Regression Methods for Prediction of PECVD Silicon Nitride Layer Thickness

Hendrik Purwins[1,2], Ahmed Nagi[1], Bernd Barak[3], Uwe Höckele[3], Andreas Kyek[3], Benjamin Lenz[3], Günter Pfeifer[3], Kurt Weinzierl[3]

[1]PMC Technologies, Münster, Germany
[2] Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[3]IFX Infineon, Regensburg, Germany

August 25 IEEE Conference on Automation Science and Engineering, Trieste

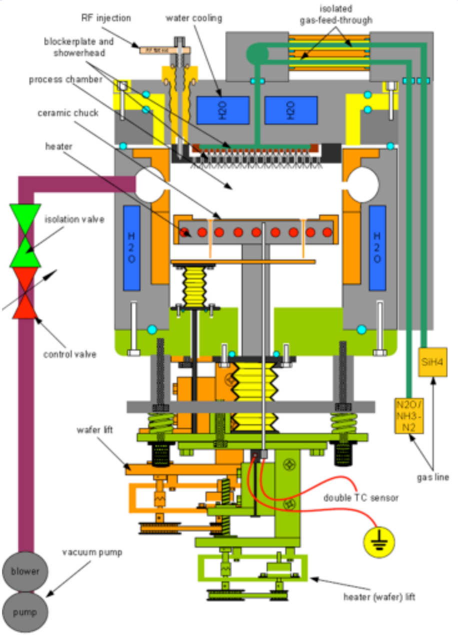
Overview

Chemical Vapor Deposition

Deposition

- Fabrication of integrated circuits
- Step in wafer processing
- Deposition of SiO_2 and Si_3N_4 on metal stack

CVD Process Chamber: Sensor and Context Predictive Variables



- Gas line input: SiH_4, NH_3 / N_2
- Reaction activated by electrical field at radio frequency 13.56 MHz
- Reflected radio frequency measured
- Pressure controlled by control valve
- Temperature: 200 °C - 500 °C
- Wafers counted since last chamber wet clean
- Deposition times set by R2R controller (prediction on physical system including self-regulation)

PMC Technologies infineon improve

Optical Layer Thickness Measurement

- Measurement via Beam Profile Reflectometry
- From the intensity of a reflected monochromatic light beam the layer thickness can be deduced
- Average error in measurement assessed by measuring the same data twice: 0.16*nm*
- Reference measurement to train/evaluate virtual measurement model

PMC infineon improve

Regression

Virtual Measurement of Si_3N_4 Layer Thickness

- Sensor and context predictive variables $\mathbf{x}$ measured in CVD process chamber or given by other equipment
- Si_3N_4 layer thickness y can be physically measured by optiprobe equipment with high costs
- A virtual measurement $\hat{y}$ is a function stochastically depending on $\mathbf{x}$ that approximates y

$$y \sim \hat{y}(\mathbf{x}) \tag{1}$$

- Saves cost of actual measurement
- Used for monitoring process, input for R2R controler

Multi Linear Regression (MLR)

- $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})' \in \mathbb{R}^d \quad (i = 1, \cdots, n)$: predictor variables
- $y_i \; (i = 1, \cdots, n)$: measured variable (averaged Si_3N_4 thickness)
- $\mathbf{w} = (w_1, \ldots, w_d)'$: coefficients
- b : intercept
- $\hat{y}_i$: virtual measurement
- n_i : noise term

$$y_i = b + w_1 x_{i1} + w_2 x_{i2} \ldots w_d x_{id} + n_i = \underbrace{b + \mathbf{w}'\mathbf{x}_i}_{\hat{y}_i} + n_i \qquad (2)$$

Ordinary Least Square Estimate

- Find best coefficients **w** and intercept b to minimize mean squared error

$$\arg\min_{b,\mathbf{w}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \arg\min_{b,\mathbf{w}} \sum_{i=1}^{n} (y_i - (b + \mathbf{w}'\mathbf{x}_i))^2, \qquad (3)$$

Coefficient and Intercept Estimation

- Centerize predictive variables and measurement

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 - \bar{\mathbf{x}} \\ \vdots \\ \mathbf{x}_n - \bar{\mathbf{x}} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \quad (4)$$

- Coefficient parameter estimation:

$$\hat{\mathbf{w}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y} \quad (5)$$

- Intercept estimation:

$$\hat{b} = \bar{y} - \hat{\mathbf{w}}'\mathbf{x} \quad (6)$$

Univariate (Simple) Linear Regression

- Choose component x_{ik} of $\mathbf{x}_i$ with lowest squared error
- Regression only with single variable x_{ik} (1-dimensional regression):

$$x_{ik} w + n_i = y_i \tag{7}$$

Ridge Linear Regression

- When predictor variables have approx. linear dependence, $\mathbf{X}'\mathbf{X}$ becomes close to singular
- Ridge Regression:

$$\hat{\mathbf{w}} = (\mathbf{X}'\mathbf{X} + r\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}, \tag{8}$$

  with ridge parameter r and identity matrix $\mathbf{I}$
- Small positive values of r improve conditioning of problem and reduce variance of estimates (comparable to regularization in kernel methods)
- Biased estimate, but often smaller mean square error than Least-Squares Estimates

Partial Least Square Regression

- PLS Regression for correlated predictor variables
- Constructs new predictor variables (components) as linear combinations of the original predictor variables, while considering the physical measurements
- Mixture of *Multiple Linear Regression* and *Principal Component Analysis*:
  - Multiple linear regression finds a combination of the predictors that best fit a measurement.
  - Principal Component Analysis finds combinations of the predictors with large variance, reducing correlations. Makes no use of measurements.
  - PLS finds combinations of the predictors that have a large covariance with measurements
- PLS combines information about variances of predictors and responses + correlations among them
- PLS can be combined with Ridge Regression (RLR)

Data

Filtering of Historical Data Set

- $>$ 100 variables of CVD production equipment
- $>$ 50 variables from measurement equipment
- Data set selected by experts
- Data filtering: removal of
  - Sensor variables with missing, (almost) constant values
  - Context variables being redundant or without process relevance
  - Instances with missing predictive variables,
  - Instances with inconsistent predictive variables
- Training set (98 instances)
- Test set (39 instances)

Results

Results

- ANOVA and data set filtering
- Training set evaluation
- Test set evaluation

Considering Most Frequent Chamber/Basic Types

- ANOVA reveals bias of *process chamber* and *basic design type* of processed wafer for Si_3N_4 layer tickness
- Build model on statistically significant number of examples
- We will **only consider the most frequent** *process chamber* for further analysis
- We will **only consider** *design types* **with at least 8 instances** for the remaining of the analysis

Methods Comparison (Training Set)

- Control variable set comparison (98 instances):
  - Control variable with most predictive power
  - 3 important predictor variables
  - Expert selected predictor variable set (17 numerical + 5 binary)
  - Full set of predictor variables (36 numerical + 5 binary)
- Method comparison:
  - Simple Linear Regression (SLR)
  - Multi-linear Regression (MLR)
  - Partial Least Squares Regression (PLR)
  - Rigid Linear Regression with Partial Least Square Estimate (RLR)
- Hyperparameter optimization with iterative grid search based on minimal validation error
- Average RMSE and Standard Deviation are given over 5 randomizations and 10-fold cross validation

Cross Validation Root Mean Squared Error (Training Set)

Variable Set	Method	CV (nm)	Std (nm)
Waf. Dep. Time	Simple LR	6.20	0.13
TTB	Multi LR	10.56	17.60
	PLS LR	**2.71**	0.03
	Ridge LR	2.75	0.07
Expert Sel.	Multi LR	2.41	0.10
	PLS LR	**2.23**	0.07
	Ridge LR	2.24	0.08
Full Filtered	PLS LR	**2.57**	0.06
	Ridge LR	2.58	0.09
GR& R	Uncond. Acc.	3.51	

- PLR performs in same range as RLR
- Passes unconditional acceptance criteria to be used as virtual metrology in production

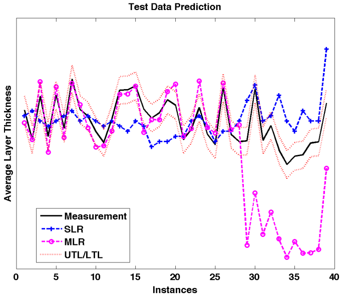
Test Set Evaluation

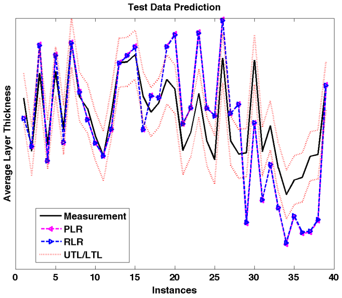
- Selection of best performing variable set (expert selected)
- Hyperparameter optimization and model training on training set; trained models applied to test set

Method	RMSE (*nm*)
Simple LR	6.93
Multi LR	13.12
Ridge LR	5.26
PLS LR	**5.19**

- PLR/RLR best and similar performance
- MLR unstable

Test Set Evaluation





- UTL/LTL: upper/lower tolerance limits for unconditional approval of method
- No. 29: chamber wetclean maintenance action, performance change

Discussion

Defining Measures for Virtual Metrology Evaluation

- Penalization of outliers
- Define error relative to target variability and noise of optical measurement
- Asymmetric evaluation measures that give different penalty for too high or too low virtual measurement

The Treatment of Time

- Try to further reduce residual n_i
- $$y_i = \hat{\mathbf{y}}(\mathbf{x}_i) + n_i = \hat{\mathbf{y}}(\mathbf{x}_i) + \hat{\hat{\mathbf{y}}}(y_{i-1}, y_{i-2}, \ldots, y_1) + n'_i, \quad (9)$$
- In order to apply auto correlation, moving average, auto regressive moving average:
- Discretize and resample points to a uniformly sampled sequence, introducing *NAN* values for non-existent sample points.
- Problem: too few and too irregularly sampled predictor variables
- Time manifests itself as degradation of
  - chamber
  - throttle valve

Conclusion

Conclusion and Future Work

- In cross validation on training set RLR performs well, in the same range as PLR, within unconditional acceptance range for inline production
- Problems: if online data out of range of training data ⇒ more training data, out-of-range test for influential features
- More investigation on feature selection (implicitly done)
- Kernel methods for regression will be explored
- Gaussian noise is assumed ⇒ other methods more suitable accounting for fat tail error distribution and giving individual confidence intervals
- Time could be modeled mainly as degradation of chamber and throttle valve
- Revisit evaluation measure

References

1. Bischop, C. (2006) Pattern Recognition and Machine Learning
2. Purwins, H., Barak, B., Hoeckele, U., Kyek, A., Pfeifer, G., Weinzierl, K.: Prediction of Average PECVD Silicon Nitride Layer Thickness (Annex D.2.4.3)
3. Purwins H., Nagi, A., Barak, B., Höckele, U., Kyek,A., Lenz, B., Pfeifer, G., Weinzierl, K.: Regression Methods for Prediction of PECVD Silicon Nitride Layer Thickness, CASE 2011
4. Purwins H., Cherla, S., Nagi, A., Barak, B., Höckele, U., Kyek,A., Lenz, B., Pfeifer, G., Weinzierl, K.I: Generalization of Average PECVD Silicon Oxide and Nitride Layer Thickness Prediction, (Annex D.2.4.3 V2)