A Novel Statistical Method for Automatically Partitioning Tools According to Engineers' Tolerance Control in Process Improvement

Kevin Kai-Wen Tu, Jack Chao-sheng Lee, and Henry Horng-Shing Lu

***Abstract*—In the semiconductor industry, tool comparison is a key task in yield or product quality enhancements. We develop a new method to automatically partition tools. The new method is called tolerance control partitioning (TCP). The advantages of TCP include 1) taking into account of unbalanced tool usage in manufacturing processes; 2) further partitioning these tools into several homogenous groups by related metrology results instead of detecting only the significant difference; and 3) partitioning these tools according to engineers' tolerance controls to avoid too many groups with small differences. TCP also could be applied in all similar cases such as experimental recipe or material comparisons. Therefore, using TCP, engineers could speed up yield or product quality ramping.**

Two simulation cases illustrate the advantages of TCP method. We also applied TCP to two real cases for yield and Cp/Cpk enhancement in the semiconductor industry. The results confirm the practical feasibility of this method.

***Index Terms*—APC, Bayesian fit, CART, C_p, C_{pk}, data mining, process capability, reversible jump Markov chain Monte Carlo, yield enhancement.**

I. Introduction

THE IMPORTANCE of semiconductor technology in today's world cannot be exaggerated. Semiconductor devices are essential components in all electronic products. However, since building a modern wafer fabrication facility costs approximately \$3 billion, rapid yield ramp to volume production is becoming an increasingly important source of competitive advantage in the ultra-competitive world of semiconductor manufacturing [1].

Tool comparison is one key task for engineers in facilitating yield ramping. The tools could be compared using metrology results from processed lots. The metrology results could include physical or electrical data, such as film thickness, film uniformity, critical dimension, overlay, defect particle count, voltage, current, and wafer sort. Because engineers must analyze massive amounts of data in a very time-consuming fashion [2], effective and efficient analysis in comparing tools is critical to rapidly ramping up yield.

ANOVA and Kruskal–Wallis [3] tests are the two most common methodologies used in statistical analysis [2]. These tests can show engineers where statistical differences exist. However, after finding the significant differences among tools, engineers still need to partition these tools into several homogenous groups to identify the best or the most problematic groups of tools to enhance product quality or to reject the worst tools, respectively [4]–[6]. Other disadvantages of using these statistical tests are that 1) nonuniform usage in most tools results in unbalanced data and diminishes the accuracy of these statistical methods [7] and 2) statistical significances may detect small differences as large sample sizes [3], [7].

Data mining [8], [9] using a classification and regression tree (CART) [10] or neural networks [11] are other popular methods. Many commercial engineering data analysis tools use these functions for yield enhancements. They help engineers partition tools into several homogenous groups and identify the best or worst groups of tools. However, it is difficult for engineers to set up related parameters for the CART or neural network methods with respect to engineering tolerances, since different engineers' decisions may result in different partitions [10].

To sum up, there are three major challenges in such tasks as follows: 1) to take into account unbalanced tool usage in manufacturing processes; 2) to further partition these tools into several homogenous groups by related metrology results instead of detecting only the significant differences; 3) to partition these tools according to engineers' tolerance controls to avoid too many groups with small differences.

The Bayesian approach is one method that solves the unbalanced data issue. Chaloner [12] got better estimations of variance components by using the Bayesian approach for unbalanced data cases. However, when different partitioning results are shown, the partition models are different. Therefore, our problem is related to model determination. The reversible jump Markov chain Monte Carlo (RJMCMC) proposed by Green [13] is a new and powerful method to deal with Bayesian model determination problems. Applications of RJMCMC include change point problems and factorial experiments [13], mixture problems [14], and so forth. Nobile and Green [15]

Manuscript received August 02, 2007; revised July 06, 2008; accepted July 08, 2008. First published July 07, 2009; current version published August 05, 2009. This work was supported in part by grants from the National Science Council of Taiwan.

K. K.-W Tu is with Macronix International Co., LTD., Hsinchu, Taiwan, and also with the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan (e-mail: kevintu@mxic.com.tw).

J. C. Lee (deceased) was with Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan.

H. H.-S. Lu is with the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan (e-mail: hslu@stat.nctu.edu.tw).

Digital Object Identifier 10.1109/TSM.2009.2025812

also applied this method in a factorial experiment by using mixture modeling. Bergeret and Gall [16] applied RJMCMC to detect the change point of a yield trend in semiconductor manufacturing and also to solve the yield issue when a failure occurs at a problematic process stage and there are different yield performances before and after the failure time.

We will further develop the TCP method by applying RJMCMC to partition tools into several homogenous groups for identifying problematic or golden tools, and we also build a new connection between statistical parameters and engineering tolerances. Using this new connection, the TCP method can automatically partition tools into several homogenous groups according to engineering tolerances. We also provide guidelines to choose initial values of parameters and hyperparameters to facilitate the practice of TCP in the semiconductor industry and other industries. As a result, engineers will be able to easily practice TCP after setting up their tolerances using their experience, knowledge or product specifications. This will provide more valuable information about tool comparisons that will help engineers enhance product yield and process capability.

The remainder of the paper is structured as follows. In Section II, we introduce our model in detail. Furthermore, we will provide guidelines for the choice of prior parameters and hyperparameters in practice. In Section III, we provide two simulation cases to illustrate the advantages of the TCP method. Here, we will also compare the partitioning results of tolerance controls with the pruning results of the CART method. In Section IV, we analyze two real illustrative examples in the semiconductor industry including 1) yield enhancement and 2) Cp/Cpk enhancement and propose an idea to integrate TCP with statistical dashboard [2] and automatic process control (APC) [17]–[19], respectively. In Section V, extensions of the TCP method are discussed.

II. Methodology

A. Modeling a Partition Problem

For each observation, Y denotes the response variable such as yield and x denotes the categorical predictor with J categorical levels such as tools. The conditional distribution of Y given x is, $Y|x \sim \mathrm{Normal}\ (\theta, \sigma^2)$, where $\theta = \theta_j$ if x is the jth tool, and $(\theta_1, \theta_2, \ldots, \theta_J)$ is the unknown parameter vector. The TCP method is used to partition the tools into several homogenous groups (that is, partition $(\theta_1, \theta_2, \ldots, \theta_J)$ into several homogenous groups). And the random effects of the tools $(\theta_1, \theta_2, \ldots, \theta_J)$ are the same if the tools belong to the same group.

A partition of J tools is a collection $g = \{S_1, S_2, \ldots, S_\kappa\}$ of subsets of $\{1, 2, \ldots, J\}$, which we call groups with $\cup_{k=1}^{\kappa} S_k = \{1, 2, \ldots, J\}$ and $S_i \cap S_j = \phi$ for $i \neq j$. The degree $\kappa (<= J.)$ of partition g is the number of groups into which J tools are divided by g. Within each group S_k, the parameter θ_j for $j \in S_k$ is drawn independently from the same normal distribution with hyperparameters μ_k and τ, that is, $\theta_j \sim \mathrm{Normal}\ (\mu_k, \tau^2)$, $j \in S_k$, $j = 1, 2, \ldots, J$, $k = 1, 2, \ldots, \kappa$.

Furthermore, using a Bayesian approach, another prior distribution for our model takes $\mu_k|\tau^2 \sim \mathrm{Uniform}\ (a, b)$, $k = 1, 2, \ldots, \kappa$, where the μ_k's are conditionally independent given τ^2.

τ^2 is distributed independently as scaled inverse Chi squared (ν, s^2), where $s^2 = (tolerance/6)^2$ in which the $tolerance$ parameter would be defined by engineers to stand for the acceptable difference between tools.

σ^2 is distributed independently as scaled inverse Chi squared (ν_1, s_1^2).

Finally, following Consonni and Veronese [20], g is independent and its prior distribution is taken as $P(g) \propto (\kappa^{-1}/\#\{g' : \kappa(g') = \kappa\})$, where κ is the degree of partition g.

When we have J tools, the jth tool has n_j observations for $j = 1, 2, \ldots, J$. We denote y_{ij} as the response variable of the ith observation of the jth tool for $i = 1, 2, \ldots, n_j$, and $j = 1, 2, \ldots, J$. The joint distribution of all variables will be specified by

$$P(g, \widetilde{\mu}, \tau^2, \widetilde{\theta}, \sigma^2, \widetilde{y}) = P(g)P(\tau^2)P(\widetilde{\mu}|\tau^2, g)P(\widetilde{\theta}|\widetilde{\mu}, \tau^2)P(\sigma^2)P(\widetilde{y}|\widetilde{\theta}, \sigma^2)$$

where $\widetilde{\mu} = (\mu_1, \mu_2, \ldots, \mu_\kappa)$, $\widetilde{\theta} = (\theta_1, \theta_2, \ldots, \theta_J)$, and $\widetilde{y} = (y_{ij})$, for $k = 1, 2, \ldots, \kappa$, $i = 1, 2, \ldots, n_j$, and $j = 1, 2, \ldots, J$.

Based on the above Bayesian models for our problem, we need to estimate the values of the unknown priors $(g, \widetilde{\mu}, \tau^2, \widetilde{\theta}, \sigma^2)$ with pre-determined hyperparameters a and b, and parameters ν, s^2, ν_1, and s_1^2. In Section II-C, we will provide the guidelines to set up initial values of these hyperparameters a and b, and parameters ν, s^2, ν_1, and s_1^2 in order to facilitate the method.

We apply RJMCMC to get the stationary posterior distribution of the unknown priors $(g, \widetilde{\mu}, \tau^2, \widetilde{\theta}, \sigma^2)$. By the RJMCMC algorithm, we could construct a Markov chain that converges to a unique stationary posterior distribution of the priors $(g, \widetilde{\mu}, \tau^2, \widetilde{\theta}, \sigma^2)$, and then estimate the unknown priors $(g, \widetilde{\mu}, \tau^2, \widetilde{\theta}, \sigma^2)$ by the modes of the stationary posterior distributions, respectively.

B. The TCP Algorithm for Partition Problems

Please refer to [13] and [21] to see the algorithm and further details of RJMCMC. The TCP algorithm is as follows.

1) Randomly choose one of the five prior parameters to update.
2) If choosing to update the partition g, we could also randomly choose one of birth and death move types.
   2.1) If choosing birth move type, one group is increased from its current partition. Then we randomly select one group which contains at least two tools to split randomly into two groups and calculate the acceptance probability R [13], based on current and new partitions, g and g'. Choose a random variable

u from Uniform (0, 1). If $u < R$, then accept the new partition g', else keep the current partition g.

2.2) If choosing death move type, one group is decreased from its current partition. Then we randomly select two groups to merge and calculate the acceptance probability, $1/R$ [13], based on current and new partitions, g and g'. Choose a random variable u from Uniform (0, 1). If u $< 1/R$, then accept the new partition g', else keep the current partition g.

3) If we choose another prior parameter to update, we update $\{\theta_1, \theta_2, \ldots, \theta_J\}$, σ^2, $\{\mu_1, \mu_2, \ldots, \mu_\kappa\}$, or τ^2 directly by the Gibbs sampler [22].

C. Guidelines for Choosing Initial Values of Parameters and Hyperparameters in TCP

To facilitate the practice of TCP in the semiconductor industry and other applications, we will provide guidelines to set up the related parameters and hyperparameters for related prior distributions of the TCP method as follows:

1) $\mu_k|\tau^2 \sim \mathrm{Uniform}\ (a, b)$. We let $a = \min\{\overline{y_{.j}}\}$ and $b = \max\{\overline{y_{.j}}\}$;
2) $\tau^2 \sim$ scaled inverse Chi squared (ν, s^2). Let $\nu = 20$, $s^2 = (tolerance/6)^2$;
3) $\sigma^2 \sim$ scaled inverse Chi squared (ν_1, s_1^2). Let $\nu_1 = 20$, and $s_1^2 =$ variance of $\{(y_{ij} - \overline{y_{.j}}), j = 1, 2, \ldots, J, i = 1, 2, \ldots, n_j\}$.

The values of ν and ν_1 are only recommendations; they impact the convergence speed of the TCP method.

By using the above guidelines for choosing initial values of parameters and hyperparameters, engineers only need to determine the numerical level of the $tolerance$ parameter. Furthermore, the tolerance concept is widely used in the semiconductor industry and other applications. For example, errors inevitably result from metrology systems and minor deviations with respect to product specifications. Hence, one can set up the $tolerance$ using engineers' knowledge, product specifications, or tool limitations. Then, the TCP method could search the optimal partition according to the acceptable level of $tolerance$. Therefore, we could integrate the engineering concept of tolerance with the TCP to help engineers implement this method in practice. In Section III-B, we present the results of sensitivity analysis to show that the TCP method could generate optimal partitions with respect to different levels of $tolerance$.

III. Simulation Studies

As an illustration of the performance of the TCP method, we provide two simulation cases. In Section III-A, we show the limited impact of unbalanced usage in manufacturing. In Section III-B, sensitivity analysis using the different tolerance controls and a comparison with the pruning results from using the CART method are reported.

A. Unbalanced Data for Unbalanced Tool Usages

We generate unbalanced data to represent a situation in which there are three kinds of yield distributions among five tools and the usages of the five tools are different. The purpose of this investigation is to partition the five tools according to their yield distributions based on a specific level of tolerance control. Following the notation of modeling in Section II-A, the simulation models are described in detail as follows: $y_{ij} \sim \mathrm{Normal}$ (θ_j, σ^2), where $i = 1, 2, \ldots, n_j$ and $j = 1, 2, \ldots, 5$, with $(n_1, n_2, n_3, n_4, n_5) = (15, 150, 20, 250, 20)$ (so the usages of five tools are unbalanced), $(\theta_1, \theta_2, \ldots, \theta_5) = (3, 3, 4, 4, 7)$, and $\sigma = 1$. Thus, there are three groups for the five tools and the true partition is {(T1, T2), (T3, T4), T5}, which we denote as (11223). Since the numbers of observations for each tool are unequal, we also present the influences on the TCP method from unbalanced data. For setting the parameters and hyperparameters of the prior distributions in this case, we use the guidelines in Section II-C with $tolerance = 1$.

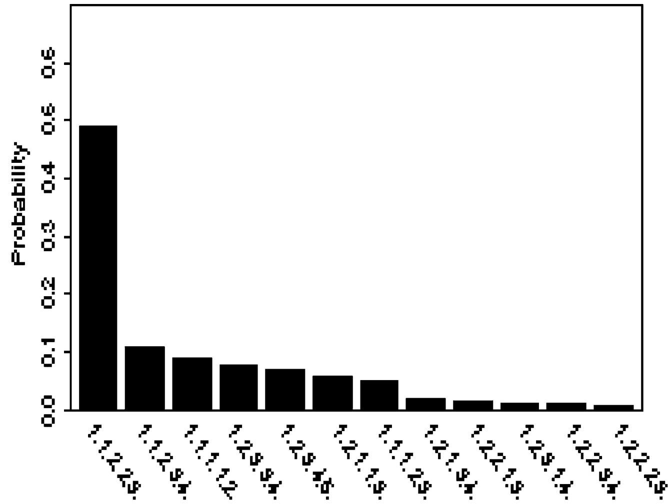


Fig. 1. The posterior distribution of partition in case of Section III-A (we only display probability $>= 0.005$).

By the TCP method, the correct partition (denoted as (11223)) is the mode of the posterior distribution with probability 0.49. The stationary posterior distribution of partitions is given in Fig. 1. As a result, we find that the impact of unbalanced data is very limited. Therefore, the unbalanced tool usage in semiconductor manufacturing will not influence our methods.

B. Sensitivity Analysis With Different Tolerance Controls and its Comparison With the Pruning Results of the CART Method

In sensitivity analysis, the data is generated from the same model as in case of Section III-A, but the number of observations for each tool is equal to 30. Thus, the true partition is {(T1, T2), (T3, T4), T5} (denoted as (11223)). Here, we apply the TCP method to different values of $tolerance (= 0.5, 1, 2, 3, 4, 5, 6)$ to show its influence on the results of the TCP method. The results of 30 simulations for every $tolerance$ level are given in Table I. Now, the group means are 3, 3, 4, 4, and 7 and the within-group standard deviations are 1 in this simulation. When the $tolerance$ is 0.5 or 1, the target partition of (11223^{*}) will be the most plausible because the between group differences can be as large

TABLE I
THE POSTERIOR PROBABILITIES OF DIFFERENT PARTITIONS WITH RESPECT TO DIFFERENT TOLERANCE. THE AVERAGE OF THE POSTERIOR PROBABILITIES FROM 30 SIMULATIONS FOR EVERY TOLERANCE LEVEL ARE REPORTED WITH THE STANDARD ERROR REPORTED IN PARENTHESES. THESE SIMULATION RESULTS CONFIRM THAT THE TARGET PARTITION, $\{11223^{*}, 11112^{**}, 11111^{***}\}$, WILL HAVE THE HIGHEST AVERAGE OF POSTERIOR PROBABILITIES AT DIFFERENT TOLERANCE LEVELS AS EXPLAINED IN SECTION III-B. FURTHERMORE, THE DISTRIBUTIONS OF POSTERIOR PROBABILITIES INDICATE THE NEXT PLAUSIBLE PARTITION UNDER THE SPECIFIC TOLERANCE CONTROL, WHICH CAN BE USED TO FORM THE CREDIBLE SET IN THE BAYESIAN APPROACH. (WHEN THE POSTERIOR PROBABILITY IS SMALLER THAN 0.0001, IT IS REGARDED AS 0. WHEN THE MAXIMUM VALUE OF THE AVERAGE POSTERIOR PROBABILITIES FOR ALL TOLERANCES IN ONE POSSIBLE PARTITION IS SMALLER THAN 0.005, THIS PARTITION IS OMITTED IN THIS TABLE.)

	Tolerance=0.5	**Tolerance=1**	**Tolerance=2**	**Tolerance=3**	**Tolerance=4**	**Tolerance=5**	**Tolerance=6**
11111*	0(0)	0(0)	0.00003(0.00017)	0.00356(0.00505)	0.05045(0.01344)	0.23946(0.02234)	**0.44427(0.01808)**
11112	0.00498(0.00564)	0.07873(0.01763)	**0.40841(0.01421)**	**0.49478(0.01448)**	**0.4496(0.01567)**	**0.30016(0.01678)**	0.15932(0.01319)
11122	0(0)	0(0)	0.00005(0.00018)	0.00205(0.00141)	0.01261(0.0031)	0.02722(0.00395)	0.03162(0.00423)
11123	0.00218(0.00312)	0.01763(0.00503)	0.04692(0.00481)	0.05001(0.00437)	0.0463(0.0037)	0.03282(0.00236)	0.01926(0.0024)
11212	0(0)	0(0)	0.00024(0.00072)	0.00262(0.00197)	0.01312(0.00377)	0.02986(0.00478)	0.03264(0.00556)
11213	0.00288(0.00377)	0.02111(0.00674)	0.05162(0.0068)	0.05227(0.00642)	0.04672(0.0034)	0.03219(0.00317)	0.01914(0.0029)
11222	0(0)	0(0)	0.00003(0.00015)	0.00053(0.00074)	0.00406(0.00147)	0.01235(0.00239)	0.01687(0.00311)
11223*	**0.68238(0.02295)**	**0.54151(0.03205)**	0.18024(0.01374)	0.0932(0.00844)	0.06211(0.00595)	0.03777(0.00475)	0.02006(0.00255)
11234	0.12316(0.01295)	0.11821(0.00626)	0.05699(0.00607)	0.03498(0.00417)	0.02869(0.00367)	0.01917(0.00218)	0.01193(0.00175)
12112	0(0)	0(0)	0(0)	0.00018(0.00037)	0.00221(0.00107)	0.00724(0.00203)	0.01149(0.00198)
12113	0.00042(0.00092)	0.00612(0.00263)	0.026(0.00343)	0.03118(0.00292)	0.03223(0.00231)	0.02358(0.00363)	0.01487(0.00183)
12123	0.0002(0.00044)	0.00329(0.00158)	0.02334(0.00385)	0.03377(0.00371)	0.03535(0.00356)	0.0264(0.00334)	0.01649(0.0027)
12134	0.00043(0.00055)	0.00433(0.00148)	0.0125(0.0019)	0.01563(0.00197)	0.01678(0.00206)	0.01289(0.00202)	0.00907(0.00161)
12213	0.00039(0.00096)	0.00333(0.00154)	0.02293(0.0029)	0.03487(0.00281)	0.03677(0.00329)	0.02617(0.00349)	0.01576(0.00315)
12221	0(0)	0(0)	0(0)	0.00012(0.00026)	0.00135(0.00077)	0.00552(0.00151)	0.00898(0.00176)
12223	0.00862(0.00715)	0.0277(0.00777)	0.04103(0.0059)	0.03733(0.00416)	0.03447(0.00371)	0.02449(0.00327)	0.01369(0.00223)
12234	0.00338(0.00239)	0.00997(0.00271)	0.01717(0.00264)	0.01699(0.00233)	0.01738(0.00172)	0.01394(0.00174)	0.00856(0.00173)
12314	0.00065(0.00084)	0.00439(0.00156)	0.01317(0.00213)	0.0157(0.00221)	0.01734(0.00176)	0.01319(0.00243)	0.00923(0.00169)
12324	0.00453(0.00272)	0.01101(0.00316)	0.01826(0.00206)	0.01804(0.00252)	0.01721(0.00206)	0.01327(0.00246)	0.00938(0.00158)
12334	0.09566(0.01151)	0.07927(0.00796)	0.0336(0.00294)	0.02163(0.00254)	0.0182(0.00271)	0.01253(0.002)	0.0084(0.00176)
12345	0.07013(0.00919)	0.0734(0.0055)	0.04738(0.00481)	0.03869(0.00379)	0.03931(0.00368)	0.03129(0.00383)	0.02345(0.0029)

as 1. When the $tolerance$ is 2, 3, 4, or 5, the target partition of (11112^{**}) will be the most plausible because the between group differences can be as large as 4 and within group standard deviation is 1. When the $tolerance$ is 6, the target partition of (11111^{***}) will be the most plausible because the between group differences are smaller than 6. The posterior probability of the target partition turns out to be the mode with the highest average of posterior probabilities in 30 simulations for every $tolerance$ level. Furthermore, the standard errors are very small, demonstrating the robustness of the posterior modes. Consequently, the averages of the posterior probabilities of three partitions, $\{11223^{*}\}$, $\{11112^{**}\}$, and $\{11111^{***}\}$, will change according to the levels of different $tolerances$. The average of posterior probabilities of the partition $\{11223^{*}\}$ decreases when the $tolerance$ increases. The average of posterior probabilities of the partition $\{11112^{**}\}$ increases when the $tolerance$ increases from 0.5 to 3 and decreases when the $tolerance$ increases from 3 to 6. The average of posterior probabilities of the partition $\{11111^{***}\}$ increases when the $tolerance$ increases. That is, we find that the probability of merging tools increases as the $tolerance$ increases. Hence, the posterior distributions of the partitions reflect the levels of tolerance controls and the posterior probabilities of the next plausible partitions and other partitions indicate the strength of plausibility. These posterior probabilities can be used to form a credible set in the Bayesian approach.

The results of sensitivity analysis do support the evidence that the partitioning results of TCP will reflect the levels of tolerance controls. Using the above simulation results, we show that the TCP method provides plausible partitioning results by in-

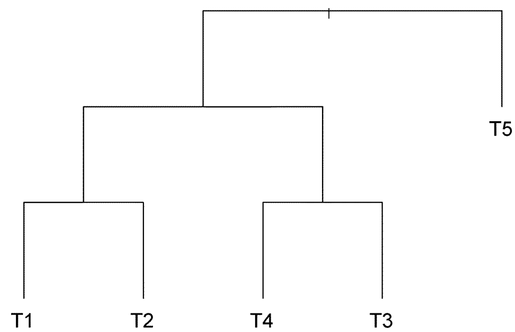


Fig. 2. A tree obtained by the CART model for case in Section III-B.

TABLE II
THE DIFFERENT PARTITIONING RESULTS WITH RESPECT TO DIFFERENT VALUES OF COST COMPLEXITY IN THE CART MODEL FOR CASE IN SECTION III-B

Partitioning result	Cost-complexity
(1,2,3,4,5)	0
(1,1,2,2,3)	5
(1,1,1,1,2)	25

tegrating the RJMCMC method and the engineers' control of tolerance.

We also analyze the data by the CART method. We use the S-Plus functions of tree and prune.tree to find the standard CART results [23]. The complete tree result is given in Fig. 2, and the best partitioning results with respect to different cost-complexities are given in Table II. Combining the results in Table I and Table II, one can generate reasonable partitioning results by different tolerances and approach the same partition from pruning trees in the CART method. However, the TCP method can generate the partition automatically according to the practice of tolerance control by engineers.

IV. Two Applications in the Semiconductor Industry

A. Ramp Up Yield by the TCP Method

The data below is from a company in the semiconductor industry in Taiwan. The term Srow is one of the key test items in wafer sort testing. The larger the Srow, the worse the yield performance is. The data consists of 439 lots with $\mathrm{mean} = 5.98$ and $\mathrm{standard\ deviation} = 1.85$. Engineers aim to reduce the mean and variance of Srow. The box plots and related statistics for various tools of the problematic step are shown in Fig. 3 and Table III, respectively.

Based on the fact that engineers set their acceptance $tolerance = 1$, the TCP method partitions the set of tools, {SPU16, SPU14, SPU07, SPU05, SPU04, SPU03} into three groups, {SPU16, SPU14, SPU07, SPU04}, {SPU05}, and {SPU03} (denoted as (111213)) with maximum probability 0.2244. The stationary posterior distribution of partitions is given in Fig. 4. The same partition result was obtained by the CART method with $\mathrm{cost-complexity} = 30$. It is consistent with the tree result using CART as in Fig. 5.

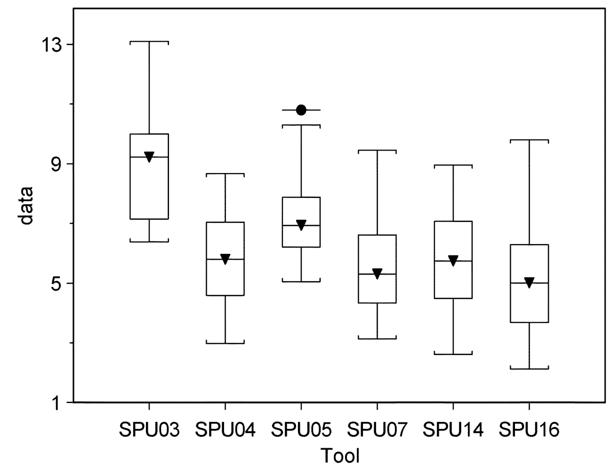


Fig. 3. The box plot of Srow by different tools for case in Section IV-A.

TABLE III
THE MEAN, STANDARD DEVIATION, AND COUNT OF EACH TOOL FOR CASE IN SECTION IV-A

Tool	Mean	Std	Count
SPU03	9.05	1.85	14
SPU04	5.79	1.51	48
SPU05	7.17	1.27	96
SPU07	5.54	1.4	36
SPU14	5.83	1.7	93
SPU16	5.16	1.78	152

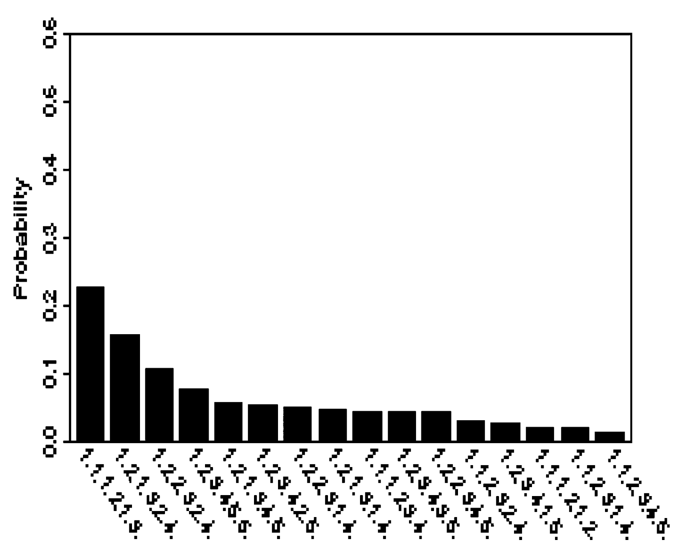


Fig. 4. The posterior distribution of partition in case of Section IV-A (we only display probability $>= 0.005$).

After looking further into this problematic step, engineers find that there are two different tool types, one type includes SPU03 and SPU05 and another type includes SPU16, SPU14, SPU07, and SPU04. Because the different tool types use different process chemicals, the contaminated chemical is the root cause of the bad performances of SPU03 and SPU05. In comparison, the TCP does separate SPU03 and SPU05 into different groups that are consistent with the findings of engineers.

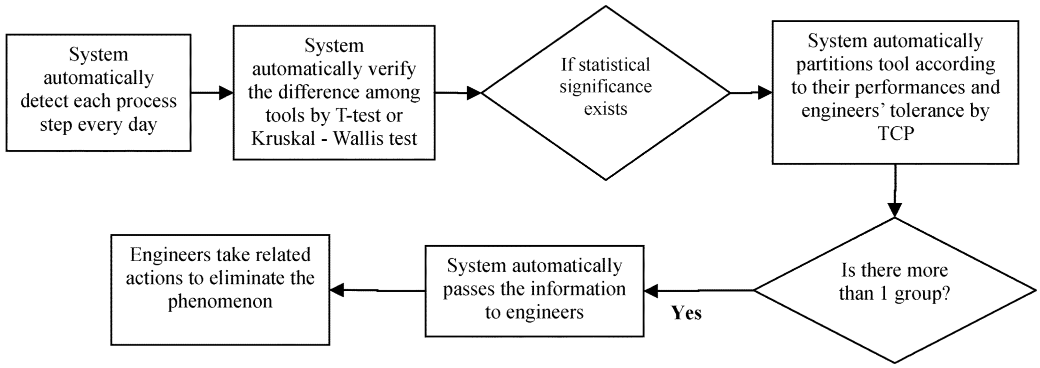


Fig. 6. Engineer daily trouble shooting flow according to autodetection mechanism.

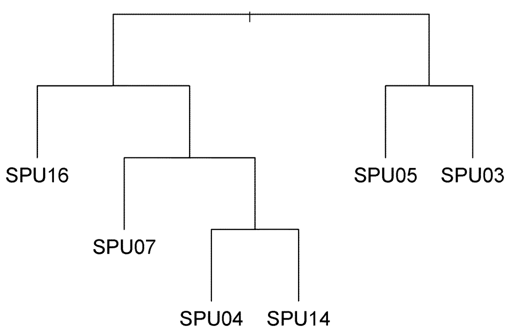


Fig. 5. A tree obtained by the CART model for case of Section IV-A.

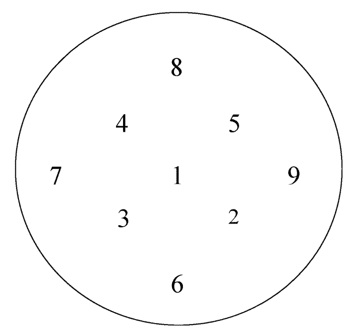


Fig. 7. Site locations on each wafer

TABLE IV
CONCEPTUAL REPORT TO AUTOMATICALLY DETECT THE PERFORMANCE DIFFERENCE AMONG TOOLS FOR EACH PROCESS STEP

Step	P value of T or Kruskal Wallis Test	TCP Result (Group, Tool List; Mean)
Step 10	**0.000005**	**(1.SPU03; 9.05);(2.SPU05; 7.17); (3.SPU04,SPU07,SPU14,SPU16; 5.58)**
Step 15	**0.0003**	**(1.TEC02; 8.32);(2.TEC01; 8.08);**
Step 2	0.06	(1.ACE01,ACE02; 8.1);
Step 4	0.08	(1.PHO01,PHO0202, PHO03; 8.21);
.		
.		

We suggest integrating the TCP method with statistical tests into a statistical dashboard [2] to form an analysis flow as in Fig. 6. After building automatic systems according to the analysis flow, systems could execute the analysis automatically at night for each item of each product; then engineers could quickly detect problems, like those shown in Table IV, at the beginning of their daily work. This will dramatically shorten the time for engineers to increase yield ramp up.

B. Enhance Process Capability Indices, $C_p$and C_{pk}, by Applying the TCP Method

Process capability indices C_p and C_{pk} have been widely used in the manufacturing industry to provide numerical measures on process performance and product quality. They are also common indices to demonstrate a company's own process capability to their customers and competitors. The two indices are defined as $C_p = (USL - LSL)/6\sigma C_{pk} = \min\{(USL - \mu)/3\sigma, (\mu - LSL)/3\sigma\}$, where LSL and USL are the lower and upper specification limits, respectively, defined by process engineers or product designers, where μ is the process mean and σ is the process standard deviation. Conventionally, we use $\overline{X} = (\sum_{i=1}^{n} X_i)/n$ and $s = [\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)]^{1/2}$ as estimators of μ and σ, respectively, so the natural estimators of C_p and C_{pk} could be $\hat{C}_p = (USL - LSL)/6s$, $\hat{C}_{pk} = \min\{(USL - \overline{X})/3s, (\overline{X} - LSL)/3s\}$. In general, the minimum requirement for $\hat{C}_p$ and $\hat{C}_{pk}$ is 1.33 or 2. Therefore, enhancing $\hat{C}_p$ and $\hat{C}_{pk}$ is one of the major tasks for process engineers in the semiconductor industry.

After completing each process step, one wafer is selected from every lot and the process parameter is measured at nine predetermined sites on each wafer as Fig. 7 indicates. The process parameter is the critical oxide thickness after one important diffusion process step at a semiconductor manufacturing company in Taiwan. Based on product specifications, $USL = 1850$ and $LSL = 1550$ and the C_p and C_{pk} are equal to 1.132, and 1.1053, respectively.

To apply the TCP method to partition these sites, we define the $tolerance$ to be 10% of specification $300(= 1850 - 1550)$, and we find the partition {site1, {site2, site3, site4, site5, site6, site7, site8, site9}} (denoted as (122222222)) with maximum probability 0.5971. This result matches the site phenomenon shown by the box plots in Fig. 8 and related statistics in Table V.

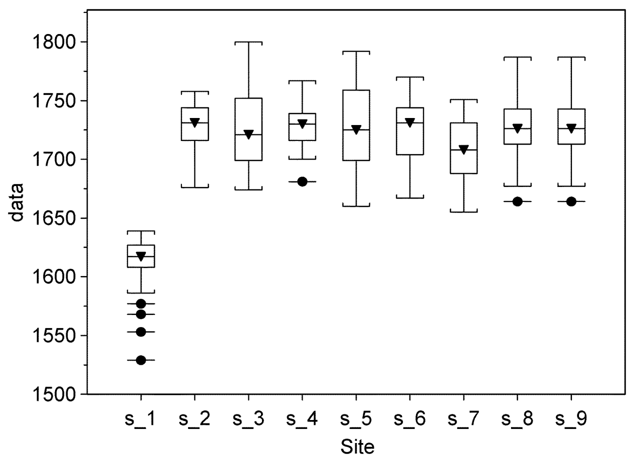


Fig. 8. The box plot of Oxide thickness by different sites for case of Section IV-B.

TABLE V
THE MEAN, STANDARD DEVIATION, AND COUNT OF OXIDE THICKNESS BY DIFFERENT SITES FOR CASE IN SECTION IV-B

site	Mean	Std	Count
s_1	1611	26.19	35
s_2	1727	20.21	35
s_3	1726	31.08	35
s_4	1728	19.06	35
s_5	1727	34.56	35
s_6	1726	26.34	35
s_7	1708	35.17	35
s_8	1713	22.11	35
s_9	1725	25.87	35

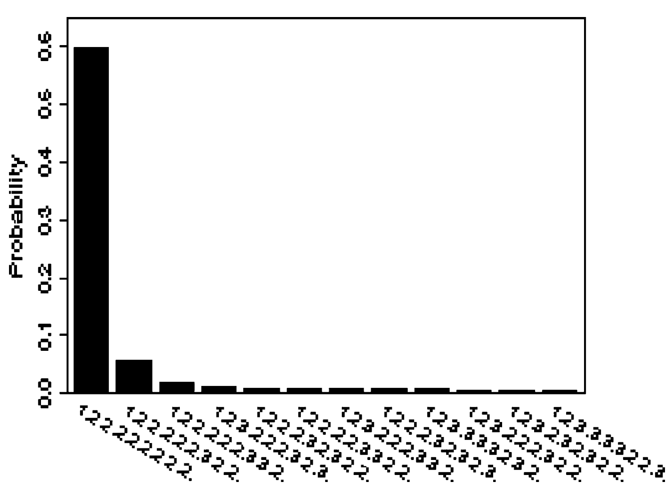


Fig. 9. The posterior distribution of partition in case in Section IV-B (we only display probability $>= 0.005$).

The same partition result was obtained by the CART method with $\mathrm{cost} - \mathrm{complexity} = 9775$. According to the posterior distribution shown in Fig. 9, this result is similar to the tree result of CART in Fig. 10. After fine-tuning the process recipe to eliminate the site difference, we can improve the C_p and C_{pk} to 1.86 and 1.56, respectively.

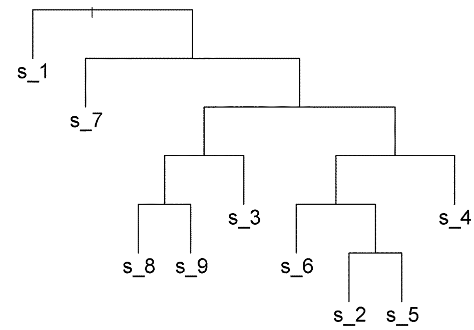


Fig. 10. A tree obtained by the CART model for case in Section IV-B.

Using an automatic system analysis flow, we can automatically find the site profile and then integrate the capability of auto fine-tuning in the most advanced tools [17]–[19]. Engineers could further build an automatic process control mechanism as in Fig. 11 to simultaneously enhance process capability and reduce the workloads of engineers.

V. Conclusion And Discussion

Instead of only detecting the statistically significant difference by using statistical tests, the TCP method can automatically partition tools into several homogenous groups with the built-in control of tolerance in engineering. It can resolve the difficulty of determining the related parameters for the CART method. Instead of using a hierarchical tree structure as CART does, TCP uses the posterior distribution of the partition to discover the partitioning structure of tools. The TCP method also performs well for unbalanced data. Finally, we suggest a method of choosing initial values of parameters and hyperparameters to facilitate the TCP method in practice. With two real applications from the semiconductor industry, we not only show that our method could provide correct information for engineers to enhance yield/process capability, but also provide an idea to build a practical mechanism by integrating engineers' daily work flow and the capability of the most advanced tools. TCP will make it much easier for engineers to realize the performance differences among tools and enhance the effectiveness and efficiency for yield / process capability enhancement and experimental analysis.

The TCP method could also be applied to similar cases such as recipes or material comparisons to extend their value. There are many potential extensions of the TCP method. Tools with unequal variances and other types of distribution models are natural extensions for practical engineering situations. All the above cases could also be extended to a multivariate situation for engineers to simultaneously perform multiple comparisons of a collection of responses. These will be investigated in future studies.

Another important extension is to develop an automatic system to combine our method with the ANOVA or the Kruskal-Wallis test in order to automatically alarm the possible excursion and apply TCP in automatic process control (APC) [17]–[19]. Thus we would provide more real benefits for the semiconductor industry.

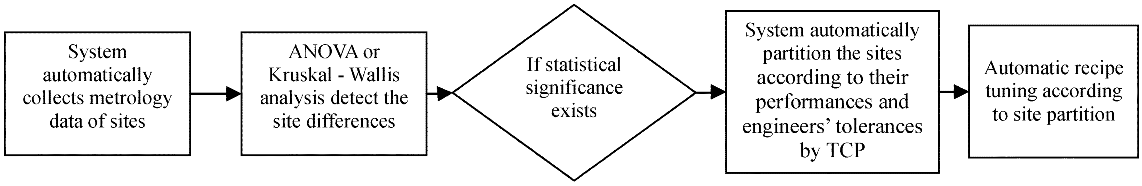


Fig. 11. Auto process capability enhancement mechanism by integrating our method, ANOVA, and the auto-recipe-tuning tool.

ACKNOWLEDGMENT

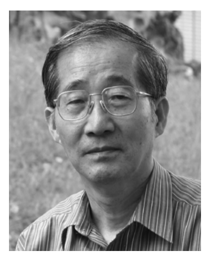
The authors thank the editor and reviewers for their valuable comments.

REFERENCES

[1] P. K. Chatterjee and R. R. Doering, "The future of microelectronics," *Proc. IEEE*, vol. 86, no. 1, pp. 176–183, Jan. 1998.
[2] F. Bergeret and Y. Chandon, "Improving yield in IC manufacturing by statistical analysis of large data base," *Micro*, pp. 59–75, 1999.
[3] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion analysis of variance," *J. Amer. Statist. Assoc.*, vol. 47, pp. 583–621, 1952.
[4] G. Kong, "Tool commonality analysis for yield enhancement," in *Proc. 2002 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, 2002, pp. 202–205.
[5] L. K. Garling and G. P. Woods, "Determining equipment performance using analysis of variance," in *Proc. 1990 Int. Semiconductor Manufacturing Science Symp.*, 1990, pp. 85–89.
[6] T. McCray, J. McNames, and D. Abercrombie, "Locating disturbances in semiconductor manufacturing with stepwise regression," *IEEE Trans. Semicond. Manuf.*, vol. 18, no. 3, pp. 458–468, Aug. 2005.
[7] D. C. Montgomery, *Design and Analysis of Experiments*, 3rd ed. New York: Wiley, 1991.
[8] R. M. Gardner, J. Bieker, and S. Elwell, "Solving tough semiconductor manufacturing problems using data mining," in *Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf. and Workshop*, Boston, MA, Sep. 2000, pp. 46–55.
[9] F. Mieno, T. Sato, Y. Shibuya, K. Odagiri, H. Tsuda, and R. Take, "Yield improvement using data mining system," in *Proc. IEEE Int. Symp. Semiconductor Manufacturing Conf.*, Santa Clara, CA, Oct. 1999, pp. 391–394.
[10] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
[11] H. S. Stern, "Neural networks in applied statistics," *Technometrics*, vol. 38, no. 3, pp. 205–213, 1996.
[12] K. Chaloner, "Bayesian approach to the estimation of variance components for the unbalanced one-way random model," *Technometrics*, vol. 29, no. 3, pp. 323–337, 1987.
[13] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
[14] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *J. Roy. Statist. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.
[15] A. Nobile and P. J. Green, "Bayesian analysis of factorial experiments by mixture modeling," *Biometrika*, vol. 87, pp. 15–35, 2003.
[16] F. Bergeret and C. L. Gall, "Yield improvement using statistical analysis of process data," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 3, pp. 535–542, Aug. 2003.
[17] J. Maeritz and A. Schels, "Production enhancement of lithography through APC methods," *Future Fab*, vol. 14, 2003.
[18] E. K. Lada, J. C. Lu, and J. R. Wilson, "A wavelet—Based procedure for process fault detection," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 1, pp. 79–90, Feb. 2002.
[19] S. T. Tseng, A. B. Yeh, F. Tsung, and Y. Y. Chan, "A study of variable EWMA controller," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 633–643, Nov. 2003.
[20] G. Consonni and P. Veronese, "A Bayesian method for combining results from several binomial experiments," *J. Amer. Statist. Assoc.*, vol. 90, pp. 935–944, 1995.
[21] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Markov chain Monte Carlo in practice," in *Interdisciplinary Statistics*. London, U.K.: Chapman & Hall, 1996.
[22] G. Casella and E. I. George, "Explaining the Gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1998, C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
[23] L. A. Clark and D. Pregibon, "Tree based models," in *In Statistical Models*, J. Chambers and T. Hastie, Eds. Belmont, CA: Wadsworth, 1992.

Kevin Kai-Wen Tu received the B.S. degree in mathematics from National Tsing Hua Unversity, Taiwan, in 1991, the M.S. and Ph.D. degree in statistics from National Chiao Tung Unversity, Hsinchu, Taiwan, in 1993 and 2008, respectively.

He has worked on engineering data analysis in semiconductor industry since 1996. He is currently a Deputy Manager in the Decision Support System Development Division, Macronix International Co., LTD. (MXIC), Science Park, Hsinchu, Taiwan. MXIC is famous on its nonvolatile memory products. He leads a team to develop engineering data analysis systems for supporting yield enhancement, product/technology development, SPC implementation, statistical consulting and data mining by integrating the statistical method, the information technology and semiconductor engineering knowledge. He also provided the course about statistical practice in the Institute of Statistics, National Chiao Tung University, Taiwan since 2009.



Jack Chao-sheng Lee (deceased) received the B.S. degree in management from National Taiwan University, Taiwan in 1964, M.A. degree in economics from University of Rochester, Rochester, NY, in 1969, and the Ph.D. degree in statistics from the State University of New York at Buffalo.

Since 2002, was University Chair Professor of Graduate Institute of Finance and Institute of Statistics at National Chiao Tung University, Taiwan. He had published many papers, and his research subjects included many fields in statistics, including multivariate analysis, time series analysis, growth curves, Bayesian inference and classification and pattern recognition. He also published in other areas such as portfolio management and option evaluation, with particular application to finance.

Dr. Lee was a Fellow of American Statistical Association, elected Ordinary Member of International Statistical Institute, and was President of International Chinese Statistical Association in 1992. He received Distinguished Service Award from Chinese Statistical Association in 1997, Service Award from International Chinese Statistical Association in 1993, and Award of Excellence from Bell Comm. Research in 1987. He passed away on March 2, 2007.

Henry Horng-Shing Lu received the B.S. degree in electric engineering from National Taiwan University, Taiwan, in 1986, the M.S. and Ph.D degrees in statistics from Cornell University, Ithaca, NY, in 1990 and 1994, respectively.

He is a Professor in the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. He has been a visiting scholar at UCLA, Harvard University, and University of Chicago. His research interests include statistics, medical images, and bioinformatics. He and collaborators have journal papers published or accepted in a wide spectrum of journals, including *Journal of the American Statistical Association*, *Journal of Multivariate Analysis*, *Statistica Sinica*, *Journal of Computational and Graphical Statistics*, IEEE TRANSACTIONS ON RELIABILITY, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MEDICAL IMAGING, *Pattern Recognition*, *Ultrasound in Medicine and Biology*, *Trends in Genetics*, *Proceedings of the National Academy of Sciences of the United States of America*, *Journal of Computational Biology*, *Bioinformatics* and so forth. He has served as an Associated Editor for *Statistica Sinica* and a guest editor for *Journal of Data Science*.