7th Annual IEEE Conference on Automation Science and Engineering (CASE), 2011

# Regression Methods for Prediction of PECVD Silicon Nitride Layer Thickness

Hendrik Purwins, Ahmed Nagi, Bernd Barak, Uwe Höckele, Andreas Kyek, Benjamin Lenz, Günter Pfeifer, Kurt Weinzierl

*Abstract*— **Different approaches for the prediction of average Silicon Nitride cap layer thickness for the Plasma Enhanced Chemical Vapor Deposition (PECVD) dual-layer metal passivation stack process are compared, based on metrology and production equipment Fault Detection and Classification (FDC) data. Various sets of FDC parameters are processed by different prediction algorithms. In particular, the use of high-dimensional multivariate input data in comparison to small parameter sets is assessed. As prediction methods, Simple Linear Regression, Multiple Linear Regression, Partial Least Square Regression, and Ridge Linear Regression utilizing the Partial Least Square Estimate algorithm are compared. Regression parameter optimization and model selection is performed and evaluated via cross validation and grid search, using the Root Mean Squared Error. Process expert knowledge used for a priori selection of FDC parameters further enhances the performance. Our results indicate that Virtual Metrology can benefit from the usage of regression methods exploiting collinearity combined with comprehensive process expert knowledge.**

## I. Introduction

In a fab, a plant that manufactures semiconductor devices, starting with an uniformly doped bare silicon wafer, the fabrication of integrated circuits needs hundreds of sequential process steps which can be assorted into 7 main process areas: *lithography*, *etching*, *deposition*, *chemical mechanical planarization*, *oxidation*, *ion implantation* and *diffusion* [4]. In order to increase the efficiency of these processes, an advanced fab is required to have online quality monitoring tools enabling the preservation of sufficiently high process capability of the production equipment. In current practice, process quality is regularly monitored by the sampling of production wafers. This approach assumes that the process quality of production equipment does not change abruptly and that the measurement result of the sampled wafers is a good representative of the actual production quality [7]. This practice may not be able to timely detect equipment performance drift happening between the scheduled measurements. This might cause quality degradation of the production wafers, an increase of the duration of production cycles, and therefore increased production cost.

To overcome this problem, a Virtual Metrology (VM) approach has been developed in order to assess the quality of every production wafer by using related process and production equipment data without physically conducting quality measurements. The prediction ability of VM allows the quality of a wafer to be estimated right after being processed, and the objective of real-time quality monitoring wafer-to-wafer is possibly achieved [5]. The VM approach has attracted the attention of researchers, and given its applicability to the various steps of wafer processing, the work that has been done pertaining to the *deposition* step will be reviewed. More specifically, the proposed data modeling, relating process and production equipment data (input variables) to the measure of production process quality in terms of deposited layer thickness (output variable) will be considered.

In the deposition steps of the manufacturing process, a multitude of layers of different materials are deposited onto the production wafers. In the chemical vapor deposition (CVD), a chemical reaction of a gas mixture at the surface of the wafer is taking place at high temperatures. In order to avoid the need of high temperature, in a Plasma Enhanced Chemical Vapor Deposition (PECVD), the chemical reaction is enhanced by means of electrical fields at radio frequency. An important aspect of this technique is the well defined and reproducible composition and thickness of the deposited film, achievable with reasonable effort by control of the signicant process parameters [4].

In [1], the authors make use of the Monte Carlo simulation to enhance the Design of Experiment (DoE) data sets, and model the relation between the input variables and the output variable using a back propagation neural network. In [10], [7], the authors

H. Purwins is with PMC Technologies GmbH, 48151 Münster, Germany and with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain `hendrik.purwins@upf.edu`

A. Nagi is with PMC Technologies GmbH, 48151 Münster, Germany, `ahmed.nagi@gmail.com`

B. Barak, Uwe Höckele, A. Kyek, B. Lenz, G. Pfeifer, K. Weinzierl are with Advanced Process Control, Infineon Technologies AG, 93049 Regensburg, Germany, `bernd.barak,uwe.hoeckele,andreas.kyek,` `benjamin.lenz,guenter.pfeifer,` `kurt.weinzierl1@infineon.com`

This work has been performed within the context of the European project IMPROVE (Implementing Manufacturing science solutions to increase equiPment pROductiVity and fab pErformance) contributing to the ENIAC Joint Undertaking, Sub-Programme SP8, Equipment & Materials for Nanoelectronics. The funding by the EU under Grant Agreement No. 120005 and the German BMBF (Federal Ministry of Education and Research) under the Contract No. 13N10417 is gratefully acknowledged. 

propose to use a radial basis function neural network to model the dependence of the output variable on the input variable, however, in [7], the models are derived using process parameters data from real production equipment, instead of DoE data sets. In [12], the authors propose a scheme for choosing between the back propagation neural network and multiple linear regression, based on the error terms obtained as a result of applying both models to a data set. Comparing a multiple linear regression model (using a stepwise procedure for selecting the input variables) with a multi-layer perceptron neural network and a radial basis function neural network is the focus of a study carried out in [2]. The authors of [9] use the multiple linear regression with stepwise selection in order to determine a set of input variables, which are in turn fed, as input variable, into a back propagation and a simple recurrent neural network model to predict the output variable. Finally, in [6] the data set is divided into an in-spec and an out-of-spec data set, and the Classification And Regression Tree (CART) is used to predict when a production wafer will be inside or outside of the defined specification limits.

The present paper is structured as follows: In Section II, we explain the PECVD process for the formation of a dual-layer metal passivation stack. In Section III, we describe the preparation and preselection of the available FDC data set. Section IV gives an introduction of Simple Linear Regression (SLR), Multi Linear Regression (MLR), Partial Least Square Regression (PLR), and Ridge Linear Regression (RLR) utilizing the Partial Least Square Estimate (PLSE) algorithm, followed by a description of the performed regression parameter optimization as well as model selection via cross validation and grid search. The results are presented in Section V. Finally, the drawn conclusions are summarized in Section VI.

## II. Manufacturing Process

### A. Chemical Vapor Deposition

CVD is used in the semiconductor industry for the deposition of thin films of solids onto substrates (i.e.wafers) by chemical reaction of a certain mixture of process gases within a process chamber. To initiate the deposition process as well as to increase the deposition rate, the reaction gases can be activated thermally, electrically (by plasma), chemically or by photons. The properties of the deposited film are determined by the method of activation, the applied energy, the amount and chemical properties of the supplied gases, the duration of the deposition process, as well as the temperature and the material properties of the substrate.

In *Plasma Enhanced CVD* (PECVD), the reaction gas mixture is activated by an electrical field at high frequency, e.g. at a radio frequency RF of 13.56 MHz, capacitively coupled into the process chamber. In this RF field, applied to two opposing equally sized electrodes, gas molecules are dissociated, radicalized and energetically excited as well as positively charged by impact ionization through accelerated free electrons, initially generated by collisions of the gas molecules and, for sufficiently high field intensity, additionally by extraction from the cathode. The accelerated heavy gas ions are not fast enough to reach the cathode before the turn-over of the RF field, but enhance the ionization of the reaction gas through collisions with other gas molecules. For a certain mixture of process gases, the impact ionization rate in the reaction gas depends on the injected RF-power and the total gas pressure. [3]As a consequence of the activation of the reaction gas mixture by means of a RF-plasma, the required process temperature i.e. wafer temperature is comparably low in a range of typically 200 °C to 500 °C.

The investigated *PECVD metal passivation process* comprises the primary deposition of a Silicon Oxide base layer onto a metal layer stack and the subsequent deposition of the considered Silicon Nitride cap layer onto the Silicon Oxide base layer (Figure 1). Each of the various manufactured product technology types of wafers comprises numerous different basic types of wafer design which classifies wafers with the same physical properties. However, all basic design types of wafers are processed with the same product technology type specific recipe sequence for the Silicon Nitride-Oxide deposition process. In the present work one product technology type with its various basic design types of processed wafer is considered.


Fig. 1. Metal Passivation Layer Structure: Silicon Nitride cap layer and Silicon Oxide base layer, deposited in a PECVD process sequence for passivation of the underlying metal layer stack.

### B. Optical Layer Thickness Measurement (Metrology)

After the PECVD metal passivation process sequence, several wafers are selected and measured for every production lot of the considered product technology type, at least one wafer for each of the three process chambers of the considered PECVD production equipment. For each sampled wafer the Silicon Nitride layer thickness is individually measured basic design type specific at several measurement points evenly distributed over the wafer. As an indicator for the quality of each measurement result the Goodness of Fit (GoF) is used. The mean calculated from these measurements is the average thickness of the Silicon Nitride cap layer. Based on this measured layer thickness the deposition time for the next lot of wafers of the same design type is calculated by a Run-to-Run (R2R) controller running for each process chamber in closed loop mode on the PECVD production equipment.

## III. FDC Data Preparation and Preselection

The initial selection of the FDC parameters has been performed by the CVD and metrology experts according to the relevance for the regarded Silicon Nitride metal passivation step in the processing of the selected product technology type. Further preselection of both FDC parameters and wafer-data sets has been executed by applying the following sequence of filtering rules to the initially available FDC data set:

- Removal of FDC parameters with entirely or predominantly missing values
- Removal of FDC parameters with constant values
- Removal of context being redundant or without process relevance
- Removal of instances with missing FDC parameter values
- Removal of instances with parameter values outside of specified limits (*mean GoF*, *minimum GoF* and *data gap count*)
- Removal of parameters being used for threshold filtering of instances (*mean GoF*, *minimum GoF* and *data gap count*)
- Removal of instances with inconsistent FDC parameter values

The resulting FDC data set comprises in total 410 wafer-data sets (instances of processed and measured wafers) completely filled with consistent values for two context and 39 numerical FDC parameters.

## IV. Statistical Methods

### A. Regression

*1) Simple Linear Regression (SLR):* This method chooses the most important FDC parameter and performs regression only with this single parameter.

*2) Multiple Linear Regression (MLR):* Given that the problem of multicollinearity in the data set is a common one, there has been a number of techniques that deals with such problem.

In *Ordinary Least Square Estimate*, the coefficient parameters are given by the following expression:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{1}$$

where $\mathbf{X}$ is the design matrix (in our case containing the FDC parameter vectors as the rows) and $\mathbf{Y}$ containing the actually measured values $y_n$ (in our case the average Silicon Nitride layer thickness of corresponding row vector $\mathbf{x}_n$).

*3) Ridge Linear Regression:* When the columns of the design matrix $\mathbf{X}$ have an approximate linear dependence, the matrix $\mathbf{X}'\mathbf{X}$ becomes close to singular. *Ridge Regression* addresses this problem of multicollinearity by solving the following expression instead of the above one:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X} + r\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \tag{2}$$

where $r$ is the ridge parameter and $\mathbf{I}$ is the identity matrix. Small positive values of $r$ improve the conditioning of the problem and reduce the variance of the estimates. While biased, the reduced variance of ridge estimates often results in a smaller mean square error when compared to Least-Squares Estimates.

*4) Partial Least Squares (PLS) Estimate: PLS Regression (PLR)* is a technique used with data that contain correlated predictor variables. This technique constructs new predictor variables, known as components, as linear combinations of the original predictor variables. PLS constructs these components while considering the observed response values, leading to a parsimonious model with reliable predictive power.

The technique appears to be like a mixture of *Multiple Linear Regression* and *Principal Component Analysis*:

- Multiple linear regression finds a combination of the predictors that best fit a response.
- Principal Component Analysis finds combinations of the predictors with large variance, reducing correlations. The technique makes no use of response values.
- PLS finds combinations of the predictors that have a large covariance with the response values.

Therefore, PLS combines information about the variances of both the predictors and the responses, while also considering the correlations among them. This method can be considered an implicit way of FDC parameter selection/dimension reduction as an alternative to the expert parameter selection described in Section V-A.3.

When using the method of PLS to estimate the coefficients, usually, the number of components to be used has to be specified. The applied choice is commented in the next paragraph.

*5) Model Selection, Stepwise Regression and Grid Search:* Stepwise regression is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression. The method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models.

Generally, there are two approaches for model selection: forward and backward. The forward model approach starts with a very simple model and builds up model complexity. For the backward approach, the opposite is true: it starts with a complex model and removes predictors which do not improve the model statistically. Given the common problem of multicollinearity explained above, forward modeling is often preferred.

Since the problem of multicollinearity is treated by using *ridge linear regression combined with PLS Estimation of the coefficients (RLR)*, there is no preference in choosing either forward or backward modeling. To find the optimal model, regression models for various values of $r$ (*ridge parameters*) and various numbers of components (for using the Partial Least Square Estimation) are constructed and the best model chosen out of them. The optimal model is determined by cross validation using the mean absolute error as evaluation function.

### B. Evaluation Schemes

A grid is spanned on the regression parameters usually on the log10 scale. The grid determines the best regression parameter combination based on majority of the cases when the error was minimal.

### C. Regression Parameter Optimization

For RLR, 2-dimensional grid search on the number of components of the PLS filter and on the ridge parameter $r$ in Equation 2 is used. For $N$ FDC parameters, the test is performed on the following numbers of components: $n = 1, 2, 3, \ldots, \lfloor \frac{2 \cdot N}{3} \rfloor$. For the ridge parameter, the following values are tested:

$$r = 10^{-10}, 10^{-9}, 10^{-8}, \ldots, 10^{5}, \quad (3)$$

allowing a grid extension up to 3 units in case the best regression parameter constellation is on the grid border.

For evaluation, a 10-fold cross validation is employed, repeated 5 times with different randomization of the the order of the data. Then the average RMSE and the standard deviation is calculated. For the PLR the number of components is chosen that had performed best with RLR. The *WEKA* [11] toolkit is used as an implementation.

## V. Results

In order to estimate average Silicon Nitride layer thickness, the FDC data prepared in Section III are investigated with the statistical methods introduced in Section IV. First, an ANOVA for the context *basic design type of processed wafer* is performed in order to assess the influence of this FDC parameter on the prediction of average layer thickness. After that, some statistical cleaning is performed on the available FDC data. Finally, the performance of various regression methods (SLR, MLR, PLR, RLR) is compared.

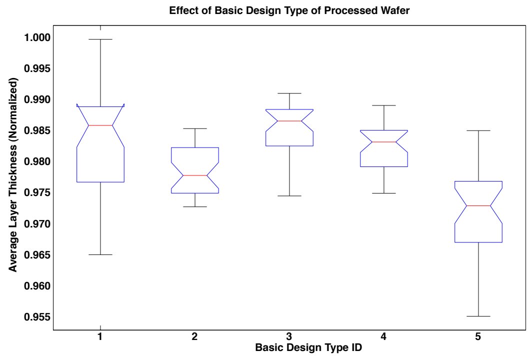



Fig. 2. Box Plot result of one-way ANOVA for the 5 most frequent *basic design types*. The tops and bottoms of each "box" are the 25th and 75th percentiles of the samples, respectively. The line in the middle of each box is the sample median. The whiskers, lines extending above and below each box, are drawn from the ends of the interquartile ranges to the furthest observations. The red points indicate outliers. Notches display the variability of the median between samples. The width of a notch is computed so that box plots whose notches do not overlap have different medians at the 5% significance level. The significance level is based on a normal distribution assumption, but comparisons of medians are reasonably robust for other distributions, when the assumption of normality is not valid. [8]

### A. Preprocessing of the available FDC Data

*1) Impact of Basic Design Type of Processed Wafer on the Mean of the Normalized Average Layer Thickness:* The purpose of *one-way analysis of variance (ANOVA)* is to find out whether data from several groups have a common mean. Therefore, this analysis is applied to the normalized values of the parameter *measured Silicon Nitride layer thickness (mean)* in the initially available FDC data set (cf. Section III), in order to assess whether the bias (intercept) term for regression modeling has to depend on the context *basic design type of processed wafer*. Figure 2 shows the results of the ANOVA.

For the 5 most frequent *basic design types of processed wafers* in the FDC data set (Figure 2), the returned p-value of the analysis is 2.8e-10, indicating that normalized average layer thickness corresponding to the different *basic design types* do not have a common mean. Hence, for *production equipment process chamber* specific regression modeling (to adapt the actual mode of the R2R controller, cf. Section II-B), the bias term has to depend on the individual basic type of the processed wafer design. [8]

*2) Filtering and Conversion of the preselected FDC Data Set:* Starting from the available preselected FDC data set as specified in Section III, the consideration of wafer-data sets for prediction modeling is consequently further restricted to the process chamber with the maximum number of instances in the FDC dataset. With regard to the ANOVA result in terms of impact of the *basic design type* on the mean of the normalized layer thickness, as discussed above, and in order to base prediction modeling on an overall statistically significant number of wafer-data sets, only *basic design* *types* with an occurrence of more than 7 instances in the FDC data set are considered i.e. 5 different *basic design types* with a frequency ranging from 8 to 30. Thus, the number of remaining instances available for statistical analysis totals to 98.

Subsequently, in a concluding scan of the data set for FDC parameters with constant values associated with this filtering of the instances two parameters are removed from the FDC data set, i.e. the number of remaining FDC parameters available for prediction modeling totals to 38 (1 context and 37 numerical FDC parameters). In order to utilize the remaining context *basic design type of processed wafer* as a numerical predictor variable, the 5 different basic design types are converted into the possible values of an accordingly 5-dimensional binary vector. Additionally, the values of all parameters in the FDC data set are normed to the range $[0, 1]$. Thus overall 42 normalized predictor variables are finally available for further statistical analysis.

*3) Selection of FDC Parameters for Prediction Modeling:* For the investigation of the different statistical methods for average layer thickness prediction modeling, three different sets of selected FDC parameters are used as predictor variables:

- *FF*: the Filtered Full FDC parameter set as described above (cf. Section V-A.2 - except the parameter *measured silicon nitride layer thickness (mean)* as the to be predicted variable) comprising 36 parameters and the context *basic design type of processed wafer*
- *ES*: the Expert Selected FDC parameter set as listed and specified in Table I consisting of 17 parameters and the context *basic design type of processed wafer* - the selection of the FDC parameters based on the FF FDC parameter set has been performed by CVD process experts.
- *TTB*: the FDC parameter set composed of the two parameters *recipe set point for wafer deposition time from R2R-controller* and *temperature of processed wafer (mean)* as well as the context *basic design type of processed wafer* - the selection of the FDC parameters based on the ES FDC parameter set has been performed by CVD process experts motivated by the intention of FDC parameter reduction to a minimum of most important predictor variables

### B. Prediction Results

The performance of the four methods, SLR, MLR, PLR, and RLR, based on the TTB, ES, and FF FDC parameter sets is compared in Table II and Figure 3. The Root Mean Squared Error (RMSE) and the standard deviation are given.

*1) Prediction based on FDC Parameters Deposition Time, Temperature and Basic Design Type of Processed Wafer:* The prediction provided by the SLR model is solely based on the most important FDC parameter *wafer deposition time*, as determined by the algorithm itself. For SLR, the RMSE is more than 2.25 times higher than for PLR as well as RLR. The RMSE for MLR is almost 4 times as high as for PLR and RLR. In this case, the performance depends heavily on the randomization of the training/test partitions. Whereas in one randomization the RMSE is as high as 42.04, for the other randomizations MLR outperforms PLR and RLR, indicating that the results of MLR without collinearity analysis and feature selection are very sensitive to different randomizations of the data set.

TABLE I

EXPERT SELECTED (ES) FDC PARAMETER SET.

| Context and Numerical FDC Parameter |
|---|
| *Basic design type of processed wafer (5 binary indicators)* |
| *Process chamber pressure (M,Std)* |
| *Process chamber pressure control valve position (M,Std)* |
| *Nitrogen ($N_2$) gas flow into process chamber (M,Std)* |
| *Monosilan ($SiH_4$) gas flow into process chamber (M,Std)* |
| *RF-power forwarded into process chamber > limit (M,Std)* |
| *RF-power reflected from process chamber > limit (M,Std)* |
| *Normalized dev. of refl. RF-power (Std,M) from batch median* |
| *Temperature of processed wafer (Std,M)* |
| *Recipe set point for wafer deposition time from R2R-ctrl.* |
| *Count of processed wafers since process chamber wet clean* |

M: Mean / Std: Standard Deviation

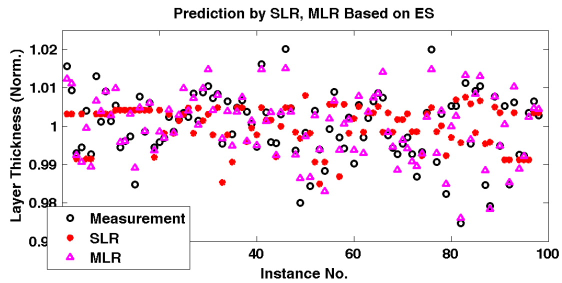
*2) Prediction Based on Expert Selected FDC Parameter Set:* SLR is not displayed since it performs equally as for the TTB FDC parameter set. PLR and RLR perform almost equally well with small Std. Compared to the TTB results, the performance of PLR is improved by 17 % (cf. Figure 3). PLR performs 7 % better than MLR.

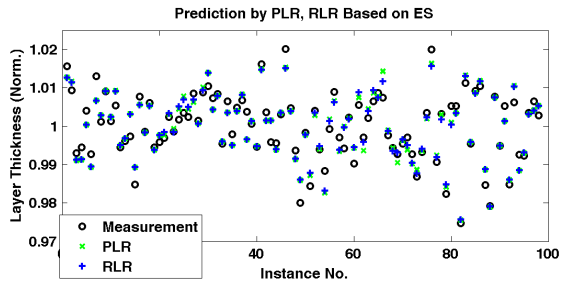
TABLE II

PREDICTION OF AVERAGE PECVD SILICON NITRIDE LAYER THICKNESS WITH SLR, RLR, MLR, PLR BASED ON THE TTB, ES AND FF FDC PARAMETER SET. AVERAGE RMSE AND STANDARD DEVIATION ARE GIVEN OVER 5 RANDOMIZATIONS AND 10-FOLD CROSS VALIDATION.

| FDC Parameters | Method | RMSE | Std |
|---|---|---|---|
| Waf. Dep. Time | SLR | 6.20 | 0.13 |
| TTB | MLR | 10.56 | 17.60 |
| | PLR | **2.71** | 0.03 |
| | RLR | 2.75 | 0.07 |
| ES | MLR | 2.41 | 0.10 |
| | PLR | **2.23** | 0.07 |
| | RLR | 2.24 | 0.08 |
| FF | PLR | **2.57** | 0.06 |
| | RLR | 2.58 | 0.09 |

*3) Prediction Based on Filtered Full FDC Parameter Set:* RLR and PLR differ by 0.1 % with small std. With respect to the results obtained for the ES FDC parameter set, the performance of RLR as well as PLR is degraded by 13 %, but still slightly better (5-6%) than the results for the TTB FDC parameter set (cf. Figure 3). MLR is not applied to the FF FDC parameter set since the



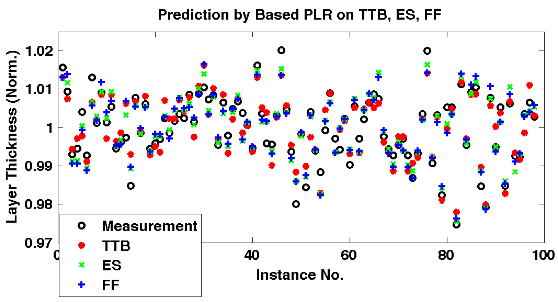




Fig. 3. Measured versus predicted normalized average PECVD Silicon Nitride layer thickness. On the best performing data set (the expert selected data set ES), different regression methods (*top* : SLR, MLR, *middle*: PLR, RLR) are compared. For the best performing algorithm (PLR), different data sets (*bottom*: TTB, ES, FF) are compared. Whereas for SLR the predicted values are nearly all within a small range around the mean, ignoring measured values far-off the mean, especially for ES, PLR and RLR predict extremely outlying measurements reasonably well. (cf. Table II)

related computational effort exceeds the actually given limitations.

## VI. Conclusion

The best prediction results are obtained for the FDC parameter set ES (with 17 parameters and 5 binary context indicators), selected by CVD process experts based on the FF parameter set (with 36 parameters and 5 binary indicators). The restriction to considerably less but important FDC parameters improves perceptibly the performance. On the other hand, further reduction of the employed FDC parameter set to the three most important FDC parameters identified by the CVD experts (TTB: *wafer deposition time*, *wafer temperature* and *basic design type of processed wafer*) again degrades the prediction performance to RMSE values slightly above the results obtained for the FF parameter set. Furthermore, the prediction based only on a single FDC parameter, i.e. the *wafer deposition time* determined as the most important parameter by SLR, results in a RMSE value more than twice as high as for the prediction using PLR or RLR based on the TTB parameter set.

Except for the FF FDC parameter set, the comparison of PLR and RLR on the one hand with MLR on the other hand indicates noticeably degraded performance results and thus confirms the robustness of PLR and RLR. In conclusion, our results indicate that VM can benefit from the usage of robust statistical methods combined with comprehensive process expert knowledge in terms of appropriate selection of the predictor variables.

## References

[1] Y.-T. Chen, H.-C. Yang, and F.-T. Cheng. Multivariate simulation assessment for virtual metrology. *IEEE ICRA*, pages 1048–1053, 2006.

[2] A. Ferreira, A. Roussy, and L. Conde. Virtual Metrology models for predicting physical measurement in semiconductor manufacturing. In *ASMC. IEEE/SEMI*, pages 149–154, 2009.

[3] A. Grill. Plasma-deposited diamondlike carbon and related materials. *IBM Journal of Research and Development*, 43(1-2):147–161, 1999.

[4] C. Hollauer. *Modelling of Thermal Oxidation and Stress Effects*. Phd dissertation, The Vienna University of Technology, 2007.

[5] H.-C. Huang, Y.-C. Su, F.-T. Cheng, and J.-M. Jian. Development of a Generic Virtual Metrology Framework. *IEEE CASE*, pages 282–287, 2007.

[6] Y.-T. Huang, F.-T. Cheng, and M.-H. Hung. Developing a product quality fault detection scheme. In *IEEE ICRA*, pages 927–932, 2009.

[7] M.-H. Hung, T.-H. Lin, F.-T. Cheng, and R.-C. Lin. A Novel Virtual Metrology Scheme for Predicting CVD Thickness in Semiconductor Manufacturing. *IEEE/ASME Transactions on Mechatronics*, 12(3):308–316, June 2007.

[8] B. Jones and I. MathWorks. *MATLAB: Statistics Toolbox User's Guide*. MathWorks, 1993.

[9] 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 Transactions on Semiconductor Manufacturing*, 22(1):204–211, Feb. 2009.

[10] T.-H. Lin, M.-H. Hung, R.-C. Lin, and F.-T. Cheng. A virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing. In *IEEE ICRA*, pages 1054–1059, 2006.

[11] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Pub, 2005.

[12] W.-M. Wu, F.-T. Cheng, D.-L. Zeng, T.-H. Lin, and J.-f. Chen. Developing a selection scheme for dual virtual-metrology outputs. *IEEE CASE*, pages 230–235, 2008.