Prediction of Package Chip Quality Using Fail Bit Count Data of the Probe Test

Jin Soo Park and Seoung Bum Kim(✉)

Department of Industrial Management Engineering, Korea University, Seoul, Korea
{dmqmpjs,sbkiml}@korea.ac.kr

Abstract. The quality prediction of the semiconductor industry has been widely recognized as important and critical for quality improvement and productivity enhancement. The main objective of this paper is to establish a prediction methodology of semiconductor chip quality. Although various research has been conducted for predicting a yield, these studies predict a yield by lot-level and do not consider characteristics of the data. We demonstrate the effectiveness of the proposed procedure using a real data from a semiconductor manufacturing.

Keywords: Quality prediction · Probe test · Smote · Nonparametric variable selection

1 Introduction

The quality is directly connected to the competitiveness of the companies. High quality products improve the reliability and customer satisfaction. For this reason, many manufacturing companies are currently working on an effort to improve the quality [1]. In particular, semiconductor market is rapidly growing, and manufacturers are focusing on early development of new products, mass production improvement, and quality management to strengthen business competitiveness power. Quality control can be divided into prediction and follow-up service. The former is predicting the quality by using manufacturing process parameters and the product properties to prevent defects in advance. The latter is to correspond and action when the customer claims occur. To improve competitiveness of the company, quality prediction and pre-detection of defects is very important.

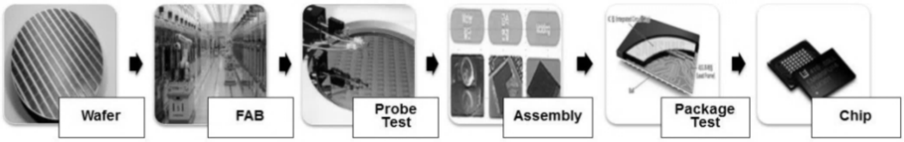

Fig. 1. The semiconductor manufacturing process

The semiconductor manufacturing usually consists of 200–300 process steps, and it takes three to four months to produce the final chips. As shown in Fig. 1, the semiconductor manufacturing processes can be generally divided into four basic

M. Ali et al. (Eds.): IEA/AIE 2015, LNAI 9101, pp. 655–664, 2015.
DOI: 10.1007/978-3-319-19066-2_63

processes: fabrication (FAB), probe test, assembly, and package test [2]. The FAB process forms hundreds to thousands of chips on the pure wafer by going through a process unit such as the photo and the etch process. Probe test, known as wafer test, provides key information about the performance of the wafer fabrication process. It involves testing of individual chips for their functionality based on different electrical probes. Assembly step separates the chips from a wafer and packs them to protect physical impacts from the outside. Packaged chips are sent for packaging test to determine the quality of chips by checking in harsh environments than users use [3].

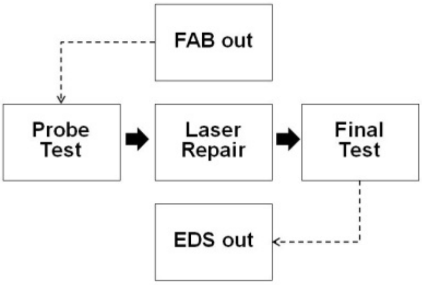

Fig. 2. Probe test process

Probe test is the first step in which the chip-level data are generated and thus an important process. Fig. 2 describes the probe test process in which the fail bit count (FBC) data are generated. Probe test is referred to as electric die sorting (EDS) process. As described in Fig. 3, the semiconductor memory chip is made up of the Giga cell, and hence defects can be present. Redundancy cells are a set of spare cells for repairing. The redundancy cells consist of spare rows and columns. Laser repair replaces the defect cells with the redundancy cells. Through this process, the yield can be significantly enhanced[4].

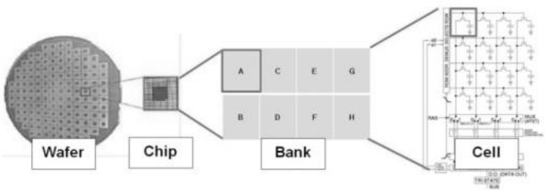

Fig. 3. The architecture of the DRAM chip

In the semiconductor manufacturing industry, much research has been conducted to improve the quality while maintaining high yield rates. As mentioned earlier, the probe test plays a significant role in the prediction of final chip quality because the probe test first generates the chip-level data. Through the chip quality prediction of the probe test step, high-quality chips and low-quality chips can be classified. Quality grading can lead to a dual package test and a dual manufacturing process. Through this, test time in the manufacturing process is reduced and the improvement in the yield can be achieved.

Thus, the studies using a probe test data were conducted because the probe test process is important. A semiconductor yield prediction using stepwise support vector machine [5], a package test yield prediction using wafer bin map [6] are conducted, however, these studies predicted yield by lot-level and focused on the accuracy of the overall model, not the sensitivity. Generally, the performance measure in classification is the accuracy, however, in the quality level, sensitivity should be taken into consideration. In particular, the semiconductor data have an imbalance problem. Data imbalance problem occurs when the number of high-quality chips are much larger than low-quality chips [7]. In this paper, we propose an efficient quality prediction methodology considering the characteristic of the FBC data from the probe test.

The rest of this paper is organized as follows. In Section 2, we illustrate the data imbalance problem. Section 3 presents the nonparametric variable selection. In Section 4, we present the quality prediction methodology with the FBC data from the probe test. In Section 5, we give some concluding remarks.

2 Data Imbalance Problems and Solutions

The FBC data from the probe test exhibit an imbalance problem. However, most classification algorithms are well trained under the assumption that the number of observations in each class is roughly equal [7]. In general, to deal with the problem caused by the imbalanced data, three methods have been previously proposed.

First, undersampling methods [8] address imbalance problems by sampling a small number of observations of the majority class. Not only can undersampling methods enhance the classification performance, but also reduce the computational costs since they samples a small number of observations from the majority class. However, applying undersampling methods has a possible drawback of biasing the distribution of the majority class. If the sampled observations from the majority class do not follow the original distribution, it may decrease in the classification performance. This possible disadvantage can be happened if the number of minority class observations is very small.

Second, oversampling methods [9, 10] solve the imbalance problems by copying observations from the minority class. In contrast with the undersampling methods, since oversampling methods contain all of the information on the original observations, it can accomplish a comparatively high performance. However, the computational costs training the classification models increase since the number of observations used in training is much larger than the number of the original observations.

The third methods, which are a combination of the two methods above, deal with the imbalance problems by sampling a small number of observations of the majority class and copying observations of the minority class [11]. The combination of these two methods is not always a good thing than using only undersampling methods. If we replicate the minority class, the decision region of the minority class becomes very specific, but does not spread into the majority class region, which makes the decision region as overfitted. One of the solutions to resolving this overfitting problem is the synthetic minority oversampling technique (SMOTE). SMOTE is a method that the minority class is oversampled by creating synthetic observations rather than by oversampling with replacement. For more details, please refer to Chawla (2002) [9].

3 Nonparametric Variable Selection

In the high-tech manufacturing process such as semiconductor process, a large amount of variables that are correlated with each other is generated. In this case, variable selection is critical. The main objective of variable selection is to identify a subset of variables that are most predictive of a given response variable. Variable selection is particularly of interest when the number of candidate explanatory variables is large, and when many redundant or irrelevant variables are thought to be present [12].

Therefore a dimensionality reduction process, which find the significant variables, is essential. In addition, since the FBC data from the probe test used in this paper have a number of variables, the dimensionality reduction technique can be necessary. In general, by performing the two-sample t-test [13] for each variable, the important variables will be selected. However the two-sample t-test is based on the parametric assumption. The FBC data from the probe test do not follow the normal distribution, as shown in Fig. 4. Therefore, we use a nonparametric variable selection technique. In this paper, we use the nonparametric resampling t-test among the various variable selection techniques [14].

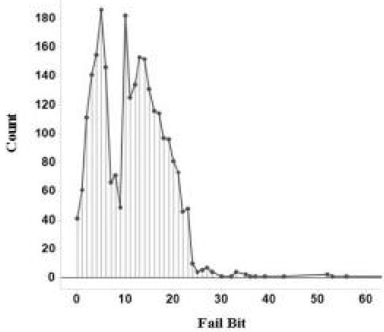

Fig. 4. The distribution of the fail bit count data

The nonparametric resampling t-test is how to find statistically significant variables in comparison to original t-statistics and resampling t-statistics without distinction of groups. Now we calculate the statistics of all the variables using following equation for each groups by employing a two-sample t-test.

$$t_i = \frac{\overline{X_{i,normal}} - \overline{X_{i,abnormal}}}{\sqrt{\frac{S^2_{i,normal}}{n_{normal}} + \frac{S^2_{i,abnormal}}{n_{abnormal}}}} \tag{1}$$

for i=1,2,…,29. i is the index of the predictor variables, $\overline{X_{i,normal}}$ and $S^2_{i,normal}$ are the sample mean and variance of normal groups. Likewise, $\overline{X_{i,abnormal}}$ and $S^2_{i,abnormal}$ are taken from abnormal groups. Next, we calculate p-values using a permutation method due to repeated measurements and thus we cannot clearly assume that each t_i follows a t-distribution. Under the assumption that there is no differential fail bit count level between the two classes (normal and abnormal groups), the t-statistic should have the same distribution regardless of how we make the permutation of fail bit count. Therefore, we can permute (shuffle) the two groups, and re-compute a set of t-statistics for each individual fail bit count feature based on the permuted dataset. If this procedure is repeated N times, we can obtain N sets of t-statistics as follows : $t_1^n, t_2^n, \ldots, t_{29}^n, n = 1,2,\ldots,N$. The nonparametric p-value for i=29 and N=1,000 is obtained by

$$p_i = \sum_{n=1}^{1000} \frac{\#\{k: |t_k^n| \geq |t_i|, k=1,2,\ldots,29\}}{29 \cdot 1000}. \tag{2}$$

Finally, we conduct a variable selection by applying the false discovery rate (FDR) [18] using these p-values. We can summarize the procedure of the variable selection based FDR as follows:

- Ordering the p-values in ascending $(p(1) \leq p(2) \leq \ldots \leq p(29))$
- Select a desired FDR level($= \alpha$) between 0 and 1 in this paper, we select 0.05
- Calculate the largest i denoted as w

$$w = \max\left[i: p(i) \leq \frac{i}{m}\frac{\alpha}{\delta}\right], \tag{3}$$

where m is the total number of variables (here m=29) and δ denotes the proportion of true null hypothesis. Many studies discuss the assignment of δ. In this paper, we use δ=1, the most conservative value.

- The threshold of the p-value is $p_{(w)}$, and declare the fail bit count feature t_i significant if and only if $p_i \leq p_{(w)}$.

4 The Quality Prediction Using the FBC Data from the Probe Test

4.1 Data Description and Performance Measure

The data used in this study are obtained from the real FBC data from the probe test in semiconductor manufacturing. The dataset contains 29 variables and 2,623 observations (2,000 high-quality chips, 623 low-quality chips). As shown in Table 1, the predictor variables, X are discrete count data, and the response variable, Y is binary, indicating whether the corresponding chip is high quality or low quality. Fig. 5 describes a three-dimensional principal component score plot [15] showing that high-quality observations and low-quality observations are overlapped with each other.

Table 1. The fail bit count data of probe test

Wafer	X1	X2	...	X28	X29	Y
1	9	21	...	1	0	0
2	8	16	...	0	0	0
⋮	⋮	⋮	...	⋮	⋮	⋮
2622	5	12	...	0	1	1
2623	3	6	...	2	0	1

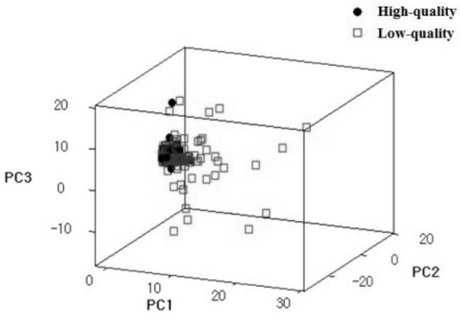

Fig. 5. A three-dimensional principal component score plot

We applied a 10-fold cross validation [15] to obtain reliable results. The performance of the proposed method is evaluated by sensitivity [16]. That can be calculated by the following equation:

$$\text{Sensitivity} = \frac{TP}{TP+FN}. \tag{4}$$

Table 2. Confusion matrix

		Predicted	
		Abnormal	Normal
Actual	Abnormal	TP	FN
	Normal	FP	TN

- TP (True Positive): the number of positive examples correctly predicted by the classification model.
- FN (False Negative): the number of positive examples wrongly predicted as negative by the classification model.
- FP (False Positive): the number of negative examples wrongly predicted as positive by the classification model.
- TN (True Negative): the number of negative examples correctly predicted by the classification model.

4.2 A Prediction Methodology of Package Chip Quality

In this study, we used the three steps to construct prediction models. Step 1 solves the imbalance problem using SMOTE. Step 2 identifies important variables by using the non-parametric techniques. Finally, Step 3 determines the relevant values of the parameters such that the sensitivity is maximized. We provide detailed descriptions of these steps as follows:

Step 1. Solving the Data Imbalance Problem

As mentioned in Chapter 2, the FBC data from the probe test have an imbalance problem. Because we are more interested in detecting low-quality chips of the minority class, the imbalance problem should be resolved [16]. In this paper, we used the SMOTE algorithm.

Fig. 6 shows the sensitivity values from various classification algorithms before and after applying the SMOTE. It can be clearly seen that the sensitivity is improved by using SMOTE.

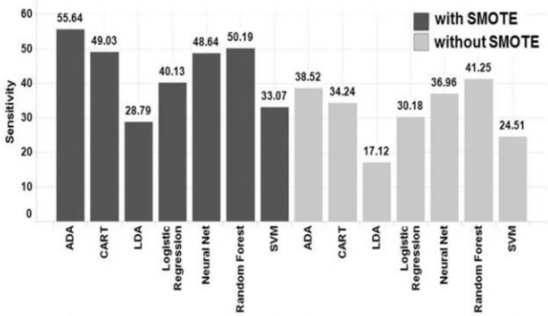

Fig. 6. The sensitivity with SMOTE technique

Step 2. Variable Selection

The number of predictor variables considered in this study is 29. Here we used a non-parametric approach to select important variables because the probability distribution of the data is unknown [17]. Especially, we performed the nonparametric resampling t-test [14] to obtain a set of p-values for each variable. We applied the false discovery rate (FDR) [18] using these p-values to select a significant variables.

As a result of variable selection, 13 of the 29 variables were selected. Looking at the non-selected variables, their correlation is high or they do not affect the outcome variable since a high-quality class and a low-quality class are almost equal.

Step 3. Adjusting the Parameters of the Model

As shown in Fig. 5, we recognized that high-quality observations and low-quality observations are overlapped with each other. In the quality prediction, the classification accuracy of the low-quality observations is very important. Therefore, we adjust the parameters of the model for increasing the sensitivity. In this study, we adjust the threshold of a logistic regression [15]. The results of the logistic regression come with probability values that belong to a particular category, as shown in the following equation.

$$\mathrm{p(x)} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{5}$$

Threshold is a criteria for determining whether observations belong to one category of the two categories. In general use of logistic regression, the threshold is set at 0.5 if the number of observations in each class is roughly equal. However, the FBC data have an imbalance problem. Therefore, if we use 0.5 as the threshold, the minority class can not be classified properly. Hence, we should find an appropriate threshold to maximize the sensitivity. Furthermore, the classification accuracy of the high-quality observations is important, thus, we adjust the threshold that the specificity is accomplished at least 50%. The specificity can be calculated by the following equation:

$$\text{Specificity} = \frac{TN}{FP+TN} \tag{6}$$

Fig. 7 is the box plot which describes the probability distribution of the two classes. If we use 0.5 as the threshold, the high-quality observations are almost classified properly. However, low-quality observations cannot be correctly classified. In this study, we set the threshold at 0.36 which is a Q1 (the first quartiles) of the low-quality observations. Fig. 8 is a diagram showing the change in sensitivity. Sensitivity is improved about 30% after changing the threshold 0.5 to 0.36.

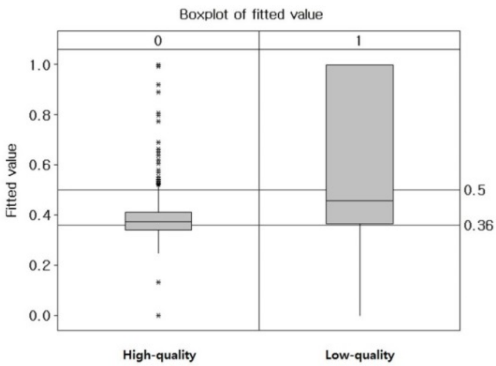

Fig. 7. Fitted value of the logistic regression

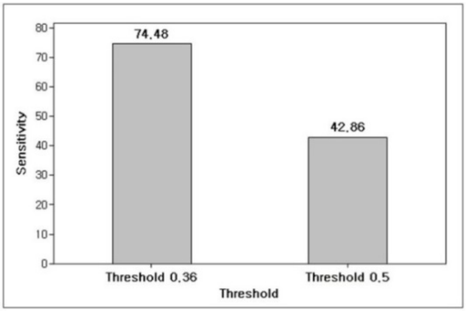

Fig. 8. The sensitivity difference between thresholds

4.3 Experiment Results

Table 3 shows the sensitivities of applying the SMOTE, the nonparametric variable selection and adjustment of the parameters of the model. The sensitivity was improved about 10% by the SMOTE technique. Besides, important variables are selected with eliminating redundant information, which eventually raise the sensitivity of the prediction model. Finally, by adjusting the threshold of the logistic regression, the sensitivity was improved from 42.86% to 74.48%. This results clearly demonsrate the effectiveness of the proposed procedure to predict the final chip quality based on FBC data obtained from the probe test.

Table 3. The result of experiment

	Sensitivity
Original data	30.18%
SMOTE	40.13%
SMOTE + Variable Selection	42.86%
SMOTE + Variable Selection + Adjusting the threshold	74.48%

5 Conclusions

The probe test is a critical step in the prediction of final chip quality. In this paper, we propose a quality prediction methodology using the FBC data obtained from the probe test. Most classification algorithms are well trained with balanced data. However, the FBC data from the probe test are highly imbalanced, and hence, we proposed to use the SMOTE algorithm to address imbalanced problems. In addition, since the FBC data from the probe test do not follow the normal distribution, we used a nonparametric variable selection technique to identify the important variables for prediction. Finally, by adjusting the parameters of the prediction model, the high sensitivity can be obtained.

References

1. Kang, J., Kim, S.: Bootstrap-Based Fault Identification Method. Journal of Korean Society for Quality Management **39**(2), 234–244 (2011)
2. Chen, H., Harrison, J.M., Mandelbaum, A., Wein Lawrence, M.: Empirical evaluation of a queueing network model for semiconductor wafer fabrication. Oper. Res. **36**(2), 202–215 (1988)
3. Kumar, N., Kennedy, K., Gildersleeve, K., Abelson, R., Mastrangelo, C.M., Montgomery, D.C.: A review of yield modelling techniques for semiconductor manufacturing. International Journal of Production Research **44**(23), 5019–5036 (2006)
4. Cunningham, S.P., Spanos, C.J., Voros, K.: Semiconductor yield improvement: Results and best practices. IEEE Transactions on Semiconductor Manufacturing **8**(2), 103–109 (1995)
5. An, D., Ko, H.H., Gulambar, T., Kim, J., Baek, J.G., Kim, S.S.: A semiconductor yields prediction using stepwise support vector machine. In: IEEE International Symposium on Assembly and Manufacturing, ISAM 2009, pp. 130–136. IEEE (2009)
6. 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 **107**(1), 88–103 (2007)
7. Kang, P., Cho, S.: EUS SVMs: ensemble of under-sampled SVMs for data imbalance problems. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4232, pp. 837–846. Springer, Heidelberg (2006)
8. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: one-sided selection. In: Proceedings of the Fourteenth International Conference on Machine Learning, pp. 179–186 (1997)
9. Chawla, N.V., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Oversampling Techniques. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
10. Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W.: SMOTEBoost: improving prediction of the minority class in boosting. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) PKDD 2003. LNCS (LNAI), vol. 2838, pp. 107–119. Springer, Heidelberg (2003)
11. Solberg, A., Solberg, R.: A Large-scale evaluation of features for automatic detection of oil spills in ers sar images. In: International Geoscience and Remote Sensing Symposium, pp. 1484–1486 (1996)
12. Shmueli, G., Patel, N.R., Bruce, P.C.: Data mining for business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons (2007)
13. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K.: Probability and statistics for engineers and scientists, vol. 5. Macmillan, New York (1993)
14. Kim, S.B., Chen, V.C., Park, Y., Ziegler, T.R., Jones, D.P.: Controlling the False Discovery Rate for Feature Selection in High-resolution NMR Spectra. Statistical Analysis and Data Mining **1**(2), 57–66 (2008)
15. Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R.: The Elements of Statistical Learning, vol. 2(1). Springer, New York (2009)
16. Tan, P., Steinbach, M., Kumar, V.: Introduction to data mining, 1st edn. Addison Wesley (2006)
17. Conover, W.J.: Practical Nonparametric Statistics, 3rd edn. Wiley, New York (1999)
18. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289–300 (1995)