JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A Study on the Fault Process and Equipment Analysis of Plastic Ball Grid Array Manufacturing Using Data-Mining Techniques

Hyun Sik Sim*

Abstract

The yield and quality of a micromanufacturing process are important management factors. In real-world situations, it is difficult to achieve a high yield from a manufacturing process because the products are produced through multiple nanoscale manufacturing processes. Therefore, it is necessary to identify the processes and equipment that lead to low yields. This paper proposes an analytical method to identify the processes and equipment that cause a defect in the plastic ball grid array (PBGA) during the manufacturing process using logistic regression and stepwise variable selection. The proposed method was tested with the lot trace records of a real work site. The records included the sequence of equipment that the lot had passed through and the number of faults of each type in the lot. We demonstrated that the test results reflect the real situation in a PBGA manufacturing process, and the major equipment parameters were then controlled to confirm the improvement in yield; the yield improved by approximately 20%.

Keywords

Fault Process and Equipment Analysis, Logistic Regression, Plastic Ball Grid Array Manufacturing Process, Yield Management

1. Introduction

A printed circuit board (PCB) consists of electronic components that are connected electrically through coppered circuits. PCBs are widely used in general household appliances, such as TVs, and in precision machineries such as smartphones. A general household appliance that performs one simple function only requires a simple PCB structure. However, the latest precision machineries, such as smartphones, use more complex PCBs for high functionality and segmentation. These PCBs include various types of packages such as leaded, chip scale, plastic ball grid array (PBGA), and flip-chip ball grid array (FCBGA) packages. Complex substrate structures degrade the manufacturing competitiveness of companies by not only increasing the production cost but also decreasing the yield of their products. To maintain high yields and quality while decreasing the defect rate, it is critical to maintain stable operation and preservation of equipment and to identify major equipment parameters that can achieve the best management efficiency at limited cost and resources.

The PBGA manufacturing process involves copper clad lamination (CCL), copper (Cu) plating, first

patterning, lay-up and drilling, Cu plating, second patterning, solder resist (SR) coating, gold (Au) plating, and routing processes (refer to Fig. 1). First, the substrates, which are the raw materials, are cut for each product type and fed into the process. In the first patterning, a surface plating for the conductive layer is formed on the raw material substrate, a dry film is attached, and after ultra violet (UV) exposure, the desired pattern is formed by the Develop, Etch, and Strip processes. In the lay-up, a multilayered substrate is formed by adding copper foils to the patterned substrate. The lay-up is followed by drilling, where holes for electrical connections are drilled on the top, bottom, inside, and outside of the multilayered substrate. Cu plating is performed for the circuit connections after drilling is completed. In the second patterning, the circuit is configured on the newly formed layer from the lay-up process after the Develop, Etch, and Strip processes. In the SR coating process, an SR layer is used to protect the surface circuit of the product. Finally, Au plating is performed to improve the electrical conductivity and corrosion resistance of the solder ball. Then, the production is completed through a routing process.
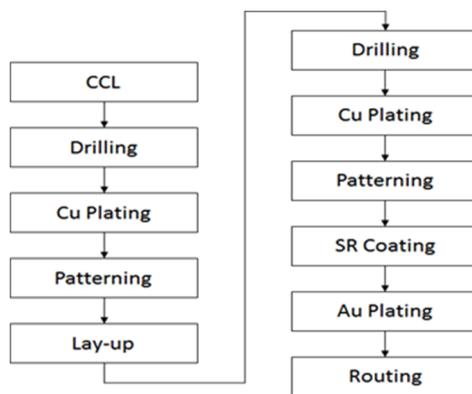

Fig. 1. PBGA manufacturing process flowchart.

In PCB manufacturing, quality has been managed for many years through the statistical process control, which verifies faults by measuring the substrate plating thickness or wire width after the production process is completed. In particular, the yield is determined by checking the connection of the circuit line and the condition of the circuit after the first patterning. Faults in the first patterning process are important factors that increase the product cost as further costs are incurred in the post-processing of defective products. Therefore, it is necessary to maximize the yield by minimizing the faults. In particular, in the inspection process, it is critical to accurately identify the processes and equipment that have many faults and lead to low yields.

However, it is very difficult to identify the processes and equipment in which the key fault factors of the inspection process appear because there are many equipment parameters in the PCB manufacturing process. Considering the equipment performance, analysis resource, and time, it is very difficult to extract and analyze all these parameters [1]. To enhance the manufacturing competitiveness, it is most effective to analyze the path through which the product has passed the equipment, with no equipment parameter data, and to identify and manage the fault-suspected processes and equipment. Thus far, logistic regression using univariate and multivariate analysis methods has been applied in various areas including finance, marketing, and customer surveys [2].

However, no study has been conducted to identify fault-suspected processes and abnormal equipment

that affect the yield of manufacturing processes within the PCB industry.

Previously, major processes and equipment were randomly selected based on the knowledge of engineers, and key equipment parameters were assumed and managed. Nevertheless, many studies have been conducted on fault detection and classification (FDC) for monitoring the sensor data of major equipment and detecting faults. Cherry and Qin [3] analyzed 18 major measurement parameters by applying a multivariate analysis approach (principal component analysis) to the post-lithography measurement process. Yan [4] applied principal component analysis to the control data of the aluminum gate complementary metal oxide semiconductor, which has 36 parameters. Goodlin et al. [5] proposed a type-specific control chart that can immediately identify the cause of an abnormality based on the fault type when a fault is detected, regardless of the control chart. Although they applied this chart to the etching process with six fault types and 19 process parameters, they noted that more research is still required. Lee et al. [6] analyzed the important parameters affecting the quality and increased the yield by applying the FDC-CNN model to the chemical vapor deposition (CVD) process, which is a semiconductor process. They found 31 important parameters through the Wilcoxon rank sum test, clustering, and regression methods for three chamber datasets from the CVD process. Their results helped to improve the process yield by setting and managing a new standard limit line. Unlike the FDC approach, which monitors the equipment sensor data in real time to detect anomalies, this study identifies the processes and equipment that have the greatest effects on the final yield by analyzing the process data accumulated from all processes in the entire manufacturing process.

The purpose of this study is to reduce the fault rate and improve the yield by analyzing fault factors affecting the yield of the product, finding suspicious processes and equipment causing faults, and performing improvement activities. In other words, through equipment analysis conducted by process, suspected processes and equipment that affect faults are analyzed, and product faults are improved through intensive management of suspected equipment. To this end, the fault factors that influence the yield of the primary circuit process were first selected to identify the major processes and equipment. Next, fitness was measured using logistic regression [7] to analyze the equipment path (categorical data). Lastly, the effects of the selected major processes and equipment on the key fault factors were verified.

The rest of this paper is organized as follows. In Section 2, the target products and processes are selected; in Section 3, the analysis model for selecting the major processes and equipment for each fault factor is described; in Section 4, the major processes and equipment that affect the faults are analyzed using the proposed analysis model; in Section 5, the conclusions of this study are outlined, and potential future data measurements and analyses are discussed.

2. Analysis Target

2.1 Product and Process Selection

The yield of the PBGA manufacturing process is an important factor that affects the product quality and cost, and it is used to evaluate the manufacturing competitiveness. The yield is affected by numerous factors because different types of products are produced in one manufacturing process, and several hundred units of process equipment are used to manufacture the product. The product yield with a uniform center value can be maximized depending on the process control status (i.e., the process

capability index). The yield of each product is determined by the product design specifications, manufacturing process difficulty, and process and equipment parameters.

In this study, a product with a below-average yield was selected among products with large monthly production levels. The yield of the selected product was determined by analyzing the inspection results of the patterning process of the produced lots. Although yields are calculated and managed after each process, the post-process cost increases if a fault occurs in the first patterning process.

Further, it becomes difficult to identify and treat the cause of the fault. This suggests that yield management of the first patterning process is critical. Thus, the selected goal was to improve the yield of the first patterning process.

2.2 Key Fault Factor Analysis

After the patterning process, products with normal characteristics are selected through several hundred inspection items in the inspection process. If the yield is low, the number of faults for each product is recorded, and the key fault factors are selected. The key fault factors and defect rates that were determined by analyzing the fault factors of the produced lots of the selected product during the study are presented in Fig. 2. Six key fault factors that had the largest effects on the product yield, including Short (F2), Defect (F4), and Open (F6), were selected. The fault factors for which the fault rates were at least 1% were selected. The fault factors other than F1–F6 were excluded from this study because they accounted for less than 1% of the total faults.
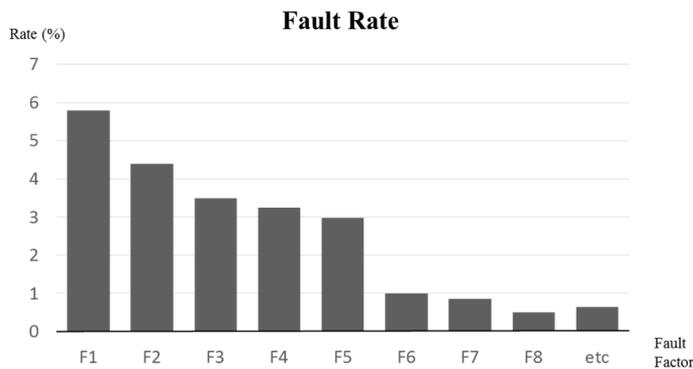

Fig. 2. Key fault factors.

In this study, the yield recorded in the inspection process after patterning and the fault factors that affected the yield in the first patterning process were analyzed. The fault-suspected processes and equipment that had the greatest effects on each fault factor were analyzed. Six fault factors were selected, as described previously. In addition, ten target processes that affected the first patterning were selected. For each process, two to five pieces of equipment that perform the same processes were deployed. As shown in Table 1, "p" and "f" denote process and equipment, respectively.

The tasks of the PBGA manufacturing process are performed in lot units. One lot consists of tens of panels, and each panel contains hundreds of chips. Each lot is processed along the equipment path. Several panels are selected from the lot after first patterning is completed, and their chips are inspected. Thus, when the inspection process is completed, the equipment path of the lot and the number of faulty

chips are determined for each fault factor. In this study, the total number of faults for each fault factor was calculated from the inspection quantity and converted to a defect rate. For the equipment path, a value of "1" was assigned to a lot that passed through a specific piece of equipment and "0" if not.

Table 1. Process and equipment status

Process and equipment status									
Process and processing equipment									
p10	**p20**	**p30**	**p40**	**p50**	**p60**	**p70**	**p80**	**p90**	**p100**
f11	f21	f31	f41	f51	f61	f71	f81	f91	f101
f12	f22	f32	f42	f52	f62	f72	f82	f92	f102
f13	-	f33	f43	f53	f63	f73	f83	f93	-
-	-	f34	f44	-	-	f74	-	-	-
-	-	f35	-	-	-	f75	-	-	-

p30 (Cu Plating), p50 (Exposure), p60 (Develop), p70 (Plating), p90 (Strip).

3. Logistic Regression

The linear regression model is applied when analyzing the correlations between one dependent variable and several independent variables. In general, a dependent variable in the linear regression model is assumed to be a factor that has consecutive values. However, if "good" or "faulty" is determined by an independent variable, or when the analysis involves a binomial event, where the dependent variable is either "0" or "1," the correlations between dependent and independent variables cannot be explained sufficiently by conventional linear regression analysis [8]. For such cases, the model must be restructured to explain the probability rather than the quantity of the dependent variable. In general, a regression analysis model can be expressed as a linear equation relating the independent variables $(x_1, x_2, \cdots, x_k)$ and the dependent variable (y), as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_k \tag{1}$$

However, to handle a binary dependent variable, such as the occurrence or nonoccurrence of a defect, y must have a probability value (P) between 0 and 1. In the linear regression model in Eq. (1), both the left and right sides can have the same range of (-∞, +∞). If P is substituted for y, the ranges of the dependent and independent variables may not coincide with each other. To solve this problem, a linear regression model with a uniform range can be created through a logit transformation using the logistic function in Eq. (2). Besides logit transformation, other models such as those by Gompertz and Probit may be used [9].

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_k)} \tag{2}$$

The logistic function in Eq. (2) is a linear equation based on the odds concept, which uses a ratio of the probability of occurrence to that of non-occurrence [10], as follows:

$$\ln\left(\frac{P}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_k \tag{3}$$

Odds is a relative measure of event occurrence, and the odds ratio is the degree of change in the odds of the dependent variable when the independent variable increases by one unit. In conventional linear

regression analysis, the coefficient β is estimated by minimizing the sum of squares of the residuals (i.e., the least squares method). However, in logistic regression, the coefficient is estimated using the maximum likelihood method that maximizes the probability of the occurring event [7].

In this study, logistic regression was used to achieve a high yield in the PBGA manufacturing process. Thus, the existence or absence of defects was used as the dependent variable. Considering the independent variable, the process equipment was selected and converted to a discrete variable that had the value "1" if the lot passed through the equipment and "0" otherwise. Next, the correlation between the dependent and independent variables was measured. In the logistic regression model of the binary variable, the defect probability (P) values larger than 0.5 were classified as group "1" (Fault) and those smaller than 0.5 as group "0" (Good). Here, the P value sets a classification reference value higher than the mean (0.5), so that the data can be classified sufficiently into group "1". Furthermore, the stepwise selection method based on the last estimated logistic regression model was used to discover the key equipment factors that affect the defects. The stepwise selection method successively adds factors that have large effects on the dependent variable. Whenever a new factor is added, an existing factor is deleted, or when a factor is deleted, a step-by-step review is conducted to determine whether the importance of an already deleted factor has increased and can be added back. In this experiment, the factor that had the largest chi-square value with the P value below 0.05 was selected first.

4. Case Study

4.1 Analysis of Fault-Suspected Process

In this study, the optimal model equation was statistically created according to the aforementioned model. The factors that had large chi-square values with P values lower than 0.05 were selected using a significance test for each independent variable. Here, the chi-square value indicates the degree of effect of an independent variable (process or equipment) on the dependent variable (defect rate).

Table 2 lists the processes that affect F2. As shown in Table 2, Exposure (p50), Develop (p60), and Strip (p90) processes were selected as the key processes that have large effects on F2 in descending order of the level of effect. In this study, the cumulative influence of each process was calculated, and three processes that accounted for approximately 70% of the total influence were selected (the P values of p80 and p100 were greater than 0.05).

Table 2. Analysis results of F2

Process	Number of equipment	χ^2	*P*-value	Impact (%)	Cumulative impact (%)
p50	**3**	**151.2**	**<0.0001**	**29.1**	**29.1**
p60	**3**	**103.8**	**<0.0001**	**20.0**	**49.1**
p90	**3**	**100.3**	**<0.0001**	**19.3**	**68.4**
p40	4	48.7	<0.0001	9.4	77.8
p30	5	37.9	<0.0001	7.3	85.1
p70	5	36.9	<0.0001	7.1	92.2
p10	3	34.8	<0.0001	6.7	98.9
p20	2	4.59	<0.0320	0.9	99.8

4.2 Fault-Suspected Process based on Fault Factor

The influence of each process was calculated by analyzing other fault factors using the afore-mentioned method, and the fault-suspected processes were selected in descending order of their effects. Thus, the processes that had large effects on each fault factor were extracted, as presented in Table 3.

The chi-square values of the factors were calculated, and the processes that had large chi-square values with P values lower than 0.05 were selected (three processes with 70% of the total influence were selected). The processes that had large effects on three or more fault factors were Chemical Cu Plating (p30), p60, Plating (p70), and p90. In addition, processes p10, p40, and p80 have no effect at all, and processes p20, p50, and p100 affect one or two factors. Therefore, it is expected that managing the four processes (p30, p60, p70, and p90) commonly affecting all the faults would be highly effective. Next, an analysis to identify faulty equipment was performed. This would determine the equipment that affected the processes influencing each fault factor the most.

Table 3. Suspected process analysis results

Process	Factors					
	F1	F3	F4	F5	F2	F6
p10						
p20	○3					
p30	●1	●3	●1			
p40						
p50					○1	
p60		●2		●2	●2	●3
p70	●2		●3	●1		●2
p80						
p90			●2	●3	●3	●1
p100		○1		○3		

●=processes that affected three or more fault factors in common, ○=processes that only affected specific fault types, 1,2,3=process (priority) with a large cumulative impact by fault type.

4.3 Analysis of Fault-Suspected Equipment

To identify the fault-suspected equipment, the processes that were performed by two or more pieces of equipment were analyzed. The processes were selected by using the stepwise selection method described in Section 4.2. The coefficient of each independent variable (equipment) in the logistic regression model was estimated using the maximum likelihood method. The equipment with a smaller estimated value has a smaller defect probability. In other words, when there are multiple pieces of equipment in the same process, the equipment with a small estimated value of the independent variable is considered as normal equipment, whereas that with a large estimated value is considered as abnormal equipment. In the case of F2, analyzed in Section 4.2, the results for the fault-suspected equipment are outlined in Table 4 (the reference equipment for each process is excluded in this table).

When the estimated coefficient of each independent variable (equipment) in Table 4 is substituted into the regression equation, Eq. (3), the following equation is obtained:

$$\log(\tfrac{P}{1-P}) = -4.43 + 0.113 \cdot f12 + 0.029 \cdot f21 - 0.245 \cdot f32 + 0.155 \cdot f34 + 0.311 \cdot f41 + 0.192 \cdot f42 -$$
$$0.119 \cdot f43 - 0.178 \cdot f51 - 0.090 \cdot f52 - 0.351 \cdot f61 + 0.124 \cdot f72 - 0.097 \cdot f73 - 0.084 \cdot f74 - \quad (4)$$
$$0.162 \cdot f92$$

Table 4. Analysis of suspected equipment (F2)

Analysis of maximum likelihood estimate						
Parameter	Equipment	DF	Estimate	Standard error	χ^2	$Pr > \chi^2$
Intercept	-	1	-4.430	0.429	106.61	<0.0001
p10	f11	1	-0.065	0.034	3.66	0.0557
	f12	1	0.113	0.022	25.94	<0.0001
p20	f21	1	0.029	0.013	4.59	0.0320
p30	f31	1	0.046	0.041	1.26	0.2603
	f32	1	-0.245	0.059	16.95	<0.0001
	f33	1	0.056	0.145	0.14	0.7009
	f34	1	0.155	0.048	10.42	0.0012
p40	f41	1	0.311	0.067	21.50	<0.0001
	f42	1	0.192	0.078	6.28	0.0121
	f43	1	-0.119	0.044	7.20	0.0073
p50	f51	1	-0.178	0.019	80.67	<0.0001
	f52	1	-0.090	0.030	8.98	0.0027
p60	f61	1	-0.351	0.034	100.40	<0.0001
	f62	1	0.029	0.024	1.53	0.2153
p70	f71	1	0.029	0.044	0.45	0.5018
	f72	1	0.124	0.030	16.98	<0.0001
	f73	1	-0.097	0.030	10.52	0.0012
	f74	1	-0.084	0.023	13.14	0.0003
p90	f91	1	0.017	0.086	0.03	0.8431
	f92	1	-0.162	0.047	11.84	0.0006

4.4 Verification

For the processes with large influence on each fault factor, normal equipment was distinguished from abnormal equipment based on the analysis results presented in Section 4.3. Specifically, the three processes with large cumulative influence that were selected in Section 4.2 were distinguished once more by normal and abnormal equipment. Table 5 outlines the results (based on Tables 3 and 4) for the identification of the fault-suspected process and equipment for each fault factor. Further, the pieces of equipment that had large influence were derived for each process. The processes that commonly had an influence on all the fault factors were p30, p60, p70, and p90. The normal equipment and abnormal equipment for each process are listed in Table 5. As shown in this table, in p30, equipment f32 was good for F1 and F4.

Hence, this equipment had no influence on F1 and F4. Similarly, in p70, equipment f72 had no influence on F4, F5, and F6. However, a few pieces of equipment showed conflicting results. For example, in the case of p90, equipment f91 had no influence on F4 and F6, but it influenced factor F5. This situation occurred because the fault factors had mutually incompatible characteristics. For example, when the connection thickness of the PBGA increased, the open circuit defects decreased, but the short circuit defects increased.

To identify the process that had the largest influence on the fault factors, the data for process factors (line width, plating thickness, etc.) managed in real manufacturing sites were analyzed. The results confirmed that the processes indeed influenced the corresponding fault factors. Moreover, in the plating

process, it was theoretically verified that the process factor value was changed, and the fault factors were influenced by the equipment condition. Further, while it is very difficult to distinguish normal equipment from abnormal equipment through direct experiments, additional verification by analyzing the parameters of each equipment is needed because process factors change and influence fault factors depending on the equipment condition and status.

Table 5. Suspected processes and equipment

Fault factors	Equipment	Suspected processes and equipment					
		p20	**p30**	**p50**	**p60**	**p70**	**p90**
F1	Normal	f22	f32			f75	
	Abnormal	f21	f33			f71	
F2	Normal			f51	f61		f92
	Abnormal			f53	f63		f93
F3	Normal		f35		f62		
	Abnormal		f31		f63		
F4	Normal		f32			f72	f91
	Abnormal		f34			f75	f93
F5	Normal				f63	f72	f93
	Abnormal				f61	f73	f91
F6	Normal				f63	f72	f91
	Abnormal				f61	f73	f92

5. Conclusion

This study was conducted to ensure manufacturing competitiveness by improving the yields and productivity of the PBGA manufacturing process. To achieve these improvements, the processes and equipment that affect the yield of the current widely distributed PBGA products were analyzed. To this end, the data on the defects and equipment variables in the PBGA manufacturing process were analyzed to identify the processes that affect the yield. This paper also proposes a technique to analyze the processes to determine the equipment that has the most adverse effects. The processes and equipment that are classified as critical factors need to be managed intensively with input from field engineers. In addition, further advanced studies on factor selection may be conducted, focusing on the analyses of equipment variables.

If additional key parameters (linear values) of the equipment are analyzed together, the use of mutual information feature selection may be considered because nonlinear factors are also present.

Research on the methods of collecting equipment data in the PCB production process and utilizing them for productivity improvement has been limited. Thus, new data analysis techniques that consider the actual process environment must be developed.

Acknowledgement

References

[1] D. H. Lee, J. K. Yang, C. H. Lee, and K. J. Kim, "A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data," *Journal of Manufacturing Systems*, vol. 52(Part A), pp. 146-156, 2019.

[2] G. Gasso, "Logistic Regression," 2019 [Online]. Available: https://moodle.insa-rouen.fr/pluginfile.php/7984/mod_resource/content/6/Parties_1_et_3_DM/RegLog_Eng.pdf.

[3] G. A. Cherry and S. J. Qin, "Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 2, pp. 159-172, 2006.

[4] L. Yan, "A PCA-based PCM data analyzing method for diagnosing process failures," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 4, pp. 404-410, 2006.

[5] B. E. Goodlin, D. S. Boning, H. H. Sawin and B. M. Wise, "simultaneous fault detection and classification for semiconductor manufacturing tools," *Journal of the Electrochemical Society,* vol. 150, no. 12, pp. 778-784, 2003.

[6] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135-142, 2017.

[7] J. Arkes, *Regression Analysis: A Practical Introduction*. Oxon, UK: Routledge, 2019

[8] M. Pal and P. Bharati, "Introduction to correlation and linear regression analysis," in *Applications of Regression Techniques*. Singapore: Springer, 2019, pp. 1-18.

[9] C. H. Cheon, *Data Mining Techniques*. Seoul, Korea: Hannarae Publishing, 2015.

[10] S. Lee, "Logistic regression procedure using penalized maximum likelihood estimation for differential item functioning," *Journal of Educational Measurement*, vol. 57, no. 3, pp. 443-457, 2020.

Hyun Sik Sim https://orcid.org/0000-0001-5031-2289

He received Ph.D. degree in Information & Industrial Engineering from Yonsei University in Seoul, Korea. He is now professor in the Department of Industrial & Management Engineering at Kyonggi University, Korea. He is also an Editorial committee of *Journal of the Semiconductor & Display Technology*. Dr. Sim worked as a group leader for Samsung Electronics Co.