Product’s Quality Prediction with respect to equipments data

M. Melhem¹, B. Ananou¹, M. Djeziri¹, M. Ouladsine¹, and J. Pinaton²
¹LSIS, Aix-Marseille University, Marseille, Provence-Alpes-Cote-d’Azur, France
²ST-microelectronics, Rousset, Provence-Alpes-Cote D’Azur, France

Abstract – The semiconductor manufacturing process is a complex process that consists in a big number of equipments and enormous data. This paper presents a Least Absolute Shrinkage and Selection Operator (LASSO) based method for predicting the product’s quality with respect to data of many equipments. The ability of the prediction model allows the product’s quality to be estimated in real-time instead of a sampling inspection. An application to data provided by semiconductor manufacturing is presented and the results show the ability of the proposed method to predict the product quality efficiently and effectively with an improvement of more than 90% compared to the multivariate linear regression.

Keywords: Prognosis, RUL Prediction, Semiconductor Manufacturing

1 Introduction

Semiconductor manufacturing is a complex process in which a Wafer goes through hundreds of sequential process steps with different recipes to produce a collection of chips. This process consists mainly of seven steps: Lithography, etching, deposition, chemical mechanical planarization, oxidation, ion implantation, and diffusion. This process is characterized by different types of equipments which are associated with FDC (Fault Detection and Classification) databases collected by real-time measurements. For monitoring the production quality, product quality parameters are gathered by sampling testing after manufacturing steps and are applied to evaluate the quality. Thus only the quality of monitored wafers is grasped. Therefore, a failure occurred during the non-sampling periods where other wafers are processed may severely deteriorate the final product’s quality, which results in a large number of scrapped products and thus a huge loss of Yield of fabrication. To overcome this problem, an efficient way is to predict the product quality with respect to process parameters and sensor data (Figure 1).

Since final product’s quality depends on how it was processed, value series of sensor measurements recorded at each processing step might contain quality-related patterns. Therefore, it is useful to identify quality deviations as early as possible and in real-time by data mining tools on distributed sensor measurements along the process chain [1]. As the number of sensors is enormous, the product’s quality prediction is very complicated. Furthermore, the lack of knowledge about the relevant variables that affect the quality makes the problem more difficult. Different types of methods for the prediction of the product’s quality are presented in the literature, which can be divided in three categories: expertise-based methods, model-based methods and data-based methods. Although different methods have established huge popularity in the industry, they have some limitations for the semiconductor manufacturing. Engineering knowledge is not always sufficient for building prediction models in this domain, and physical models can’t be constructed due to the complexity of the process. So, preferred methods for the wafer’s quality prediction are the data-based methods. Data-based methods can be divided in two categories: Statistical based methods and artificial intelligent methods.

Fig 1 : Description of semiconductor manufacturing process and wafer’s quality prediction

A number of previous works have been proposed for modelling the manufacturing processes and predicting the associated product quality prediction. A DPNN-based process management system is proposed [2] to predict four quality parameters associated to the ingot fabrication corresponding to control parameters. Multiple regression models and a Bootstrap algorithm were applied to generate sufficient data for ingot prediction. A polynomial neural network is applied...
in [3] to construct a predictive model of Plasma Etch process. Two Chemical Vapor Deposition (CVD) predictive models are constructed by a Radial Basis Function Neural Network [4] and a Support Vector Machine [5]. Bayesian Networks were used in [6] to generate causal relations between process variables and wafer quality. Regression methods are applied in [7] [8] to predict the CVD thickness. A quality prognostic scheme is developed in [9] to estimate the sputtering thickness as a processing quality with respect to processing parameters and sensor data in the TFT-LCD manufacturing process. For this purpose, neural networks and Weighted Moving Average algorithms are applied.

The works cited above construct a prediction model of the product’s quality at any stage of the process without taking into account the cumulative effect of previous stages. A method is developed in [10] for continual prediction of manufactured microprocessor quality with respect to sparsely sampled control measurements prior to final testing by using an average prediction of linear regression and boosted trees. But this work doesn’t consider any data characterizing the fab for prediction. A useful idea is to consider FDC data as additional powerful predictors.

In this paper, a regularized regression model (LASSO) is applied to investigate the influence of equipments FDC data on the measured quality parameters while taking into account the relationship between the quality specifications of previous stages. This method can be considered as a combination of a multiple regression model and a variable selection method and thus, it can construct a prediction model that take into account the cumulative effect of many equipments and avoid overfitting caused by the complex models.

The remaining of this paper is organized as follows: Section 2 summarizes the methods used in the literature for the product’s quality prediction with respect to many equipments, and it particularly explains the LASSO-based regression. Section 3 presents the proposed method for online prediction based on the LASSO-based regression. Section 4 provides an example with application in semi conductor manufacturing process to illustrate the feasibility of the proposed method. Finally, section 5 concludes the paper and identifies future work for improvements.

2 Model description

2.1 Literature review

In the semi-conductor manufacturing, a huge amount of high-dimensional and correlated data are collected through many equipments and requires a reduction. Multivariate statistical techniques can be used for feature extraction, like Principal Component Analysis (PCA) [11] and Partial Least Squares (PLS). PCA is used to develop a prediction model from a historical database when product quality data are not available [12]. However, it is able to analyze the correlation between variables in a particular manufacturing stage and thus, it consider the whole manufacturing as happened in a single stage. A solution is recommended in [13] to estimate the effect of each stage on output quality of the next stage by a regression model, and it is applied to mobile phone production line. However, the semiconductor process is complicated and the quality measurements are not always available. For considering the correlation between manufacturing stages, a Cascade Quality Prediction Model is developed in [12] based on the PCA and decision trees. But this requires a significant expert knowledge.

To overcome the shortcomings of the existing methods, a sequential feature extraction method based on the regularized least-squares regression algorithm so called LASSO is proposed in this paper which will improve the prediction accuracy. This algorithm is well suitable to control the large number of variables, it reduces the observable variables to fewer numbers of factors by shrinking the non pertinent variables to zero.

2.2 LASSO regression

Standard linear regression models formulated as (1) work by identifying a set of regression coefficients that minimize the Residual Squared Error between the observed values and the fitted values from the model (equation 2) to obtain the Ordinary Least Square (OLS) estimate.

\[
y = X\beta + \epsilon \\
\min_{\beta} \|y - X\beta\|^2_2
\]

Where X(N×P) is the matrix of process variables, y(N×L) is the matrix of quality measurements, \(\beta(P \times L)\) is the vector of regression coefficients, and \(\epsilon\) is the residual vector. Multiple linear regressions are a particular case where a combination of the predictors that best fit the response is identified.

Given the problem of data correlation and the fact that the number of process variables is very large in many manufacturing processes, many techniques has been developed that deals with such problems. Partial Least Squares regression is used to deals with correlated predictor variables by constructing new components as linear combination of them. This method is usually used when the columns of X are highly correlated and their number is very large. The idea is to decompose the matrices X and Y like in Principal Component Analysis:

\[
\begin{align*}
X &= TP^T + E \\
Y &= UQ^T + F \\
T &= XW^* 
\end{align*}
\]
Where T and U are the component or factor matrices, P and Q are the orthogonal loading matrices, and E and F are the error terms.

In this way, this technique appears as a mixture of Multiple Linear Regression and Principal Component Analysis, and thus, it can be considered as a way of features dimension reduction.

Many extensions of the PCA/PLS methods were used in the literature for the end-product quality prediction. The Multi-way Partial Least Squares is the most famous method with good applications. However, the MPLS takes all the process as happened in a single stage, and involves all process variables in the model no matter they are critical to the end-product-quality or not. A Least absolute Shrinkage and Selection Operator (LASSO) type regularization were developed in [14] to predict the end-product quality and it overcomes the problems of the MPLS by selecting the critical-to-quality phases.

Least absolute Shrinkage and Selection Operator (LASSO) is a regularization method originally proposed for variable selection and it is demonstrated to be the best subset selection method [15]. It introduces an additional term to the minimization problem, which is the L1-norm of the regression coefficients vector multiplied by a weight parameter between zero and one, which tends to produce sparse models, which verifies its use as a variable selection tool.

\[
\hat{\beta}_{\text{lasso}} = \arg\min_{\beta} \left\{ \| y - X\beta \|^2 + \lambda \| \beta \|_1 \right\}
\]

The tuning parameter \( \lambda \) controls the strength of the penalty. A value of zero is equivalent to a standard linear regression, and as it increases, regression model coefficients are shrunk toward zero. To find the optimal model, regression models for various values of \( \lambda \) are evaluated and the best model is chosen by a Cross Validation as having the smallest Mean Squared Error.

3 Proposed method

3.1 Data description

As explained above, a huge amount of FDC data is collected from the semiconductor manufacturing process. FDC data are usually stored in a three-way matrix \( X(I \times J \times K) \), where \( I \) is the number of monitored wafers stored in the FDC database, \( J \) is the number of process variables, and \( K \) is the number of observations of each variable for each wafer. At first, \( X \) is unfolded into a two-dimensional matrix with \( I \) rows and \( P=J \times K \) columns before applying the regression model.

Quality parameters \( Y(I \times L) \) are obtained by periodically testing a sample of products with measurement equipments after the completion of critical stages of manufacturing for monitoring the production. They contain various items such layer thickness mean or uniformity, etch rate... However, measurement steps are performed on randomly selected lots, on at least one wafer within each sampled lot. Thus, most of the data items are missing in the quality database. A drift happening between the scheduled measurements cannot be detected, and the quality of other processed wafers is unknown and need to be estimated for maintaining high yield of production.

The purpose of our study is to predict the quality parameters corresponding to FDC data of many types of equipments. As unnecessary inputs can affect the prediction results, selection of critical parameters is necessary to improve model performance. The restriction to considerably less but the most pertinent FDC parameters improves substantially the performance. This can be achieved by using a LASSO regression model.

3.2 Method description

At every production stage, a LASSO-based model is used to estimate the missing quality data for each no-measured wafer. In this way, the quality parameters data are completed.

The equipment FDC data and the estimated quality already obtained from the previous equipments can be considered as inputs in modelling. Meanwhile, one quality parameter is used as output. The measured sampled quality parameters are used to evaluate the model quality.

A Low Pass Filter is used to remove noise from the modelling signals, and then the filtered signal at a particular manufacturing step is used instead of the original signal as additional input to the prediction model which will be constructed at the next stage. An overview of the applied methodology is shown in (2).
The Mean Squared Error (MSE) is adopted here as the evolution criterion to evaluate the prediction accuracy and it is described by the following expression:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$  \hspace{1cm} (5)$$

Where n is the number of sampled measured products, $y_i$ and $\hat{y}_i$ are the measured and the estimated values, respectively. Closer is the MSE value to zero, better the prediction accuracy is. But as the length of the measured sample is small, a standard procedure for estimating the performance of the model is the n-Cross-Validation. The original sample is randomly partitioned into n equally sized subsamples and each subset is used once as a hold-out set for testing the model, and the remaining n-1 samples are used for training. The total accuracy is calculated as the average of the accuracies on all the hold-out subsets.

As already said, the quality measurements are sparsely sampled, our objective is to estimate the product's quality where it is processed in a particular equipment with respect to FDC data and the measured and estimated quality corresponding to the previous equipments. Whenever a quality measurement is available for one of the previous equipment, the model is updated and evaluated for improving outcomes.

4 Application results

4.1 Results of the proposed methods

The developed method is applied on data provided by three equipments (A, B and C) in the semiconductor manufacturing process. A LASSO model is applied for each equipment where the FDC parameters are considered as input and a quality parameter is considered as output of the model. Regression parameters optimization is performed and evaluated via Cross Validation, using the Mean Squared Error. The figure (3) displays the relationship between the tuning parameter $\lambda$ and the Cross Validated Mean Squared Error of the LASSO model for the equipment A. The dots show the MSE of the corresponding model. The vertical line segments stretching out from each dot are error bars for each estimate. The line on the right is drawn at the minimum CV error, the other is drawn at the maximum value of $\lambda$ within 1 SE of the minimum. Vertical bars depict 1 standard error.

For model construction, 1500 wafers are used to construct the model, while the remaining 500 wafers are used for testing. Figure 4 shows the estimated model signal and the filtered signal for the training and testing set. Figure 5 represent the difference between the measured and the predicted quality data for equipment A. The filtering of the model signal allows obtaining the estimated quality data as shown in figure 6. We can notice that the error after filtering data is slightly larger than the one before filtering, and therefore, we do not lose a lot of information by using the Low pass filter.

Fig. 3. Cross-Validated MSE for different values of Lambda of LASSO fit for equipment A.

![Cross-Validated MSE for different values of Lambda of LASSO fit for equipment A.](image3)

Fig. 4. Estimated signals for quality data obtained by LASSO for equipment A before and after filtering. The first 1500 wafers represent the training set, while the remaining of the signal represents the predicted quality data for the testing set.

![Estimated signals for quality data obtained by LASSO for equipment A before and after filtering.](image4)

The obtained filtered signal shown in figure 4 has been used as input with the FDC data of the equipment B to construct the signal shown in figure 7. This figure shows also the filtered signal. The figure 8 shows the sampled and the predicted quality measurements.

This procedure has been repeated for the equipment C where the results shown in figures 9 and 10 are obtained. And thus, the LASSO model can perform a variable selection and a quality prediction model for each equipment with respect to equipments data and the output of the previous equipments.
Fig. 5. Measured and estimated quality data by LASSO for equipment A before filtering.

Fig. 6. Measured and predicted quality data after filtering by LASSO for equipment A.

Fig. 7. Estimated signals for quality data obtained by LASSO for equipment B before and after filtering.

Fig. 8. Measured and predicted quality data by LASSO for equipment B.

Fig. 9. Estimated signals for quality data obtained by LASSO for equipment C before and after filtering.

Fig. 10. Measured and predicted quality data by LASSO for equipment C.
The model was evaluated for 5 cases according to the availability of the quality data. The first case consists of estimating the product’s quality for the equipment C if the first 1500 wafers are used for training and the remaining wafers are used for validating the model. A quality measure is available at the 1600th wafer for example for the equipment C, the model is updated by adding this measure in the training data. This procedure is repeated at each point where one quality measure is available. The evolution of the MSE is represented in (11) for the training data and in (12) for the testing data according to the available quality data. For the 1750th product, two patterns were added in the training data in estimating the product’s quality for equipment C that are two measured products for the equipments A and C. By taking into account these measurements in addition to the new available FDC data, the quality prediction shows an improvement of 99% for the testing data as shown in the table 3 and the figure 12. The table 1 and 2 show the MSE of the training and test data, and the figures 11 and 12 show the evolution of the MSE.

Fig. 11. MSE evolution for the training data with the LASSO model

Fig. 12. MSE evolution for the validation data with the LASSO model.

4.2 Comparison with multivariate linear regression

The developed LASSO method is compared with a multivariate linear regression where we obtain the results shown in (13) and Table 3. The LASSO model presents an improvement of more than 90% compared to the multivariate regression model, which implies the great importance of variable selection in case of multivariate and correlated data.

<table>
<thead>
<tr>
<th>Product</th>
<th>LASSO</th>
<th>Multivariate Regression model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600</td>
<td>0.0044</td>
<td>0.0582</td>
</tr>
<tr>
<td>1750</td>
<td>5.1116*10⁻⁵</td>
<td>0.0083</td>
</tr>
<tr>
<td>1800</td>
<td>2.1797*10⁻⁴</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

Fig. 13. Comparison between the MSE of the testing data of the LASSO and the multivariate regression model
5 Conclusions

In this paper, we proposed a LASSO-based method for predicting the product’s quality in a manufacturing process composed of many equipments, and an application of semiconductor manufacturing process is given. This method works as an interpolating method that is updated every time a new quality measurement is available. Results of our work have validated the effectiveness of using the LASSO regression method and show an improvement of more than 80% compared to a multivariate regression model. This method provides benefits for our application, but in the other hand, it has limitations that need to address in future. Firstly, as the prediction model is constructed by a statistical regression model, it lacks a physical significance, so the relationships between process parameters and quality data need to be evaluated by process engineer. Furthermore, the lack of measured quality data may cause an over-fitted model.

6 References


