New concept to compute confidence of reported information level for logic diagnosis

Quoc Bao Duong, Éric Zamaï, Khoi Quoc Tran Dinh

9th *International Conference of Modeling, Optimization and Simulation - MOSIM'12*
June 6-8, 2012 - Bordeaux - France
"Performance, interoperability and safety for sustainable development"

NEW CONCEPT TO COMPUTE CONFIDENCE OF REPORTED INFORMATION LEVEL FOR LOGIC DIAGNOSIS

Quoc Bao DUONG; Eric ZAMAÏ

G-SCOP / Grenoble - INP
46 avenue Félix Viallet,
38031 Grenoble Cedex, France
Quoc-Bao.Duong@grenoble-inp.fr, Eric.Zamai@grenoble-inp.fr

Khoi Quoc TRAN DINH

Danang University of Technology
54 Nguyen Luong Bang Street,
Danang city, Vietnam
tdkquoc@gmail.com

ABSTRACT: *This paper proposes a model to compute confidence of reported information level (CRIL) in the domain of logic diagnosis. This level of confidence is provided by a diagnosis module allowing to quickly identify the origin of equipment failure. We studied the factors affecting CRIL, such as measurement system reliability, production context, position of sensors in the acquisition chains, type of product, reference metrology, preventive maintenance and corrective maintenance based on historical data and reported information generated by production equipment. We have introduced a new 'CRIL' concept based on the Bayesian Network approach, Naïve Bayes model and Tree Augmented Naïve Bayes model. Our contribution includes an on-line confidence computation module for production equipment data, and an algorithm to compute CRIL. We suggest it be applied to the semiconductor manufacturing industry.*

KEYWORDS: *Diagnosis, Confidence, Bayesian networks, Naïve Bayes, Semiconductor Manufacturing, Discrete-event system.*

1 INTRODUCTION

Nowadays, Semiconductor Manufacturing operates in an intense competitive environment. Companies working in this industry are striving to improve process quality while also improving production equipment effectiveness. However, virtual metrology, dynamic control plan, and maintenance, etc are remain challenging areas. In this paper, we introduce the concept of CRIL to improve the method for an on-line diagnosis to quickly detect the origin of equipment failures. The CRIL is computed from the production equipment data with Bayesian Network. An automated tool is also developed and proposed to be used in the semiconductor manufacturing industry.

This paper is divided in 6 sections. The semiconductor manufacturing process and CRIL concept are presented in section- 2 and section- 3, respectively. Bayesian network approach is discussed in section- 4 followed by the algorithm and CRIL computation model in section- 5. Section- 6 includes conclusion and the future works.

2 SEMICONDUCTOR MANUFACTURING SYSTEM

Semiconductor manufacturing is a complex process, based on a variety of equipment (as shown in Figure 1). They include production and metrology equipment, that continuosly demonstrate a natural drift. If this drift becomes larger than the threshold value, it might result in propagation of significant failures, immediately affecting the production process and leading to a large number of products in the manufacturing process being scrapped. Therefore, it is critical to precisely and quickly locate the causes of failures for repair and maintenance purposes.

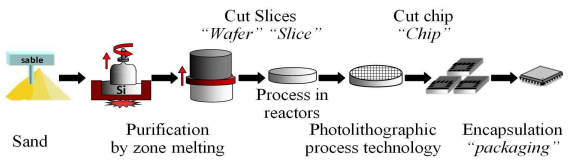

Figure 1: Semiconductor manufacturing process

Semiconductor manufacturing is an Automated Manufacturing System (AMS), structured around CIM architecture (Jones & Saleh 1989) with three main parts: controlled system, control system and product flow (Figure 2). The controlled system is a set of elementary functional chains (FCs) (Deschamps & Zamai 2007) where its operating parts are controlled by the control system based on the information collected from the controlled system. Consequently, the behavior of the control architecture is generic and is based on the principle of observability

(Cambacau 1991). It allows use of the remote procedure call (RPC) principle to launch the requests which are sent to the lowest level (customer request) i.e. level 1 within the control system.

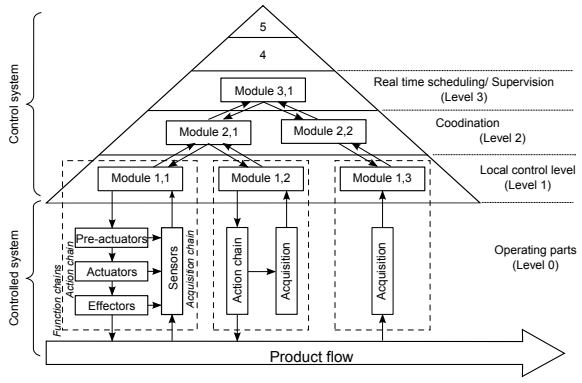

Figure 2: Diagram on an AMS

In reality, an AMS consists of hardware, software, organizational and human elements. It is subjected to uncertainties due to its operating parts (failures of sensors, actuators, etc.) and customer requests (variations in production and product specifications). In order to guarantee reactivity of the AMS, the reactive loop in the control system and dynamic reconfiguration are proposed in papers (Michel. & Courvoisier 1990) and (Henry, Zamai & Jacomino 2012). The reactive loop is characterized by collaboration of several supervision, monitoring and control (SM&C) functions such as detection, diagnosis, prognosis, decision, and automatic control (Zamai, Chaillet-Subias & Combacau 1998). Consequently, depending on the operating mode (normal or abnormal running), the purpose of the coordination level is to manage a set of FCs by using services offered by these FCs (Figure 2). In case of propagated failure in the product, detected by metrology equipment, the coordination level has to locate the origin of failure in the production equipment used in the failure.

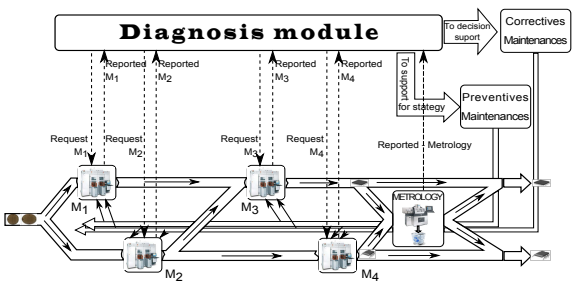

Figure 3: Semiconductor production system

Figure 3 illustrates a classical production system of a semiconductor with four production (M1, M2, M3, M4) and one metrology equipment item. A final product in such a system often requires more than 700 processing steps (Byungwhan & May 1995). Product quality is ensured by the inspections carried out in a large number of steps during the production process with the metrology equipment (inside M1, M2, M3, M4). Each control step returns a report on process execution status to the coordination control module. Nevertheless, this metrology equipment in manufacturing systems implies a lack of confidence in reported information derived from production equipment, essentially due to location and quantity of sensors in production equipment.

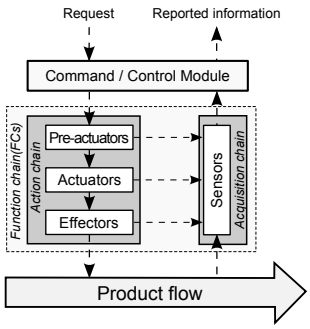

Figure 4: Internal Equipment Structure

In Figure 4, we present the problem associated with the open loop that may introduce doubts (uncertainties) as to the success of the operated service, thus implying an increased risk of (Not OK) product parts not being observed. This will result in the default response (OK) by the control module. A failure detected by the metrology equipment guarantees that one or more process steps have failed. Hence, to locate the origin of failure within production equipment, a large number of approaches (Lafortune, Teneketzis, Sampath, Sengupta & Sinnamohideen 2001, Deschamps & Zamai 2007, Fanti & Seatzu 2008) have been proposed. Particularly, (Deschamps & Zamai 2007) proposed the on-line diagnosis function providing information on the capacities of operating parts and incorporating generic rules for fault diagnosis. In this method, the response against each executed request is inserted in the diagnosis model finding the possible origin of the failure (inconsistent operation execution) and its consequence on the other services based on the doubt propagation principle. Doubt corresponds to the information that must be qualified as suspect: this mechanism (Deschamps & Zamai 2007) is referred to propagation-before and propagation-after of a diagnosis model. This approach offers a binary evaluation of the reported confidence depending on presence or absence of the product sensors within the equipment, which is a drawback. Therefore, we propose refining this confidence to between 0 and 1 via the Bayesian Network. In this approach, we consider significant factors directly impacting the confidence value such as reliability of many sensors in measurement systems, production context, maintenance activities, human expertise and type of products and its historical metrology, etc.

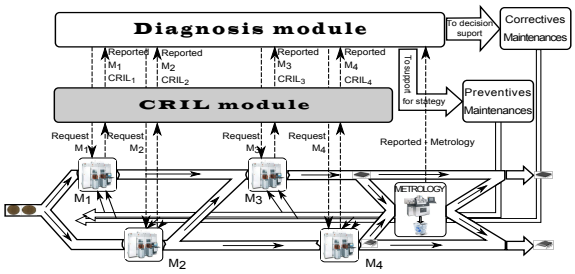

Figure 5: Module for the computation of CRIL

Accurate computation of confidence of reported information level based on collected data plays an important role in the final diagnosis. We thus suggest the on-line module for computation of the CRIL in real time (Figure 5). Furthermore, the BN approach allows probability computation based on knowledge gained (Niculescu & Mitchell 2006, Ben-Gal 2007). Next, we present the definition of the confidence of reported information level (CRIL), the characterization of CRIL and its model.

3 THE CONFIDENCE OF REPORTED INFORMATION LEVEL (CRIL)

Confidence of the reported information received from the local controlled system of Automated Manufacturing Systems (AMS) is based on previous and current data of the operating parts, followed by the computation of CRIL at coordination level.

3.1 Definition

The Confidence of Reported Information Level (CRIL) corresponds to the capacity of the functional chains that have correctly performed the requested services. It is a probability value between 0 and 1. The purpose of CRIL computation is to support diagnosis and provide relevant information about correct actions confidence of reported information.

3.2 Characterization of CRIL

To study confidence of reported information derived from equipment, we have developed a partnership with the internationally reputed semiconductor manufacturer within the European IMPROVE project. This project aims at improving European semiconductor fabrication efficiency by providing better methods and tools to control process variability, thus reducing cycle time and enhancing equipment effectiveness. Based on the information provided by industrial partners (STMicroelectronics, LPFundary, INTEL), we mainly focus on analyzing the factors which directly affect the CRIL based on Fault Detection Classification (FDC), Failure Modes and Effects Analysis (FMEA), and Statistical Process Control (SPC), etc. An analysis of equipment life from FDC data (Table 1) is given as an illustration, and represents monitoring of evolution of equipment parameters. For confidentiality reasons the table is voluntarily limited and some information is hidden. For a given item of equipment, table 1 shows the start and end time of the event, the type of product, the type of maintenance processing, etc.

EQ	Event	Start Time	End Date	Maint state name	Maint previous state	...
EQ01	FAILURE	9:18:12	10:15:03	REPAIR	FAIL	...
EQ01	MAINT	10:58:16	11:21:38	PM	IDLE	...
EQ01	PROD	18:27:19	19:22:46	NONE	NONE	...
...	...	...	...	...	...	...

Table 1: Equipment life from FDC data

Initial data analysis and brainstorming sessions held with engineers have resulted in the following seven main parameters with high impact on the CRIL:

- Measurement system reliability (R).

- Production context (C).

- Position of sensor (P) in the acquisition chain.

- Type of product (TP).

- Reference metrology for each type of product (Me).

- Preventive maintenance activities (PM).

- Corrective maintenance activities (CM).

These parameters are clearly different from the behavior issues as the whole production process is fully automated in most of the semiconductor production facilities. The production context and the product mix are well recorded, and preventive maintenance is clearly characterized. However, sensor data from the production and metrology equipment have an inherent temporal value that must be utilized to improve confidence of reported information level. The reliability of a measurement system (Maquin, Huynh, Luong & Ragot 1994) is highly temporal and varies according to usage and operating conditions. It is thus difficult to determine the relationship between measurement system reliability and real time reported information. In this paper, we propose CRIL as an estimation of the probability distribution function of a metrology system with respect to reported information. It is highly impacted by corrective maintenance as this type of maintenance cannot be scheduled. However, we present a methodology to improve and compute a real time CRIL. To accurately model the relationship between the above mentioned parameters affecting the CRIL, we have relied on the ex-

pert's knowledge and on probabilistic methods e.g. Expectation Maximization (EM) (Dempster, Laird & Rubin 1977), Markov chain Monte Carlo (MCMC) (Gilks, Richardson & Spiegelhalter 1995), Neural Network (Nauck, Klawonn & Kruse 1997). Some important probabilistic analysis methods are mentioned (Bouaziz, Zamai, Duvivier & Hubac 2011) in the European IMPROVE project. We have adopted a Bayesian approach to model the CRIL. A real time CRIL computation module, presented in this paper, combines knowledge representation in a graphical form (direct dependence relationships: cause → effect → failure) and probabilistic knowledge uncertainty (Populaire 2000). This method is used to model a directed acyclic graph (DAG) as causal dependencies of information even if they are imperfect or missing. It is thus ideally adapted to the context of our study, as learning and inference of this real time CRIL computation module are powerful features enabling merging of incomplete data with assistance of an expert. We assume that the effects of these elements on CRIL are dependent and discrete.

4 BAYESIAN NETWORK

BNs are a family of probabilistic graphical models providing joint distribution for a set of random variables (Ben-Gal 2007). Known as a DAG, they are used to represent uncertain knowledge in artificial intelligence (Korb & Nicholson 2004). The structure of a DAG combines sets of nodes and arcs where nodes represent a set of random variables from a domain. A set of directed arcs (or links) connects pairs of nodes, representing direct dependencies between variables, where variables are defined over several states. Assuming these are discrete variables, the strength of the relationship between variables is estimated by conditional probability distributions associated with each node. BNs are applied in cases of uncertainty, when we know certain conditional probabilities and seek unknown probabilities for given specific conditions. To achieve this goal, one of the BN models is widely used as a Naïve Bayes model (Lowd & Domingos 2005). This model is based on the simplest assumption that variables are conditionally independent in a given node (class): the Naïve Bayes model is presented by a single common parent node to all the variable nodes. It has certain advantages such as an intuitive technique that does not require a large amount of data before learning can begin, and fast computation, etc (Ben-Gal 2007). Therefore, the Naïve Bayes model provides reasonably good results in some practical problems and is particularly suitable for our analysis and hypotheses in this paper for real time CRIL computation. We will describe Naïve Bayes in more detail in the next section.

4.1 Naïve Bayes models for probability

The naïve Bayes method is a widely used classification and clustering method based on the Bayesian theory. It is a special form of Bayesian network relying on an important simplifying assumption of independence. It has a single node (class) that directly influences other variables, and other variables are independent for a given class.

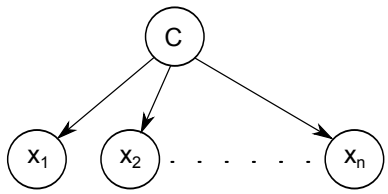

Figure 6: The structure of Naïve Bayes classifier

Figure 6 shows a Naïve Bayesian classifier for the class variable C. Using Bayes's theorem, we have:

$$P(C|\mathrm{x_i}) = \frac{P(\mathrm{x_i}|\mathrm{C})P(C)}{P(\mathrm{x_i})} \tag{1}$$

Where: C : hypothesis; $P(C)$: prior probability (probability of hypothesis C before seeing any data); $P(\mathrm{x_i}|\mathrm{C})$: conditional probability (likelihood probability); $P(\mathrm{x_i})$: probability of occurrence of record $\mathrm{x_i}$; $P(C|\mathrm{x_i})$: posterior probability estimating the probability of C given $\mathrm{x_i}$.

Let us have a set of classes $C = \mathrm{c_1, c_2, .., c_m}$ representing the observed training set. All variables (training set) $\mathrm{X} = \mathrm{x_1, x_2, \ldots, x_n}$ are assumed to be mutually independent given C. If variable C is observed in the training data, Naïve Bayes can be used by assigning training set $(\mathrm{x_1, x_2}, \ldots, \mathrm{x_n})$ to compute maximum $\mathrm{P}\,(\mathrm{C}|\mathrm{x_i})$. If C is unobserved then the data can be clustered using the Expectation Maximization (EM) algorithm (Dempster et al. 1977) alternating between computing expectations for unobserved values using current parameters and the maximum likelihood. Since the training set has n independent features, we estimate by the conjunction of all conditional probabilities of the features as shown in (2).

$$P(X|\mathrm{C}) = \prod_{i=1}^{n} \mathrm{P}(\mathrm{x_i}|\mathrm{C}) \tag{2}$$

The Naïve Bayes model is a simple and efficient approach for classifying new training set instances. Naïve Bayes is very efficient in computation models for the probability of classifying new training sets as its structure is easily constructed by an expert. Besides, it outperforms analysis of a set of sophisticated classifiers over a large set of data, notably where features are not strongly correlated (Lowd & Domingos 2005). Unfortunately, their features are

not always independent and we can find a correlation. Tree Augmented Naïve Bayes (TAN) was introduced by (Friedman, Geiger & Goldszmidt 1997) as a natural extension to the Naïve Bayesian classifier.

4.2 Tree Augmented Naïve Bayes (TAN) models

Just like Naïve Bayes models, TAN models are a special family of Bayesian networks that allow computation with correlated features. A specific TAN model presents dependence features. In this case, we show 7 features in the model as they are suitable for our study in the following chapters. Figure 7 comprises nodes $\mathrm{C}, \mathrm{x}_1, \mathrm{x}_2, ..\mathrm{x}_7$ and arcs from C to all nodes. These nodes are dependent on features $\mathrm{x}_1, \mathrm{x}_2, ..\mathrm{x}_7$ and we can compute $\mathrm{P}\,(\mathrm{C}|\mathrm{x}_1, \mathrm{x}_2, ..\mathrm{x}_7)$ for each feature as the evidence node.

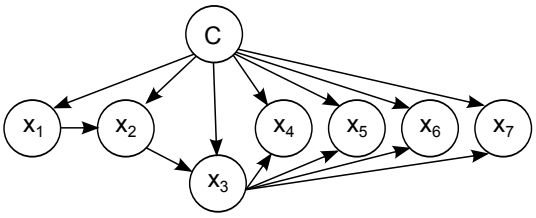

Figure 7: The structure of specific TAN

Based on the approach presented by (Friedman et al. 1997) and the transformation for coherence with the model in Figure 7, we get the posterior probability as follows:

$$\begin{aligned} &P(C\,|x_1, x_2, x_3, x_4, x_5, x_6, x_7) = \\ &= \frac{P(x_1,x_2,x_3,x_4,x_5,x_6,x_7|C\,).P(C)}{\sum_j^m P(x_1,x_2,x_3,x_4,x_5,x_6,x_7,C=c_j)}; \\ &C = \{c_1, c_j, ...c_m\} \end{aligned} \tag{3}$$

Where:

$$\begin{aligned} &P(x_1, x_2, x_3, x_4, x_5, x_6, x_7\,|C\,) = \\ &= P(x_1\,|C\,) \times P(x_2, x_3, x_4, x_5, x_6, x_7\,|C\ , x_1) \\ &= P(x_1\,|C\,) \times P(x_2\,|C\ , x_1) \times P(x_3\,|C\ , x_2)\times \\ &\times P(x_4\,|C\ , x_3) \times P(x_5\,|C\ , x_3)\times \\ &\times P(x_6\,|C\ , x_3) \times P(x_7\,|C\ , x_3) \end{aligned} \tag{4}$$

Equation 4 is derived from the Bayes theorem and the structured model in Figure 7.

The basic TAN yields good model probability that could address the factors in a reasonable fashion, as well as weak independence assumption in the Naïve Bayesian approach.

Through the analysis of the probability computation method described in this section, we build a CRIL computation model in section 5 below.

5 CRIL COMPUTATION MODEL

5.1 CRIL model

We propose the CRIL computation model based on our analysis in section 3 & 4 using a heritage approach. We use the NBC model (Figure 6) as a basic model to build a computation model. Then, to remove dependence assumptions of the Naïve model, we consider application of the TAN model (Figure 7) and build a better computation model. In this section, we present the methods for building a CRIL computation model based on the TAN approach.

The computation model is built based on Tree Augmented Naïve Bayes models presented in section 4.2. CRIL is the posterior probability that estimates the probability of CRIL with impact factors such as R, P, C, TP, PM, CM and Me (see 3.2). According to the arguments and assumptions given in the previous section, we can model each of these effects-elements as a node (variable). In particular, the *Report* (Re) variable is the parent node (C) in the structure of the TAN model for probability (Figure 7). An arrow from the generic node *Report* (Re) to node R or P, etc means that R or P is conditionally dependent on the reported information level. For each node, a conditional probability quantifies the effect of the parents on that node. One thing to note in Figure 8 is that dependence between nodes, e.g. type of product (TP) at time *t*, has a specific metrology (Me) value. The structure of the TAN model in this case is developed by experts from correlation in data and experience.

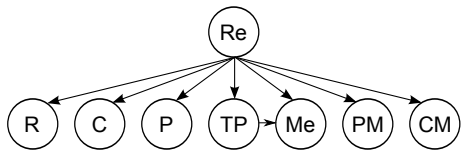

Figure 8: Modeling the CRIL by a TAN approach

Mathematically speaking, CRIL is the posterior probability computed as follows:

$$CRIL(\mathrm{Re}) = P(\mathrm{Re}|\mathrm{R}, \mathrm{P}, \mathrm{C}, \mathrm{TP}, \mathrm{Me}, \mathrm{PM}, \mathrm{CM}) \tag{5}$$

Combining (3) with (4) to compute CRIL(Re), $x_1, ..., x_7$ and corresponding to R, P, C, TP, Me, PM, CM, respectively, we have (6).

$$\begin{aligned} &P(\mathrm{Re}\,|R, C, P, TP, Me, PM, CM) = \\ &= \frac{\Omega}{\sum_j P(R,C,P,TP,Me,PM,CM,\mathrm{Re}_j)} \\ &\Omega = P(\mathrm{Re}) \times P(\mathrm{R}\,|\mathrm{Re}) \times P(C\,|\mathrm{Re})\times \\ &\times P(P\,|\mathrm{Re}) \times P(TP\,|\mathrm{Re}) \times P(Me\,|\mathrm{Re}, TP)\times \\ &\times P(PM\,|\mathrm{Re}) \times P(CM\,|\mathrm{Re}) \end{aligned} \tag{6}$$

According to equation 3 we have the expression for the probability that (Re Corresponding to C

in equation 3) will take on (Re = OK) or (Re = Not OK). In equation 6 we need to compute $P(Re), P(R|Re), P(C|Re), P(P|Re), P(TP|Re)$, $P(Me|Re, TP), P(PM|Re)$ and $P(CM|Re)$ where P(Re) is computed from the training set (data) by counting the number of occurrences of the *Reported* event, for example *(Reported = OK)* or *(Reported = Not OK)*. The probability $P(C|Re), P(P|Re), P(TP|Re), P(PM|Re), P(CM|Re)$ can be estimated by counting how often each value C, P, TP, PM, CM occurs within a class in the training set.

The computation model for the CRIL that we present in this section takes into account the observed data such as: C, P, TP, PM and CM. However, reliability of the measurement system varies over time. We thus propose a model for the measurement system and the approach EM algorithm to find the correlation between measurement system reliability and Report ($P(R|Re)$. This is presented in section 5.1.1. The probability $P(Me|Re, TP)$ in equation 6 is developed in section 5.1.2.

5.1.1 Measurement system reliability

Sensor reliability is defined as the probability r(t) of the non-failure of the sensor at time t, which represents the intrinsic quality of the sensor. It is a main factor in CRIL calculation as lower sensor reliability means lower CRIL. $r(t) = 1 - \int_0^t f(t)dt$, f(t): failure density function.

Sensor time to failure is described by the probability density function. For exponentially distributed times to failure of sensor (Dhillon 2002), sensor reliability can be written as, λ: failure per a time unit. The measurement system is made up of many sensors, and is represented by a block diagram. The probability of failure or success of one of these sensors is estimated to calculate the probability of failure or success of the overall system. In this case, the system consisting of m sensors with respective reliabilities $r_i(t)$ may define the reliability of a measurement system by $R = f(r_i(t), r_2(t), \ldots, r_m(t))$.

Depending on the functions and tasks of the measurement system, the system block diagram could be series, parallel or bridge systems (Dhillon 2005). Therefore, we need to determine the structure to calculate the associated reliability of a measurement system. For example, in an automated production as shown in figure 9, we consider the activity part of the machine M2 with sensor setting as in figure 9 to define the block diagram. This is supposed to transfer the wafer process in reactors. Our goal is to find reliability of a measurement system over time. With a constant failure rate and exponentially distributed times to failure of sensor i *(i=1,..5)*, at time t, the equation of the parallel system for dependent sensor reliability is presented as follows:

$$R(t) = 1 - (1 - e^{\lambda_1 t})(1 - e^{\lambda_2 t})^2(1 - e^{\lambda_4 t})^2 \quad (7)$$

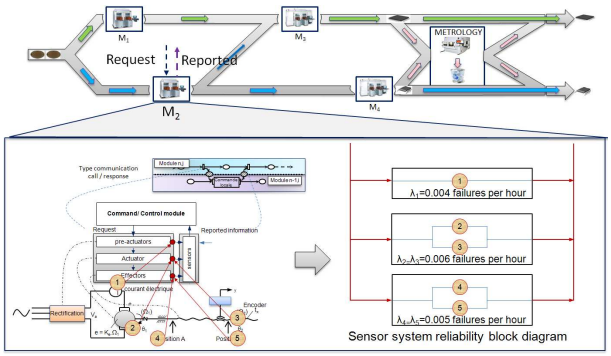

Figure 9: The structure of measurement system

How does reliability of the measurement system (R(t)) affect the CRIL? This factor is one of many factors that we need to consider. The problem is how to determine the relationship between them from the information provided by historical production data. This relationship should be standardized according to a specific function, taking change over time into account in real time. At a certain time with fixed reliability of the measurement system R(t), we can identify a *Report* event that is (OK) or (Not OK). However, at a random time in the report, it is difficult to find the probability of the *Report* event. To achieve this, relying on historical production data, we compute the probability distribution function R(t) and the probability of Report event occurrence. Now, we use the EM algorithm (D'Souza 2002) to determine the parameters of the Gaussian Mixture such as τ_i, μ_i, σ_i:

$$f(x) = \sum_{i=1}^{k} \tau_i \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (8)$$

Where: τ_i: mean and μ_i: covariance, σ_i: variables are indicator variables that are multinomial distributions, $(\sum_i^k \tau_i = 1)$.

We performed this task on MATLAB. To ensure that accuracy and computation times were not too long and complex, we chose the numbers of Gaussian as 3. We then obtained the following results.

In Figure 10, the X-axis represents measurement system reliability at time t, while the Y-axis represents probability of the $P(R|Re)$ one.

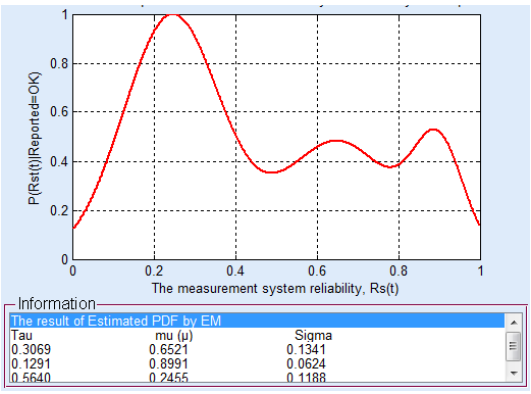

Figure 10: The relationship between measurement system reliability and reported information

5.1.2 *Metrology activities for each type of product*

As presented in the above section, the metrology machine is located at the end of a process to decide whether end products are OK or Not OK. Our goal in this part is to analyze the relationship between the metrology decision, the type of product, and the reported information (Re) to show elements affecting confidence of reported information. We assume correct report by the metrology machine, random testing in a batch of products, and existence of a fixed late time comparison with the report of manufacturing machines.

Date and time	Type of product	Reported information
21:57 06-Jan-2005	TypeA	Not OK
22:04 06-Jan-2005	TypeC	Ok
01:32 07-Jan-2005	TypeA	Not OK
02:44 07-Jan-2005	TypeB	Not OK
02:49 07-Jan-2005	TypeC	Not OK
03:27 07-Jan-2005	TypeC	Ok
08:09 07-Jan-2005	TypeB	Ok
15:09 07-Jan-2005	TypeA	Not OK
16:00 07-Jan-2005	TypeC	Ok
16:12 07-Jan-2005	TypeA	Not OK
17:50 07-Jan-2005	TypeB	Ok
18:05 07-Jan-2005	TypeB	Not OK
01:42 08-Jan-2005	TypeB	OK
02:56 08-Jan-2005	TypeC	Not OK

Table 2: Production data

For example (as shows in Table 2, 3), we consider three types of products, A, B and C, which pass through production machines M3. The historical production data for three dates are shown in Table 2. The first column shows data on the random time receiving reports from M3. The results of the metrology machine are shown in Table 3. The metrology machine randomly tests all types of products with, on average, one time per date. Column 3 in table 3 supplies two possible values *(Metrology = Pass)* or *(Re = Not Pass)*. *Pass* means that the quality of a product manufactured by the machine is good, while *Not Pass* means the opposite. Considering a particular case at 01:42:00, 08-jan-2005, product type B after the production process at machine M3 had received the report *(Re = OK)*. However, after a fixed late time at 17-jan-2005, the metrology report received was *Not Pass*. This means that there is a difference between the reports of M3 and the metrology machine resulting in an uncertainty of reported information on production machines. We thus need to analyze the impact of reported information and metrology in the CRIL model.

Date	Type of product	Metrology
08-Jan-2005	TypeA	Pass
17-Jan-2005	TypeB	Not Pass
24-Jan-2005	TypeC	Pass

Table 3: Metrology data

In Equation (6), to calculate the CRIL, we need to define $P(Me|Re, TP)$.

Considering the Bayes' theorem, we obtain as follows:

$$P(Me\,|TP, Re) = \frac{P(Me,TP,\mathrm{Re})}{P(TP,Re)} = \frac{P(Me,TP,\mathrm{Re})}{\sum_{j}^{U} P(TP,Re,Me_j)};$$
$$U = \{\mathrm{Pass}, Not\_Pass\} \tag{9}$$

Production and metrology data are required in the same survey period to compute P(Me,TP,Re) from equation (9). However, we faced difficulties as to the time reported in Tables 2 and 3 such as the difference in format and the number of reports (production data time reports on average 10 times/day, whereas metrology data reports once every day). We can solve this problem by integrating the data (Table 4) on the sampling principle. The result of metrology in the production system reflects the quality of the product in one day. In other words, the time before the new results or the result in *(t-1)* will be true at any time before it.

P(Me,TP,Re) is computed from Table 4 by counting the number of occurrences (simultaneous appearance of values) in Table 4, e.g. $\{\mathrm{TypeA}, \mathrm{OK}, \mathrm{Pass}\}$.

Date and time	Type of product	Report	Metrology
21:57 06-Jan-2005	TypeA	Not OK	Pass
22:04 06-Jan-2005	TypeC	Ok	Pass
01:32 07-Jan-2005	TypeA	Not OK	Pass
02:44 07-Jan-2005	TypeB	Not OK	Not Pass
02:49 07-Jan-2005	TypeC	Not OK	Pass
03:27 07-Jan-2005	TypeC	Ok	Pass
08:09 07-Jan-2005	TypeB	Ok	Not Pass
15:09 07-Jan-2005	TypeA	Not OK	Pass
16:00 07-Jan-2005	TypeC	Ok	Pass
16:12 07-Jan-2005	TypeA	Not OK	Pass
17:50 07-Jan-2005	TypeB	Ok	Not Pass
18:05 07-Jan-2005	TypeB	Not OK	Not Pass
01:42 08-Jan-2005	TypeB	OK	Not Pass
02:56 08-Jan-2005	TypeC	Not OK	Pass

Table 4: Mixing the data

5.2 Algorithm and Implementation

5.2.1 CRIL Computation Algorithm

In this section, we propose an algorithm to compute CRIL in real time from the structure presented in Figure 16, based on the characteristics and formulas given by the CRIL TAN Model. We started with data analysis to identify the relationships and impacts on reported information through the experience and knowledge of experts who built the Bayesian network structure. The same data are used to compute conditional probabilities from the training data set during the learning step. The CRIL is computed in real time by the computation model to report the (report + training set) information prior to sending it to the diagnosis module.

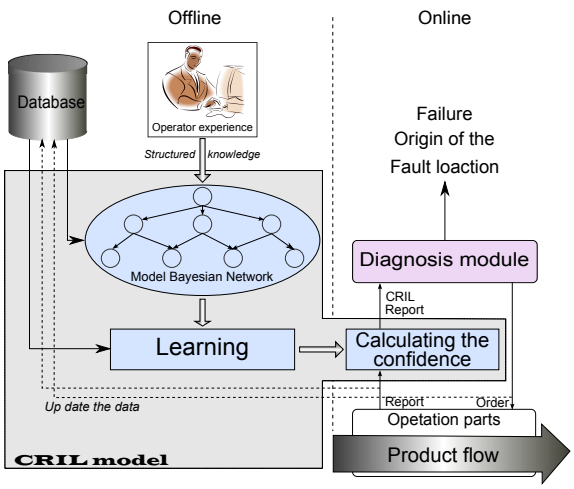

Figure 11: The structure of the CRIL Computing Algorithm

The proposed algorithm presented in figure 12 performs two main tasks: (i) learning and (ii) testing. With respect to learning, it estimates the conditional probabilities of each component by multiplying it with equations 6, whereas for testing, after receiving a new vector of parameters, it computes the conditional probabilities based on the TAN model. The final value output of the algorithm is CRIL.

Input: *Training data set and a set new vector of parameters (C, P,TP,PM,CM,Re)*

Learning

From given data:

- *Compute a prior probability P(Re)*
- *Compute a probability P(R|Re) over time*
- *Compute a probability P(C|Re);P(P|Re); P(TP|Re);P(PM|Re);P(CM|Re)*
- *Compute a probability P(Me|TP,Re) with respect to TP*
- *Compute a probability*

$$P(R,C,P,TP,Me,PM,CM|Re) = P(Re).P(R|Re).P(C|Re).P(P|Re).P(TP|Re).P(Me|Re,TP).P(PM|Re).P(CM|Re)$$

Testing

For each new vector of parameters (C, P,TP,PM,CM,Re):

- *Compute the probability*

$$P(Re|R,C,P,TP,Me,PM,CM) = \frac{P(Re).P(R|Re).P(C|Re).P(P|Re).P(TP|Re).P(Me|Re,TP).P(PM|Re).P(CM|Re).}{\sum_j P(R,C,P,TP,Me,PM,CM,Re_j)}$$

Output: *probability P(Re|R,C,P,TP,Me,PM,CM)*

Figure 12: Computing CRIL TAN algorithm

5.2.2 Implementation program on MATLAB

In an AMS as shown in Figure 5, we consider the historical data collected from production equipment M3 and the metrology data fed from the files as shown in frame (1) on Figure 13. After completion of the production process on machine M3, its control system reports process end at the coordination level. This report provides information about the parameters of M3 and its current operating status. If the received report is *(OK)* then the control system ends the process considering the processed product to be OK. However, if the report is *(Not Ok)*, the product in M3 will be considered *(Not Ok)*. In both cases, we consider the effects of metrology equipment on the production process and its associated confidence of reported information.

After implementing the program with algorithms and mathematical formulas expression, we obtain the results of the interface of the CRIL computing model as shown in Figure 13.

In Figure 13: R = $\{0 \div 100\%\}$; P = {Open-Loop (OL), Pre-Actuator (PA), Actuator (AC), Pre-Actuator + Actuator (PA.AC), Effectors (EF), Effectors + Pre-Actuator (EF.PA), Effectors + Actuator (EF.AC), Effectors + Actuator + Pre-Actuator

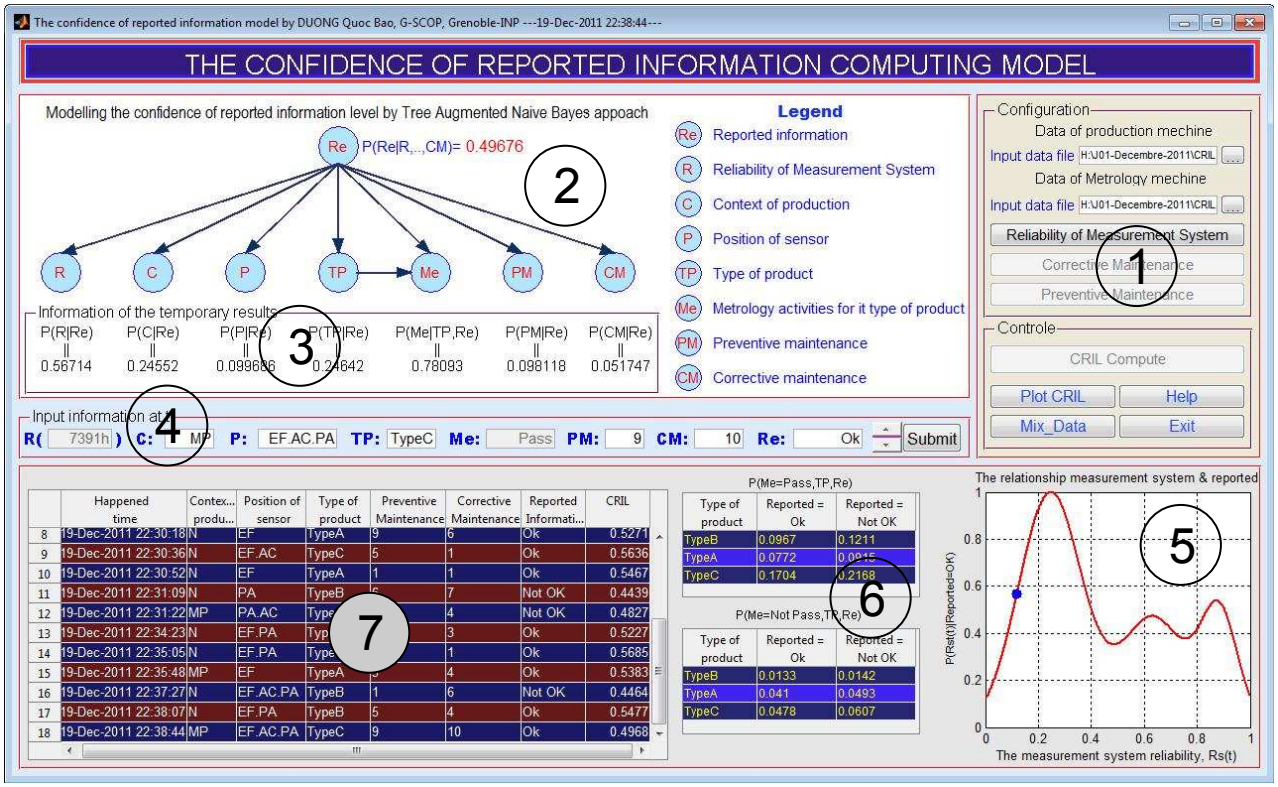

Figure 13: Interface of confidence of reported information computing model

(EF.AC.PA)}; C={ Normal production (N), Mass production (MP), Change recipes (CR) }; PM = $\{0 \div m\}$; CM = $\{0 \div k\}$.

We introduce some main frames in the interface as follows: (1). Input data (Production and Metrology data). (2). Description of the model structure. (3). The temporary result of equation 6 . (4). The reported information at time t. (5). Presentation of probability $P(R|Re)$ over time. (6). The temporary result of probability *P(Me,TP,Re)*. (7). Show all the calculated results.

In line 18, of frame 7, the CRIL of *Report* is (OK) with the current parameters of M3, e.g. R(7391h) = 0.1196. This means that reliability of the measurement system, considering its operated time as 7391 hours is 0.1196. We can infer value $P(R|Re)$ = 0.56714 from R(7391h). C = MP, P = EF.AC.PA, TypeC, PM = 9 CM = 10, and association with the backward reference metrology activity (Me = Pass) is computed as 49.68%.

In this case, the CRIL of equipment M3 that has correctly performed the requested services (Reported = OK) is 49.68% at the time of computation, with the current equipment parameters and the production context. It helps automation engineers to locate the process equipment leading to product failure detected by the Metrology. Change over time in the current parameters of M3 results in different CRILs(frame 7).

Based on process data and the expert's knowledge, we are now able to evaluate the confidence of all the reported information taken from the equipment. The resulting CRIL is a value between 0 and 1 that extends the diagnosis inference proposed by (Deschamps & Zamai 2007).

6 CONCLUSIONS

This paper proposes a concept of confidence of reported information level (CRIL) to help automation engineers locate the process equipment leading to product failure detected by the Metrology. The proposed CRIL index is a value between 0 and 1 and is computed based on the BNs approach which is widely accepted as a methodology to learn uncertainties. Furthermore, we have developed an algorithm and a computation module for real time computation of the said CRIL index. Based on the algorithms and the TAN model, a tool developed in Matlab is presented and proposed for use in the semiconductor manufacturing industry. Our work is an extension to the diagnosis approach proposed by (Deschamps & Zamai 2007). Based on the our analysis, we have highlighted seven main factors that impact the CRIL. Our proposed TAN model processes the inherent uncertainty reported information and obtain the posterior probabilisty of the reported information. It is thus able to provide the diagnosis module with the information it need to facilitate locate of the origin

of an equipment failure in a given production process. In future, we focus on validating the model on the data collected from a world reputed semiconductor manufacturing industry (partner in the European IMPROVE project). Furthermore, we are working on the extension of the proposed TAN model with the inclusion of continuous and temporal variable using Dynamic Bayesian network.

REFERENCES

References

Ben-Gal, I. (2007). Bayesian networks, *in* F. Ruggeri, R. Kenett & F. W. Faltin (eds), *Encyclopedia of statistics in quality and reliability*, John Wiley and Sons.

Bouaziz, M. F., Zamai, E., Duvivier, F. & Hubac, S. (2011). Dependability of complex semiconductor systems: Learning bayesian networks for decision support, *Dependable Control of Discrete Systems (DCDS), 2011 3rd International Workshop on*, pp. 7–12.

Byungwhan, K. & May, G. S. (1995). Real-time diagnosis of semiconductor manufacturing equipment using neural networks, *Electronics Manufacturing Technology Symposium, 1995. 'Manufacturing Technologies - Present and Future', Seventeenth IEEE/CPMT International*, pp. 224–231.

Cambacau, M. (1991). *Commande et surveillance des systèmes à événement discrets complexes: appliaction aux ateliers flexible*, PhD thesis.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1): 1–38.

Deschamps, E. & Zamai, E. (2007). Diagnostis for control system reconfiguration, *The 4th International Federation of Automatic Control Conference on Management and Control of Production and Logistics.*

Dhillon, B. (2002). Reliability, *Engineering maintenance*, CRC PRESS.

Dhillon, B. (2005). Dynamic reliability evaluation models, *in* B. Dhillon (ed.), *Reliability, Quality, and Safety for Engineers*, CRC PRESS.

D'Souza, A. A. (2002). Using em to estimate a probablity density with a mixture of gaussians.

Fanti, M. P. & Seatzu, C. (2008). Fault diagnosis and identification of discrete event systems using petri nets, *Discrete Event Systems, 2008. WODES 2008. 9th International Workshop on*, pp. 432–435.

Friedman, N., Geiger, D. & Goldszmidt, M. (1997). Bayesian network classifiers, *Mach. Learn.* **29**(2-3): 131–163.

Gilks, W. R., Richardson, S. & Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC.

Henry, S., Zamai, E. & Jacomino, M. (2012). Logic control law design for automated manufacturing systems, *Engineering Applications of Artificial Intelligence* (0).

Jones, A. & Saleh, A. (1989). A multi-layer/multi-level control architecture for computer integrated manufacturing systems, *Industrial Electronics Society, 1989. IECON '89., 15th Annual Conference of IEEE*, pp. 519–525 vol.3.

Korb, K. B. & Nicholson, A. E. (2004). *Bayesian Artificial Intelligence.*

Lafortune, S., Teneketzis, D., Sampath, M., Sengupta, R. & Sinnamohideen, K. (2001). Failure diagnosis of dynamic systems: An approach based on discrete event systems, *The American Control Conference*, Vol. 25-27, Arlington, pp. 2058–2071.

Lowd, D. & Domingos, P. (2005). Naive bayes models for probability estimation.

Maquin, D., Huynh, C. T., Luong, M. & Ragot, J. (1994). Observability, redundancy, reliability and integrated design of measurement systems, *2nd IFAC Symposium on Intelligent Components and Instruments for Control Applications, SICICA'94*, p. . SURFDIAG.

Michel., C. & Courvoisier, M. (1990). A hierarchical and modular structure for fms control and monitoring, *The 1st Conference AI simulation and planning in high autonomy systems*, Tucson, Arizona.

Nauck, D., Klawonn, F. & Kruse, R. (1997). *Foundations of Neuro-Fuzzy Systems*, Wiley.

Niculescu, R. & Mitchell, T. (2006). Bayesian network learning with parameter constraints.

Populaire, S. (2000). Introduction aux réeaux bayésiens, *Rapport interne du laboratoire SIME*, Université de Technologie de Compiègne.

Zamai, E., Chaillet-Subias, A. & Combacau, M. (1998). An architecture for control and monitoring of discrete events systems, *Computers in Industry* **36**(1-2): 95–100.