Wafer Probe Yield Prediction Modeling based on General Regression Neural Network to Improve DRAM Processes

Chien-Chih Wang^{a,*}, Chi-Hung Kuo^a, Bing-De Wu^a

^aDepartment of Industrial Engineering and Management Ming Chi University of Technology, New Taipei, Taiwan

*Corresponding Author: ieccwang@mail.mcut.edu.tw

Abstract

In this paper, the General Regression Neural Network proposed to build a predicted model of the wafer probe yield that the key variables form the Wafer Acceptance Test (WAT). Using the actual situation of Taiwan DRAM production processes as an example. From the analysis results, it can get the 82 significant variables of the WAT, and the R^2 of GRNN prediction model is 0.73. After discussing with the manufacturer, the model was introduced into the actual process, and the results can effectively increase the yield and reduce quality cost in the DRAM manufacturing.

Keywords: feature selection, wafer acceptance test, data analysis.

1. Introduction

The yield is a major comprehensive indicator that measures financial affairs, fabrication ability, and the stable supply of products. A high yield can reduce costs and increase the marginal benefits of companies. Fabrication with a steady yield can facilitate production arrangement and attract orders, which can enhance the competitiveness of enterprises. The previous research on yield models for wafer concentrated on defect clustering ⁽¹⁾, productivity optimization ⁽²⁾, and interconnect yield analysis ⁽³⁾. A few researches emphasized on the yield prediction. Lee and Ha (2009) pointed out the key factors of yield in semiconductor fabrication are particles or contaminants on the wafer, substances in the manufacturing instruments, manufacturing process parameters, process engineers' attitudes, and the design of semiconductors ⁽⁴⁾. This study focused on the wafer probe test yield.

The Wafer Acceptance Test (WAT) is the fabrication variable test after wafer manufacturing and before the wafer

probe test ⁽⁵⁾. With an electrical property measurement, it aims to check if the circuit parameters of the wafer are in the acceptable range. Ke and Rao proposed the back propagation Artificial Neural Network model designed to infer electrical test parameters from the given list of parameters with the intention of reducing test time, to enhance throughput, and to improve cycle time ⁽⁶⁾. The most important function of the WAT is to ensure the normality of the chip and avoid a low yield rate. Additionally, WAT can reflect problems in the production line and judge the quality of metal according to the measurement results.

The Artificial Neural Network (ANN) takes time, and it is hard to interpret the results. Worse still, the determined key variables are not based on the correlation between the variables. ANN is a calculation system that uses a large number of simply connected neurons to stimulate biological neural networks and frequently used the method to improve quality and predict the yield of wafers ⁽⁷⁾⁽⁸⁾. Featuring high learning accuracy, fast recall, and high non-linear mapping, the back-propagation ability of ANN is a prediction model of monitoring-based learning. Hsu and Chien suggested integrating statistics with ANN to analyze the wafer cluster in WBM to locate defected wafers and apply them to the yield rate tests of wafer factories ⁽⁹⁾.

The research is a case study of an industry-university cooperative research project. At present, the primary reason for the underestimated prediction results of the cooperation of manufacturers is direct yield prediction; however, manufacturers are most concerned about the level of the yield. In the past, scholars used the number and area of wafer defects and defect distribution as the input variables of the yield rate prediction mode. However, such information is not obtained from the wafer probe test, and wafer manufacturers hope to get the information regarding yield distribution as early as possible. Therefore, this study concentrated on establishing the classification of the yield

grade of DRAM wafers and predicted the yield through common variables.

In this study, adopted the WAT variables to establish a yield grade classification model, which can identify yield of different grades, as well as the relationship among the WAT measurement variables. Moreover, it can select the common variables of the yield of different classes by the MTS classification mode to establish the Generalized Regression Neural Network (GRNN) yield prediction model. With the yield grade classification model and yield rate prediction model, it is possible to monitor quality in WAT fabrication. And find the crucial new WAT variables related to yield. The achievements of these two parts can eliminate the current bottleneck and significantly improve fabrication that will help enhance the production of manufactures and reduce quality costs.

2. Research Method

In this study, MTS and the GRNN were used for the selection of key variables and establishment of the classification and prediction models, respectively. There were four stages in the analysis process. In Stage 1, the WAT data obtained from the Engineer Database was appropriately pre-processed. In Stage 2, the reasons for the different yield grades were considered, and the information was grouped. Meanwhile, the yield was divided into five groups according to the suggestion of engineers, including “100%~85%”, “85%~80%”, “80%~40%”, “40%~2%”, and “2%~0%”. In Stage 3, MTS was adopted to find the key variables, from the 164 variables. In the calculation, the group with a yield of “100%~85%” was taken as the benchmark group and used to determine the critical variable combinations and classification models in the different groups. In Stage 4, the results of MTS were adopted as the input variables, and GRNN was used to establish the predictive model.

GRNN is a monitoring-based learning network derived from the Probability Neural Network. Its dynamic learning mode can be used for prediction and control, and shows high prediction ability, regardless of whether the regression model is linear or non-linear. Meanwhile, it can predict the pass rate according to previous information, and without hypothetical distribution. The Framework of GRNN, an input unit is a distribution unit, which distributes all x measurement values among all the pattern units in the second layer, where each pattern unit represents a training

example. If a new x vector enters the network, the squared value of the gap between the vector and the vector of the training sample will be aggregated and added into the non-linear function. The value obtained from this function will be the output value of the pattern unit. Then, the output value will be sent to the summation unit, which will aggregate the weight vectors and the result of all the observation values of y multiplied by the value of the weight vectors. For the last step, the evaluated value of y can be obtained by the division between the two output values of the summation unit. The only thing in GRNN that must be determined is a smooth parameter. In general, the holdout method, as proposed by Specht (1991), was adopted to define a smooth parameter.

3. Case study

The engineers selected 25,000 copies of information from the engineering information system to demonstrate the proposed method. WAT parameters include all electrical measurements about devices or transistors. To avoid superfluous data and reduce analysis time, it is essential to select relevant and useful WAT parameters. In this study, each of the 164 WAT variables refers to an electrical or physical feature value such as I_{sat} (saturation current), I_{off} (lower leakage), V_t (threshold voltage), R_c (contract resistance) and BV (breakdown). During the first three weeks, the data was taken as the training sample; in the fourth week, it was taken as the testing sample. Given that the key variables that influence the yield are not necessarily the same. The WAT data was divided into five groups according to the yield grade after the discussion of the engineers, including “100%~85%”, “85%~80%”, “80%~40%”, “40%~2%”, and “2%~0%”. Taking the group with the yield of “100%~85%” as the benchmark group, the researcher undertook MTS analysis of the remaining four groups.

First, the study conducted analysis where the group with the yield of “100%~85%” was taken as the benchmark group, and the group with the yield of “85%~80%” was regarded as an abnormal group. The use MTS to classify of yield grade. Taking (100%~85% vs. 85%~80%) as an example, the Trial and Error method was adopted to evaluate the threshold. Meanwhile, the number of WAT variables gradually declined to 134, 130, 122, 109, 74, 43, 27, 16, and 5, according to the different effect gains (>0 , >0.05 , >0.1 , >0.15 , >0.2 , >0.25 , >0.3 , >0.35 , >0.4). The

highest precision ($=0.7882$) was chosen to determine the variable combination, and the original 164 WAT variables were reduced to 134 variables. Likewise, the thresholds of (100%~85% vs. 40%~20%), (100%~85% vs. 20%~2%), and (100%~85% vs. 2%~0%) were 1.2, 1.2, and 1.7, respectively. The numbers of their WAT variables were 130, 128, and 114, respectively. Their precision was 82.29%, 82.98%, and 89.98%, respectively. Then, all the results were evaluated on the basis of the test samples. The precision of the four groups, (100%~85% vs. 85%~80%), (100%~85% vs. 40%~20%), (100%~85% vs. 20%~2%), and (100%~85% vs. 2%~0%) was 95.92%, 95.24%, 94.18%, and 95.31%, respectively. The selected 82 common WAT variables and GRNN were used to establish the pass rate prediction model. The researcher wrote a Matlab® program with an initial smooth parameter of 0.5. The program could automatically modify the smooth parameter to reduce the error to a minimum level. In this way, a smooth parameter value with the minimum error could be obtained, as shown in Table 1. Through comparison, the prediction model ($R^2=73.12\%$), as established with common variables, is better than the current prediction model ($R^2=33.84\%$).

Table 1. The result of GRNN between the variables by MTS and current variables

	Smooth paramet er α	Training sample			Testing sample		
		R^2	MSE	MAE	R^2	MSE	MAE
Variables by	0.39435	0.75	0.00	0.01	0.73	0.00	0.01
MTS	29	09	2	1	12	2	6
Current	0.18470	0.42	0.00	0.05	0.33	0.00	0.06
variables	59	12	8	91	84	83	27

4. Conclusions

The production of semiconductors involves thousands of complicated steps, which are closely related to each other. Any error in a process step will result in a lower wafer yield rate. With the WAT data offered by a DRAM manufacturer in Taiwan as the research subject. In this paper, the MTS and GRNN proposed to establish a classification and prediction yield model of a wafer probe test.

This study gives specific suggestions for practitioners to improve their WAT monitoring mechanism. Through a demonstration, the result can increase the wafer yield rate

and reduce quality cost in the DRAM manufacturing. Improving yield would significantly reduce the manufacturing cycle time. An accurate yield prediction model would help to prevent the production of non-conforming wafers before a malfunction is detected in the process. Practically, it is rather time-consuming and destructive to undertake the Chip Probe (C/P) with IC. Thus, effective monitoring of WAT variables can check if there is any problem in wafer fabrication, reduce or gradually replace the C/P test. Ensure the normal functions of Chip and avoid a low pass rate. For future research and application, it is suggested to adopt a multivariable control chart to monitor WAT variables. Development of big data analysis for quality problems in the production line or machine parameters can provide more efficient quality monitoring in fabrication.

References

- (1) Michalka, T. L., Varshney, R. C., & Meindl, J. D. "A discussion of yield modeling with defect clustering, circuit repair, and circuit redundancy", IEEE Transactions on Semiconductor Manufacturing, Vol. 3, No. 3, pp.116-127, 1990.
- (2) Stapper, C. H., McLaren, A. N., & Dreckmann, M. "Yield model for productivity optimization of VLSI memory chips with redundancy and partially good product", IBM Journal of Research and Development, Vol. 24, No. 3, pp.398-409, 1980.
- (3) Ghaida, R. S., & Zarkesh-Ha, P. "A layout sensitivity model for estimating electromigration-vulnerable narrow interconnects", Journal of Electronic Testing, Vol 25, No. 1, pp. 67-77, 2009.
- (4) Lee, J. H., & Ha, S. H. "Recognizing yield patterns through hybrid applications of machine learning techniques," Information Sciences, Vol. 179, No. 6, pp. 844-850, 2009.
- (5) Chien, C. F., Hsu, C. Y., & Hsiao, C. W. "Manufacturing intelligence to forecast and reduce semiconductor cycle time", Journal of Intelligent Manufacturing, Vol. 23, No. 6, pp. 2281-2294, 2012
- (6) Ke, J., & Rao, M. V. C. "Extraction of electrical test parameters by artificial neural network", International Journal of Industrial Engineering: Theory, Applications and Practice, Vol. 13, No. 1, pp. 71-80, 2006.
- (7) Sinanoglu, C. "Design of an artificial neural network for assembly sequence planning system", International Journal of Industrial Engineering: Theory,

Applications and Practice, Vol. 15, No. 1, pp. 92-103, 2008.

- (8) Tsai, T. N., & Tsai, C. W. "Development of a closed-loop diagnosis system for reflow soldering using neural networks and support vector regression", International Journal of Industrial Engineering: Theory, Applications and Practice. Vol. 21, No. 1, pp.19-33, 2014.
- (9) 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 Economics, Vol. 107, No.1, pp. 88-103, 2007.