Construct an Intelligent Yield Alert and Diagnostic Analysis System via Data Analysis: Empirical Study of a Semiconductor Foundry

Yi-Jyun Chen, Yen-Han Lee, Ming-Chuan Chu

Construct an Intelligent Yield Alert and Diagnostic Analysis system via Data analysis: Empirical Study of a Semiconductor Foundry

Yi-Jyun Chen, Yen- Han Lee and Ming-Chuan Chiu

National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, Taiwan
mcchiu@ie.nthu.edu.tw

Abstract. As semiconductor manufacturing technology advances, the process becomes longer and more complex. A critical issue is to determine how to avoid yield loss at an early stage or to diagnose the cause of yield loss soon, in order to save more money. Traditional statistical regression analysis and correlation analysis are unable to quickly and easily figure out the causes of process anomalies and potential problems. This study aims to construct an intelligent yield alert and diagnostic analysis framework combined within a big data analysis architecture. Through an intelligent detection and early warning mechanism, instant detection of yield anomalies and automatic diagnostic analysis based on good/bad wafer classification, we can effectively and rapidly find out the factors that may cause process variation to help quickly clarify the causes of abnormal product yield. The case study in this paper uses real-world data from a foundry in Taiwan. We hope to provide engineers and domain experts with a reference framework for building a yield analysis system to help improve the yield of semiconductor manufacturing and enhance the competitiveness of high-tech industries.

Keywords: Big Data Analysis, Yield Analysis, Regression Analysis, Correlation Analysis.

1 Introduction

The concepts of Smart Factory and Industry 4.0 have recently drawn increasing attention in Taiwan's semiconductor industry. The semiconductor industry in Taiwan plays an important role in the global market and has a leading position especially in the fields of wafer foundry, packaging, and testing. Due to strong global competition, implementation of smart manufacturing to increase the efficiency of production processes, increase the automation of equipment, and enhance data analysis capabilities is not just an option but is a necessary action.

The advances in semiconductor process technology required to reach the Nano-node have made the process more complex and lengthier. In the wafer manufacturing process, huge amounts of production records and process parameters data have accu-

mulated and must be analyzed. A semiconductor fab must use various types of dynamic data analysis collected from the automated production machines to early predict the equipment status of possible production problems and abnormal state by referring to the history data. If it is possible to adjust the equipment immediately to correct the wrong status, the fab can reduce huge losses caused by interruption of production due to errors in the production flow.

It is critically important to understand how to utilize the huge production records and process parameters data. This research aims to propose an intelligent yield alert and diagnostic analysis framework that can intelligently detect and alert the abnormal situations at an early stage for the semiconductor industry. The framework is based on and validated with the data collected from shop floor of a semiconductor foundry company in Taiwan.

The second section of this study is a review of the literature concerning big data, and the applications of big data in the industry. The third section proposes to construct an intelligent yield warning and diagnosis analysis model using a big data analysis architecture and apply this model to the semiconductor wafer manufacturing industry. The Section 4 verifies the validity and feasibility of the research model with a semiconductor foundry's products, which come from two stages of trial production and mass production. Section 5 concludes the contribution of this research and suggests future research directions.

2 Literature review

Big Data is a concept that has been used extensively over the past decade in the enterprise for data analytics, business intelligence and statistical applications. The field has matured due to the rapid growth in the quantity of data, the decline in data storage costs, the advancement of software technology and maturity of cloud environment Data analysis from previous insights of historical data can predict future outcomes, or even innovate, creating business models never seen before.

The characteristics of big data are known as 3Vs: Volume, Velocity and Variety (Eaton *et al.,* 2012). These are similar to the data types in semiconductor industry. The first "V" represents that the massive amount of data. The second one is about how fast the data is generated. The last one means there are various types of data. Due to the lengthy and complicated wafer process, a lot of data is generated during manufacturing. Each machine at each station in each step of process could generate data, so there is a high velocity of generating data. Data generated from different machine could be in different forms, such as images or numerical.

As a result, some papers utilize big data to solve semiconductor factory problems. The applications of big data can be roughly divided into three types: realizing existing data, enhancing the competitiveness of enterprises through data and using data as the foundation and core services to subvert traditional industries (Li, 2015). Learning how to extract useful knowledge from the data is very important. The authors designed experimental data mining to extract helpful information from data automatically collected in the process. (Chien *et al.,* 2013) The next challenge is to use the data gener-

ated in semiconductor manufacturing. The authors recommend utilizing FDC and MES data to enhance the overall usage effectiveness, and have developed a framework for that data. (Chien *et al.*, 2014) Other tools for data mining in semiconductor industry have been developed for decades. (Kusiak, 2001). Or combine two data mining approaches to extract pattern of data in semiconductor industrial to improve yield (Hsu *et al.*, 2007). Further propose a framework to solve the problems such as root cause detection in semiconductor industrial (Chien *et al.*, 2014). Gradually develop a systematic approach to solve problems.

3 Intelligent yield alert and diagnostic analysis framework in semiconductor industry

This research proposed an intelligent yield alert and diagnostic analysis system (see Fig.1). The first stage of this model is the abnormality detection and early warning mechanism. The detection is based on each piece of wafer yield test data returned daily from the production line. The second stage of the model is to diagnose the cause and alert the operator to any suspicious factors.

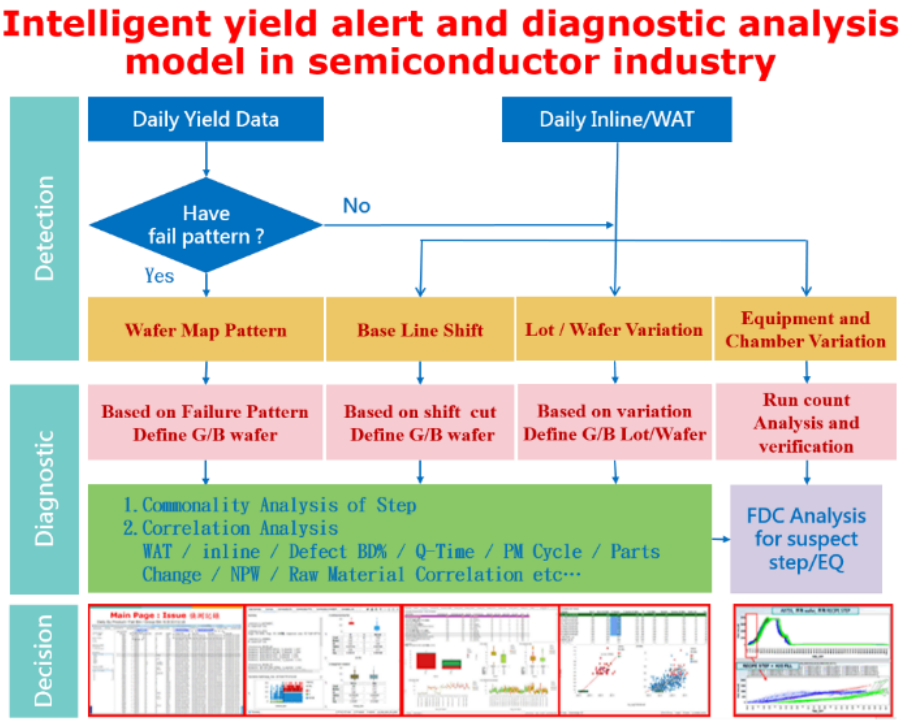


Fig. 1. Intelligent Yield Alert and Diagnostic Analysis Model

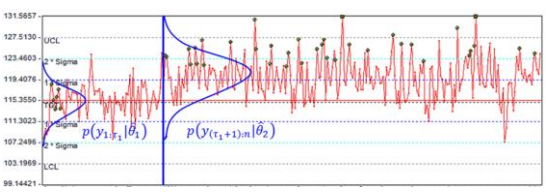
3.1 Abnormality detection and early warning mechanism

The yield of semiconductor wafers is calculated based on the ratio of defective and functioning wafers. A probe is used to test the electrical characteristics of each of the crystal grains in the wafer and their connections in the circuit. There is a special pattern in the Wafer Bin Map (WBM) to determine whether there is an abnormality in the process of wafer manufacturing. The yield data includes different fault codes (Bin #/ Bin Group), which respectively represent the meanings of different electrical characteristics of the grain. Therefore, it is possible to classify the yield abnormalities according to different fault codes. The WBM is not the only data source we can use to find different abnormal phenomena in the yields. Other methods we investigated include: base Line Shift, lot to lot variation, wafer to wafer variation, and equipment and chamber variation.

Wafer Bin Map. First, we used Novelty Detection (Pimemtal, 2014) to see if there are unknown patterns in the new wafer bin map to determine whether the classification model needs to be retrained. If there is no need to retrain the model, wafer yield data will undergo Radon Transform to do data transformation and feature extraction. We can get important features from this graph. Then we use an SVM as a classification model. We separate the two groups of wafers into Random Pattern (G0) and Non-Random Pattern (G'). Finally, we divide the Non-pattern (G') group into clusters of G1, G2, ... Gk according to their wafer pattern characteristics. We use Calinski-Harabasz Criterion (Calinski, 1974) to determine the number of clusters.

Baseline Shift. In practice, the yield test data returned by the production line every day are sorted according to the test time, and the product yield rate will vary over time. If the exact position of the yield change can be accurately identified, it can quickly clarify the time when the abnormalities occur. This study detected baseline deviations through Change Point Analysis (Rebecca, 2014) as shown in Fig 2.

Lot-to-lot & Wafer-to-wafer. In practice, we will find that products often have variations between lots to lots, or variations within batches, from wafer to wafer. Therefore, this study uses quadrant analysis as shown in Figure 3 to classify and select the variations between lots or wafers. The plot shows the Lot Mean (Fail Bin Loss Lot Mean) along the X axis and Lot Stdev (Fail Bin Loss Lot Stdev) as the Y axis. Two-dimensional quadrants are the median of Lot Mean and median of Lot Stdev as shown in Fig.2.



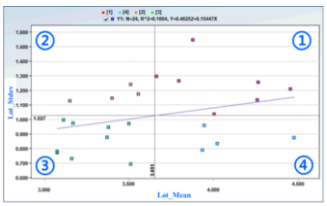


Fig. 2. Baseline Shift detection **Fig. 3.** Quadrant analysis of lot mean and lot Stdev.

Equipment and Chamber Variation. On the actual production line, there are many differences between the individual machines in the same class of machine group, and these differences may alter the yield of products. Therefore, this study uses Statistic Matching Methodology and Mean Matching to check if there are differences between the machines. The concept of this method is as shown in Fig.3, there are significant variations of machine Tool4 so the yield loss of products produced by Tool4 may come from those variations. In the second stage of the model we will do further analysis to confirm this hypothesis.

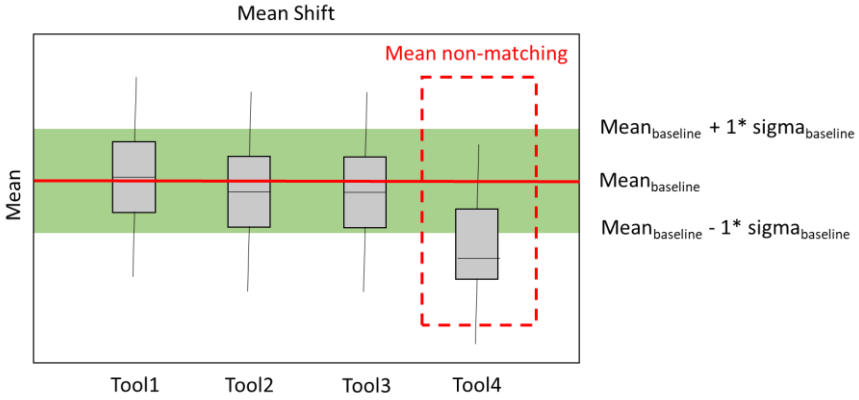


Fig. 3. Equipment and Chamber Variation

3.2 Suspicious factor decision model

Due to the complexity of the process, our first step was to identify doubtful sites that may cause abnormality in yields through commonality analysis and correlation analysis. Then, our second step as to go through the intelligent yield alert and diagnostic analysis model- suspicious factor decision model to further confirm and filter out the real cause of the abnormality. There are three steps to the suspicious factor decision model:

Long Trend Performance Analysis. Since each step of the diagnostic analysis can only explain part of the abnormal phenomenon, all wafers within the same time period must be sorted by production time to show the trends of yield changes.
Run Count Analysis. This step of the suspicious factor decision model is to find equipment and chamber variation. After listing the possible variability of the machines, we will analyze the run count of the wafers passing through the machine to further verify that the wafer yield will significantly drop when the number of wafers passing the same machine increases.
Fault Detection and Classification (FDC). The basic FDC is to inspect the specifications of equipment and to use statistical process control (SPC) to control machine parameters. The advanced FDC is to use a time-series model according to univariate

analysis or multivariate statistical analysis to view the feature information for a particular time segment in the process. Based on the data from process equipment, measurement equipment, wafer acceptance test (WAT), and yield, data mining and analysis are used to find diagnostic rules, modeling, and perform immediate diagnosis and control.

4 Case study

This study validates the framework with an empirical study conducted in a fab in Taiwan. The model is designed to quickly clarify issues in the pilot production yield, speed up the rate of increasing yield, and quickly diagnose the problem of mass production anomalies and avoid their abnormal expansion.

The empirical data covers historical production records of plot production products (product A) and mass production products (products B and C). We prepared the data by arranging it, establishing whether it is relevant and integrating it in advance, then storing it in the big data analysis platform.

We used the intelligent yield alert and diagnostic analysis framework in semiconductor industry mentioned in section 3. In the first stage, we use the Wafer Bin Map Pattern for identification and classification, the Base Line Shift testing, variation test between batches and wafers, Equipment and Chamber Variation test and other methods to construct the yield abnormality detection and early warning system. In the second stage, each abnormal type of Good and Bad Lot / Wafer are selected to carry out a diagnostic analysis of the yield abnormal factors' commonality and correlation. For the factors found by the diagnosis, we used a cross-validation of the machines run count and FDC analysis of long-term trends to determine the final suspicious factors to help engineers narrow the scope of exploration yield anomalies.

The experimental results show that Equipment and Chamber Variation can be effectively aberration and other anomalies. The problems discovered are shown in Table 1., which were found by the yield abnormalities detection and early warning system, Wafer Bin Map Pattern, baseline shift, lot to lot variation, and wafer to wafer variation.

These anomalies cover nearly 80% of the daily abnormalities generated in the plant, so the system can effectively help engineers to quickly sort out the abnormalities of daily products. At the same time, through the diagnostic analysis system, the scope of suspicion of anomalies can effectively be reduced to help engineers to quickly identify and confirm the causes of daily product anomalies.

Table 1. Description of empirical case

Case	Product	Product type	Problem type	Yield (↑)
A1	A	Pilot	Baseline Shift	0.7%
B1	C	Mass	Baseline Shift	0.19%
B3	A	Pilot	Baseline Shift	0.09%
B2	A	Pilot	Map Pattern	0.05%

A3	B	Mass	Baseline Shift	0.03%
A3	B	Mass	Baseline Shift	0.03%

This model was applied to 78 actual cases a year which are from important customers. The performance is as shown in Table 2. The average analysis time of a case was reduced from 13 days to 3.5 days, which greatly improved the efficiency of the analysis of the abnormality. At the same time, product A improved the yield rate by 7.08%, and the improvement of yield of product B was 3.94%. Every die in average improved the yield rate by 3%, bringing a potential profit of 10 million to the company.

Table 2. Improvement of the cases

#	Improvement	Before	After
1	Average analysis time/ per case	13 days	3.5 days
2	Product A yield	d1	d1 + 7.08%
3	Product B yield	d2	d2 + 3.94%

5 Conclusion

This study builds on the manufacturing and process data collected during the wafer fabrication process to provide an early warning of abnormal production, and helps diagnose anomalies and thus increase yield. According to the semiconductor wafer manufacturing yield testing data, this paper proposes a model construction of yield anomaly warning and analysis of abnormal factors diagnosis. Through the two stages of the model, we can find out the types of abnormal phenomena and yield variation at an early stage, and then conduct a diagnostic analysis upon the abnormal yield factor and cross-validate the hypothesis to find out the final suspicious factors causing issues.

The contribution of this study is to develop an effective yield anomaly detection and diagnosis model. It can shorten the analysis time to find out cause of defect, and help engineers to narrow down scope of suspicious factors and to improve the quality of decision making.

Due to the lengthy and complicated characteristics of semiconductor manufacturing process, many variables influence the model, which also may affect each other. In future research, we can focus on the process of interaction between the front-end and back-end to study the diagnosis of cross-site interaction impact factor. At the same time, we advise keeping a record of correct analyses found in previous each anomaly case. This helps to build an abnormal case database and establish an index that can be used to improve the accuracy of detection and diagnosis through machine learning and search techniques. In addition to finding anomaly factors, the wafer fab's most frequently asked question is whether or not to look for ways to study the golden path for best-in-class wafers, and we can do more to study the most effective process flow to find that path.

References

1. Chien, C.-F., Chang, K.-H., Wang, W.-C.: An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. Journal of Intelligent Manufacturing 25(5), 961–972(2014).
2. Chien, C.-F. , Chuang, S.-C.: A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics, IEEE Transactions On Semiconductor Manufacturing 27(4), 485-488 (2014).
3. Chien, C.-F., Diaz, A. C., Lan, Y.-B.: A data mining approach for analyzing semiconductor MES and FDC data to enhance overall usage effectiveness (OUE). International Journal of Computational Intelligence Systems 7(2), 52-65(2014).
4. Chien, C.-F. , Hsu, C.-Y..: Data Mining & Big Data Analytics. Future Culture, New Taipei City (2014).
5. Davenport, T.H.. Big Data. Global Views-Common Wealth Publishing group, Taipei (2015).
6. Eaton, C., Deroos, D., Deutsch, T., Lapis, G., Zikopoulos, P.: Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Companies, New York (2012).
7. Hsu, C.-Y.: Clustering Ensemble for Identifying Defective Wafer Bin Map in Semiconductor Manufacturing. Hindawi Publishing Corporation Mathematical Problems in Engineering (2015), Article ID 707358, http://dx.doi.org/10.1155/2015/707358.
8. Hsu, S.-C., Chien, C.-F.: Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing. International Journal of Production Economic 107(1), 88-103 (2007).
9. Kusiak, A.: Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing. IEEE Transactions On Electronics Packaging Manufacturing 24(1), 44-50(2001)
10. Li, J.: Industrial Big Data the Revolutionary Transformation and Value Creation in Industry 4.0 Era. CommonWealth, Taipei (2016).
11. Nakata, K., Orihara, R., Mizuoka, Y., Takagi, K.: A Comprehensive Big-Data-Based Monitoring System for Yield Enhancement in Semiconductor Manufacturing. IEEE Transactions On Semiconductor Manufacturing 30(4), 339 – 344 (2017).
12. Pyle, D.: Data Preparation for Data Mining. Morgan Kaufmann Publishers, San Francisco (1999).
13. Shirota, M.: Big data profit model: Graphic. Case. Strategy. Actual combat. EcoTrend Publications, Taipei (2013).
14. Wang, Y.-C., Li, Y.-F., Yuan, Q.-J.: International Big Data Research Hotspot and Frontier Evolution Visualization Analysis. Journal Of Engineering Studies (3), 282 – 293 (2014).
15. Yoo, Y., Park, S.-H., An, D., Kim, S.-S., Baek, J.-G.: A Spatial Point Pattern Analysis to Recognize Fail Bit Patterns in Semiconductor Manufacturing. International Scholarly and Scientific Research & Innovation. 8(2), 465-469 (2014).