creative commons

COMMONS DEED

저작자표시-비영리-변경금지 **2.0** 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

데이터사이언스석사 학위논문

Finding Optimal Equipment Combination using Semiconductor Metadata of FAB Process

반도체 FAB 공정 Meta 데이터를 활용한 최적 장비 조합 도출

2022년 8월

서울대학교 데이터사이언스대학원

정 지 원

Finding Optimal Equipment Combination using Semiconductor Metadata of FAB Process

반도체 FAB 공정 Meta 데이터를 활용한 최적 장비 조합 도출

지도 교수 김용대

이 논문을 데이터사이언스석사 학위논문으로 제출함

2022년 6월

서울대학교 데이터사이언스대학원

정 지 원

정지원의 석사 학위논문을 인준함

2022년 7월

위 원 장	차 상 균	(인)
부위원장	김 용 대	(인)
위 원	박 현 우	(인)
위 원	오 민 환	(인)

Abstract

Semiconductor manufacturing goes through hundreds of complex processes, and improving low yield is an essential task in the industry. In addition, the manufacturing equipment greatly influences the yield of products produced in the semiconductor manufacturing process. Therefore, predicting the yield through the combination of the equipment will help improve the yield by finding wafers that need improvement in advance. Moreover, considering the complex characteristics of the semiconductor process, we can find a combination of semiconductor equipment using models with good predictive performance, such as the deep neural network (DNN) model. However, using the DNN model creates a computationally difficult problem that requires exploring all combinations of variables, and the complexity of the model does not help much in analyzing the issue of the low-yield products and in what direction to improve. Therefore, in this paper, we propose a methodology to find optimal manufacturing equipment combination by applying metaANOVA, which allows us to interpret the complex prediction model by approximating an ANOVA model with multi-order interactions to the prediction model. In particular, we want to help identify the characteristics of each yield group by classifying the high and low yield groups based on the wafer test yield and exploring the combination of equipment representing each yield group.

Keywords: Semiconductor, Deep Learning, Machine Learning, Equipment Combination

Student ID: 2020-24952

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation

The semiconductor is a vital component of the high-tech industry in the 21st century. The importance of semiconductors has been highlighted as the Biden administration in the U.S. has scheduled to make massive investments in semiconductor manufacturing upon its launch. The semiconductor market's growth, which was slowed down for a while due to the U.S.-China trade war and the COVID-19 Pandemic, has gradually increased with the recovery of its previous market size. According to SEMI's announcement, the semiconductor market is expected to grow further in 2022 and achieve the largest ever. The importance of semiconductor manufacturing has become an essential factor in future national competitiveness, and studies related to semiconductor manufacturing have been continued.

In semiconductor manufacturing, yield is the key performance index which indicates the proportion of successful outputs from the inputs, and the production yield is highly affected by the production equipment. In actual industrial sites, equipment with low yield is identified during the wafer test stage. To solve this problem, engineers in the semiconductor industry temporarily reduce the production of the equipment to increase the yield in the short term, then search for issues with the equipment and fix them in the long term. To overcome this limitation, many studies suggest that defected wafers may be found in advance and converted into a high-yield product if the production equipment can predict the yield. As a result, maximizing production efficiency can be expected by reducing the time and cost required for semiconductor manufacturing. According to Baek et al. [6], applying traditional statistical methods and engineers' empirical analysis has a limitation in identifying root factors that cause yield degradations. This is due to the increasingly complicated process of semiconductor manufacturing that the process becomes physically small as they enter the nano-scale, making it more challenging to rely solely on the analysis of the domain experts. Therefore, the necessity of diagnosing degrading factors such as manufacturing equipment by intellectual methodology emerges as a pivotal improvement. Moreover, offering a data-driven solution to engineers can enhance productivity so that finding the best equipment for each process independently and combining equipment that is in interactive relation can have significant meaning to the semiconductor industry.

In general, it is known that machine learning methodologies such as boosting, random forest, and deep neural network (DNN) models have high prediction power. However, these models are usually complex and unsuitable when analyzing the problem of low yields products and how they should be improved. In other words, they are deficient in knowledge representation, making them known as black-box models. To solve this issue, Shin [1] proposes a method to provide the most similar case to the predicted query using a hybrid system that combines machine learning with memory-based reasoning (MBR). However, it has limitations in offering symbolic knowledge because it simply provides similar previous cases to the current query [1]. In addition, considering the complex characteristics of the semiconductor manufacturing process, the equipment combination can be found using models with relatively higher prediction performance, like the DNN model. However, using the DNN model for the semiconductor-related data will cause a substantial computation cost problem in exploring all equipment combinations. To overcome this drawback, we leverage a new approach to reduce the scope of navigation for finding equipment combinations and add interpretive power to the black-box model. And through this method, we can analyze the semiconductor equipment and offer data-driven solutions to the semiconductor domain experts.

1.2 Purpose

In the semiconductor manufacturing industry, the fabrication (FAB) process is a sequential step with multiple chemical and photolithographic process that turns a silicon wafer into integrated circuits. In general, the FAB process contains the following processing steps: lithography, etching, deposition, chemical mechanical planarization, oxidation, ion implantation, and diffusion. Each process step in the FAB process uses various equipment. In addition, interactions can exist between process steps, and finding those interactions can help improve productivity. Therefore, we expect to increase the wafer test (WT) yield, which is used after the FAB process as an index of productivity, by finding the interaction between the process steps and the optimal combination of equipment used in each process step in the FAB process.

As a goal of this paper, we intend to maximize the WT yield through a combination of equipment used in the processing steps in the FAB process. To be more specific, we want to identify the characteristics to be included in each tier divided into a high-yield group and a low-yield group based on the WT yield and explore the equipment combinations representing each tier. Finding the optimal combination of equipment that best represents each tier in semiconductor manufacturing can provide insight into improving yield through process equipment. In the case of searching for optimal equipment combinations for each high and low-yield group, all combinations of variables must be explored if the DNN model is used. Hence, we apply a novel method called metaANOVA, which approximates the complex predictive model like deep neural network to an ANOVA model with multi-order interactions. Meta ANOVA uses the characteristics that the ANOVA model is segmentable to significantly reduce the scope of navigation, enabling an efficient search for representative combinations. By comparing the predicted probability of belonging to each tier from the optimal equipment combination and the existing combinations, we expect to see whether it is possible to find the best equipment combination of the FAB process. Furthermore, we expect that finding the optimal combinations will help identify the characteristics that the semiconductor process should have in order to increase its productivity. Moreover, since the ANOVA model used in this study is a linear model, it is model-agnostic in that we can directly interpret the model through its coefficients. Therefore, essential factors in the model can be found based on the variance of the coefficients for each factor, and interpretation of the predicted values of each observed data is also available.

2. Related Works

Many prior studies have been conducted on semiconductor yield. Applying traditional statistical methods to semiconductor yield problems seems to be no longer effective due to semiconductor data's vast dimensions. Therefore, machine learning models have been used in the semiconductor manufacturing field in recent studies. Kim et al. [5] employs seven different machine learning models and three different dimensionality reduction methods to detect faulty wafers. The purpose of this study is to decrease manufacturing cost and increase lead time by finding faulty wafers in advance. Although statistical process control (SPC) method and virtual metrology (VM) have been applied for the detection, several limitations from these methods and the characteristics of the dataset bring the necessity of novel experimental approach. As one of the limitations, for example, SPC method uses each variable independently. But in fact, multiple variables have interactive influences each other. As an experimental setting, Kim et al. [5] applies three dimensionality reduction methods: stepwise linear regression, stepwise one class support vector machines (1-SVM), and principal component analysis

(PCA). In addition, the following machine learning based methods are used: gaussian density estimation, gaussian mixture model, parzen window, k-means clustering, 1-SVM, PCA, and KPCA. As a result, the study concludes that stepwise 1-SVM shows the best dimensionality reduction performance and 1-SVM has the best TPR-FPR result.

Jang et al. [8] propose a yield prediction method for new wafer maps by using deep learning algorithms on spatial features of semiconductor dies before the FAB process. This proposed study uses five spatial features of semiconductor dies, the wafer's central axis and four coordinates, as input variables for the DNN. As a result, this methodology improves the model's predictive power by learning yield data and helps design a new Wafer Map to improve productivity by 8.59%.

Jiang et al [4] introduces a method for predicting the final test (FT) yield at wafer fabrication (FAB) stage using machine learning techniques. Predicting FT yield at FAB stage can detect low yield wafers in advance which will eventually enhance the productivity of semiconductor. The study proposes a robust solution that uses all manufacturing related parameters including both numerical and categorical data. Jian et al [4] applies gaussian mixture models, one hot encoder and label encoder techniques as a preprocess step. Then model selection and model ensemble are adopted with using F1 macro method as a score metric. The study has an advantage in data-driven decision-making process by using automatic handling of manufacturing data because this novel framework can overcome the limitations of engineers' empirical analysis that manually reviews all production related data. The study compares the performance of seven different machine learning techniques with F1 macro (average) score: the machine learning techniques are support vector machine classifier, k-nearest neighbor, gaussian process classifier, logistic regression, extra tree classifier, gradient boost, and XGBoost model. In addition, three different pre-processing is adopted as an experimental condition which are label encoder, one hot encoder, and dropping categorical input. This generic framework selects top three models, extra tree classifier, gaussian process classifier, and XGBoost model, and carry out importance feature using feature importance analysis.

An et al. [2] propose a yield prediction model using SVM to address the limitations of neural network models that have excellent predictive power but are difficult to explain. Instead of indiscriminate testing in the two Probe Tests, which are required from FAB OUT to the final test, selectively classifying lots into high and low yield and dualizing the probe test conditions according to the above classification can reduce the time and cost required for manufacturing. After the first probe test is completed, this method uses a Stepwise SVM (SSVM) classifier model to distinguish between a high-yield lot and a low-yield lot. Then potential defective chips can be found after applying an adjusted second probe test to each yield lot. Therefore, the final test yield rate can be improved by finding potential defects in advance in the final test process. SSVM is a method of classifying data by adjusting parameters by expanding the classification interval step by step, which can improve predictive power over a traditional SVM model. As a result of this study, SSVM demonstrates superior performance in classification, showing higher classification accuracy and lower misclassification rates than traditional SVM, LDA, and QDA.

Shin [1] proposes a hybrid machine learning technique that combines a neural network model with memory-based reasoning (MBR), which uses the k-nearest neighbor (k-NN) method for case retrieval. The results of this hybrid system are implemented by a prediction query manager (PQM) method that simultaneously compares the results of the neural network and MBR when a new query is requested and returns the predicted value. This system has two appealing features that it can be applied to both classification and regression tasks without the mechanism conversion. Moreover, due to its on-line learning property, the system can monitor the process which is suitable for the semiconductor manufacturing process. However, there are some limitations to this method. Firstly, many previous cases are required in advance, so it is challenging to utilize the proposed system with few cases. In addition, the system has limitations in providing explicit symbolic knowledge because the system delivers information similar to the current query by offering previous cases.

3. Methods

3.1 Data

Semiconductor manufacturing involves a complex process that goes through hundreds of processes. In general, semiconductor manufacturing goes through the FAB, PKG, and Module processes. The wafer test (WT), package test, and module test are conducted accordingly at the end of each process to verify that each process is performed correctly, and defects or abnormal reactions occurring in each step degrade the yield/quality of the final products. Among the processes, the FAB process is the most complicated one which takes the longest time during the whole manufacturing process; therefore, the FAB process has the greatest influence on semiconductor yield [6]. The FAB process refers to the entire process of turning a raw silicon wafer into an integrated electronic circuit by forming several types of materials on the surface of a wafer and repeating the process of selectively removing a specific part using a mask already made. As shown in Figure 1, silicon wafers with non-conductor properties are made of semiconductors with electrical properties through the FAB process. Wafer test is the first stage to test the wafer from the FAB process, and it plays a role in detecting and repairing initial defects. In addition, WT yield, which represents the number of final products compared to the input wafers during the WT process, is a crucial evaluation index indicating the production efficiency of the product.

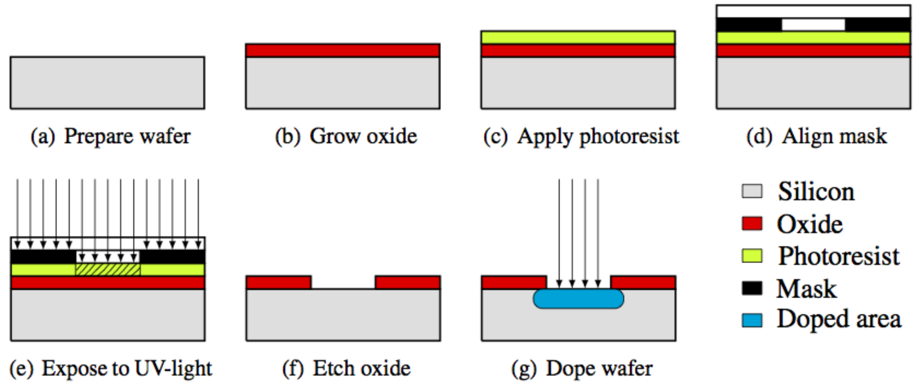


Figure 1. Basic steps of the FAB process in semiconductor manufacturing

For this study, SK Hynix provides the FAB related dataset which consists nearly 79,800 wafer observations. Moreover, the 3,700 FAB process-related can be divided into three types of variables. First, the information about each wafer is recorded with its lot id, wafer id, and the WT yield. The second is continuous variables with a cue time value representing the time cost in each processing step. The last is categorical variables which include meta information about the process such as equipment, recipe, reticle, etc. As shown in Figure 2, most categorical variables contain 10 to 40 unique categories. There are 3,192 lots and each lot in the dataset contains up to 25 wafers. Each wafer has a different WT yield value because of the different elements used to produce a wafer during the process. The most critical information in the FAB dataset is meta-information which consists of 2870 variables with categorical data of FAB process. Since the study focuses on equipment of processing steps in the FAB process, we extract variables of the equipment and WT yield. Hence, we get equipment combination dataset with their WT yields that is represented as a row in the dataset. Unfortunately, the FAB process-related dataset has been encoded for security reasons, so the exact names of the variables and values are provided unknown.

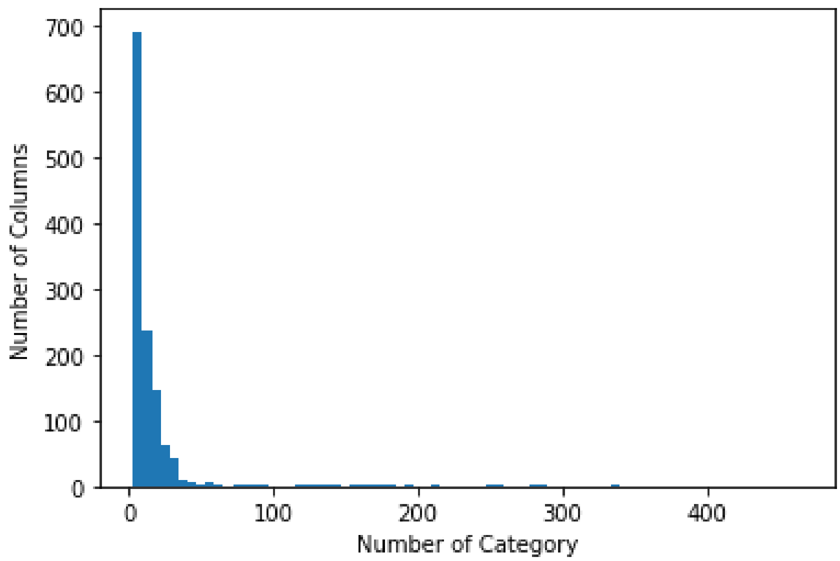


Figure 2. Distribution of unique category per each categorical variable of processing steps in the FAB process

The original FAB process-related dataset contains several missing values with some outliers. As a preprocessing step, we remove data whose yield value is less than 80 and variables with a single category. When handling the missing data, we realize that all data have missing values in at least one variable. Therefore, we first drop variables that contain 30% or more missing values and then remove row-wise data with missing values. As a result of preprocessing step, we get nearly 40,000 wafer observations with 240 variables which contain equipment information of each processing steps in the FAB process. In other words, we could extract equipment combinations used in the FAB process. In consideration of the characteristics of the semiconductor process, the WT yield has an imbalanced distribution structure. However, the imbalanced structure of the dataset is solved since the WT yields are classified into a high yield group and a low yield group according to the rank value after removing the outlier.

3.2 Experiment Framework

Previous studies use various machine learning models for predicting semiconductor yields. Considering the complex manufacturing process of semiconductors, using machine learning models with good prediction performance, such as the DNN, is essential. However, a high computation cost is required to navigate all possible combinations of semiconductor equipment variables if we use the DNN model. Moreover, it is difficult to interpret or explain the model due to its complexity. Gunning [3] presents the relationship between the prediction accuracy and explainability of machine learning models through Figure 3. We can observe a trade-off that the explainability of the model decreases as the prediction accuracy increases and vice versa.

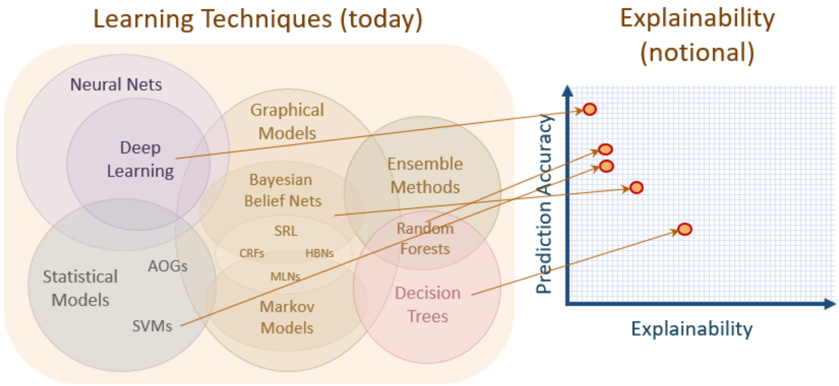


Figure 3. Relation between performance and explainability

In this paper, we try to solve the computation issue and interpretability problem by training an ANOVA model with multi-order interactions and using it to find a combination of the FAB process equipment representing each tier. The experimental framework is presented in Figure 4. First, the data are divided into two tiers based on the WT yield, the high yield group and the low yield group. Next, we use the DNN model as a classification model for each tier group. As mentioned above, however, the computation cost of the model will dramatically increase if we use the DNN model as it is to explore combinations of all variables. Thus, we obtain the equipment combination with the highest probability of belonging to each tier by approximating the DNN model to the ANOVA model. The process of training ANOVA model that approximates the target model is conducted by applying metaANOVA algorithm. In this study, the probability that data belongs to each tier is used as a performance index because the higher the probability indicates a better expression of each tier's characteristic. Hence, the optimal equipment combinations for each tier are re-used as the classification model's inputs, and each optimal combination's performance is evaluated by comparing the maximum probability that the original FAB data belongs to each tier group.

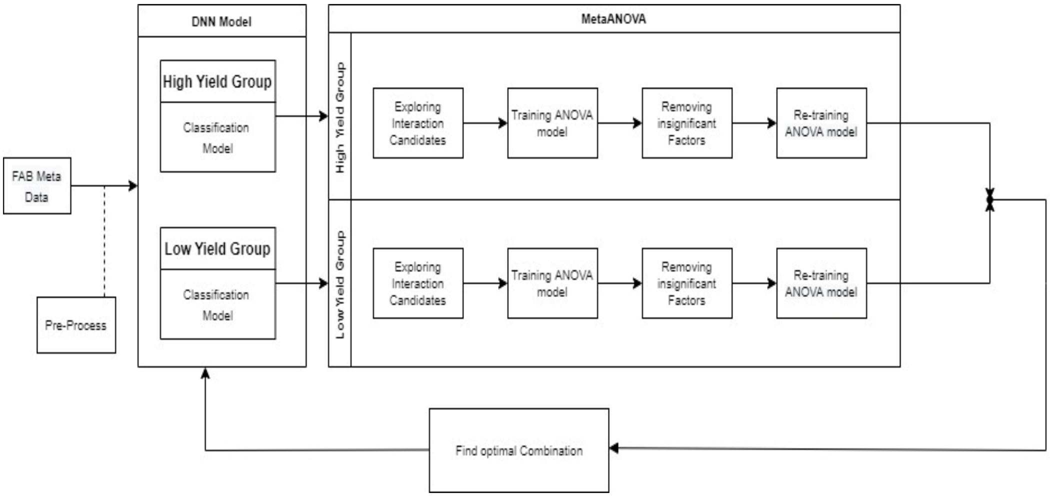


Figure 4. Framework of proposed method

3.3 MetaANOVA

In general, the machine learning model is called a black-box model because the model becomes more complex and harder to interpret as the predictivity of the model increases. MetaANOVA is a methodology that allows a straightforward interpretation of black-box models by approximating a complex model to an ANOVA model with multiple-order interactions, i.e., creating an ANOVA model that predicts the output value of a given model [9]. Therefore, meta ANOVA involves finding interactions within a complex black-box model and learning the ANOVA model based on the interactions. At this time, an unnecessarily large number of coefficients are used when all interactions are included; thus, we eliminate insignificant interactions in meta ANOVA. Given the predictive model to be approximated, metaANOVA makes the ANOVA model through the following three steps [9].

1. Explore candidate interactions
2. Learn the ANOVA model with main effects and explored interactions
3. Eliminate non-significant factors from learned ANOVA models

The most significant step of metaANOVA is exploring candidate interactions, and the used equation is as follows. The following equations related to metaANOVA are introduced by Kim et al. [9]. Let f be the function to be approximated is and $x \in \mathbb{R}^p$ is the input vector. Suppose $[p] = \{1, \ldots, p\}$, then $k \in [p]$ and $J_k = \{j \subset [p]: |j| = k$, and with given $j \subset [p]$, $x_j!$ means $\prod_{j \in j} x_j$ and the input variable x_j is either 0 or 1 [9]. Knowing that the input variable contains binary value, we assume that there is an ANOVA model g that $f(x) = g(x)$ where f is a black box model to be approximated. Therefore, to screen interactions, metaANOVA considers the Equation (1), which is an ANOVA model that includes all interactions.

$$g(x) = \beta_0 + \sum_{j \subset [p]} \beta_j x_j! \tag{1}$$

If $\beta_j \neq 0$, interaction j is valid for a given $g(x)$, so we have to find j that satisfies $\beta_j = 0$ [9]. With the given j, the above equation can be summarized as the Equation (2).

$$g(x) = \beta_0 + \sum_{j' \subset j} x_{j'}! \left\{ \beta_{j'} + \sum_{j_2 \subset j^c} \beta_{j' \cup j_2} x_{j_2}! \right\} + \sum_{j_3 \subset j^c} \beta_{j_3} x_{j_3}! \tag{2}$$

In addition, if we summarize $g_{j',\ j}(x_{j^c}) = \beta_{j'} + \sum_{j_2 \subset j^c} \beta_{j' \cup j_2} x_{j_2}!$, we can write the expression as the Equation (3).

$$g(x) = \beta_0 + \sum_{j' \subset j} x_{j'}!\ g_{j',j}(x_{j^c}) + \sum_{j_3 \subset j^c} \beta_{j_3} x_{j_3}! \tag{3}$$

Using the above equation, Kim et al. [9] introduce two theorems:

Theorem 1 *For a given* j*,* $\beta_{j'} = 0$ *for all* $j' > j$ *if and only if* $g_{j,j}(x_{j^c})$ *is a constant function for all* x_{j^c}*, where* $j' > j$ *means* $j' \supset j$ *but* $j' \neq j$*.*

Theorem 2 *For any* j*,* $g_{j,j}(x_{j^c})$ *can be represented as follows. Note that the function* $f(x)$ *is the result of* f *at* $x_{j'} = 1$ *and* $x_{j-j'} = 0$*.*

$$g_{j,j}(x_{j^c}) = \sum_{j' \subseteq j} (-1)^{|j-j'|}\ f(x\colon x_{j'} = 1, x_{j-j'} = 0) \tag{4}$$

With these theorems, we can efficiently get the interaction set S that is $\bigcup_{k=1}^{K} S_k$ by eliminating unnecessary high-dimensional interactions. Then, the ANOVA model can be trained by using the set of main effects and the searched interactions S. For a function to be approximated $f(x)$, we can train ANOVA model by minimizing the following the Equation (5).

$$\sum_{i=1}^{N} \left(f(x_i) - f^{a}(x_i;\ \beta)\right)^2 \quad (5)$$
$$subject\ to\ sum - to - zero\ condition$$

where ANOVA model including the interaction can be summarized as the Equation (6).

$$f^{a}(x;\ \beta) = \beta_0 + \sum_{j=(j_1,\dots,j_k)\in S} \sum_{a_{j1}\in X_{j1}} \cdots \sum_{a_{jk}\in X_{jk}} \beta_{a_{j1},\dots,a_{jk}} I(x_{j1} = a_{j1}, \dots, x_{jk} = a_{jk}) \quad (6)$$

3.4 Experiment Setting

Label Smoothing

This experiment divides the FAB process data into two tiers based on the WT yield and then finds the most suitable equipment combination for each tier group. Therefore, WT yield should be converted to the probability that the data belongs to each tier. As a first step, whether the data belongs to each tier is labeled as a binary value in additional variables: the high-yield and low-yield groups. Since the dependent variable is a probability, the tier variables should have a value between 0 and 1 according to the weight of the data instead of the binary value. Hence, we apply label smoothing to weigh the dependent variable with a probability value. In addition, label smoothing gives the normalization effect by converting hard targets to soft targets so that we can expect performance improvement in the model. Let K be the number of categories, y_k be a hard target composed of 0 and 1, and ϵ be given a hyperparameter; then, we can get the soft target y_k' by using the Equation (7).

$$y_k' = (1 - \epsilon)\, y_k + \frac{\epsilon}{K} \tag{7}$$

By dividing each of the two converted variables by tier through label smoothing above, we can have two datasets that contain the probability of belonging to each tier as a dependent variable.

Exploring interaction candidates

The most important part of the study is metaANOVA algorithm. To train ANOVA model, we first find interactions and explore candidate interactions sequentially according to order in the following way.

1. Given the searched k order interactions, calculate a score indicating the likelihood that there will be more than $k+1$ order interactions involving each interaction ($k=1$ means the main effect)

2. The k order interaction with a low score determines that there is no interaction more than the ($k+1$) difference involving itself, so eliminate all corresponding interactions

We consider using second-order interactions in this study. Figure 5 shows the score for the main effects of the high-yield group in high order, and the score represents the importance of factors in each equipment variable. In the case of the high-yield group, we can see that the graph becomes gentle when the number of the main effect is 30 in Figure 5. Therefore, interactions that include the main effect below 30 can be removed. Similarly, Figure 6 represents the sorted score for the main effects of the low-yield group, and the interactions including the corresponding main effects are deleted below 30.

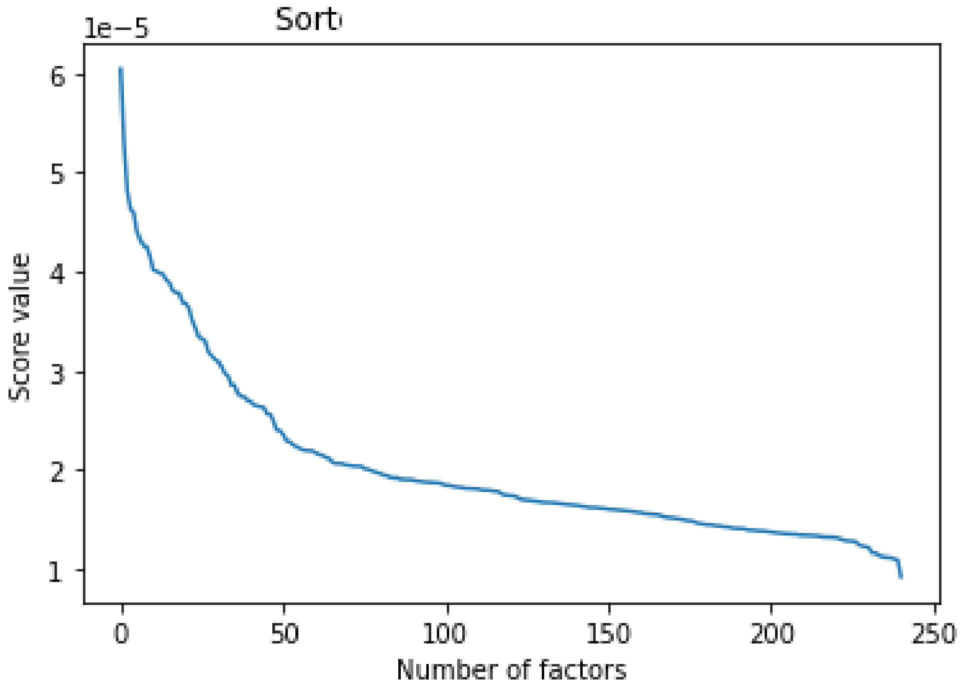


Figure 5. Score for the main effects of high-yield group in high order

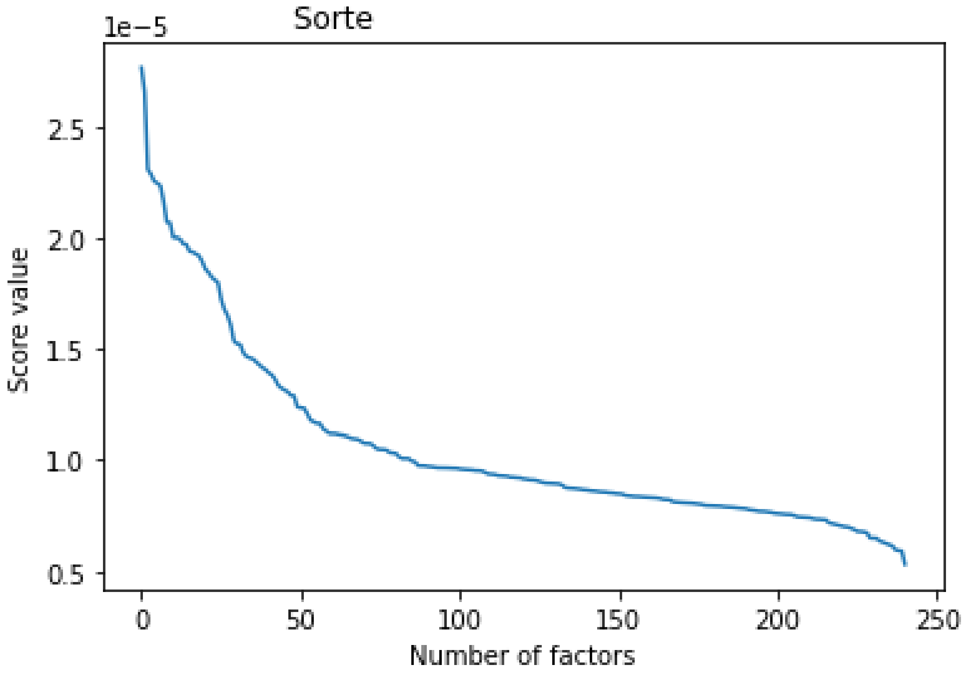


Figure 6. Score for the main effects of low-yield group in high order

Learning ANOVA model and eliminating non-significant factors

We train the ANOVA model based on the candidate interactions explored above as a next step. Therefore, we learn the linear model that predicts the predicted value of the target function, and we limit the sum of all the coefficients of each factor to zero by applying the sum-to-zero constraint.

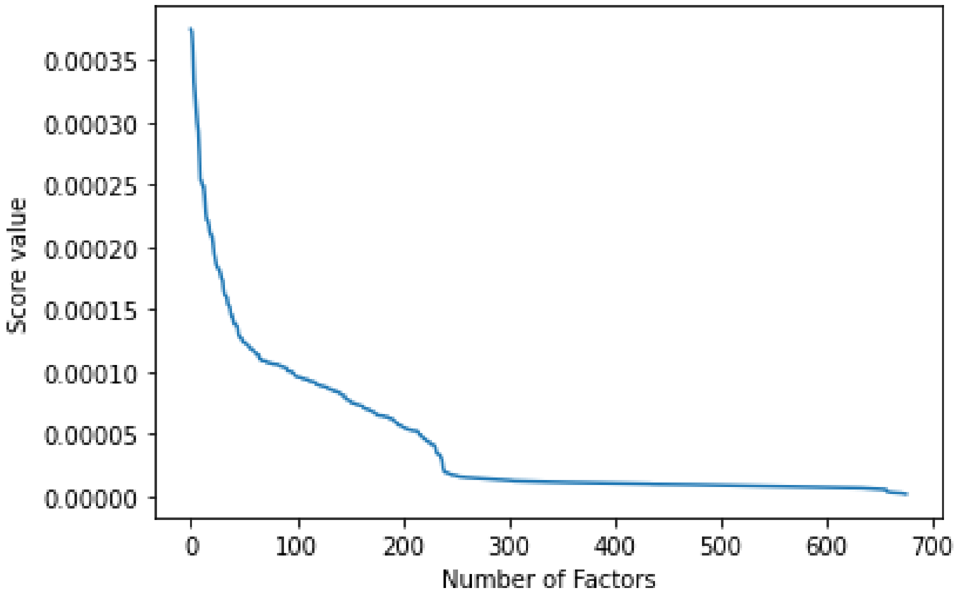


Figure 7. Importance of main effects and Interactions in high order for high-yield group

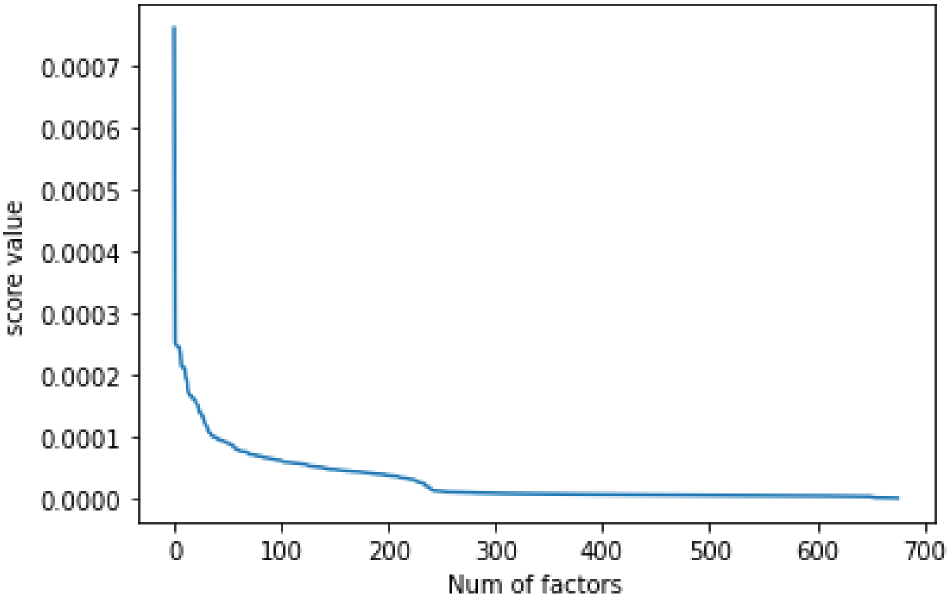


Figure 8. Importance of main effects and Interactions in high order for low-yield group

If the predicted value from the fitted ANOVA model changes substantially depending on a factor, the factor can be considered significant. This is equivalent to the large variance of the coefficients for the factor. Therefore, the variance of the coefficient can be considered as the importance of the factor, and insignificant factors can be removed based on their importance. Figure 7 shows the importance of the high yield tier model. According to Figure 7, factors below the 300-th rank can be considered insignificant and removed so that 241 main effects and 59 interactions remain.

An ANOVA model is learned in the same way for the low yield tier corresponding to the result shown in Figure 8. Using the fitted ANOVA model, the mean squared error (MSE) of the predicted logit outcome of the DNN model can be seen in Table 1.

Table 1. Variance of DNN logit & MSE of ANOVA

Yield Group	Variance (DNN logit)	MSE of ANOVA model
High Yield	0.5687	0.0017
Low Yield	0.4867	0.0038

Using the variance and MSE in Table 1, we can compute the coefficient of determination as shown in the equation below. Since we have relatively high variance and low MSE, the value of the coefficient of determination increases, indicating that the independent variables have great explanatory power for the dependent variable.

$$R^2 = 1 - \frac{SSE}{SST} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \mu_y\right)^2} = 1 - \frac{MSE}{Var(y)} \tag{8}$$

Finding representative combinations by tier

We can find optimal equipment combinations through ANOVA model that using the characteristics of the model can make the problem more explicit. Consider a network with variables as nodes and interactions as edges. If we find groups of variables in this network, there is no interaction between the groups. Therefore, we can arrange the ANOVA model according to the group of variables to form a group-by-group function. We can then switch to the problem of finding the combination that maximizes the functions of each group. In Figure 9, for example, variables numbered from 1 to 15 are marked as nodes, and edges connect nodes with interactions. Note that variables numbered 4 to 7 have interactions and are grouped into one cluster. Moreover, we can consider independent nodes like node 12 as a main effect variable. To sum up, we can search for the optimal equipment combination by finding a combination that maximizes the functions of the group.

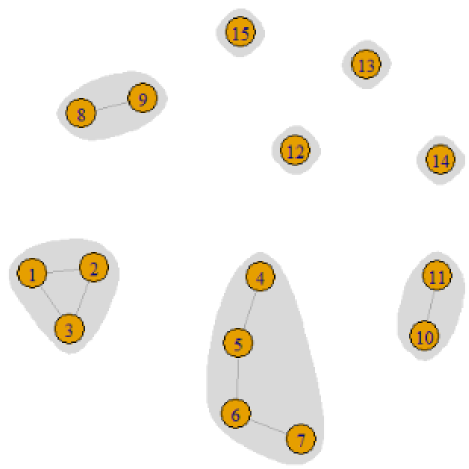


Figure 9. Example of grouping variables

We can select equipment variables with maximum coefficients for main effect variables. For the group of variables that include interactions, we should get predictions for all combinations implementing the network method mentioned above. However, calculating all combinations in this study is difficult due to the vast number of variables in a group. So multiple combinations are randomly generated to find the combination with the highest predictions. In this experiment, the variables with second-order interactions formed one large group. We select the combination with the highest predicted value among 50,000 randomly generated combinations in the existing data.

After finding the optimal semiconductor equipment combination, we use the DNN classification model, which the ANOVA model approximates, to evaluate the performance of the equipment combination. By inserting optimal equipment combination data as input into each high-yield group and low-yield group's DNN classification model, we can obtain the probability values belonging to each yield group. Moreover, we evaluate the performance of the optimal equipment combination by comparing with the results of the existing equipment combinations in the data.

4. Results

Using importance, we can select variables to be observed with particular attention. Since each variable represents the FAB process, we can consider the importance as a score for each process. The following Table 2 and Table 3 show the three most important main effects and interactions in the high and low-yield model with their importance. Unfortunately, variables are encoded for security reasons, so the exact name of the process cannot be known.

Table 2. Top three main effects and its importance score

Yield Group	Process Variable	Importance
High Yield	I2061000E_lhi_tulh_daj_ln	0. 000375
	H9047000U_lhi_tulh_daj_ln	0.000372
	J9031000Q_lhi_tulh_daj_ln	0.000358
Low Yield	I4031100X_lhi_tulh_daj_ln	0.000760
	W1032000Q_lhi_tulh_daj_ln	0.000253
	I4156000M_lhi_tulh_daj_ln	0.000249

Table 3. Top 3 interactions and its importance score

Yield Group	Process Variable	Importance
High Yield	(W1032000Q_lhi_tulh_daj_ln, I7095000E_lhi_tulh_daj_ln)	1.9087e-05
	(W1032000Q_lhi_tulh_daj_ln, I5010200A_lhi_tulh_daj_ln)	1.8476e-05
	(I7095000E_lhi_tulh_daj_ln, I7046200M_lhi_tulh_daj_ln)	1.8247e-05
Low Yield	(W1032000Q_lhi_tulh_daj_ln, J1034000M_lhi_tulh_daj_ln)	1.4995e-05
	(W1032000Q_lhi_tulh_daj_ln, I7095000E_lhi_tulh_daj_ln)	1.4433e-05
	(I7095000E_lhi_tulh_daj_ln, J1034000M_lhi_tulh_daj_ln)	1.3732e-05

We can explain the impact of each equipment on having the maximum predicted value by using the coefficient of the ANOVA model. Table 4 and 5 below show the main effects and interactions that have the greatest influence on predicting both high and low-yield group's optimal equipment combinations. For example, we can interpret that equipment named 4QTLQ302 in "M1051600B_lhi_tulh_daj_ln" process contributes the most to have the maximum predicted value.

Table 4. Top 3 main effects in optimal combination

Yield Group	**Process Variable**	**Equipment**	**Coefficient**
High Yield	Q1030000B_lhi_tulh_dah_ln	4ARA4703	0.1552
	L9511300B_lhi_tulh_daj_ln	EFP404	0.1179
	I4166100M_lhi_tulh_daj_ln	4QMQ0306	0.1139
Low Yield	L9515100U_lhi_tulh_daj_ln	4QLPQ702	0.0863
	J8020200E_lhi_tulh_daj_ln	4QTLM921	0.0845
	I4157600M_lhi_tulh_daj_ln	4JLQF507	0.0802

Table 5. Top 3 interactions in optimal combination

Yield Group	Process Variable	Equipment	Coefficient
High Yield	(W1032000Q_lhi_tulh_daj_ln, I2031000B_lhi_tulh_daj_ln)	(EFLZ02, 4ARA4905)	0.0874
	(H9047000U_lhi_tulh_daj_ln, I5010200A_lhi_tulh_daj_ln)	(4LDU2211, 4ARA5006)	0.0769
	(M1056200Q_lhi_tulh_daj_ln, H9047000U_lhi_tulh_daj_ln)	(4QTLM104, 4LDU2211)	0.0755
Low Yield	(X3065400M_lhi_tulh_daj_ln , W1032000Q_lhi_tulh_daj_ln)	(4ESW1005 , 4ELLB402)	0.0481
	(W1032000Q_lhi_tulh_daj_ln , I5010200A_lhi_tulh_daj_ln)	(4ELLB402 , 4ARA5108)	0.0463
	(M1023000E_lhi_tulh_daj_ln , W1032000Q_lhi_tulh_daj_ln)	(4QMS0704 , 4ELLB402)	0.0452

Table 6 summarizes the predicted probability of original and optimal combinations in both high and low-yield groups. More specifically, it shows the highest predicted probability that each original combination from training data and optimal equipment combination belongs to both high and low yield groups. As shown in Table 6 and Figure 10, the optimal equipment combination contains a higher probability value in all yield groups, which are increased by 14% and 21%, respectively. Therefore, we can conclude that metaANOVA successfully finds the critical equipment in each process and interactions that have a crucial impact on the WT yield.

Table 6. Predicted probability of training data and optimal equipment combination

Yield Group	Original Combination	Optimal Combination
High Yield	0.8504	0.9906
Low Yield	0.7213	0.9261

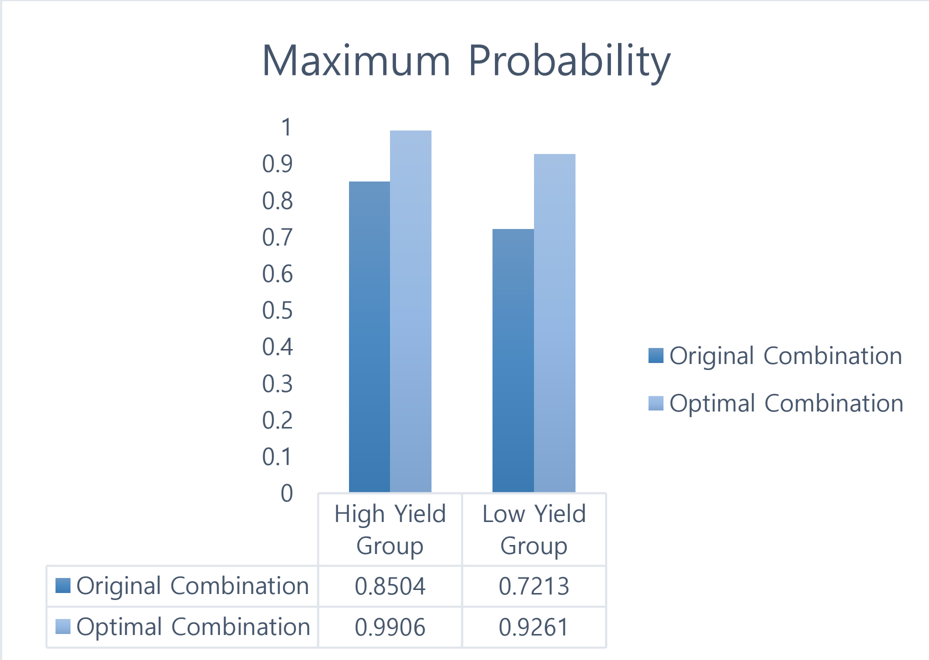


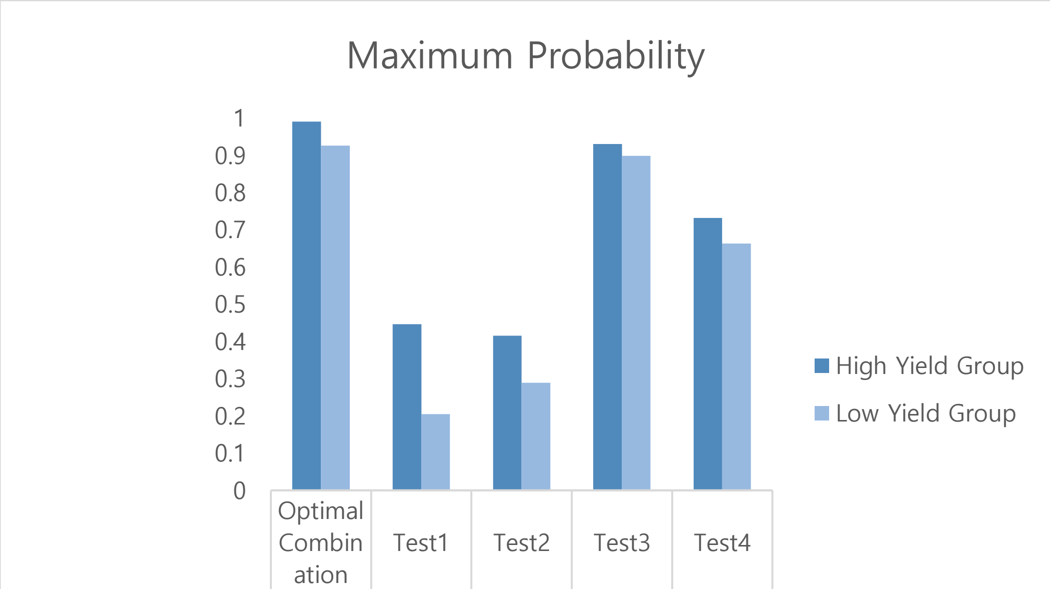
Figure 10. Comparison of probability belonging to each high and low yield group between training data and optimal equipment combination

To verify the results above, we conducted additional experiments, and the result summary is represented in Table 7. Firstly, we ignore both main effects and interactions by applying the same constant value to all process variables. By doing so, we can see that the predicted probabilities decrease dramatically. In a second experiment, interaction variables are only used that we insert the same constant value to the main effect variables. Moreover, the main effects are solely used by using the same constant value to interaction variables in the third experiment. As shown in Table 7, we realize that main effect variables can be considered more significant than those of interaction. This is due to the differences between the predicted probability from the second and third experiments which are approximately 57% and 63% in the high and low yield groups, respectively. Lastly, we generate 50,000 random equipment combinations by sampling in a uniform distribution in all process variables.

Table 7. Comparison of maximum predicted probability that each equipment combination from additional experiments belongs to both yield group.

Experiment	High Yield Group	Low Yield Group
#1. Not using main effects and interactions	0.4466	0.2046
#2. Using interactions only	0.4152	0.2893
#3. Using main effects only	0.9303	0.8983
#4. Random combinations	0.7315	0.6634

Figure 11 shows the results of the highest predicted probability of the equipment combinations from the additional tests mentioned above. In addition, we compare the predicted probabilities from the additional tests to the one from the optimal equipment combination which is found by applying the metaANOVA method. As shown in Fig. 10, we can validate that the optimal equipment combinations best represent both the high and low yield groups because they have the highest predicted probabilities.



	Optimal Combination	Test1	Test2	Test3	Test4
High Yield Group	0.9906	0.4466	0.4152	0.9303	0.7315
Low Yield Group	0.9261	0.2046	0.2893	0.8983	0.6634

Figure 11. Comparison of probability belonging to each high and low yield group among the optimal equipment combination and four different tests.

5. Discussion

Semiconductor manufacturing is complicated because hundreds of processes and various equipment are used. In addition, the yield indicating the productivity of the manufactured product is greatly influenced by the combination of equipment used. Therefore, while solid predictive performance can be expected with the DNN model, exploring all equipment combinations creates a problem because the computational volume becomes significantly large. Furthermore, studies have shown that models with more substantial predictive power, such as the DNN model, lack interpretability. In this study, we intend to enhance the WT yield of the semiconductor FAB process by improving the equipment combination. To compensate for the mentioned shortcomings, we propose to create an ANOVA model with multi-interactions by applying the metaANOVA algorithm. We can find representative combinations efficiently by significantly reducing the scope of the search using the divisible nature of the model. In addition, direct interpretation of the model through coefficients is possible because the ANOVA model is linear. Then we find an optimal equipment combination in the FAB process with the highest predicted probability of corresponding to each high and low yield group.

The proposed method firstly has a process of training a DNN classification model with pre-processed data, and then we train an ANOVA model that approximates the DNN model by applying the metaANOVA algorithm. In short, MetaANOVA trains the ANOVA model through three steps: exploring candidate interactions, training the ANOVA model including main effects and explored interactions, and removing non-significant factors from the trained ANOVA model. As a result, the optimal equipment combinations, including main effects and interactions in the ANOVA model for both high and low yield groups, are compared with the original equipment combinations of the data. Moreover, we get the predicted probability by inserting these equipment combinations into the DNN classification model. As a result, we confirm that the optimal equipment combination in each yield group has a higher probability of belonging to the yield group than the original equipment combinations from the data. We expect that this method, which identifies essential process factors through the interpretable ANOVA model and finds the optimal equipment combination, will be helpful in various processes in the semiconductor manufacturing field. Still, we can improve the performance of this study by applying higher interaction order that we only use second-order interaction in this study due to the lack of computation power. Without the limitation, we can find the proper interaction order that fits the given data, eventually finding a more optimal combination of semiconductor FAB process equipment.

Bibliography

[1] C. K. Shin and S. C. Park, “A machine learning approach to yield management in semiconductor manufacturing,” *International Journal of Production Research*, vol. 38, no. 17, pp. 4261–4271, 2000.

[2] D. An, H.-H. Ko, T. Gulambar, J. Kim, J.-G. Baek, and S.-S. Kim, “A semiconductor yields prediction using stepwise support vector machine,” *2009 IEEE International Symposium on Assembly and Manufacturing*, 2009.

[3] D. Gunning and D. Aha, “DARPA’s Explainable Artificial Intelligence (XAI) program,” *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[4] D. Jiang, W. Lin, and N. Raghavan, “A novel framework for semiconductor manufacturing final test yield classification using Machine Learning Techniques,” *IEEE Access*, vol. 8, pp. 197885–197895, 2020.

[5] D. Kim, P. Kang, S. Cho, H.-joo Lee, and S. Doh, “Machine learning-based novelty detection for faulty wafer detection in semiconductor

manufacturing," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4075–4083, 2012.

[6] D.-H. Baek, I.-J. Jeong, and C. H. Han, "Application of data mining for improving yield in wafer fabrication system," *Computational Science and Its Applications – ICCSA 2005*, pp. 222–231, 2005.

[7] M. Rovitto, "Electromigration reliability issue in interconnects for three-dimensional integration technologies," Doctoral. dissertation, Technische Universität Wien, Vienna, Austria, 2016.

[8] S.-J. Jang, J.-S. Kim, T.-W. Kim, H.-J. Lee, and S. Ko, "A wafer map yield prediction based on machine learning for Productivity Enhancement," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 4, pp. 400–407, 2019.

[9] Y. Kim, D. Kim, and Y. Choi, "MetaANOVA: an efficient algorithm for approximating a discriminative model to obtain interpretability," working paper, 2021.

국문 초록

반도체 제조는 수백 가지의 복잡한 과정을 거치며, 낮은 수율을 개선하는 것이 중요한 과제라고 할 수 있다. 또한 반도체 제조 공정에서 생산되는 제품의 수율은 제조 장비에 의해 큰 영향을 받기 때문에, 장비 조합을 통해 수율을 예측할 수 있다면 개선이 필요한 웨이퍼를 사전에 발견하여 수율을 개선하는 데 도움을 줄 것이다. 그리고 반도체 공정의 복잡한 특성을 고려하면 딥러닝과 같이 예측 성능이 좋은 모델을 사용하여 장비 조합을 찾을 수 있을 것이다. 하지만 딥러닝 모델을 그대로 사용한다면 모든 변수의 조합을 탐색해야 하는 계산적으로 매우 어려운 문제가 발생하며, 모델의 복잡성 때문에 탐지된 저수율 제품의 문제가 무엇인지 그리고 어떠한 방향으로 개선해야 하는지 분석할 때 큰 도움이 되지 않는다. 따라서 본 연구에서는 복잡한 예측 모델을 다차 교호작용을 포함하는 ANOVA 모델로 근사시켜 간단하게 해석할 수 있도록 하는 방법론인 metaANOVA를 적용하여 예측 모형을 해석하는 방안을 제시한다. 특히, wafer test 수율을 기준으로 고수율 군과 저수율 군을 분류하고 각 수율 군을 대표하는 장비 조합을 탐색하여, 각 수율 군에 포함되기 위해 가져야 할 특징을 파악하는 데 도움을 주고자 한다.

주요어: 반도체, 머신러닝, 딥러닝, 장비 조합

학번: 2020 - 24952