Golden Path Search Algorithm for the KSA Scheme

Ching-Kang Ing, Chin-Yi Lin, Po-Hsiang Peng, Yu-Ming Hsieh, and Fan-Tien Cheng, *Life Fellow, IEEE*

***Abstract*— The concepts of Industry 4.1 for achieving Zero-Defect (ZD) manufacturing were disclosed in *IEEE Robotics and Automation Letters* in January 2016. ZD of all the deliverables can be achieved by discarding the defective products via a real-time and online total inspection technology, such as Automatic Virtual Metrology (AVM). Further, the Key-variable Search Algorithm (KSA) of the Intelligent Yield Management (IYM) system developed by our research team can be utilized to find out the root causes of the defects for continuous improvement on those defective products. As such, nearly ZD of all products may be achieved. However, in a multistage manufacturing process (MMP) environment, a workpiece may randomly pass through one of the manufacturing devices with the same function in each stage. Different devices of the same type perform differently in each stage, where the performances will be accumulated through the designated manufacturing process and affect the final yield. KSA can only identify the influence of univariate variables (i.e., single devices) on the yield, yet it cannot detect the manufacturing paths that have significant influence on the yield. In order to cope with this deficiency such that the golden path with a better yield amongst all the MMP paths can be found, this research proposes the Golden Path Search Algorithm (GPSA), which can plan golden paths with high yield under the condition of the number of variables being much larger than that of samples. As a result, it makes the improvement of manufacturing yield be more comprehensive.**

***Note to Practitioners*—Traditional scheduling only considers the capacity of the manufacturing devices for allocation; while, the impact on the yield is rarely considered. In fact, in a production process, the production deviations will be gradually accumulated and affect the product quality along with the processing influence of each device. Therefore, the purpose of this paper is to propose the GPSA scheme to quickly search for high-yield manufacturing paths before the production. Manufacturers can then configure production devices based on these paths. According to the experimental results of real manufacturers' data, GPSA can not only quickly nail down the high-yield paths from a large amount of historical data, but also alert the users to avoid paths that are prone to defective rates for their reference.**

Manuscript received August 9, 2021; revised October 20, 2021; accepted November 7, 2021. This article was recommended for publication by Associate Editor C.-Y. Lee and Editor J. Li upon evaluation of the reviewers' comments. This work was supported in part by the Intelligent Manufacturing Research Center (iMRC) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan; and in part by the Ministry of Science and Technology of Taiwan under Contract MOST 109-2218-E-006-007, Contract MOST 108-2221-E-006-210-MY3, and Contract MOST 110-2218-E-006-027. *(Corresponding author: Fan-Tien Cheng.)*

Ching-Kang Ing and Po-Hsiang Peng are with the Institute of Statistics, National Tsing Hua University, Hsinchu 300, Taiwan (e-mail: cking@stat.nthu.edu.tw).

Chin-Yi Lin, Yu-Ming Hsieh, and Fan-Tien Cheng are with the Institute of Manufacturing Information and Systems, National Cheng Kung University (NCKU), Tainan 701, Taiwan (e-mail: chengft@mail.ncku.edu.tw).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2021.3129528.

Digital Object Identifier 10.1109/TASE.2021.3129528

***Index Terms*— Zero defects (ZD), key-variable search algorithm (KSA), intelligent yield management (IYM), golden path search algorithm (GPSA), multistage manufacturing process (MMP), intelligent and flexible manufacturing, semiconductor manufacturing.**

I. Introduction

ZERO Defects (ZD) has been one of the quality-improvement objectives for enhancing manufacturing quality since late 1960s [1]. Through prevention methods, ZD aims to boost production and minimize waste. ZD is based on the concept that the number of mistakes a worker makes doesn't matter since inspectors will catch them before they reach customers.

Thus, how to maintain their feasible production yield becomes an important issue. In this paper, the yield is defined as the final yield rate that equals to the number of functioning dies per wafer divided by the total number of dies per wafer.

Observing the changing curves of yield and cost during the product life cycle shown in Fig. 1, the product yield (blue solid line) gradually rises up in the research-and-development (R&D) phase and ramp-up phase and then keeps steady in the mass-production phase. On the contrary, the product cost (red solid line) continuously decreases during the production life cycle. If a company can improve its changing curves of yield and cost from the solid lines into their corresponding segmented lines, the company's competitiveness would be enhanced effectively.

This implies that rapidly increasing the yield in the R&D phase to transfer products into the mass-production phase, and then assuring the yield in the mass-production phase while promptly finding out and resolving the root causes of yield losses could be a feasible strategy for increasing the company's competitiveness.

Recently, many scholars have proposed machine learning and/or data mining methods to search for the root causes of the yield loss. Some related papers are summarized and presented below.

Chen and Liu [2] explored the use of a neural-network approach to classify defects in the semiconductor manufacturing processes. If the same defect occurs next time, the same reason can be used for fast and effective process monitoring and tool control to shorten the yield improvement cycle and reduce yield loss.

Lee *et al.* [3] proposed an advanced FDC model with automatic feature selection and fault-diagnosis functions that used a convolutional neural network (CNN) to detect the multivariate sensor signals from a semiconductor manufacturing process.

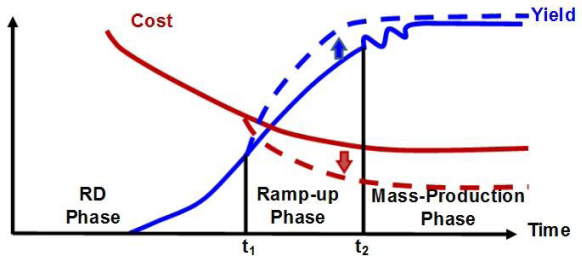


Fig. 1. Changing curves of yield and cost during the product life cycle.

However, the above-mentioned method does not solve the problem on how to find the key devices/parameters that lead to yield loss when the number of samples is limited in the R&D and ramp-up phases. It is statistically called the high-dimensional variable selection problem ($p \gg n$ problem), where p is the number of variables and n is the number of samples.

As shown in Fig. 2 [4], Cheng *et al.* [5] proposed the Key-variable Search Algorithms (KSA) scheme of the Intelligent Yield Management (IYM) system. The inputs of the IYM system should cover all the big data related to the entire manufacturing process of this product, including workpiece production paths provided by the manufacturing execution system (MES), process data of all the production tools, in-line measurement data of all the metrology tools, as well as the final yield inspection data. The KSA scheme composed of possess data preprocessing module, reliance index module [5], blind stage module [6], and interaction-effect search module [7] can quickly find out the root causes for solving the high-dimensional variable selection problem.

However, instead of considering all stages to avoid product defects simultaneously, the existing studies mentioned above can only identify the key variables with a univariate analysis that affects the yield for a single device or stage. As such, this single-variable consideration method in the above-mentioned studies may not find the golden path with a better yield amongst all the multistage manufacturing process (MMP) paths with multistage performance in consideration. At present, most studies [8]–[10] on production scheduling aim to maximize productivity but rarely consider the impact on yield.

To check the overall processing performance, the characteristics of the MMP shall be considered. MMP usually contains multiple consecutive stages to produce the final products. In order to meet the capacity requirement, each stage may have multiple devices that perform the same task, so each workpiece may randomly be assigned to different devices in each stage by the scheduling strategy. In the entire processing stages of each workpiece, such a sequence record is called the manufacturing path.

In fact, even a device of the same type is assigned to perform the same task in each stage, the actual performance of the device may still be different. Since the workpieces processed by different devices of the same stage on the production line may not be the same, multiple manufacturing paths could exist on the production line. Therefore, due to the characteristics of MMP, the accumulated performances may vary from path to path and lead to defects in products.

In the literature concerning MMP, Jin and Liu [11] proposed a data-driven method of piecewise linear regression trees to interrelate the variables for MMP with a multimode variation, yet it couldn't search out the path from the multimode variation. Lee *et al.* [12] proposed a procedure to select the process path with the best yield. The paper [12] proposed to use ANOVA to identify the process with the greatest impact on the yield and perform the machine path for the process with the largest impact on the yield. It [12] used historical paths of key processes to analyze yield data, and grouped machine paths according to high and low yields to search for a group of paths with good yields. However, when the number of samples is limited, the verification power of ANOVA will lessen as the number of samples decreases. As a result, the searching process may generate estimation errors. More importantly, it can only search for historical data, but it is unable to effectively predict a never-appeared path that might generate good yield.

Taha [13] proposed a novel framework of identifying defective fabrication path (IDFP) to find out the manufacturing path that caused the defective wafers by converting a fabrication path tree into an equivalent tree. However, the tree algorithm usually requires big data to ensure the prediction accuracy.

Lee *et al.* [14] proposed a heuristic approach to derive golden paths from an MMP to improve product quality. However, the golden path only considered the yield of the historical path but not the individual device performance on the yield, thus, a large amount of samples were required to build a model. Moreover, it cannot suggest a golden path that has never been passed; in other words, it cannot predict the yield of the manufacturing path.

To solve those problems mentioned above, this study aims to develop a Golden Path Search Algorithm (GPSA) to improve the performance of KSA. By suggesting the golden paths, various paths that are expected to produce high yield products can be generated. Moreover, through golden paths, more high-performance paths can be identified and utilized to increase the flexibility and yield of production.

The remainder of this paper is organized as follows. Section 2 briefs the original KSA scheme. Section 3 describes the bumping manufacturing process. Section 4 explains the GPSA scheme. Section 5 presents the illustrative example. Section 6 discusses the requirement of data quantity. Finally, summary and conclusions are stated in Section 7.

II. The KSA Scheme of the Intelligent Yield Management

The IYM system was developed to effectively figure out the root causes that affect the yield of the final products. To achieve this goal, all the big data related to the entire manufacturing processes of the product, such as the workpiece production paths provided by the manufacturing execution system (MES), process data of all the production tools, in-line measurement data of all the metrology tools, and the final yield inspection data, should all be included as the inputs of the IYM system.

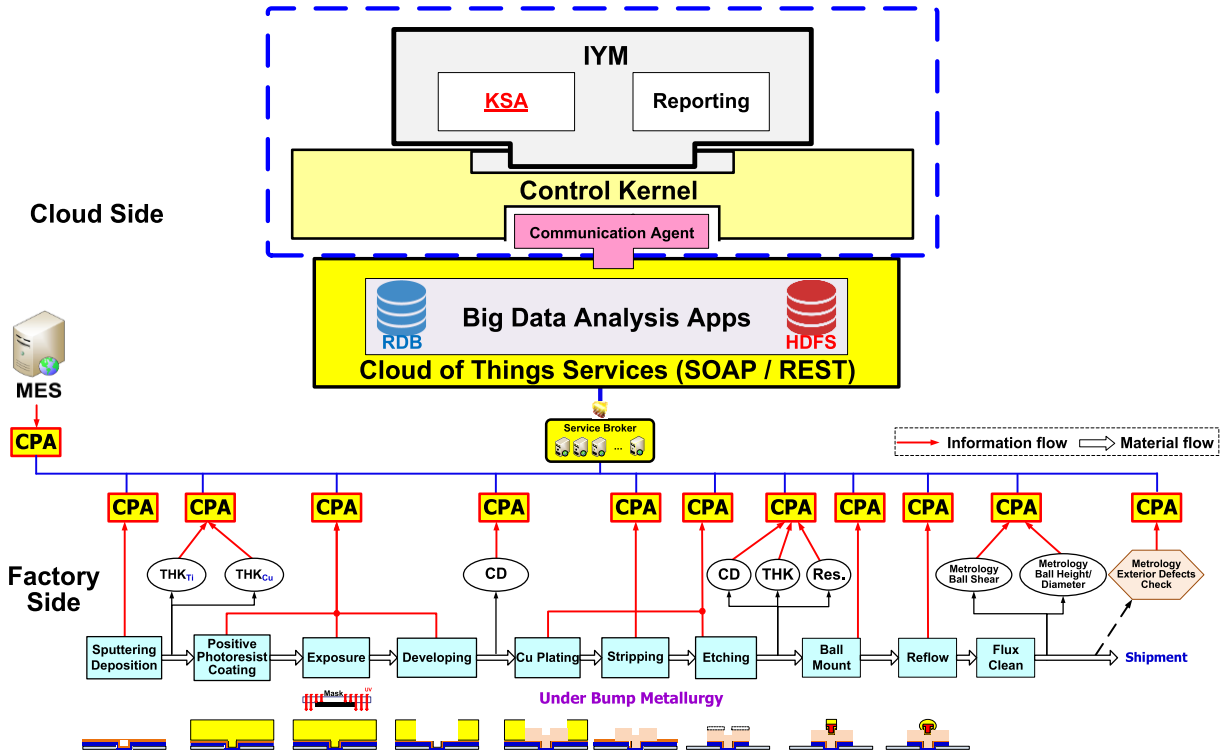


Fig. 2. KSA scheme of the intelligent yield management system [4].

As shown in Fig. 2, the IYM system [4] consists of the KSA scheme and the reporting module, along with numerous Internet-of-Things (IoT) devices denoted as Cyber-Physical Agents (CPAs) [4]. CPAs are adopted by the factory side to perform input data collection for the KSA scheme.

Observing Fig. 3, excluding the so called GPSA module shown at the bottom-right corner of Fig. 3 (GPSA will be added into the KSA scheme later), the original KSA scheme contains various data preprocessing modules, blind-stage search algorithm (BSA) module [6], KSA module [5], reliance index module for KSA ($\mathbf{RI_K}$), interaction-effect search algorithm (IESA) module [7], and reliance index module for IESA ($\mathbf{RI_I}$).

The inputs of the KSA scheme include production paths ($\mathbf{X_R}$), process data ($\mathbf{X_P}$), and in-line metrology values (y) as well as the final inspection data ($\mathbf{Y}$). Defect detection is important in the semiconductor industry, as defects ($\mathbf{D}$) (e.g. scratch, particle, etc.) may happen to any production stage. Thus, defects should also be taken into consideration for the KSA analysis.

The outputs of the KSA scheme are the blind-stage index: BS; KSA search results of Triple Phase Orthogonal Greedy Algorithm (TPOGA): $\mathrm{KS_O}$ and that of Automated Least Absolute Shrinkage and Selection Operator (ALASSO): $\mathrm{KS_L}$, and their accompanying reliance index: $\mathrm{RI_K}$; as well as interaction-effect index: $\mathrm{IE_O}$ via TPOGA, $\mathrm{IE_T}$ via Tree, and IESA reliance index: $\mathrm{RI_I}$.

However, the original KSA scheme was mainly designed to search for the key variables only. KSA is still unable to search for the best device path combined with the best yield, which is defined as the Golden Path in this study. If the device path combination that generates the best yield can be found before each production, manufacturers can improve the yield effectively. Therefore, the purpose of this paper is to integrate the so-called GPSA module shown at the bottom-right corner of Fig. 3 to improve the performance of KSA.

III. Introducing the Bumping Manufacturing Processes

The semiconductor manufacturing phase can be divided into two steps: "frontend" and "backend." The frontend step contains the wafer-fabrication and chip-probing processes; while the backend step includes packaging and testing processes. Flip chip is one of the packaging technologies and wafer bumping is essential to flip-chip semiconductor packaging.

Bumping is an advanced wafer-level process technology where "bumps" or "balls" made of solder are formed on the wafers in a whole wafer form before the wafer is being diced into individual chips. Those "bumps," composed of eutectic, lead-free, high lead materials or Cu pillar on the wafer, are the fundamental interconnect components that will interconnect the die and the substrate together into a single package. These bumps not only provide a connected path between die and substrate but also play an important role in the electrical, mechanical, and thermal performance of the flip-chip package.

The production line of the bumping process shown in Fig. 4 consists of two sub-layers, i.e., Re-Distribution Layer (RDL) and Under Bump Metallurgy (UBM).

The RDL and UBM bumping processes contain the following stages: Sputtering Deposition, Photoresist (including Positive Photoresist Coating, Edge Bead Remover, Exposure,

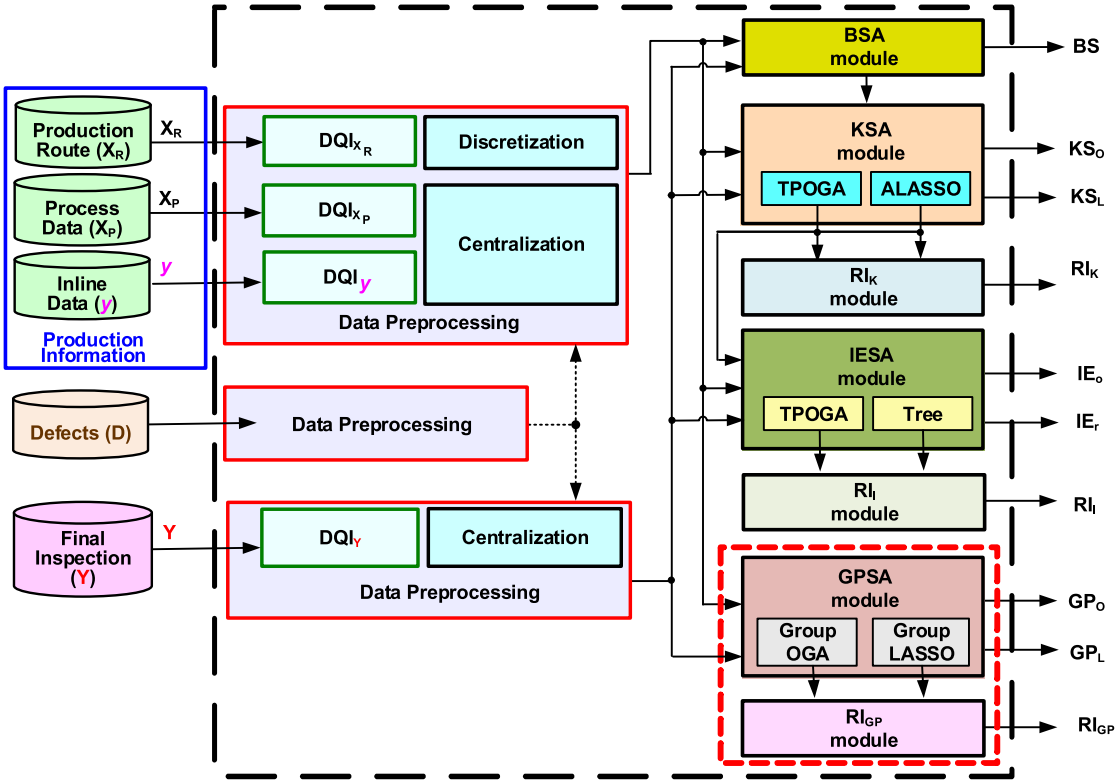


Fig. 3. The KSA scheme with the GPSA module.

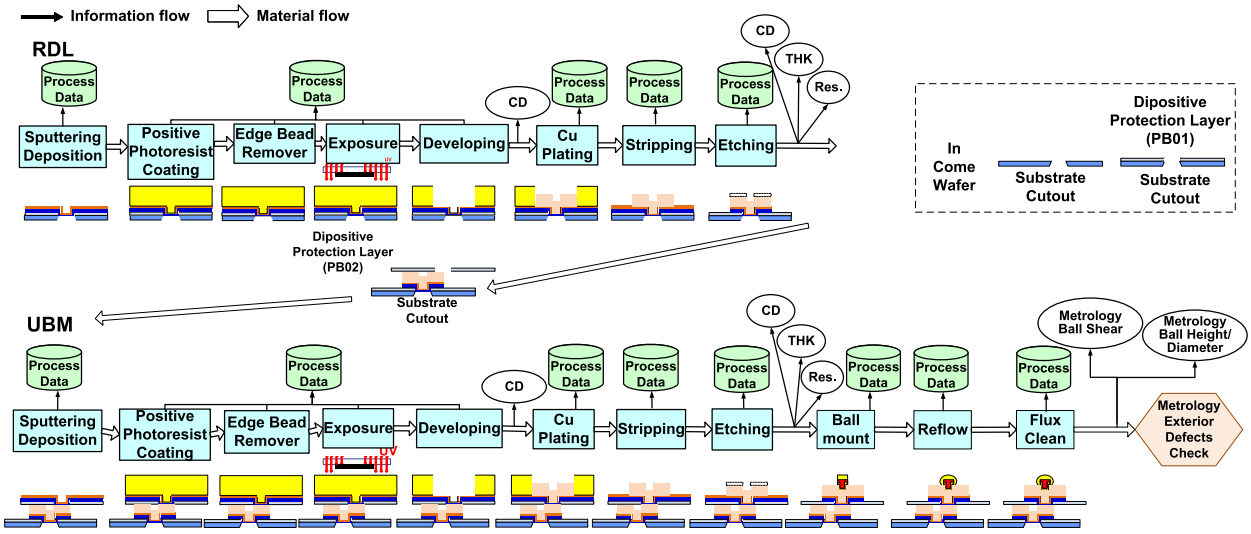


Fig. 4. Production line of the bumping process.

and Developing), Cu Plating, Striping, Etching, Ball Mount, Reflow, and Flux Clean.

There are roughly 25 production stages in the illustrative bumping process, and each stage has about 1 to 5 manufacturing devices as shown in Table I. Thus, as shown in Fig. 5, a total of $4 \times 3 \times 4 \times \ldots \times 3 \times 5 \times 4 = 6,449,725,440$ paths could possibly affect the yield of the bumping process. If the information of device maintenance and different material sources is also considered, the number of yield-affecting parameters, i.e., p, is even higher. In the R&D phase of the product life cycle, the number of samples, i.e., n, is relatively small, which thereby leads to a challenge of finding the root causes of poor yield.

IV. Golden Path Search Algorithm

Whether in semiconductor, TFT-LCD, or other high-tech industries, pursuing high-yield products is a common goal. These high-tech industries have one thing in common: com-

TABLE I
TOTAL NUMBER OF DEVICES IN THE BUMPING MANUFACTURING PROCESS

Layer	Stage	# of device
RDL	sputtering deposition	4
	coating	3
	edge bead remover	4
	exposure	4
	developing	3
	cu plating	3
	stripping	2
	etching	4
	descum	3
	Furnace	3
UBM	sputtering deposition	3
	coating	3
	edge bead remover	2
	exposure	2
	developing	2
	cu plating	3
	stripping	2
	etching	4
	ball mount	3
	reflow	5
	flux clean	4

plex manufacturing processes. In the semiconductor manufacturing process, hundreds to thousands of procedures are usually required to obtain the final output. However, each batch of products passes through one or more production devices in each production stage, and the combinations of production devices for each batch of products are not the same. Since the condition of the device used in each production process will directly impact the yield of the product; thus, how to find the best device combination path to increase product yield is an important issue.

Considering the real data of a semiconductor factory, the number of possible device combinations in the entire factory may reach billions. This research focuses on finding the best device combination that can improve the product yield, which is the golden path. However, considering the issue of production capacity, if only one optimal device combination path is found, the actual production can only be optimally configured under limited production capacity. Therefore, this study will also 1) generate all possible manufacturing paths of the key-stage devices and predict their corresponding yields, 2) consider the processing variation of the device combination through the path, then 3) calculate the estimated values and rank each manufacturing path based on the predicted yield and variation. As such, manufacturers may plan the production in advance according to the path ranking with better yield and less variation to avoid the manufacturing paths with poor yield and large variation.

However, some stages have rather minor impact on the yield during the production process, that is to say, no matter which device is arranged for the stage to process (e.g., testing and inspection stages, …, etc.), its impact on the yield is the same. Therefore, GPSA will prioritize the selection of the devices in the key-stage for path sorting, without considering the stages of insignificant influence. Also, GPSA does not allow reentrance; it means that each path can only pass through one of the devices in the stage. Moreover, due to insufficient statistical verification power, devices with a pass rate of less than 5% are not taken into consideration.

As mentioned previously and shown in Fig. 5, if the device combinations of all stages are considered, there are 6, 449, 725, 440 manufacturing path types. After selecting the important stages of GPSA, the device combinations will be greatly reduced. The main purpose of GPSA is to complement KSA to find golden paths with better yield, and eliminate paths with poor yield. Based on the above goals, this research proposes a two-phase GPSA scheme with Phase I: identify the key stages and Phase II: generate all manufacturing paths of the key-stage devices and predict their corresponding yields. Finally, all key-stages that may affect the yield are prioritized according to the yield index values of the golden paths for manufacturers' reference. The flow chart of the GPSA scheme is shown in Fig. 6, and the explanations of each step are stated below.

A. Conduct Data Preprocessing and Quality Check

To begin with, data preprocessing and quality check must be performed to ensure the quality of the input data: production routes ($\mathbf{X_R}$) and final inspection ($\mathbf{Y}$). In (1), $\mathbf{X_R}$ represents the production routes, p is the number of parameters indicating the total device number a workpiece may go through, and n is the number of samples. As for (2), $\mathbf{Y}$ means the final inspection data and again n is the number of samples.

$$\mathbf{X_R} = \begin{bmatrix} x_{1\mathrm{R}1} & \dots x_{1\mathrm{R}j} \dots & x_{1\mathrm{R}p} \\ \vdots & \ddots & \vdots \\ x_{n\mathrm{R}1} & \dots x_{n\mathrm{R}j} \dots & x_{n\mathrm{R}p} \end{bmatrix} \tag{1}$$

$$\mathbf{Y} = \begin{bmatrix} \mathrm{Y}_1 & \dots & \mathrm{Y}_n \end{bmatrix}^{\mathrm{T}}. \tag{2}$$

Then, Phase I is executed. The three steps of Phase I are shown in Fig. 7.

B. Phase I: Identify the Key Stages

Step I-1: Filter out the devices with passing rate $< r$ %

Filter out the devices with too few passing samples, as these devices may generate non-representative results. Users can filter out the devices of r % passing rate based on their experience. The default would be 5% passing rate.

Step I-2: Select the key stages

In order to eliminate minor stages, this step selects important stages through the feature selection methods. Moreover, GPSA contains two algorithms to evaluate the searching reliability, which are Group Least Absolute Shrinkage and Selection Operator (GLASSO) and Group Orthogonal Greedy Algorithm (GOGA), as described below.

a) GLASSO: LASSO [15] is a popular variable selection method based on l_1 regularization. However, the selected model may be difficult to interpret in some cases. In particular, when the predictors are grouped naturally or based

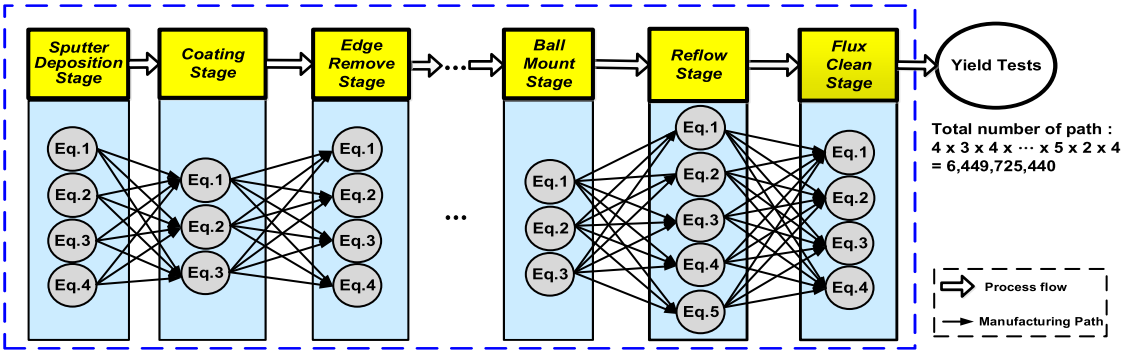


Fig. 5. Illustration of process routes.

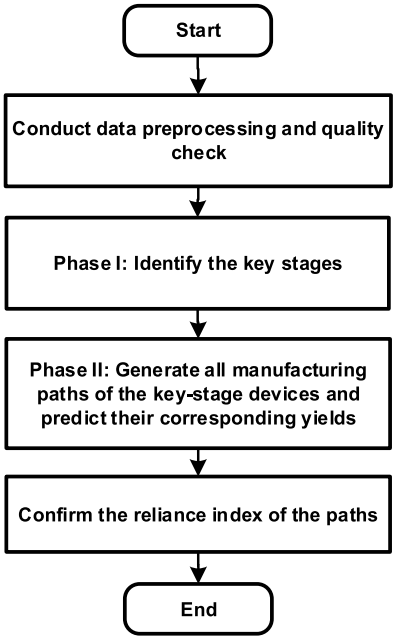


Fig. 6. Flow chart of the GPSA scheme.

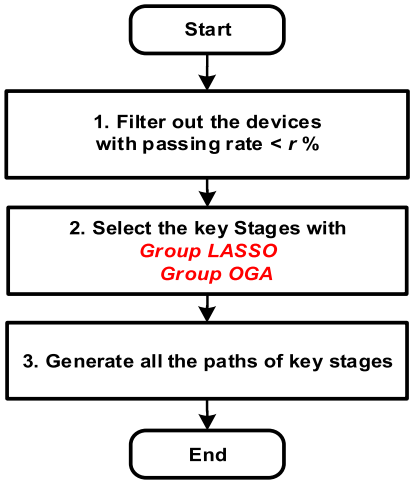


Fig. 7. Phase-I flow chart of the GPSA scheme.

on previous knowledge, LASSO may not jointly include or exclude these groups of features. To overcome this problem, Yuan and Lin [16] proposed the Group LASSO (GLASSO), which ensures that the predefined groups of predictors can be jointly selected into or eliminated from a model. Define $\boldsymbol{X_R} = \{x_{R1}, x_{R2}, \ldots, x_{Rp}\}$ as the design matrix, and $J = \{J_1, J_2, \ldots, J_m\}$ as the set of stages, where $J_l = \{x_{R\theta}, x_{R\theta+1}, \ldots, x_{Rj}\}$ with $1 \le l < m$ and $1 \le \theta < j < p$. The GLASSO formulation is stated as follows.

$$\hat{\beta}^{group\ lasso} = \underset{\hat{\beta}}{argmax}\ (\left|\left|\mathbf{Y} - \mathbf{X_R}\hat{\beta}\right|\right|_2^2 + \lambda \sum_{l=1}^{m} ||\hat{\beta}_{J_l}||) \quad (3)$$

where

$\hat{\beta}^{group\ lasso}$: optimized coefficients of $\hat{\beta}$;

λ: penalty operator;

$\hat{\beta}_{J_l}$: regression coefficient of $\mathbf{Y}$ for J_l.

Let λ_{max} be the smallest value of λ that makes $\hat{\beta} = 0$. Choose a proper λ to keep the set of important stages $\hat{L}_s$ and discard the insignificant ones.

b) GOGA: Another popular variable selection method is through greedy forward selection. Ing and Lai [17] proposed Orthogonal Greedy Algorithm (OGA) to achieve selection consistency in high-dimensional linear regression models. However, like LASSO, OGA cannot ensure that the variables in one group will be jointly selected into or eliminated from a model. Chan *et al.* [18] proposed Group OGA (GOGA) to tackle this difficulty. In particular, GOGA allows us to simultaneously include or exclude grouped variables. The steps of GOGA are briefly described below.

(i) Define $U^{(0)} = \mathbf{Y} = (\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_n)^{\mathrm{T}}$, choose the variable defined as $J_{s_1} = \{x_{\mathrm{R}\theta}, x_{\mathrm{R}\theta+1}, \ldots, x_{\mathrm{R}j}\}$, which is the most correlated with $U^{(0)}$. Then, the corresponding residual will be:

$$U^{(1)} = U^{(0)} - \hat{\beta}_{s_1}^{(1)} J_{s_1} \quad (4)$$

where

J_{s_1}: highest correlation variable with $U^{(0)}$;

$\hat{\beta}_{s_1}^{(1)}$: regression coefficient of $U^{(0)}$ for J_{s_1}.

(ii) Choose another variable, J_{s_1}, which is the most correlated with $U^{(1)}$. The corresponding residual will be:

$$U^{(2)} = U^{(0)} - \hat{\beta}_{s_1}^{(2)} J_{s_1} - \hat{\beta}_{s_2}^{(2)} J_{s_2} \quad (5)$$

where

$\hat{\beta}_{s_1}^{(2)}$: regression coefficient of $U^{(1)}$ for J_{s_1};

$\hat{\beta}_{s_2}^{(2)}$: regression coefficient of $U^{(1)}$ for J_{s_2}.

(iii) Return to *(ii)* and repeat S times so as to stepwise choose $J_{s_1}, J_{s_2}, \ldots, J_{s_s}$ and calculate the corresponding

regression coefficients $\hat{\beta}_{s_1}^{(s)}, \hat{\beta}_{s_2}^{(s)}, \ldots, \hat{\beta}_{s_s}^{(s)}$, respectively. Then, the corresponding residual will be:

$$U^{(k)} = U^{(0)} - \hat{\beta}_{S_1}^{(s)} J_{s_1} - \hat{\beta}_{S_2}^{(s)} J_{s_2} - \cdots - \hat{\beta}_{S_s}^{(s)} J_{s_s}. \quad (6)$$

However, it is hard to select the exact S stages to recover the entire original signal. Therefore, Ing and Lai [17] proposed a termination condition, HDGIC, to choose along the GOGA path that has the smallest value of a suitably chosen criterion. Let

$$\text{HDGIC}(J_s) = m\, log\hat{\sigma}_J^2 + \#(J_s)\log m(\log m - \log\log m) \quad (7)$$

with

$$\hat{\sigma}_J^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y} - \hat{\mathbf{Y}}_{J_s})^2 \quad (8)$$

where

J_s: set of variables selected in the model $(J_{s_1}, J_{s_2}, \ldots, J_{s_s})$;
$\hat{\sigma}_J^2$: mean square error of the corresponding model;
$\hat{\mathbf{Y}}_{J_s}$: predictive value of final inspection corresponding to J_s.

Define J_s to be the set of all the variables selected after executing GOGA + HDGIC via (7), where S represents the number of stages selected in the model. However, there are still some irrelevant stages in J_S. The method of removing irrelevant variables is necessary.

The threshold estimation procedure trims the solution set J_S to eliminate irrelevant groups as follows. Based on the concept of GOGA + HDGIC, define a subset $\hat{j}_S$ of J_S, where $\hat{j}_S$ is the real relevant variables in the model:

$$\hat{j}_S = \{J_l : \text{HDGIC}(J_S - \{J_l\}) > \text{HDGIC}(J_S), 1 \le l \le S\}, \quad \text{if } S > 1. \hat{j}_S = \{\hat{J}_1\}, \quad \text{if } S = 1. \quad (9)$$

If $\hat{j}_S$ is greater than the HDGIC value, it means that the group has enough influence and needs to stay in the model; if it is less than the HDGIC value, it will be removed. The selected groups are sequentially removed through trimming, and the remaining S-1 groups are used to calculate the HDGIC value. Finally, combining the processes of GOGA, HDGIC and trimming to select the key stages.

Step I-3: Generate all the paths of key stages

If only one optimal manufacturing path is found by the algorithm in actual production, the production can only be conducted under limited capacity. To resolve this issue, all the key-stage devices are utilized to create manufacturing paths.

Define the device combination of S stages as a S-ary Cartesian power [19] of a set $\hat{S}_s = \{p_1, p_2, \ldots, p_N\}$, where p_q represents the q^{th} path, $q = 1, 2, \ldots, N$. The S-ary Cartesian power formulation is stated as follows:

$$\hat{S}_s = S_1 \times S_2 \times \ldots \times S_s = \left\{ (x_{\text{R}\theta}, x_{\text{R}\theta+1}, \ldots, x_{\text{R}p}) \middle| \begin{array}{l} (x_{\text{R}\theta}, x_{\text{R}\theta+1}, \ldots) \in S_1 \wedge \ldots \\ \wedge(\ldots, x_{\text{R}k-1}, x_{\text{R}k}) \in S_s \end{array} \right\} \quad (10)$$

where S_ν is the key stages selected by the GOGA $(\hat{j}_s)$ or GLASSO's result $(\hat{L}_s)$, $\nu = 1, 2, \ldots, S$.

By applying Phase-I steps, N important key-stage paths will be generated; yet which ones work better on yield is still unknown. In Phase II, the scheme creates the models for all the paths and predicts their yields, and then prioritizes them accordingly.

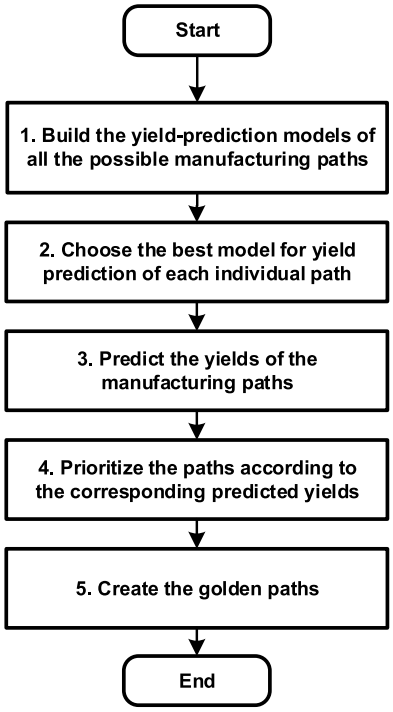


Fig. 8. Phase-II flow chart of the GPSA scheme.

C. Phase II: Predict the Corresponding Yields of All Manufacturing Paths of the Key-Stage Devices

According to the devices of the key-stages, the prediction models are created such that the yields of all the paths are predicted. Phase II is composed of 5 steps as shown in Fig. 8.

Step II-1: Build the yield-prediction models of all the manufacturing paths

The following regression model is used to predict the yield of a specific manufacturing path:

$$\text{Y}_t = \beta_0 + \beta_1 x_{t\text{R}1} + \beta_2 x_{t\text{R}2} + \cdots + \beta_k x_{t\text{R}k} + \varepsilon_t \quad (11)$$

where Y_t is the yield of the t^{th} sample and $x_{t\text{R}1}, \ldots, x_{t\text{R}k}$ are the devices in this specific path. The value of $x_{t\text{R}k}$ is 0 or 1 depending on whether the k^{th} device is utilized, and ε_t is the random noise.

This is the most intuitive but powerful model to predict the yield; however, this model is usually mis-specified. More features may be created to reduce the effect of misspecification. The simplest way to do this is to adopt the interaction model shown below:

$$\text{Y}_t = \beta_0 + \sum\nolimits_m \beta_m x_{t\text{R}m} + \sum\nolimits_{k,m} \beta_{km} x_{t\text{R}km} + \varepsilon_t \quad (12)$$

where $x_{t\text{R}km} = x_{t\text{R}k} x_{t\text{R}m}$. Note that this model also takes the effect of interaction terms into account. In particular, it considers the effect of passing through $x_{t\text{R}k}$ and $x_{t\text{R}m}$ at the same time. It is also straightforward to consider the features of the third-order model: $x_{t\text{R}kml} = x_{t\text{R}k} x_{t\text{R}m} x_{t\text{R}l}$. Like the

two-way interaction model, it considers the effect of passing through $x_{t\mathrm{R}k}$, $x_{t\mathrm{R}m}$ and $x_{t\mathrm{R}l}$ at the same time. This model can be extended up to the O^{th} order. Since the interaction involves too many factors, it is difficult and unnecessary to consider the higher-order O value, thus, $O < 5$ is adopted in this paper.

Step II-2: Choose the best model for yield prediction of each individual path

The proper order of interaction is decided in this step. The order can be chosen by the Akaike Information Criterion (AIC) [20]. In statistics, AIC is a criterion for model selection among a finite set of models. Specifically, let k be the number of parameters and $\hat{L}$ be the maximum value of the likelihood for the model. Then AIC value of the model can be calculated as follows:

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}). \tag{13}$$

Given the order O, by assuming the normality of random noise, the AIC value can be obtained. It is worth mentioning that the order can be different for different manufacturing paths.

Step II-3: Predict the yields of the manufacturing paths

Once the predicted model is decided for each path, yield prediction of each specific path can be conducted. Indeed, by the definition of $x_{t\mathrm{R}m}$, the predicted yield is simply the sum of all coefficients of the model:

$$\hat{\mathrm{Y}}_{p_q} = \beta_0 + \sum\nolimits_m \beta_m x_{\mathrm{R}m} + \sum\nolimits_{k,m} \beta_{km} x_{\mathrm{R}km} \tag{14}$$

where $m \epsilon p_q$, $k \epsilon p_q$, and $\hat{\mathrm{Y}}_{p_q}$ is the predicted yield of the q^{th} path.

Step II-4: Prioritize the paths according to the corresponding predicted yields

After estimating the yield ($\hat{\mathrm{Y}}_{\boldsymbol{p_q}}$), the manufacturing paths will be sorted by the predicted yields. Higher predicted yield value represents lower defect occurrence; on the contrary, lower predicted yield value indicates higher chances of defects.

Step II-5: Create the golden paths

The top 10 paths will be selected as the golden paths and output to users, the results may contain paths that have never been passed in the historical data for reference. However, these never-passed paths may not be trustworthy. Therefore, the reliance index module is required to confirm the results next.

D. Confirm the Reliance Index of the Paths

The concept of the $\mathbf{RI_{GP}}$ module refers to the $\mathbf{RI_K}$ module of KSA [5] to inform users of the reliance degree of the results. The $\mathbf{RI_{GP}}$ value is between 0 and 1, and the calculation of $\mathbf{RI_{GP}}$ is as follows:

$$\mathbf{RI_{GP}}(kp_q) = 1 - \frac{\hat{Y}_{p_q} - \sum_{\theta=1}^{i} Corr(x_{\mathrm{R}\theta})}{\mathbf{Y}_{max} - \mathbf{Y}_{min}} \tag{15}$$

where

$\mathbf{RI_{GP}}(kp_q)$: reliability index of the q^{th} machine path combination;

$Corr(x_{\mathrm{R}\theta})$: correlation between $x_{\mathrm{R}\theta}$ and $\mathbf{Y}$;

$\mathbf{Y}_{max}$: maximum of $\mathbf{Y}$;

$\mathbf{Y}_{min}$: minimum of $\mathbf{Y}$.

In the following, two remarks are added. Remark 1 addresses the reason why linear regression models adopted in the proposed method are sufficient for the purpose of golden path search; while Remark 2 explains how the computation cost scales with the number of key stages.

Remark 1: The main reason why linear models are used is that the output variable (yield) under consideration is of continuous type, in which linear models generally perform well. In fact, non-linearity is also taken into account in this paper by considering the high-order interaction terms in Phase II procedure. A data-driven method is suggested to select suitable high-order terms that can better describe the data and predict the yields. A non-linear extension of the proposed method to output the variables of discrete type can be done by implementing the Chebyshev greedy algorithm [21], which is a non-linear counterpart of OGA in the Poisson regression models. This kind of extension, however, is out of scope of this paper.

Remark 2: The computational burden of the procedure grows rapidly along with the number of key stages because it leads to an exponential increase in the number of paths. Let $\bar{M}$ be the largest computationally affordable number of key stages, and $\bar{K}$ be the number of key stages identified in Phase I. When $\bar{K} > \bar{M}$, a modified procedure, which is still "greedy" in nature, is offered to substantially save the computational cost. The modification goes as follows.

Step R2-1: Set $h = M + 1$ and let S denote the set of key stages formed by the top M key stages ranked by GOGA, where $1 \le M \le \bar{M}$ is a prescribed positive integer.

Step R2-2: Determine the golden paths based on S (and Phase II), and compute the average yield from the golden paths.

Step R2-3: Among the remaining $\bar{K} - h + 1$ key stages, find the one that generates the largest average yield from the golden paths after it is added into S. Denote this key stage by $\hat{j}_h$.

Step R2-4: If the average yield obtained in Step R2-3 is smaller than the one obtained based on S, stop the algorithm; otherwise, set $S = S \cup \{\hat{j}_h\}$ and $h = h + 1$.

Step R2-5: If $h > \bar{M}$, stop the algorithm; otherwise, return to Step R2-3.

V. Illustrative Example

The bumping process data for semiconductor package manufacturing at the ASE group, Inc., Taiwan, are adopted as the illustrative example. A wafer is the basic tracking unit of a bumping process which can be divided into redistribution layer (RDL) and under bump metallurgy (UBM) layer with each layer consisting of sputtering, photo, plating, stripping, and etching stages. The GPSA scheme shown in Figs. 6-8 are applied and the results are presented below.

A. Conduct Data Preprocessing and Quality Check

To begin with, the GPSA scheme will check the data quality and the data format. The final yield test is the resistance test

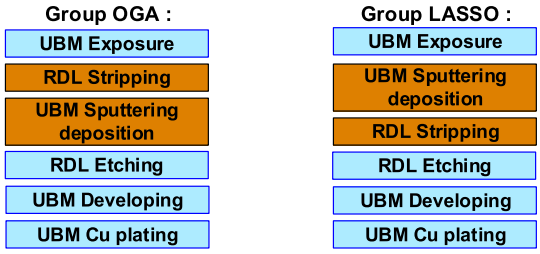


Fig. 9. Results of the key-stage search.

(the final yield equals to the number of functioning dies per wafer divided by the total number of dies per wafer), and all wafers in the resistance layer are included for analysis ($n = 345$). Twenty-one stages out of 25 in total ($m = 21$) are actually analyzed after excluding the inspection stages. There are totally 282 devices fab-wide. **Y**is the final yield data. For example, the 1st wafer has a yield rate of 93.23%, and the 2nd wafer has a yield rate of 95.46%, then the yield-test result of **Y** will be [93.23, 95.46, . . . , Y_n].

B. Phase I: Identify the Key Stages

In this case, 5% of filter rate is utilized in Step I-1 to eliminate some unbalanced devices so as to increase the model accuracy. The total number of the devices is decreased from 282 to 174. In Stage I-3, 174 devices are classified into 21 groups. Then, GOGA and GLASSO are conducted. As shown in Fig. 9, all the stages searched by GOGA and GLASSO are identical except that the sequence of Top 2 and Top 3 stages differs; while UBM Exposure is the Top 1 stage by both algorithms. The devices of the six key stages shown in Fig. 9 will then be utilized to generate the production paths.

In Step I-3, find the device numbers of those six key stages shown in Table I and apply (10) to generate the production paths:

$$\begin{aligned}\hat{S}_6 &= S_{Top1} \times S_{Top2} \times \ldots \times S_{Top6} \\ &= 2 \times 2 \times 3 \times 4 \times 3 \times 1 = 144.\end{aligned}$$

As a result, the number of the original paths (as shown in Fig. 5) will be reduced to 144, which means the complexity of the total path scheduling is significantly lessened.

C. Phase II: Predict the Corresponding Yields of All Manufacturing Paths of the Key-Stage Devices

There is no way to know which kind of model can predict the most accurate yield before model creation; hence, in Step II-1, the O^{th} order model of each path (14) is created according to the historical data, and various models are adopted to fit the best prediction results. Take Path 6 as the example, different kinds of models are created as shown in Table II.

The next step is to decide the proper order. For instance, the 4^{th} order of Path 6 has the smallest AIC [20] value as shown in Table III, this model is therefore chosen as the prediction model.

TABLE II
EXAMPLE OF THE PATH-6 MODEL

Order	Model
1^{st}	$Y_t = \beta_0 + \sum_a \beta_a x_{tRa} + \varepsilon_t$
2^{nd}	$Y_t = \beta_0 + \sum_a \beta_a x_{tRa} + \sum_{a,b} \beta_{a,b} x_{Rab} + \varepsilon_t$
3^{rd}	$Y_t = \beta_0 + \sum_a \beta_a x_{tRa} + \sum_{a,b} \beta_{a,b} x_{Rab} + \sum_{a,b,c} \beta_{a,b,c} x_{Rabc} + \varepsilon_t$
4^{th}	$Y_t = \beta_0 + \sum_a \beta_a x_{tRa} + \sum_{a,b} \beta_{a,b} x_{Rab} + \sum_{a,b,c} \beta_{a,b,c} x_{Rabc} + \sum_{a,b,c,d} \beta_{a,b,c,d} x_{Rabcd} + \varepsilon_t$

TABLE III
AIC OF THE PATH-6 MODEL

	1^{st}	2^{nd}	3^{rd}	4^{th}
AIC	2058.57	2048.56	1995.79	1988.24

After determining the prediction model of each path, the yield prediction of each specific path ($\hat{Y}_{p_q}$) can be conducted in Step II-3. The prediction value of Path 6 ($\hat{Y}_{p_6}$) is the sum of all the coefficient values of the 4^{th} order model. The coefficients of the 6th path model are listed in Table IV. In fact, not only the effects of all devices that go through Path 6, but the yield rate generated by the 4^{th} interaction of all these devices are also taken into consideration. The sum of the coefficient values is 99.991 as shown in Table V, which indicates that under the effect of passing through Path 6, the expected yield rate would be 99.991.

In Step II-4, the golden paths are sorted in sequence according to the predicted values of yield rate ($\hat{Y}_{p_q}$). The path with the highest predicted value would be chosen as the Top 1 golden path, and the paths with lower predicted values are suggested to be avoided. The results are displayed in Table V.

Finally, observing the TOP 1 row in Table V, which in fact shows the data of Path 6, since its $\hat{Y}_{p_6}$ is the highest value (99.991), Path 6 is chosen as the Top 1 golden path. However, this path has not been utilized yet, thus the historical average yield rate ($\bar{\mathbf{Y}}_{p_{\mathbf{Top1}}}$) shows NaN in Table V.

D. Confirm the Reliance Index of the Paths

Applying (15), $\mathbf{RI_{GP}} = 0.9976$, which indicates that the difference between this prediction result and the correlation of the device and **Y** is little; therefore, the result is reliable. Moreover, as shown in the last row of Table V, the largest value of the last path $\hat{Y}_{p_{\mathbf{Last1}}}$ is 83.85, which is identical to the

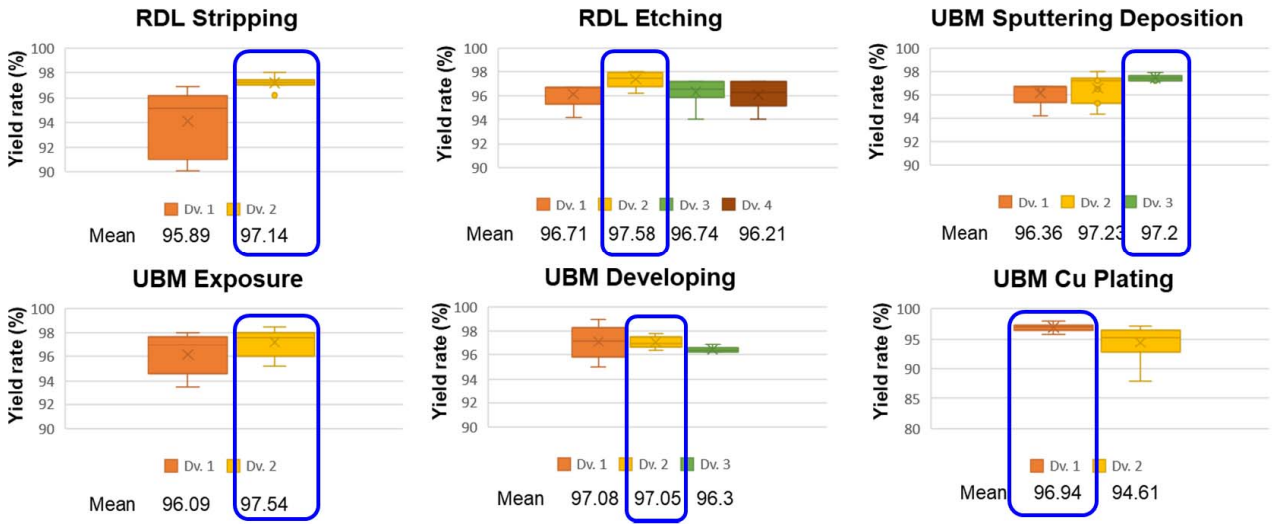


Fig. 10. Yield-rate box plots of all the devices in the six key stages.

TABLE IV
COEFFICIENTS OF THE PATH-6 MODEL

β_1	β_2	β_3	β_4	β_5	β_6
-102.94	-100	-146.98	-101.58	-102.26	-103.48

$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{2,6}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{3,6}$	$\beta_{4,5}$	$\beta_{4,6}$	$\beta_{5,6}$
-100.33	149.26	102.31	104.61	104.19	146.98	-101.21	101.95	100.45	141.56	-110.72	146.26	101.88	103.02	102.89

$\beta_{1,2,3}$	$\beta_{1,2,4}$	$\beta_{1,2,5}$	$\beta_{1,2,6}$	$\beta_{1,3,4}$	$\beta_{1,3,5}$	$\beta_{1,3,6}$	$\beta_{1,4,5}$	$\beta_{1,4,6}$	$\beta_{1,5,6}$	$\beta_{2,3,4}$	$\beta_{2,3,5}$	$\beta_{2,3,6}$	$\beta_{2,4,5}$	$\beta_{3,4,5}$	$\beta_{3,4,6}$	$\beta_{3,5,6}$	$\beta_{4,5,6}$
102.14	103.74	-102.22	-101.92	-145	105.41	-149.9	104.8	-104.67	-105.49	102.82	-104.39	-149.32	100.24	107.86	-140.98	107.06	-105.21

$\beta_{1,2,3,4}$	$\beta_{1,2,3,5}$	$\beta_{1,3,4,5}$	$\beta_{1,3,4,6}$
-101.04	108.75	-109.2	140.65

TABLE V
RESULTS OF GOLDEN PATHS

Path ranking	RDL stripping	RDL etching	UBM sputtering deposition	UBM exposure	UBM developing	UBM Cu plating	Predicted yield ($\bar{Y}_{p_q}$)	RI_{GP}	Order of optimal model	# of passed wafers	Mean of yield ($\bar{Y}_{p_q}$)
Top 1 (Path 6)	Dv. 2	Dv. 2	Dv. 3	Dv. 2	Dv. 2	Dv. 1	99.991	0.9976	4	0	NaN
Top 2	Dv. 2	Dv. 3	Dv. 3	Dv. 2	Dv. 1	Dv. 1	99.56	0.9961	3	1	99.55
Top 3	Dv. 2	Dv. 2	Dv. 3	Dv. 2	Dv. 1	Dv. 1	99.94	0.997	3	0	NaN
Top 4	Dv. 2	Dv. 1	Dv. 3	Dv. 2	Dv. 1	Dv. 1	99.93	0.9881	3	1	100
Top 5	Dv. 2	Dv. 3	Dv. 1	Dv. 2	Dv. 2	Dv. 1	99.9	0.9987	3	0	NaN
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Last 5	Dv. 1	Dv. 1	Dv. 1	Dv. 1	Dv. 2	Dv. 2	87.94	0.9773	4	0	NaN
Last 4	Dv. 1	Dv. 4	Dv. 1	Dv. 2	Dv. 3	Dv. 2	87.11	0.9834	4	0	NaN
Last 3	Dv. 1	Dv. 4	Dv. 1	Dv. 1	Dv. 2	Dv. 2	86.45	0.9977	4	1	86.65
Last 2	Dv. 1	Dv. 4	Dv. 1	Dv. 2	Dv. 2	Dv. 2	85.49	0.996	4	0	NaN
Last 1	Dv. 1	Dv. 4	Dv. 1	Dv. 1	Dv. 3	Dv. 2	83.85	0.9935	4	2	83.85

actual average yield rate ($\bar{\mathbf{Y}}_{p_{\text{Last1}}}$) of passing through this path. Thus, it proves that GPSA can also successfully identify the path with poor yield.

Figure 10 depicts the yield-rate box plots of all the devices in the six key stages. Also, the devices (Dv.2-Dv.2-Dv.3-Dv.2-Dv.2-Dv.1) in Top 1 golden path are marked with blue blocks.

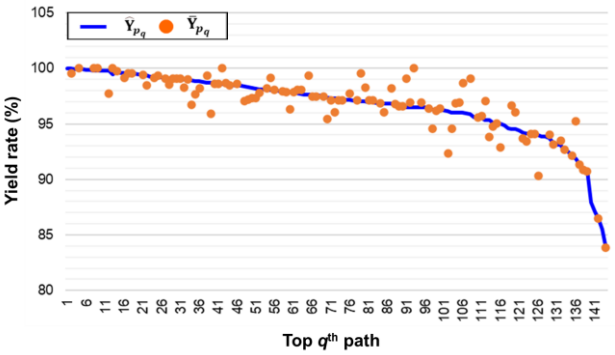


Fig. 11. Yield-rate comparison between the actual ($\bar{Y}_{p_q}$) and predicted paths ($\hat{Y}_{p_q}$) of the illustrative example.

Observing those devices in the blue blocks, besides Dv. 2 of UBM Developing, all the other devices possess the highest yield rate in each individual key stages. As for Dv. 2 of UBM Developing, its yield rate is the 2nd highest while its variance is smaller than that of UBM-Developing's Dv. 1, which has the best yield rate. Based on the fact stated above, Path 6 is verified to be selected as the Top-1 golden path. By the same token, it can be verified that the yield rates of Dv.1-Dv.4-Dv.1-Dv.1-Dv.3-Dv.2, which belong to the last path, are indeed the ones with the poorest yields as shown in Fig. 10.

A comparison between the predicted values ($\hat{\mathbf{Y}}_{p_q}$) of all 144 paths sorted from large to small and the 106 paths actual values ($\bar{\mathbf{Y}}_{p_q}$) that wafers actually pass through is depicted in Fig. 11. The predicted values of the 144 paths appear to have better prediction trend than some of the 106 values, and the extreme values of the worst paths can be detected as well. In short, yield enhancement can be achieved by adopting the paths with good predicted yields and excluding the ones with poor predicted yields prior to production.

Comparing GPSA with the traditional ANOVA method, it is discovered that the difference between these two methods lies in Phase I: GPSA identifies the key stages by GOGA, whereas ANOVA points out key devices with p-values less than 0.05 and then identifies the corresponding key stages. Table VI shows the historical yield rates of the top and last 5 paths selected by GPSA and ANOVA, respectively. Note that since ANOVA only identifies 4 key stages, every path selected by ANOVA has already been utilized. On the contrary, 6 paths out of the 10 paths ranked by GPSA have not been used before. Another important fact is that, the top 5 paths selected by GPSA have higher yield rates than those selected by ANOVA; and the last 5 paths identified by GPSA have lower yield rates than those chosen by ANOVA. One possible explanation is that ANOVA misses some key stages so that it cannot separate the good paths from the bad ones, which further highlights the importance of GOGA.

VI. Simulation Study

The data quantity requirement of the proposed algorithm is defined through a simulation study, by considering the

TABLE VI
COMPARISON OF GPSA & ANOVA RESULTS

	GPSA	ANOVA
	$\bar{\mathbf{Y}}_{p_q}$ / $\hat{\mathbf{Y}}_{p_q}$	$\bar{\mathbf{Y}}_{p_q}$
Top 1	NaN / 99.991	99.10
Top 2	99.55 / 99.56	96.39
Top 3	NaN / 99.94	98.01
Top 4	100 / 99.93	98.26
Top 5	NaN / 99.9	98.98
Last 5	NaN / 87.94	94.14
Last 4	NaN / 87.11	93.25
Last 3	86.65 / 86.45	92.13
Last 2	NaN / 85.49	91.28
Last 1	83.85 / 83.85	86.61

TABLE VII
SIMULATION OF DATA QUANTITY REQUIREMENT

n	p	Number of Success
100	150	57
100	300	60
100	450	48
200	150	78
200	300	80
200	450	77
300	150	80
300	300	85
300	450	78

following regression model:

$$Y_t = \sum_{j=1}^{p} x_{tRj}\beta_j + \epsilon_t, \quad t = 1, \ldots, n,$$

where x_{tRj} is 1 or 0 depending on whether device j is used or not, $\beta_j \sim \text{Unif}(-3, 3)$ for $j = 1, \ldots, 15$, $\beta_j = 0$ for $j = 16, \ldots, p$, and ϵ_t's follow standard normal distribution. Note that p is a multiple of 3 and the i^{th} stage contains devices $3i-2$, $3i-1$, and $3i$, for $i = 1, \ldots, p/3$. To generate Y_t, one device from each stage is picked randomly. A simulation experiment is successful if one of the top 3 best paths is selected by the golden paths obtained from our procedure. More specifically, define

$$R(j_1, \ldots, j_5) = \sum_{i=1}^{5} \beta_{j_i}$$

where $j_i \in I_i = \{3i-2, 3i-1, 3i\}, i = 1, \ldots, 5$. Then the top 3 best paths are those formed by $\{x_{j_1}, \ldots, x_{j_5}\}$ with the corresponding $R(j_1, \ldots, j_5)$ belonging to the three largest values in $\{R(j_1, \ldots, j_5) : j_i \in I_i, i = 1, \ldots, 5\}$. Table VII reports the number of successes in 100 simulations of varying n and p. It can be observed that when n is larger than 200, the number of successes is nearly 80. In contrast, when n is less than 200, the number of successes can sometimes go down to around 50. The results suggest that the proposed method is more reliable when the sample size exceeds 200.

VII. Summary and Conclusion

The original KSA scheme can only identify the influence of univariate variables (i.e., single devices) on the yield, yet it cannot identify the manufacturing paths that have significant influence on the yield. To remedy this flaw, the GPSA scheme is proposed, which consists of two phases with Phase I identifying the key stages and generating all manufacturing paths of the key-stage devices; Phase II predicting their corresponding yields. Finally, all key-stages that may affect the yield are prioritized according to the yield index values of the golden paths for manufacturers' reference. The illustrative example shows that GPSA possesses promising potential in predicting the golden paths for yield enhancement.

References

[1] J. F. Halpin, *Zero Defects: A New Dimension in Quality Assurance*. New York, NY, USA: McGraw-Hill, Nov. 1966.

[2] F.-L. Chen and S.-F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. Semicond. Manuf.*, vol. 13, no. 3, pp. 366–373, Aug. 2000.

[3] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[4] F.-T. Cheng, *Industry 4.1: Intelligent Manufacturing With Zero Defects*. New York, NY, USA: Wiley, Oct. 2021.

[5] F.-T. Cheng, Y.-S. Hsieh, J.-W. Zheng, S.-M. Chen, R.-X. Xiao, and C.-Y. Lin, "A scheme of high-dimensional key-variable search algorithms for yield improvement," *IEEE Robot. Autom. Lett.*, vol. 2, no. 1, pp. 179–186, Jan. 2017.

[6] F.-T. Cheng, C.-Y. Lin, C.-F. Chen, R.-X. Xiao, J.-W. Zheng, and Y.-S. Hsieh, "Blind-stage search algorithm for the key-variable search scheme," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 1840–1847, Oct. 2017.

[7] C.-Y. Lin, Y.-M. Hsieh, F.-T. Cheng, Y.-R. Yang, and M. Adnan, "Interaction-effect search algorithm for the KSA scheme," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2778–2785, Oct. 2018.

[8] K. Ding, X. Zhang, F. T. S. Chan, C.-Y. Chan, and C. Wang, "Training a hidden Markov model-based knowledge model for autonomous manufacturing resources allocation in smart shop floors," *IEEE Access*, vol. 7, pp. 47366–47378, 2019.

[9] T. Zhou, D. Tang, H. Zhu, and L. Wang, "Reinforcement learning with composite rewards for production scheduling in a smart factory," *IEEE Access*, vol. 9, pp. 752–766, 2021.

[10] E. Yashchin, "Statistical monitoring of multi-stage processes," in *Frontiers in Statistical Quality Control 12*. Cham, Switzerland: Springer, 2018, pp. 185–209.

[11] R. Jin and K. Liu, "Multimode variation modeling and process monitoring for serial-parallel multistage manufacturing processes," *IIE Trans.*, vol. 45, no. 6, pp. 617–629, 2013.

[12] C.-H. Lee, D.-H. Lee, Y.-M. Bae, and K.-J. Kim, "Determining golden process routes in semiconductor manufacturing process for yield management," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2017, pp. 2366–2470.

[13] K. Taha, "An effective approach for identifying defective critical fabrication path," *Cogent Eng.*, vol. 6, no. 1, Jan. 2019, Art. no. 1575636.

[14] C.-H. Lee, D.-H. Lee, Y.-M. Bae, S.-H. Choi, K.-H. Kim, and K.-J. Kim, "Approach to derive golden paths based on machine sequence patterns in multistage manufacturing process," *J. Intell. Manuf.*, vol. 1, no. 1, pp. 1–17, Sep. 2020.

[15] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[16] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., B, Statist. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[17] C.-K. Ing and T. L. Lai, "A stepwise regression method and consistent model selection for high-dimensional sparse linear models," *Stat. Sinica*, vol. 21, no. 4, pp. 1473–1513, Oct. 2011.

[18] N. H. Chan, C.-K. Ing, Y. Li, and C. Y. Yau, "Threshold estimation via group orthogonal greedy algorithm," *J. Bus. Econ. Statist.*, vol. 35, no. 2, pp. 334–345, Mar. 2017.

[19] J. M. Marstrand, "The dimension of Cartesian product sets," *Math. Proc. Cambridge Phil. Soc.*, vol. 50, no. 2, pp. 198–202, Apr. 1954.

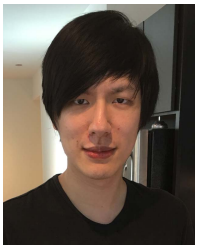
[20] H. Akaike, "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting," *Biometrika*, vol. 66, no. 2, pp. 237–242, 1979.

[21] V. N. Temlyakov, "Greedy approximation in convex optimization," *Constructive Approx.*, vol. 41, no. 2, pp. 269–296, Jan. 2015.

Ching-Kang Ing received the Ph.D. degree in statistics from the Institute of Statistics, National Tsing Hua University. He is currently a Distinguished Professor with the National Tsing Hua University. Through academia-industry cooperation programs, he has tailored statistical methods for detecting defective tools and improving yields for several leading semiconductor and display manufacturing companies. He has been interested in applying his research in high-dimensional data analysis to Taiwan's high-tech industries during recent years. His areas of expertise include mathematical statistics, model selection in time series, and high-dimensional data analysis. He has gained international recognition and has won major awards in Taiwan, including the Outstanding Research Awards of the Ministry of Science and Technology (MOST) in 2008 and 2013, the Academia Sinica Investigator Award in 2011, the Science Vanguard Research Award of the MOST in 2016, the Outstanding Scholar Award of Foundation for the Advancement of Outstanding Scholarship in 2017, and the Sun Yat-Sen Academic Award of Sun Yat-Sen Academic and Cultural Foundation in 2020.

Chin-Yi Lin was born in Taipei, Taiwan, in September 1992. He received the M.S. and Ph.D. degrees from the Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan, Taiwan, in 2017 and 2020, respectively.

He has been focused on applying his research in high-tech industries, such as ASE Group, INNOLUX, and Infineon Technologies. He is currently a Post-Doctoral Researcher with the Intelligent Manufacturing Research Center (iMRC), National Cheng Kung University. His research interests include intelligent yield management, predictive maintenance, and automatic virtual metrology.



Po-Hsiang Peng received the B.S. degree from Soochow University, Taipei, Taiwan, in 2016, and the master's degree from The University of Chicago, Chicago, IL, USA, in 2018. He is currently pursuing the Ph.D. degree in statistics with the National Tsing Hua University, Hsinchu, Taiwan.

His research interests include high-dimensional model selection and time series analysis.

Yu-Ming Hsieh was born in Penghu, Taiwan, Republic of China, in May 1983. He received the M.S. degree from the Department of Industrial Management and the Institute of Industrial Engineering and Management, National Formosa University, Yunlin, Taiwan, in 2007, and the Ph.D. degree from the Institute of Manufacturing Engineering, National Cheng Kung University, in 2015. Soon after, he graduated from the Department of Computer Science and Information Engineering, National Cheng Kung University (NCKU), in 2015. He got in the e-Manufacturing Research Center (eMRC), NCKU, serving as a Post-Doctoral Researcher. He focuses on the execution and implementation of the cooperation projects mainly in the fields, such as semiconductor automation or artificial intelligence. He also applies the research results, including the intelligent yield management (IYM) and intelligent predictive maintenance (IPM) systems in the cooperative partners to assist them to achieve the goal of lights-out factory.

Fan-Tien Cheng (Life Fellow, IEEE) received the B.S. degree from the National Cheng Kung University (NCKU), Tainan, Taiwan, in 1976, and the master's and Ph.D. degrees from The Ohio State University, Columbus, OH, USA, in 1982 and 1989, respectively. After his graduation of the B.S. degree in 1976, he started to work at the Chung-Shan Institute of Science and Technology (CSIST), Ministry of National Defense, Taoyuan, Taiwan, in 1995. He then began his academic career at the Institute of Manufacturing Engineering and the Department of Computer Science, NCKU. He has been a Chair Professor with NCKU since 2009. He has dedicated himself to applying the technologies of automation and system integration that he learned at CSIST to the Intelligent Manufacturing of both high-tech and traditional industries to exceed industry 4.0 so as to reach the goal of industry 4.1, meaning zero defects of all products. He received major honors and awards domestically and internationally, namely the 2011 Award for Outstanding Contributions in Science and Technology from the Executive Yuan; the National Science Council (NSC) Outstanding Research Awards for three times in 2006, 2009, and 2013; the National Invention and Creation Award for three times in 2011, 2012, and 2018; the 2008 University Industry Economy Contribution Award–Individual Award from Ministry of Economic Affairs (MoEA); the Outstanding Industry-University-Cooperation (IUC) Award from the Ministry of Education in 2003; the 2010 TECO Award from the TECO Technology Foundation; the 2014 Outstanding Research Award from the Pan Wen Yuan Foundation; the 2014 K.-T. Li Science and Humanities Chair-Honorary Scholar Award; the 20th Outstanding Achievement Award from the Phi Tau Phi Scholastic Honor Society in 2015; the 2013 IEEE Inaba Technical Award for Innovation Leading to Production (for contributions to the development of the AVM System); the IEEE ICRA Best Automation Paper Award for two times in 1999 and 2013; the IEEE CASE Best Application Paper Award in 2017; and the IEEE ICRA 2018 Best Paper in Automation—Finalist. He is currently in his second term (11th and 12th) of the President of the Chinese Institute of Automation Engineering. He was the Program Chair of the IEEE CASE 2014 and the Award Chair of CASE 2016 and ICRA 2017. He has been a Senior Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING since October 2017.