Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

BIG DATA ANALYTICS FOR MODELING WAT PARAMETER VARIATION INDUCED BY PROCESS TOOL IN SEMICONDUCTOR MANUFACTURING AND EMPIRICAL STUDY

Chen-Fu Chien
Ying-Jen Chen

Department of Industrial Engineering & Engineering Management
National Tsing Hua University
101 Section 2 Kuang-Fu Road
Hsinchu 30013, TAIWAN

Jei-Zheng Wu

Department of Business Administration
Soochow University
56 Section 1 Kueiyang Street
Taipei 10048, TAIWAN

ABSTRACT

With the feature size shrinkage in advanced technology nodes, the modeling of process variations has become more critical for troubleshooting and yield enhancement. Misalignment among equipment tools or chambers in process stages is a major source of process variations. Because a process flow contains hundreds of stages during semiconductor fabrication, tool/chamber misalignment may more significantly affect the variation of transistor parameters in a wafer acceptance test. This study proposes a big data analytic framework that simultaneously considers the mean difference between tools and wafer-to-wafer variation and identifies possible root causes for yield enhancement. An empirical study was conducted to demonstrate the effectiveness of proposed approach and obtained promising results.

1 INTRODUCTION

Driven by Moore's law (Moore 1965), semiconductor manufacturing technology has evolved continually from a mature process to an advanced process. Because complementary metal-oxide-semiconductor (CMOS) transistors are scaled to a nanometer feature size range, in-line process control plays a critical role from the viewpoint of yield enhancement. Indeed, tool-induced process variations affect parametric integrated-circuit (IC) product yield. Parametric yield is a measure of the quality of functioning systems, whereas functional yield measures the proportion of functioning units produced by a manufacturing process (May and Spanos 2006). In particular, variations in transistor parameters at the wafer acceptance test (WAT) stage are the main cause of parametric yield losses.

In practice, end-of-line quality control at the WAT stage follows a sequential detection and diagnosis approach (Fan et al. 2000). Transistor parameters are first monitored in the acceptance sampling test module and the statistical process control (SPC) module. If an out-of-spec signal is detected in the acceptance sampling test module or an out-of-control signal is detected in the SPC module, a preliminary diagnosis is performed, and potential process steps are screened. According to the screening results and stratification data from these critical steps, an in-depth diagnosis on the basis of domain knowledge is performed. However, this process is time consuming, and its performance varies from person to person. A systematic and data-driven tool is required to shorten the response time while maintaining high confidence in terms of identification of possible root causes upon the receipt of an alarm. In addition, multiple components of variation exist in transistor parameters—lot-to-lot, wafer-to-wafer, site-to-site, and residual—which are not considered adequately in traditional analysis of variance (ANOVA) techniques.

Focusing on the needs of digital decisions at the WAT stage, this study proposes a big data analytic framework for modeling transistor parameter variations induced by process tools to support yield enhancement in semiconductor manufacturing. In particular, the proposed approach integrates forward

stepwise screening for heteroscedastic regression, the least absolute shrinkage and selection operator (Lasso) (Tibshirani 1996), and bootstrap techniques (Efron 1979; Efron and Tibshirani 1993) to consider the mean difference between tools and wafer-to-wafer variation simultaneously with high confidence. An empirical study was conducted to validate the proposed approach in a leading semiconductor company in Taiwan. The results show the viability of the proposed approach.

2 LITERATURE REVIEW

In the semiconductor industry, yield is defined as the fraction of total input transformed into shippable output. In practice, total yield loss can be divided into three categories: line yield loss, which is the fraction of wafers discarded before reaching the WAT; die yield loss, which is the fraction of dice on wafers not discarded before reaching assembly and the final test; and final test yield loss, which is the fraction of semiconductor devices that are unacceptable for shipment. In particular, die yield loss can be decomposed into functional yield loss and parametric yield loss. Functional yield loss consists of dice that do not function, whereas parametric yield loss consists of dice that do function but not according to specification. Functional yield losses are usually caused by particulate defects, scratches, and contamination during the manufacturing process. By contrast, parametric losses are usually caused by process variations that cause the die to perform differently from specifications, including a lower frequency, slower speed, and incompatible voltage range (Cunningham et al. 1995; May and Spanos 2006). With the shrinking feature size of semiconductor devices, manufacturing technologies including equipment and environments have advanced considerably to reduce yield loss. Hence, reducing parametric yield losses caused by process variations to support yield enhancement is becoming increasingly critical in modern semiconductor manufacturing. Because the transistor parameter variation at the WAT stage is usually highly correlated with parametric yield losses, modeling and control of the transistor parameter variations is important.

The use of SPC for monitoring transistor parameters at the WAT stage has been reported. Fan et al. (2000) proposed a methodology for generating robust design parameters to simultaneously apply Shewhart and Exponentially Weighted Moving Average (EWMA) control charts to WAT data. Currently, data mining and big data analytics approaches have been developed to extract potentially useful information and manufacturing intelligence from massive data in semiconductor manufacturing, including demand forecasting (Chien et al. 2010), human resource management (Chen and Chien 2011), troubleshooting (Chien et al. 2007; Chien and Chuang 2014), advanced process control/advanced equipment control (Chien et al 2013; Chien et al. 2015), and wafer bin map/defect image classification (Chien et al. 2013; Liu and Chien 2013; Chen et al. 2013; Chen et al. 2016). In particular, Hwang and Lee (2014) proposed the use of hidden variable logistic regression to identify critical process variables and address missing observations for modeling parametric yield. Pan et al. (2011) developed a virtual metrology system for predicting end-of-line electrical properties by using a multivariate analysis of covariance (MANCOVA) model with tools clustering. However, little research has been conducted to address issues related to the modeling of transistor parameter variations induced by process tools from the viewpoint of reducing parametric yield loss.

3 PROPOSED APPROACH

3.1 Problem Definition

A typical semiconductor manufacturing process contains a WAT stage at the end of the process line to ensure outgoing quality. At the WAT stage, more than 100 transistor parameters are required for inspection. Such parameters are usually associated with electronic properties, and variations in them cause yield loss. Such loss is termed parametric yield loss (Agarwal et al. 2007). To reduce parametric yield loss, it is critical to effectively and efficiently identify process tools that cause variations by using big data analytics for further decision making.

For each transistor parameter, the contributors to variation can be decomposed into several levels: within-die variation, die-to-die variation, wafer-to-wafer variation, lot-to-lot variation, and tool-to-tool variation. Indeed, process tools usually have their own characteristics, and tool misalignment considerably affects process variations in semiconductor manufacturing (Chien et al. 2015). In addition, wafer-level is the most commonly used granularity for analyzing advanced process data in practice; therefore, wafer-to-wafer variation is critical and should be addressed. Hence, the present study focuses not only on tool-to-tool variation but also on wafer-to-wafer variation. Furthermore, this study defined the following terminologies:

- A process tool has a location effect if it is one of the root causes of tool-to-tool variation.
- A process tool has a dispersion effect if it is one of root causes of wafer-to-wafer variation.
- A transistor parameter follows a location-dispersion model if there exist process tools with location and/or dispersion effects.
- A transistor parameter follows a location-only model if there exist process tools with location effects.

3.2 Data Preparation

In the data preparation phase, users must first choose a target transistor parameter as the response variable for analysis and the query-related process stages as features, on the basis of domain knowledge. Two key data preparation issues were addressed in this study for categorical features: missing value imputation and collinearity exploration to enhance data quality and model results.

Although modern semiconductor manufacturing fabs are fully automated, process tool values are usually missing according to a nonrandom missing mechanism and time-domain missing patterns. In advanced semiconductor processes, the data volume is usually small because of ramping, and elimination of data containing missing values may result in a scenario in which no available data can be used. To resolve this difficulty, this study proposes a forward k-nearest neighbor algorithm for the effective imputation of categorical data as follows:

1. Compute the similarity matrix between observations for the original data set.
2. Compute the number of missing values for each observation and arrange them in ascending order.
3. According to the order, follow sequentially the following steps for each observation:
   (a) Obtain stage names with missing values.
   (b) Obtain the descending order of observations according to similarity.
   (c) Take k nearest neighbors and use majority vote to impute missing values for each stage.

In addition, v-fold cross validation can be employed to validate the proposed imputation algorithm.

Fixed process tool combination in short loops is another common situation in the ramping phase of advanced processes, and this is attributed to yield concerns. This situation causes stage collinearity in data structures and renders statistical models unreliable. We propose the use of a hierarchical clustering approach based on Cramer's V coefficient (Cramer 1946) to provide an overview of stage collinearity before model construction. Given the threshold of Cramer's V coefficient, we automatically combine stages that are highly correlated and use process tool combination for further analysis to enhance model reliability. Indeed, engineers must endeavor to distinguish the most likely root causes among stages with collinearity. The tool with the most observations is the baseline (golden tool) for each stage.

3.3 Model Construction

The model construction process involves two subprocesses: stage-level screening and effective tool identification. The objective of stage-level screening is to narrow down the range of suspected stages that could contain misaligned process tools from the entire data set. Effective tool identification is then employed to further identify specific tools with significant evidence. In this phase, two possible models are derived, namely a location-dispersion model and location-only model, and the model used depends on

whether the dispersion effect exists. In particular, the location-only model is equivalent to a classical linear regression model under the following assumptions:

$$Y_i = E(Y_i \mid \mathbf{x}_i) + \varepsilon_i = \beta_0 + \sum_j x_{ij}\beta_j + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} Normal(E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2) \quad (1)$$

That is, the residual is independently and identically distributed normally. By contrast, the location-dispersion model is a heteroscedastic linear model with the following assumptions:

$$Y_i = E(Y_i \mid \mathbf{x}_i) + \varepsilon_i = \beta_0 + \sum_j x_{ij}\beta_j + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} Normal(E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma_i^2) \quad (2)$$

$$E({\sigma_i}^2 \mid \mathbf{x}_i) = \exp(\gamma_0 + \sum_l x_{il}\gamma_l) \quad (3)$$

To model nonconstant variance in the location-dispersion model, a generalized linear model (GLM) with Gamma distribution and log link (Myers et al. 2010) is proposed for parameter estimation. To inspect potential dispersion effects in the data set, the Breusch-Pagan test (Breusch and Pagan 1979) is performed after the construction of a classical linear regression model.

3.3.1 Stage-level Screening

Semiconductor manufacturing contains hundreds of process stages with parallel process tools; the sample size is quite small in the ramping phase. In other words, it is an ultrahigh dimensional data structure, and stage-level screening is highly difficult. Forward stepwise regression is a frequently used and classical variable screening method, and it has been shown to identify all relevant predictors consistently, even if the predictor dimension is substantially larger than the sample size (Wang 2009). Hence, the forward stepwise strategy is applied to screen active stages with the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as follows:

$$\text{AIC} = -2\log L + 2 \cdot p \quad (4)$$

$$\text{BIC} = -2\log L + [\log N] \cdot p \quad (5)$$

where L represents the likelihood function based on the location-dispersion model, as given by (6); N represents the sample size; and p represents the number of parameters in the model.

$$L(\beta_j, \gamma_l \mid y_i) = \prod_{i=1}^{n} \left(\sqrt{2\pi} \exp\left[\tfrac{1}{2} \sum_l x_{il}\gamma_l \right] \right)^{-1} \exp\left[- \frac{\left(y_i - \sum_j x_{ij}\beta_j \right)^2}{2\exp(\sum_l x_{il}\gamma_l)} \right] \quad (6)$$

Indeed, the BIC has a higher penalty than the AIC does on a number of parameters. Hence, the BIC tends to entail selecting several active stages with fewer process tools, whereas the AIC usually includes stages containing minor-effect tools that may not be adequately significant in the regression model. This implies that the BIC is a conservative criterion for achieving a lower false alarm rate of selected stages, whereas the AIC may detect a greater number of actual root causes. Given the trade-off between the AIC

and BIC, it is necessary to consider decision makers' preference when judging which criterion would be more suitable in practice.

3.3.2 Effective Tool Identification

If the dispersion effect is not significant, a typical linear regression is constructed for the location-only model, and the proposed method calculates a signal-to-noise index (SNI), according to Equation (7), to examine the fitness of the linear regression. When the SNI exceeds the specified threshold, the regression is considered to have adequate capability for the location-only model. Thus, statistical inference with the derived confidence interval provides evidence to support the decision of identifying effective tools among selected stages. Conversely, random forest (RF) (Breiman 2001), a machine learning–based approach, is used to identify effective tools among all stages because the linear regression lacks evidence of screening results.

$$SNI = \frac{\max_j \left|\hat{\beta}_j\right|}{s^2} \tag{7}$$

RF is an ensemble learning method that can handle a multivariate problem with high dimensionality and collinearity by aggregating several decisions or predictions of weak learners (Breiman 2001; Verikas et al. 2011). RF is appropriate for evaluating factor importance in semiconductor data because it can achieve a favorable trade-off between the explanationability and singularity of factors on the basis of the measurement of the increase in the mean square of predicted errors (Chien and Chuang 2014; Chien, Liu, and Chuang 2015). The present study proposes a hybrid RF–AIC procedure for identifying effective tools.

Regarding the location-dispersion model after stage-level screening, an iterative Lasso procedure is proposed for identifying effective tools and estimating tool effects. The Lasso procedure is vital for variable selection and estimation in high dimensions by shrinking coefficients in a linear model to achieve a trade-off between diminished variance and increased bias. Therefore, the Lasso approach can be computed efficiently even when p is extremely high, and it often improves the accuracy of predictions (Hastie et al. 2009). The Lasso approach can be extended to heteroscedastic regression, as has been conducted in the context of bioinformatics (Daye et al. 2011).

Equation (8) shows Lasso estimators considering heteroscedasticity in the location-dispersion model.

$$\left(\tilde{\beta}_j, \tilde{\gamma}_l\right) = \arg\min\{\log L\} \qquad \text{subject to } \sum_j \left|\beta_j\right| < t_1, \sum_l \left|\gamma_l\right| < t_2 \tag{8}$$

where L is given by Equation (6), and t_1 and t_2 are tuning parameters pertaining to location and dispersion, respectively. If the tuning parameters are infinity, the Lasso estimators are equivalent to the result of stage-level screening. By contrast, the Lasso estimators tend to shrink the estimated results in stage-level screening toward zero when the tuning parameters are small. Therefore, only a subset of coefficients is nonzero when tunning parameters are given. In particular, with the shrinkage of the tanning parameters, minor effects are more possibly restricted to zero, whereas only significant effects are presented.

3.4 Result Evaluation and Interpretation

To validate the reproducibility of effective tool identification using Lasso, the bootstrap technique, a general tool for assessing statistical accuracy, is used to provide more evidence for effective tools. The basic idea of the bootstrap technique is to create m replications with the same sample size by using

sampling with replacement at first, and then fit the proposed iterative Lasso model with the same parameter setting for each replication. If a tool is identified by the Lasso model in each bootstrap replication, we can conclude that the tools have location effects or dispersion effects with strong evidence. In addition, a scatter plot between actual values y_i and fitted values $\hat{y}_i$ with an adjusted R-squared is used for the location-dispersion model (LD-adj. R^2) index to present the fitness of the location-dispersion model for overall model assessment. In particular, the fitted values of the location-dispersion model are defined by Equation (9).

$$\hat{y}_i = \hat{y}_i^{location} + sign(y_i - \hat{y}_i^{location}) \cdot \tilde{\sigma}_i \tag{9}$$

where $\hat{y}_i^{location}$ represents the fitted value from the location model and $\tilde{\sigma}_i$ represents the fitted value from the dispersion model. Therefore, LD-adj. R^2 is defined as Equation (10).

$$LD - adj.\ R^2 = \frac{\sum_i (y_i - \hat{y}_i)^2 \Big/ (N-q)}{\sum_i (y_i - \bar{y})^2 \Big/ (N-1)} \tag{10}$$

If $LD - adj.\ R^2$ is close to one, a major variation in the data set can be modeled by process tools, and engineers should focus on tools that have high reproducibility in bootstrapped Lasso for troubleshooting and process control; otherwise, domain knowledge should be engaged to explore latent root causes that are not included in the model.

4 EMPIRICAL STUDY

This study analyzed a real case to demonstrate the proposed approach. The entire data set contained a given transistor parameter as the analysis target, 29 stages derived after domain knowledge screening, and 5500 wafers with a missing rate of approximately 40%. The distribution of the transistor parameter was approximately normal, with a mean of 0.5363 and standard deviation of 0.0244. Domain experts believe that tool-induced variation exists in the transistor parameter; therefore, the objective was to identify possible root causes for decision support.

For data preparation, this study used the proposed forward k-nearest neighbor algorithm with 10-fold cross validation to impute missing values. The results showed that the imputation accuracy for this data set was 82.5%. In addition, this study applied the hierarchical clustering approach on the basis of Cramer's V coefficient to provide an overview of stage collinearity. As shown in Figure1, the data set used herein did not have a severe stage collinearity issue; hence, no transformation to reduce the collinearity effect was required.

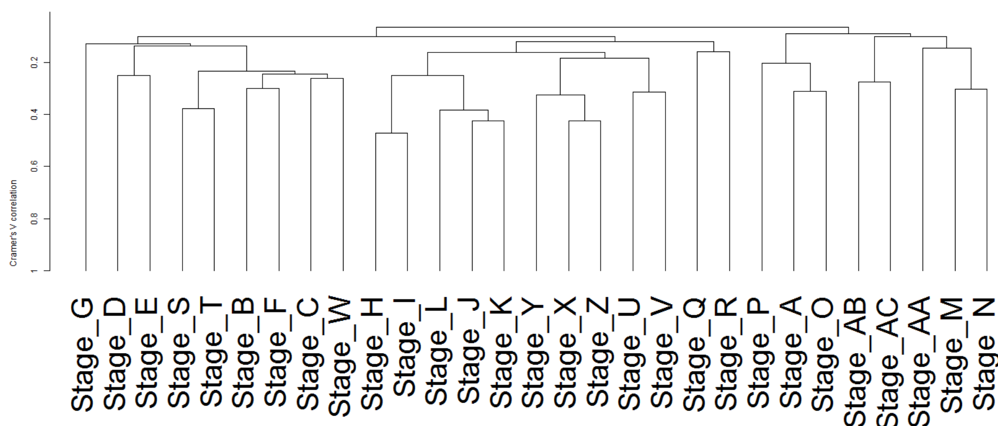

Figure 1. Collinearity exploration.

For model construction, this study applied the location-dispersion model to the data set because a significant dispersion effect was detected by the Breusch-Pagan test. This study involved four stages (Stage_F, Stage_N, Stage_P, and Stage_S) for location effects and two stages (Stage_U and Stage_Z) for dispersion effects after stage screening. Furthermore, this study used the proposed iterative Lasso approach with the bootstrap technique to identify effective tools for location and dispersion, as shown in Figures 2 and 3, respectively. In particular, tools denoted by red points have 100% reproducibility in every bootstrap replication, whereas those denoted by orange points have 70% reproducibility. In addition, Figure 4 presents the overall model assessment in terms of a scatter plot of the actual values and the predicted values in Figure 4(a) and a normal Q-Q plot of model residuals in Figure 4(b).

Since the LD-adj. R^2 is 0.692 and the Q-Q plot satisfies the normality assumption, the analysis results can provide reliable evidence for making domain judgments. Regarding location, 6 effective tools were identified from 58 tools in 4 stages. Tool_N6 and Tool_N5 had the highest positive effects, whereas Tool_P7 had a negative effect for mean shift. By contrast, regarding dispersion, 3 effective tools were identified among 22 tools in 2 stages. Tool_U11, Tool_Z5, and Tool_Z9 may be assignable causes for variance shift. On the basis of these results, domain experts can trace the historical events of these tools and fix the problem quickly.

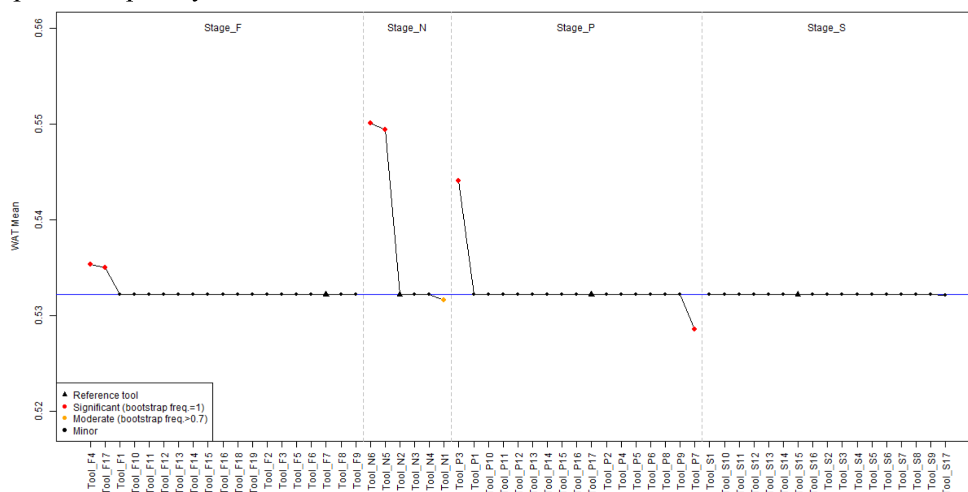

Figure 2. Visualization of location-effective tools.

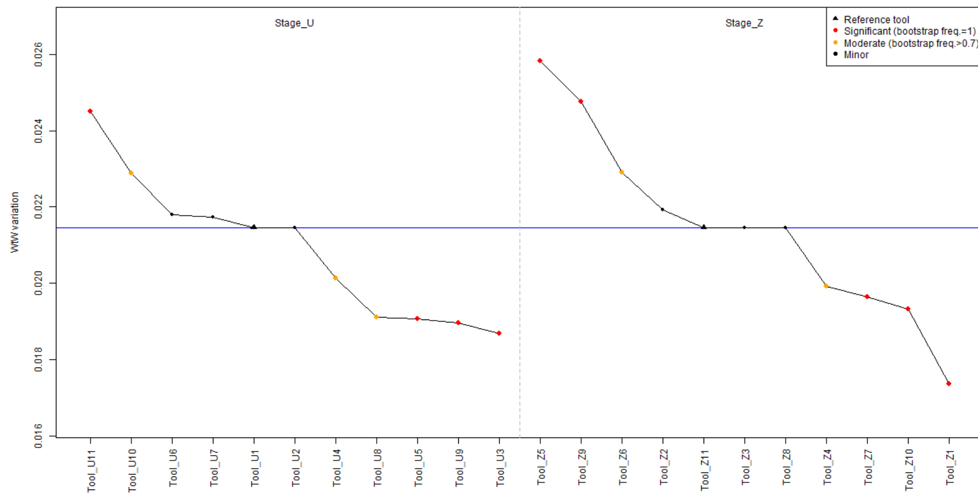

Figure 3. Visualization of dispersion-effective tools.

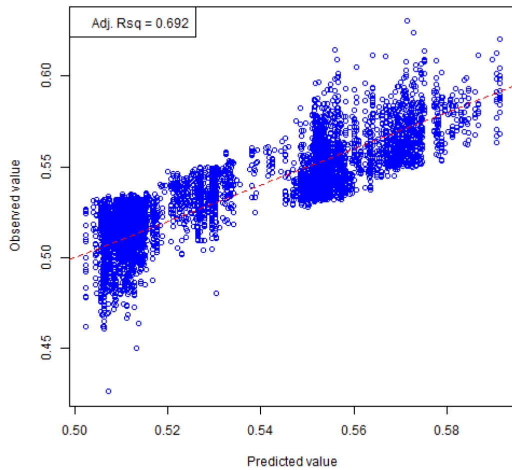

Figure 4(a). Scatter plot.

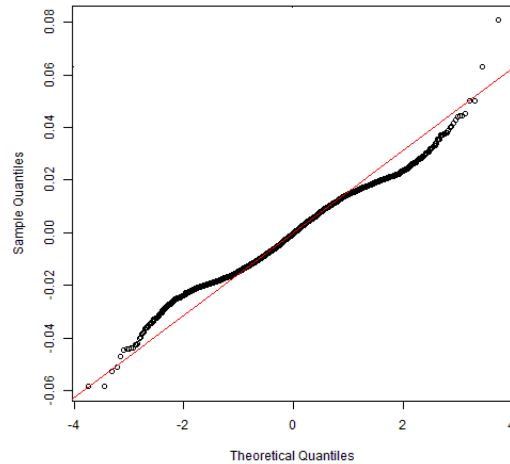

Figure 4(b). Normal Q-Q plot.

5 CONCLUSION

Because wafer fabrication is reaching nanotechnology nodes, developing data-driven tools to support yield enhancement decisions effectively and efficiently is necessary. Moreover, modeling of transistor parameter variation at the WAT stage is critical for reducing parametric yield losses. Therefore, this study proposes a big data analytic framework that integrates various tools including forward stepwise, Lasso, and RF to derive appropriate models for identifying effective tools for location and dispersion. Through an empirical study and experimental design based on a real data set collected from a leading semiconductor manufacturing company, this study validated the proposed approach and showed that it outperforms the individual methods. As semiconductor fabs become more intelligent, future studies can focus on developing fab-wide advanced process control and advanced equipment control techniques based on the results extracted from big data analytics to empower manufacturing intelligence. In addition, the use of more complex statistical models such as generalized linear mixed models (GLMMs) is suggested for modeling die-level data without violating model assumptions (Krueger and Montgomery 2014).

ACKNOWLEDGMENTS

This research was supported by the Ministry of Science and Technology, Taiwan (MOST 103-2218-E-007-023; MOST 104-2622-E-007-002; MOST 105-2622-8-007-002-TM1; MOST 104-2410-H-031-033-MY3). The authors thank domain experts for their assistance with empirical studies and validation.

REFERENCES

Agarwal, K., R. Rao, D. Sylvester, and R. Brown. 2007."Parametric Yield Analysis and Optimization in Leakage Dominated Technologies." *IEEE Transactions on VLSI* 15:613–623.

Breiman, L. 2001."Random Forests." *Machine Learning* 45:5–32.

Breusch, T.S., and A.R. Pagan. 1979."A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47:1287–1294.

Chen, L.-F., and C.-F. Chien. 2011."Manufacturing Intelligence for Class Prediction and Rule Generation to Support Human Capital Decisions for High-Tech Industries." *Flexible Services and Manufacturing Journal* 23: 263-289.

Chen, Y.-J., C.-Y. Fan, and K.-H. Chang. 2016."Manufacturing Intelligence for Reducing False Alarm of Defect Classification by Integrating Similarity Matching Approach in CMOS Image Sensor Manufacturing." *Computers & Industrial Engineering*, DOI: 10.1016/j.cie.2016.05.009.

Chen, Y.-J., T.-H. Lin, K.-H. Chang, and C.-F. Chien. 2013."Feature Extraction for Defect Classification and Yield Enhancement in Color Filter and Micro-lens Manufacturing and an Empirical Study." *Journal of Industrial and Production Engineering* 30: 510-517.

Chien, C.-F., Y.-J. Chen, and C.-Y. Hsu. 2015."A Novel Approach to Hedge and Compensate the Critical Dimension Variation of the Developed-and-Etched Circuit Patterns for Yield Enhancement in Semiconductor Manufacturing." *Computers & Operations Research* 53: 309-318.

Chien, C.-F., Y.-J. Chen, and J.-T. Peng. 2010."Manufacturing Intelligence for Semiconductor Demand Forecast Based on Technology Diffusion and Product Life Cycle." *International Journal of Production Economics* 128: 496-509.

Chien, C.-F., and S.-C. Chuang. 2014."A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics." *IEEE Transactions on Semiconductor Manufacturing* 27: 475-488.

Chien, C.-F., C.-Y. Hsu, and P.-L. Chen. 2013."Semiconductor Fault Detection and Classification for Yield Enhancement and Manufacturing Intelligence." *Flexible Services and Manufacturing Journal* 25: 367-388.

Chien, C.-F., S.-C. Hsu, and Y.-J. Chen. 2013."A System for Online Detection and Classification of Wafer Bin Map Defect Patterns for Manufacturing Intelligence." *International Journal of Production Research* 51:2324-2338.

Chien, C.-F., C.-W.Liu, and S.-C.Chuang.2015."AnalysingSemiconductor Manufacturing Big Data for Root Cause Detection of Excursion for Yield Enhancement." *International Journal of Production Research*, article in press, DOI:10.1080/00207543.2015.1109153.

Chien, C.-F., W. Wang, and J. Cheng. 2007."Data Mining for Yield Enhancement in Semiconductor Manufacturing and an Empirical Study." *Expert Systems with Applications* 33: 192-198.

Cramér, H. 1946. *Mathematical Methods of Statistics*.Princeton: Princeton University Press.

Cunningham, S.P., C.J. Spanos, and K. Voros. 1995."Semiconductor Yield Improvement: Results and Best Practices." *IEEE Transactions on Semiconductor Manufacturing* 8: 103-109.

Daye, Z.J., J. Chen, and H. Li. 2012."High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis." *Biometrics* 68:316-326.

Efron, B. 1979."Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7: 1-26.

Efron, B, and R. Tibshirani. 1993. *An Introduction to the Bootstrap*.London: Chapman and Hall.

Fan, C.-M., R.-S. Guo, S.-C. Chang, and C.-S. Wei. 2000."SHEWMA: An End-of-Line SPC Scheme Using Wafer Acceptance Test Data." *IEEE Transactions on Semiconductor Manufacturing* 13: 344-358.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York:Springer.

Hwang, J.Y.; and H.C. Lee. 2014."Parametric Yield Modeling Using Hidden Variable Logistic Regression." *Journal of Quality Technology* 46: 323-339.

Krueger, D.C., and D.C. Montgomery. 2014."Modeling and Analyzing Semiconductor Yield with Generalized Linear Mixed Models." *Applied Stochastic Models in Business and Industry* 30: 691-707.

Liu, C.-W., and C.-F. Chien. 2013."An Intelligent System for Wafer Bin Map Defect Diagnosis: An Empirical Study for Semiconductor Manufacturing." *Engineering Applications of Artificial Intelligence* 26: 1479-1486.

May, G.S., and C.J. Spanos. 2006. Fundamentals of Semiconductor Manufacturing and Process Control.

Moore, G.E. 1965."Cramming More Components onto Integrated Circuits." *Electronics Magazine* 38:114-117.

Myers, R.H., D.C. Montgomery, G.G. Vining, and T.J. Robinson. 2010. *Generalized Linear Models: With Applications in Engineering and the Sciences*. 2nd ed. Hoboken, N.J: John Wiley & Sons.

Pan, T.-H., B.-Q. Sheng, S.-H. Wong, S.-S. Jang. 2011."A Virtual Metrology System for Predicting End-of-Line Electrical Properties Using a MANCOVA Model With Tools Clustering." *IEEE Transactions on Industrial Informatics* 7: 187-195.

Tibshirani, R. 1996."Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267-288.

Verikas, A., A. Gelzinis, M. Bacauskiene. 2011."Mining Data with Random Forests: A Survey and Results of New Tests." *Pattern Recognition* 44: 330–349.

Wang, H. 2009."Forward Regression for Ultra-High Dimensional Variable Screening." *Journal of the American Statistical Association* 104:1512-1524.

Zou, H., T. Hastie, R. Tibshirani. 2007."On the 'Degrees of Freedom' of the Lasso." *Annals of Statistics* 35:2173-2192.

AUTHOR BIOGRAPHIES

CHEN-FU CHIEN is a Tsing Hua Chair Professor in NTHU, Taiwan. He is also the Principal Investigator of the Semiconductor Technologies Empowerment Partners Consortium (STEP Consortium) and the NTHU-TSMC Center for Manufacturing Excellence. From 2005 to 2008, he was on leave as the Deputy Director of Industrial Engineering Division at Taiwan Semiconductor Manufacturing Company (TSMC), the largest semiconductor foundry in the world. His research efforts center on decision analysis, modeling and analysis for semiconductor manufacturing, manufacturing strategies, and manufacturing intelligence. He has received 11 U.S. patents on semiconductor manufacturing and has published three books, more than 140 journal papers, and a number of case studies with Harvard Business School. He is an Area Editor of *Flexible Services and Manufacturing Journal*, Advisory Board Member of *OR Spectrum*, and Editorial Board Member of a number of international journals including *Computers & Industrial Engineering* and *Journal of Intelligent Manufacturing*. His email address is cfchien@mx.nthu.edu.tw.

YING-JEN CHEN is a postdoctoral researcher with the STEP Consortium & NTHU-TSMC Center for Manufacturing Excellence in Hsinchu, Taiwan. He received his PhD in Industrial Engineering and Engineering Management (IEEM) from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2013. He received the Best Paper Award at the CIIE Annual Meeting (2010, 2012-2014) and the Best Presentation Award at the 2015 International Symposium on Semiconductor Manufacturing Intelligence

(ISMI2105). His works have been published in *IEEE Transactions on Automation Science and Engineering*, *Computers & Operations Research*, *Computers & Industrial Engineering*, and *International Journal of Production Research*. His research interests include quality engineering, advanced process control, data mining & big data analytics, and yield enhancement. His email address is yjchen@ie.nthu.edu.tw.

JEI-ZHENG WU is Associate Professor in the Department of Business Administration, Soochow University (SCU), Taipei, Taiwan. He received the Quality Paper Award from Chinese Society for Quality, Award for Distinguished Performance in Industry–Academia Collaboration from National Science Council, Outstanding Researcher Scholarship from National Science Council, Research Award from Soochow Business Administration Education Foundation, Research Publication Prize from Soochow University, the Best Paper Award at the Twelfth Asia-Pacific Industrial Engineering & Management Systems (APIEMS 2011) conference, the Best Paper Award at the CIIE Annual Meeting (2011 and 2010), and the Young Scientist Prize at the Intelligent Manufacturing & Logistics Systems International Conference in 2008. His main research interests include manufacturing strategy, operations management, supply chain management, decision analysis, metaheuristics, decision support systems, and management and applications of telematics. Dr. Wu serves as Associate Editor of *International Journal of Industrial Engineering: Theory, Applications and Practice* (IJIETAP) (SCIE). He has also served as Guest Editor of Journal of Quality (EI). His works have been published in SCIE/SSCI–indexed journals including *International Journal of Logistics Management, Computers & Industrial Engineering*, *OR Spectrum*, *IEEE Transactions on Semiconductor Manufacturing*, *International Journal of Production Research*, *Journal of Intelligent Manufacturing*, *International Journal of Shipping and Transport Logistics*, *Growth and Change*, *Expert Systems and Applications*, *INFORMATION—An International Interdisciplinary Journal*, *NTU Management Review,* and other international journals. His email address is jzwu@scu.edu.tw**.**