

# 運用資料探勘技術建構半導體封裝業之品質改善系統

研究生：吳瑞琦

指導教授：陳瑞順博士

國立交通大學資訊管理研究所

## 摘 要

在半導體封裝產業中，製程的良率一直是同業競爭中的關鍵因素，有鑑於此，本研究乃針對提高製程良率而建置一套品質改善系統。個案將以液晶顯示器驅動積體電路之捲帶式封裝製程為研究主體，透過改善關鍵製程來提高良率，以減少因為未將製程參數最佳化所產生的瑕疵。

首先，我們將蒐集來的失效產品資料經由群集演算法探勘獲得瑕疵種類分群的數量，然後決定出各群的定義，並導入資料倉儲中。接著再採用適合分類之演算法，如決策樹、類神經網路、羅吉斯迴歸、貝氏分類與關聯規則等，利用多準則決策從中評估與比較以挑選出最佳的演算法來進行關鍵變數之分析。

由本研究的結果，顯示出不良產品的關鍵因素並探討出問題解決對策，以提供品質管制工程人員做為製程參數最佳化之參考。經導入本品質改善系統後，大大地提昇了整體產品良率、減少了重工/停機次數與縮短了產品生產週期，因此對半導體封裝業而言，本研究是提昇競爭力以因應瞬息萬變市場的有利工具。

**關鍵字：**資料探勘、資料倉儲、群集、決策樹、類神經網路、羅吉斯迴歸、貝式、關聯法則、線上分析處理、捲帶式封裝、多準則決策

# Using Data Mining Technology to Construct a Quality Improvement System for Semiconductor Packaging Industry

Student: Ruey-Chyi Wu

Advisor: Dr. Ruey-Shun Chen

Institute of Information Management  
National Chiao Tung University

## Abstract

Yield enhancements are always the key factor among competitors in the semiconductor packaging industry. In view of this, this paper attempts to construct a quality improvement system to increase process yield. Case study focuses mainly on the tape carrier package process of liquid crystal display driver IC by improving key process for the increase of yield rate. The parameter settings should be taken into account to avoid that the defects happen due to process parameters not being optimized.

First, we collect large amounts of data of defective products and use clustering algorithm to obtain the amount of defective categories. Next, the definitions of these clusters are determined and loaded into the data warehouse. Then, the suitable classification algorithms, decision tree, neural network, logistic regression, Bayesian, and association rule, are used for evaluation and comparison with multiple criteria decision making to select a better algorithm for the analysis of key variables.

From the result of this research, major causes of defective products are investigated and countermeasures of the problems will provide know-how for the optimization of process parameters to quality control engineers. After the proposed quality improvement system is applied, the benefits are obtained to improve product yield reduction of rework/holding lot frequency and product cycle time. Therefore, to semiconductor packaging industry, this study provides a most beneficial tool to improve competitiveness for responding to the ever changing marketplace.

**Keywords:** Data mining; Data warehouse; Clustering; Decision tree; Neural network; Logistic regression; Bayesian; Association rule; OLAP; TCP; Multiple criteria decision making

## 誌 謝

回顧四年多來的博士班學術研究生涯，身邊有許多幫助我成長的人，在此致上深深的謝意。

首先，感謝我的論文指導教授，在他熱忱與耐心的指導下，使我可以順利的完成所有的學分科目、英文檢定與期刊發表等，以合乎畢業規章。再者，在畢業論文計畫書撰寫期間，感謝口試委員們的細心指導，使得論文不足之處得以增進，在正式論文口試時，教授們更不吝撥冗、不辭辛勞地參與論文審查，並毫無保留地提供許多寶貴的意見，使得論文內容得以更趨完善，在此表達對教授們的誠摯謝意。

同時，也感謝我任職單位的長官和同仁們對我的體諒，使得本人能利用工作之餘順利完成交大之繁重課程。

最後，我要特別感謝內人與兒子，由於他們的包容與支持，使得我可以全力以赴取得博士學位。

# Contents

<b>Abstract (in Chinese)</b> .....	I
<b>Abstract (in English)</b> .....	II
<b>Acknowledgement</b> .....	III
<b>Contents</b> .....	IV
<b>List of figures</b> .....	VI
<b>List of tables</b> .....	IX
<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Motivation.....	1
1.2 Goal.....	1
1.3 Research Architecture .....	3
1.4 A Guide to this Dissertation .....	4
<b>Chapter 2 Related Works</b> .....	<b>6</b>
2.1 Data Warehouse .....	6
2.2 Data Mining .....	7
2.3 Traditional Semiconductor Packaging Processes.....	10
<b>Chapter 3 Tape Carrier Package Technology</b> .....	<b>12</b>
3.1 LCD driver IC Packaging Processes.....	12
3.2 Tape Carrier Package Processes.....	13
<b>Chapter 4 System Design</b> .....	<b>19</b>
4.1 Design a Framework for the Quality Improvement System .....	19
4.2 Establishing a Data Warehouse System .....	21
4.2.1 Establishing Data Warehouse Architecture .....	21
4.2.2 Establishing Data Warehouse Procedures .....	21
4.2.3 Establishing a Schema for Data Warehouse.....	22
4.2.4 Establishing a Multidimensional Model .....	25
4.3 Integrating Decision Analysis and Data Mining Systems.....	26
4.4 Establishing Data Mining Systems .....	27

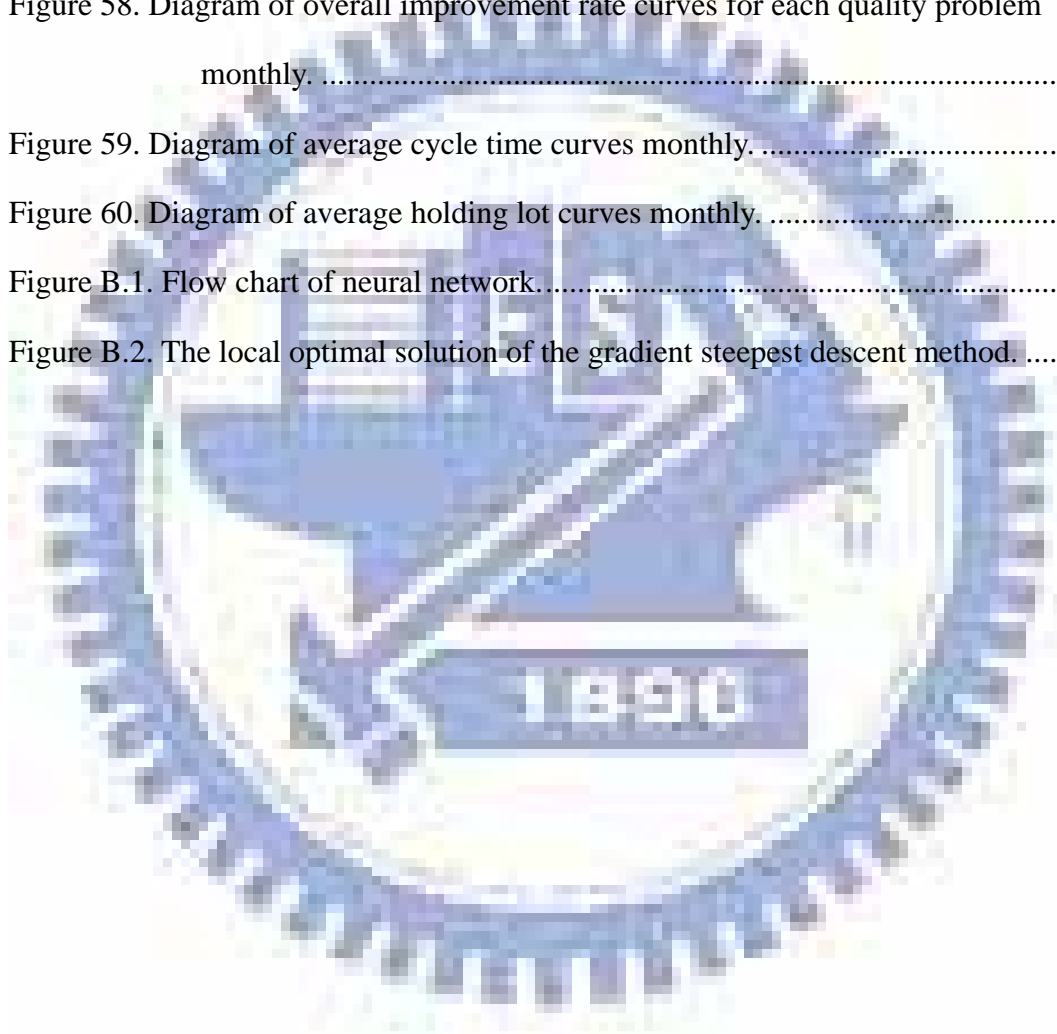
4.4.1 Set Up of the Clustering Data Mining Engine .....	27
4.4.2 Classification and Predictions Procedures .....	29
4.5 Establishing Visual Figures of Data Mining .....	42
<b>Chapter 5 System Implementation.....</b>	<b>43</b>
5.1 System Implementation Environment.....	43
5.2 Data Collection .....	43
5.3 Data Cleaning.....	47
5.4 New Data Generation.....	47
5.5 Establishing a Data Warehouse.....	48
5.6 Clustering and Defining Meanings .....	50
5.7 Implementing Proposed Algorithms .....	55
5.7.1 Decision tree .....	55
5.7.2 Neural network.....	60
5.7.3 Logistic regression.....	64
5.7.4 Bayesian.....	67
5.7.5 Association rule.....	70
5.8 Comparing and Evaluating Data Mining Models .....	72
5.9 OLAP Analysis .....	79
5.10 Practical Operation of a Better Model .....	81
5.11 Benefits Analysis.....	85
<b>Chapter 6 Contributions.....</b>	<b>88</b>
6.1 Contributions.....	88
6.2 Future Works .....	88
<b>Reference.....</b>	<b>90</b>
<b>Appendix A Decision Tree .....</b>	<b>94</b>
<b>Appendix B Neural Network.....</b>	<b>98</b>

## List of Figures

Figure 1. Tape carrier package processes analyzed by this research. ....	3
Figure 2. The research architecture. ....	4
Figure 3. Knowledge discovery process. ....	8
Figure 4. Traditional semiconductor packaging processes. ....	11
Figure 5. LCD driver IC back-end packaging processes. ....	12
Figure 6. Tape carrier package processes. ....	14
Figure 7. Tape carrier package tape. ....	14
Figure 8. Illustration of tape carrier package processes technology. ....	15
Figure 9. Illustration of inner lead bonding processes flow. ....	16
Figure 10. Major sealing processes using liquid sealing material. ....	17
Figure 11. A framework of the quality improvement system for semiconductor packaging industry. ....	19
Figure 12. The data warehouse system established through the top-down model. ....	21
Figure 13. The framework for data warehouse procedures. ....	22
Figure 14. A starflake schema for the data warehouse. ....	24
Figure 15. Time periods for the data warehouse. ....	24
Figure 16. Example of a three-dimension data cube. ....	26
Figure 17. Integration of decision-making analysis and data mining systems. ....	26
Figure 18. Calculation flow of E-M algorithm. ....	28
Figure 19. Setting up model and applying procedures. ....	30
Figure 20. Process of growth of decision tree rules. ....	30
Figure 21. Feed-forward network flow chart. ....	36
Figure 22. Logit function. ....	36
Figure 23. The environment and framework of product quality improvement system. .....	43
Figure 24. Major causes of product defects. ....	47

Figure 25. Microsoft SQL Server 2005 SSIS diagram. ....	48
Figure 26. The quality improvement system starflake schema.....	48
Figure 27. Illustration of clustering classification. ....	52
Figure 28. Display of decision tree model classification. ....	57
Figure 29. Matrix classification of decision tree. ....	58
Figure 30. Lift chart of decision tree. ....	59
Figure 31. Processing time of decision tree. ....	59
Figure 32. Missing value automatically filled in missing with decision tree. ....	59
Figure 33. The architecture of neural network by the research. ....	61
Figure 34. Display of neural network model classification. ....	61
Figure 35. Matrix classification of neural network.....	62
Figure 36. Lift chart of neural network.....	62
Figure 37. Processing time of neural network. ....	63
Figure 38. Missing value automatically filled in missing with neural network.....	63
Figure 39. Display of logistic regression model classification.....	65
Figure 40. Matrix classification of logistic regression.....	65
Figure 41. Lift chart of logistic regression.....	66
Figure 42. Processing speed of logistic regression. ....	66
Figure 43. Missing value automatically filled in missing with logistic regression. ....	66
Figure 44. Display of Bayesian classification.....	68
Figure 45. Matrix classification of Bayesian. ....	68
Figure 46. Lift chart of Bayesian ....	69
Figure 47. Processing speed of Bayesian.....	69
Figure 48. Missing value automatically filled in missing with Bayesian. ....	69
Figure 49. Display of association rule classification. ....	71
Figure 50. Matrix classification of association rule.....	71
Figure 51. Lift chart of association rule.....	72

Figure 52. Processing speed of association rule. ....	72
Figure 53. Lift chart of proposed data mining algorithms. ....	74
Figure 54. Process progress and speed for mining structures. ....	75
Figure 55. Drill-down of OLAP display for the three-dimension cube.....	80
Figure 56. Roll-up of OLAP display for machine’s attributes.....	80
Figure 57. The decision tree for training data.....	83
Figure 58. Diagram of overall improvement rate curves for each quality problem monthly. ....	86
Figure 59. Diagram of average cycle time curves monthly. ....	86
Figure 60. Diagram of average holding lot curves monthly. ....	87
Figure B.1. Flow chart of neural network.....	98
Figure B.2. The local optimal solution of the gradient steepest descent method. ....	100





## List of Tables

Table 1 Most used data mining/analytic methods.....	2
Table 2 Classification of data mining technology.....	2
Table 3 Advantages and disadvantages comparisons among TCP, COG and COF.....	13
Table 4 Parameters setting by neural network algorithm.....	34
Table 5 Defective types in the manufacturing process .....	44
Table 6 Four attributes contrast table.....	46
Table 7 Manufacturing fact table .....	49
Table 8 Quality dimension table .....	49
Table 9 Man dimension table.....	49
Table 10 Machine dimension table .....	49
Table 11 Material dimension table.....	50
Table 12 Method dimension table.....	50
Table 13 Parameters of clustering algorithm setting by this research .....	51
Table 14 Number of all defective clusters and their percentage.....	52
Table 15 Contrast table of numbers of attribute variables and quality problems .....	54
Table 16 Independent coding table for all attribute variables and judgment items .....	55
Table 17 Parameters of decision tree algorithm setting by this research.....	56
Table 18 Total and accurate entries of decision tree .....	58
Table 19 Parameters of neural network algorithm setting by this research .....	60
Table 20 Total and accurate entries of neural network .....	62
Table 21 Parameters of logistic regression algorithm setting by this research .....	64
Table 22 Total and accurate entries of logistic regression .....	65
Table 23 Parameters of Bayesian algorithm setting by this research.....	67
Table 24 Total and accurate entries of Bayesian .....	68
Table 25 Parameters of association rule algorithm setting by this research .....	70
Table 26 Total and accurate entries of association rule.....	71

Table 27 Accuracy rate of classification matrix.....73  
Table 28 Summy of evaluating data mining models.....76  
Table 29 ILB training data set.....81

