 2014.

[21] H. Purwins, B. Barak, A. Nagi, R. Engel, U. Hockele, 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 Trans. Mechatronics*, vol. 19, no. 1, pp. 1–8, Feb. 2014.

[22] S. Hwang, M. K. Jeong, and B.-J. Yum, ''Robust relevance vector machine with variational inference for improving virtual metrology accuracy,'' *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 1, pp. 83–94, Feb. 2014.

[23] T. Hirai and M. Kano, ''Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares,'' *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 2, pp. 137–144, May 2015.

[24] G. A. Susto, S. Pampuri, A. Schirru, A. Beghi, and G. De Nicolao, ''Multi-step virtual metrology for semiconductor manufacturing: A multilevel and regularization methods-based approach,'' *Comput. Oper. Res.*, vol. 53, pp. 328–337, Jan. 2015.

[25] S. Park, S. Jeong, Y. Jang, S. Ryu, H.-J. Roh, and G.-H. Kim, ''Enhancement of the virtual metrology performance for plasma-assisted oxide etching processes by using plasma information (PI) parameters,'' *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 3, pp. 241–246, Aug. 2015.

[26] L. Puggini and S. McLoone, ''Extreme learning machines for virtual metrology and etch rate prediction,'' in *Proc. 26th Irish Signals Syst. Conf. (ISSC)*, Carlow, Ireland, Jun. 2015, pp. 1–6.

[27] S. Kang, D. Kim, and S. Cho, ''Efficient feature selection-based on random forward search for virtual metrology modeling,'' *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 4, pp. 391–398, Nov. 2016.

[28] C. Park and S. B. Kim, ''Virtual metrology modeling of time-dependent spectroscopic signals by a fused lasso algorithm,'' *J. Process Control*, vol. 42, pp. 51–58, Jun. 2016.

[29] A. U. Haq and D. Djurdjanovic, ''Virtual metrology concept for predicting defect levels in semiconductor manufacturing,'' in *Proc. 49th CIRP Conf. Manuf. Syst.*, vol. 57, Jan. 2016, pp. 4–580.

[30] M. F. Abdullah, M. K. Osman, N. M. Somari, A. I. C. Ani, S. P. R. S. Appanan, and L. K. Hooi, ''*In-situ* particle monitor using virtual metrology system for measuring particle contamination during plasma etching process,'' in *Proc. 6th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Batu Ferringhi, Malaysia, Nov. 2016, pp. 507–511.

[31] S. Kang and P. Kang, ''An intelligent virtual metrology system with adaptive update for semiconductor manufacturing,'' *J. Process Control*, vol. 52, pp. 66–74, Apr. 2017.

[32] X. Jia, Y. Di, J. Feng, Q. Yang, H. Dai, and J. Lee, ''Adaptive virtual metrology for semiconductor chemical mechanical planarization process using GMDH-type polynomial neural networks,'' *J. Process Control*, vol. 62, pp. 44–54, Feb. 2018.

[33] C. Park, Y. Kim, Y. Park, and S. B. Kim, ''Multitask learning for virtual metrology in semiconductor manufacturing systems,'' *Comput. Ind. Eng.*, vol. 123, pp. 209–219, Sep. 2018.

[34] S. Kang, ''On effectiveness of transfer learning approach for neural network-based virtual metrology modeling,'' *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 149–155, Feb. 2018.

[35] S. Umeda, K. Nogi, D. Shiraishi, and A. Kagoshima, ''Advanced process control using virtual metrology to cope with etcher condition change,'' in *Proc. Int. Symp. Semiconductor Manuf. (ISSM)*, Tokyo, Japan, Dec. 2018, pp. 1–4.

[36] S. Kang, ''Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing,'' *J. Intell. Manuf.*, vol. 31, no. 2, pp. 319–326, Feb. 2020.

[37] K. Suthar, D. Shah, J. Wang, and Q. P. He, ''Next-generation virtual metrology for semiconductor manufacturing: A feature-based framework,'' *Comput. Chem. Eng.*, vol. 127, pp. 140–149, Aug. 2019.

[38] T. Tsutsui and T. Matsuzawa, ''Virtual metrology model robustness against chamber condition variation using deep learning,'' *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 428–433, Nov. 2019.

[39] J. Choi and M. K. Jeong, ''Deep autoencoder with clipping fusion regularization on multistep process signals for virtual metrology,'' *IEEE Sensors Lett.*, vol. 3, no. 1, pp. 1–4, Jan. 2019.

[40] W.-T. Yang, J. Blue, A. Roussy, J. Pinaton, and M. S. Reis, ''A structure data-driven framework for virtual metrology modeling,'' *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1297–1306, Jul. 2020.

[41] M. Maggipinto, A. Beghi, S. McLoone, and G. A. Susto, ''DeepVM: A deep learning-based approach with automatic feature extraction for 2D input data virtual metrology,'' *J. Process Control*, vol. 84, pp. 24–34, Dec. 2019.

[42] S. Kang, D. An, and J. Rim, ''Incorporating virtual metrology into failure prediction,'' *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 553–558, Nov. 2019.

[43] C.-H. Chen, W.-D. Zhao, T. Pang, and Y.-Z. Lin, "Virtual metrology of semiconductor PVD process based on combination of tree-based ensemble model," *ISA Trans.*, vol. 103, pp. 192–202, Aug. 2020.

[44] H. Cai, J. Feng, Q. Yang, W. Li, X. Li, and J. Lee, "A virtual metrology method with prediction uncertainty based on Gaussian process for chemical mechanical planarization," *Comput. Ind.*, vol. 119, Aug. 2020, Art. no. 103228.

[45] X. Wu, J. Chen, L. Xie, L. L. T. Chan, and C.-I. Chen, "Development of convolutional neural network based Gaussian process regression to construct a novel probabilistic virtual metrology in multi-stage semiconductor processes," *Control Eng. Pract.*, vol. 96, Mar. 2020, Mar. 104262.

[46] C. Chien and Y. Chen, "Manufacturing intelligence and smart production for industry 3.5 and empirical study of decision-based virtual metrology for controlling overlay errors," in *Proc. Int. Symp. VLSI Design, Automat. Test (VLSI-DAT)*, Hsinchu, Taiwan, Apr. 2016, pp. 1–4.

[47] C. Chien, Y. Chen, C. Hsu, and H. Wang, "Overlay error compensation using advanced process control with dynamically adjusted proportional-integral R2R controller," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 473–484, Apr. 2014.

[48] G. Sharma and P. S. Rao, "Process capability improvement of an engine connecting rod machining process," *J. Ind. Eng. Int.*, vol. 9, no. 1, pp. 1–9, 2013.

[49] A. Bleakie and D. Djurdjanovic, "Feature extraction, condition monitoring, and fault modeling in semiconductor manufacturing systems," *Comput. Ind.*, vol. 64, no. 3, pp. 203–213, 2013.

**TZE CHIANG TIN** received the M.Sc. degree in computer science from the University of Malaysia, Sarawak, in 2018. In 2012, he joined X-FAB Sarawak Sdn. Bhd. He is currently an Engineer with the Photolithography Manufacturing Department.

**SAW CHIN TAN** (Senior Member, IEEE) received the M.Sc. degree in information technology from Coventry University, U.K., and the Ph.D. degree in information technology from Multimedia University, Malaysia, in 2008. Since 2002, she has been engaged as a Lecturer with the Faculty of Computing and Informatics. In 2008, she became a Senior Lecturer with the Faculty of Computing and Informatics, Multimedia University. Her research interests include software defined networking, optical communication, and ant colony optimization.

**HING YONG** received the B.E. degree in electronic and computer from the University of Malaysia, Sarawak, in 2013. In 2013, he joined X-FAB Sarawak Sdn. Bhd. He is currently a Process Engineer with the Photolithography Department.

**JIMMY OOK HYUN KIM** received the B.E. degree in mechanical engineering from Kyung Hee University, South Korea, in 2000. In 2006, he joined X-FAB Sarawak Sdn. Bhd. He is currently the Manager with the Photolithography Process Group, Photolithography Department.

**ERIC KEN YONG TEO** received the B.S. degree in physics from Universiti Teknologi Malaysia, in 1999. In 2010, he joined X-FAB Sarawak Sdn. Bhd. He is currently the Manager of the Photolithography Department.

**CHING KWANG LEE** (Senior Member, IEEE) received the B.S. degree from the School of Information Science and the M.Sc. and Ph.D. degrees from the University of Kent, Canterbury, U.K., in 1982 and 1987, respectively. From 1988 to 1990, he was a Research Fellow in microwave antennas with a major in frequency-selective surfaces (FSS), University of Kent. From October 1990 to July 1991, he was a Research Scientist with the Electro-Optic Group, Division of Radio Physics, Commonwealth Scientific Industrial Research Organization (CSIRO), Australia. Since 1991, he has been the Chartered Engineer. From July 1991 to July 2010, he was a Faculty Member with the School of Electronic Engineering, Nanyang Technological University, Singapore. Since November 2010, he has been with the Faculty of Engineering, Multimedia University.

**PETER THAN** joined X-FAB Sarawak Sdn. Bhd., in 2000. He is currently an Engineer with the Computer Integrated Manufacturing (CIM) Group of IT Department.

**ANGELA PEI SAN TAN** received the B.S. degree in information technology from Universiti Malaysia Sabah, in 2004. In 2004, she joined X-Fab Sarawak Sdn. Bhd. She is currently an Engineer with the Computer Integrated Manufacturing (CIM) Group of IT Department.

**SIEW CHEE PHANG** received the B.E. degree in chemical engineering from Universiti Teknologi Malaysia, in 2002. In 2002, he joint X-FAB Sarawak Sdn. Bhd. He is currently the director of manufacturing and IT departments.