Regression Methods for Prediction of PECVD Silicon Nitride Layer Thickness

Hendrik Purwins¹,², Ahmed Nagi¹, Bernd Barak³, Uwe Höckele³, Andreas Kyek³, Benjamin Lenz³, Günter Pfeifer³, Kurt Weinzierl³

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

August 25 IEEE Conference on Automation Science and Engineering, Trieste
Overview

1. Chemical Vapor Deposition
2. Regression
3. Data
4. Results
5. Discussion
6. Conclusion

Purwins et al.: Regression for Prediction of Si₃N₄ Thickness
Deposition of Silicon Nitride Layer

Deposition

- Fabrication of integrated circuits
- Step in wafer processing
- Deposition of $SiO_2$ and $Si_3N_4$ on metal stack
- 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)
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 \text{nm}
- Reference measurement to train/evaluate virtual measurement model
Chemical Vapor Deposition

Regression

Data

Results

Discussion

Conclusion

Purwins et al.: Regression for Prediction of Si$_3$N$_4$ Thickness
Virtual Measurement of Si₃N₄ Layer Thickness

- Sensor and context predictive variables \( x \) measured in CVD process chamber or given by other equipment
- Si₃N₄ layer thickness \( y \) can be physically measured by optiprobe equipment with high costs
- A virtual measurement \( \hat{y} \) is a function stochastically depending on \( x \) that approximates \( y \)

\[
y \sim \hat{y}(x)
\]  

- Saves cost of actual measurement
- Used for monitoring process, input for R2R controller
Multi Linear Regression (MLR)

- \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})' \in \mathbb{R}^d \) \( (i = 1, \cdots, n) \): predictor variables
- \( y_i \) \( (i = 1, \cdots, n) \): measured variable (averaged \( \text{Si}_3\text{N}_4 \) thickness)
- \( \mathbf{w} = (w_1, \ldots, w_d)' \): coefficients
- \( b \): intercept
- \( \hat{y}_i \): virtual measurement
- \( n_i \): noise term

\[
y_i = b + w_1 x_{i1} + w_2 x_{i2} \ldots w_d x_{id} + n_i = b + \mathbf{w}' \mathbf{x}_i + n_i
\]

\[
\hat{y}_i
\]
Find best coefficients $\mathbf{w}$ and intercept $b$ to minimize mean squared error

$$
\arg \min_{b, \mathbf{w}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \arg \min_{b, \mathbf{w}} \sum_{i=1}^{n} (y_i - (b + \mathbf{w}' \mathbf{x}_i))^2, \quad (3)
$$
Purwins et al.: Regression for Prediction of Si₃N₄ Thickness

Multi Linear Regression (MLR)

Coefficient and Intercept Estimation

- Centerize predictive variables and measurement

\[
X = \begin{pmatrix}
    x_1 - \bar{x} \\
    \vdots \\
    x_n - \bar{x}
\end{pmatrix},
\quad
y = \begin{pmatrix}
    y_1 - \bar{y} \\
    \vdots \\
    y_n - \bar{y}
\end{pmatrix}
\quad (4)
\]

- Coefficient parameter estimation:

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

- Intercept estimation:

\[
\hat{b} = \bar{y} - \hat{w}'x
\quad (6)
\]
Univariate (Simple) Linear Regression

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

$$x_{ik}w + n_i = y_i$$ (7)
Ridge Linear Regression

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

$$\hat{w} = (X'X + rI)^{-1}X'y,$$  \hspace{1cm} (8)

with ridge parameter $r$ and identity matrix $I$
- 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 (PLR and RLR)

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)
<table>
<thead>
<tr>
<th>Data</th>
</tr>
</thead>
</table>

*Purwins et al.: Regression for Prediction of Si$_3$N$_4$ Thickness*
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
ANOVA reveals bias of *process chamber* and *basic design type* of processed wafer for $Si_3N_4$ layer thickness.

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)

<table>
<thead>
<tr>
<th>Variable Set</th>
<th>Method</th>
<th>CV (nm)</th>
<th>Std (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waf. Dep. Time</td>
<td>Simple LR</td>
<td>6.20</td>
<td>0.13</td>
</tr>
<tr>
<td>TTB</td>
<td>Multi LR</td>
<td>10.56</td>
<td>17.60</td>
</tr>
<tr>
<td></td>
<td>PLS LR</td>
<td>2.71</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Ridge LR</td>
<td>2.75</td>
<td>0.07</td>
</tr>
<tr>
<td>Expert Sel.</td>
<td>Multi LR</td>
<td>2.41</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>PLS LR</td>
<td>2.23</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Ridge LR</td>
<td>2.24</td>
<td>0.08</td>
</tr>
<tr>
<td>Full Filtered</td>
<td>PLS LR</td>
<td>2.57</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Ridge LR</td>
<td>2.58</td>
<td>0.09</td>
</tr>
</tbody>
</table>

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

**Purwins et al.: Regression for Prediction of Si₃N₄ Thickness**
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

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple LR</td>
<td>6.93</td>
</tr>
<tr>
<td>Multi LR</td>
<td>13.12</td>
</tr>
<tr>
<td>Ridge LR</td>
<td>5.26</td>
</tr>
<tr>
<td>PLS LR</td>
<td>5.19</td>
</tr>
</tbody>
</table>

- PLR/RLR best and similar performance
- MLR unstable
- 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 \( n_i \)

\[
y_i = \hat{y}(x_i) + n_i = \hat{y}(x_i) + \hat{y}(y_{i-1}, y_{i-2}, \ldots, y_1) + n'_i,
\]

- In order to apply auto correlation, moving average, auto regressive moving average:
- Discretize and resample points to a uniformly sampled sequence, introducing \( \text{NAN} \) 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.

*Purwins et al.: Regression for Prediction of Si$_3$N$_4$ Thickness*
References


