Bayesian Inference Technique for Data Mining for Yield Enhancement in Semiconductor Manufacturing Data

Conference Paper · October 2015

CITATIONS
0

READS
104

3 authors, including:

Marzieh Khakifirooz
Tecnológico de Monterrey
59 PUBLICATIONS **553** CITATIONS

SEE PROFILE

Bayesian Inference Technique for Data Mining for Yield Enhancement in Semiconductor Manufacturing Data

Marzieh Khakifirooz, Chen-Fu Chien and Ying-Jen Chen

***Abstract*: The yield management in semiconductor manufacturing is one of the interesting areas that data mining approaches to find useful applications. The abundant steps and complex workflows during wafer manufacturing automatically generate large volumes of data and, hence, engineers who rely on personal domain knowledge cannot find possible root causes of defects quickly and effectively.**

The complexities involved in semiconductor manufacturing have always delayed the dream of creating a reliable process to produce 100% yield. Although the manufacturing recipes are carefully designed and revised to maximize yield, yield is still affected by errors that are reported by systematic factors (e.g., defective tools or interactions between tools) or random factors (e.g., dust particles). Furthermore, experiments have shown that most insidious and dangerous defects come from the interactions between components of a complex system - that cannot be detected by human diagnostic at an individual developer level. Although, generally, selecting the process tool, chamber set and recipe name, eventuate based on a series of previous experience, however, these practical intuitions don't have any seat in computerized process mining for defect detection.

This study aims to develop a framework for data mining and knowledge discovery from a database that consists of three phases: data preparation, data dimension reduction and the model construction and evaluation based on Bayesian Variable Selection (BVS) to figure the effect of practical intuitions and investigate the huge amount of semiconductor manufacturing data and infer possible causes of faults and manufacturing process variations. The proposed approach has been validated by an empirical study, eventually replicated Cross-validation has emerged as the preferred method to estimate the accuracy of the proposed approach on a particular data set and the results have shown its practical viability.

***Keywords*:** Bayesian Variable Selection (BVS), Data Mining, Yield Enhancement.

I. INTRODUCTION

In the age of digital information, Big Data, mining, and analytics are the principal components of strategic decision-making. Investments in data management and analytics are growing whereby helping companies to predict process behavior, to identify and detach defective tools and recipes to help improve yield.

Semiconductor manufacturing is among the most demanding businesses, which has one of the complex production processes, this complexity heightens the allure of data mining analytics, which it can sieve through complex data and improve efficiency, yield and decision making.

The yield learning curve of semiconductor manufacturing [1], [2], have demonstrated that in addition to data analytics, cumulative engineering training and experience significantly enhanced yield improvement, hence the integrated yield management methods [3], [4], [5], supported by historical experience and statistical data management are widely applicable in industry.

Although during wafer fabrication, yield engineers for selecting the machinery tools or chambers, trust to their cumulative skills and analytic methods simultaneously, however, this integration makes a lack of convenience to embrace the independence condition among the operation variables for statistical test.

Additionally, typically the chip industry batch production process, brings affiliation among the process variables. Statisticians entitle this issue as Multicollinearity.

The points of Multicollinearity and empirical variable selection behavior plus the high volume of variables persuade us to reflect on the prior distribution for semiconductor manufacturing data frame and the purpose of mining production data to extract discovery knowledge of defect diagnosis and eventually yield improvement.

This work is organized as follows. Section 2 presents the fundamental material for our application to semiconductor manufacturing. Section 3 proposes a research framework with detail procedures. Section 4 validate the framework with empirical study. Section 5 summarizes the main results, gives the conclusion, and describe some areas for further research.

II. FUNDAMENTAL BASIS

Consider that the window of the production cycle of wafer divided into segments or steps. These steps represent processes applied to all wafers. Fig. 1 illustrates a fragment of the life cycle of a wafer. The wafer would complete sequentially by the passing couple of hundred steps, at each step the wafer passes by a particular process tool. Many alternative tools and chambers may be qualified for performing the same action on a single step, however, only one of the many similar tools-chambers is applied to a wafer.

In this study for simplification, we are used to denoting the compound of tools and chambers processed with singular nominal factors, where even the probability of random clipping, depends on the other factors at each step.

A. Categorical Distribution for Nominal Data

In probability theory and statistics, a categorical distribution, also called "Multinomial distribution", is a probability distribution that describes the possible results of a random event that can take on one of *k* possible outcomes, with the probability of each outcome separately specified [6].

Let $\mathbf{Y} = (Y_1,...,Y_k)$, where Y_i is the number of n independent trials that result in the category $i, i = 1,\ldots k$.

M. Khakifirooz, C-F Chien and Y-J Chen are with Department of Industrial Engineering & Engineering Management, National Tsing Hua University, Hsinchu, Taiwan.

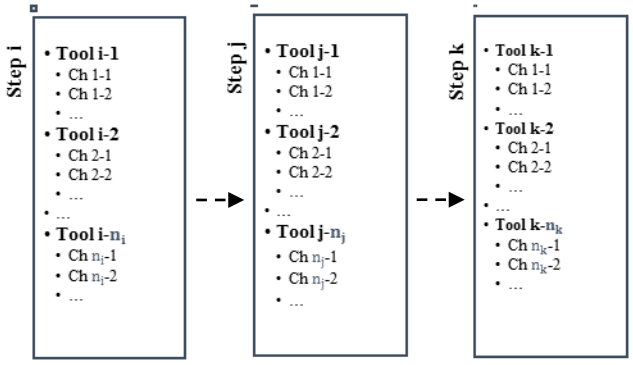

Figure 1. Schematic of batch production of a wafer

The probability distribution function of this multinomial distribution is:

$$f(y_1,\ldots,y_k;n,p_1,\ldots,p_k)=\Pr(Y_1=y_1,\ldots,Y_k=y_k)$$
$$=\begin{cases}\dfrac{n!}{y_1!\ldots y_k!}p_1^{y_1}\ldots p_k^{y_k} & \text{when } \sum_{i=1}^{k}y_k=n\\ 0 & \text{otherwise,}\end{cases} \quad (1)$$

B. Bayesian Models for Multinomial Data

Bayesian models can represent the dependency between variables. In the Bayesian paradigm, current knowledge about the model parameters is expressed by placing a probability distribution on the parameters, called the prior distribution, often written as $P(x)$, when new data θ become available, the information they contain regarding the model parameters is expressed in the likelihood, which is proportional to the distribution of the observed data given the model parameters, written as $P(\theta \mid x)$. This information is then combined with the prior to producing an updated probability distribution called the posterior distribution.

In Bayesian statistics, if the posterior distributions $P(\theta \mid x)$ are in the same family as the prior probability distribution $P(x)$, the prior and posterior are then called conjugate distributions and the prior is called a conjugate prior for the likelihood function, the Dirichlet distribution [7] is the conjugate prior for the multinomial distribution, consider the (1), then the corresponding likelihood function can be expressed using the gamma function as:

$$f(y_1,\ldots,y_k,p_1,\ldots,p_k)=\frac{\Gamma\left(\sum_i y_i+1\right)}{\prod_i\Gamma(y_i+1)}\prod_{i=1}^{k}p_i^{y_i}, \quad (2)$$

which is the form of probability density function of Dirichlet distribution.

C. Approximate Inference for Bayesian Model with Gibbs Sampling

The aim of Bayesian inference is to impound the posterior probability distribution over a set of random variables. However, using this distribution often needs intractable computing. Gibbs sampling [8] is one Monte Carlo Markov Chain (MCMC) technique suitable for this task. The idea in Gibbs sampling is to generate posterior samples by eliminating each variable to sample from its conditional distribution with the remaining variables fixed to their current values. For instance, consider the random variables Y_1, Y_2 and Y_3, we proceed as follows:

1: start by setting the initial values for each variables $y_1^{(0)}$, $y_2^{(0)}$ and $y_3^{(0)}$

2: at iteration *i*, sample

$$y_1^{(i)} \sim P\left(Y_1=y_1 \mid Y_2=y_2^{(i-1)}, Y_3=y_3^{(i-1)}\right)$$
$$y_2 \sim P\left(Y_2=y_2 \mid Y_1=y_1^{(i)}, Y_3=y_3^{(i-1)}\right)$$
$$y_3 \sim P\left(Y_3=y_3 \mid Y_1=y_1^{(i)}, Y_2=y_2^{(i)}\right)$$

3: iterate the above step until the sample values have the same distribution as if they were sampled from the true posterior joint distribution.

The most common reason of Gibbs sampling popularity, it works well in the presence of Multicollinearity and high dimensionality.

D. Cohen's Kappa Coefficient

Cohen's kappa [9] is a statistic which measures levels of agreement between two raters which each classifies into several exclusive categories.

The value of Kappa is defined as:

$$\kappa=\frac{P_0-P_e}{1-P_e}, \quad (3)$$

when P_0 is the relative observed agreement among raters and P_e is the expected probability of chance agreement.

Kappa measures the percentage of data values in the main diagonal of the contingency table and then adjusts these values for the amount of agreement that could be expected due to chance alone.

A brief overview of nonparametric techniques discerns that kappa is most generally applied to predictive models build from unbalanced data. In this study, we utilize kappa coefficient for the purpose of data clearance and classification.

E. Repeated Random Sub-sampling Validation

To estimate how accurately our predictive model will perform in practice, we employed the repeated random sub-sampling validation (repeated-cv) technique. This method involves the following steps:

1: Randomly assign each observation into one of two groups: training and validation.

2: Fit the model to the observations in the training set.

3: Use the observations from the validation set to test the model's performance.

The method will repeat number of times and the ultimate results are then averaged over the slots.

III. PROPOSED FRAMEWORK

In this study, we constructed a data mining framework to explore large volumes of semiconductor manufacturing data for prognosticating defective tools and chambers at a determined production time. This framework includes four major steps: problem definition, data preparation, data mining and key factor screening and model construction, evaluation and interpretation as shown in Fig. 2.

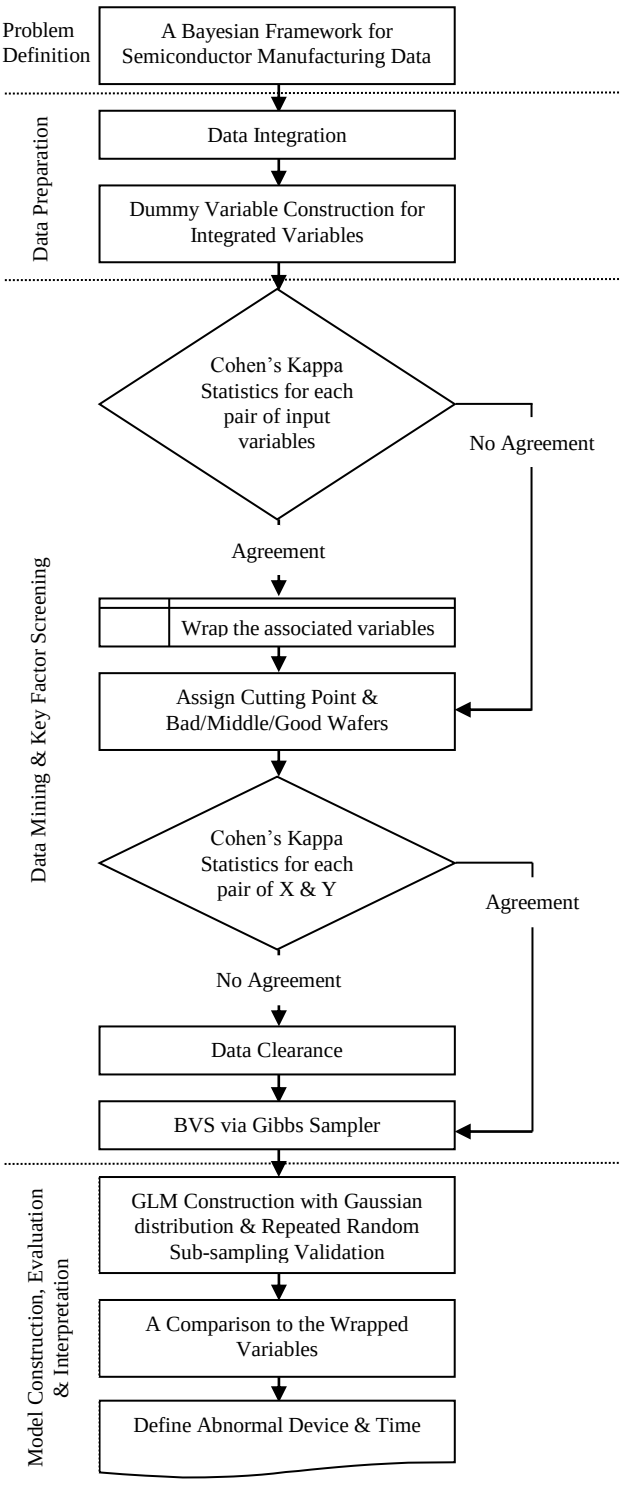

Figure 2. Research Framework

A. Problem Definition:

In practice, both results from knowledge-based and data-driven inference serve on diagnosing the yield-loss factors [10] where the rule-based expert system based on knowledge-based inference generates a priority chance for selecting the appropriate tools-chambers. This research is to identify the extraordinary process variables regards to their prior probability.

B. Data Preparation

As illustrated in Table 1, for our diagnosis objective, a simplified and comprehensive spreadsheet for the massive amount of information utilized to address the pairs of step-tool-chamber feature, the actual value will be binary, true (1) or false (0), indicating whether that tool-chamber was used in that step. This approach is able to dominate with the technical problem of missing information.

TABLE I. TRANSFORMED SAMPLE DATA

Wafer ID	**$Step_1$-$tool_1$-$chamber_1$**	**$Step_1$-$tool_1$-$chamber_2$**	**…**	**$Step_k$-$tool_j$-$chamber_i$**
w_1	0	1	…	1
…	…	…	…	…
w_n	1	0	…	1

C. Data Mining & Key Factor Screening

This study employed various types of statistical tools to wrap the associated variables, filter the unimportant factors and key factor screening via the following technique:

Cohen's Kappa Statistics for each pair of input variables: we use Cohen's Kappa as a measure of agreement between the two individuals (true (1) or false (0)) for each pair of binary predicted variables.

Wrap the associated variables: as a result of the kappa's interpretation, variables with high level of agreement (0.6~1) wrap with their peers in the same group where it is possible that a variable appears in more than one group.

Assign cutting point and bad, middle or good wafers: create a new dummy variable as an indicator of wafers level

Cohen's Kappa Statistics for each pair of X & Y: once again employ the Cohen's Kappa to remove insignificant variables, albeit this time for each set of response and predictor variables.

Data Clearance: follow the last step, predictor variables with the low level of agreement (0~0.2) will eliminate.

Bayesian Variable Selection via Gibbs Sampler: to determining which variables are included in the generalized linear model, we consider Bayesian strategies for performing this election. In particular, we focus on approaches based on the Gibbs sampler.

D. Model Construction, Evaluation & Interpretation

At the last step, firstly the generalized linear model (GLM) algorithm with Gaussian distribution is employed to construct the proper model via the selected key factors, henceforth, to evaluate the efficiency of model repeated random sub-sampling validation is adopted.

Secondly, recall factors associated to achieve the impressionable group factors.

Finally, we construct a time series graph to analyze the outputs of the tool-chamber machines at each process step. This phase is to explore the extensive process information to identify the possible root causes for specifying time cycle in the semiconductor manufacturing process.

IV. VALIDATION

Following the process framework, we implemented an empirical study and tested its performance in the task of the root causes detection of high and low yield. This yielded to reducing the cost and time caused by trial and error method.

A. Problem Definition:

The present problem involved 500 wafers of 20 lots with one CP yield as response variable and 100 process stages as predictor variables, which each lot passes through all the stages. As shown in Fig. 3 this problem induce the both high and low productivity in fab, engineers had to recall the related fabrication data with large varieties and complexity, find the root causes and replace the inadequate tools or chambers with lucrative ones.

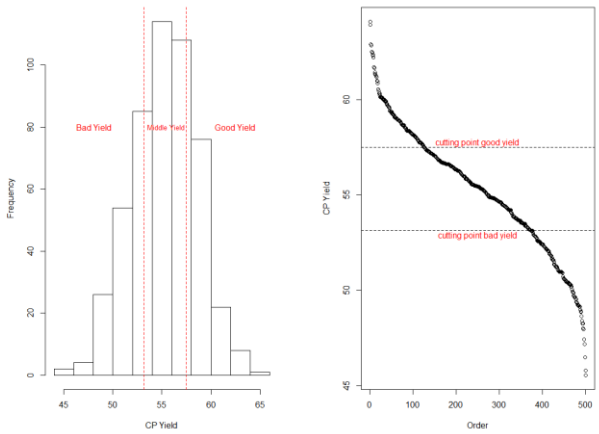

Figure 3. Scatter and histogram plot of sorted CP Yield

B. Data Preparation

To conform the framework and deal with nominal factors, information regards to 100 stage transfer to dummy variables, the transformed data include 1988 factors which each factor consists of the history of stages, tools, and chambers.

Since the raw data contained a lot of missing elements, data preparation was performed including imputation of missing elements, there were 1460 factors remained.

C. Data Mining & Key Factor Screening

After problem definition and data preparation, we use Cohen's Kappa statistics and Gibbs sampler to help us identify abnormal process stages and machines and provide this useful information to engineers as a reference for troubleshooting and defect diagnosis.

Phase 1: Totally 1,065,070 Cohen's Kappa Statistics computed for each pair of input variables, the distribution of the Kappa attribute is listed in Table 2.

Phase 2: The 25th and 75th quantiles of CP yield exploited as cutting points to classify the good, middle and bad yield. There were 250 wafers in the middle group, 125 wafers in the bad group and others were good groups. For analysis convenience, create a new variable containing the yield groups, the wafers in the middle group were marked as 2, in bad group as 1, and others (good group) were marked as 0. Cutting points are shown in Fig. 3. The cutting points were at 53.12% and 57.51% of yield rate. These marks assist us to distinguish more clearly if the bad or good wafers were fabricated from the same process stage. The descriptive statistics of the three groups are summarized in Table 3.

TABLE II. THE CLASS DISTRIBUTION FOR THE KAPPA TEST FOR EACH PAIR OF INPUT VARIABLES

Almost perfect agreement	**Substantial agreement**	**Moderate agreement**
3[a]	109	1,764
Fair agreement	**Slight agreement**	**No agreement**
24,539	280,081	758,574

a. Number of pairs at each level of agreement

TABLE III. BASIC STATISTICS OF CP YIELD GROUPS

Group	**Mean (%)**	**Standard deviation (%)**
Good wafers	59.33	1.41
Middle wafers	55.38	1.25
Bad wafers	50.88	1.58

Phase 3: Because there were too many process factors, we used Kappa statistics at this step to narrow the number of factors. The Kappa was applied to find out possible process factors with an appropriate measure of reliability. Similar to phase 1, Kappa statistic was used to compare the rating of the grouped yield with each individual dummy variable, to eliminate the influence of root cause factors, observations with mid-range value were removed. We wiped out process variables with no level of agreement $(\kappa \le 0.2)$. Indeed, after this phase, 411 predictor variables were identified as input for the next step.

Phase 4: To apply Bayesian inference using Gibbs sampler, we used the well-crafted "BayesVarSel" R package. Then, prior probabilities are estimated from the sample frequencies for each variable as follow:

$$\text{prior probability for j-th variable} = \frac{\sum_{i=1}^{n} I_{V_j}}{n}, \qquad (4)$$

when I_{V_j} is the j-th indicator variable and n denotes the sample size. Implementing a Gibbs sampler, reduces the number of important factors to 33.

D. Model Construction, Evaluation & Interpretation

The general linear model for finding the best linear relationship between the predictors and response variable was employed where the Gaussian family accepted to identify the response variable. To evaluate the effectiveness and practical viability of the proposed approach, two other conventional approaches, generalized boosted regression model (GBM) and random forest (RF), were selected for comparison. Through the adopted repeated-cv method, residual mean square error (RMSE) and adjusted R-squared were dedicated as evaluation criteria. The results are

summarized in Table 4, in which the importance of repetition and sample size nested with the result of cross-validation. From comparing the sampling distributions for the four models, it is apparent that, in this case, the GLM combined with Bayesian variable selection technique has an advantage.

TABLE IV. SUMMARIZING RESULTS FROM THE DISTRIBUTIONS OF EACH MODEL

Model[a]	**RMSE**			**Adjusted R-squared**		
	Min	*Median*	*Max*	*Min*	*Median*	*Max*
Gibbs + GLM	2.709	2.806	2.845	0.292	0.297	0.332
GBM + GLM	2.900	3.032	3.164	0.1782	0.181	0.183
RF + GLM	3.058	3.067	3.076	0.142	0.177	0.221
GLM	150.4	186.9	218	0.004	0.008	0.011

a. Number of resamples 2, Number of iterations 1

Model[a]	**RMSE**			**Adjusted R-squared**		
	Min	*Median*	*Max*	*Min*	*Median*	*Max*
Gibbs + GLM	1.842	2.653	2.841	0.046	0.371	0.711
GBM + GLM	2.534	3.051	3.332	0.000	0.053	0.337
RF + GLM	2.268	2.838	3.660	0.016	0.293	0.507
GLM	7.951	34.60	139.8	0.000	0.029	0.214

a. Number of resamples 20, Number of iterations 2

Seeking the validation and advantage of our framework, to evaluating the reliability of selected variables, reverted to the associated predictor variables by Cohen's kappa, comparison is shown 3 pairs of substantial agreement, 74 pairs of moderate agreement, 500 pairs of fair agreement and 6528 pairs of slight agreement between the selected and other predictor variables. We consider all factors with moderate and higher levels of kappa as the potential critical factors which can influence the growth of the yield enhancement.

The final stage of the framework is to mine the critical time spots. Fig. 4, illustrates the supporting time series for the selected rules by Gibbs sampler and critical factors selected by Cohen's kappa. The problem or perfection would be detected at a certain cycle time when a relatively large number of wafers deviate significantly at the high or low yield scope.

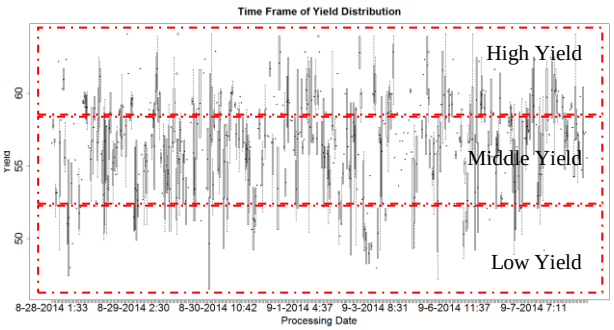

Figure 4. Scatter boxplot of selected factor's CP yield over the time

To promote the use and exploitation of the proposed framework, a part of final critical rules as a core structure for decision making summarized in Table 5.

V. CONCLUSION

The huge volume of data recorded on the line during the manufacture of chips makes it a natural application for data mining. Many studies have reported on their efforts to data mining. There have been notable successes in these efforts, mostly in the detective work for finding the cause of an unresolved problem in the fab. There, some specialized samples are collected to find root causes [11], [12], [13], [14] and [15].

Besides the data preparation, data clearance and variable selection are important works in the data mining process although, this part is very time-consuming, it cannot be ignored and needs much patience.

The proposed framework combines the Bayesian approach with traditional statistical methods and data mining viewpoint to explore huge semiconductor manufacturing data. Based on the empirical results, we validate that the proposed approach has practical viability, which means adding the efficacy of domain knowledge and experience to the system could improve results. Furthermore, using the domain knowledge might be to restrict conjunctions in rules to tools, chambers and steps that are related or occur within a reasonable time frame.

The data are not sampled from a stationary population, hence, over the time, the results may change significantly, and in addition some empirical answers might be reject based on engineer domain knowledge, which doesn't mean that the result is incorrect. Rather, the result may be a proxy for one or more events that are occurring elsewhere or at the other periods of the time, hence, the simulation study is an essential tool for evaluating the accuracy of our proposed method. Also, as a part of the simulation study future study can be done to develop the effect of prior probability to consider the performance of the Gibbs sampler.

TABLE V. TELIC DECISION TABLE CORE STRUCTURE

Factors	**Date**	
	Bad	*Good*
Stage10 - Tool2 - Chamber3	befor 8/29/2014 2:32	after 8/29/2014 12:50
Stage12 - Tool2 - Chamber1	between 8/30/2014 3:26 & 8/30/2014 3:43	before 8/29/2014 10:55
Stage12 - Tool2 - Chamber4	after 8/29/2014 7:36 till 8/30/2014 3:44	before 8/29/2014 7:36
Stage13 - Tool5 - Chamber2	-	generally effected the high yield
Stage17 - Tool2 - Chamber2	after 8/30/2014 12:21	befor 8/30/2014 10:37
Stage23-Tool3-Chamber2	-	generally effected the high yield
Stage44 - Tool7.- Chamber2 and Chamber3	at 9/3/2014	at 9/1/2014
Stage49 - Tool1.- Chamber4	at 9/3/2014	at 9/2/2014
Stage57 - Tool1.- Chamber3	-	generally effected the high yield

ACKNOWLEDGEMENTS

This research is supported by the Ministry of Science and Technology, Taiwan (MOST 103-2218-E-007-023; MOST 104-2622-E-007-002) and the Advanced Manufacturing and Service Management Research Center of National Tsing Hua University, Taiwan (104N2074E1).

REFERENCES

[1] L. Argote and D. Epple, "Learning curves in manufacturing," Science, vol. 247, no. 4945, pp. 920–924, 1990.

[2] N.W. Hatch and D.C. Mowery, "Process innovation and learning by doing in semiconductor manufacturing," Management Science, vol. 44, no. 11, pp. 1461-1477, 1998.

[3] M. Effron, "Integrated yield management: a systematic approach to yield management," Advanced Semiconductor Manufacturing Conference and Workshop, Cambridge, MA, pp. 397-403, 1996.

[4] K. W. Tobin, T. P. Karnowski, and S. S. Gleason, "Integrated Yield Management," Analytical and diagnostic techniques for semiconductor materials, Leuven, Belgium, pp. 298-310, 1999

[5] D.-D. Hu, C.-M. Liu, C.-M. Huang and L.-C. Chen, *Integrated defect yield management and query system*, USA Invention Patent, US 6,314,379 B1, 2001

[6] K. P. Murphy, *Machine learning: a probabilistic perspective*, p. 35. MIT press. 2012

[7] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions. Volume 1: Models and Applications*, New York: Wiley, 2000

[8] W. R. Gilks, S. Richardson and. D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996

[9] J. Cohen, "A coefficient of agreement for nominal scales". Educational and Psychological Measurement, vol.20, no.1, pp. 37–46, 1960

[10] C.-M. Fan and Y.-P. Lu, "A Bayesian Framework to Integrate Knowledge-Based and Data-Driven Inference Tools for Reliable Diagnoses", Proceedings of the 40th conference on winter simulation, Austin, TX, pp. 2323 – 2329, 2008

[11] M. Gardner, J. Bieker, "Data Mining Solves Tough Semiconductor Manufacturing Problems", Proceedings of the 6th ACM SIGKDD international conference on knowledge discovery and data mining, Boston, MA, pp. 376-383, 2000

[12] D.-H. Baek, I.-J. Jeong and C. H. Han, "Application of Data Mining for Improving Yield in Wafer Fabrication System", Computational Science and Its Applications, vol. 3483, no. 4, pp. 222-231, 2005

[13] C.-F. Chien, W.-C Wang and J.-C. Cheng, "Data Mining for Yield Enhancement in Semiconductor Manufacturing and an Empirical Study", Expert Systems with Applications, vol.33, no.1, pp. 192-198, 2007

[14] S. M. Weiss, R. J. Baseman, F. Tipu, C. N. Collins, W. A. Davies, R. Singh and J. W. Hopkins, "Rule-Based Data Mining for Yield Improvement in Semiconductor Manufacturing", Applied Intelligence, vol.33, no.1, pp.318-329, 2010

[15] C.-F. Chien and S.C Chuang, "a Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analysis", IEEE Transactions on Semiconductor Manufacturing, vol.27, no.4, pp.475-488, 2014