A Novel Bayesian Network-Based Fault Prognostic Method for Semiconductor Manufacturing Process

Guodong Wang*, Ramin M. Hasani*, Yungang Zhu[†‡§] and Radu Grosu*

*Institute of Computer Engineering, Vienna University of Technology, Austria

{guodong.wang, ramin.hasani, radu.grosu}@tuwien.ac.at

[†]College of Computer Science and Technology, Jilin University, China

[‡]State Key Laboratory of Applied Optics, Chinese Academy of Sciences, China

[§]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, China

zhuyungang@jlu.edu.cn

***Abstract*—Fault prognostic in various levels of production of semiconductor chips is considered to be a great challenge. To reduce yield loss during the manufacturing process, tool abnormalities should be detected as early as possible during process monitoring. In this paper, we propose a novel fault prognostic method based on Bayesian networks. The network is designed such that it can process both discrete and continuous variables, to represent the correlations between critical deviations and quality control data. Such a network enables us to perform high-precision multi-step prognostic on the status of the fabrication process given the current state of the sensory info. Additionally, we introduce a layer-wise approach for efficient learning of the Bayesian-network parameters. We evaluate the accuracy of our prognostic model on a wafer fabrication dataset where our model performs precise next-step fault prognostic by using the control sensory data.**

I. Introduction

One of the great challenges in the semiconductor manufacturing industry is fault detection in various production steps [1]. Verification of semiconductor chips consists of pre-silicon fault diagnostic approaches [2–4], post-silicon fault detection phases [5–7] and defect prognostic during the fabrication process of the chip [8–10]. Particularly, the fabrication process is comprised of many process steps on a variety of different products [10]. To improve the efficiency and quality of the fabrication process, one major goal is to simultaneously detect critical deviations as early as possible in the production process. There is a great need in the semiconductor industry to efficiently monitor processes in real time and extract useful profile information to support process monitoring and fault detection.

To reduce yield loss during the manufacturing process, tool abnormalities should be detected early through process monitoring. Today, Statistical Process Control and Advanced Process Control methods are used to monitor and control processes and equipment to avoid critical deviations [11–14]. However, many critical deviations on product level are only found later after parameter control monitoring and wafer test measurements on the chip-level. An automatic approach for finding correlations between critical deviations in wafer-test data at the end of the process and process-control data early in the process would help to identify the critical processes together with the machine parameters. The results could be used to improve the process control setup and configurations, leading to earlier detection of deviations in the production flow.

Bayesian network is an important tool for uncertainty processing, and it is potential to decompose a complex problem into smaller and simpler ones through conditional independence, hence it provide a potential methodology for dealing with big data using divide and conquer strategy.In this paper, we propose a prognostic method for semiconductor manufacturing based on Bayesian networks. We designed a Bayesian network to represent the relationships among the status of the current process step and the fault of current step and next step. The network is designed in an hybrid fashion meaning that it process both continuous and discrete variables. Subsequently, we introduce a method to learn the Bayesian network from a wafer-fabrication process dataset. Additionally, we generilize our technique to a multi-step prognostic model by stacking up pre-trained Bayesian networks together. We eventually demonstrate the accuracy of our prognostics on the status of the fabrication process, whether it is in a normal or abnormal state, by employing our prognostic model.

The rest of the paper is organized as follows: Section II gives the definition and notations of the problem and provides a brief background on Bayesian networks. The proposed approach is described in Section III and Section IV. Section V presents the experiments for verifying the proposed method. Finally, the conclusions are summarized in Section VI.

II. Problem definition and Bayesian networks preliminaries

Let us assume that the semiconductor fabrication process consists of m steps. In each step, there are some sensors to monitor its status. There are n_t sensors at the time-step t. At each step, our aim is to use the value of the sensors to detect and predict whether there will be a fault in future steps or not.

A Bayesian network is a graphical model representing the joint probability distribution of n variables $X = \{X_1, X_2, ..., X_n\}$. It includes two components: G and Θ. G

is a directed acyclic graph where each node of the graph represents a random variable. Each arc represents a dependency relationships between variables. $\Theta = \{P(X_i|\Pi_i), 1 \le i \le n\}$ represents the conditional probabilities distribution of each node given the values of its parent nodes where Π_i stands for the parent set of X_i in the network [15].

III. Model construction for single step

A. Structure design of the model

For step t, we construct a Bayesian network as shown in Figure 1: This Bayesian network consists of n_t continuous

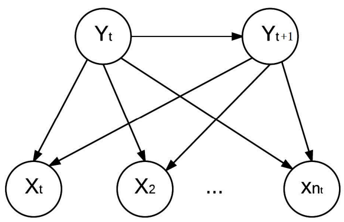

Fig. 1. Proposed Bayesian network for prognostic.

variables and 2 discrete variables. The continuous variables are placed in lower levels. $X_1, ..., X_{n_t}$ representing the n_t sensors which monitor the status at step t. The value of these variables are the sensor outputs. The discrete variables on the top level, Y_t and Y_{t+1}, predict whether there is a fault in the current step and next step, respectively. Y_t and Y_{t+1} are binary variables where a value of one implies that the status is faultless, while a zero suggests a faulty status. Note that the described network is not a general Bayesian network, it realizes a modified network which contains both continuous and discrete variables.

The structure of the Bayesian network is designed as follows: A fault varies the value of the sensors, so there are arcs from Y_t and Y_{t+1} to each variable of $X_1, ..., X_{n_t}$. A defect in the current step may result in the occurrence of a fault in the next step. Therefore, there exist an arc from Y_t to Y_{t+1}. One can use $P(Y_t|X_1, ..., X_{n_t})$ to detect whether there is a fault in the step t, and employ $P(Y_{t+1}|X_1, ..., X_{n_t})$ to predict whether there the next step is faulty or not.

B. Parameter learning of the proposed Bayesian network

Since the structure of the proposed Bayesian network is fixed, we only need to obtain the parameters before utilizing the model. We learn the parameters by using the data stored by the semiconductor corporation.

The parameters of a Bayesian network are the conditional probability of each variable given the values of its parent variables. For X_1, ... , X_{n_t} , it is often assumed that $P(X_i|Y_t, Y_{t+1})$ follows a distribution. To be precise:

$$P(X_i = x|Y_t = j, Y_{t+1} = k) = \frac{1}{\sqrt{(2\pi)}\sigma_{ijk}} exp(-\frac{(x-\mu_{ijk})^2}{2\sigma_{ijk}^2}) \quad (1)$$

where $1 \le i \le n_t$, x is the value of X_i, j and k stand for the possible value of Y_t and Y_{t+1}, respectively. σ_{ijk} represents the variance of the normal distribution and μ_{ijk} appears for the mean of the distribution.

To obtain the conditional distribution, we compute the mean and variance for each variable in X_1, ..., X_{n_t} as follows:

$$\mu_{ijk} = \frac{\sum_{l=1}^{M}(x_{i,l} \cdot \delta_{jk,l})}{\sum_{l=1}^{M}\delta_{jk,l}} \quad (2)$$

$$\sigma_{ijk}^2 = \frac{\sum_{l=1}^{M}((x_{i,l} - \mu_{ijk}) \cdot \delta_{jk,l})}{\sum_{l=1}^{M}\delta_{jk,l}} \quad (3)$$

where

$$\delta_{jk,l} = \begin{cases} 1, & \text{if } Y_{t,l} = j \text{ and } y_{t+1,l} = k, \\ 0, & \text{otherwise.} \end{cases}$$

$1 \le i \le nt$, M is the number of training data, $x_{i,l}$ is the value of X_i in sample l of the training data set, and $y_{t,l}$ and $y_{t+1,l}$ are the value of Y_t and Y_{t+1} in sample l of the training data set, respectively.

For Y_t and Y_{t+1}, a conventional method for the learning of the Bayesian network parameters is performed where $P(Y_t)$ and $P(Y_{t+1}|Y_t)$ are obtained as follows:

$$P(Y_t = j) = \frac{N_j}{M} \quad (4)$$

where $N_j = \sum_{l=1}^{M}\delta_{j,l}$, and

$$\delta_{j,l} = \begin{cases} 1, & \text{if } Y_{t,l} = j, \\ 0, & \text{otherwise.} \end{cases}$$

in another words, N_j is the number of cases (samples) in which $Y_t = j$.

$$P(Y_{t+1} = k|Y_t = j) = \begin{cases} 0, \text{ if } j\text{=0 and } k\text{=1}, \\ \dfrac{N_{jk}}{N_J}, \text{ otherwise} \end{cases} \quad (5)$$

where $N_{jk} = \sum_{l=1}^{M}\delta_{jk,l}$, and

$$\delta_{jk,l} = \begin{cases} 1, & \text{if } y_{t,l} = j \text{ and } y_{t+1,l} = k, \\ 0, & \text{otherwise.} \end{cases}$$

i.e. N_{jk} is the number of samples in which $Y_t = j$ and $Y_{t+1} = k$.

C. Fault detection, prognostic and diagnosis

As mentioned before, we can use $P(Y_t|X_1, ..., X_{n_t})$ to detect whether there is fault in step t, we first derive the formula of $P(Y_t|X_1, ..., X_{n_t})$ as follows:

$$P(Y_t|X_1, ..., X_{n_t}) = \frac{P(Y_t, X_1, ..., X_{n_t})}{P(X_1, ..., X_{n_t})} \propto P(Y_t, X_1, ..., X_{n_t})$$

$$P(X_1, ..., X_{n_t}) = \sum_{Y_{t+1}} P(Y_T, Y_{t+1}, X_1, ..., X_{n_t})$$

$$= \sum_{Y_{t+1}} P(Y_t)P(Y_{t+1}|Y_t)\prod_{i=1}^{n_t} P(X-i|Y_t, Y_{t+1})$$

So we obtained,

$$P(Y_t|X_1,...,X_{n_t}) \propto \sum_{Y_{t+1}} P(Y_t)P(Y_{t+1}|Y_t)\prod_{i=1}^{n_t} P(X_i|Y_t,Y_{t+1}) \tag{6}$$

where each probability term in equation 6 is the conditional probability distribution of each variable in the Bayesian network, and can be obtained from data using the method in Section III.B. In above derivation, an important property of Bayesian network is used, that is the joint probability distribution of Bayesian network can be represented as the product of the conditional probability distributions of all the variables in Bayesian network. We can use $P(Y_{t+1}|X_1,...,X_{n_t})$ to predict whether there will be fault in next step t+1, we then derive the formula of $P(Y_{t+1}|X_1,...,X_{n_t})$ as follows:

$$\begin{aligned} P(Y_{t+1}|X_1,...,X_{n_t}) =& \frac{P(Y_{t+1},X_1,...,X_{n_t})}{P(X_1,...,X_{n_t})} \\ &\propto P(Y_{t+1},X1,...,X_{n_t}) \\ P(Y_{t+1},X_1,...,X_{n_t}) =& \sum_{Y_t} P(Y_{t+1},X_1,...,X_{n_t}) \\ = \sum_{Y_t} P(Y_t)P(Y_{t+1}|Y_t)&\prod_{i=1}^{n_t} P(X_i|Y_t,Y_{t+1}) \end{aligned}$$

So we obtained,

$$\begin{aligned} &P(Y_{t+1}|X_1,...,X_{n_t}) \propto \\ &\sum_{Y_t} P(Y_t)P(Y_{t+1}|Y_t)\prod_{i=1}^{n_t} P(X_i|Y_t,Y_{t+1}) \end{aligned} \tag{7}$$

where each probability term in equation 7 is the conditional probability distribution of each variable in the Bayesian network.

Additionally, we can specify which sensor caused the defect by calculating the divergence of the sensor value form the normal mean value:

$$\arg\max_{i=1,...,n_t} |x_i - \mu_{i11}| \tag{8}$$

IV. Discussion about model construction for all steps

In this section, we will discuss a idea for predicting a possible fault in further step. The semiconductor fabrication process comprises m steps. Accordingly, we combine m Bayesian networks of each step and create a relatively large Bayesian network as shown in Figure 2.

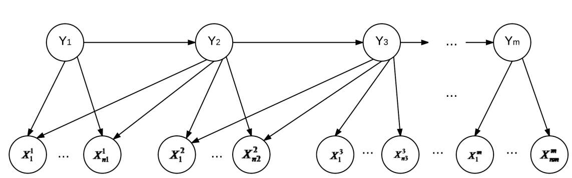

Fig. 2. Proposed Bayesian network for all steps.

To obtain such Bayesian network, we learn the small Bayesian network of each step separately, and then stack up the networks together. Such an approach is called a divide-and-conquer approach and It realizes a cost-efficient procedure since if we learn an entire Bayesian network that represents the relationships among all the sensors and faults variables of all the steps from the whole data, the computational cost will be tremendously go high.

Using the proposed stacked Bayesian network, we can make further prognostic. For instance, at step t, we can predict the possibility of having a defect at $t+2$, $t+3$ and so on. For example, we can predict whether there will be a fault in step 5 given the status values of step 1 by $P(Y_5|X_1^1,...,X_{n_1}^1)$.

We can generalize the prognostic to the kth step $t+k(k>0)$ given the status of step t by the probability of $P(Y_{t+k}|X_1^t,...,X_{n_t}^t)$. This probability can be calculated through Bayesian inference, however, it is not time efficient. Therefore, we present an approximation method as follows:

$$\begin{aligned} &P(Y_{t+k}|X_1^t,...,X_{n_t}^t) \propto \\ &\sum_{Y_tY_{t+1}...Y_{t+k-1}} P(Y_t)\prod_{i=t}^{t+k-1} P(Y_{i+1}|Y_i)\prod_{j=1}^{n_t} P(X_j|Y_t,Y_{t+1}) \end{aligned} \tag{9}$$

Assuming the final product is unqualified, one can also diagnose in which step the fault is occurred, by computing $P(Y_t|Y_m=0)(1 \leq t \leq m)$.

V. Experimental results

We assess our prognostic model on a wafer fabrication dataset [16]. In the dataset, six parameters have been identified by domain experts, the parameters are crucial for monitoring the state of the manufacturing process. The parameters include: radio frequency forward power, radio frequency reflected power, chamber pressure, 405 nanometer (nm) emission, 520 nanometer (nm) emission, and direct current bias. Radio frequency forward power and radio frequency reflected power are used to measured the electrical power applied to the plasma, and the parameter chamber pressure calculates the pressure in the etch chamber, the 405 and 520 nanometer emission measure the intensity of two different wavelengths (i.e., colors) of light emitted by the plasma, and the direct current bias parameter calculates the direct current electrical potential difference in the tool.

Based on prior knowledge on the parameters of the dataset, we reshape and reconstruct the data such that we can apply our Bayesian net. Therefore, the data is a simulated version of the dataset. In the dataset, we divide the process of semiconductor fabrication into six steps; in each step, there are six parameters (sensors) monitoring the state of the manufacturing process for quality control. The dataset contains the value of six parameters and whether there is a fault in the current step and next step. To our knowledge, most existing methods are to detect a fault of current step in real time, few methods can predict for further step, so in this paper we only test the performance of the proposed method. We use such data to train our Bayesian network while employing ten-fold cross validation in order to validate the accuracy of the proposed

TABLE I
PROGNOSTIC ACCURACY USING 10-FOLD CROSS VALIDATION

Predict Accuracy(%) / fold	step 1	step 2	step 3	step 4	step 5
1	94	54	100	100	100
2	100	100	100	97	56
3	35	92	64	97	93
4	100	100	79	100	100
5	100	100	79	95	100
6	100	85	36	61	100
7	100	100	93	97	100
8	100	92	86	97	100
9	100	92	86	100	100
10	100	92	100	82	70
average	92.9	90.8	82.1	92.6	91.9

method. Consequently, the dataset of each step is divided into ten groups, at each time we determine nine sets to be the training set and one set to be the test set. The status (values of the six parameters) in each step are used to predict the occurrence of a defect in the next step. Table I illustrates the prognostic accuracy at each time-step together with their average a precision at each step.

In Table I, the columns denote the process step. For instance, values in step1 column show the prognostic accuracy where given the sensory values in step1, the Bayesian net provides a prognostic on occurrence of fault in step2. The rows represent each test of the 10-fold cross validation together with the average accuracy for each step. Such results suggest a qualified precision is achieved by means of our stacked Bayesian network.

VI. CONCLUSIONS

In order to detect a fault as early as possible of the production process in semiconductor manufacturing industry efficiently, we proposed a Bayesian network based solution. We designed a new Bayesian network structure which contains both discrete and continuous variables representing the relationships among the status of current process and the probability of being faulty in the current step and/or next steps. We then trained the network in order to get the optimal parameters from the data. The experimental results illustrate the effectiveness of the proposed method. Our prognostic model is accurate and well-functional whenever adequate training data containing the value of sensors together with information on whether the state is normal or abnormal are provided.

For the future work, we aim to assess our approach on more experiments examples to provide further validation and efficiency, conduct more experiments on the proposed diagnosis method and further-step prognostic method, and integrate the comparisons with other methods proposed in previous works.

ACKNOWLEDGMENT

This work is funded by the Doctoral College - Cyber-Physical Production Systems (CPPS) Project, Vienna University of Technology, Austria, the NSFC (61502198), Science and Technology Development Program of Jilin Province (20150520066JH), China. Authors would like to thank Mr. Hamidreza Mahyar for his constructive suggestions on the preparation of this work. Prof. Radu Grosu and Dr. Yungang Zhu are the corresponding authors for further discussion.

REFERENCES

[1] D. Stanisavljevic and M. Spitzer, "A review of machine learning in manufacturing," in *In Proceedings of International Conference on Knowledge Technologies and Data-driven Business, Graz, Austria*, 2016.

[2] W. M. Lowe, "Transaction based windowing methodology for pre-silicon verification," Jun. 6 2000, uS Patent 6,073,194.

[3] A. Adir, S. Copty, S. Landa, A. Nahir, G. Shurek, A. Ziv, C. Meissner, and J. Schumann, "A unified methodology for pre-silicon verification and post-silicon validation," in *2011 Design, Automation & Test in Europe*. IEEE, 2011, pp. 1–6.

[4] R. M. Hasani, D. Haerle, and R. Grosu, "Efficient modeling of complex analog integrated circuits using neural networks," in *Ph. D. Research in Microelectronics and Electronics (PRIME), 2016 12th Conference on*. IEEE, 2016, pp. 1–4.

[5] K.-h. Chang, I. L. Markov, and V. Bertacco, "Automating post-silicon debugging and repair," in *2007 IEEE/ACM International Conference on Computer-Aided Design*. IEEE, 2007, pp. 91–98.

[6] S. Mitra, S. A. Seshia, and N. Nicolici, "Post-silicon validation opportunities, challenges and recent advances," in *Proceedings of the 47th Design Automation Conference*. ACM, 2010, pp. 12–17.

[7] D. Leey, T. Kolanz, A. Morgenshteinz, V. Sokhinz, R. Moradz, A. Ziv, and V. Bertaccoy, "Probabilistic bug-masking analysis for post-silicon tests in microprocessor verification," in *Proceedings of the 53rd Annual Design Automation Conference*. ACM, 2016, p. 24.

[8] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of process control*, vol. 7, no. 6, pp. 403–424, 1997.

[9] Q. Wang, G. Barnett, R. M. Greig, and Y. Cheng, "System and method for performing real time data acquisition, process modeling and fault detection of wafer fabrication processes," 1999, uS Patent 5,859,964.

[10] S.-K. S. Fan, S.-C. Lin, and P.-F. Tsai, "Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, adaboost and decision tree," *Journal of Industrial and Production Engineering*, vol. 33, no. 3, pp. 151–168, 2016.

[11] C.-F. Chien, C.-Y. Hsu, and P.-N. Chen, "Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence," *Flexible Services and Manufacturing Journal*, vol. 25, no. 3, pp. 367–388, 2013.

[12] S. Munirathinam and B. Ramadoss, "Predictive models for equipment fault detection in the semiconductor

manufacturing process," *International Journal of Engineering and Technology*, vol. 8, no. 4, p. 273, 2016.

[13] A. M. Ison, W. Li, and C. J. Spanos, "Fault diagnosis of plasma etch equipment," in *Semiconductor Manufacturing Conference Proceedings, 1997 IEEE International Symposium on*. IEEE, 1997, pp. B49–B52.

[14] A. Ison and C. J. Spanos, "Robust fault detection and fault classification of semiconductor manufacturing equipment," in *Proceedings of the Fifth International Symposium on Semiconductor Manufacturing*. Citeseer, 1996, pp. 1–4.

[15] Y. Zhu, D. Liu, H. Jia, and D. Trinugroho, "Incremental learning of bayesian networks based on chaotic dual-population evolution strategies and its application to nanoelectronics," *Journal of Nanoelectronics and Optoelectronics*, vol. 7, no. 2, pp. 113–118, 2012.

[16] R. T. Olszewski, "Generalized feature extraction for structural pattern recognition in time-series data," DTIC Document, Tech. Rep., 2001.