

IDENTIFYING ILL TOOL COMBINATIONS VIA GIBBS SAMPLER FOR SEMICONDUCTOR MANUFACTURING YIELD DIAGNOSIS

Yu-Chin Hsu
Rong-Huei Chen

Chih-Min Fan

National Taiwan University
No. 1 Sec. 4 Roosevelt Rd.
Taipei, TAIWAN, R.O.C.

Yuan Ze University
No. 135 Yuan Tung Rd.
Taoyuan, TAIWAN, R.O.C.

ABSTRACT

In semiconductor manufacturing, all up-to-date tool commonality analysis (TCA) algorithms for yield diagnosis are based on greedy search strategies, which are naturally poor in identifying combinational factors. When the root cause of product yield loss is tool combination instead of a single tool, the greedy-search-oriented TCA algorithm usually results in both high false and high miss identification rates. As the feature size of semiconductor devices continuously shrinks down, the problem induced by greedy-search-oriented TCA algorithm becomes severer because the total number of tools is getting large and product yield loss is more likely caused by a specific tool combination. To cope with the tool combination problem, a new TCA algorithm based on Gibbs Sampler, a Markov Chain Monte Carlo (MCMC) stochastic search technique, is proposed in this paper. Simulation and field data validation results show that the proposed TCA algorithm performs well in identifying the ill tool combination.

1 INTRODUCTION

In semiconductor manufacturing, there are hundreds of processing steps with multiple tools at most steps. Any tool excursion in a processing step may result in product yield loss and decrease manufacturing profit. Though various in-line inspections are established to monitor individual tools, it is not guaranteed to successfully isolate all the root causes of product yield loss (Leachman and Ding 2011). Some tool excursions relative to yield loss cannot be identified by in-line inspections but are only observable through end-of-line tests such as electrical test (E-test) and circuit probing (CP) test. Once an end-of-line yield loss event is detected, how to effectively identify a specific tool excursion from hundreds of processing steps as the root cause is a permanent challenge to a modern semiconductor manufacturing fab.

Tool commonality analysis (TCA) is an emerging topic for the effective identification of tool excursions using end-of-line yield data. Given a yield loss event with affected wafer yield and associated tool usage data, TCA iteratively conducts statistics hypotheses on individual tools in each processing step and pinpoints which tool may cause the product yield loss. The output of TCA is a ranking list of suspected tools based on their contributions to product yield loss from statistical perspective. Due to the fact that diagnosis of product yield loss with respect to tool excursions relates to several departments, the soundness of TCA has a high impact on the effectiveness of cross-department yield diagnosis, which further relates to the competitive advantage of a semiconductor manufacturing fab.

However, in industry TCA practice, the root cause of product yield loss usually does not receive top rank orders. One of the reasons is that all up-to-date TCA algorithms are based on greedy search strategies, which are naturally poor in identifying combinational factors. When the root cause of product yield loss is tool combination instead of a single tool, the greedy-search-oriented TCA algorithm usually results in both high false and high miss identification rates. Incorrect identification of root cause not only loses engineer's trust on TCA but also delays the process improvement opportunity. As the feature size of sem-

iconductor devices continuously shrinks down, the problem induced by greedy-search-oriented TCA algorithm becomes severer because the total number of tools is getting large and product yield loss is more likely caused by a specific tool combination.

To cope with the tool combination problem, a new TCA algorithm based on Gibbs Sampler, a Markov Chain Monte Carlo (MCMC) stochastic search technique, is proposed in this paper. In specific, a tool health indicator with binary value is defined for each tool to determine if it should be involved in the tool combination as root cause. Instead of conducting exhausted search on the whole space, an indirect sampling approach statistically traversing the whole space is applied in the Gibbs Sampler to find the combinations of tool health indicators that explains the product yield loss with maximum joint probability. With the Gibbs Sampler, the computation complexity is expected to reduce from $O(2^n)$ to about $O(n^2)$, where n is the number of tools.

The proposed TCA method has a direct benefit to reduce both the miss-identification and false-identification rates, whereas maintain the high computation efficiency at the same time. An indirect benefit of the proposed TCA method is that it enables the effectiveness of cross-department yield diagnosis, which further enhances the competitive advantage of a semiconductor manufacturing fab.

2 LITERATURE SURVEY

The formal term of TCA for semiconductor manufacturing yield diagnosis is first proposed by Kong (2002). He indicates that as the complexity of process and the number of process steps increase, it is definitely a very challenging task to pin point which tool is the source of problem and at which process step it occurs. Kong (2002) discusses and summarizes the critical elements of successful TCA, including sample size selection, raw data classification, statistical analysis, time series and analysis of tools with multiple entry points within the same process flow. His analysis is successfully applied on yield enhancement effort in an advanced volume manufacturing fab. However, his fundamental assumption of TCA is that a single processing equipment tool excursion resulted in yield lost of multiple semiconductor wafer lots. The case of tool combination resulted in yield loss is not addressed in Kong (2002).

McCray et al. (2005) points out that analysis of variance (ANOVA) is widely applied to TCA, which consists of conducting a series of hypothesis tests at each processing step. Their research first highlights that the intuitive conduction of ANOVA on each processing step may lead to excessive false alarms and limited sensitivity. They describe how the TCA problem can be framed as a subset selection problem, and specifically indicate that non-uniform tool usage and transition probabilities can cause these subsets from different steps to have significant overlap. They show that non-uniform tool usage and transition probabilities make subsets from different steps less statistically distinguishable. Two TCA methods based on stepwise regression to solve the distinguishable problem are therefore proposed.

Vries and Chandon (2007) further study the performance of nonparametric ANOVA using Kruskal–Wallis statistic with respect to non-uniform tool usage and transition probabilities. They show that non-uniform tool usage and transition probabilities will lead to the problem of unbalanced sample size, which leads Kruskal–Wallis statistic too conservative and gives an incorrect false-positive rate of TCA results. Alternative approximations of Wallace are investigated and shown to perform significantly better than Kruskal–Wallis statistic.

Chien et al. (2007) apply the decision tree, a data mining technique, to solve the TCA problem. They propose a framework of decision-tree-based TCA and demonstrate how it works through a real case study. Unlike the previous TCA work with focus on the effects of individual tools, the decision-tree-based TCA aims at identifying the effects of tool interactions on product yield loss.

Chen and Hong (2010) argue that decision tree will suffer the sample size depletion problem, which makes the TCA results unreliable. They clearly point out that a specific tool combination is usually the root cause of a yield loss event and is a special case of tool interaction. However, the correct tool combination as root cause may not be identified by the iterative construction of decision tree based on greedy search strategy. A comprehensive method to search for the tool combination that mostly significant to yield loss is then proposed by Chen and Hong (2010). A major assumption of their method is that indi-

vidual tools involved in the tool combination as root cause will be also significant to yield loss. Nevertheless, such an assumption may not hold in real cases.

This paper aims at further improving the performance of TCA on identifying a specific tool combination as the root cause of product yield loss. In specific, the method proposed in this paper does not need the assumption that individual tools involved in the tool combination as root cause will be also significant to yield loss.

3 TCA VIA GIBBS SAMPLER

3.1 Formulation of TCA Problem

The TCA problem is formulated as explaining the variation of end-of-line yield among a set of wafers with respect to their corresponding in-line tool usage information. Let $\mathbf{y} = \{y_1, y_2, \dots, y_K\}$ be a continuous variable following a normal distribution and denote the wafer yield measured at E-test or CP test, $\mathbf{x}_j = \{x_{1,j}, x_{2,j}, \dots, x_{K,j}\}$ be a binary indicator of tool usage to denote whether k^{th} wafer is processed by the tool j (with 1 indicating the j^{th} tool is used and 0 indicating otherwise), where $j = 1, 2, \dots, J$ and J is the total number of tools involved in TCA and $k = 1, 2, \dots, K$ and K is the total number of wafer lots.

To facilitate the formulation of tool combination as root cause of yield loss, tool health indicator γ_j is defined for each tool to indicate whether j^{th} tool is included in the combinational root cause ($\gamma_j = 1$ indicating the j^{th} tool should be included and $\gamma_j = 0$ indicating otherwise).

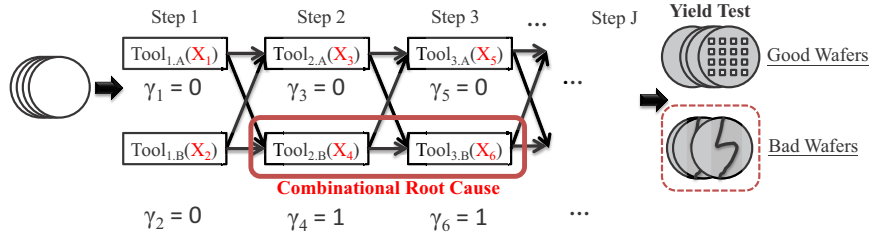


Figure 1: Relationship between health indicator and root cause

With both the definitions of usage indicator \mathbf{x}_j and health indicator γ_j for individual tools, a combinational tool usage indicator of root cause is further defined as

$$\mathbf{C} = \prod_j x_j^{\gamma_j}, j = 1, 2, \dots, J. \quad (1)$$

\mathbf{C} is a collection of situation of k wafers; a wafer with $\mathbf{C}_k = 1$ means that it is processed by the combinational root cause and is expected to be with low yield. On the contrary, a wafer with $\mathbf{C}_k = 0$ will perform the normal yield. Therefore, the wafer yield can be characterized by two distributions

$$y_k = \begin{cases} \text{Normal}(\mu_a, \sigma_a^2), & \text{if } C_k = 1 \\ \text{Normal}(\mu_b, \sigma_b^2), & \text{if } C_k = 0 \end{cases}, k = 1, 2, \dots, K \quad (2)$$

Given a yield loss event with affected wafer yield and associated tool usage data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J\}$, The objective of TCA is to find the set of tool health indicator, $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_J\}$, so that it can maximize the joint probability

$$\boldsymbol{\gamma}^* = \arg \max_{\boldsymbol{\gamma}} \Pr(\boldsymbol{\gamma} | \mathbf{X}, \mathbf{y}). \quad (3)$$

3.2 Gibbs-Sampler-Based TCA

Instead of conducting exhausted search on the whole space of γ , the technique of Gibbs Sampler (Casella and George 1992) is applied to find the combinations of tool health indicators that explains the product yield loss with maximum joint probability in (3). The Gibbs Sampler is one type of Markov Chain Monte Carlo (MCMC) stochastic search techniques. It conducts an indirect sampling approach to iteratively and statistically traverse the full solution space. The Gibbs Sampler is proved to be efficient in resembling the joint probability in high dimension space.

In each MCMC iteration, the drawing of a sample in Gibbs Sampler is based on the conditional distribution of a specific tool health indicator given that all the other tool health indicators are the same as the former MCMC iteration, which can be derived by Bayes rule as

$$\Pr(\gamma_j | \gamma_{(j)}, \mathbf{X}, \mathbf{y}) \propto L(\mathbf{y} | \gamma_j, \gamma_{(j)}, \mathbf{X}) \Pr(\gamma_j) \quad (4)$$

where $\gamma_{(j)}$ represent the set of tool health indicators that exclude γ_j ; that is, $\gamma_{(j)} = \{\gamma_a; a=1, 2, \dots, j-1, j+1, \dots, J\}$. $L(\cdot)$ is the likelihood function based on (2), and $\Pr(\gamma_j)$ is the prior probability of γ_j . Given the prior probabilities of individual γ_j 's, the iterative drawing of random samples based on (4) will statistically traverse the full solution space. The vector of random samples γ receiving the highest frequency in Gibbs Sampler process is equal to the tool health indicator γ that maximize the joint probability in (3). So the objective function in (3) is transformed to

$$\gamma^* \sim \arg \max_{\gamma} \text{FrequencyCount}(\gamma) \quad (5)$$

Note that the drawing of random samples of tool health indicator based on (4) is carried out many times and in two stages. The first stage is a warm-up period at the end of which it is assumed that the Gibbs Sampler has converged to the joint distribution of $\Pr(\gamma | \mathbf{X}, \mathbf{y})$. The second stage is an estimation period and the set of tool health indicator γ simulated during this period with maximum frequency is identified as the optimal solution to (3) and (5).

The performance of Gibbs-Sampler-Based TCA highly depends on the setting of method parameters. In the method of Gibbs-Sampler-Based TCA, only the prior probability of tool health indicator, $\Pr(\gamma_j)$, needs to be pre-specified. Note that the main purpose of prior probability in Gibbs-Sampler-Based TCA is to calculate the conditional probability in (4) so that it can facilitate the MCMC simulation process. The setting of prior probability is not necessary to incorporate the domain knowledge. Therefore, in the real applications, it is assumed that no domain knowledge is involved, i.e. the prior probability of tool health indicator, $\Pr(\gamma_j)$, is set as 0.5. In this way, the calculation of conditional probability in (4) is just based on the likelihood function.

3.3 Gibbs Sampler Transition Mechanism

The procedure of Gibbs-Sampler-based TCA is to update γ_j sequentially for every j and to apply Bayes's Rule in (4). Each of the J tool health indicators is sampled according to the full conditional distribution of γ_j given $\gamma_{(j)}$, yield-loss event data \mathbf{y} and tool usage data \mathbf{X} . The transition of each γ_j from current value to future value is to take a draw from the conditional distribution, that is, the posterior probability.

Hence, the posterior distribution of γ_j given $\gamma_{(j)}$, \mathbf{X} , \mathbf{y} can be obtained through

$$L(\mathbf{y} | \gamma_j = 0, \gamma_{(j)}, \mathbf{X}) \Pr(\gamma_j = 0) = f_0, L(\mathbf{y} | \gamma_j = 1, \gamma_{(j)}, \mathbf{X}) \Pr(\gamma_j = 1) = f_1 \quad (6)$$

based on (1), (2) and the posterior distribution (Bernoulli Distribution) for j^{th} health indicator is obtained

$$\Pr(\gamma_j = 0 | \gamma_{(j)}, \mathbf{X}, \mathbf{y}) = \frac{f_0}{f_0 + f_1}, \Pr(\gamma_j = 1 | \gamma_{(j)}, \mathbf{X}, \mathbf{y}) = \frac{f_1}{f_0 + f_1}. \quad (7)$$

According to the posterior distribution, a uniform distribution random variable will be drawn to decide the transition of current health indicator. The procedure for Gibbs Sampler for each iteration is below.

```

Class GibbsSamplerProcess{
  Randomly Select j from S = {1, 2, 3, ..., J}
  Compute posterior distribution for jth health indicator by (7)
  Generate random variable T  $\leftarrow$  U(0, threshold)
   $\begin{cases} \gamma_j = 0, & \text{if } T < \Pr(\gamma_j = 0 | \gamma_{(-j)}, \mathbf{X}, \mathbf{y}) \\ \gamma_j = 1, & \text{otherwise} \end{cases}$ 
  Record  $\gamma$  and Remove j from S
  Continue until S is empty};

```

Based on the Gibbs Sampler Process, the overall algorithm is developed as below

```

Class MainFunction{
  Input yield-loss event data and corresponding tool usage data
  Create tool health indicators ( $\gamma = \{\gamma_j; j=1,2,...,J\}$ )
  Do while  $\gamma$  is not convergence
    GibbsSamplerProcess
  End
  For each i, i = 1, 2, 3, ..., I
    GibbsSamplerProcess
  Go to next iteration until i = I
  Provide  $\gamma$  that sorted by frequency count};

```

4 TRANSITION MECHANISM

Markov Chain Monte Carlo is a well-known stochastic search technique; however, there is no clear definition on state transition mechanism. Acceptance probability is proposed in most paper to determine the transition from state to state ($\gamma^{(t)} \rightarrow \gamma^{(t+1)}$ in this paper), but the definition is vague and non-intuitive.

In this paper, we try to deal with the combinational root cause and confounding tool. Confounding tool is the tool with similar tool usage data as root cause; if drawing the transition random number from U(0, 1), there is half chance that confounding tool will be included in γ and receive higher frequency count for the reason that $\Pr(\gamma_j = 0 | \gamma_{(-j)}, \mathbf{X}, \mathbf{y})$ will close to 0.5. However, there are some problem when drawing transition random number from U(0, 1) to determine whether the x is included under $\gamma_j = 0 \ \forall j$.

4.1 Problem of Random Variable (T) in Transition Mechanism

The intuitive idea is to draw the transition random number from U(0, 1) and successfully detect problematic tools under yield-loss event; however, some normal tools is detected to be problematic tool under no yield-loss event. To clarify to situation, the simulated under no yield-loss event is presented as follows.

Figure 2 shows two facts that $\Pr(\gamma_j = 0 | \gamma_{(-j)}, \mathbf{X}, \mathbf{y})$ will be 0.5 under normal case and that permutation effect exists to produce extreme case; that is, yield data will be grouped and test significantly by chance.

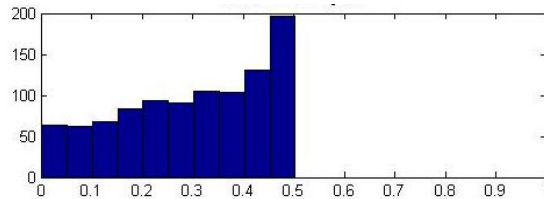


Figure 2: Histogram for the posterior probability under no yield-loss event via likelihood

If we want to exclude normal case with $\Pr(\gamma_j = 0 | \gamma_{(-j)}, \mathbf{X}, \mathbf{y}) = 0.5$ and draw Random Variable T with U(0,0.5), there is still about half of the chance that normal tools will be detected as problematic tools due

to the permutation condition. Thus, smaller interval of uniform distribution ($U(0, \text{threshold})$) for Transition Random Variable (T) is desired to decrease the false identification rate under no yield-loss event.

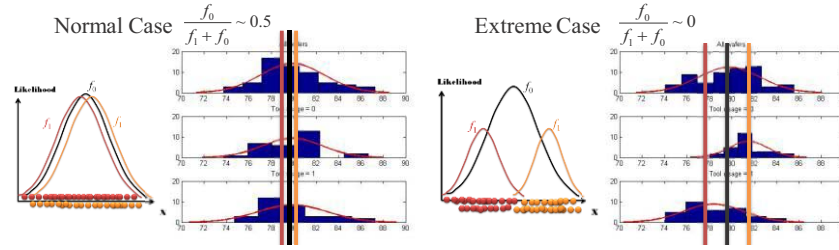


Figure 3: Extreme and normal case for posterior probability $\Pr(\gamma_j = 0 | \gamma_{(j)}, \mathbf{X}, \mathbf{y})$

In order to observe whether the given uniform distribution ($U(0, \text{threshold})$) will provide the lower false identification under no yield-loss event, a simulation is applied to test the detecting capability when the corresponding tool usage data is Root cause $\mathbf{x}_{\text{RootCause}}$ without combinational effect given yield-loss event data \mathbf{y} and shift size from 0σ to 4σ .

Table 1: Type I/II error when computing the post prob. via likelihood

Likelihood Approach: Power Curve Result (Threshold vs. Shift Size) Replication = 1000									
Threshold	Shift Size (Sigma)								
	0	0.5	1	1.5	2	2.5	3	3.5	4
0.001	0.001	0.034	0.413	0.939	0.998	1	1	1	1
0.003	0.002	0.064	0.57	0.975	0.999	1	1	1	1
0.005	0.005	0.088	0.633	0.976	1	1	1	1	1
0.007	0.005	0.1	0.688	0.982	1	1	1	1	1
0.009	0.007	0.127	0.709	0.981	1	1	1	1	1
0.01	0.006	0.109	0.715	0.988	1	1	1	1	1
0.03	0.013	0.212	0.827	0.993	1	1	1	1	1
0.05	0.022	0.285	0.857	0.994	1	1	1	1	1
0.07	0.036	0.331	0.893	0.996	1	1	1	1	1
0.09	0.045	0.368	0.893	0.998	1	1	1	1	1
0.1	0.073	0.376	0.926	0.998	1	1	1	1	1
0.2	0.117	0.516	0.945	1	1	1	1	1	1
0.3	0.204	0.59	0.962	1	1	1	1	1	1
0.4	0.297	0.722	0.969	1	1	1	1	1	1
0.5	0.382	0.729	0.964	0.999	1	1	1	1	1
0.6	0.485	0.763	0.989	1	1	1	1	1	1
0.7	0.581	0.825	0.981	1	1	1	1	1	1
0.8	0.613	0.817	0.986	0.999	1	1	1	1	1
0.9	0.618	0.873	0.992	1	1	1	1	1	1
1	0.687	0.862	0.987	1	1	1	1	1	1
Interpretation	Type I Error	1 - Type II Error (Testing Power)							

The simulation result is presented in table 1; the result shows the probability of detecting \mathbf{x} as root cause with respect to \mathbf{y} ; thus, the probability represents the Type I Error for non-yield-loss data (shift size $= 0\sigma$) and the Testing Power for yield-loss data (shift size $\neq 0\sigma$).

The Type I Error of setting 0.05 as threshold is 0.022; that is, there are 22 false identifications out of 1000 detections. The performance is better than setting 0.5 as threshold. If we expect further lower Type I Error and pick 0.01 as threshold, the testing power for $\mathbf{x}_{\text{RootCause}}$ will be reduce to 0.109; that is, it become harder to include $\mathbf{x}_{\text{RootCause}}$ under yield-loss event with smaller shift size. Hence, engineers might not set 0.01 as threshold owing to the consideration of small shift size.

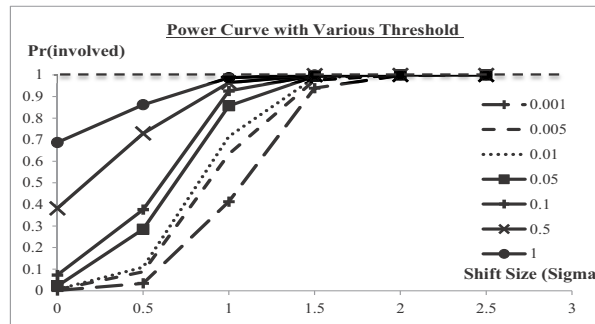


Figure 4: Power Curve when computing the post probability

The detecting power is close to 100% when the shift size of single problematic tool is 1.5σ . Compared to the 7 threshold candidate, 1, 0.5 and 0.1 are eliminated due to the higher Type I Error and 0.01 is rejected because of the lower testing power. Therefore, ideal setting is to draw the transition random number from $U(0, 0.05)$ for all the $\gamma_j = 0$ and from $U(0,1)$ for at least a $\gamma_j = 1$.

5 SIMULATION CASE STUDY

5.1 Design of Simulation Cases and Evaluation Metrics

Without loss of generality, the yield loss event is defined as yield shift-down. In specific, the end-of-line yield data is modeled as the mixture of two normal distributions in (2). Therefore, a yield shift-down case is characterized by four parameters (C, J, K, S) , where C is the combinational tool usage indicator C in (1) and controls the mixture of two normal distributions, J is the total number of tools involved in TCA, K is the total number of wafers involved in TCA, and S is the ratio of shift-down size to noise.

Most literatures focus on study the TCA performance with respect to the last three parameters (J, K, S) . To the authors' best knowledge, no study addresses the relationship between TCA performance and the combinational tool usage indicator C . However, the combinational tool usage indicator C performs three features that have high impacts on the identification rate of TCA.

(F1). Tool-related issue

Suppose the root cause of a yield loss event is characterized by the combinational tool usage indicator C . A confounding tool is a tool j not involved in the root cause, but with its tool usage indicator \mathbf{x}_j highly correlated to the combinational tool usage indicator C . Such a concept is implicitly introduced in McCray et al. (2005) by saying that non-uniform tool usage and transition probabilities can cause the subsets of wafers from different tools to have significant overlap, i.e. highly correlated. Note that the confounding tool may perform significant contributions to yield variation and be falsely identified as the root cause. Pure data driven inference methods cannot completely solve the confounding problem.

(F2). Wafer-related issue

Small proportion of low-yield wafers and incomplete combinational root cause condition will result in unbalanced sample size problem. Vries and Chandon (2007) study the impact of unbalanced sample size on TCA and propose to use Wallace approximation for nonparametric ANOVA applications. However, their study just focus on individual tools but not tool combinations. When the root cause of a yield loss event is a specific tool combination, the small proportion of low-yield wafers may result in poor performance of Greedy-Search-Oriented TCA. In Chen and Hong (2010), it is claimed that the most significant tool combination may not be found by Greedy-Search-Oriented TCA because of incomplete combinational root cause condition. For example, suppose $\mathbf{x}_1 \times \mathbf{x}_2$ has the most significant combined effect. If good wafers passed part of the combinational root cause, \mathbf{x}_1 and \mathbf{x}_2 individually may not be the most significant factor. Therefore, \mathbf{x}_1 and \mathbf{x}_2 may never be identified by Greedy-Search-Oriented TCA. However, Chen and Hong (2010) did not explicitly state that the poor performance of Greedy-Search-Oriented TCA on identifying tool combinations is due to incomplete combinational root cause condition.

(F3). Yield-related issue:

Yield data can be divided into three parts – Low, Normal and High yield data which represent the wafer passing through ill, normal and golden tool respectively. The yield difference among Low, Normal and High yield reflect the effect of ill, normal and golden tool; in the large-scaled production line, the effect of tools is diverse, that is, difference among Low, Normal and High yield is diverse as well. Such phenomena affect the statistical testing power while the difference is small so that the significance of individual combination root cause will be reduced and be ignored. Here, we focus on the ill and normal tool effect in simulation case study.

As a result, the characterization of a yield shift-down case is extended to six parameters (C, J, K, S, CC, CP), where the new parameter CC is a tool usage indicator \mathbf{x}_j to denote if there is a confounding tool j which is not involved in the root cause but highly correlated with the combinational tool usage indicator C , and CL is another new parameter to denote the proportion of low yield wafers.

With the six parameters defined above, four cases of simulation are designed to demonstrate the advantages of Gibbs-Sampler-Based TCA over Greedy-Search-Oriented TCA. Table 1 lists two cases with corresponding features and parameters. Note that in each of the four cases, the root cause of yield shift-down is identically set as the combination of tool 1 and tool 2, i.e. the combinational tool usage indicator C is equal to $\mathbf{x}_1 \times \mathbf{x}_2$.

Table 2: Design of Simulation Cases

Case ID	Features	C	J	K	S	CC	CP	Note
1	F1	$\mathbf{x}_1 * \mathbf{x}_2$	40	100	3	\mathbf{x}_4	0.5	\mathbf{x}_4 is highly correlated with $\mathbf{x}_1 * \mathbf{x}_2$; \mathbf{x}_4 is more significant than $\mathbf{x}_1 * \mathbf{x}_2$
2	F2	$\mathbf{x}_1 * \mathbf{x}_2$	40	100	3	none	0.1	\mathbf{x}_2 has a small probability to be different from \mathbf{x}_1

The implementation of Gibbs-Sampler-Based TCA adopts 8000 MCMC iterations for the estimation period. Furthermore, for each case in Table 1, TCA will generate four evaluation metrics as follows.

- Frequency: count the tool health indicator γ simulated during the estimation period
- Rank List: rank the specific tool health indicator γ according to the frequency.
- Log Likelihood: the likelihood of the corresponding tool health indicator γ
- P-value: ANOVA test of the yield grouping by combinational tool usage indicator C

5.2 Study of Confounding Effect in Case 1

As designed in Table 1, the generation of confounding effect with tool 4 is by simulating a tool usage indicator \mathbf{x}_4 highly correlated with $\mathbf{x}_1 \times \mathbf{x}_2$. The model for generated \mathbf{x}_4 is based on a random variable $CF \sim U(0, 1)$ and a threshold $p, 0 < p < 1$, so that

$$\mathbf{x}_4 = \mathbf{x}_1 \times \mathbf{x}_2, \text{ if } CF < p; \quad \mathbf{x}_4 = 1 - (\mathbf{x}_1 \times \mathbf{x}_2), \text{ otherwise} \quad (8)$$

The results of case 1 by greedy-search-oriented TCA and Gibbs-Sampler-Based TCA are demonstrated in Tables 2 and 3 respectively. It can be found that the confounding factor \mathbf{x}_4 will be falsely identified when applying the greedy-search-oriented TCA. Due to the fact that \mathbf{x}_4 is highly correlated with the root cause $\mathbf{x}_1 \times \mathbf{x}_2$, the greedy-search-oriented TCA will select \mathbf{x}_4 simply because of its smallest p-value. But the smallest p-value doesn't mean \mathbf{x}_4 is the root cause of the system.

On the contrary, the Gibbs-Sampler-Based TCA will generate three candidates of tool combinations as potential root causes. Comparing among the three candidates of tool combinations in Table 3, we can observe that \mathbf{x}_2 simultaneously appears in the three candidates of tool combinations, while \mathbf{x}_1 and \mathbf{x}_4 compete for each other. So we can conclude \mathbf{x}_2 is involved in the root cause while \mathbf{x}_1 and \mathbf{x}_4 may be confounded with each other. It demonstrates that Gibbs-Sampler-Based TCA not only highlights the root cause but is also capable of informing engineers the possible confounding effects.

Table 3: The result of case 1 by greedy-search-oriented TCA

Iteration	Selected Factors	p-value
1	\mathbf{x}_4	1.32e-07
2	$\mathbf{x}_4, \mathbf{x}_2$	1.03e-12
3	$\mathbf{x}_4, \mathbf{x}_2, \mathbf{x}_1$	3.54e-13

Table 4: The result of case 1 by Gibbs-Sampler-Based TCA algorithm

Ranking	Frequency	Log Likelihood	Candidate Factors	p-value
1	3795/8000	-150.2472	$\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_4$	3.54e-013
2	2333/8000	-151.2627	$\mathbf{x}_1 \mathbf{x}_2$	9.88e-013
3	1872/8000	-151.3123	$\mathbf{x}_2 \mathbf{x}_4$	1.03e-012

5.3 Study of The Effect Due to Small Proportion of Low-Yield Wafers in Case 2

The generation of small proportion of low-yield wafers is by manipulating the proportion of tool usages involved in the root cause. Let \mathbf{x}_1 and \mathbf{x}_2 be the root causes of the combination effect. The generation of \mathbf{x}_1 and \mathbf{x}_2 is based on a random variable $U \sim U(0, 1)$ so that

$$\begin{aligned} \mathbf{x}_2 &= 1 - \mathbf{x}_1, \text{ if } U < CP, \text{ when } \mathbf{x}_1 = 1; & \mathbf{x}_2 &= 1 - \mathbf{x}_1, \text{ if } U < CP, \text{ when } \mathbf{x}_1 = 0; \\ \mathbf{x}_2 &= \mathbf{x}_1, \text{ otherwise, when } \mathbf{x}_1 = 1; & \mathbf{x}_2 &= \mathbf{x}_1, \text{ otherwise, when } \mathbf{x}_1 = 0 \end{aligned} \quad (9)$$

Based on (7), there is small proportion of low-yield (both \mathbf{x}_1 and \mathbf{x}_2 equal to 1) when the threshold value, CP, is with a small value. The results of case 2 by greedy-search-oriented TCA and Gibbs-Sampler-Based TCA are demonstrated in Tables 4 and 5 respectively. It can be observed that the greedy-search-oriented TCA algorithm will find wrong tool with insignificant p-value because of the small combination set (passing the specific steps/tools usage set sequentially) of the root cause. The phenomena of proportion of the low-yield wafers results in the insignificant of p-value when testing the root cause individually.

On the contrary, the Gibbs-Sampler-Based TCA successfully identifies the root cause. It is because the Gibbs-Sampler-Based TCA considers the various tools integrally rather than individually. In this case, the tools involved in root cause are individually insignificant but is significant when the tool combination is tested simultaneously.

Table 5: The result of case 2 by greedy-search-oriented TCA algorithm

Iteration	Selected Factors	p-value
1	\mathbf{x}_4	0.235

Table 6: The result of case 2 by Gibbs-Sampler-Based TCA algorithm

Ranking	Frequency	Log Likelihood	Candidate Factors	p-value
1	8000/8000	-134.8780	$\mathbf{x}_1 \mathbf{x}_2$	1.12e-005

6 SEMICONDUCTOR MANUFACTURING CASE STUDY

6.1 Real data characteristics and the algorithm initial setting

Gibbs-Sampler-Based TCA algorithm is validated through the following actual semiconductor manufacturing yield case. In this case, engineers suspect there are combination effects in the observed data.

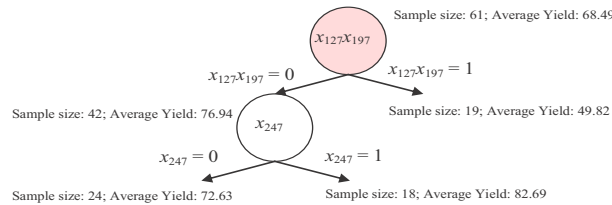


Figure 5: Tree structure for combinational effect identified by SERT

The objective, in this case, is to detect tool combination on the circuit probe (CP) yield by the historical lot processing dataset. The observed dataset is provided the response of CP yield with processed tools categorical indicators (binary variables x_j); the dataset characteristic consists 252 binary variables x_j and 61 lots with average CP yield 68.50% (range from 89.48% to 23.59%). According to the analysis result of Chen and Hong (2010), the tree structure is shown in figure 5.

The most popular TCA such as CART, Greedy-Search-Oriented TCA algorithm and SERT algorithm are widely used but with hidden drawbacks while implementing. CART introduced by Breiman (Breiman et al. 1984) is not suitable to deal with combinational effect among the large-scale production line.

Greedy-Search-Oriented TCA algorithm is the most intuitive way to dispose combinational effect based on the assumption that tools in combinational root cause will be significant individually; however, one-factor-at-a-time analysis will lead the false identification when the idea is violated.

Although SERT algorithm successively solves combinational effect and improves the drawback of one-factor-at-a-time analysis by Benjamini Hochberg Procedure (Benjamini et al. 1995; Benjamini et al. 2001), the performance will be affected by the setting candidate threshold and (F1), (F2) and (F3).

To implement and compare the performance among algorithms, the initial setting is below:

S1. Gibbs-Sampler-Based TCA algorithm

γ_{initial} : all the health indicator γ are 0; $\text{Prior}_{\text{initial}}$: the prior probability of health indicator are both 0.5

S2. Greedy-Search-Oriented TCA algorithm and CART

SignificanceThreshold: the significance of initial x_j must lower than 0.05.

S3. Sample-Efficient Regression Tree algorithm

CandidateThreshold: set the threshold of Benjamini Hochberg Procedure as 0.15.

6.2 Real Data Manipulation Method

The case 1 and 2 are created by adding additional wafer data to test the robustness of TCA. 22 wafer data are added in case 1 to create smaller shift size (27.11 to 12.55) between normal and ill tools. In case 2, 42 wafer data are added to obtain higher unbalance tool usage index (0.475 to 0.689). The unbalance tool usage index is defined as the number of passing part of root cause over the total number of wafers (e.g. unbalance tool usage index of case 1 = 29 / 61).

Table 7: Overall manipulated data characteristic

Case ID	Sample Size	Number of tool	Average Yield		Shift Size	Unbalanced Tool Usage Index
			normal	ill		
0 (Real Data)	61	248	76.94	49.83	27.11	0.475
1 (Artificial Data)	83	248	76.13	63.58	12.55	0.614
2 (Artificial Data)	103	248	76.84	62.07	14.77	0.689

The comprehensive analysis result is summarized in the table 8; some analysis properties and result is provided to evaluate the performance of each algorithm and the analysis result is shown in table 8.

E1. Significance: the p-value by KW test with Wallace approximation of the detected tool combination

E2. Complexity: for * is number of KW with Wallace approximation and for ** is number of pattern

E3. Tool combination: the tool index of the detected tool combination

Table 8: Analysis result for three cases

Method	CART*			Greedy Search*			SERT*			Gibbs-Sampler**		
Case ID	0	1	2	0	1	2	0	1	2	0	1	2
Significance	1.20E-08	3.26E-03	8.91E-04	8.20E-12	5.54E-05	7.50E-05	8.20E-12	5.54E-05	7.50E-05	8.20E-12	8.67E-06	3.33E-07
Complexity	247	247	247	753	986	1230	5456	992	1240	43	988	816
Tool Combination	Low			Medium			Medium			Low ~ Medium		
	X197	X130	X226	1. X197; 2. X127	1. X130 2. X85 3. X179	1. X226 2. X119 3. X10 4. X49	1. X197; 2. X127	1. X130 2. X85 3. X179	1. X226 2. X119 3. X10 4. X49	1. X197; 2. X127	1. X197; 2. X127	1. X197; 2. X127
Result	Fail in three cases			Fail in case of small shift size and unbalance tool usage						Success in all three cases		

For complexity aspect, CART identifies the most significant tool by go through all the tool one by one, so the number of statistical test is the total number of tool. For each suspected tool, Greedy Search & SERT attempt to improve the significance by adding additional tool. Thus, the complexity is increased with the number of suspected tool and the number of tool involving in the tool combination.

The pattern γ , in Gibbs sampler, is determined by comparing random sample with posterior probability. Once the pattern is root cause, it hard to change to another pattern when the significance of root cause is relatively large. Thus, the number of pattern become larger for the reason that the shift size is shrink from case 0 to 2. But the complexity of Gibbs sampler is relative small. With the increasing unbalance tool usage index, CART, Greedy Search and SERT algorithm are affected and induce the false identification. On the contrary, Gibbs sampler is more robust.

Table 9: Detection for confounding tool

Case ID	Frequency Count	Tool Combination	p-value	Case ID	Frequency Count	Tool Combination	p-value	Case ID	Frequency Count	Tool Combination	p-value
0	520	X22 X37 X127 X197	8.20E-12	1	59	X22 X127 X197 X231	8.67E-06	2	276	X22 X37 X127 X197 X231	3.33E-07
	511	X127 X197	8.20E-12		54	X37 X160 X197 X231	7.23E-05		227	X22 X37 X127 X197	3.33E-07
	370	X22 X127 X197 X231 X237	8.20E-12		51	X22 X37 X127 X197 X231	8.67E-06		220	X127 X197 X231	3.33E-07
	317	X37 X127 X197	8.20E-12		41	X37 X127 X197	8.67E-06		220	X127 X197	3.33E-07
	255	X22 X127 X197	8.20E-12		41	X160 X197	7.23E-05		217	X37 X127 X197 X231	3.33E-07
	235	X127 X197 X237	8.20E-12		38	X22 X160 X197	7.23E-05		193	X22 X127 X197 X231	3.33E-07
	204	X22 X37 X127 X197 X231	8.20E-12		32	X127 X197	8.67E-06		174	X22 X127 X197	3.33E-07
	71	X22 X127 X197 X231	8.20E-12		32	X22 X37 X160 X197 X231	7.23E-05		157	X37 X127 X197	3.33E-07
	29	X22 X127 X197 X237	8.20E-12		30	X37 X160 X197	7.23E-05		146	X22 X37 X127 X160 X197	0.000128
	302	X22 X127 X160 X197	0.000186		25	X37 X92 X160 X197 X231	6.91E-05		113	X22 X37 X127 X160 X197 X231	0.000128
1	286	X22 X37 X127 X160 X197	0.000186	15	23	X92 X160 X197	6.91E-05	10	101	X127 X160 X197 X231	0.000128
	243	X37 X127 X160 X197	0.000186		22	X37 X92 X160 X197	6.91E-05		94	X37 X127 X160 X197	0.000128
	211	X127 X160 X197	0.000186		21	X127 X197 X231	8.67E-06		81	X37 X127 X160 X197 X231	0.000128
	202	X22 X37 X127 X160 X197 X231	0.000186		21	X22 X127 X197	8.67E-06		77	X22 X127 X160 X197 X231	0.000128
	197	X22 X127 X160 X197 X231	0.000186		20	X22 X37 X92 X160 X197	6.91E-05		74	X22 X127 X160 X197	0.000128
	165	X37 X127 X160 X197 X231	0.000186		16	X92 X127 X160 X197 X231	0.002339		72	X127 X160 X197	0.000128
	124	X127 X160 X197 X231	0.000186		16	X37 X92 X127 X160 X197	0.002339		14	X92 X160 X197 X231	0.001575
	69	X37 X127 X197 X231	8.67E-06		15	X22 X37 X127 X197	8.67E-06		10	X22 X37 X92 X160 X197 X231	0.001575

The confounding tool cannot be highlighted by CART, Greedy search and SERT. However, the analysis result of Gibbs sampler highlights confounding tool among three cases in table 9.

Because of (F1), (F2) and (F3), ideal threshold setting is difficult. Once an inappropriate threshold is set such as significance threshold in CART, Greedy search and candidate threshold of Benjamini Hochberg Procedure in SERT, the analysis result will receive high false/miss identification rate. Gibbs sampler not only ease the effect of unbalance tool usage but also eliminate the requirement of threshold.

7 CONCLUSIONS

When the root cause of product yield loss is tool combination instead of a single tool, the greedy-search-oriented TCA algorithm usually results in both high false and high miss identification rates. Although SERT algorithm attempts to overcome the false and miss identification rate, confounding tool and incomplete combinational root cause condition will still affect performance. To cope with combinational effect, a new TCA algorithm based on Gibbs Sampler, a Markov Chain Monte Carlo stochastic search technique, is proposed. The preliminary study showed that Gibbs-sampler-based TCA is superior to the CART, greedy-search-oriented TCA algorithm and SERT algorithm in identifying the tool combination, especially when the incomplete combinational root cause condition occurs or there are other normal but confounding tools processing most of low-yield wafers. We will improve the robustness of Gibbs-sampler-based TCA by relaxing the normal distribution assumption in the near future.

In semiconductor manufacturing with intensive cross-departmental cooperation, false identification of root cause will destroy the harmony in the organization because of result-based accountability; thus, false identification not only loses engineer's trust on TCA but also induces cross-departmental conflicts. Besides, yield improvement is important and urgent; miss identification will receive longer detection time and delay the process improvement opportunity. Gibbs-sampler-based TCA provides more robust analysis result and diminish the difficulty of setting threshold; hence, some difficulties of yield analysis in industrial practices is overcome efficiently.

REFERENCES

- Benjamini, Y., and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.
- Breiman, L., and Friedman, J., and Olshen, R., and Stone, C. 1984. Classification and Regression Trees, Wadsworth. New York.
- Casella, G., and George, E. I. 1992. Explaining the Gibbs sampler. *American Statistician*, 167-174.
- Chen, A., and Hong, A. 2010. Sample-Efficient Regression Trees (SERT) for Semiconductor Yield Loss Analysis. *Semiconductor Manufacturing, IEEE Transactions on*, 23(3), 358-369.
- Chien, C. F., and Wang, W. C., and Cheng, J. C. 2007. Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192-198.
- de Vries, D. K., and Chandon, Y. 2007. On the False-Positive Rate of Statistical Equipment Comparisons Based on the Kruskal-Wallis H Statistic. *Semiconductor Manufacturing, IEEE Transactions on*, 20(3), 286-292.
- Fan, C. M., and Guo, R. S., and Chen, A., and Hsu, K. C., and Wei, C. S. 2001. *Data mining and fault diagnosis based on wafer acceptance test data and in-line manufacturing data*.
- George, E. I., and McCulloch, R. E. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 881-889.
- Kong, G. 2002. *Tool commonality analysis for yield enhancement*.
- Langford, R. E., and Hsu, G., and Sun, C. 2000. *The identification and analysis of systematic yield loss*.
- Leachman, R. C., and Ding, S. 2011. Excursion yield loss and cycle time reduction in semiconductor manufacturing. *Automation Science and Engineering, IEEE Transactions on*, 8(1), 112-117.
- McCray, A. T., and McNames, J., and Abercrombie, D. 2005. Locating disturbances in semiconductor manufacturing with stepwise regression. *Semiconductor Manufacturing, IEEE Transactions on*, 18(3), 458-468.
- Nicholson, M. 2003. System and method for finding an operation/tool combination that causes integrated failure in a semiconductor fabrication facility: Google Patents.
- Pierre, E. R. S., and Tuv, E., and Borisov, A. 2008. *Spatial patterns in sort wafer maps and identifying fab tool commonalities*.
- Tu, K. K. W., and Lee, J. C., and Lu, H. H. S. 2009. A Novel Statistical Method for Automatically Partitioning Tools According to Engineers' Tolerance Control in Process Improvement. *Semiconductor Manufacturing, IEEE Transactions on*, 22(3), 373-380.

AUTHOR BIOGRAPHIES

CHIH-MIN FAN is an Assistant Professor in the Department of Industrial Engineering and Management at Yuan Ze University, Taiwan. He received a master's degree in Electrical Engineering and a Ph.D. in the same subject from National Taiwan University, Taiwan. His current research and teaching interests are in statistical data analysis, knowledge representation and reasoning, engineering knowledge mining, and on-demand knowledge service with applications to engineering chain collaboration in semiconductor manufacturing industry.

YU-CHIN HSU received the B.S. degree in the Department of Industrial Engineering and Management at Yuan Ze University, Taiwan. He is currently a master student in the Graduate Institute of Industrial Engineering at National Taiwan University, Taiwan. His research interest is on the statistical data mining and semiconductor yield analysis, especially in the case of combinational effect in yield-loss event.

Rong-Huei Chen received a master's degree in Electrical Engineering and He is currently a Ph.D. student in the same subject from National Taiwan University, Taiwan. His research interest is on semiconductor yield analysis and related issue.