



HAL
open science

Event-based probabilistic approach for proactive maintenance to improve production capacities in HMLV industries

Asma Abu Samah

► **To cite this version:**

Asma Abu Samah. Event-based probabilistic approach for proactive maintenance to improve production capacities in HMLV industries. Automatic. Université Grenoble Alpes, 2016. English. NNT : 2016GREAT122 . tel-03332417

HAL Id: tel-03332417

<https://theses.hal.science/tel-03332417>

Submitted on 2 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Automatique - Productique (AP)**

Arrêté ministériel : x

Présentée par

Asma' ABU SAMAH

Thèse dirigée par **Eric Zamaï**
et codirigée par **Muhammad Kashif Shahzad**

préparée au sein du **Laboratoire GSCOP** et **Univ. Grenoble Alpes**
et de l'**Ecole doctorale EEATS**

Event-based probabilistic approach for proactive maintenance to improve production capacities in HMLV industries.

Thèse soutenue publiquement le **8 Novembre 2016**,
devant le jury composé de :

M. Michel TOLLENAERE

Professeur, Grenoble-INP, Président

M. François, PERES

Professeur, Ecole Nationale d'Ingénieurs de Tarbes (ENIT), Rapporteur

M. Philippe WEBER

Professeur, Université de Lorraine, Ecole Supérieure des Sciences et Technologies de L'Ingénieur de Nancy (ESSTIN), Rapporteur

M. Sha'ri MOHD YUSOF

Professeur Ir., Université de Technologie Malaisie de Kuala Lumpur, Examineur

M. Stéphane HUBAC

Manufacturing Science Senior Expert Advanced Process and Equipment Control, STMicroelectronics-Crolles, Invité

M. Eric ZAMAÏ

Maître de conférence HDR, Grenoble-INP, Directeur de thèse

M. Muhammad Kashif SHAHZAD

Docteur, Institut National des Sciences Appliquées (INSA) de Lyon, Co-Directeur de thèse



None of us got to where we are
alone.

Harvey Mackay

Acknowledgement

This dissertation is the result of the work that has been conducted during three years of research with the G-SCOP laboratory, and under the Univ. Grenoble Alpes, France. In various ways, it has involved several persons and organisations that I am most eager to acknowledge.

First and foremost, I would like to thank my supervisors. To Eric Zamaï for his trust and confidence in me which goes a long way to my master degree year, for his constructive comments, directions, and the continuous moral support. To Muhammad Kashif Shahzad for his availabilities, helpful and relevant comments, as well as fruitful discussion and corrections.

Secondly, I would like to express the gratitude to the European project INTEGRATE and region RhoneAlpes for the funding of the research. To STMicroelectronics in particular for their support and provision of data for case studies. I would also like to thank Michel Tollenaere, Stéphane Hubac and Farouk Bouaziz for their direct and indirect participation and investment in this work, for their kindness and valuable guidelines.

My sincere thanks also go to the jury members who have agreed to evaluate this dissertation. I am honoured and thankful to Francois Pérès and Philippe Weber for their interest in this work and by agreeing to be the reporters. I also thank Yusof Sha'ari, a professor in Malaysian University for his acceptance to be my viva examiner.

I express my appreciation to the staffs, my colleagues and friends at the GSCOP Laboratory who participated and contributed in one way or another. In particular, to Marie-Joséphé, Fadila, Christine, Souad, Myriam, Kevin, Jean-Yves, Sébastien, Hanna, Anis, Emna, Widad, Trinh and Hoang. Thank you for your support, for the various exchanges, convivialities and your help during these years.

Finally, I would like to convey my heartfelt appreciation to my families. To my mother Aishah and my late father, Abu Samah for their prayers, encouragement and blessing, to my parents in law, Azman and Zainoha for their support and wisdoms, and to my one and only grandmother, Zainon for The wait. But most importantly, to my husband, Abdul Hadi who is currently in his final sprint of thesis writing and our sons, Talhah and Suhayb, not only for their patience and endless support, but also for their love and

their sharing of every moment along this memorable journey. To my brothers, sisters, friends and their families whom I can't possibly mention all their names, thank you very much.

Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Winston Churchill

*Asma' Abu Samah
Grenoble, August 2016*

Contents

List of Figures	iii
List of Tables	iv
Acronyms and Notations	v
General Overview	1
I INTRODUCTION	6
1 Research Context	7
1.1 Industrial Revolution and Emerging Production Environments	8
1.2 High-Mix Low-Volume Production Characteristics	9
1.3 Research Goal and Objectives	11
1.3.1 Goal	12
1.3.2 Objectives	12
1.4 Unscheduled Breakdown Triggering Mechanisms	13
1.5 Challenges in High-Mix Low-Volume Production	14
1.5.1 Limitation in minimizing Corrective Maintenance . . .	15
1.5.2 Limitations in improving Preventive Maintenance . . .	16
1.6 Conclusion	17
2 State of the Art	18
2.1 Role of Maintenance Strategies	18
2.1.1 in reducing Corrective Maintenance (CM)	19
2.1.2 in improving Preventive Maintenance (PM)	20
2.2 Failure/Cause Diagnosis Framework to reduce Corrective Main- tenance	21
2.2.1 Source of Product Quality Drifts	22
2.2.2 Equipment Failure/Cause Diagnosis	23
2.3 Predictive Maintenance Framework to improve Preventive Main- tenance	24
2.3.1 Online Failure Prediction Approaches	25
2.3.2 Failure Signature based Approaches	27
2.4 Scientific Gaps	29
2.5 Research Questions	31
2.6 Conclusion	31

3	Positioning of Contributions and Thesis Organization	33
3.1	Contribution and Positioning	33
3.1.1	Integrated 2-Level Product Quality Drift and Failure/Cause Diagnosis Methodology	33
3.1.2	Methodology to Extract and Validate Time Bound Fail- ure Signatures for Predictive Maintenance	35
3.2	Thesis Schematic	36
3.3	Thesis Organization	38
3.3.1	Part-I: Introduction	38
3.3.2	Part-II: Choice of Modelling Approach and Proposed Methodologies	38
3.3.3	Part-III: Application and Results	39
3.4	Conclusion	40
 II CHOICE OF MODELLING APPROACH AND PROPOSED METHODOLOGIES		41
4	Modelling Approach	42
4.1	Existing Modelling Approaches	42
4.1.1	for Failure/Cause Diagnosis	43
4.1.2	for Extraction of Failure Signatures	44
4.2	Discussion on the Choice of Modelling Approach	45
4.3	Bayesian Network as a Choice for Modelling Integrated Drift Source and Failure/Cause Diagnosis and Time Bound Failure Signatures	46
4.3.1	Bayesian Principle and Bayesian Network (BN)	46
4.3.2	BN for Classification	49
4.3.3	Bayesian Network Structure Learning	49
4.3.4	Estimating Conditional Probability Tables (CPT)	53
4.4	Conclusion	54
5	Integrated Drift Source and Failure/Cause Methodology	56
5.1	Key Elements for Modelling Equipment Failure	56
5.2	The Methodology	59
5.2.1	Level-1: Product Quality Drift Source Diagnosis	60
5.2.2	Level-2: Equipment Failure and Causes Diagnosis	62
5.3	Discussion	64

6	Methodology for the Extraction and Validation of Time Bound Failure Signatures	66
6.1	The methodology	66
6.1.1	Step-1: Dataset Preparation	68
6.1.2	Step-2: Equipment Failure Probability Plots	70
6.1.3	Step-3: Extraction of Time Bound Failure Signatures (Patterns/Rules)	72
6.1.4	Step-4: Validate Rules and Compute Predictability Index	73
6.2	Discussion	74
III	APPLICATION AND RESULTS	75
7	Case Study Description	76
7.1	Introduction to Semiconductor Industry (SI)	76
7.1.1	Characteristics of Integrated Circuits (IC) Production Environment	76
7.1.2	Production Line multi-Source Data	80
7.2	Presentation of Case Study	82
7.2.1	Thermal Treatment Workshop	82
7.2.2	Reactor Module of TASMI-03	84
7.3	Discussion	84
8	Application and Results of Integrated Drift Source and Failure/Cause Methodology	85
8.1	Level-1: Product Quality Drift Source Diagnosis	86
8.1.1	Identification and Classification of Predictors	86
8.1.2	Learn and Train Semi-Supervised Bayesian Network	88
8.1.3	Test Bayesian Network for Product Quality Drift Source Diagnosis	90
8.2	Level-2: Equipment Failure/Cause Diagnosis	93
8.2.1	Identification of Failure/Causes at Module and Equipment Levels	94
8.2.2	Learn and Train Semi-Supervised Bayesian Network at Module/Equipment Levels	94
8.2.3	Test and Compare Bayesian Network Performance at Module/Equipment Levels	95
8.3	Conclusions and Discussions	98

9 Application and Results of Online Failure Prediction Methodology with Time Bound Failure Signatures	103
9.1 Step-1: Dataset Preparation	104
9.1.1 Identification and Selection of Variables	105
9.1.2 Interval Discretization	105
9.1.3 Dataset Pre-Processing	106
9.2 Step-2: Equipment Failure Probability Plots	107
9.2.1 Learn/Optimize Bayesian Network Structure and compute CPT	107
9.2.2 Plot Failure Probability from Bayesian Network for each Discretized Interval	108
9.3 Step-3: Extraction of Time Bound Failure Signatures (Rules) .	110
9.4 Step-4: Validate Rules and Compute Predictability Index . . .	113
9.5 Conclusions and Discussion	113
Conclusions and Future Works	117
A Bayesian Network Learning	121
A.1 Notation	121
A.2 Optimize Structure Learning using Equivalence Class (EQ) and Tabu List	122
A.3 Minimum Description Length (MDL) Score	126
B The INTEGRATE PROJECT	127
B.1 Introduction	127
B.1.1 Project management	127
B.1.2 STMicroelectronics Crolles	129
Bibliography	130
Résumé français	139

List of Figures

1.1	Product mix, commonality and differentiation vs. equipment utilization.	10
1.2	Production capacities optimization in view of unavailability linked to CM and PM (due to CM).	12
1.3	Example of process control in production.	13
1.4	Failure counts, durations and occurrences.	15
1.5	Example of process control in production.	17
2.1	Evolution of maintenance strategies.	21
2.2	Elements of time and their relation in online failure prediction.	28
3.1	Thesis schematic.	37
4.1	Example of BN.	47
4.2	Example of Inference Algorithms.	48
4.3	Illustration of EQ search strategy.	52
4.4	Illustration of Tabu list search strategy.	53
5.1	Equipment states classification in the SI.	57
5.2	Relationship between equipment states at parent and child level.	59
5.3	Sub-steps in the identification of product quality drift source.	60
5.4	Bayesian Network structure for product quality drift diagnosis.	61
5.5	Example of reasoning and its use in Bayesian Network.	61
5.6	Associated sub-steps for the integrated equipment failure cause diagnosis.	62
5.7	Conceptual framework of BN structure for integrated failure and cause diagnosis.	63
6.1	The methodology.	67
6.2	Summary of Step-1.	68
6.3	Validation scheme using historical data.	69
6.4	Conceptual framework for module level BN construction to plot failure probabilities.	70
6.5	Summary of Step-2.	70
6.6	Summary of Step-3.	72
6.7	Summary of Step-4.	73
7.1	Global sales revenues of SI.	77

7.2	Front-end proces flow.	79
7.3	Description of information systems of Crolles, 300mm production line at STMicroelectronics, Crolles.	81
7.4	Chart representing a hierachy of equipment composition in the SI.	82
7.5	Rules on inherited failure within different level of components.	82
7.6	General structure of the LPCVD equipment (left) and the reactor module (right) from TT workshop.	83
8.1	Proposed 3-step methodology for reducing failure/cause misdiagnosis in the SI.	86
8.2	Classification of predictors.	88
8.3	BN model for FS identification.	89
8.4	Representative nodes for the proof of concept.	90
8.5	Proof of concept of FS (Failure Source) identification.	91
8.6	Failure source standard deviation and the box plot graph of true positive prediction performances.	92
8.7	Prediction accuracy with ROC curves for all FS.	93
8.8	Gain curves for all FS.	94
8.9	Module level BNs for failure/cause diagnosis.	96
8.10	Mainframe BN for failure/cause diagnosis.	97
8.11	Proofs of concept for Reactor1 and Reactor2 BN.	98
8.12	Proofs of concept for Mainframe.	99
8.13	Equipment level BN for failure/cause diagnosis.	99
8.14	Proof of concept from equipment level BN.	100
8.15	Gain curves for module level BNs.	100
8.16	Gain and ROC curves for equipment Level BN.	101
8.17	Equipment level standard deviation and the box plot graph of true positive prediction performances, reliability (left) and precision (right).	101
8.18	Reliability and precision comparison by failures and module/equipment.	102
8.19	Gain in prediction accuracy for module level BNs over equipment level.	102
9.1	Summary of the 4-steps failure signature extraction methodology with respective colours for elaboration.	104
9.2	Step-1: Identification of Predictors as failure predictors and dataset pre-processing.	104
9.3	Data interval discretization rules.	106
9.4	Step-2: Unsupervised learning and optimization of Bayesian Network.	107

9.5	BN model for failure probabilities plot.	109
9.6	Failure inferences in BNL&RE testing dataset, data interval [2835,4725].	109
9.7	Failure inferences in BNL&RE testing dataset, data interval [3480,3490].	110
9.8	Step-3&4: Extraction of rules from each failure probability plot.	111
9.9	Proof of concept for rule(s) extraction on failure.	112
9.10	Predictability Index for selected rules.	114
9.11	Proof of concept for identification of potential causes linked to <i>failure_a</i>	115
A.1	Flow hcart for EQ+Tabu list search strategy.	122
B.1	The workpackages.	128

List of Tables

2.1	Classification of RUL methods and tools	26
5.1	Occurrence confusion matrix example.	63
5.2	Precision matrix example.	64
5.3	Reliability matrix example.	64
6.1	Preparation of data for BN training.	69
8.1	Precision and reliability matrices of FS BN.	92
8.2	List of failures and failure modes's states and causes nodes. . .	95
9.1	List of identified Predictors and failures for TASMI03_01. . . .	105
9.2	Summary of results from rule(s) extraction.	113
9.3	Example of results and information for rule(s) testing.	114
A.1	Notation	121

Acronyms and Notations

Acronyms

APC	Advanced Process Control
BN	Bayesian Network
CBM	Condition Based Maintenance
CM	Corrective Maintenance
CPT	Conditional probability Table
EQ	Equivalence Class
FDC	Fault Detection and Classification
FMECA	Failure Mode, Effects and Criticality Analysis
HMLV	High-Mix Low-Volume
LPCVD	Low Pressure Chemical Vapor Deposition
MDL	Minimum Description Length
ML	Maximum Likelihood
PdM	Predictive Maintenance
PM	Preventive Maintenance
RUL	Remaining Useful Life
SI	Semiconductor Industry
SPC	Statistical Process Control

Notations

Δtl	Lead time
Δtw	Minimal Warning Time
Δtp	Prediction Period
Δtd	Data Window
BNL	Bayesian Network Learning
FM	Failure Mode
FS	Failure Source
Q	Quarter
RE	Rules Extraction
V	Validation

General Overview

The post industrial revolution era has resulted in the emergence of high-mix low-volume production environment which is challenged by increasing demand volume and diversity. Consequently, product life cycles has shortened. In this fragile, complex and highly competitive production setting, market share can only be increased and sustained by developing new technologies and improved utilization of existing production capacities. However, frequent changes in process recipes has led the drifting equipment behaviours and increasing unscheduled equipment breakdowns which disrupts and reduces the production capacities. There exist multiple solutions to this problem. For example, we can invest to buy additional production equipment and neutralize the challenges of high-mix low-volume. This will be an artificial transformation to low-mix low-volume but highly capital intensive. Moreover, sustainable market share requires new technologies in short but regular intervals; therefore, we have a risk of technological incompatibility with these newly employed production equipment. These challenges can also be addressed by incorporating automation system and computerized maintenance management. However, the hostile equipment behaviour requires intelligent than programmable automation systems to improve production capacities.

The increasing unscheduled equipment breakdown results in increasing corrective and preventive maintenance. This not only unstabilizes but also reduces the production capacities. As a matter of fact, shortened product life cycles also result in shortened technology life cycle; therefore, increased and sustainable market share cannot be ensure by incorporating only new technology. The core problem lies in unstable and reduced production capacities. In this research thesis, we are objectively focused on stabilizing and improving production capacities in high-mix low-volume production settings. We argue that this can be achieved by improving increased corrective (CM) and preventive maintenance (PM). This argument served as the basis for initial investigation using data collected from the production line of a world reputed high-mix low-volume production line. The analysis (see chapter 1, section 1.2, subsections ??, 1.5.2) highlighted that the reason for increasing CM is the misdiagnosis in identifying the source of product quality drifts and failure(s)/cause(s). This results in increasing failure durations, failure counts and repair actions. This has direct impact on increasing PM. It is because of the fact that during the execution of CM actions, misdiagnosis often result in the execution of PM actions by technicians, in anticipation to put machine

back in production as early as possible.

In high-mix low-volume production settings, most commonly employed maintenance practices are CM, systematic PM and Conditional Based Maintenance (CBM). The CM refers to run to failure maintenance strategy whereas systematic PM refers to time or other elapsed unit based maintenance with an objective to keep the equipment in functional condition. These maintenance strategies could easily result in under or over engineering and lead to the emergence of CBM where maintenance actions are triggered based on the condition monitoring information available from the past until decision time. The CBM is implemented with FDC (fault detection and classification system) and is used to control equipment stoppage in case of abnormal behaviour detected through continuous monitoring of signals from equipment sensors. It is accomplished in two steps, first the detection of the fault, and second the identification of respective causes for the fault. Moreover, success of CBM based on FDC depends on repeatable failures or equipment drifts. It can be concluded that this maintenance strategy is not very useful in high-mix low-volume because of changing equipment behaviours. In comparison to these maintenance strategies, PdM (predictive maintenance) has emerged as a promising solution where efforts are made to predict failures before their occurrences. There is a high degree of uncertainty associated with the PdM approaches. In this research thesis, we are subjectively focused on addressing misdiagnosis in product quality drift source and failure(s)/cause(s) which result in increasing CM and PM. Consequently, production capacities becomes unpredictable and reduced. We argue that an intelligent system, besides automation and computerized maintenance management, is required to help addressing this misdiagnosis issue to reduce CM and PM initiatives.

In literature (see chapter 2), we found that equipment is believed to be the only source of product quality drifts. However, in high-mix low-volume production settings, product quality drift source can vary from equipment, product, process (technology) to maintenance. Therefore, we argue that in the absence of multi source diagnosis model for product quality drifts, it is not possible to reduce increasing unscheduled equipment breakdowns and respective CM and PM actions. We also found that most of the failure(s)/causes(s) diagnosis approaches proposed are based on equipment sensor data. Moreover, these models are developed and benchmarked at the equipment level. We argue that the sensor reliability issues could easily add up to misdiagnosis rather than helping in reducing associated failure durations and repair actions executed during an unscheduled equipment breakdown. We also argue that equipment are composed of modules also known as assemblies which are linked to one another in parent child relation and has significant effect on the functioning of each other. The emerging and promising PdM based

approaches rely on the computation of RUL (remaining useful life) or identification of failure signatures. The RUL based approaches use physical degradation models to predict the potential failure and are benchmarked at the component level whereas failure signature based approaches are data driven in nature and rely on machine learning and artificial intelligence techniques. The drawback associated with the failure signature based approaches is that these failure signatures do not guarantee the failure alert well before time so that proactive/corrective measures can be either executed to avoid failure occurrences or planned to improve resources utilization. Based on these arguments, we fixed the scope of this research thesis in finding answers to (i) how to address the misdiagnosis at the source product quality drift?, (ii) how to improve failure(s)/cause(s) diagnosis and (iii) how to extract time bound failure signatures to improve not only CM but also PM. These answers will enable us to move towards implementing proactive maintenance framework to improve production capacities in high-mix low-volume production environment. Moreover to avoid equipment sensor reliability issue, we restricted to recent based contextual information for the validation using case study from world reputed semiconductor manufacturer.

The existing modelling approaches used for failure/cause diagnosis and failure predictions are data driven (see chapter 4). These can be further broadly classified as statistical (Principal Component Analysis, Discriminant Analysis, Partial Least Square etc.) and machine learning (Support Vector Machines, Fault Tree, Markov Model, Artificial Neural Network, Bayesian Network etc.) techniques. The choice of BN as modelling approach, in this research thesis, is based on the fact that we use event based contextual information instead of temporal data from sensors. This limits the direct usage of approaches which are based on purely temporal data. However, since sensor based temporal data can be discretized; however, we argue that this will result in aggregation of sensor bias leading to poor decisions. Moreover, our objective is not to automate the product quality drift source, failure/cause diagnosis and failure prediction but to help maintenance teams with decision support mechanisms. Inherently, uncertainty is also associated with the occurrence of failures as well as failure/cause diagnosis due to the complexity of production line operations (high-mix low-volume) and multi purpose fully automated production equipment. The BN is directed acyclic graph composed of nodes and edges. The nodes represent variables whereas edges present the influence of one variable on the other depending on their direction. Therefore, we chose BN as our principal modelling approach as it gives more accurate results with event based contextual information along with probabilistic inference. Moreover, it provides a graphical visualisation of variables of interest along with their influence structure. The drawback with

BN is its limitation in treating large number of variables as it results into NP hard problem when conditional probabilities are compute at each node. This can be controlled by selecting most important and critical variables. However, there existing statistical approaches e.g. mutual information, entropy, to rank most important variables in a given dataset under study. In this research thesis, we based the selection of variables on the experts' opinion due to time constraints. Moreover, computation time for the construction of BN and computation of conditional probabilities was further reduced by enforcing structure learning rules. The variables are grouped into classes and structure learning and computation or probabilities between these classes are constrained.

To the best of our knowledge, there do not exist literature which is focused on identifying the source of product quality drift instead of believing on the hypothesis that equipment is the only source of product quality drifts. Therefore, we proposed, developed and validated a Bayesian network (BN) to reduce misdiagnosis in the source of product quality drift as the first line of defence against increasing unscheduled equipment breakdowns. The failure(s)/cause(s) diagnosis with BN is developed using event based contextual information collected from product, process, equipment and maintenance data sources at both equipment and module (assembly) levels. This serves as the second line of defence against increasing unscheduled equipment breakdowns. The results from the case study demonstrates that BN significantly identifies the product quality drift source other than equipment and that module (assembly) level BN gives more accurate predictions than the equipment level. Besides this, we proposed, developed and validated time bound failure signature extraction methodology based on BN and contextual information. The role of BN in this methodology is to plot failure probabilities over discretized and aggregated contextual information time line. This methodology offers flexibility to extract multiple rules with varying associated lead times such that proactive/corrective measures can be executed to either avoid the failure occurrences or plan maintenance resources. The lead time computation associated with maintenance actions required is subjective and depends on the competence of maintenance team; hence, this is left at the discretion of equipment experts. This methodology offers flexibility to maintenance teams to relax or tighten the corresponding failure alerts based on the production line drift indicators. This is important because an inappropriate choice could result in over or under engineering.

This thesis is divided into three parts. The part I presents introduction to problem context, research goal and objectives, relevant literature review and positioning of the scientific contributions. The part II is focused on the choice of modelling approach and presentation of integrated product quality

drift source and failure/cause methodology, and extraction of time bound failure signatures. In part III, we present the case study and validation results of the proposed methodologies as presented in part 2. The detailed thesis organization is presented in part 1, chapter 3.

Part I
INTRODUCTION

CHAPTER 1

Research Context

Introduction

The objective of this chapter is to present research context, problem background, research goal and objectives. In the start, evolution in production environments from pre to post industrial revolution eras is presented which highlights that industries are challenged with unstable and reduced production capacities in high-mix low-volume production settings. This is due to increasing Corrective (CM) and Preventive Maintenance (PM). Further data collected from a world reputed semiconductor manufacturer, a high-mix low-volume production environment, is analysed to identify causes against increasing CM and PM. The results highlight increasing unscheduled equipment breakdown, repair count, failure occurrences and failure durations as the key contributors.

Contents

1.1 Industrial Revolution and Emerging Production Environments	8
1.2 High-Mix Low-Volume Production Characteristics	9
1.3 Research Goal and Objectives	11
1.3.1 Goal	12
1.3.2 Objectives	12
1.4 Unscheduled Breakdown Triggering Mechanisms	13
1.5 Challenges in High-Mix Low-Volume Production	14
1.5.1 Limitation in minimizing Corrective Maintenance	15
1.5.2 Limitations in improving Preventive Maintenance	16
1.6 Conclusion	17

1.1 Industrial Revolution and Emerging Production Environments

The era from 1760 till 1850 is recognized as industrial revolution era which is characterized as low-mix high-volume production (Flinn, 1974). This is attributed to factory based manufacturing with shift from human to machine power. The pre-industrial revolution era is known as low-volume low-mix human based manufacturing whereas post industrial revolution era is powered by industrial automation that resulted in the emergence of high-mix low-volume production environment.

Contrary to low-mix high-volume production (industrial revolution era), where dedicated production lines are designed for specific products, in high-mix low-volume production (post industrial revolution era):

- equipment are shared among different production tasks
- product life cycles are short
- automation and control technologies result in high capital tied up in production equipment
- uncertain and diversified customer demands make it difficult to achieve stable production
- experts knowledge is available but it is challenged by unknown failures and causes

The competitiveness and performance of manufacturers in high-mix low-volume production environment depend on the availability, reliability and productivity of production equipment. In this context, many researchers have emphasized the importance of the maintenance function of its role in keeping and improving equipment performance and product quality (Al-Najjar, 1996; Al-Najjar and Alsyof, 2003). However, depending on the industry, maintenance activities and associated resources also represent 15-60% of the cost of produced products (Mobley, 2002). This includes many unnecessary or improperly carried out maintenance that significantly reduces production time. Therefore, growing importance of effective maintenance has generated an increasing interest in developing and implementing optimal maintenance strategies to improve system availability, prevent occurrence of system failures, and reduce maintenance costs of deteriorating systems.

In high-mix low-volume production, manufactures need production flexibility to cope up with uncertain volume and product changes to sustain production capacities (Iravani et al., 2005) as well as optimization in equipment utilization (Pehrsson and Al-Najjar, 2005). The equipment breakdowns

significantly reduce and disrupt the production capacities. In case of unscheduled equipment breakdown, more data and information is required; however, experts' knowledge on such unwanted events are becoming more difficult to obtain. This gives rise of interest in having an efficient and effective equipment utilization which can only be ensured through appropriate maintenance strategies.

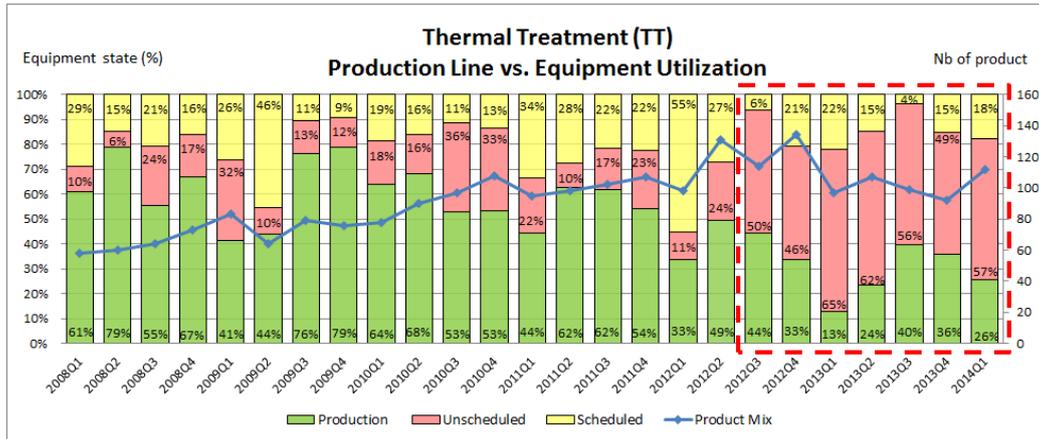
1.2 High-Mix Low-Volume Production Characteristics

This research thesis is completed in collaboration with STMicroelectronics Crolles, France which is one of the world's largest Semiconductor Industry (SI), known for Integrated Circuit (IC) chip manufacturing. At present, the SI is challenged by increasing demand diversity and volume that resulted in short product life cycles and has emerged as a high-mix low-volume production environment. The SI is fragile and fastest growing manufacturing domain that has revolutionized our daily lives with IC chips. On average, we use more than 250 chips and one billion transistors per day per person. These are installed in almost all the equipment around us ranging from dish washer, microwave ovens and flat screens to office equipment. These ICs are manufactured on automated production line, comprising of hundreds of production and metrology/inspection equipment, grouped as different workshops, based on type of operations. These chips are manufactured on silicon wafers of 200/300 mm diameter that undergo up to 1100+ elementary operations, depending on the technology. These are processed in the lots of 25 wafers where each wafer contains around 900 chips and cost 6K to 12K US dollars.

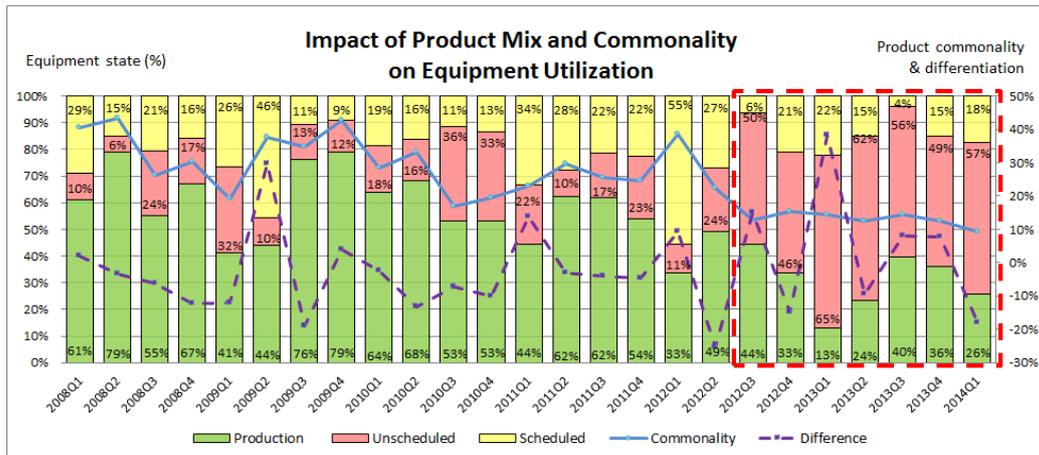
The SI production line comprise of 8 workshops; however, a field study was carried out at the Thermal Treatment (TT) workshop. This is because of the fact that TT is one of the most complex workshops in the SI and carries out 30% of the production operations. Moreover, real time operations and confidentiality limited us to one workshop in the production line. The production capacities in the form of three types of equipment states (production, scheduled equipment down and unscheduled equipment breakdown) were plotted against the evolution in product mix¹ (Figure 1.1a) and then

¹Product mix as the number of product variety corresponding to new technology developed every 2-3 years

against the product commonality² and difference³ (Figure 1.1b), as evidence. The data is aggregated at quarter (Q) level and spans over six years (2008Q1 to 2014Q1). For confidentiality, data is manipulated; however, scale is kept constant to maintain original trends.



(a)



(b)

Figure 1.1: Product mix, commonality and differentiation vs. equipment utilization.

It can be seen that during 2008Q1 and 2012Q2, production capacities are significantly larger than both scheduled and unscheduled breakdowns (Figure 1.1a). In this period, we also observe slight increase in the product

²Product commonality as the number of common product between two consecutive period

³Product difference as the number of different product between two consecutive period

mix that decreases production capacities. The data till 2014Q1 shows that with the fluctuation of the product mix, the production capacities suffer instability and a significant decline. The Figure 1.1b presents impact of commonality (Equation 1.1) and differentiation (Equation 1.2) in product mix on equipment from two consecutive quarters. The difference in product mix is plotted on secondary y-axis. This can be positive or negative and ranges from -25% to +38%; whereas, product commonality is plotted on the primary y-axis, that ranges from 49% to 92%.

$$\text{Commonality} = \frac{\text{Number of common product}}{\text{Quarter's product mix}} \quad (1.1)$$

$$\text{Differentiation} = \frac{\text{Previous quarter's product mix} - \text{Current quarter's product mix}}{\text{Quarter's product mix}} \quad (1.2)$$

It can be seen that production capacities increase with increase in product commonality and are inversely proportional to unscheduled breakdowns. Therefore, production learning curves against demand diversity can be improved by reducing not only the unscheduled breakdowns but also by stabilizing them. It is because instability in the capacities result quick changes in production planning and reshuffling of production lots. In addition, time constraint lots result in scrap that impacts not only cost but also cycle time. The Figure 1.1a also shows that in last two years, an increase in product mix with differentiation and short product life cycles have resulted in 30% reduction of production capacities with high instability. Consequently, unscheduled breakdowns have increased exponentially.

1.3 Research Goal and Objectives

From the analysis, presented in section 1.2 (Figures 1.1a and 1.1b), it can be concluded that the SI is challenged by increasing demand diversity and volume which resulted in short product life cycles. Moreover, production capacities are unstable and reduced due to increasing unscheduled equipment breakdowns, giving rise in unavailability due to corrective (CM) and preventive (PM) maintenance due to CM. The latter is due to the fact that while carrying out diagnosis in CM, technicians often perform PM procedures. These actions give rise to PM actions in a high-mix low-volume production environment.

1.3.1 Goal

The success of the SI in a competitive production environment, challenged by increasing demand volume and diversity, depends on improved, stable and sustainable production capacities. Therefore, research goal in this PhD thesis is to “improve and stabilize production capacities”. Illustration of goal can be referred to Figure 1.2.

1.3.2 Objectives

Increasing corrective and preventive maintenance (due to CM) have resulted in not only reduced but also highly unstable production capacities, attributed to increasing unscheduled equipment breakdowns. This give rise of huge interest in putting efforts to reduce these increasing corrective and preventive maintenance (due to CM); hence, objectives in this PhD thesis are set as under:

- (a) minimize Corrective Maintenance (CM)
- (b) improve Preventive Maintenance (PM) due to CM

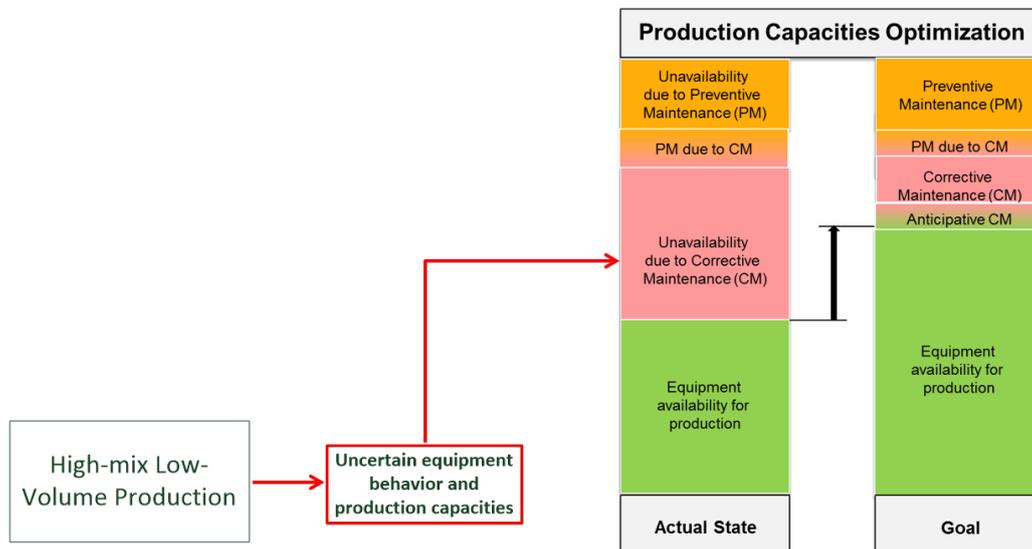


Figure 1.2: Production capacities optimization in view of unavailability linked to CM and PM (due to CM).

1.4 Unscheduled Breakdown Triggering Mechanisms

In the SI production line, maintenance interventions are triggered by two process control functions, Fault Detection and Classification (FDC) and Statistical Process Control (SPC). The definitions are provided in Mili (2009) and an exemplary image which links these controls is provided by Bouaziz (2012) as presented in Figure 1.3.

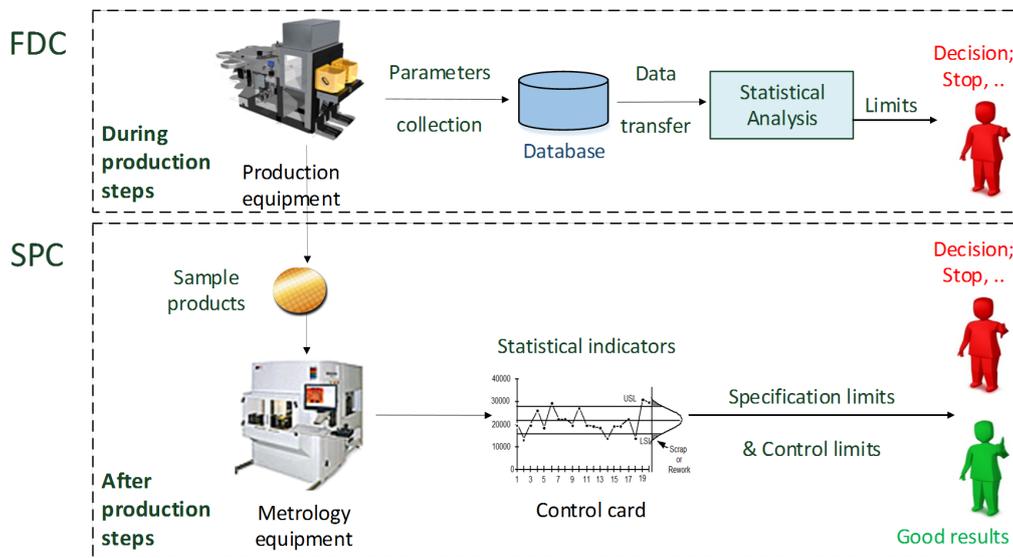


Figure 1.3: Example of process control in production.

- **SPC:** provides control using statistical indicators based on products' (i.e. wafer) physical measures (e.g. thickness, length and height of transistors) and other features. For each critical step, a wafer from a process lot is automatically measured just after the production step to check quality conformance (centring and dispersion). This control methodology is applied to three different types of measures:
 - *Physical measurements:* performed at critical steps (etching, chemical mechanical polishing, ...) through metrology equipment.
 - *Intermediate electrical measurements:* performed at the end of transistors manufacturing.
 - *Final electrical measurements:* performed at the end of production operations and are also known as EWS (Electrical Wafer Sort).

These are associated to two types of limits, (i) specification limits to ensure the proper functioning of ICs and (ii) control limits to verify the success of manufacturing step and detect equipment's drift. If control or specification limits are crossed, the equipment is stopped as unscheduled equipment breakdown. However, in case of out of control, the production lot is stopped for inspection whereas in case of out of specs, production lot is scrapped.

- **FDC**: allows to monitor in real time the evolution of process equipment parameters (e.g. temperature and pressure). Throughout production, equipment parameters are collected for each wafer. The aim is to control an abnormal equipment operation using statistical analysis where equipment is stopped if necessary, and is qualified as unscheduled breakdown. The FDC is applied to control equipment stoppage. It is divided into two processes, first the detection of the fault, and second the identification of the causes for the fault. This is of particular interest for the SI as the root causes of faults in a complex process may be dozens, even hundreds and in a dynamic and highly automated production, experts face the difficulty in understanding the correlation between abnormal events, causes and effects (Susto et al., 2012).

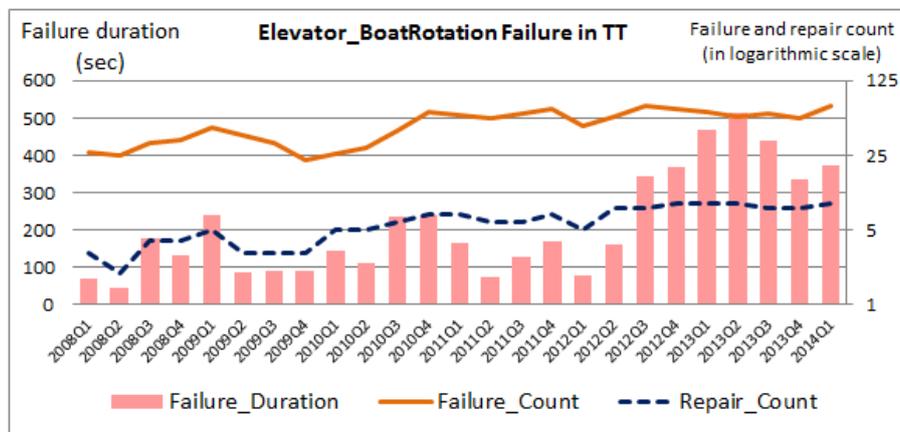
In summary, in the SI production line, unscheduled breakdown can be triggered either at the equipment or product levels upon statistical limits detection. At equipment level, unscheduled breakdown is triggered by FDC system; however, misdiagnosis can occur due equipment failure undetection. The reasons are sensor bias or absence of FDC rules in its statistical analysis. At product level, unscheduled breakdown is triggered by SPC upon product quality drift; however, misdiagnosis can occur due to unrelated equipment failure. The reasons are absence of product drift source detection mechanism. The triggered unscheduled breakdown can be characterized by several indicators such as the number of triggered failure alarms (Failure count), stoppage duration due to triggered failure alarms (Failure duration) and number of repair actions upon triggered alarm (Repair count).

1.5 Challenges in High-Mix Low-Volume Production

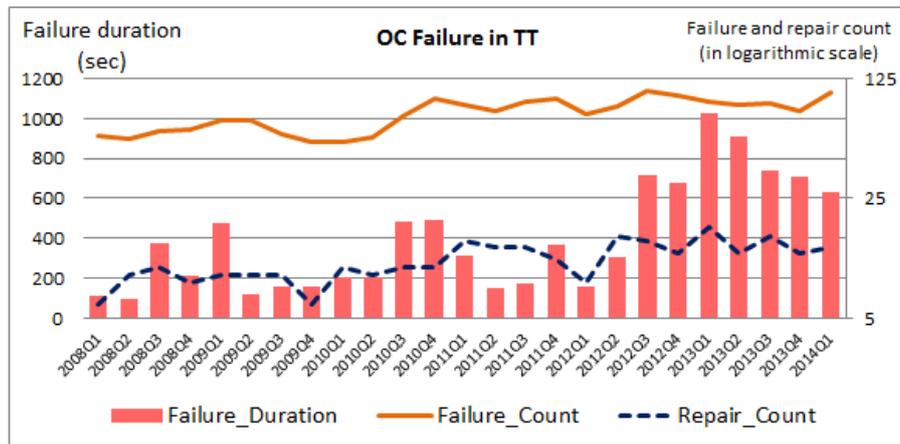
The challenges (increasing demand volume and diversity) faced by the SI results in increasing unscheduled equipment breakdown which consequently results in an increase of CM and PM (due to CM). Therefore, these increasing CM and PM must be further investigated to identify the respective causes.

1.5.1 Limitation in minimizing Corrective Maintenance

Further analysis using data from TT workshop is presented below in Figure 1.4 to identify factors for recent huge increase in unscheduled equipment breakdowns and resulting CM. This analysis is extended on two type of failures (unscheduled breakdown events), as presented in Figure 1.4a and 1.4b. The failure duration (sec) is plotted on the primary y-axis whereas failure occurrences and number of repair actions in each occurrence are plotted on the secondary y-axis.



(a) Elevator boat rotation failure



(b) Out of Control (OC) failure

Figure 1.4: Failure counts, durations and occurrences.

The data in Figure 1.4 is plotted for two significant failures: (a) Elevator boat rotation and (b) Out of Control (OC) and is manipulated for confidentiality reasons. It can be seen that failure count and average number of repair

actions in each occurrence are inversely proportional to product commonality that reduces process variations and result in stable capacities and controlled unscheduled breakdowns. In addition, OC is 30% higher than elevator boat rotation in its occurrence and failure duration. Therefore, increase in failure duration, occurrence and number of repair actions are the key indicators proving misdiagnosis in increasing unscheduled breakdowns by the current practice of FDC and SPC. Beside FDC equipment failures misdiagnosis, we argue that misdiagnosis can also occur while identifying sources of product quality drifts. In a complex production environment, e.g. SI, the likely sources can be product itself (imperfections from previous process steps or poor design), process (poor operations recipes design) or maintenance (poor execution of maintenance actions). However, at present, the equipment is blamed for all product quality drifts.

1.5.2 Limitations in improving Preventive Maintenance

From the analysis results in section 1.5.1, it can be suggested indirectly that increase in CM increases PM as well due to the fact that while carrying out diagnosis, technicians often perform PM procedures. These actions give rise to PM actions in a high-mix low-volume production environment. Both CM and PM due to CM can be improved by mastering the increasing uncertain equipment behaviour using failure prediction, gearing the PM towards Predictive Maintenance (PdM). At present, PdM is carried out by online failure prediction approaches based on failure signatures and Remaining Useful Life (RUL) computation. The RUL approaches uses physical models of degradation benchmarked at component level. The failure signature based approaches are promising but lacks time bound element that can ensure prediction of a failure well before its occurrence such that proactive measures can be taken to avoid failure and its propagation. To illustrate the importance of time bound element in failure prediction, a scenario from the same production line is considered (Figure 1.5).

A production equipment is either single or multi chamber equipment with 5,000 sensors and data frequency at 10ms. In a complex and HMLV production environment, production runs are long (around 4 to 6 hours) and cannot be stopped once launched. Without the capability to project failure f_1 at time t_1 such that f_1 can corrected before starting of production run 2, then this failure will propagate into other faults and failures. Each fault is converted to work request and each failure is converted into work order which result in higher failure counts, failure duration, CM, PM, unstable production capacities and million USD loss and waste of maintenance resources. Therefore, it is critical to introduce a time bound element through failure prediction to

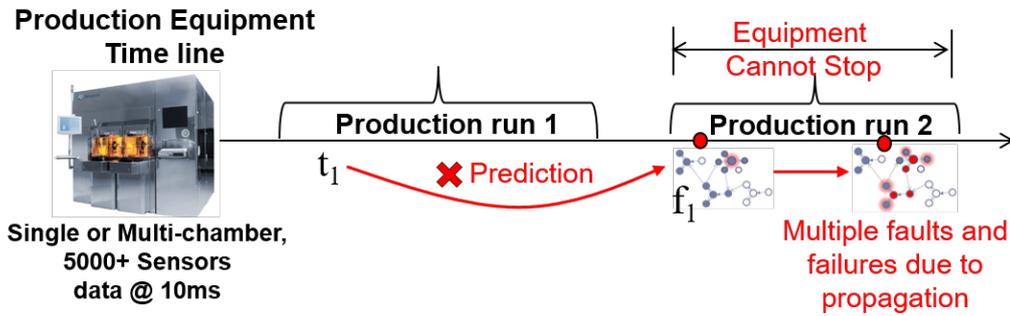


Figure 1.5: Example of process control in production.

avoid such failure propagation in a long and unstoppable production run.

1.6 Conclusion

The high-mix low-volume production environments are challenged by increasing demand diversity which lead to increasing unscheduled equipment breakdowns. This results in increasing CM and PM actions, destabilizing and reducing production capacities. In the competitive environment, like SI, the success depends on our ability to improve and sustain production capacities. This give rise of interest in improving and smoothing production capacities while minimizing CM and improving PM. The analysis carried out using data of Thermal Treatment workshop, from one of the world's largest IC manufacturer, advocates that potential causes for increasing CM in high-mix low-volume production environment is the misdiagnosis. This leads to increasing failure durations, failure counts and repair counts. It is also observed that in most of the equipment stoppages, equipment is not the source of product quality drift whereas drifts source can also be traced to product design, process and maintenance. This highlights that the challenges in reducing CM are misdiagnosis of product quality drift sources and failure/causes during unknown failures. Moreover, PM due to CM is also increased due to the fact that while carrying out diagnosis for unscheduled breakdown, technicians often performs PM procedures. The challenge in reducing PM is lack of ability to predict online failure using time bound failure signatures, well before time, required to proactively correct causes and hence, avoid failure propagation.

CHAPTER 2

State of the Art

Introduction

This chapter presents literature review on the role of existing maintenance strategies in reducing CM and improving PM in order to stabilize and sustain production capacities. Moreover, existing approaches for diagnosis of product quality drifts and failure and causes are briefly presented along with online failure prediction to highlight scientific gaps in the context of high-mix low-volume production. This chapter concludes with research questions focused in this research thesis.

Contents

2.1	Role of Maintenance Strategies	18
2.1.1	in reducing Corrective Maintenance (CM)	19
2.1.2	in improving Preventive Maintenance (PM)	20
2.2	Failure/Cause Diagnosis Framework to reduce Cor-	
	rective Maintenance	21
2.2.1	Source of Product Quality Drifts	22
2.2.2	Equipment Failure/Cause Diagnosis	23
2.3	Predictive Maintenance Framework to improve Pre-	
	ventive Maintenance	24
2.3.1	Online Failure Prediction Approaches	25
2.3.2	Failure Signature based Approaches	27
2.4	Scientific Gaps	29
2.5	Research Questions	31
2.6	Conclusion	31

2.1 Role of Maintenance Strategies

Maintenance strategy is defined as a collection of maintenance activities combined with different methods, tools and techniques to reduce maintenance

costs while increasing reliability, availability, maintainability and security of equipments. These activities comprise of fault detection, isolation and identification, fault/failure prognostics, regular/ad hoc inspections, etc... The goal of a maintenance strategy is to become aware of events, understand them and act upon them appropriately.

To become aware of maintenance events, we refer to the standard french maintenance terminology¹ of 2001 for the concepts fault and failure. These two concepts are employed largely in literature review for failure analysis supporting the implementation of different maintenance strategies. **Fault** is the state of an item characterized by inability to perform a required function, excluding the inability during preventive maintenance or other planned actions, or due to lack of external resources. **Failure** on the other hand is the termination of the ability of an item to perform a required function. It is to be noted that fault is a state whereas failure is an event and after the occurrence of a failure, equipment becomes faulty which may be complete or partial. Moreover, the concept **Defect** refers to the difference between actual and expected characteristic of an entity which exceeds the limits of acceptability. This does not affect the overall ability of the equipment to perform a required function; therefore, defect does not always lead to a failure. However, a failure systematically leads to a defect but this may also result in defect (quality drift) of products.

In order to understand the events, it is important to distinguish the existing phenomena used when modelling an item's fault and/or failures. Pérès and Noyes (2003) give an extensive review of different forms and rates of failures, consequently propose to analyse data to link the different types of failure in a system to varying repair time. The most common of these phenomena is the continuous degradation which is the irreversible process in one or more characteristics of a component with either time, use or an external cause. This degradation may lead to failure and is often referred as wear out.

Finally, to act consequently and appropriately to events, the evolution and role of maintenance, strategies classified as CM and PM, is presented in the next two subsections.

2.1.1 in reducing Corrective Maintenance (CM)

Industrial maintenance practices have gone through a process of change in the last decades due to the increasing awareness of the importance of maintenance management. Starting the evolution in the 1950s, CM (or run-to-

¹NF EN 13306, retrieved on 3rd May 2016 from: <http://maint.t.i.b.free.fr/Files/Other/NF%20EN%2013306.pdf>

failure) was the prevalent maintenance strategy. This strategy is reactive in nature which involves inspection of faults/failures; however, maintenance actions are triggered upon the occurrence of a fault/failure. These actions can be either replacement or repair. The corrective maintenance actions are divided between palliative actions which focus on correcting problem temporarily without solving them permanently whereas curative actions involve not only problem identification but also definitive repair.

The constant changes in the industrial production environment and field of maintenance have enabled new and innovative maintenance strategies, refer to section 2.1.2, however CM strategy can never be fully replaced as equipment failure behaviour is stochastic (cannot be 100% predicted) and breakdowns are unforeseen.

2.1.2 in improving Preventive Maintenance (PM)

In the 1960s, systematic preventive (time or other elapsed unit count based) maintenance became popular when a group of Japanese engineers began a new maintenance concept which was to simply follow recommendations from equipment manufacturers to avoid failures. In the PM, regular and systematic equipment component replacements are scheduled to avoid any potential unscheduled failure regardless of the real time health status and equipment performance. However, in the latter half of the 1980s, the PM resulted in over engineering problem (i.e. replacement of components with potentially longer RUL) (Pintelon and Parodi-Herz, 2008). This and due to the fact that sensors and the underlying techniques in equipment function became more available, have resulted in condition monitoring and diagnostic techniques which led the next class of PM strategy known as the Condition Based Maintenance (CBM). The CBM introduces continuous or periodic measurement and interpretation of equipment's condition data to determine the need for maintenance based on decision thresholds. The objective of CBM is to avoid unnecessary maintenance activities by triggering these actions only when there is an evidence of failure (e.g. deterioration) or abnormal behaviour. A CBM program, if properly established and effectively implemented, can reduce maintenance cost by reducing the number of unnecessary systematic PM actions (Pintelon and Parodi-Herz, 2008).

Recently, prognostic, which deals with fault and failure prediction before occurrence, made its way as a concept to support proactive maintenance strategy. This maintenance strategy monitors parametric evolution and determines whether a fault/failure is impending and estimates how soon and likely it will occur (Jardine et al., 2006). This results in another classification of maintenance approaches as prognostics and diagnostics. The prognostic

approaches are based on prior event analysis whereas diagnostic approaches rely on posterior event analysis. A maintenance strategy which incorporates prognostics in the decision process is defined as a Predictive Maintenance (PdM). This means that, in comparison to CBM, PdM incorporates more information into the maintenance decision process. The benefit of using information about future degradation over currently observed information is illustrated in different publications (Zhou et al., 2007; Yang et al., 2008).

Figure 2.1 synthesized the differences between CM and PM strategies (Castanier, 2001). The implementation of these strategies provide the most basic distinctions to the different criteria for decision making in the field of maintenance.

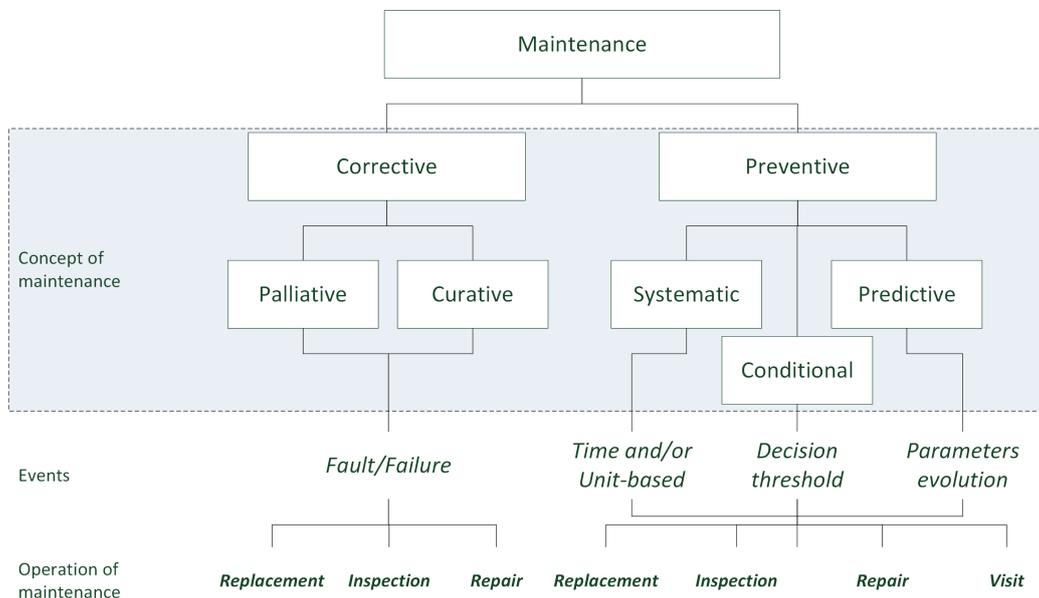


Figure 2.1: Evolution of maintenance strategies.

2.2 Failure/Cause Diagnosis Framework to reduce Corrective Maintenance

This section is dedicated to the presentation of two aspects for misdiagnosis analysis as identification of product quality drifts sources leading to equipment stoppage and failure/causes diagnosis during unscheduled breakdowns.

2.2.1 Source of Product Quality Drifts

In any modern production environment, equipment and product are two major actors and have a mutual link between them. The equipment is responsible to ensure transformation of raw material into products which are compliant to customer specifications. However, the performance of an equipment is highly dependent on historical maintenance actions and their execution in compliance with operating procedures. Moreover, product design itself can also be vulnerable to production line variations. As production is carried out in production steps; therefore, outcome of the previous production steps on subsequent steps can cause serious compliance issues with reference to customer specifications or product functionality. Similarly, poorly designed process (technology) can also cause serious non compliance issues. Contrary to these potential product quality drift sources, in literature, most of the research is focused on identifying the causes against product failure within equipment (Doty, 1996; Smith, 1998). They use causes classification by Ishikawa (1990), a division of six assignable categories: man, machine (equipment), method, material, measure and environment that explain abnormal situations. It is a well-known qualitative method, frequently used in the diagnosis domain, but requires long brainstorming sessions with experts and is performed for new excursion. Therefore, it cannot be used in complex production environment for all excursions. Sarkar (2004) has combined cluster analysis with engineering knowledge to classify big sets of equipment failure events into small number of categories and use experts' knowledge to find root causes for each cluster. These are specific for the equipment related failures and do not take into account other potential sources of drifts. Weidl et al. (2005) model industrial process and product failure control system using generic object oriented Bayesian Network (BN) that proposes corrective maintenance actions with explanation of root causes. Their set of root causes contains all possible hypotheses on the failure sources or conditions coming from the equipment sensors, process conditions and basic failures in maintenance. This approach does not take into account the product related failure events as the cause of product quality drifts. Besides this, the BN structure is also defined by an expert; however, experts' knowledge might need renewal due to dynamically changing production environment behaviour.

Additionally, in the context of high-mix low-volume production, metrology/test equipment allows control on product specifications compliance across production steps. In certain cases, metrology equipment are not installed after every production step for economical reasons and due to limited control capacity. This introduces the phenomena of failure propagation where detection of an equipment from multiple consecutive production steps as the

product quality drift source becomes challenging. Nguyen (2015) develops a diagnostic logic and BN as a solution to this problem. The author also concludes that product itself can be one of the causes for equipment failure apart from human, recipe (process) and equipment anomalies.

As presented above, it can be concluded from literature that equipment is blamed to be the only source of product quality drifts. Therefore, high-mix low-volume production environment, challenged by increasing unscheduled equipment breakdowns, directly contributes to increase respective CM. This is because of the fact that an equipment stopped due to product quality drift requires diagnosis and qualification procedures to be executed by maintenance teams before it is added back to the production capacity pool. This implies that equipment are stopped for diagnosis and repair for product quality drifts when in fact the source probably is not the equipment. This argument needs to be integrated in the modelling of equipment failure as it serves as the first line of defence against the exponential increase in unscheduled equipment breakdowns, failure durations and number of repairs in each failure.

2.2.2 Equipment Failure/Cause Diagnosis

The commonly used techniques to optimize production operations are Advanced Process Control (APC) methods that include Run to Run (R2R) loops, Statistical Process Control (SPC) and Fault Detection and Classification (FDC). Yue and Tomoyasu (2004); Lacaille and Zagrebnev (2007); He and Wang (2007) used FDC approach to detect and classify equipment failures by calculating several statistics of collected parameters from FDC sensors data, on predefined time windows. These result in indicators which are then monitored through SPC control charts to detect sources of variation in the form of shift or drift of equipment signals. A comparable approach has been proposed by Chen and Blue (2009) using EWMA (Exponentially Weighted Moving Average) chart as a function of variance and covariance of relevant parametric distributions to assess the quality of equipment. However, this approach is objectively different from the above approaches as it integrates all sensors to generate one single index that reflects the overall equipment health against the product quality and is argued to be more robust to recipe and operation changes. Chang et al. (2012) proposed a fault detection and classification methodology for the SI using a sequential SVDD (Support Vector Data Description) classifier algorithm.

A careful analysis of the existing approaches, methods and techniques, highlights that till today, to model equipment behaviour, a significant number of its parameters (status variable identifications) from sensors are col-

lected during wafer processing. With this data, the principal objective is to improve fault detection and failure diagnosis on the equipment. However, due to frequent changes in recipes (process) and diversity of operations, in a high-mix low-volume production environment, overall equipment condition is very difficult to be evaluated. Because of these changes, Blue et al. (2012) highlighted in their research that FDC indicators based on pattern modelling for specific recipes can have reliability issues as they are not appropriate to represent the equipment conditions continuously. Hence, they proposed Generalized Moving Variance (GMV) technique in a hierarchical scheme for monitoring; aiming for a robust estimation of overall equipment condition based on the similar variations in FDC sensors data.

In literature, proposed approaches for failure/causes diagnosis use FDC data which can be misleading in dynamic production environment. They also focus on either estimating the components health or the overall equipment health through an index for equipment failure diagnosis; whereas in most of the cases, the cause assignment is left at the maintenance engineer's judgement. As the second line of defence, failure and causes prediction on an unscheduled equipment breakdown must be modelled at module level rather than at equipment level, using contextual and statistical information. Equipment in the SI is composed of modules and sub modules. Each one representing a system, however in previous works, it is considered as a unique system. A comparison of diagnosis models at equipment and module level is crucial to provide us with the more reliable model as maintenance decision support for engineers. Consequently, reduction in failure durations and number of repair actions should stabilize and improve capacities with reduction in the unscheduled equipment breakdowns.

2.3 Predictive Maintenance Framework to improve Preventive Maintenance

Prediction of failure occurrence is the key task of PdM. Limiting the scope to fault/failures, there are several areas where the term prediction is used. For example, in reliability theory, the goal of reliability prediction is to assess future reliability of an item from its design or specification. Denson (1998) gives an overview of reliability prediction techniques applied to electronic devices. These are concerned with long-term predictions based on the failure rates, architectural properties, or the availability indicators such as MTTR and MTBR. However, the focus of our survey is to identify during runtime (online) whether a failure will occur in the near future based on an assess-

ment of the monitored current system's parameters evolution. Online failure prediction is a domain concerned with this short-term assessment that allows to decide, whether there will be a failure ahead in time or not (Salfner et al., 2010).

This section is dedicated to the presentation of two key approaches from the online failure prediction, those using Remaining Useful Life (RUL) estimation and those that rely on failure signatures.

2.3.1 Online Failure Prediction Approaches

Since the introduction of prognostics and PdM, failure prediction is traditionally and most commonly defined as being the RUL estimation of a system knowing his current health, failure signatures (representation of system degradation) and/or future conditions of use. The estimation of RUL can be achieved by various methods using different processing tools, modelling techniques and type of analysis. Several attempts to classify these are proposed in the literature. A synthesis of the RUL classification in the literature can be referred in the Table 2.1. The key contributions on this concept can be classified into methods that evaluate the content of input data (raw, featured data/information or expert's knowledge)(Vasile, 2008) and the nature of input data (events-based, states-based or both events and states) (Heng et al., 2009). Furthermore, a classification according to the type of input data (fault, error, symptom and failures) was proposed in (Salfner et al., 2010) which includes the RUL estimation but also the failure signature approach. These classification can be further categorized into methods that evaluate continuous and regular temporal data and those that rely on discrete and irregular inputs.

With the development of sensors technology and real-time data collection, most research on failure prediction is placed in the temporal input category, where system's condition is usually modelled as periodic assessment using structural time series followed by forecasting deterioration to functional failure using regression methods (Lu et al., 2007; García et al., 2010; das Chagas Moura et al., 2011). This type of evaluation generally uses features from sensors and they focus on discerning anomalies and estimation of RUL without identification of state and/or type of failure of the equipment. This limits its use in multi state equipment with multi failure types (failure modes under different failure mechanisms) such as the case in high-mix low-volume production. Moreover, due to diversity in operations, overall equipment condition is exposed to sensor biases and is difficult to be evaluated accurately (Blue et al., 2012).

It is also observed that the concept of online failure prediction through

Table 2.1: Classification of RUL methods and tools

Criteria for classification	The classification	Works
Initial and most basic classification	<ul style="list-style-type: none"> - Based on a model - Guided by data - Based on experience 	Lebold and Thurston, 2001
The content of input	<ul style="list-style-type: none"> - Raw data - Featured data or Information - Expert's knowledge 	Vasile, 2008
Nature of the input data	<ul style="list-style-type: none"> - Events-based - States-based - Both events & states 	Heng et al., 2009
Decision based on 3 criteria: The objectif, future context utilisation, origin of knowledge	Combination of <ul style="list-style-type: none"> - 3 types of objectifs - 3 types of future context - 4 types knowledge 	Cocheteux, 2010
Decision based on four criteria: The complexity, cost, accuracy and applicability (Referenced to Lebold and Thurston classification)	<ol style="list-style-type: none"> 1) Based on model, guided by data and based on AI 2) Based on model, guided by data, based on knowlegde and fusion of models 3) With vs. without models 4) Based on model, guided by data and hybrid of model & data 	Jardine, 2006 Peng et al., 2010 Zio and Di Maio, 2010 Lee et al., 2012

RUL estimation targets its application on component or system composed of several components (Kothamasu et al., 2009; Sikorska et al., 2011; El-Koujok et al., 2014). This estimation at component level is inevitable for its ability to detect needs for components replacement, nevertheless the concept is misleading at the equipment level and suggests that no prevention is possible. The existing approaches primarily rely on sensors and continuous monitoring of data to calculate features (or indicators) exceeding pre determined threshold and subsequently real time failure prediction is used to plan corrective actions.

In addition to the estimation of RUL there are few published examples of prediction models being applied on complex systems. Only Wang (2002); Lin et al. (2003) considers that in some situations, it is more desirable to predict the chance that an item operates without failure up to some future time (e.g. next inspection interval) given the current equipment condition and historical operation. Actually, in any situation, the probability that an equipment operates without failure until next intervention (e.g. the systematic preventive maintenance (PM) interval) could be a good reference for maintenance teams to determine whether the inspection interval is appropriate or not. In 2013, Bouaziz et al. (2013) propose this type of prediction in the SI in his proposition to project next states of Equipment Health Factor (EHF) to predict future states of the equipment. This approach is promising, however

in high-mix low-volume production environment, these PM intervals are not optimized.

2.3.2 Failure Signature based Approaches

Another key approach in online failure prediction is the prediction without estimation of RUL using failure signatures involving discrete and irregular inputs. Alternatively to temporal input data, prediction of failures can be gained by working on a collection of equipment's event and log data from equipment and maintenance data sources, containing rich operational information to draw failure signatures (Gu et al., 2008; Sipos et al., 2014). Several techniques have been proposed to extract failure signatures from these type of data:

- **Classification:** In this, temporal input vectors (one or more event reports) are processed by classification techniques e.g. Nearest Neighbour (NN) and Support Vector Machines (SVM), to associate it to a class (Liang et al., 2007; Fulp et al., 2008; Fronza et al., 2013).
- **Pattern recognition:** This finds temporal sequences of events to identify patterns that indicate a failure-prone equipment state. Salfner et al. (2006) present an approach called Similar Events Prediction (SEP) which is a Semi-Markov model, where each event is represented as a state, to identify suspicious patterns of faulty events. Sequence similarity is computed by the product of state traversal probabilities. To improve this method, several works have been developed using Hidden Markov Models (HMM) to estimate sequences of hidden degradation states of a system before a failure occurs (Salfner and Malek, 2007; Zhou et al., 2010; Vrignat et al., 2015).
- **Rules extraction:** Similarly, this technique put interest in finding sequences of events; however, they express temporal ordering of events such that “if event A and B occur within a specific time window x , then target event (failure) C occurs within the next time window y with probability p ”. Several parameters such as the minimum and maximum length of the time window (Δtp and/or Δtd), type of data accepted as events and ordering requirements had to be pre-specified. Vilalta and Ma (2002) describe a method to search for predictive subsets of faults (*eventsets*) occurring prior to a target failure using plots of different fault types in a moving fixed size time window (Δtp). Their method addresses the issue of class imbalance (rare failure occurrences compared to non-failures) by filtering the *eventsets* appearing frequently

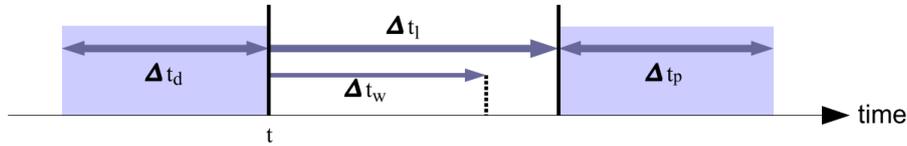


Figure 2.2: Elements of time and their relation in online failure prediction.

prior to both failure and non-failure before they are combined into a rule-based model for prediction.

The works mentioned above are distinguished following the data pre-processing to identify predictors, to create vectors, to mine the sequences of events and to label data using log files. These employ different techniques for the extraction of the signatures and the treatment of the element of times in their prediction approach. These approaches also take into account the class skewness in the datasets with small number of failure occurrences, targeting the application environment of information system. However, to the best of our knowledge, no works have attempted to find the failure signatures integrating other contextual informations coming from the process and product database in a single model. Unlike events reported in equipment and maintenance logs that provide information on equipment behaviour, contextual information of process and product data sources are only implicitly reported to faulty equipment states; however, dependencies exist between contextual information and equipment performances.

Most importantly, none of the works in this approach provides the time bound failure signatures which enable to take into account production run temporal constraint, either by preventing occurrence of a failure, or by preparing the diagnosis and repair actions in order to avoid failure propagation and reduce the duration of unscheduled breakdown. To benefit the advantage of online failure prediction more precisely, four different time elements need to be explored (see Figure 2.2)²:

- **Lead-time** Δtl defines how far from present time failures are predicted in the future. This is similar to the concept of RUL.
- **Minimal warning-time** Δtw defines the minimum lead-time such that failure prediction is of any use. If lead-time were shorter than the warning time, there would not be enough time to perform any preparatory or preventive actions.

² Figure is provided by Salfner et al. (2010)

- **Prediction-period** Δtp is the time for which a prediction holds. Increasing Δtp increases the probability that a failure is predicted correctly. On the other hand, if Δtp is too large, the prediction is of little use since it is not clear when exactly the failure will occur.
- **Data window size** Δtd defines the amount of data that is taken into account for failure prediction. Even if online failure prediction algorithms take the current system state into account, many algorithms additionally investigate what happened shortly before present time. However, in some approaches the amount of data is not determined by a time window but other measures such as, e.g. a fixed number of error events. In this case Δtd is also defined, but may vary with each prediction.

In the above figure, present time is denoted by t . Failures are predicted with Δtl , which must be greater than minimal Δtw . A prediction is assumed to be valid for some time period, Δtp . In order to perform the prediction, some data up to a time horizon of Δtd are used. Different works exploit different elements of time with different type of failure signatures, however, none takes into account the time Δtw with the boundness to correct failure while identifying the failure signatures. Significant gains in production sustainability can be achieved by adopting a proactive approach where causes are treated to avoid failure occurrences and its propagation in upcoming production run.

2.4 Scientific Gaps

As presented above, it can be concluded from literature that equipment is blamed to be the only source of product quality drifts. Qualitative (Doty, 1996; Smith, 1998; Sarkar, 2004) and quantitative (Weidl et al., 2005) methods have been developed to identify the causes against product drift but all of these efforts are focused on finding causes within the equipment. However, Nguyen (2015) highlights that product itself can be one of the causes for equipment failure apart from human, recipe (process) and equipment anomalies. The scope of the qualitative approaches e.g. brain storming, Ishikawa etc. can be equally adapted to find potential cause against the product quality drift beyond equipment. These approaches are based on subjective judgments by group of experts' and can produce most accurate results; however, its inclusion in an automated high-mix low-volume production settings is not possible. This is because of the fact that qualitative approaches are time consuming and automated production line cannot be stopped for each excursion

(increasing unscheduled equipment breakdown) on the production line. In comparison with qualitative, quantitative approaches are promising because they can produce results in short time and are based on data which is collected across the production line. These data driven statistical/probabilistic approaches focus on identifying the causes within equipment for excursions across the production line.

We argue that increasing unscheduled equipment breakdown in high-mix low-volume production settings against product quality drifts comprise of significant number of equipment stoppages where source is other than equipment. Therefore, if we can identify the source of product quality drift other than equipment, unscheduled equipment breakdowns and respective CM can be significantly reduced. This is because of the fact that an equipment stopped due to product quality drift requires diagnosis and qualification procedures to be executed by maintenance teams before it is added back to the production capacity pool. This will also act as first line of defence against increasing unscheduled equipment breakdown, failure durations and respective CM actions; and will contribute to stabilize and improve production capacities.

From the above discussion, we conclude the first scientific gap as the missing data driven methodology to identify actual source of product quality drift in high-mix low-volume production settings with an objective to reduce unscheduled equipment breakdowns and CM actions while stabilizing and improving production capacities.

The most common Advanced Process Control (APC) methods used for failure/cause diagnosis are FDC and SPC. These methods use sensor data on predefined time windows. These are data driven approaches and are primarily used with CBM maintenance strategy. Moreover, failure/cause diagnosis models are developed using statistical and machine learning approaches. The sensors data could easily add bias due to sensor reliability issues; however, there are very few approaches that take into account contextual information or hybrid of contextual and sensor data for modelling purposes. Moreover, revolution in information technology has resulted in more sensors than ever before. This increases not only the volume but also the dimensionality of data. Consequently, modelling becomes complicated because accurate models can be developed with minimum dimensions and this gives rise of interest in dimensionality reduction research. The literature on failure/cause diagnosis modelling also concludes that models are developed at equipment level.

Increasing data dimensionality and volume along with sensor reliability issues cannot cope up with dynamically changing production line behaviour (variations); therefore, we argue that contextual information if used for modelling failure/cause diagnosis can give more accurate results. Moreover, we

argue that if failure/cause diagnosis modelling is done at the module level, the accuracy of the results will be much higher. These gaps, if bridged, will help us in developing more robust and accurate failure/cause diagnosis models which will reduce not only the diagnosis durations but also respective CM actions.

Literature on predictive maintenance framework presents RUL and failure signature based approaches that objectively focus on predicting failures prior to their occurrences. The RUL based approaches use physical degradation models and benchmark component usage whereas failure signature based approaches focus on modelling and associating patterns with potential failures. The former is promising approach; however, we argue that this can be beneficial in improving preventive maintenance if and only if they predict a failure prior to its occurrence such that respective correction action or proactive measures can be executed to avoid failure. This requires identifying time bound failure signatures where each failure has predefined time required for proactive or corrective measures.

2.5 Research Questions

These scientific gaps, presented in section 2.4, are formulated into three research questions as:

1. how to reduce misdiagnosis in product quality drift sources and failure and causes in high-mix low-volume production?,
2. what is the best modelling level (equipment or module) for more accurate failure causes diagnosis?, and
3. how to extract and validate time bound failure signatures (rules) for predictive maintenance?

2.6 Conclusion

In this chapter, we have identified research gaps which are then formulated into research questions. The first research gap is that equipment is believed to be the only source of product quality drifts. In high-mix low-volume production settings, product quality drift source can vary from equipment, product, process (technology) and maintenance. Therefore, in the absence of multi source diagnosis model for product quality drifts, it is not possible to reduce increasing unscheduled equipment breakdowns and respective CM and PM actions. The second research gap, identified from literature, is the level of existing failure/cause diagnosis models. The existing failure/cause diagnosis

models use sensors data that could easily generate reliability issues and are benchmarked at the equipment level. The equipment in the SI are composed of module and sub-modules and are modelled in parent child relation. We argue that modelling failure/cause diagnosis using contextual information from multi-source data, collected across the production line, and bench marked at module level will provide more accurate and robust diagnosis model. This will help to reduce unscheduled failure/cause diagnosis durations with an objective to stabilize and improve production capacities. The last research gap is the absence of time bound failure signatures. The existing approaches to identify failure signatures do not ensure failure prediction such that corrective actions to avoid failures can be executed. Therefore, to improve PM, it is very important to identify failure signatures (rules) such that we have sufficient time to execute corrective and proactive measures to avoid failures. These research gaps are formulated as three research questions which can be referred in section 2.5.

Positioning of Contributions and Thesis Organization

Introduction

This chapter presents brief presentation of proposed contributions against research questions (section 2.5) and positioning along with thesis graphical schematic and thesis organization. The objective is to provide a picture of this research thesis to the readers about its structure and presentation.

Contents

3.1	Contribution and Positioning	33
3.1.1	Integrated 2-Level Product Quality Drift and Failure/Cause Diagnosis Methodology	33
3.1.2	Methodology to Extract and Validate Time Bound Failure Signatures for Predictive Maintenance	35
3.2	Thesis Schematic	36
3.3	Thesis Organization	38
3.3.1	Part-I: Introduction	38
3.3.2	Part-II: Choice of Modelling Approach and Proposed Methodologies	38
3.3.3	Part-III: Application and Results	39
3.4	Conclusion	40

3.1 Contribution and Positioning

3.1.1 Integrated 2-Level Product Quality Drift and Failure/Cause Diagnosis Methodology

The proposed methodology responds to research questions 1 & 2. The Level-1 in this methodology acts as first defence line against increasing unscheduled

equipment breakdowns. In this level, contextual information collected from 4 sources of data (product, process, equipment and maintenance) are used to build Bayesian Network (BN) based diagnosis model to identify actual source of product quality drifts. The objective is to intelligently optimize equipment stoppages. This means that equipment must be stopped if it is the source of product quality drift otherwise it should continue to serve production capacities. The Level-2, in this integrated methodology, proposes failure/cause diagnosis model, based on Bayesian network, both at equipment and module levels. This level gets executed only if the source of product quality drift is identified as equipment (Result from Level-1).

The Level-1 significantly reduce unnecessary equipment stoppages because, at present, in literature it is believed that equipment is the only source of product quality drifts. The potential sources of product quality drifts modelled in proposed integrated methodology are equipment, product, process (technology) and maintenance. This level provides an added advantage, over existing models in the literature, as it helps in stabilizing and improving production capacities by stopping equipment only when it is the source of drift. The Layer-2 complements the proposed integrated methodology by adding BN based failure/cause diagnosis model. Moreover, results highlight that failure/cause diagnosis modelling is more beneficial at module level than the equipment level, as benchmarked in the literature.

The proposed Integrated 2 levels product quality drift and failure/cause diagnosis methodology¹ is validated using data collected from world's reputed semiconductor manufacturer. The case study is presented in Part-III, Chapter 7. The methodology itself is presented in Part-II, Chapter 5 whereas the case study results of the methodology are presented in Part-III, Chapter 8. This methodology also provides answers to following research questions:

Q(1): how to reduce misdiagnosis in product quality drift sources and failure and causes in high-mix low-volume production?

A(1): An integrated 2 levels product quality drift and failure/cause diagnosis methodology, based on BN, is proposed and validated using multi source data (product, process (technology), equipment, maintenance) collected from world reputed semiconductor manufacturer.

Q(2): what is the best modelling level (equipment or module) for more accurate failure causes diagnosis?

¹(Abu-Samah et al., 2014, 2015b)

A(2): Results obtained from failure/cause diagnosis model, based on BN, developed at both equipment and module levels concludes that module level failure/cause diagnosis models demonstrates more accurate results.

3.1.2 Methodology to Extract and Validate Time Bound Failure Signatures for Predictive Maintenance

The proposed methodology responds to research question 3. In this methodology, we plot failure probabilities using already learned and trained BN. Subsequently, patterns and rules are extracted along with validation using historical data along with predictability index and computation of time gain for each pattern and rule. This offers flexibility to the engineering managers to select appropriate pattern/rule for each failure depending upon the prerequisite time required to execute corrective and/or proactive measures to avoid failures.

The proposed methodology helps engineering managers in identifying time bound failure signatures; however, in existing failure signature based approaches, to predict failures, this time bound element is missing. The existing approaches do not ensure sufficient time prior to the failure occurrence; hence, improvement in preventive maintenance by introducing predictive maintenance is not possible. In comparison, proposed methodology helps in identification of time bound failure signatures for subsequent usage in predictive maintenance.

The proposed methodology to extract and validate time bound failure signatures² is validated using data collected from world's reputed semiconductor manufacturer. The case study is presented in Part-III, Chapter 7. The methodology itself is presented in Part-II, Chapter 6 whereas the case study results of the methodology are presented in Part-III, Chapter 9. This methodology also provides answer to following research question:

Q(3): how to extract and validate time bound failure signatures (pattern/rules) for predictive maintenance?

A(3): A methodology to extract and validate time bound failure signatures (patterns/rules) is proposed and validated using data collected from world's reputed semiconductor manufacturer.

²(Abu-Samah et al., 2015a, 2016)

3.2 Thesis Schematic

This section presents graphical schematic of the thesis (Figure 3.1) with inputs and outputs for each chapter.

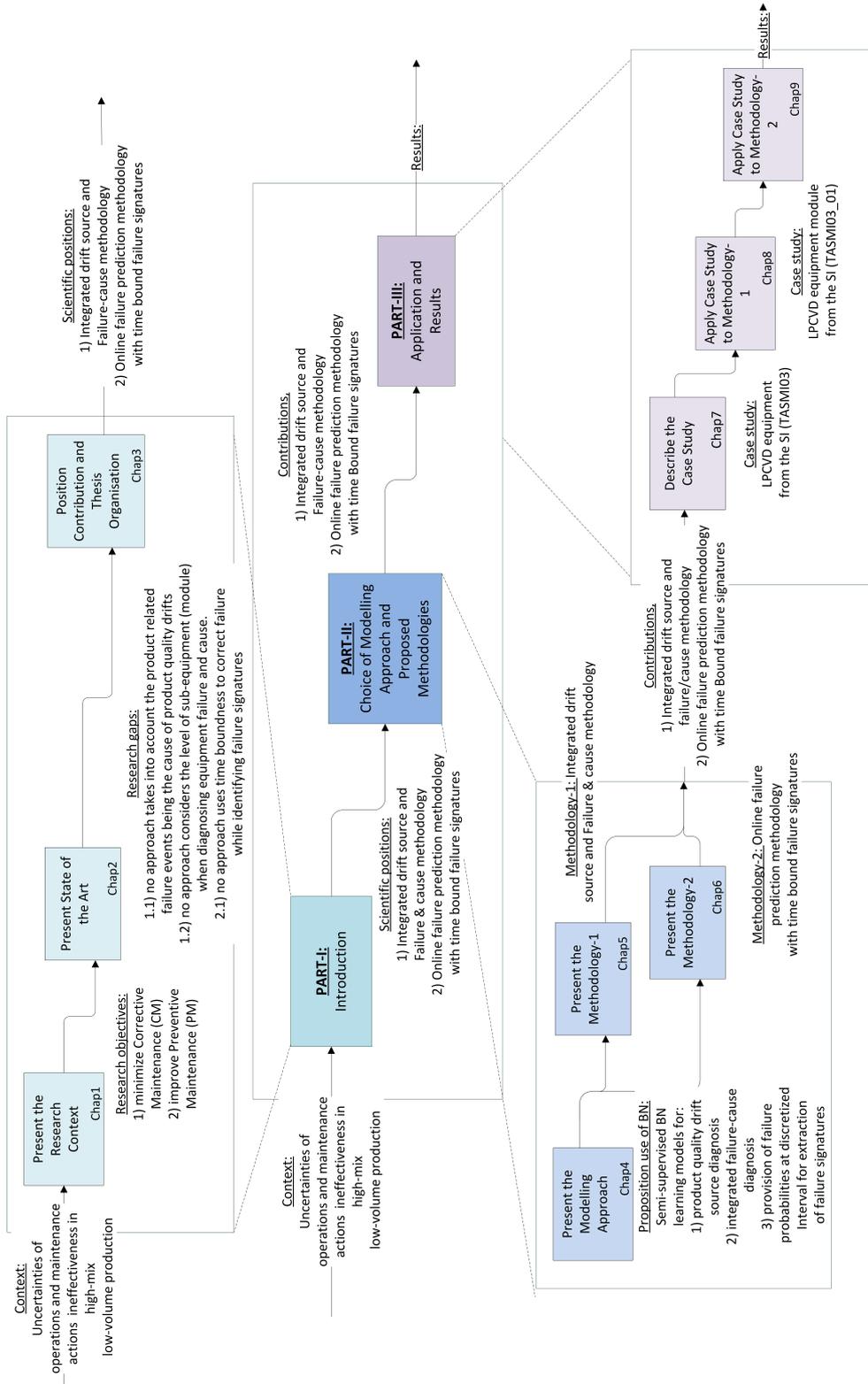


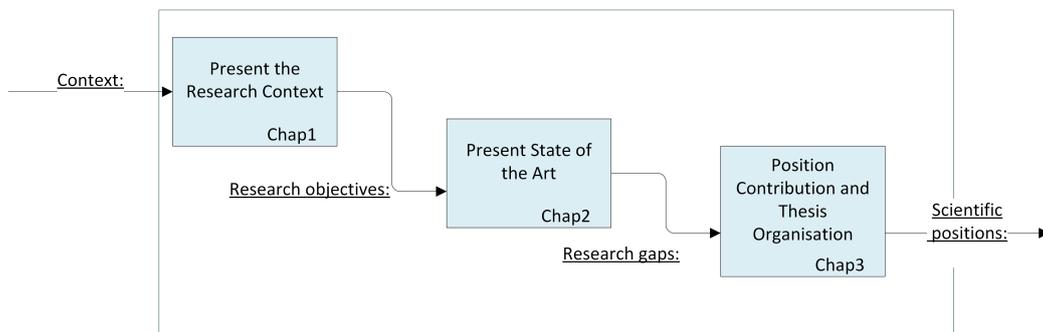
Figure 3.1: Thesis schematic.

3.3 Thesis Organization

This research thesis is organized into three Parts whereas each Part consists of 3 chapters.

3.3.1 Part-I: Introduction

The part 1 comprise of three chapters as presented below:



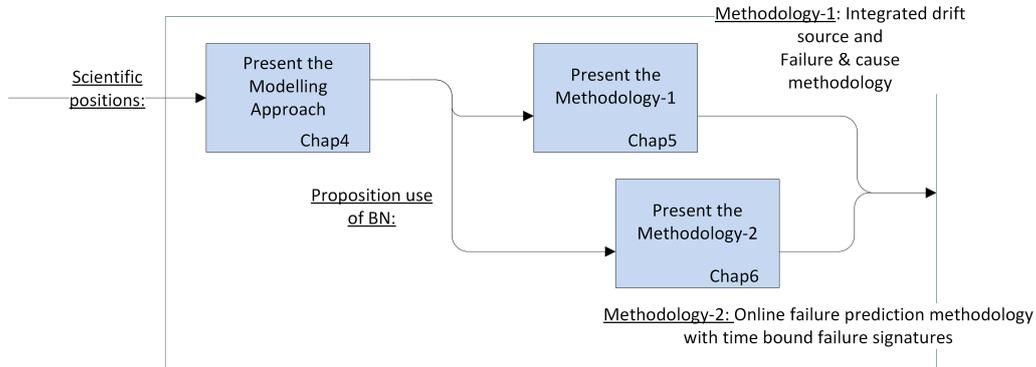
Chapter 1 presents research context, research goal and objectives and briefly summarizes the challenges faced by high-mix low-volume production settings. The research goal, objectives and challenges are input for chapter 2.

Chapter 2 briefly reviews the role of maintenance strategies and existing approaches to reduce CM and PM actions. Moreover, this chapter reviews efforts carried out in research within the framework of predictive maintenance to improve PM. The output of these chapters are scientific gaps and three research questions.

Chapter 3 presents graphical schematic of the research thesis, brief review of scientific contributions and their positioning with reference to existing approaches.

3.3.2 Part-II: Choice of Modelling Approach and Proposed Methodologies

The part-II presents choice of modelling approach and proposed methodologies.



Chapter 4 explores and justifies the use of BN as modelling approach. This follows with the brief introduction to Bayesian Network (BN), structure learning algorithms, Conditional Probability Tables (CPT) computations and inferences.

Chapter 5 presents in detail the proposed integrated 2-level methodology for product quality drift source and failure/cause diagnosis with an objective to reduce increasing unscheduled equipment breakdowns which leads to increasing CM, failure durations. The goal is to stabilize and improve production capacities in high-mix low-volume production capacities.

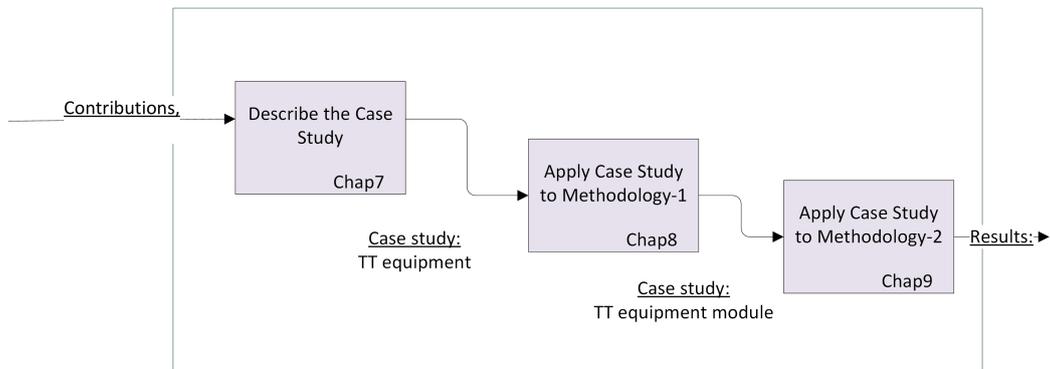
Chapter 6 presents 4-step methodology for the extraction and validation of time bound failure signatures (patterns/rules) for subsequent usage in predictive maintenance.

3.3.3 Part-III: Application and Results

The third part of this research thesis is focused on the presentation of case study and proposed methodologies results.

Chapter 7 presents a brief introduction to the Semiconductor Industry (SI) along with its characteristics. Moreover, case study and data used to validate proposed methodologies (from Part-II) is presented along with data pre-processing steps.

Chapter 8 presents the results obtained after validating the proposed methodology (Part-II, Chapter 4) using multi-source (product, process (technology), equipment, maintenance) data collected across the production line from a world reputed semiconductor manufacturer.



Chapter 9 presents the results of case study used to validate the proposed methodology (Part-II, Chapter 5) for the extraction and validation of time bound failure signatures (patterns/rules).

3.4 Conclusion

This chapter presents positions and contributions against existing approaches in literature, the graphical schematic and organization of thesis.

Part II

CHOICE OF MODELLING APPROACH AND PROPOSED METHODOLOGIES

Modelling Approach

Introduction

This chapter presents existing modelling approaches for the product quality drift source diagnosis, failure/cause identification, and online failure prediction based on failure signatures. The objective is to make a choice among existing modelling approaches to find answers to the research questions (see section 2.5). Consequently, BN is retained as the modelling approach; therefore, theoretical foundation on learning BN structure and computation of conditional probabilities for inferences is also presented. This chapter concludes with discussion and conclusion.

Contents

4.1 Existing Modelling Approaches	42
4.1.1 for Failure/Cause Diagnosis	43
4.1.2 for Extraction of Failure Signatures	44
4.2 Discussion on the Choice of Modelling Approach .	45
4.3 Bayesian Network as a Choice for Modelling In-	
 tegrated Drift Source and Failure/Cause Diagnosis	
 and Time Bound Failure Signatures	46
4.3.1 Bayesian Principle and Bayesian Network (BN)	46
4.3.2 BN for Classification	49
4.3.3 Bayesian Network Structure Learning	49
4.3.4 Estimating Conditional Probability Tables (CPT) . .	53
4.4 Conclusion	54

4.1 Existing Modelling Approaches

This section presents critical review of the existing modelling approaches on failure/causes and online failure prediction based on failure signatures. The objective is to choose appropriate modelling approach to address the scientific

gaps (see section 2.4) and find answers to the research questions (see section 2.5).

4.1.1 for Failure/Cause Diagnosis

The fault diagnosis in process control refers to the isolation task (determination of the exact fault location) as well as the identification task (of time, type and magnitude of fault) (Gertler, 1998). The methods and techniques differs depending on these tasks, but these are essentially a classification problem and are traditionally classified as the knowledge (qualitative), model (quantitative) and data driven (historical process) based approaches (Venkatasubramanian et al., 2003c,a,b). The revolutionary changes in Information and Communications Technology (ICT) and increasing data dimensionality and volumes have resulted in the emergence of data driven methods for diagnosis. These approaches range from multivariate statistical methods (Principal Component Analysis (PCA) (Chiang et al., 2000), Partial Least Squares (PLS), Discriminant Analysis (DA)) (Chiang et al., 2000, 2004), Support Vector Machine (SVM)(Chiang et al., 2004; Widodo and Yang, 2007) to black-box methods based on Artificial Neural Networks (ANN) (Samanta and Al-Balushi, 2003). These classification approaches assign abnormal observations to closely related fault/failure. Besides these statistical approaches, Verron et al. (2010); Atoui et al. (2015) proposed and used probabilistic approach, Bayesian Network (BN) as fault classifiers and isolator.

The fault/failure detection techniques are characterized by robustness (resilience to disturbances, noise, modelling errors), sensitivity (ability to distinguish between closely related different faults), detectability, detection time (faults/failures must be detected as soon as possible), and computation effort. The data driven approaches primarily need more data to make good fault/failure predictions. These approaches work well with the sensors data collected across the production line. The predictive models developed based on these approaches are impacted by sensor reliability issues. In comparison to these approaches, probabilistic methods use both sensor data and event based contextual data/information which is insensitive to the sensor reliability issues. Moreover, these approaches associate probability with the potential occurrence of an event.

Correa et al. (2009) compare artificial NN, another continuous data based classifier, with BN in the problem of product quality detection, targeting automotive and aeronautical industry. The BN technique can deal with both continuous and discrete data. Moreover, BN approach is proved to have higher classification accuracy given new sets of measured variable and a bet-

ter interpretability of resulting network. From the studies, ANN follows black box model in the sense that non-linear relationships of cause and effect are not easily interpretable. This makes it difficult to explain the underlying causal relationship behind input and output.

4.1.2 for Extraction of Failure Signatures

Similarly, the most commonly used modelling approaches for failure signature (pattern/rules) extraction are Multivariate Statistical Process Control (MSPC), probabilistic methods e.g. BN and machine learning based method e.g. Nearest Neighbourhood (NN), Support Vector Machines (SVM) and Hidden Markov model (HMM). In MSPC, abnormal process operations are detected to find the source of abnormalities. The abnormalities include faulty sensors or its measurements, product quality drift as well as equipment drift or failure. As there are hundreds of sensors that offer continuous values on the condition of equipment, enormous works to transform them into information and/or indicator have been introduced to detect the signs of abnormalities. Qin et al. (2001) for example aggregate the values from sensors to an index using Squared Prediction Error (SPE) and Hotelling's T^2 while Yue and Qin (2001) used both to produce single index. The BN is used as modelling approach by Bouaziz (2012) to compute Equipment Health Factor (EHF) in industrial context whereas this approach is demonstrated to give good results on the estimation of future equipment states based on the set of symptoms. This is an interesting approach because it provides us with probabilities associated to the future occurrence of events. The machine learning methods are data driven approaches and require large set of temporal data for the extraction of failure signatures for accurate predictions.

All of these approaches, except probabilistic methods, use temporal data or sequence of events as input where target prediction is the identification of signs of abnormality. The key limitation associated with these methods is that these approaches do not extract time bound failure signatures. We define time bound failure signature as a failure signature which guarantees to generate an alert prior to specified time of failure occurrence such that proactive measures can be taken to avoid failure occurrences. Hence, failure signature identified using existing approaches cannot ensure reduction in unscheduled equipment breakdown or improvement in preventive maintenance. Besides failure signature based approaches, RUL based approaches approximate the expected failure time of components, based on physical degradation models. These approaches do not take into account failure propagation mechanisms due to the fact that assemblies and sub-assemblies are linked in parent-child relationship. However, in comparison, failure signature based approaches

take into account this parent-child relationship but lack the prediction of failures prior to the cumulative time required to take proactive measures.

4.2 Discussion on the Choice of Modelling Approach

The BN is known to have the advantage of being relatively easy to encode prior knowledge in network form, either by fixing portions of the structure, forbidding relations, or by using prior distributions over the network parameters. Such prior knowledge can allow a system to heuristically learn accurate models from small quantity of data. Kobbacy et al. (2011) discuss the various utilities of BNs in manufacturing with emphasis on its applicability when uncertainty is the key characteristic. Weber et al. (2012) present a detailed review of BN applications in the domains of reliability, risk analysis and maintenance. For probabilistic evaluation and using event based data, comparisons can be made between BN, Fault Trees (FT) and Markov Model (MM) based models. The FT model, is limited in assessing just one top event per model as opposed to BN which can model multiple states and evaluate several outputs in the same model, a characteristic well suited for the selection of alternative actions when we have to make a decision against a problem with multiple failure modes and causes. On the other hand, MM based approaches allow the representation of multi-state variables but these are computationally expensive, even for simple problems. The system becomes complex with a large number of variables, which is the case in our context where we try to integrate different variables coming from different data sources with multiple states. With BN, the constraint is avoided since the number of parameters within the conditional probabilities table is considerably lower to MM (de Souza e Silva and Ochoa, 1992).

Additionally, MM based approach relies on the observation of sequence of events prior to decision, while a BN relies on the conditional dependencies between them regardless of the sequence of appearance. BN is defined as a probabilistic graphical model. It also carries other names such as probabilistic network. A BN is a comprehensive tool for visualization of variables and their dependencies (or independencies). It is based on both graph and probability theories. It allows to quantitatively describe the operation of a system through various probability calculations for the system variables. Moreover it offers a compact encoding of a complex distribution over a high-dimensional space using the graph based representation. The Bayesian network approach has recently become a focus for dynamic maintenance management and fail-

ure diagnosis in the SI. Yang and Lee (2012); Bouaziz et al. (2013) applied BN for diagnostics and prognostics in the SI with an objective to investigate causal relation among equipment conditions and their effects on product quality. Moreover, there exist published methods and algorithms to adapt the BN to fit to the specific case studies in the SI (Roeder et al., 2012).

In addition to their ability to represent causal relationships, BN has the ability to perform learning efficiently in uncertain environments, involving small amount of related failure data and short temporal change of states. It is based on the conditional probability theory and has a compact graphical presentation. Other advantages of using Bayesian network are its inherent ability for deduction and inter-causal reasoning (Kjærulff and Madsen, 2006). The deductive (causal) reasoning takes into account the causal links between variables, from causes to effects using dynamic detection evolution. The inter-causal reasoning is an interesting and powerful ability of BN where evidence on one possible cause disapproves other possible causes. It can also be used to represent compact joint probability distributions (Margaritis, 2003).

4.3 Bayesian Network as a Choice for Modelling Integrated Drift Source and Failure/Cause Diagnosis and Time Bound Failure Signatures

4.3.1 Bayesian Principle and Bayesian Network (BN)

Generally in BN, variables are modelled as nodes and edges are created between these variables. The edges may reflect a causal phenomenon between related variables (causal networks); however, this is not necessarily the case. Indicating an arc between two variables (a parent and a child) implies a direct dependence between them. BN can provide the state of the child variable given the states of its (if several) parents. For this, each node has a Conditional Probability Distribution (CPD), to quantify the effect of the parent node(s) on that node. For the root nodes (without parents), the probability table is not conditional; hence, prior probabilities are fixed. BN prohibits child dependencies to parents. Thus, the set of variables and edges forms a directed (edges with direction) and acyclic (no cycle) graph.

A traditional BN (Figure 4.1), $B = (G, \Theta)$ is defined by:

- A graphical (qualitative) component: a directed acyclic graph (DAG)

$G = (V, E)$, where V is the set of vertices (nodes) represents n discrete random variables $X = X_1, \dots, X_n$, and E is the set of directed arcs (edges) corresponding to conditional dependence relationships among these variables.

- A numerical (quantitative) component: presents a set of parameters $\Theta = \Theta_1, \dots, \Theta_n$ where each $\Theta_i = P(X_i | P_a(X_i))$ denotes the conditional probability distribution of each node X_i given its parents $P_a(X_i)$. These conditional probability distributions (CPD) are usually stored and organized in tables named conditional probability tables (CPTs). The latter is constructed based on the Bayes rule (Equation 4.1).

$$P(x | y) = \frac{P(y | x).P(x)}{P(y)} \quad (4.1)$$

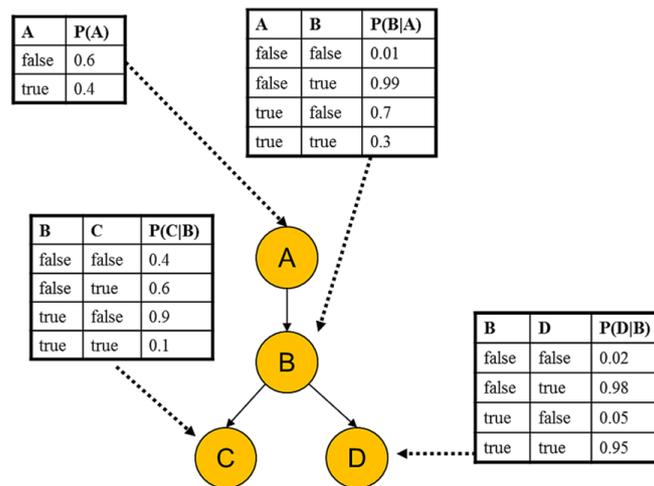


Figure 4.1: Example of BN.

The BN reasoning computes probabilistic inference. It refers to the calculation of the posterior probabilities distribution of each of the network variables given new information concerning some other variables which are set to certain states (evidence). Evidence can be 'hard' (e.g. $failure_a$ happens 100% sure) or 'soft' (e.g. $failure_a$ happens, it is 80% sure). Once the information is entered, it is propagated in the network by any chosen inference algorithms (Figure 4.2). Such a process looks for the impact of a certain information regarding some variables, on the remaining ones. The benefit of BN is that it supports omni-direction reasoning (regardless of arc directions).

For example, one can perform diagnostic reasoning, i.e. reasoning from symptoms to cause, such as when a maintenance personnel observes high frequencies of equipment failure alarms, he then updates his belief about equipment's age and whether the equipment shall be replaced. Note that this reasoning occurs in the opposite direction to the network arcs. Or again, one can perform predictive reasoning, reasoning from new information about causes to new beliefs about effects, following the directions of the network arcs. This distinction, however, only exists from the perspective of the users and based on their current needs. In terms of algorithm of inference, Figure 4.2 shows the main classification of BN Inference Algorithms. Detailed description of these algorithms and their references are provided in Guo and Hsu (2002). The probabilistic representation of a system in a graphical form with this type of interrogation allows monitoring relationships among different variables (Pearl, 2014). It is an efficient feature to model causal relationships between a set of events.

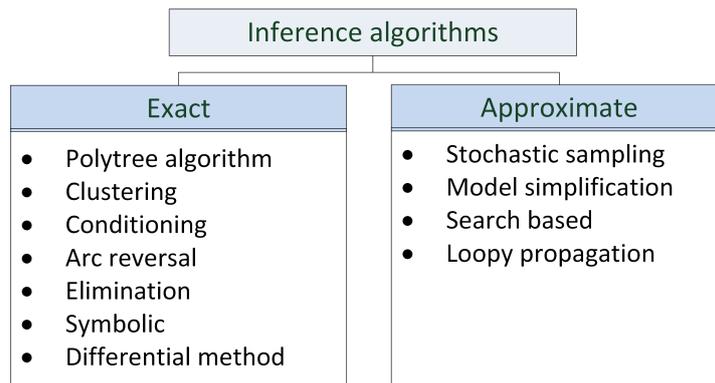


Figure 4.2: Example of Inference Algorithms.

The construction of a BN is realised following the 3 necessary tasks (Naïm et al., 2011), (i) identification of variables, (ii) definition of BN structure and (iii) definition/computation of its CPT. Moreover, BN learning from data implies structure learning and parameter estimation from complete or incomplete training data. Finally, a BN can be used in several approaches to fulfil different functions such as scoring, forecasting and classification. To fulfil the target use of BN, we focus on BN as classifier learning from complete data. The tasks (ii) and (iii) are further elaborated in the following subsection.

4.3.2 BN for Classification

In machine learning and statistics modelling, unsupervised learning refers to the knowledge discovery without any target to predict or estimate in a given large sets of data. There is no other external support other than the data that were given. While, in supervised learning, users are provided with sets of data, but are also given the target or outcome variable which is to be predicted from the given set of variables. Using these set of variables, a function is created to map the inputs to the desired outputs. The learning process continues until the model achieves a desired level of accuracy on the learning data. In this way, the machine is 'supervised' to make its decisions.

In the context of BN, unsupervised learning of structure refers to the discovery of "structural" probabilistic relationships between a large number of variables, without having to specify the input and output nodes. There are no hypotheses constraining the exploration of possible relationships between its variables. It is a practical approach for obtaining a general understanding of simultaneous relationships between many variables without prior knowledge. On the other hand, supervised learning has the same objective as many traditional modelling methods, i.e. to develop a model for predicting a target variable. If the target variable is continuous, the predicted distribution produces an expected value. For a discrete target variable, the process is called classification. The learning algorithms in the latter type of learning can focus on the characterization of the target node, rather than on a representation of the entire joint probability distribution of the learning set. With BN classification, beyond producing predictions, individual variables can be quantified and identification of cause-effect links between the variables can be established. To complete, in BN learning, it is possible to apply the hybrid approach called semi-supervised learning. It refers to the unsupervised learning algorithms that searches and discovers the relationships between the nodes that belong to a predefined distance of the target. The semi-supervised learning algorithms allows learning a network centred on the target variable. This approach is very useful for analysis that involve a lot of nodes but with a target nodes in mind, which fits our target for A(1), A(2) and A(3) (see section 3.1).

4.3.3 Bayesian Network Structure Learning

The task of finding a network structure is a combinatorial optimization problem, and is known to be NP-hard (Chickering et al., 1994), even if we restrict each node to having at most two parents. As few as 3 nodes (with two parents max) can be resulted in 25 different ways to form a network and for 5

nodes in 29281 ways. A number of literature has been produced that seeks to understand and provide methods of learning structure from data (Daly et al., 2011). The standard approach for addressing this problem is to perform heuristic search over some space of possible structures. Many methods have been proposed along these lines, varying both on the formulation of the search space, and on the mechanism used to search the space. The BN structure learning methods are typically divided into two main families. The first family tackles this issue as a constraint satisfaction problem which looks for independencies (alternatively dependencies) in the data, using statistical tests, then try to find the most suitable graphical structure with this information. Links are added or deleted according to the results of statistical tests, which identify marginal and conditional independencies. The second family, commonly known to be a better tool for learning structure from data (Friedman et al., 1999) are the search & score based methods. This second family treats learning as an optimization problem. Its search algorithms evaluate how well a structure fits to the data using a score function (e.g. K2 and Minimum Description Length (MDL) scores). So, these score-based algorithms search for the structure that maximizes this function. This score trades off network complexity against the degree of fit to the data, which is typically expressed as the likelihood of the data given the network. All search algorithms (Glover, 1986; Munteanu and Bendou, 2001; Chickering, 2002; Acid and de Campos, 2003; Teyssier and Koller, 2012) present,

1. a search space consisting of the various candidates for BN structure,
2. an effective search mechanism to move from one structure to another in the search space, comprising of
 - a mechanism to encode each of the states,
 - a mechanism to move from state to state in the search space, and
3. a scoring function to assign a score to each candidate structure in the search space.

In the context of learning the BN with clear target node, the problem is to find the most adapted algorithms to characterize the target from association and potentially causal discovery among the other chosen variables. In the course of experimenting the different type of existing algorithms in BayesiaLab 5.3¹, the combination of Equivalence Search (EQ) and Tabu al-

¹a tool distributed by Bayesia. BayesiaLab handles exact and approximate inference. It provides parameter learning using maximum likelihood estimation and a variety of structure learning algorithms. It supports missing data, prior knowledge addition, supervised, unsupervised and semi-supervised learning, etc

gorithms provides us consistently with the most accurate cross validation scores.

Many have proposed heuristic algorithms to explore a small part of the huge space of possible structures in order to find a local optimum to best suit the structure. Given resources for learning, it is recommended to employ multiple searching procedures to raise the probability of finding the nearest to the global optimum network. For example, as it has been employed in our research, a combination of EQ and Tabu. The initial structure is learned using the Equivalence Class (EQ), a heuristic algorithm to search highest scoring network explicitly across the spaces of potential BN structures that have same conditional independence relations (4.3.3). The learned structure is further optimized using Tabu (4.3.3), method that complements EQ in terms of search space and exploration strategy. Alternatively, a rather positive result has been obtained using a combination of Tabu order and Tabu algorithms.

A. Equivalence class framework The simplest formulation of search space is the set of all possible individual DAG. The intuitive way to find the best network is greedy search which starts at an initial structure in the structure space then considers all nearest neighbours of the current structure and moves to the neighbour that has the highest score. Neighbours comprise of all structures that can be generated by current structure by adding, deleting or reversing a single arc, subject to the acyclicity constraint. If no neighbours have higher score, a local maximum is reached and the algorithm stops. While the method is simple, it can be a disadvantage due to equivalence class property. Two DAGs G and G' are equivalent if for every BN,

$$B = (G, \Theta) \tag{4.2}$$

there exists a Bayesian network

$$B = (G', \Theta') \tag{4.3}$$

such that B and B' define the same probability distribution, thus the same score. This type of search can waste time re-scoring the same equivalence class and in many cases, in order for the algorithm to move from one equivalence class to another, it will have to make numerous moves within the same equivalence class. Furthermore, in large networks we can anticipate accumulation of early stage wrong decisions thus end up with a final network very different from the ideal one. In order to overcome these difficulties, we can realize the search in the space of equivalence classes.

Figure 4.3 illustrates the search strategy. This approach consists in allowing the addition of undirected edges, transforming a DAG into PDAG (Partially DAG) and complete PDAG (CPDAG). Edge orientation is delayed until the interactions between edges make possible the choice of a direction on the basis of the score. As the obtained partially directed graphs may be interpreted as equivalence classes, this solution modifies the search space. Consequently, the search algorithm explores the space of equivalence classes of BNs instead of the space of BN DAGs. When all EQ is explored, a network (DAG) with the best score is chosen as the final structure.

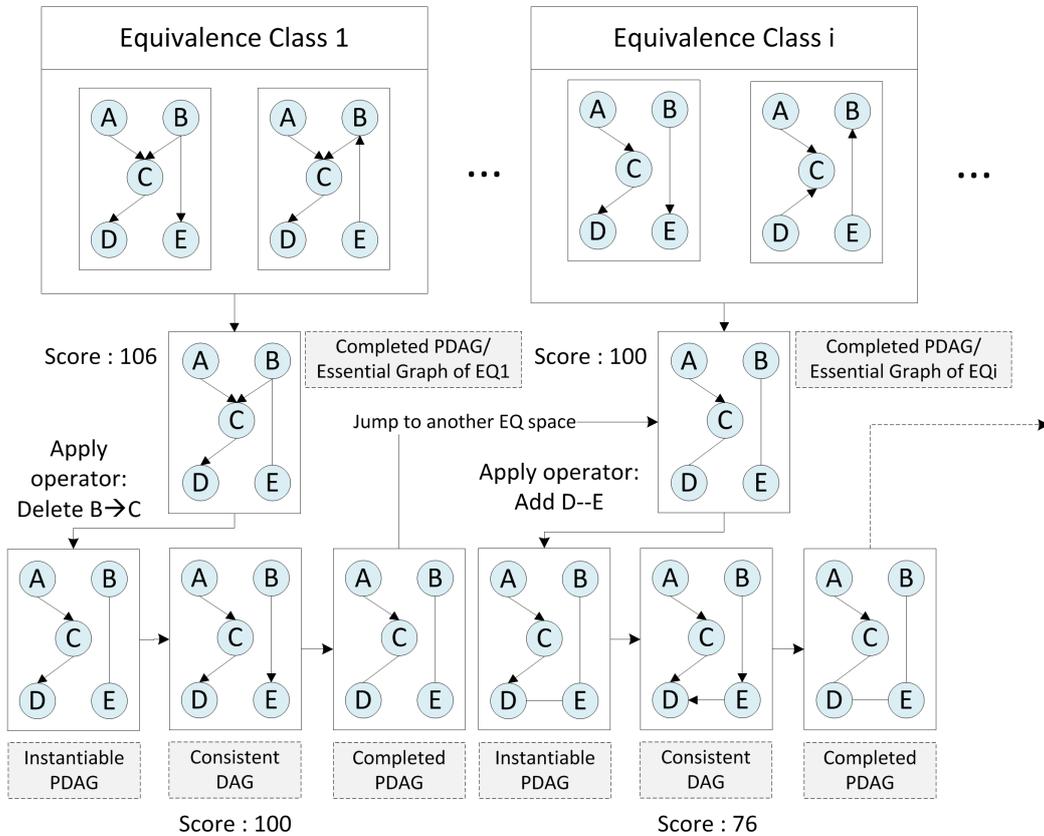


Figure 4.3: Illustration of EQ search strategy.

B. Tabu framework The Tabu is an extended form of greedy search algorithm that tries to escape from a local maximum (in the search space of all DAG) by selecting the solution that minimally decrease the value of scoring function. Immediate re-selection of local maximum, just visited, is prevented by maintaining a list of solutions (of predefined precise size) that are forbidden a.k.a. the Tabu list (Glover, 1986; Acid and de Campos, 2003). Figure

4.4 illustrates this strategy and its Tabu listed networks in dotted boxes. The search operators involved in transforming one DAG to another are addition, suppression and reversal. When sufficient changes occur but without an increase in the minimum score ever encountered during search, the algorithm terminates and the overall best scoring structure is then returned. This strategy typically requires random restarts to find the optimized solution, but using EQ as the initial network largely reduce the restart number as well as the size of necessary Tabu list. In complementary point of view to EQ method, the Tabu approach offers the exploration of solutions that might not be considered as consistent in EQ's PDAG \rightarrow DAG transformation.

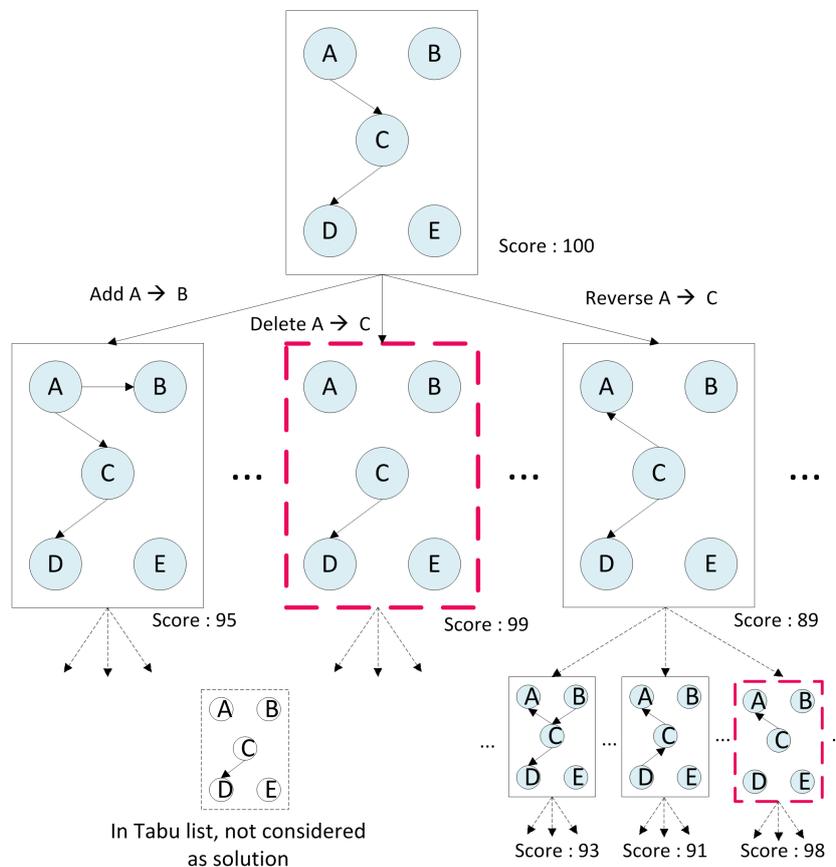


Figure 4.4: Illustration of Tabu list search strategy.

4.3.4 Estimating Conditional Probability Tables (CPT)

Once the structure is identified, the parameters are estimated to compute CPT and finalize the BN. In the case of Maximum Likelihood (ML) learning,

the parameters could be the actual probabilities in the conditional probability table attached to each node. This is a statistical estimation of the probability of an event (positive counts) from the total count of occurrences without the event in the data set. This is a fairly simple and widely used method (Heckerman, 2008) for cases with defined structure and with complete observation. The probability estimation is given by the following formula,

$$\hat{P}(X_i = x_k \mid \text{parents}(X_i) = x_j) = \frac{N_{i,j,k}}{\sum_{n=1} N_{i,j,k}} \quad (4.4)$$

$N_{i,j,k}$ is the number of positive counts where the variable X_i takes the value x_k and X_i 's parents take the value x_j .

The combination of search algorithms (EQ and Tabu) involving two types of search space and search exploration produces a final structure with the lowest score Minimum Description Length (MDL) (For elaboration refers Appendix A) and is accepted for further analysis. All BN models are learned and tested using the 10-fold cross validation strategy.

4.4 Conclusion

From the above discussion on existing modelling approaches used for failure/cause diagnosis and failure predictions, it is concluded that these are data driven approaches. Primarily, these approaches highly depend on the quality of sensor signals. Moreover, limited availability of contextual data/information also impacts the quality of decisions made. This can be improved by discretizing continuous (sensor signals) data; however, this would result in the aggregation of sensor signal bias into contextual data/information, leading to biased decisions. Therefore, in this thesis, besides the availability of sensor signals data, we rely only on the contextual data/information collected and stored across the production line databases. This reduces our choice of the modelling approach as we do not have temporal data. Moreover, during failure/cause diagnosis and failure prediction, our objective is not to automate the root cause analysis but to help maintenance teams in effective failure/cause diagnosis and predictions. Inherently, uncertainty is also associated with the occurrence of failures as well as failure/cause diagnosis due to the complexity of production line operations (high-mix low-volume) and multi purpose fully automated production equipment. Therefore, for both failure/cause diagnosis and failure prediction, we retain Bayesian network (BN) as our principal modelling approach.

Other modelling approaches are preferred by researcher due to the fact that they need large temporal data sets and existing revolution in ICT have

resulted in producing and archiving such large data sets. In BN, the nodes are connected to one another and represent either predictor or target variables whereas linkages (edges) between nodes formulates its structure. Moreover, it uses discrete contextual data/information and provide inference with probabilities based on the varying relationship between predictors. The BN is also capable of learning variance relationship between predictors (structure) with less observations. However, only drawback associated with this modelling approach is its limitation in number of predictors to be used for learning the BN structure and computing respective Conditional Probability Tables (CPTs). In this thesis, predictors are not selected by ranking them using statistical approaches like mutual information, entropy etc. however, these are selected through brainstorming sessions. The predictors are associated to the classes of variables and structure learning rules are restricted from one class to another. This is referred as semi-supervised learning of the BN structure. Consequently, structure learning is no more NP-hard as well as more predictors can be added to the BN which leads to less computation time for the CPTs.

Integrated Drift Source and Failure/Cause Methodology

Introduction

In this chapter, 2-level integrated methodology for product quality drift source and failure/cause diagnosis is presented. The proposed methodology also addresses first two research questions (see chapter 2, section 2.6) which focus on reducing misdiagnosis and identifying the best modelling level as either equipment or modules.

Contents

5.1	Key Elements for Modelling Equipment Failure . . .	56
5.2	The Methodology	59
5.2.1	Level-1: Product Quality Drift Source Diagnosis . . .	60
5.2.2	Level-2: Equipment Failure and Causes Diagnosis . . .	62
5.3	Discussion	64

5.1 Key Elements for Modelling Equipment Failure

The key elements for modelling product and equipment failures/causes in the SI are identified as linkages of product quality failures to equipment failure, equipment level (equipment, module, component) failures, failure modes, failure causes and root causes. In chapter 2, section 2.2.1 and 2.2.2, it is concluded from literature that equipment is blamed to be the source of product quality drifts as well as failures are modelled either at equipment or component levels. In the SI, equipment is composed of modules and sub-modules, linked and modelled in parent child relationship. The change in the state (failure) of parent or child has direct impact on the other. In addition to this, equipment in the SI can be of single, batch or multi chamber type. Therefore,

while modelling product quality failure source and equipment failure/cause diagnosis, these particularities must be taken into account.

The equipment failure events can be extracted from equipment database and have multiple operating modes (Mendez et al., 2005). The Figure 5.1 presents existing classification of equipment states in the SI¹. For modelling purpose, we selected five states (hold, online, fail, hold_spc, preventive maintenance) which are further aggregated into two high level states as up and down as under:

- normal (up/production) operation: This represents the state where equipment is functional to manufacture products on given specifications.
- abnormal (down/breakdown) operation: This represents the state of the equipment where it is no more functional to produce products in accordance with specifications due to faults/failure(s).

ACTIVITY TYPE	EQUIPMENT STATE	CATEGORY
ENGINEERING	Engineering time	UP
HOLD	Standby time	
ONLINE	Productive time	
FAIL	Unscheduled time	DOWN
HOLD_SPC		
PM	Scheduled time	
SETUP		
QUALIFICATION		
OFFLINE	Non scheduled time	OFFLINE

Figure 5.1: Equipment states classification in the SI.

Consequently, several physical failure mechanisms exist for each abnormal state. A failure mechanism denotes the relation of faults and system states to a failure with focus on how the faults lead to the failure. This is closely related to the term Failure Mode (FM) as defined by Stamatis (2003). We

¹Retrieved from: <http://www.semi.org/en/semi-e10-specification-equipment-reliability-availability-and-maintainability>

argue that it is important to model not only failure events, but also the failure modes inside an equipment.

The definition of failure cause can be traced to the NF EN 13306 standard², as the reason which leads to failure. The reasons may be the result of one or more of the following: design failure, manufacturing failure, installation failure, misuse failure, mishandling failure, maintenance related failure. These causes are classified as related to the different stages in equipment functional life cycle or the development phase from which any fault is seen as been originated. The failure causes are also classified sometimes as design faults, originate from bad system design as opposed to runtime faults which occur during the production phase of a system. Another classification can be found in the Ishikawa diagram (Ishikawa, 1990) or the cause and effect diagram, developed in 1943 by Japanese scientist Kaoru Ishikawa. This diagram is used to identify all potential causes (or events) that may result in a specified event. It is not only applicable to equipment failures, but also for process operation and product quality drift diagnosis. The key (5M) categories which are proposed are manpower, methods, materials, machinery (equipment), and milieu (environment). These categories however should be selected to fit actual application. Several observable information and events such as the defect in equipment parts, process recipe management or usage conditions which are the (potential) underlying causes (or root-causes) behind failure of equipment or which initiate a process that leads to failure. These data are referred as potential root cause(s) whereas cause is the first level cause detected upon the occurrence of failure from historical knowledge. Below are the 4 sources of predictors (or potential root-causes) chosen for our work with their respective potential predictors,

- *Process*: Abnormal changes in process conditions such as the influence of production recipes are normally registered by sensors. If not identified and corrected, these abnormal conditions can enable events causing an observed failure. Recipes implemented in production such as the SI are qualified and have near zero chance to become the primary source of failure. Nevertheless, due to the frequent change of recipes and the diversity of operations in specific equipment, we argue that information on its combination are of interest for root cause analysis under uncertainty and for the purpose of decision support on corrective actions. Other potential variables to be observed are the combination of process steps, the amount of engineering and R&D process executed in an equipment.

²NF EN 13306, retrieved on 3rd May 2016 from: <http://maint.t.i.b.free.fr/Files/Other/NF%20EN%2013306.pdf>

- *Product*: Similar to the process information, we propose to study the influence of different product combinations processed in the same equipment to the equipment behaviour. Other potential variables include the redo work on products, the number of rework.
- *Equipment*: A lot of data is available on the maintenance work order and work request. Analysis have shown that only a fraction of work request records gives a notice of upcoming failures. Potential variables include gradual build-up on process chamber, machine ageing, cleaning, sensor drift.
- *Maintenance*: The contextual data on maintenance operations and the generated maintenance indicators. The data can be reported by humans working on the maintenance interventions.

An example of the position of the different elements presented above in failure analysis is illustrated in Figure 5.2

Root causes	Failure cause	Failure mode	Failure type	Equipment state	Equipment level
	No failure cause	No failure mode	ELEVATOR _BOAT_ROTATION	CHILD_TOOL _FAIL	(TASMI03)
-Process based -Product based -Equipment based -Maintenance based	Reactor tube elevator boat pickup	BEARING	ELEVATOR _BOAT_ROTATION	FAIL	Module/sub- equipment level (TASMI03_1)

Figure 5.2: Relationship between equipment states at parent and child level.

5.2 The Methodology

The proposed methodology uses Bayesian Network as modelling approach (see chapter 4, section) and models increasing unscheduled equipment breakdowns at two levels. The first level is to determine the true product quality drift failure source to stop equipment only when it is identified as the true cause and the second level to diagnose potential failures (its localisation inside the equipment) and failure modes (FMs) along with causes. These two levels are presented in the sub-sections below:

5.2.1 Level-1: Product Quality Drift Source Diagnosis

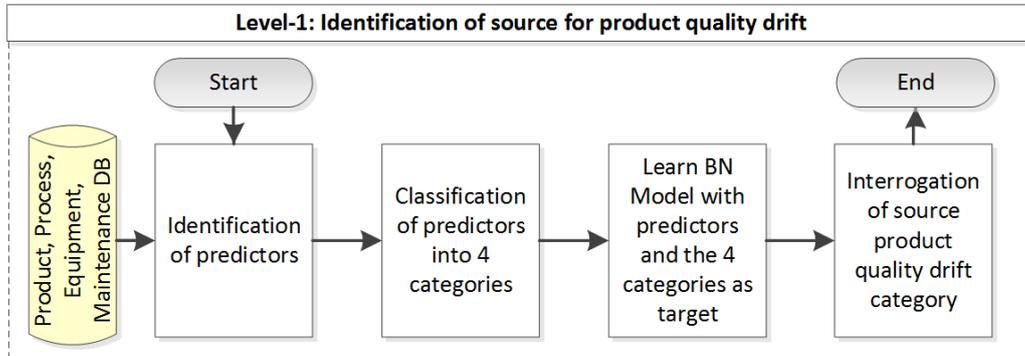


Figure 5.3: Sub-steps in the identification of product quality drift source.

Step 1.1: Identification of predictors for BN model The *predictors* (variables to be used as nodes in potential BN model) can be identified either through experts' knowledge or statistical or probabilistic approaches. This requires multidisciplinary expertise from product, process, equipment and maintenance domains. Therefore, this sub-step is completed using a task force comprising of experts from each domain through brain storming sessions. This collection of predictors is realized primarily through information gathering sessions, commonly known as brainstorming.

The predictors might be modelled as having only the binary states (e.g. normal vs abnormal), but BN can model the existing values as multiple states without jeopardizing much the complexity and precision of the model. In latter case where variables have many states with sparse occurrence of values, the states and values are proposed to be selected using the Pareto, detected by number and by cumulative time of occurrence (Hohmann, 2011).

Step 1.2: Classification of source of product quality drift into categories The objective of this level is to predict (identify) the real source of product quality drift knowing the predictors; hence, source classification will be referred henceforth as Failure Source (FS). The FS states as target nodes are identified as product, process, equipment and maintenance.

Step 1.3: Train BN model In this sub-step, data comprising of predictors and target nodes are used to learn the BN structure and respective Conditional Probability Tables (CPT). The BN structure is learned in semi

supervised fashion. The structure is illustrated in Figure 5.4. The combination of EQ and Tabu is proposed as the most adapted learning algorithms.

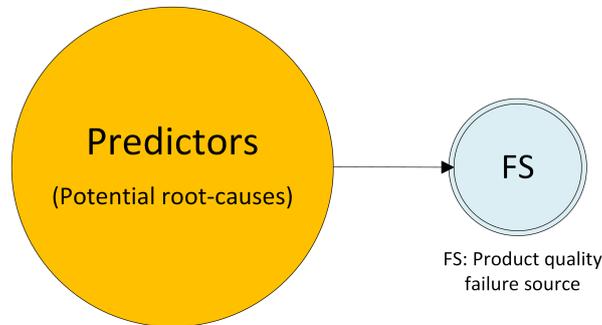


Figure 5.4: Bayesian Network structure for product quality drift diagnosis.

Knowing the predictors of the model and its graphical structure, the following sub-step in BN creation is to compute joint and conditional probability tables associated with different network predictors.

Step 1.4: Interrogation of source product quality drift category

Finally, the equipment stoppage decision is modelled as a function of FS and predictors. The resulting BN for the source identification serves as first decision support against the unscheduled equipment breakdowns. As an output to the model, probability to each FS is obtained and the state with highest probability is chosen for further investigation. It will help engineers to make more accurate decision on the equipment stoppage when product drift occurs. This serves as a defence against bad equipment stoppages. Figure 5.5 shows an example of reasoning results. As product has the highest value of probability, the equipment shall not be stopped.

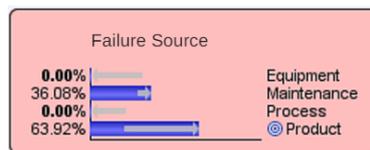


Figure 5.5: Example of reasoning and its use in Bayesian Network.

However, if equipment is identified as the source, further analysis needs to be done at the equipment level. The same set of predictors are retained to explore the localisation of the failed system inside equipment and its failure mode. This will be further explained in the next subsection.

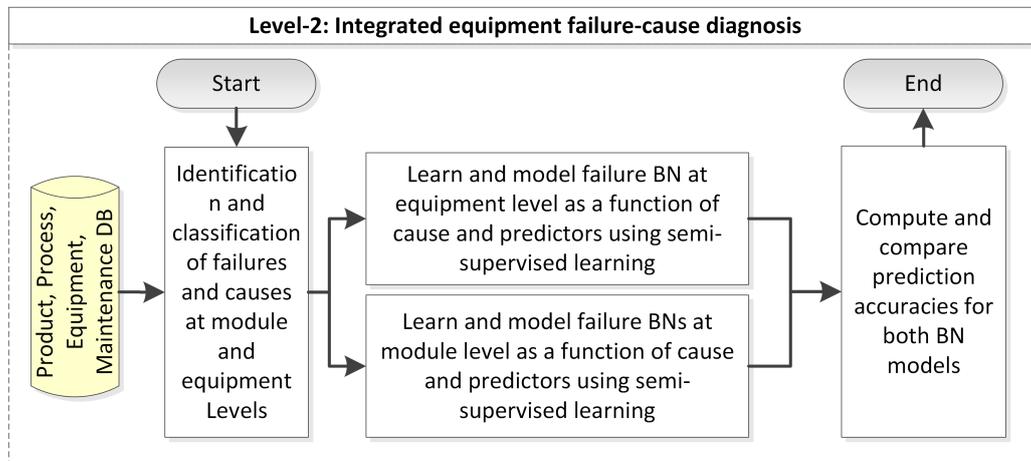


Figure 5.6: Associated sub-steps for the integrated equipment failure cause diagnosis.

5.2.2 Level-2: Equipment Failure and Causes Diagnosis

The second level of diagnosis advocates modelling the equipment localisation of failure, determination of failure mode and identification of most probable causes as a function of predictors at the module and equipment levels. The objectives are to find the model that gives accurate using event based data. The literature (see chapter 2, subsection 2.2.2) concludes that at present, failure/cause diagnosis is modelled either at the equipment or component level. However, in the last sub-step we model the failure/cause BN models at equipment and module level followed by comparison of performance. This will provide us answer to our research question 2 (see chapter 2, section 2.5).

Step 2.1: Identification and classification of failures, failure modes and causes The predictors identified for product quality drift source are retained and used for building equipment failure (localize where?), Failure Mode (FM) (which failure?) and causes (why?) BN model. The failures and failure modes will be our target node in the failure/cause diagnosis BN model. The first sub-step is to identify the existing failures, failure modes and their respective causes. This can be done, in a way using the FMECA analysis performed on respective equipment and which are recorded as well in the historical databases. The failure refers to the event on a block of system inside the equipment that renders equipment in a non functional state to produce products as per specifications, and FM defines the manner in which

the system (potentially) fails. The (potential) failure mode may also be the cause of a (potential) failure in a higher level system, or be the effect of a (potential) failure in a lower level component.

Step 2.2: Learn and model failure BNs at module vs. equipment level The objective in this sub-step is to find the modelling level (equipment or module) which gives more accurate predictions on failures, failure modes and causes. The conceptual framework for learning failure/cause BN structure is presented in Figure 5.7. The failure (localisation) is set as the target node, and the link from failure to FM is forced to imply the causality link.

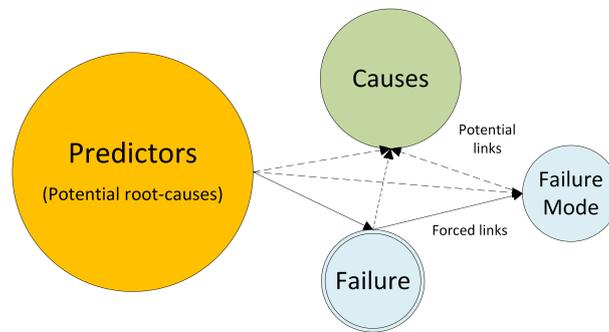


Figure 5.7: Conceptual framework of BN structure for integrated failure and cause diagnosis.

Step 2.3: Computation and comparison of prediction accuracies The BN models serve as second decision support against the unscheduled equipment breakdowns. The next sub-step is to compare the prediction accuracies for both level BN models. To evaluate BN model’s accuracy, two metrics (precision and reliability) are employed in this thesis. Given an example of following occurrence confusion matrix (Table 5.1), the definition of precision and reliability are:

Table 5.1: Occurrence confusion matrix example.

Value (Real value/prediction)	A (730)	B (27)	C (242)
A (723)	703	7	13
B (25)	5	20	0
C (251)	22	0	229

- **Precision:** the ratio of predicted positive cases to total number of the total actual cases (column),
- **Reliability:** the ratio of predicted positive cases to total number of prediction cases (line).

From the above example, precision and reliability matrices are computed and presented in Table 5.2 and 5.3.

Table 5.2: Precision matrix example. Table 5.3: Reliability matrix example.

Precision	A (730)	B (27)	C (242)
A (723)	96.3%	25.92%	5.37%
B (25)	0.68%	94.07%	0%
C (251)	3.01%	0%	94.62%

Reliability	A (730)	B (27)	C (242)
A (723)	97.23%	0.96%	1.79%
B (25)	20.0%	80.0%	0%
C(251)	8.76%	0%	91.23%

To validate the BN models, we worked in particular with the 10-fold cross validation strategy where the historical dataset is randomly partitioned into 10 equal sized subsamples. The validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the test data. For each fold, both reliability and precision matrices are computed for both BN models (equipment and module level). The results from the 10 folds are then averaged to produce averaged reliability and precision. Finally, evaluation and decision are made by comparing the accuracy for equipment BN (Equation 5.1 and module BN (Equation 5.2). The BN with higher accuracy will be considered for future modelling of failures.

$$Accuracy_{eq} = \frac{\overline{Reliability}_{eq} + \overline{Precision}_{eq}}{2} \quad (5.1)$$

$$Accuracy_{mod} = \frac{\sum_{n=1}^{Nb_{modules}} \overline{Reliability}_{mod_n} + \sum_{n=1}^{Nb_{modules}} \overline{Precision}_{mod_n}}{Nb_{modules} \times 2} \quad (5.2)$$

5.3 Discussion

In this chapter, we have presented two level integrated product quality drift source and failure/cause diagnosis methodology which is based on BN approach. The proposed methodology offers two levels of defence against increasing unscheduled equipment breakdowns. At the first level, Bayesian Network (BN) model is proposed to address misdiagnosis in product quality

drift (failure) source because we argue that equipment must be stopped only if equipment is the source of drift. The second level of defence proposes to develop an integrated failure/cause diagnosis BN model with an objective to reduce failure and diagnosis durations which will ultimately stabilize and improve production capacities. The objective of this methodology is to provide answer to first two research questions (see chapter 2, section 2.5). The validation and answers to these research questions are obtained through a case study results presented in chapter 7 and 8.

Methodology for the Extraction and Validation of Time Bound Failure Signatures

Introduction

In this chapter, we present 4-step methodology for the extraction and validation of time bound failure signatures. The objective of this methodology is to provide answer to research question 3 (see chapter 2, section 2.5) to identify time bound failure signatures for the improvement of preventive maintenance. The goal is to stabilize and improve production capacities due to increasing unscheduled equipment breakdowns in high-mix low-volume production settings.

Contents

6.1	The methodology	66
6.1.1	Step-1: Dataset Preparation	68
6.1.2	Step-2: Equipment Failure Probability Plots	70
6.1.3	Step-3: Extraction of Time Bound Failure Signatures (Patterns/Rules)	72
6.1.4	Step-4: Validate Rules and Compute Predictability Index	73
6.2	Discussion	74

6.1 The methodology

The proposed 4-step methodology for the extraction and validation of time bound failure signatures (patterns/rules) is presented in Figure 6.1. The first two steps consist of building a BN to generate and plot equipment failure probabilities at irregular intervals (data/time) and the final two steps focus

on the failure probability plots for patterns/rules extraction and their validation such that these failure signatures ensure execution of proactive actions to avoid failures occurrence. The 4-steps of the methodology are presented and elaborated below:

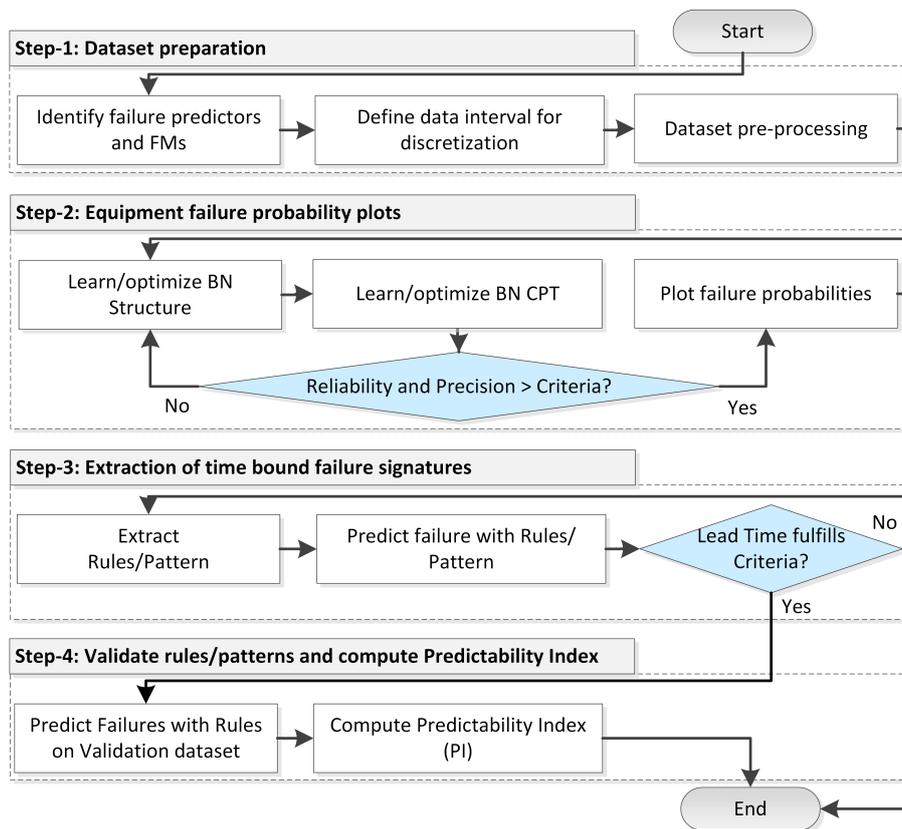


Figure 6.1: The methodology.

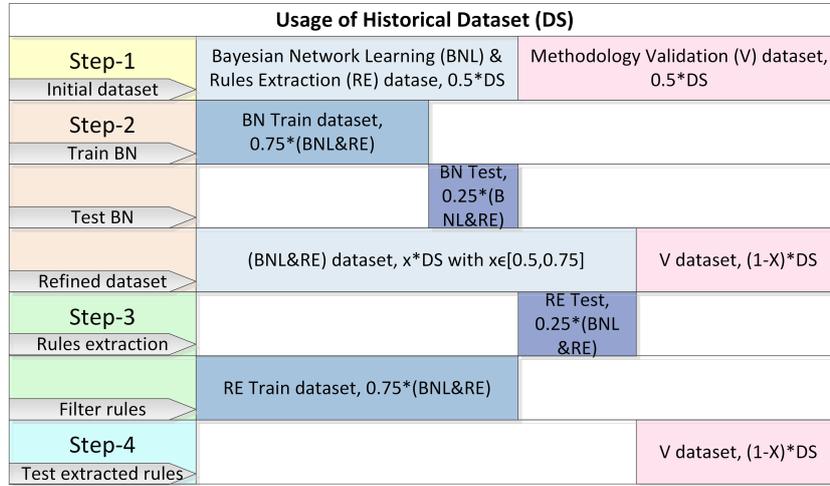


Figure 6.3: Validation scheme using historical data.

Step 1.3: Data pre-processing In this sub-step, we aggregate data collected on identified predictors from multiple databases into a single dataset. The data are sorted temporally using the definition of data intervals, with FM as the primary key. The data must be synchronized and structured in the form of Table 6.1. The first two columns however are not used to build the BN.

Table 6.1: Preparation of data for BN training.

Date-time	State duration	Failure Mode	P1	P2	...	Pn
08/03/2014 05:13	04:23:00	Failure.a				
08/03/2014 08:19	39:45:00	No.Failure				
08/03/2014 09:42	10:11:00	Failure.b				
...						

It is necessary to divide the historical dataset to validate different steps of the methodology. The division is initiated with two parts as BN Learning with Rules Extraction (BNL&RE) and Validation (V). Specifically, in this thesis, the notion of test and validation are distinguished. The test refers to cross validation task using BNL&RE sample data while validation is the final step of the methodology. The two parts of data are initially divided evenly as 50:50. The rest of division can be illustrated using the validation schema, proposed and presented in Figure 6.3.

6.1.2 Step-2: Equipment Failure Probability Plots

The BN structure proposed in the methodology is similar to previous chapter's (chapter 5 module level BN model; however, it needs to be slightly modified. The first change is the omission of causes nodes whereas second change is the inclusion of *no_failure* state in the FM. Finally, Failure and FM are modelled as the target nodes (Figure 6.4), however, failure probability will be computed upon the FM as the observed node. As a result, the modified BN can be used in real time not for failure/cause diagnosis purposes, but as failure probability plots for the purpose of time bound failure signature (patterns/rules) identification.

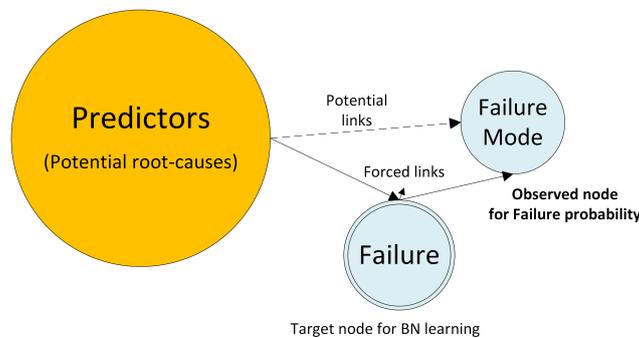


Figure 6.4: Conceptual framework for module level BN construction to plot failure probabilities.

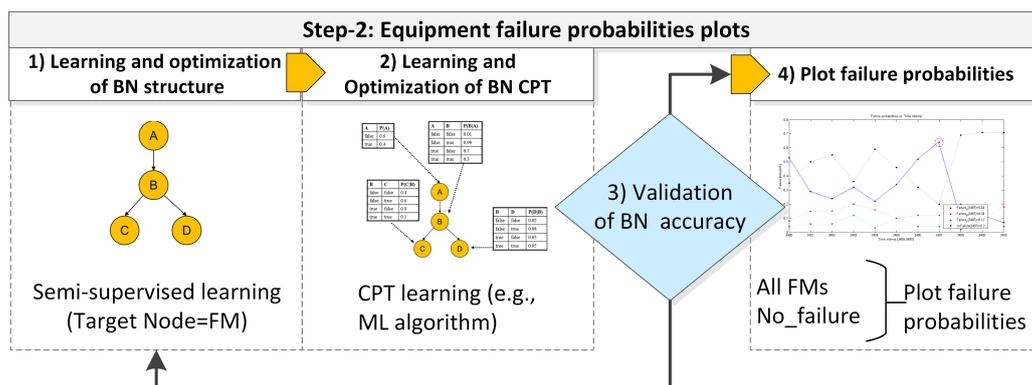


Figure 6.5: Summary of Step-2.

This step deals with the construction of BN model to plot failure probabilities for its subsequent usage to extract time bound failure signatures.

This step consists of 3 sub-steps as presented below:

Step 2.1: Learning and optimization of BN Structure The first task is focused on learning and optimizing BN structure. The structure of BN can be obtained either through experts knowledge or learned from the data using any heuristic algorithm deems fit. In this methodology, it is proposed to learn the BN structure from data using score-based semi-supervised learning algorithm that uses Minimum Description Length (MDL) as an objective function for its advantage of trade-off between data fit and model complexity (Lam and Bacchus, 1994). The search and score algorithms for structure learning are proposed as EQ (from scratch) with Tabu (for optimization).

Step 2.2: Learning and optimization of BN CPT Given the graphical structure, if the assessment of the Conditional Probability Tables (CPTs) are not done in parallel with chosen score-based algorithms, the next sub-step is to compute CPT associated with different network nodes. Like the training for structure, the probability distributions are entirely determined by automatic learning from data. In our context we used the Maximum Likelihood (ML).

Step 2.3: Validation of accuracy and precision The BN model validation is based on accuracy (reliability and precision) criteria which are set and provided by the end user. If it is not fulfilled, the initial BN structure is to be further optimized using other learning algorithm(s) and/or combination of algorithms. This is because of the fact that heuristic search algorithms might explore only a small part of a huge search space in order to find a local optimum. It does not guarantee to find the global optimum. Use of different search spaces or search strategies or both can lead to an optimized structure. The non-compliance to the user defined criteria using selected algorithms may require increasing and adjusting the ratio of BNL&RE and V dataset. This task follows recursive relearning and optimization of BN structure until user defined criteria is met or $Size(V) \leq 0.25 * Size(Complete\ Dataset)$.

Step 2.4: Plot Failure Probabilities In this sub-step, failure probabilities using BN model are computed with aggregated BNL&RE dataset as $P(FM_i | evidence_{on\ all\ predictors})$. The probabilities associated to each FM_i is plotted upon testing dataset on the discretized data intervals, and for each failure separately. These graphs are the input to next step.

6.1.3 Step-3: Extraction of Time Bound Failure Signatures (Patterns/Rules)

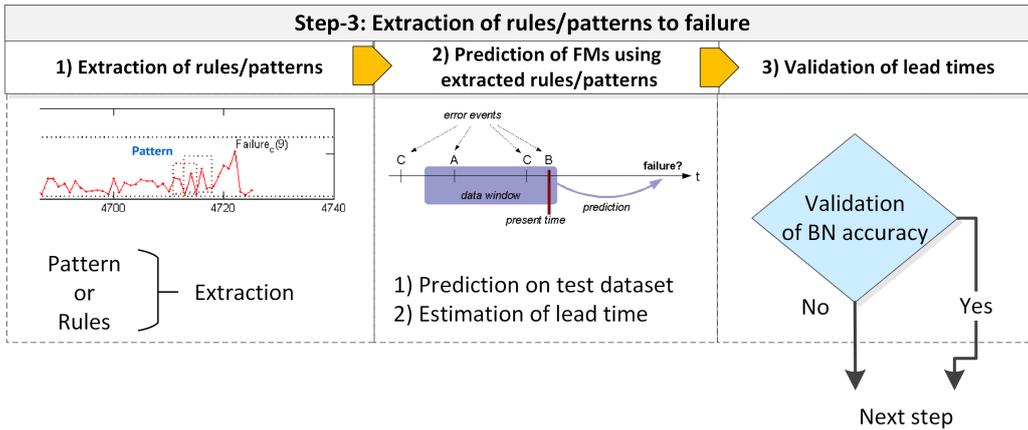


Figure 6.6: Summary of Step-3.

At this stage of the proposed methodology, we are equipped with the probability graphs for all FMs over discretized data intervals.

Step 3.1: Extraction of patterns/rules These probability plots from the BNL&RE testing set are analysed to extract patterns/rules for all type of FMs separately. Patterns are the specific rules where obvious regularities of failure probability values are observed. We make the assumption that if probability of a failure type at a given interval is superior to a certain level and to all other types of failure, including the *no_failure*, then the associated failure is attributed as the source of equipment failure. Moreover, occurrence of a failure can also be attributed depending on the consistency in failure probability plots. This is based on the fact that dependencies among the chosen predictors exist and are translated into conditional probabilities of target failures accurately. We search to identify pattern in the plots, but in case of no pattern, predictive boundaries and number of sufficient consecutive points are computed as the basis to construct rule(s) for prediction. These boundaries correspond to the lower and upper probability limits such as minimum, maximum of observed probabilities. The identified patterns may also use the predictive boundaries to refine the results.

Step 3.2: Predict failure using extracted patterns/rules The previous sub-step resumes with the prediction of failures using identified pat-

terns/rules on the training BNL&RE dataset and the estimation of lead time (the time interval from the prediction to the failure occurrence, Δtl) for each prediction.

Step 3.3: Validation of lead times to fulfil predefined criteria To be selective among generated patterns/rules, the last sub-step is the validation of average lead time of a pattern/rule falling within the acceptance interval of criteria (time bound) defined by the users. Example of time bound criteria is the warning time (the minimum lead-time such that failure prediction is of any use, Δtw). Then the patterns/rules are preserved for the next step. This approach offers flexibility over existing approaches such that the patterns/rules are extracted to predict associated failures with Δtw as defined by experts. These Δtw allow sufficient time to execute corrective/proactive measures for failure avoidance as a proactive approach.

6.1.4 Step-4: Validate Rules and Compute Predictability Index

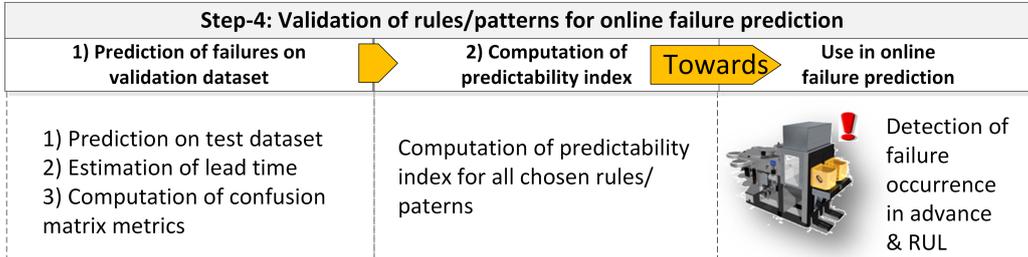


Figure 6.7: Summary of Step-4.

The validation of chosen patterns/rules on validation dataset is the final step. It is divided into 2 sub-steps.

Step 4.1: Prediction of failures on validation dataset The first task is to perform the prediction of FM using selected patterns/rules from the Step-3, but on the validation dataset. The metrics from confusion matrix and the lead time are also computed for these predictions.

Step 4.2: Computation of predictability index In this final sub-step, predictability index (PI) is computed for all chosen patterns/rules as the

average of prediction accuracy (Equation 6.1), prediction precision (Equation 6.1) and percentage lead time (Equation 6.3)¹. Prediction accuracy and precision are the most commonly used metrics for off-line evaluation in the domain of diagnostics and prognostics (Saxena et al., 2010), while lead time percentage is the additional predictive metric required in our methodology to score the prediction horizon through quantification of lead time gain and loss. The qualifying patterns/rules are then used in subsequent real time failure prediction.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (6.1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6.2)$$

$$\text{Lead time \%} = \frac{\Delta tl_{TP}}{\Delta tl_{TP} + \Delta tl_{FP}} \quad (6.3)$$

6.2 Discussion

This chapter presents methodology for time bound failure signature extraction and validation. The proposed methodology empowers engineers in selecting flexible failure signatures (patterns/rules) such that the failure prediction ensures sufficient time to execute corrective/proactive measures with an objective to avoid failure. The concept of predictability index (PI) is also proposed which represent failure warning lead time. The maintenance teams can use this methodology to extract failure signature for a given lead time. This flexibility makes the proposed time bound failure signature extraction methodology unique in its contribution.

¹TP=True Positive, TN=True Negative, FP=False Positive and FN=False Negative

Part III

**APPLICATION AND
RESULTS**

Case Study Description

Introduction

This research thesis is carried out in the 'INTEGRATE' European project in collaboration with STMicroelectronics, Crolles. In this chapter, we present introduction to the Semiconductor Industry (SI), multi source data collected across the production line, and pre-processing for subsequent usage and validation of proposed methodologies in chapters 5 and 6. A brief description of 'INTEGRATE' project can be referred in Appendix-B.

Contents

7.1	Introduction to Semiconductor Industry (SI)	76
7.1.1	Characteristics of Integrated Circuits (IC) Production Environment	76
7.1.2	Production Line multi-Source Data	80
7.2	Presentation of Case Study	82
7.2.1	Thermal Treatment Workshop	82
7.2.2	Reactor Module of TASMI-03	84
7.3	Discussion	84

7.1 Introduction to Semiconductor Industry (SI)

7.1.1 Characteristics of Integrated Circuits (IC) Production Environment

The semiconductor industry (SI) is characterized as high-mix low-volume production environment. It is fragile, fastest growing and highly complex production settings where market is driving not only the product (demand volume and diversity) but also the process (technology). The sales revenues in the SI are characterized by cyclic demand patterns and positive compound

annual growth rate (CAGR)¹ of 7.72% (Figure 7.1) in 2015 as compared to 1989. The demand for ICs is mainly driven by end user markets from the Electronics Industry (EI) e.g. data processing, communication, consumer electronics, industrial sector and automotive. PricewaterhouseCoopers² forecasts 5-years (2014-2019) growth rate of +11.2% for automotive sector. The SI forms a part of this complex interaction among these multiple industrial sectors (Kumar, 2008; Yoon and Malerba, 2010). Therefore, demand is increasing not only in volume but also in diversity that led the emergence of high-mix low-volume production environment and shortening product life cycles, in the SI. The success in the SI requires sustainable production capacities to cope up with associated challenges in this complex and highly competitive environment.

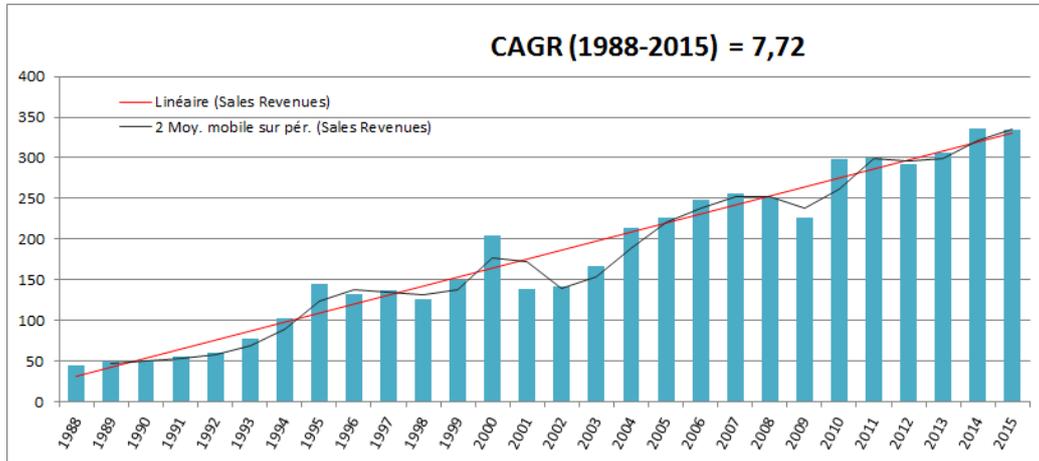


Figure 7.1: Global sales revenues of SI.

The overall semiconductor manufacturing process can be classified into three main stages,

1. *Design*: The manufacturing process begins by receiving customer demand (functional specifications) for product. This is also referred as Computer Aided Design (CAD) phase during which simulation tools are used to design the functional blocks needed to manufacture the product. The design tools are also used to evaluate the electrical description of these blocks to respect the limits of functional parameters

¹The CAGR formula is given by $CAGR = \left(\frac{\text{Ending year of sales}}{\text{Reference year of sale}} \right)^{\frac{1}{\text{Number of years}}} - 1$

²Retrieved from: <https://www.pwc.com/us/en/technology/publications/assets/semiconductor-industry-device-deal-trends.pdf>

defined with the client. The CAD tools also help transform the design into physical architecture (more commonly known as layouts) enabling the generation of masks for IC production in the second stage.

2. *Wafer fabrication (Front-end)*: In this stage, the ICs (transistors, resistors, metal interconnections, etc.) are manufactured on the surface of silicon wafers (example, of 200 mm or 300 mm diameter).
3. *Wafer assembly and packaging (back-end)*: In this stage the processed wafers are tested prior to be assembled and cut as individual ICs before being encapsulated in a plastic moulding compound or ceramic case, forming the package. Finally, the individual circuits are again tested and shipped to customers.

Among these stages, front-end represents the most complex process flow and in this environment is the main focus of our research. The wafers fabrication is a multi stages process with re-entrant flows. It has eight different production workshops performing manufacturing operations on each chip, such as cleaning and polishing wafer surfaces using Chemical Mechanical Polishing (CMP), ion implantation, oxidation and thermal treatment, deposition of dielectric, deposition of metal, photo-lithography, etching and metrology. For each product type, and depending on the technology, a wafer undergoes about 1100+ elementary operations steps over a period of several weeks (See Figure 7.2)³.

The SI production line consists of hundreds of equipment and are divided into three types,

- **Production equipment**: These equipments are capable of performing physical transformations on silicon wafers. They are used to perform any of above mentioned elementary process steps. These equipment can be of single, batch or multi process module type.
- **Metrology equipment**: These equipments measure and control the quality of the processes performed by the production equipment. Among others, these are used to measure critical dimensions and uniformity of deposition or etching and contamination of wafers.
- **Automated transport equipment**: These are the automated robots enabling the transport by batches of 25 wafers (called FOUP) from one equipment to another throughout the production areas.

In such complex production environment, unscheduled equipment breakdown is the limiting factor for sustainable production capacities. Therefore, higher level of complexity and uncertainties during production arise due to the following facts:

³Figure is adapted from Yugma et al. (2015).

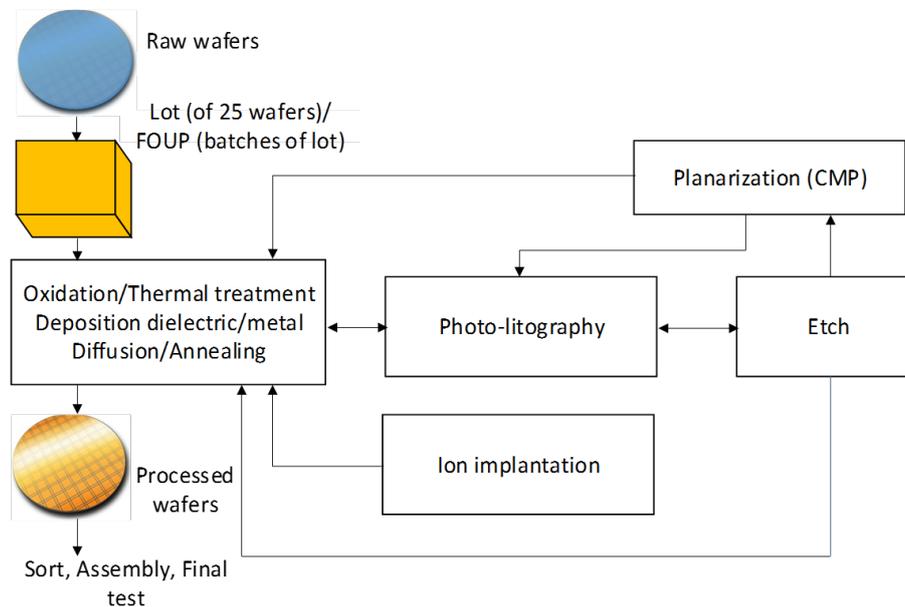


Figure 7.2: Front-end proces flow.

- Technology is developed every 2-3 years (development cost 3-5 B\$) (Shahzad, 2012) to manufacture smaller designs while ensuring and improving circuit performance. The high mix of products, but in low volume demand leads to frequent changes in process recipe and therefore the operating conditions of equipment vary due to differences in materials used, configuration or layout of devices and interconnections, feature and overall chip size.
- High cost of process and equipment drives manufacturers to maximize the use of their equipments, having as minimum break downs or idle times as possible.
- Variations in product quality produced during manufacturing often are functions of the product being produced as well as the production equipments being used. Identical equipment may process identical wafers differently based on conditions like number of lots processed since the last maintenance event, small differences or minor variations in ambient conditions.
- More and more data from multiple data sources are generated but no existing R&D methods are yet implemented to link the high interaction between different elements of the production to the unscheduled breakdown events.

7.1.2 Production Line multi-Source Data

Data is the fundamental basis for most preventive, conditional and predictive functions and applications. A big quantity of data are generated everyday in the SI to maintain the processes at their specification levels, to monitor product and equipment for potential drifts and faults. These data are either intervention log data (installation, qualification, adjustment and maintenance) or process data (production, R&D, measurements, inspections, transportation of lots) made in the clean room. All these data are stored in multiple databases to allow easy access to the data needed for different needs of analysis. Figure 7.3⁴ shows a representation of the automated production environment which consist of production and metrology equipment, and Automated Material Handling System (AMHS), information systems (consist of Automation and MES modules), existing data bases (TGV, EDA, FDC, etc) and data analysis tools (Klarity, KLAACE, etc), to manage, supervise and control the production system 300mm production line at STMicroelectronics, Crolles.

The functions of different databases can be related to either production or metrology equipment.

- Databases related to production equipment:
 - *Tool Global Visibility (TGV)*: This collects maintenance data. It archives historical data and calculate the indicators which are then used to control the maintenance activities.
 - *Fault Detection and Classification (FDC)*: This database stores equipment parameters, collected in real time from equipment sensors, during production operations.
 - *Engineering Data Analysis (EDA)*: This contains historical process data from equipment including production operations (Work In Process (WIP)), process (recipe), product and metrology.
- Metrology and inspection databases:
 - *Inline*: This stores all physical measurements made on silicon wafers during production operations and primarily serves to SPC system.
 - *Parametric Test (PT)*: This includes the electrical measurements performed on test structures located around the circuit for all production wafers before the assembly and package phase.

⁴Figure is translated to English from Said (2016).

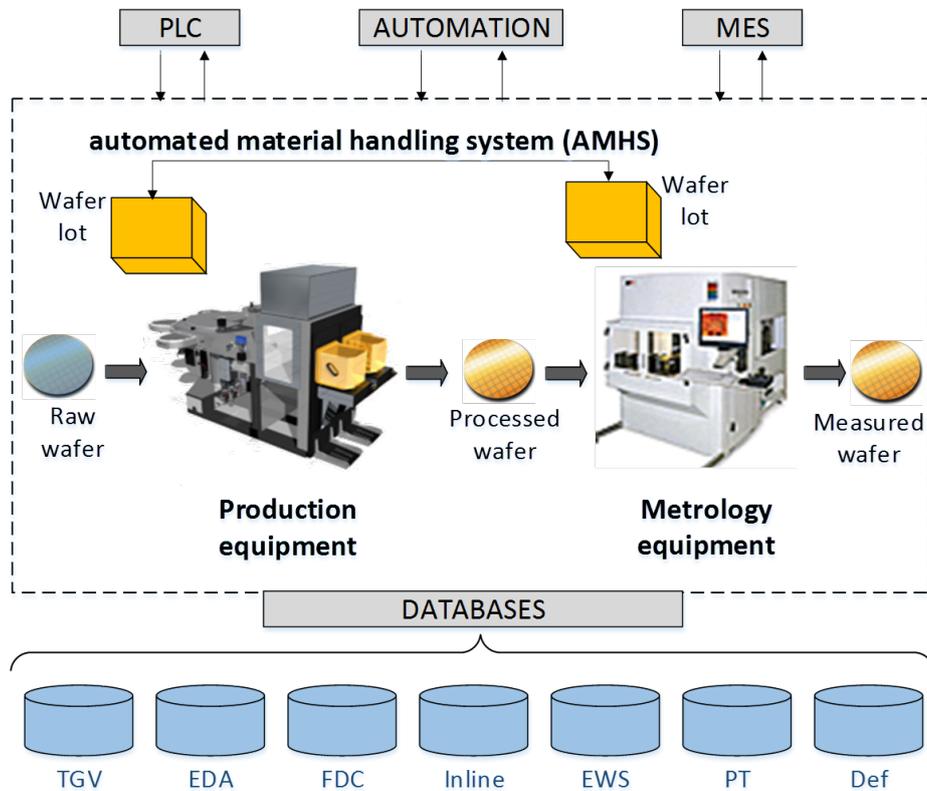


Figure 7.3: Description of information systems of Crolles, 300mm production line at STMicroelectronics, Crolles.

- *Electrical Wafer Sort (EWS)*: This database stores all electrical measurements (functional tests) made on IC chips after manufacturing.
- *Defectivity (DEF)*: This stores inspection information of particles on the surface of the wafers.

In the SI, equipment can have one or more process module(s) (a.k.a. chamber), non-process modules, several sub systems associated directly with the equipment and sub systems associated to the modules (Figure 7.4). These are related to one another in parent child relation. This complex modelling scenario results in the inherited failure concept where failure arising on either parent or child has an impact on the state of other. The rules are illustrated in Figure 7.5.

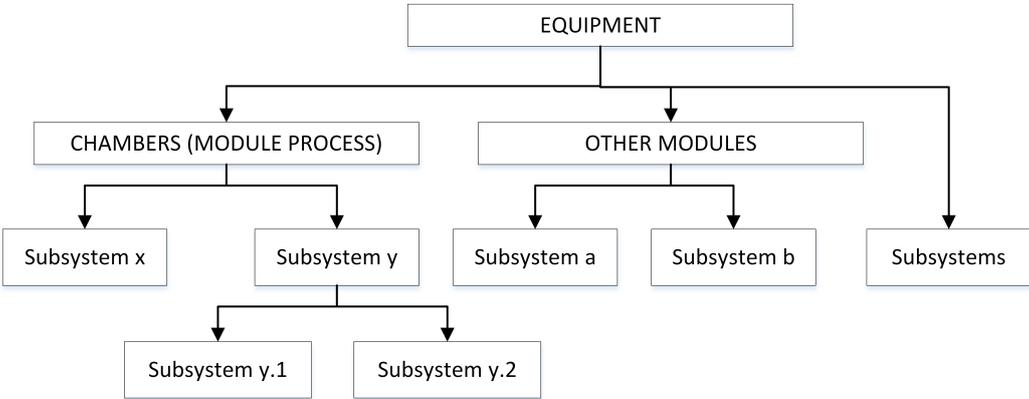


Figure 7.4: Chart representing a hierarchy of equipment composition in the SI.

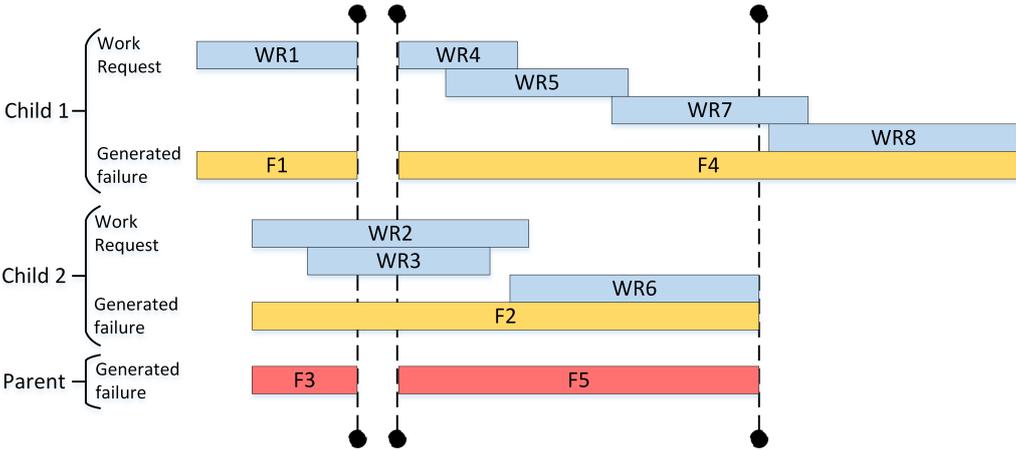


Figure 7.5: Rules on inherited failure within different level of components.

7.2 Presentation of Case Study

7.2.1 Thermal Treatment Workshop

In the Thermal Treatment (TT) workshop, two types of operation are performed:

- Deposition of oxide and nitride layers using LPCVD (Low-Pressure Chemical Vapor Deposition) technique in which a growth or deposition is carried out at a relatively high temperature (>600°C) on the surface of the silicon wafer.

- Annealing (heat treatment) process after the insulated layers deposition or ion implantations. This process is performed to densify, stabilize crystalline structure of silicon wafers and/or cure the defects in depositions, or activate the electrical dopant after ion implantation operation, prior to the next steps.

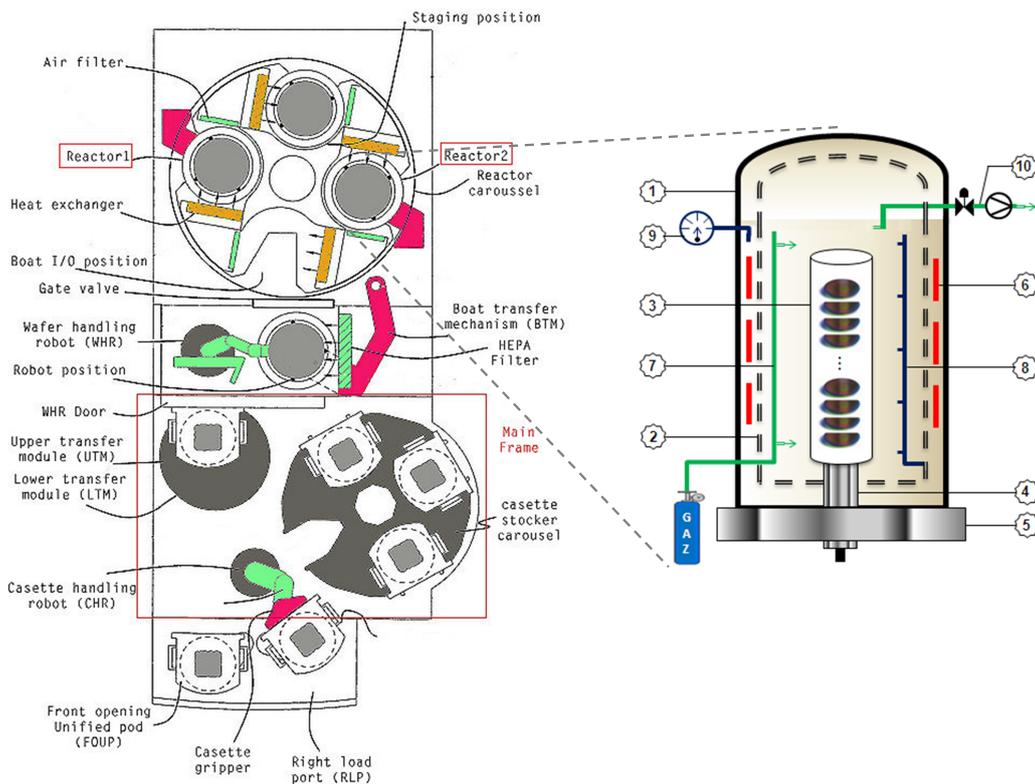


Figure 7.6: General structure of the LPCVD equipment (left) and the reactor module (right) from TT workshop.

There are eight LPCVD equipment of type 'TASMI' in the thermal treatment workshop of the production line under consideration. The general structure of this TASMI equipment is presented in Figure 7.6⁵. The equipment type in this production line is batch cluster with two process modules known as reactors. The Reactor1 (TASMI0X_1), Reactor2 (TASMI0X_2) and Mainframe (TASMI0X_MF) are the three main modules of the equipment. Mainframe can be further composed of many sub modules. Each of the TASMI0X (X refers to numbers from 1 to 8) has similar functions but with

⁵Figure is retrieved from <http://www.google.fr/patents/US7553516> on 21st March 2014)

different composition of its modules. In this thesis, we consider the three main modules from TASMI03 (TASMI03_1, TASMI03_2 and TASMI03.MF). The integrated failure/cause diagnosis BN models at module and equipment levels which is the focus of methodology proposed and validated in chapters 5 and 8 are therefore developed only for these equipment modules.

7.2.2 Reactor Module of TASMI-03

The reactor module, (see Figure 7.6)⁶, considered is composed of:

1. Exterior chamber
2. Inner chamber with quartz (liner)
3. Wafer support (boat)
4. Elevator boat rotation
5. Watertight door for loading and unloading
6. Heating elements
7. Gas panel
8. Temperature sensor
9. Pressure gauge (manometer)
10. Pressure regulator.

7.3 Discussion

This chapter presents key characteristics of the SI and the case study carried out on TASMI equipment selected from thermal treatment workshop to validate the proposed methodologies in chapters 5 and 6. The individual data dimensions (variables) used for modelling are presented in the respective validation chapters 8 and 9.

⁶Figure is retrieved from Bouaziz (2012)

Application and Results of Integrated Drift Source and Failure/Cause Methodology

Introduction

This chapter presents the validation results of 3-step methodology for integrated drift source and failure/cause diagnosis (chapter 5). This methodology primarily provides answers to first two research questions that focus on reducing misdiagnosis in the source of product quality drifts and failure/causes and finding the best modelling level (equipment/module) for more accurate failure/cause diagnosis in high-mix low-volume settings. The answers provide us the means to stabilize and improve production capacities which is the key for success in the SI.

Contents

8.1	Level-1: Product Quality Drift Source Diagnosis	86
8.1.1	Identification and Classification of Predictors	86
8.1.2	Learn and Train Semi-Supervised Bayesian Network	88
8.1.3	Test Bayesian Network for Product Quality Drift Source Diagnosis	90
8.2	Level-2: Equipment Failure/Cause Diagnosis	93
8.2.1	Identification of Failure/Causes at Module and Equip- ment Levels	94
8.2.2	Learn and Train Semi-Supervised Bayesian Network at Module/Equipment Levels	94
8.2.3	Test and Compare Bayesian Network Performance at Module/Equipment Levels	95
8.3	Conclusions and Discussions	98

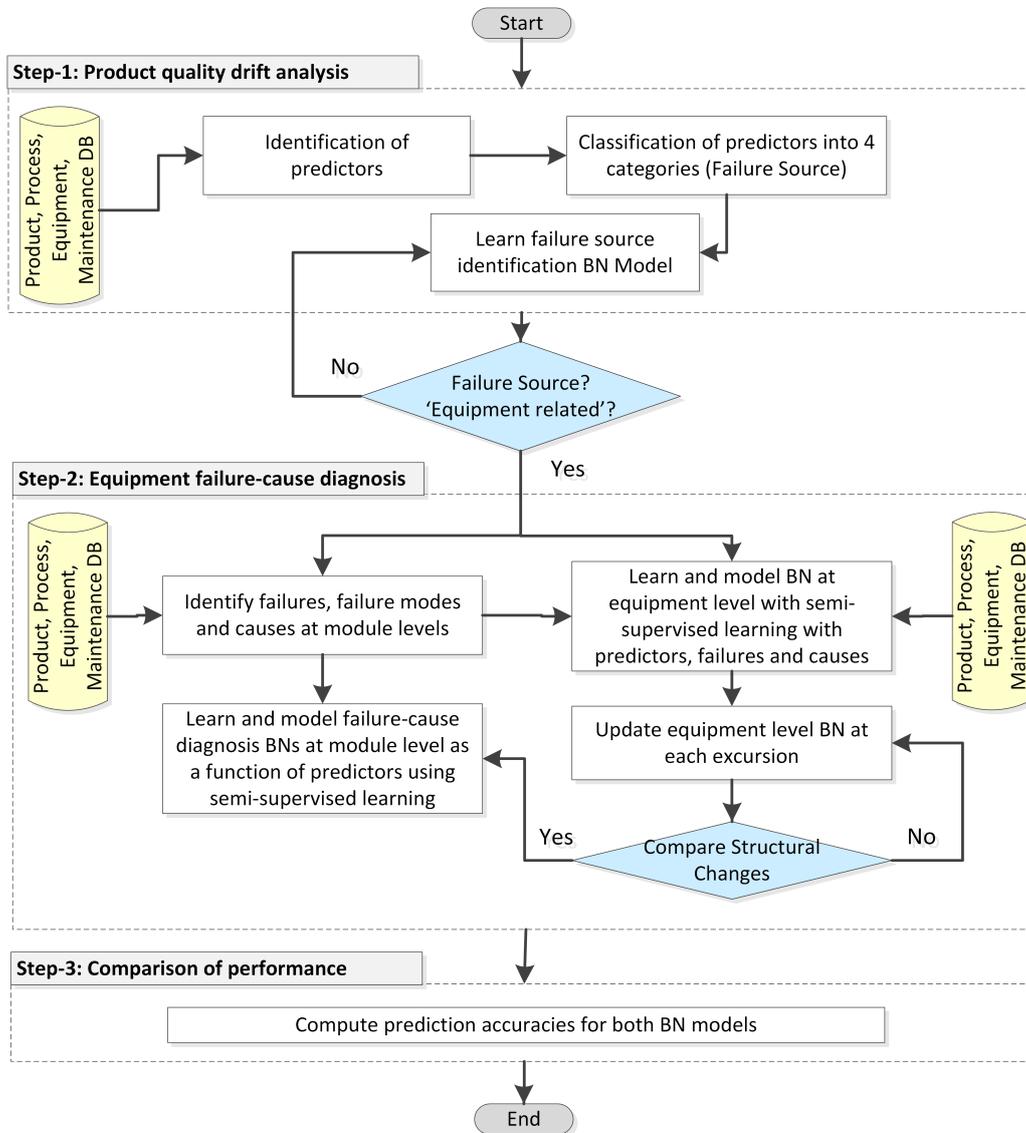


Figure 8.1: Proposed 3-step methodology for reducing failure/cause misdiagnosis in the SI.

8.1 Level-1: Product Quality Drift Source Diagnosis

8.1.1 Identification and Classification of Predictors

The dataset used in the case study spans six months (from week 27th to week 52nd of 2013) and are collected across the databases linked to both the

production and metrology equipments for TT equipment. These databases however are further distinguished as product, process, equipment and maintenance according to their specific functions. The acquired data are going to be used as the predictors, failures and causes. The first task of the step-1 is the identification of predictors from the four databases and are used as the nodes to generate the BN to accurately identify the product quality Failure Source (FS), as well as development of an integrated failure/cause diagnosis BN models at the module and equipment level (subsections 8.2.3 and 8.2.2).

The predictors are classified in four axes as *Product*, *Process*, *Equipment* and *Maintenance*. The identification and classification of potential predictors from the databases is the most difficult and complex task. It requires multidisciplinary expertise from product, process, equipment and maintenance domains; therefore, a task force with required expertise was formed. This step of the case study is implemented using brainstorming between experts which is capitalized in the Ishikawa diagram (Ishikawa, 1990), see Figure 8.2. Alternative proposition is to use Pareto analysis on available Failure Mode, Effects and Criticality Analysis (FMECA) to identify the important failure predictors across the domains. The Ishikawa approach is more beneficial if such FMECA files are updated and available at the equipment and module levels. However, due to the absence of formally capitalized and updated knowledge in the SI (Said et al., 2016), we can only rely on brainstorming without it. The brainstorming sessions resulted in the formalization of well-known Fishbone diagram to find potential predictors and links to FS from product, process, equipment and maintenance areas. The results are presented in Figure 8.2.

The TASMI equipment is of batch cluster type; hence, they process multiple lots in a given step. Therefore current/previous product combinations might influence the product quality. Number of reworks, wait time before process and defect distribution from previous steps are also identified as key product predictors linked with product quality drift. The process capability (C_p) and process capability index (C_{pk}) are the key process predictors. It is also identified that not only current recipe but also previous recipe and their respective process steps combinations could be strongly linked with product quality. The FDC sensor signals from equipment database are not directly considered; however, decisional information based on these signals is a good candidate for potential predictors. The key predictors from equipment database are equipment capability (C_m) and equipment capability index (C_{mk}); however, Overall Equipment Efficiency (OEE) indicators and counters are also included as the additional predictors. The counters are the meters associated with equipment modules (process chambers and mainframe), used for triggering preventive maintenance actions. Last category

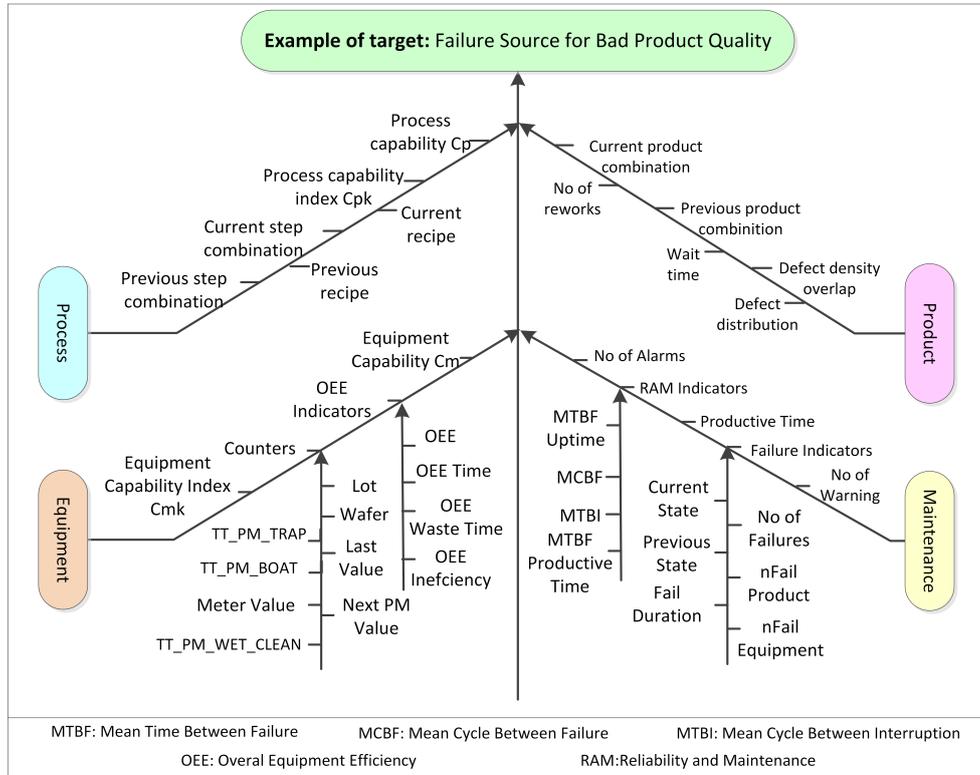


Figure 8.2: Classification of predictors.

of predictors is the maintenance where Reliability, Availability and Maintenance (RAM), and failure indicators are identified as the key predictors. The data is collected for these predictors against product quality drifts. The data for OEE, RAM, process and equipment capability, and failure indicators are aggregated on weekly basis whereas rest of the data is instantaneous for a given product and process step.

8.1.2 Learn and Train Semi-Supervised Bayesian Network

The structure of BN can be obtained either through experts' knowledge or learned from the data. The BN structure in this thesis is learned with the BayesiaLab 5.3 using score-based unsupervised learning algorithms that use Minimum Description Length (MDL) as an objective function. The structure is learned and optimized using combination of two search and score algorithms, sequentially. The initial structure is learned using the Equivalence Class (EQ), and it is further optimized using Tabu algorithm, that comple-

ment EQ in terms of search space and exploration strategy. The learned and optimized BN is presented in Figure 8.3 where FSs are modelled as the function of predictors (differentiated with different colors respecting the class of predictors). The green, pink, yellow and light brown colors represent *Product*, *Process*, *Equipment* and *Maintenance* respectively). The objective of showing this graph (Figure 8.3) is to present the complexity of resulting network.

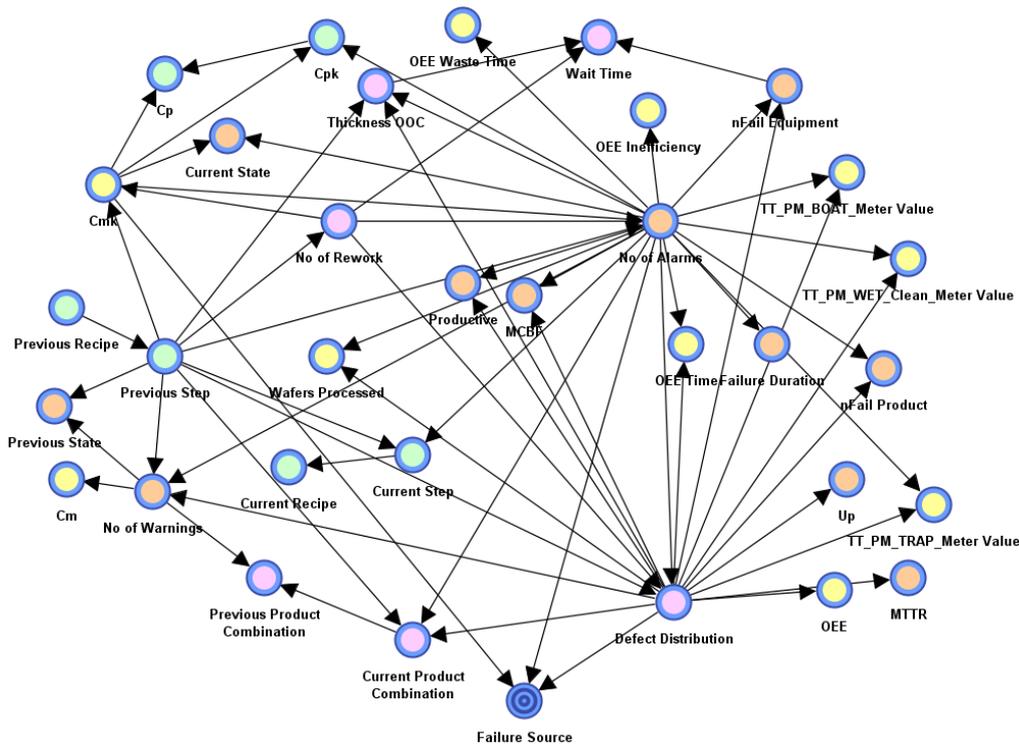


Figure 8.3: BN model for FS identification.

Figure 8.4 highlights the nodes that have direct causalities with the target. The probability to have each failure source shall differ based on different values taken by these nodes. The proofs of concept are presented in Figures 8.5. Failure source (light red background) is the result of inference given the observations of highlighted predictors (white background with distinct coloured frames highlighting different categories). Chosen predictors are the nodes with direct association. It can be seen that in the Figure 8.5a, BN identifies *Product* (64%) and *Maintenance related* (36%) as failure sources. Hence, in this situation, maintenance personnel should not stop the equipment. Similarly, the Figure 8.5b shows that maintenance is found as the only reason against given evidence from predictors; hence, BN model suggests to

stop the equipment for further investigation on failures and causes.

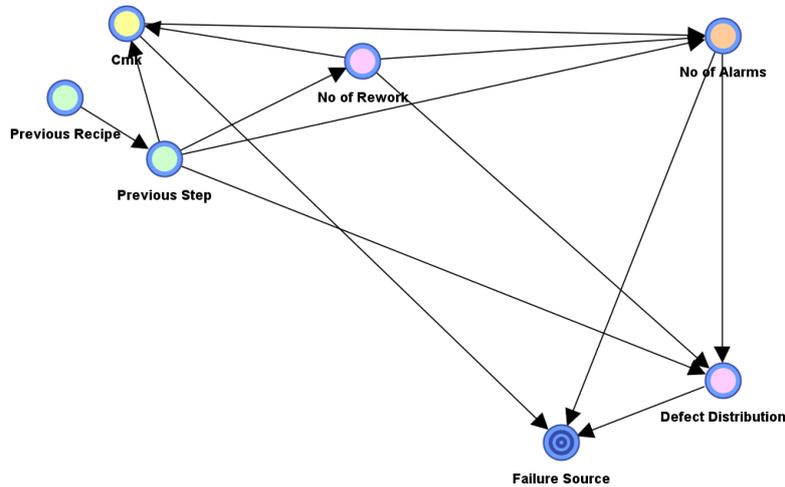


Figure 8.4: Representative nodes for the proof of concept.

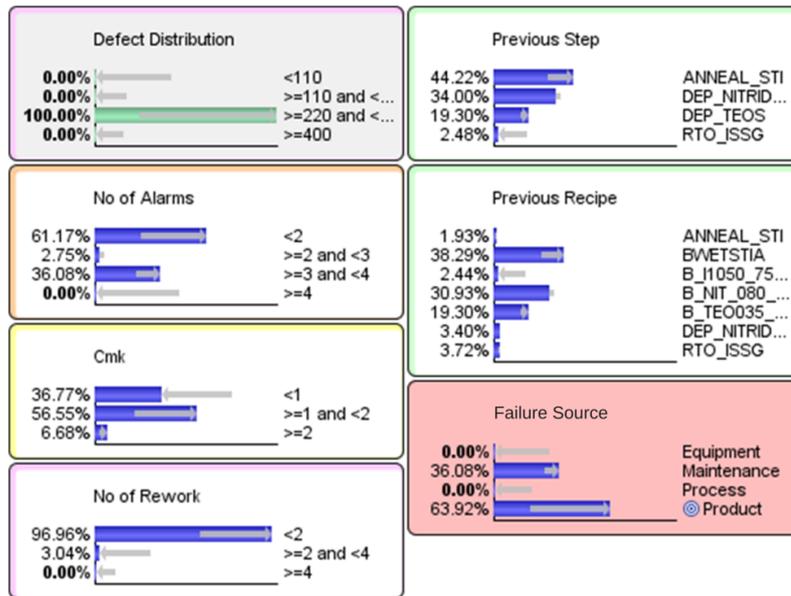
8.1.3 Test Bayesian Network for Product Quality Drift Source Diagnosis

Table 8.1 presents the results from one of the FS BN model precision and reliability matrices (refer paragraph 5.2.2 of chapter 5) based on 10-fold cross validation strategy, whereas the aggregated results are summed in Figure 8.6 with box plots to demonstrate the distribution of true positive prediction precision and reliability for each FS. The results show that the BN model is more accurate in diagnosis of equipment and process as the source of unscheduled breakdown drifts than product and maintenance. This can be due to dependencies and choice of predictors. However, the overall model accuracy is high and with low standard deviation.

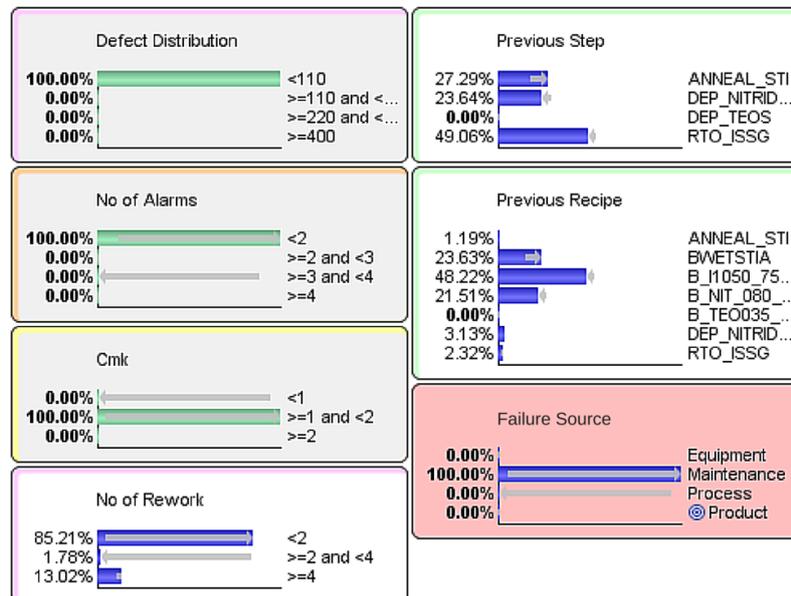
Further analyses are executed using available tests in BayesiaLab 5.3. The Figure 8.7 shows FS prediction performance with Receiver Operating Characteristic (ROC) curve, a graph to plot true positive rate¹ (Y-axis) against false positive rate² (X-axis). Its index represents the surface under the ROC curve divided by the total surface and in the graph of *product* for example, it represents a 99.66% average accuracy with 0.34% of false positive prediction. Again, equipment and process ROC (100% and 99.9% respectively) prove to provide with the better performances.

¹True positive rate measures the proportion of positives that are correctly identified

²False positive rate measures the proportion of negatives that are correctly identified



(a) Product as FS.



(b) Maintenance as FS.

Figure 8.5: Proof of concept of FS (Failure Source) identification.

Table 8.1: Precision and reliability matrices of FS BN.

Precision	Eq (1533)	Maint (1417)	Proc (1015)	Prod (1035)
Eq (1533)	100%	0%	0%	0%
Maint (1338)	0%	95.67%	0%	0%
Proc (1043)	0%	0%	100%	0%
Prod (1086)	0%	5.58%	0%	97.77%
Reliability	Eq (1533)	Maint (1417)	Proc (1015)	Prod (1035)
Eq (1533)	100%	0%	0%	0%
Maint (1338)	0%	100%	0%	0%
Proc (1043)	0%	0%	97.83%	2.68%
Prod (1086)	0%	7.27%	0%	94.03%

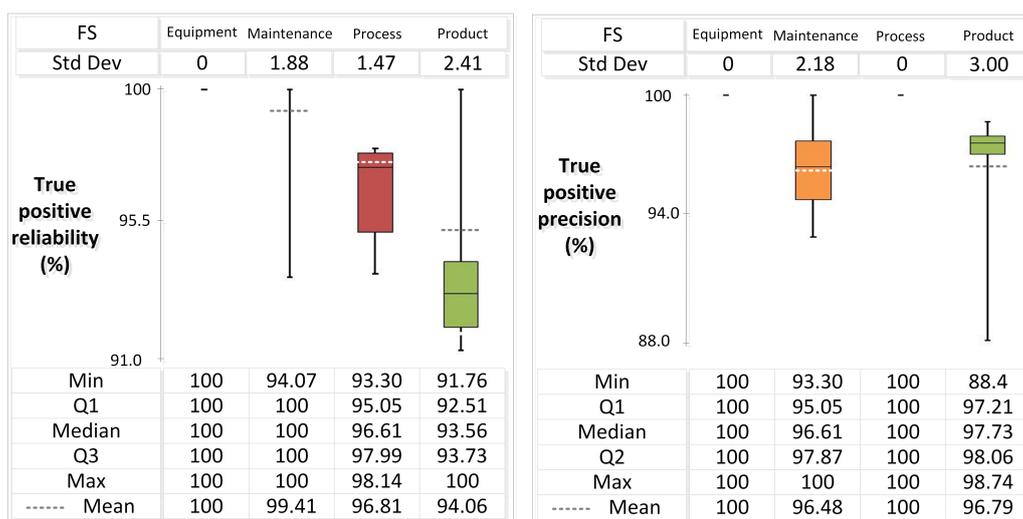


Figure 8.6: Failure source standard deviation and the box plot graph of true positive prediction performances.

Furthermore, the capability of FS identification model with gain curves is presented in Figure 8.8. The x-axis represents rate of individual cases taken into account for prediction whereas y-axis represents rate at which they are predicted accurately with target failure source. In the figure, the blue curve represents the gain curve of prediction using random policy and the red using optimal policy. For example, the figure for *product* illustrates that choosing 26% of individuals allows getting 100% of the individuals with the target variable with the optimal policy. The data mining Gini index for cross validation represents the gain over random model and is computed as

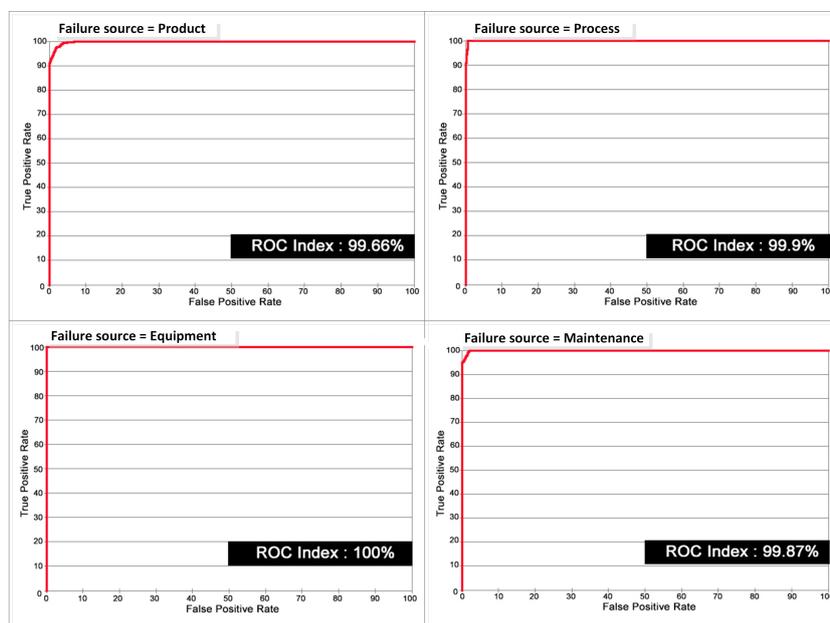


Figure 8.7: Prediction accuracy with ROC curves for all FS.

the surface between the red curve and the blue curve divided by the surface above the blue curve. The relative Gini index is computed by dividing the area within triangle formed due to crossing of red, blue and black dotted lines with area within yellow line triangle. The yellow curve is the curve that enables us to determine the percentage of individuals allowing identical value of the relative Gini index and ROC index. It is observed that FS identification capability for product and process are higher than equipment and maintenance.

8.2 Level-2: Equipment Failure/Cause Diagnosis

The FS identification model, presented in previous section, is the first step towards reducing unscheduled equipment failure breakdowns. This is complemented by failures and causes diagnosis through developed BN models at module and equipment levels. For the experiment and proof of concept, we have used three modules (i) Reactor1, (ii) Reactor2 and (iii) Mainframe.

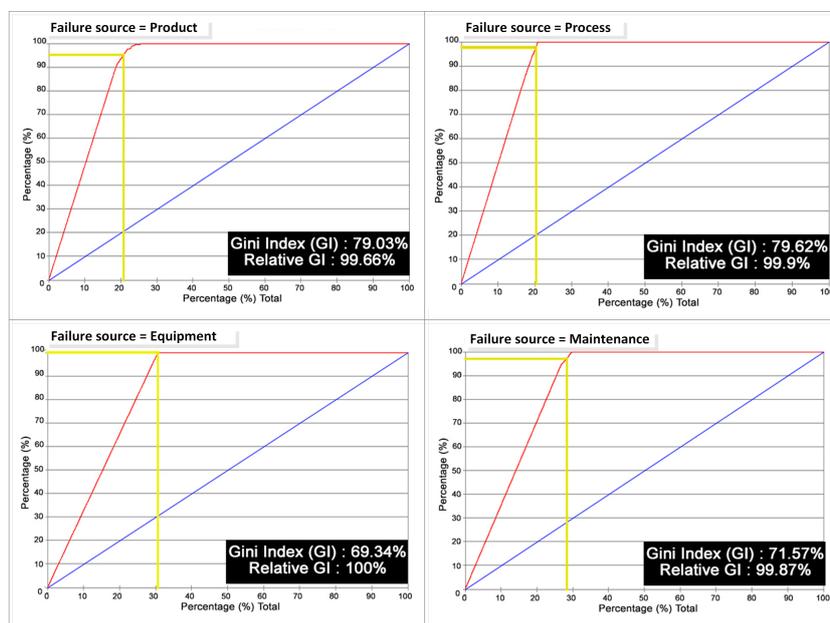


Figure 8.8: Gain curves for all FS.

8.2.1 Identification of Failure/Causes at Module and Equipment Levels

The predictors from FS identification BN model (section 8.1.1) are complemented with equipment (or module) level failures, failure modes and causes to develop the integrated failure/cause BN at equipment and module levels. For each model, the target nodes Failure (localisation), Failure mode are modelled as the function of these predictors; however, causes are also allowed to be directed from these predictors. Therefore, failures, failure modes and causes are identified from equipment and maintenance databases (see Table 8.2). The causes are modelled as binary nodes.

8.2.2 Learn and Train Semi-Supervised Bayesian Network at Module/Equipment Levels

The set of predictors resulting from section 8.2.1 are then used to learn the structure of integrated failure/cause diagnosis BN models for three modules (Reactor1, Reactor2, Mainframe). The resulting BN are presented in Figures 8.9a, 8.9b and 8.10, for Reactor1, Reactor2 and Mainframe. The color scheme for predictors classes is modified as compared to the ones presented in subsection 8.1.1. The green, light brown, dark pink, dark yellow represent

Table 8.2: List of failures and failure modes's states and causes nodes.

Node/module	FAILURE (Single node and list of its states)	FAILURE MODE (Single node and list of its states)	CAUSES as binary nodes (list of the cause nodes)
Reactor1	ELEVATOR_BOATROTATION OCAP_SPC REACTOR_CARROUSEL WHR_BTMM	BEARING LOTS MOTOR NPW WHR_BTMM Z.OTHER	Tool stop FDC generic Reactor tube elevator boat pickup Alarm Boat Elevator Integrator OCAP Pb limit calibration of robot Overheat digital input 12 Reactor Status Inhibit next load
Reactor2	ELEVATOR_BOATROTATION OCAP_SPC Z.OTHER	BEARING LOTS NPW Z.OTHER	Abort the SPV following "alarm section tube" Wait boat alarm OCAP AbortINL
MainFrame	CHC_CAROUSSEL GAS.FLOW GATE.VALVE REACTOR_CAROUSSEL TUBE.DOOR	BLOWER SWITCH Z.OTHER	Problem loading the boat Tube robot door open Analysis bp remounted in O ² JobOut fail problem on TASMIO3 Gate valve open Gate Valve impossible to be opened Pb with Pick up switch

Product, Process, Equipment and Maintenance respectively, whereas failure modes and causes are added as nodes with blue and orange colors, respectively. The nodes, not connected, in these BN models are found with zero influence on either failures, failure modes or causes.

The proofs of concept for Reactor1, Reactor2 and Mainframe are presented in the Figures 8.11a, 8.11b and 8.12. Again, in it we present only few chosen predictors for visualization purposes while the module's failure, failure mode and causes diagnosis made by BN model is presented as the function of predictors (in green, green and orange frames of the right columns). In the BN model, we have presented the key predictors having direct influence on causes and failures.

8.2.3 Test and Compare Bayesian Network Performance at Module/Equipment Levels

To find out, whether module level BN models are more accurate than equipment level model, we developed an equipment level failure/cause diagnosis model. Besides this, we added one node 'Module' (maroon node) to diagnose failures for a given module in the equipment. The model is presented in Figure 8.13. It can be observed that all nodes are connected. The nodes having zero influence in module level BNs, are connected in this network that add confusion and influence the equipment level diagnosis. Confusion is also caused due to the fact that similar modules, Reactor1 and Reactor2 share common failures such as Out of control alarms which are detected by the SPC system (*OCAP_SPC*), *Reactor carousel* and *Elevator boat rotation*.

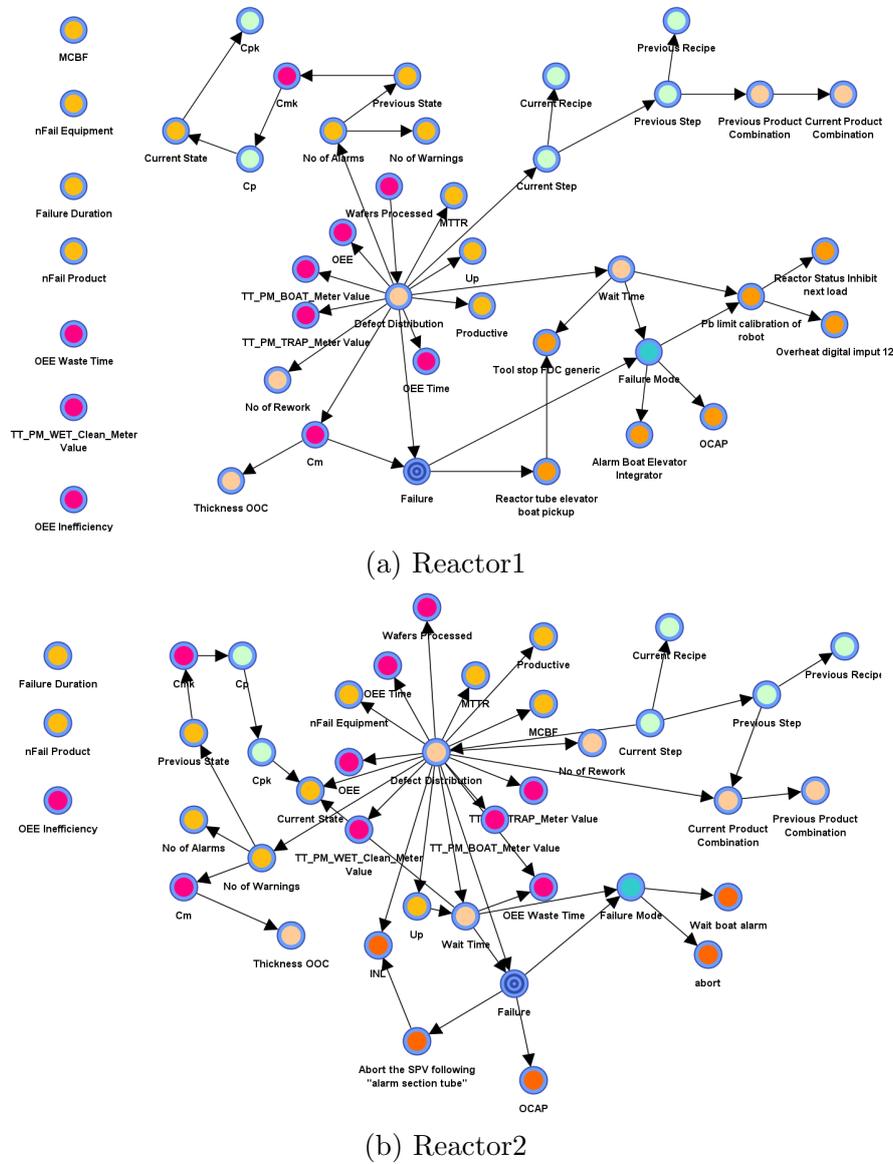
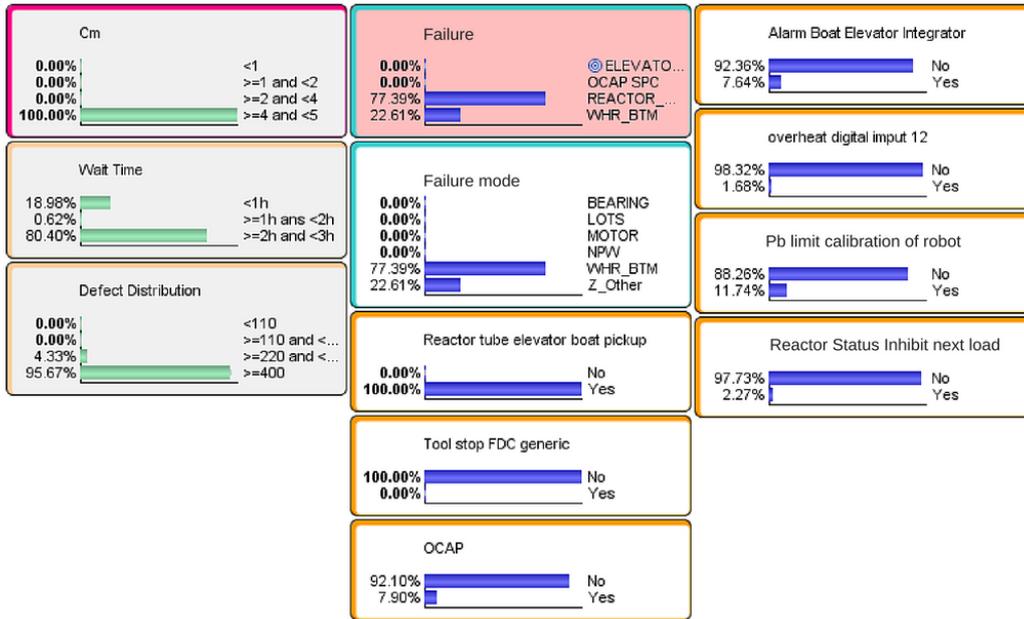


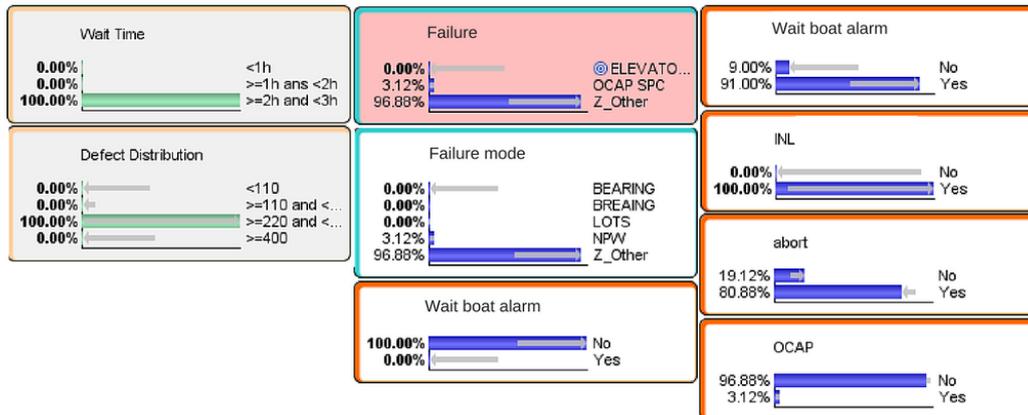
Figure 8.9: Module level BNs for failure/cause diagnosis.

Each module has different occurrences of these failures but in this network, they do overlap. It is also observed from the proof of concept (Figure 8.14) that for given predictors, all modules have 33.33% probability of occurrence that confirms the added confusion.

As the last task of the second step, the equipment level BN is modelled and proposed to be updated upon new excursions, any structural change between two consecutives equipment level BNs will be used as a signal to



(a) Reactor1



(b) Reactor 2

Figure 8.11: Proofs of concept for Reactor1 and Reactor2 BN.

offer more accuracy over equipment level BNs for failure/cause diagnosis.

8.3 Conclusions and Discussions

The key results obtained in this chapter are the answers to the two research questions. The case study results obtained using data collected across the world reputed semiconductor manufacturer demonstrates that integrated

Chapter 8. Application and Results of Integrated Drift Source and Failure/Cause Methodology

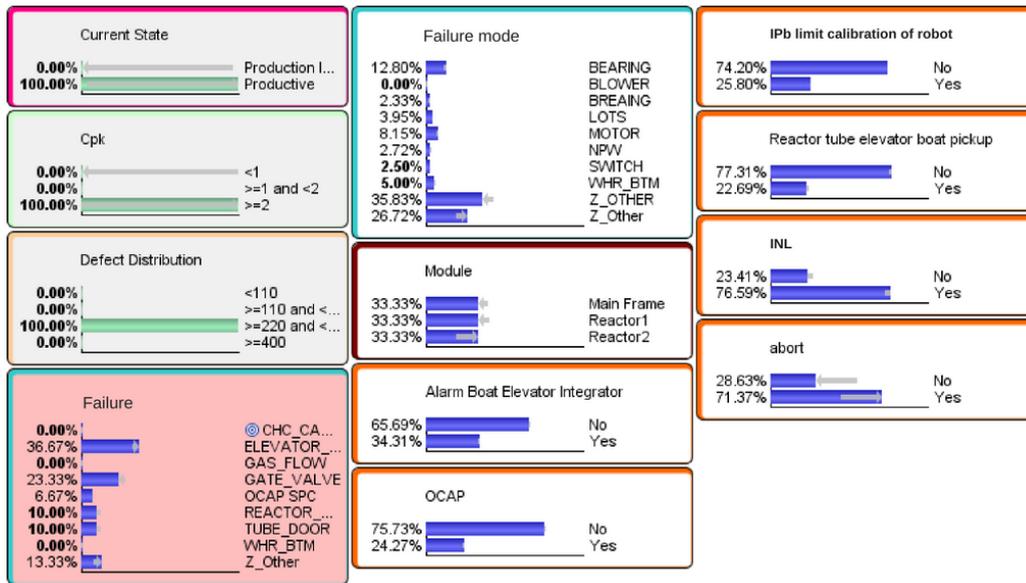


Figure 8.14: Proof of concept from equipment level BN.

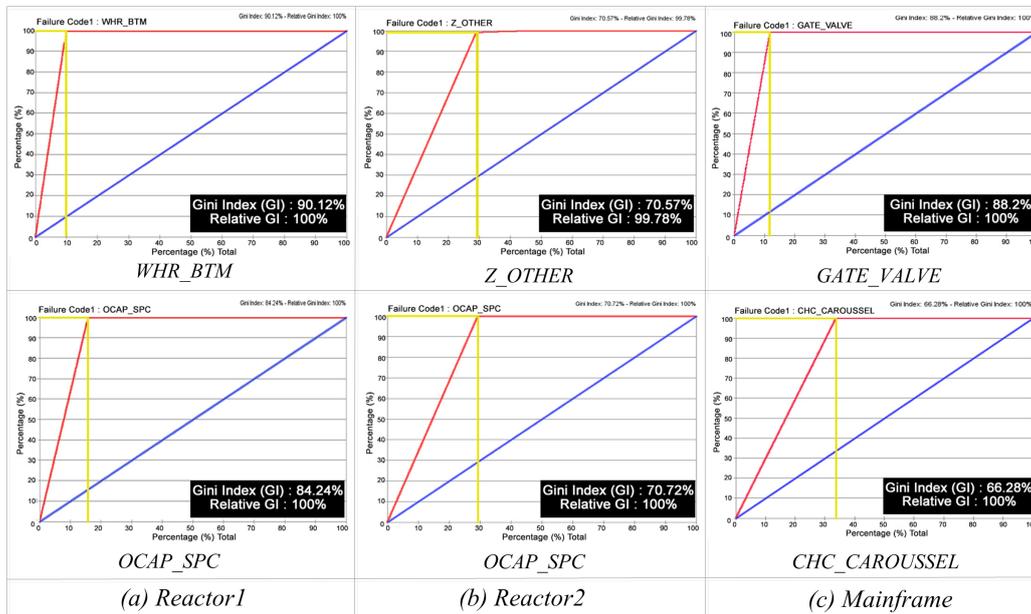


Figure 8.15: Gain curves for module level BNs.

Chapter 8. Application and Results of Integrated Drift Source and Failure/Cause Methodology

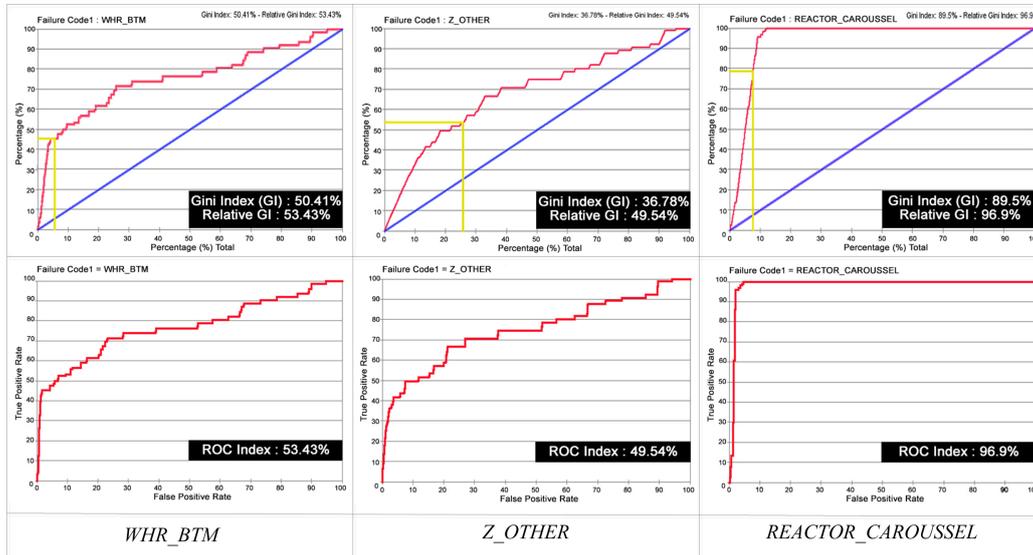


Figure 8.16: Gain and ROC curves for equipment Level BN.

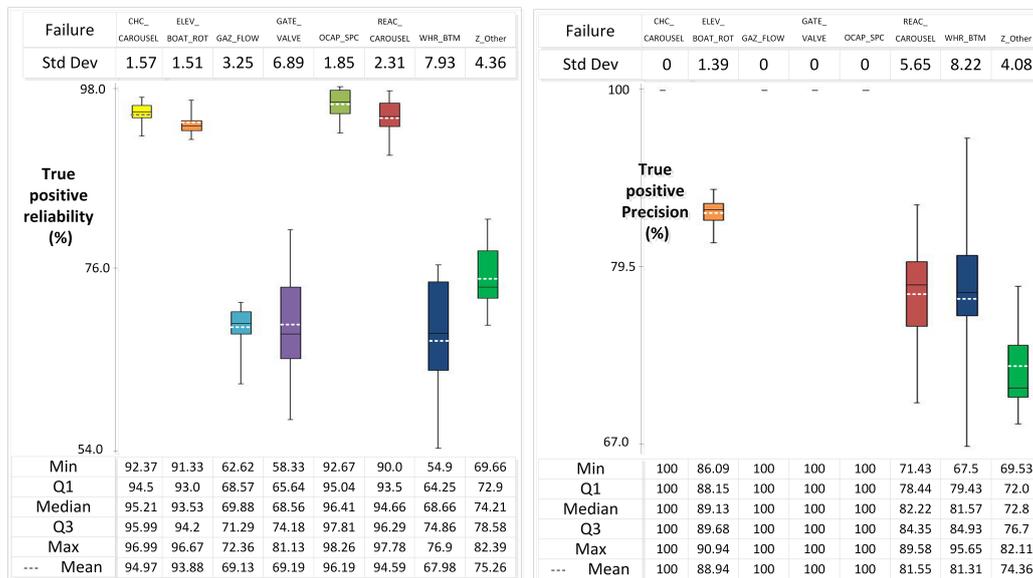


Figure 8.17: Equipment level standard deviation and the box plot graph of true positive prediction performances, reliability (left) and precision (right).

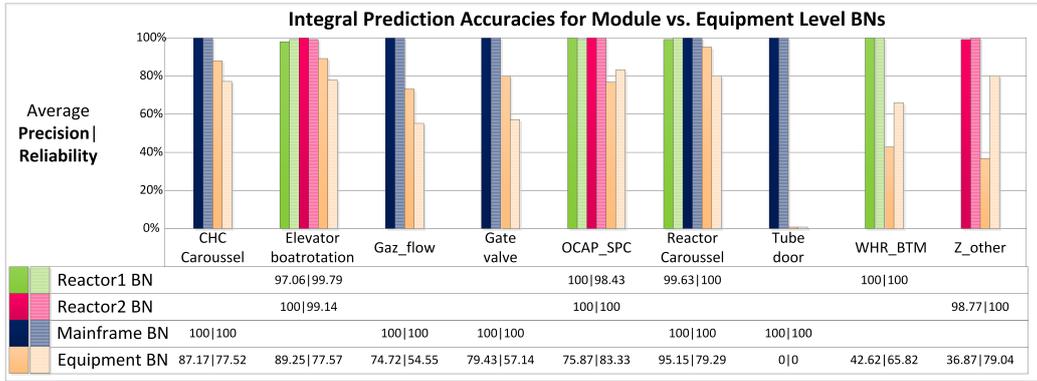


Figure 8.18: Reliability and precision comparison by failures and module/equipment.

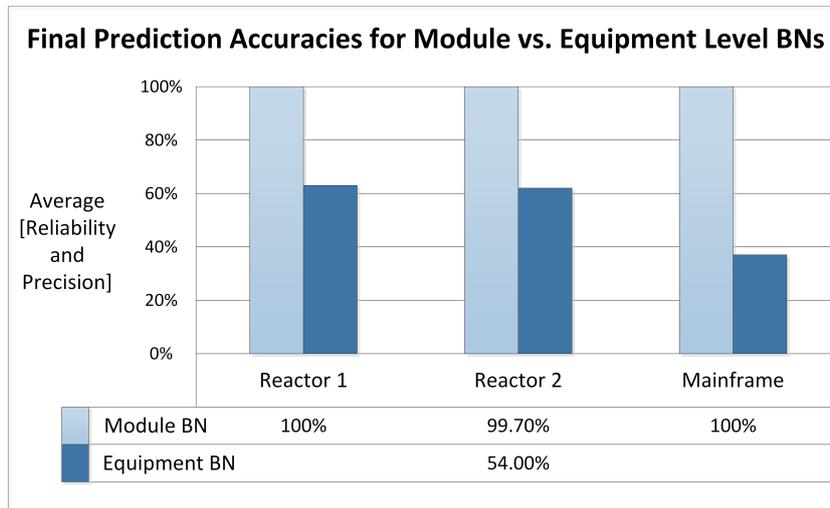


Figure 8.19: Gain in prediction accuracy for module level BNs over equipment level.

Application and Results of Online Failure Prediction Methodology with Time Bound Failure Signatures

Introduction

This chapter presents validation results of the methodology for the extraction of time bound failure signatures (patterns/rules) as presented in chapter 6. The Figure 9.1 presents the summary of the proposed methodology. The dataset used in this case study spans 10 months (from week 27th 2013 till week 16th 2014). The results are presented step by step from sections 9.1 till 9.4.

Contents

9.1 Step-1: Dataset Preparation	104
9.1.1 Identification and Selection of Variables	105
9.1.2 Interval Discretization	105
9.1.3 Dataset Pre-Processing	106
9.2 Step-2: Equipment Failure Probability Plots	107
9.2.1 Learn/Optimize Bayesian Network Structure and compute CPT	107
9.2.2 Plot Failure Probability from Bayesian Network for each Discretized Interval	108
9.3 Step-3: Extraction of Time Bound Failure Signatures (Rules)	110
9.4 Step-4: Validate Rules and Compute Predictability Index	113
9.5 Conclusions and Discussion	113

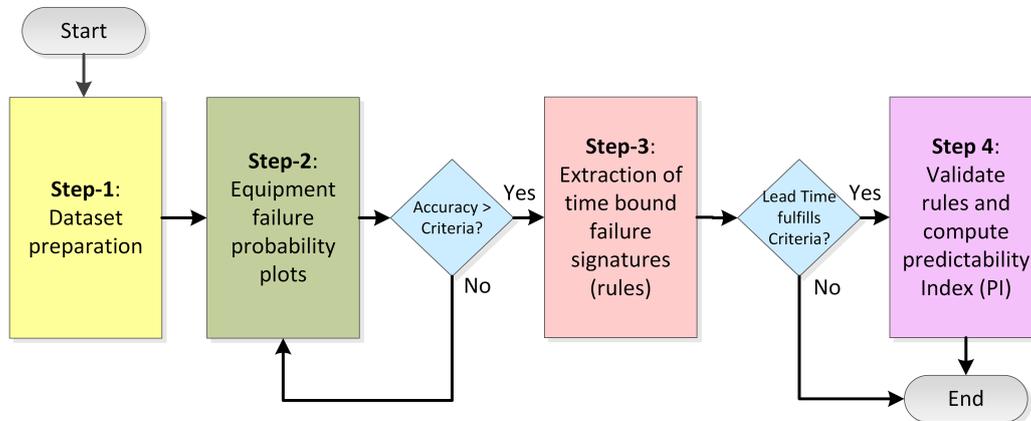


Figure 9.1: Summary of the 4-steps failure signature extraction methodology with respective colours for elaboration.

9.1 Step-1: Dataset Preparation

Each sub-step in dataset preparation step is presented in Figure 9.2.

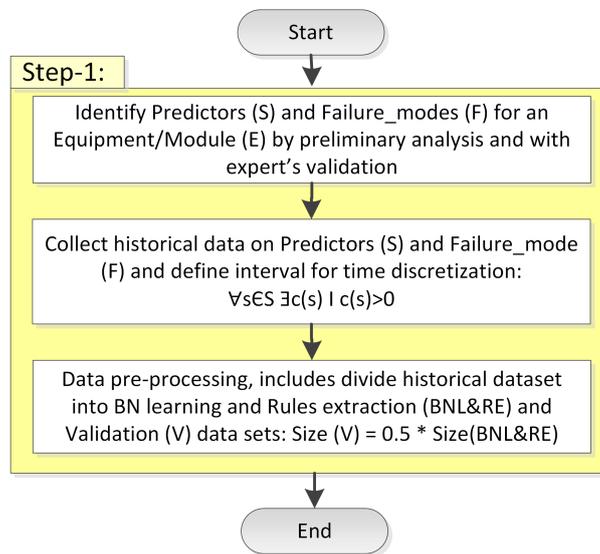


Figure 9.2: Step-1: Identification of Predictors as failure predictors and dataset pre-processing.

Table 9.1: List of identified Predictors and failures for TASMI03_01.

Failure	Product	Process
	No of rework Defect distribution Wait time Previous product combination Current product combination	Current step Previous step Current recipe Previous recipe
	Equipment	Maintenance
Failure_a Failure_b Failure_c no_failure	Previous state No of alarms No of warnings MTTR Up Productive	Wafers processed OEE TT_PM_Boat_meter TT_PM_Trap_meter OEE time Cm

9.1.1 Identification and Selection of Variables

This sub-step is based on the predictors identification step as elaborated in chapter 8, subsection 8.1.1. However, re-identification of predictors is done at equipment and module levels following the results of BN failure/cause diagnosis models (subsection 8.2.2) where certain predictors are found to have zero influence on the target node and are thus omitted from use with experts opinion. Further revision of nodes are also proposed by the experts following results of the methodology. In case of TASMI03_01, the total remaining predictors with conditional dependencies to the target nodes are 22 predictors (from initial 34) and are related to 3 significant modes of failure (a) ElevatorBoatRotation (Bearing) (b) GazPanel (Gas flow) and (c) Out of control (OC) (LOTS and NPW). The predictors and the three failure modes are presented in Table 9.1 as *Failure_a*, *Failure_b*, *Failure_c*, respectively, and are organized under respective data sources. Moreover, the *no_failure* state is modelled as the fourth failure mode.

9.1.2 Interval Discretization

After the identification of predictors, data intervals are aggregated such that we have atleast one value for all predictors. The objective is to integrally evaluate the association between predictors (from multiple data sources) and equipment failure type using BN. As a proof of concept, we aggregate predictors with multiple values in the data interval using mean and mode for continuous and discrete values, respectively. Some predictors values are updated either upon the occurrence of failure e.g. MTTR, failure mode; there-

fore, these are kept constant between respective failures for all data intervals. The data interval rules are given in the Figure 9.3.

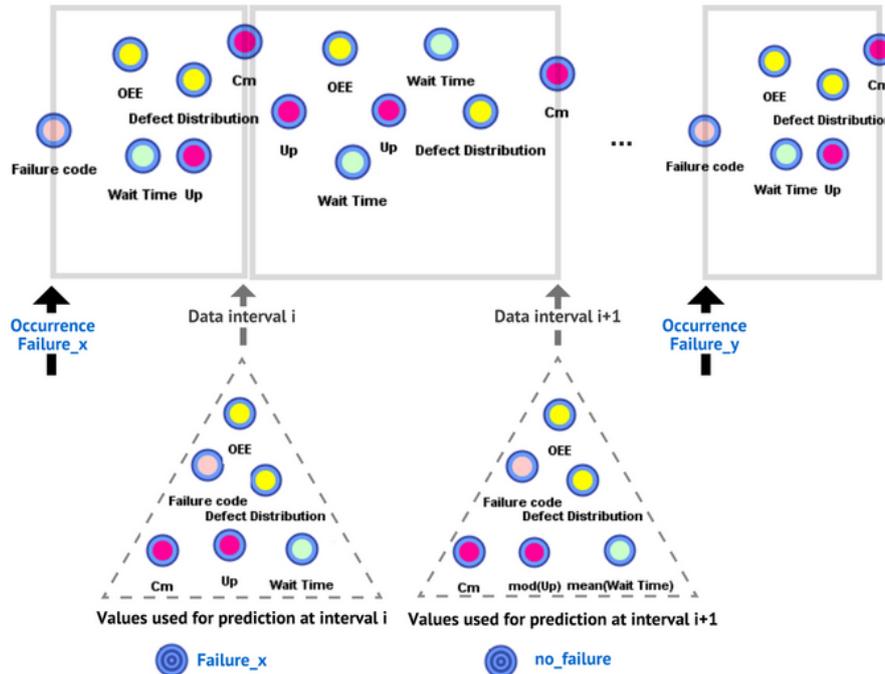


Figure 9.3: Data interval discretization rules.

9.1.3 Dataset Pre-Processing

In this sub-step, data is transformed into a single dataset which is used in BN modelling and patterns/rules extraction. Given historical data in the span of 10 months, we generated 6300 data intervals, equivalent to 82 occurrences of failure. The intervals are not uniform and ranges from few minutes to several hours. This newly constructed dataset are evenly split initially in 2 parts, Bayesian Network Learning & Rule Extraction (BNL&RE) and Validation set (V).

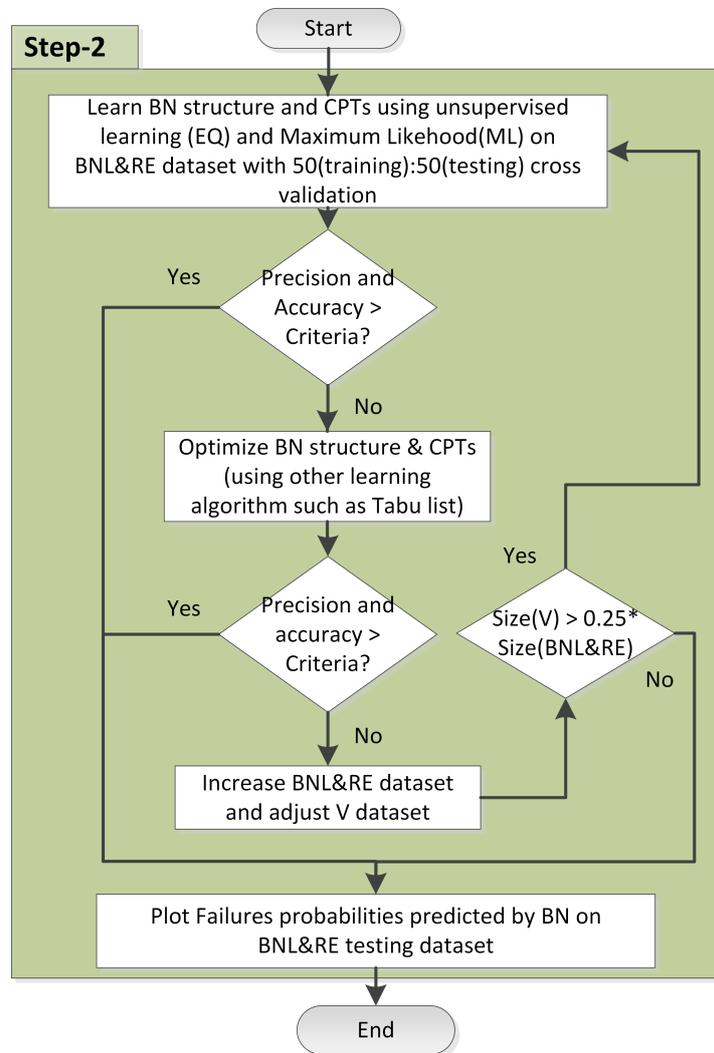


Figure 9.4: Step-2: Unsupervised learning and optimization of Bayesian Network.

9.2 Step-2: Equipment Failure Probability Plots

9.2.1 Learn/Optimize Bayesian Network Structure and compute CPT

In the proposed methodology, this step (see Figure 9.4) consist of learning and computation of BN structure and conditional probability tables. The BN structure is learned and optimized using BayesiaLab 5.3 and based on

Minimum Description Length (MDL). The criteria for model validation is set for reliability and precision to be larger than 95%. The structure is first learned using the Equivalence Class (EQ), a heuristic algorithm to search for highest scoring network in a reduced space of potential BN structures that have same conditional independence relations (Munteanu and Bendou, 2001; Chickering, 2002). Moreover, Maximum Likelihood (ML) is used to estimate Conditional Probability Tables (CPT). The precision and accuracy criteria are computed and compared to the validation criteria. If the first attempt does not fulfill the criteria, the BN structure is further optimized using Tabu algorithm (Glover, 1986; Acid and de Campos, 2003; Teyssier and Koller, 2012). The model with the lowest MDL score is accepted for further analysis.

The BN structure learning and optimization, and CPT computations are repeated until we get the given prediction accuracy and accordingly the BNL&RE and V data sets are adjusted. As a result, the BN model (Figure 9.5) was finally obtained with average accuracy of 97.2% and a final ratio of BNL&RE to V equals to 62:38. The predictors, in this model, are differentiated with different colors following their type of data sources. The light brown, green, orange and pink colors represent product, process, equipment and maintenance related predictors respectively whereas failure and failure mode are the target nodes. In the figure, it is also highlighted that the first set of predictors are removed based on our results from the integrated drift source and failure/cause methodology, section 8.2.2 (in green frame). The predictors removed by the experts before the learning process in this methodology are presented in the red frame. The final network is used therefore to plot failure probabilities on testing dataset on the discretized data intervals for each failure mode. These graphs are the input to next step.

9.2.2 Plot Failure Probability from Bayesian Network for each Discretized Interval

The resulting BN is employed to infer failure probabilities in the BNL&RE testing dataset upon observations (1890 data intervals with 24 failure occurrences). The probabilities for each failure can be distinguished by their respective colors in Figure 9.6. The occurrences can be distinguished from the set of points with high probabilities. For elaboration, refer to example on Figure 9.7 with observation of *Failure_a* (Blue curve) for data interval 3487. The failure is characterized by its high probability (0.64) compared to average probability of *Failure_a* which is 0.18, whereas *no_failure* probability drops to 0.2.

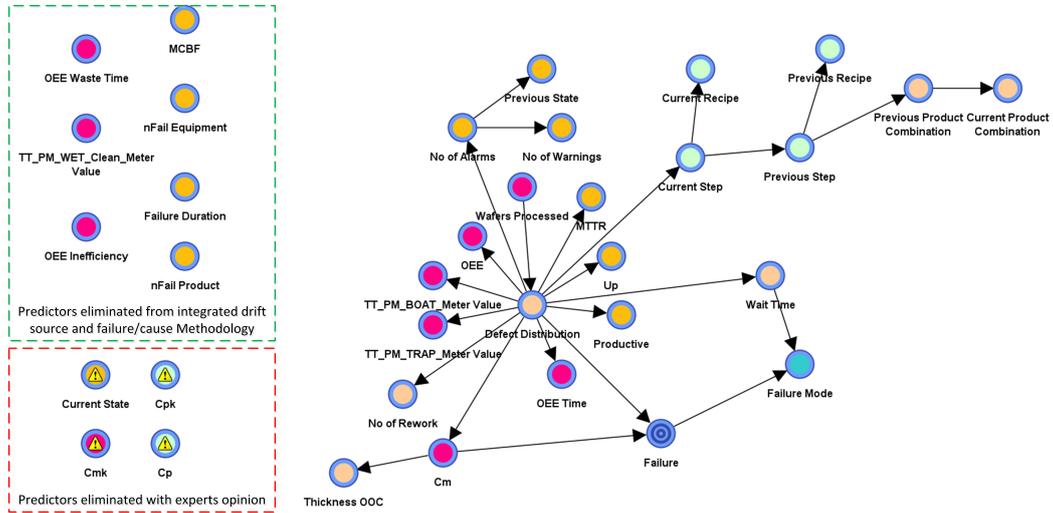


Figure 9.5: BN model for failure probabilities plot.

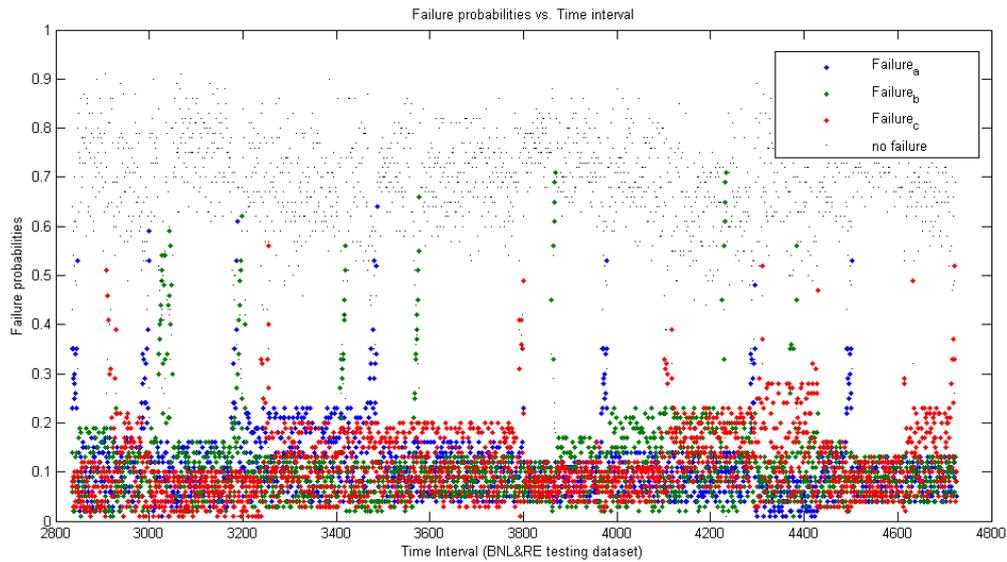


Figure 9.6: Failure inferences in BNL&RE testing dataset, data interval [2835,4725].

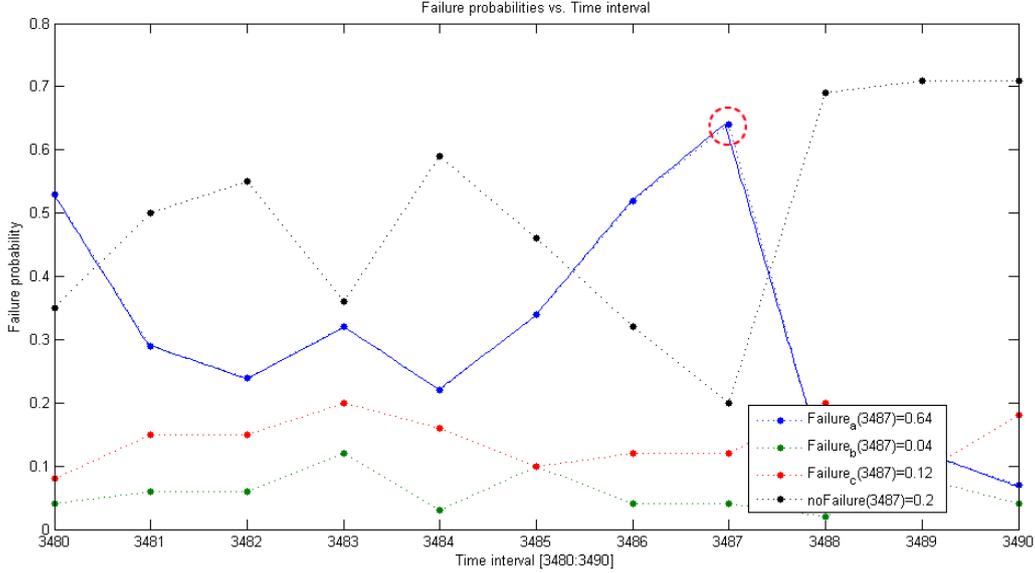


Figure 9.7: Failure inferences in BNL&RE testing dataset, data interval [3480,3490].

9.3 Step-3: Extraction of Time Bound Failure Signatures (Rules)

Plotted graphs provided in Figure 9.6 plays a pivotal role in this third step to define rules for detection of failure before its occurrence. For the application in this study case, we propose a scheme to extract the rules (Figure 9.8).

To extract the patterns/rules, first, based on the probabilities plot and observation of each failure, detect the existence of a pattern. If it exists, define predictive boundaries common to all of respective failure occurrences for pattern P_i . The limits for this region are selected as $min(probability)$ and $max(probability)$, observed among all failure occurrences. However, in case of no pattern, the boundaries are defined first with min and max limits. Instead of defining the lower limit as $min(probability)$, it is defined as $max(mean, mod, median\ of\ overall\ failure\ probabilities)$ because in search of rules without existence of obvious pattern, the focus is to look for the higher probabilities which potentially result in the reduced performance in the equipment. This follows the computation of existing min and max number of consecutive points within the predictive boundaries and the cardinality of $|\min;max|$ of the consecutive points translated into the the number of rules extracted for each failure. This step concludes with the prediction of failure using extracted

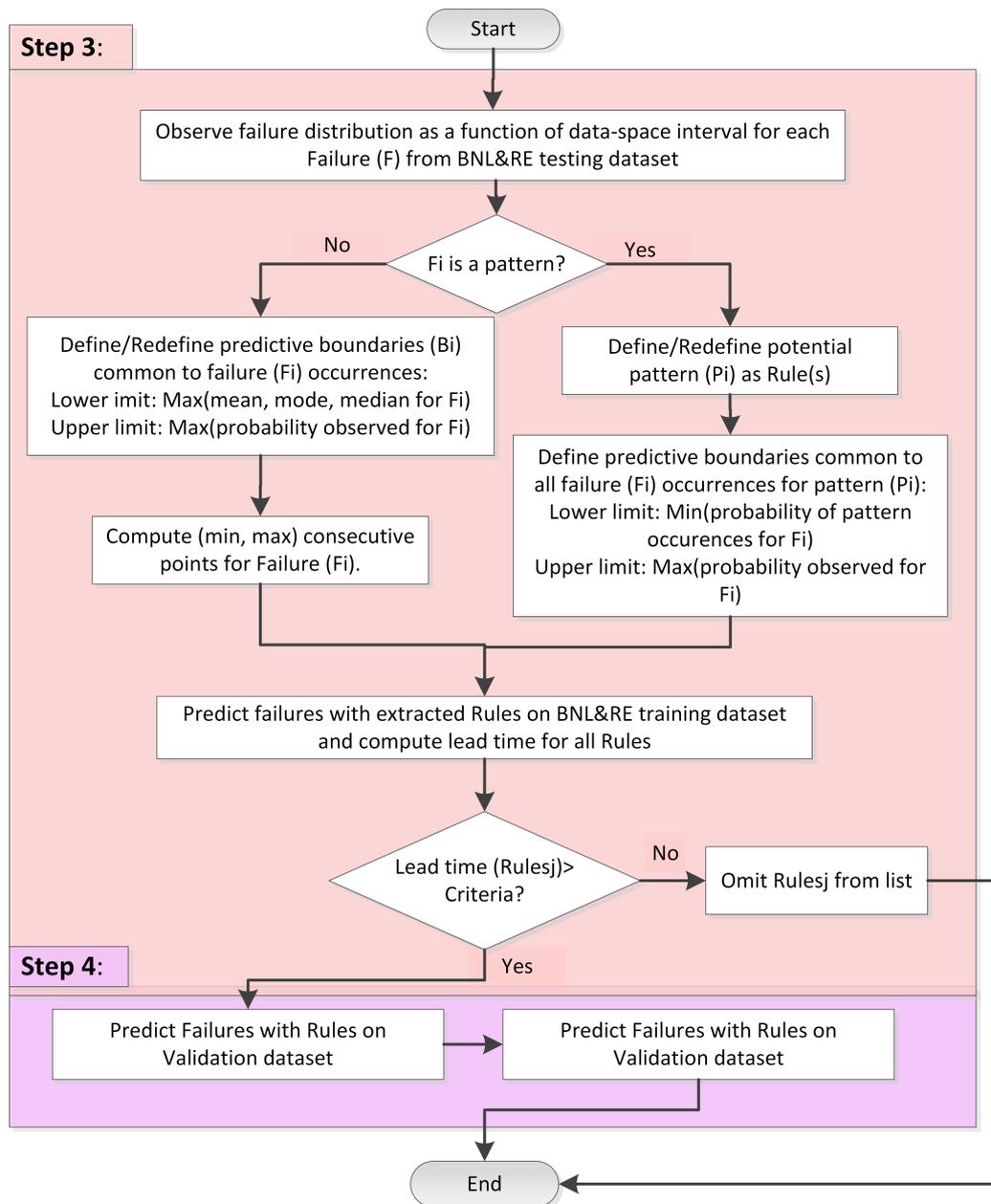


Figure 9.8: Step-3&4: Extraction of rules from each failure probability plot.

rule(s) on the training BNL&RE dataset and the computation of lead time for each prediction. Rules respecting the lead time are reserved for step 4, omitted otherwise.

As a result, patterns have been detected for occurrences of *Failure_b* and

Failure_c. *Failure_b* is continuous increasing probabilities and *Failure_c* is a W or M point-to-point pattern with consecutive values are at least 5 times smaller or bigger. Next, predictive boundaries common to all occurrences of each failure are defined. For *Failure_a*, no pattern could be extracted and predictive boundaries of all occurrences for this failure needs to be established first before assigning the rules of consecutive number of points. The probability plots for each failure in chosen data intervals with the count of occurrences is presented in brackets (refer Figure 9.9). In the graphs we can spot the horizontal lines showing the upper and lower limit of the predictive boundaries whereas in brown boxes are the patterns associated to defined rules. The Table 9.2 summarizes our results for extracted rules and their respective predictive boundaries. Column rule(s) of the table describes the extracted rules. For example, *Failure_a* is of non pattern type with rules as [Min=2; Max=22] consecutive points inside the PB. In the end, we have (22-2+1) rules corresponding to the *Failure_a*. The *Failure_b*, is a pattern based rule as sequentially increasing probabilities and [Min=3; Max=11] consecutive points. The *Failure_b* has (11-3+1) rules. The same applies to *Failure_c* which is of type pattern. These pattern/no pattern rules are further used to compute early and late predictions (hours, minutes, seconds) for each failure such that maintenance teams can employ suitable rules to ensure the execution of respective proactive actions.

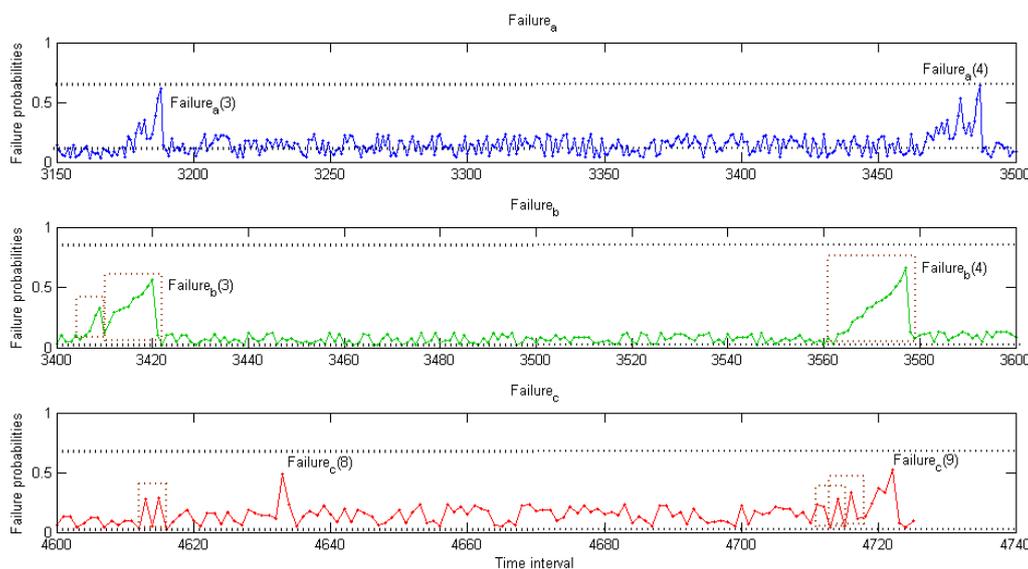


Figure 9.9: Proof of concept for rule(s) extraction on failure.

Table 9.2: Summary of results from rule(s) extraction.

Failure	Base of Rule	Predictive Boundaries (PB)		Rule(s)
		Lower limit	Upper limit	
Failure.a	Non-pattern	max(mean)=0.175	max(probability observed for the failure)=0.64	1) [Min=2;Max=22] consecutive points inside the PB.
Failure.b	Pattern	min(probability of failure occurrence)=0.08	max(probability observed for the failure)=0.71	1) Sequentially increasing probabilities &, 2) [Min=3;Max=11] consecutive points inside the PB following rule 1.
Failure.c	Pattern	min(probability of failure occurrence)=0.04	max(probability observed for the failure)=0.57	1) M/W Pattern &, 2) $\forall 2$ consecutive points of the pattern, $P(t)=\text{factor of } 5^* - P(t+1) -$

Next, prediction of failures using rule(s) plus computation of lead time, which will be categorized as early and/or late detection are made on the BNL&RE training dataset. Table 9.3 presents some of the rules with their associated earliest and latest lead time, while rules without late detection are those that appear only once in between failure intervals. *Failure.c* made 15 predictions, but 2 missed failure occurrences are recorded.

As the final sub-step of this step, the average lead time of each rule are compared to the acceptance interval defined by the users, and rules are preserved for the next step if their average lead time falls into the interval. However, in this study case, this final sub-step could not be completed with users opinion.

9.4 Step-4: Validate Rules and Compute Predictability Index

Step-4 (Figure 9.8) is composed of validation of accepted rules from Step-3 using Validation (V) dataset and of computation of PI. Due to the absence of users criteria in final sub-step of Step-3, Predictability Index (PI) was executed for all patterns/rules. Some results are presented using five selected rules in Figure 9.10. High indices are obtained for these selected rules, responding to average high prediction accuracy, reliability and lead time gain.

9.5 Conclusions and Discussion

There are several explanations on the observations from Table 9.3. *Failure.a*, rule 2 consecutive points in PB has a high number of predictions and big early detection. These are not very significant and it is due to the choice

Table 9.3: Example of results and information for rule(s) testing.

Failure	PB	Total Occ.	Rule(s)	Total Pred.	Earliest detection [min...max]	Latest detection [min...max]
Failure_a	[0.175;0.64]	7	2 cons. Points in PB	47	[7h28m ... 513h15m]	[0h02m ... 0h26m]
			4 cons. Points in PB	12	[3h26m ... 478h37m]	[1h46m ... 10h30m]
			10 cons. Points in PB	6	[1h10m ... 5h09m]	
			22 cons. Points in PB	3	[0h28m ... 2h56m]	
Failure_b	[0.08;0.71]	10	Rising & 3 cons. Points in PB	13	[35h08m ... 135h33m]	[4h22m ... 5h30m]
			Rising & 5 cons. Points in PB	10	[5h33m ... 125h37m]	[1h29m ... 3h45m]
			Rising & 11 cons. Points in PB	7	[0h39m ... 1h48]	
Failure_c	[0.04;0.57]	13	Min Factor 5 W/M Pattern in PB	15 (+ missed prediction)	[7h32m ... 78h55]	[4h22m ... 5h30m]

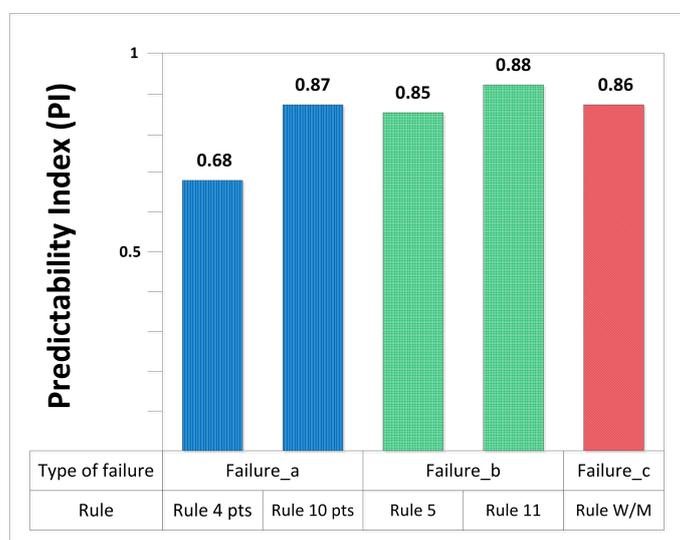


Figure 9.10: Predictability Index for selected rules.

of 2 consecutive points which appears frequently. An increase in lower limit is recommendable. Another significant remark is the missed predictions on *Failure_c* pattern. Even though failure detections are superior than the total

existing failures occurrences, missed predictions still appear and we argue that refinement on rules is required especially because *Failure_c* is related to out of control situation (unknown failure).

Figure 9.11 shows the proof of concept to potential causes behind high probability estimation of *Failure_a* (the disorientation of bearing supporting Elevator Boat Rotation in the equipment). In the proof of concept, only 7 out of 22 predictors are showed, with Defect Distribution, No of Rework and No of Alarms take the minimum values. Further diagnosis and corrective actions are meanwhile pointed to Cm (with low equipment capability), Wait Time (in between 1 and 2 hours) and the combination of Previous and Current Step (as RTO_ISSG and DEP_TEOS respectively). Each of these cause requires certain diagnosis and corrective actions while each action has its own action time. If we want to avoid *Failure_a* coming from this combination of predictors, the required warning time to be set by users as criteria is minimum equal to the cumulation of estimated diagnosis and corrective action times.

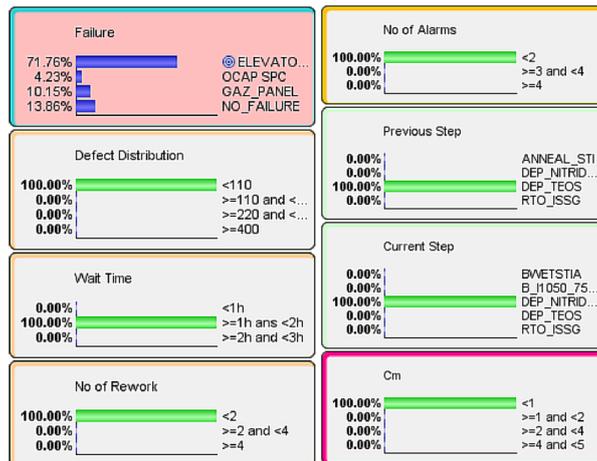


Figure 9.11: Proof of concept for identification of potential causes linked to *failure_a*.

The goals of any predictive analysis are to eliminate unplanned maintenance events thereby improving equipment predictability, and to control costs. This includes advanced warning of impending or imminent failure with enough confidence to plan response and high feasibility to act for a given period of time. The concepts of early and late prediction are important to find the balance of under and over engineering. On one hand, early maintenance and repair can eliminate potential unscheduled time caused by

equipment failure because early diagnosis and appropriate actions plans can be activated. On the other hand, early interventions can cause unnecessary corresponding cost of maintenance and resources. This is left to the choice of end user who can better judge the type of failure and associated repair duration to ultimately decide the target warning time for early failure prediction.

The acceptance of the patterns/rules are for the moment provided by the PI, a presentation of average of average from prediction accuracy and precision plus lead time percentage. This evaluation is good enough as this methodology provides the failure signatures to predict the state of system with n-intervals ahead (fixed horizon), unlike the approach with remaining useful life estimation where the assessment is exposed to different accuracies and precision at different prognostic horizon. However, other metrics and a more rounded PI is to be developed to provide the needed confidence and to assure real time prediction reliability.

Conclusions and Future Works

Conclusion

The works that have been presented in this thesis deals with the reduction of CM and PM in the context of high-mix low volume production. They take place within the European project INTEGRATE and have been developed in close collaboration with our industrial partner STMicroelectronics Crolles. They are initiated through works made by Bouaziz (2012) in this field, within the GCSP¹ and SIREP² teams of the G-SCOP³ laboratory

The contribution of our work are doubled. It has not only focused on the two proposed methodologies to apprehend research gaps and industrial problems, but also the identification of the following specificities:

- Use of event-based data/contextual information collected from product, process, equipment and maintenance data sources in BN models to avoid using sensors data which are exposed to biases,
- Work at equipment and modules decision level instead of focusing at the component and system level, and
- Use of BN as the probabilistic modelling approach to complement the widely used statistical approaches in industry.

The integrated drift source and failure/cause diagnosis methodology aims to reduce failure breakdowns due to misdiagnosis in existing maintenance practice. It uses BN at two levels of diagnosis. The methodology has several functions, each with their own purposes and can be used independently. However as a whole, one of its strength is the use of a single identification of predictors but which can be used in both levels.

1. In the first level, a BN model of product quality drift source as a function of event-based data is proposed. This step provides the decision support against equipment stoppage if equipment is not identified as the drift source.

¹Gestion et Conduite des Systèmes de Production-FR, Production System and Operation Management-EN

²Système d'Information, conception Robuste des Produits-FR, Information Systems and Multiple Product Representations-EN

³Grenoble-Sciences pour la conception, l'Optimisation et la Production-FR, Grenoble-science for the design, optimization and the production-EN

2. If equipment is identified as the failure source, further diagnosis is performed as integrated failure (localisation), failure mode and causes diagnosis. The diagnosis is modelled at the module level and at the equipment level using BN.
3. Finally, a comparison of performance from module levels and equipment level BN models are made to provide decision on which level integrated failure-cause diagnosis shall be modelled. Evidence using one of the TT equipment case study provides module level with the more accurate results. The results from this step can be benefited as input for the second methodology.

In comparison to existing approaches to learn and extract failure signatures, the time bound failure signatures extraction methodology offers extraction, selection and validation of rules/patterns which is linked to sufficient time (time bound) to execute corrective and proactive measures to avoid failures. It uses BN for equipment failure mode probability estimations and their plots. The plots are then use to extract rules/patterns, follows by computation of lead-time and PI to validate their use in subsequent online failure prediction.

1. The event-based predictors, each with their own irregularities in the scale of time are temporally characterized using data intervals for the prediction of failures prior to their occurrence. A scheme for allocation of available historical dataset into several datasets for learning, testing and extracted failure signatures validation is also provided.
2. A BN model is then constructed to its advantage using the predictors of different nature coming from multiple data sources in a single prediction model. The target node is set as the failure (localisation), but the failure mode node is observed to plot failure modes probabilities. Each failure mode and a no_failure are modelled as different states in the node, so different failure signatures, unique to each type of failure can be extracted.
3. Each of the failure mode's failure probability plot are used to extract rules and patterns as failure signatures using experts' opinion or machine learning algorithms. This are complemented with the key action of selecting rules/patterns with a time bound criteria which ensure the sufficient time for corrective and proactive reactions and measures to avoid failures.
4. Selected rules/patterns are further explored and validated using Predictability Index, an index which measure the robustness of each rule and pattern to be used in subsequent online failure prediction. The

final extracted rules/patterns are expected to have several abilities in online failure prediction such as the ability to detect impending or incipient equipment drift, to detect a failure early enough, to distinguish the failure type from one another and to have a low false alarm rate.

All steps in the methodologies have been tested and validated using one Thermal Treatment (TT) equipment (TASMI03) and its module(s) from a reputed semiconductor manufacturer.

Future works

At the end of these works conducted during three years, several considerations emerged to be considered as research perspectives. Four areas of immediate investigations may be considered,

- The methodologies have been tested using a single case study, with different amount of data (6 and 10 months). Within the project, another dataset from another workshop of 12 months have been acquired, but without experts opinion on the choice of predictors. The validation of the two methodologies for this equipment is the first consideration for future work. Similar target is the validation to similar type of equipments (with common failure behaviours) in the TT workshop.
- Secondly, the methodologies involves experts opinion and knowledge elicitation on several parts; the identification of predictors and their states, provision of accepted reliability and precision of BN model, identification of rules/patterns from probability plots and provision of time bound criteria (e.g. warning-time). Each of these task raise interest in exploration of machine learning algorithms to replace or complements experts' opinions. Comparison of different approaches and techniques are also in our interest.
- Thirdly, in the time bound failure signatures extraction methodology, data intervals having one occurrence for all predictors are proposed. This has resulted in encouraging extracted rules/patterns. However, it is in our interest to see and compare this configuration with other proposition of data intervals definition and see their impact on the Predictability Index performance.
- Fourthly, methodologies have been proposed using event based contextual information, it is desirable to see the impact of the methodologies by integrating selective FDC sensors data (with least risk of biases) as predictors.

In the medium term, we can highlight at least 4 directions of research,

- First of all, in the modelling of integrated failure-cause diagnosis, we have only taken into account the three main modules of TT equipment, when in reality it has 8 modules which are in a 'parent-child' relationship. Consideration of the whole modules with the relationship (structural and organizational) shall be considered.
- Secondly, for the proposed PI (time bound failure signature extraction methodology), it is a presentation of average of average from prediction accuracy, precision and lead-time percentage. This is only a first proposition of PI. We proposed to explore other metrics depending on requirements to assure real time prediction reliability. This is important as in failure prediction, assessment are exposed to different accuracies and precision at earlier and imminent stages of failure.
- We make the assumption that if probability of a failure mode at a given interval is superior to a certain level and to all other types of failure, including the no_failure, then the associated failure mode is attributed as the type of equipment failure. Moreover, occurrence of a failure can also be attributed depending on the consistency in failure probability plots where rules/patterns are extracted, separately, for each failure mode. The third consideration is to take this assumption out and handle the confusion if multiple failure signatures are detected in a common interval.
- With the time bound failure signatures, we can get 'trapped' very easily in over engineering situation. A study to dynamically revise the extracted signatures to minimize false alarm is highly recommendable.

Finally, in the longer term, consideration should be given to the study of,

- Reroute of product to avoid equipments which are predicted as having imminent failure using the time bound failure signature.
- compatibility of the proposed methodologies in another context of application such as the automotive, nuclear, hospital and energy distribution.

APPENDIX A

Bayesian Network Learning

A.1 Notation

Table A.1: Notation

B	The network, $B = (G, \Theta)$
D	The data
G	The DAG, $G = (V, E)$
E	The edges of G
V	The list/set of vertices (Nodes) in the network B
n	the number of vertices in the network B
Θ	$\Theta = \Theta_1, \dots, \Theta_n$
Θ_i	$\Theta_i = P(X_i P_a(X_i))$
X_i	The nodes in V
q_i	number of possible configurations of the parent set $\prod X_i$ of X_i
r_i	number of states of the finite random variable X_i
N_{ijk}	number of instances in the data D where the variable X_i takes its k^{th} value x_{ik} and the variables in $\prod X_i$ take their j^{th} configuration w_{ij}
N_{ij}	number of instances in the data T where the variables in $\prod X_i$ take their j^{th} configuration w_{ij}
N	total number of instances in the data D
H	A list of neighbours DAGs

A.2 Optimize Structure Learning using Equivalence Class (EQ) and Tabu List

The Figure A.1 presents the flow of implementing the optimized structure learning using EQ and Tabu list. The algorithm starts by ordering all the BN edges using Algorithm 1. Elaboration on Maximum Description Length (MDL) as an example of scoring function and which is used in the thesis is provided in section A.3.

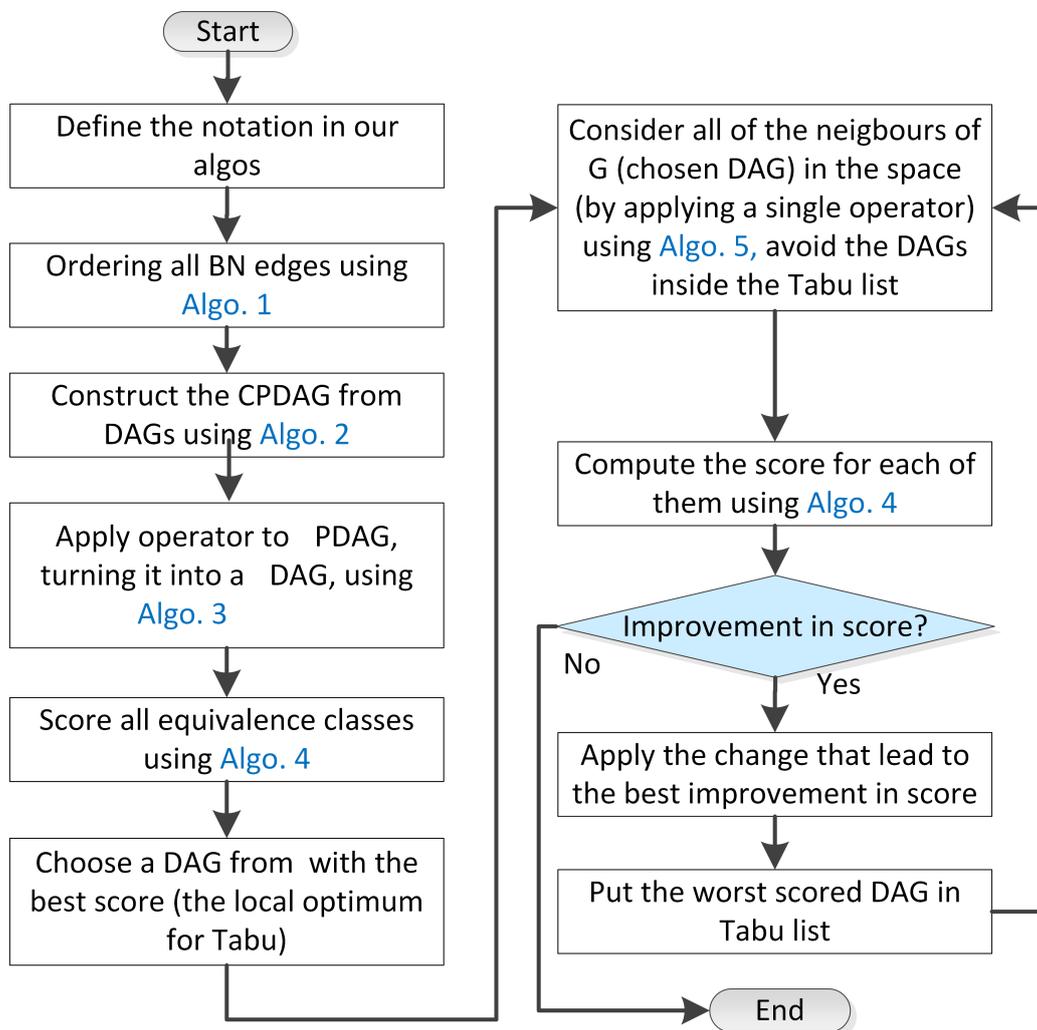


Figure A.1: Flow heart for EQ+Tabu list search strategy.

Algorithm 1 Ordering

Require: E : list of edges, V : List of nodes

Ensure: $E_{ordered}$

- 1: $Topological_Tree(X_i)$ ▷ $X_i \in V$
 - 2: $K \leftarrow 0$
 - 3: $A \leftarrow unordered(E)$
 - 4: **repeat**
 - 5: $X_j \leftarrow min_j(X_j / (X_i, X_j) \in A)$ ▷ Lowest destination node of an
unordered edge
 - 6: $X_i \leftarrow max_i(X_i / (X_i, X_j) \in A)$ ▷ Greatest source node of an
unordered edge to X_j
 - 7: $order(X_i, X_j) \leftarrow k$
 - 8: $A \leftarrow A \setminus (X_i, X_j)$
 - 9: **until** $A = \emptyset$
-

Algorithm 2 Equivalence class algorithm (DAG to CPDAG)

Require: G : a DAG

Ensure: $CPDAG_G$: The Completed Partially Directed Acyclic Graph of the DAG

```

1: Order ( $E$ )                                ▷ The set of directed edges of  $G$ 
2:  $\forall e \in E, label(e) \leftarrow \emptyset$ 
3:  $A \leftarrow unlabeled(E)$ 
4: repeat
5:    $(X_i, X_j) \leftarrow min_A(e)$            ▷ Lowest unlabeled edge
6:    $\forall K_k / label(X_k, X_i) \leftarrow irreversible$ 
7:    $End \leftarrow False$ 
8:   if  $X_k \notin paX_j$  then
9:      $label(*, X_j) \leftarrow irreversible$ 
10:     $A \leftarrow A \setminus (*, X_j)$ 
11:     $End \leftarrow True$ 
12:   else
13:      $label(X_k, X_j) \leftarrow irreversible$ 
14:      $A \leftarrow A \setminus (X_k, X_j)$ 
15:   end if
16:   if  $End = False$  then
17:     if  $\exists e(X_k, X_j) / X_k \notin pa(X_i) \cup X_i$  then
18:        $\forall (X_k, X_j) \in A$ 
19:        $label(X_k, X_j) \leftarrow irreversible$ 
20:        $A \leftarrow A \setminus (X_k, X_j)$ 
21:     else
22:        $\forall (X_k, X_j) \in A$ 
23:        $label(X_k, X_j) \leftarrow reversible$ 
24:        $A \leftarrow A \setminus (X_k, X_j)$ 
25:     end if
26:   end if
27: until  $A = \emptyset$ 

```

Algorithm 3 Algorithm (PDAG to DAG)

Require: $PDAG$

Ensure: DAG

```

1:  $B \leftarrow PDAG$ 
2:  $A \leftarrow$  list of arcs of PDAG ▷ Not sure??
3: repeat
4:   Find node  $X_i$  such that
5:   - No existence of  $X_i \leftarrow X_j$  in  $A$ 
6:   - and for all  $X_j$  such that  $X_i - X_j$  in  $A$ ,
7:   -  $X_j$  is adjacent to all other adjacent nodes of  $X_i$ 
8:   if  $X_i$  does not exist then
9:      $PDAG$  does not admit any extension to completely directed
10:  else
11:     $\forall X_j$  such that  $X_i - X_j \in A$ 
12:     $X_i \leftarrow X_j$  in  $B$ 
13:     $A \leftarrow A \setminus (X_i, X_j)$ 
14:  end if
15: until  $A = \emptyset$ 

```

Algorithm 4 Generate MDL score

Require: G : DAG

Ensure: H : MDL score ▷ cannot be provided, refer section A.3

Algorithm 5 Generate neighborhood

Require: G : a BN structure

Ensure: H : a list of neighbours DAGs

```

1:  $H \leftarrow \emptyset$ 
2: for all  $e \in G$  do
3:    $H \leftarrow H \cup (G \setminus \{e\})$  ▷ delete_edge(e)
4:   if acyclic ( $G \setminus \{e\} \cup invert(e)$ ) then
5:      $H \leftarrow (H \cup \{e\} \cup invert(e))$  ▷ invert_edge(e)
6:   end if
7: end for
8: for all  $e \notin G$  do
9:   if acyclic ( $G \cup \{e\}$ ) then
10:     $H \leftarrow H \cup (G \cup \{e\})$  ▷ add_edge(e)
11:   end if
12: end for

```

A.3 Minimum Description Length (MDL) Score

The Minimum Description Length (MDL) score (Lam and Bacchus, 1994) is a two-component score, which has to be minimized to obtain the best solution. In our machine-learning application, the “model” is a BN, consisting of a graph and probability tables. More formally, we write this score as:

$$MDL(B | D) = LL(B | D) - \frac{1}{2} \log(N) | B | \quad (\text{A.1})$$

where:

$$LL(B | D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log\left(\frac{N_{ijk}}{N_{ij}}\right) \quad (\text{A.2})$$

and

$$| B | = \sum_{i=1}^n (r_i - 1) q_i \quad (\text{A.3})$$

The log-likelihood (LL) score tends to favour complete network structures and it does not provide a useful representation of the independence assumptions of the learned network. The over fitting phenomena which might occur in LL score is trade off with the second MDL term where $| B |$ denoted complexity. Minimum value for this second term is obtained with the simplest structure, i.e. the fully unconnected network, in which all variables are stated as independent. The minimum value for the first term, is obtained with the fully connected network, i.e. a network corresponding to the analytical form of the joint probability distribution, in which no structural independences are stated. Thus, minimizing this score consists in finding the best trade-off between both terms. For a learning algorithm that starts with an unconnected network, the objective is to add a link for representing a probabilistic relationship if, and only if, this relationship reduces the log-likelihood of the data, i.e. $LL(B | D)$, by a large enough amount to compensate for the increase in the size of the network representation, i.e. $\frac{1}{2} \log(N) | B |$.

The INTEGRATE PROJECT

B.1 Introduction

European IC manufacturers are on the way to become suppliers of customized products for OEM companies, which means high flexibility and product diversity. The INTEGRATE¹ project is initiated to conduct research to support the development of new manufacturing procedures, new organizations and new information and control tools to enable IC production lines to efficiently manage a high product and technology mix and heterogeneous lots of different size and priority for development, engineering and prototyping.

This project was preceded by the IMPROVE² allowed the development of algorithms to make the move from 'reactive to predictive operations'. INTEGRATE goes further by addressing the needed improvements of the high-mix production environment allowing putting the pieces together and lifting them to the fab level. INTEGRATE will therefore especially focus on: factory operation methodologies, data acquisition and analysis concepts, factory information and control system, and process data analysis from heterogeneous samples. The objective is to develop procedures that can be implemented on existing fab information and control system. More information can be found on the internet website of the project³.

B.1.1 Project management

The INTEGRATE consortium consists in 26 partners and is composed of: 6 European semiconductor major companies, 5 SME's of solution providers, 2 equipment suppliers and 13 academic partners covering a wide range of competencies and high level expertise in statistics, data analysis and manufacturing science. To reach its objective, INTEGRATE addresses different levels of fab organization and managements Figure B.1 and is organized into five work packages (WP):

¹INTEGRATE: Solutions for Agile Manufacturing in High-mix Semiconductor Fabs

²IMPROVE: Implementing Manufacturing science solutions to increase equipment productivity and fab performance.

³<http://www.eniac-integrate.eu/>

- WP2: the optimization of the equipment usage, taking especially into account (first level) the EHF (Equipment Health Factor developed in the IMPROVE project).
- WP4: the control of the lot flow with the idea to optimize the lots trajectories with respect of the actual real time status of the production equipment (second level)
- WP3 the optimization of the process control and quality, including the actual use of Virtual Metrology algorithms developed in IMPROVE for introduction of real time wafer to wafer regulation loops (third level)
- WP5: the data analysis and data management developments where the necessary algorithms will be developed and assessed to support the previous three levels. This approach will include an exhaustive modelling of the yield.
- WP1: the collection of feedback from the other WPs based on their experience and refines the generic specifications and architectures.

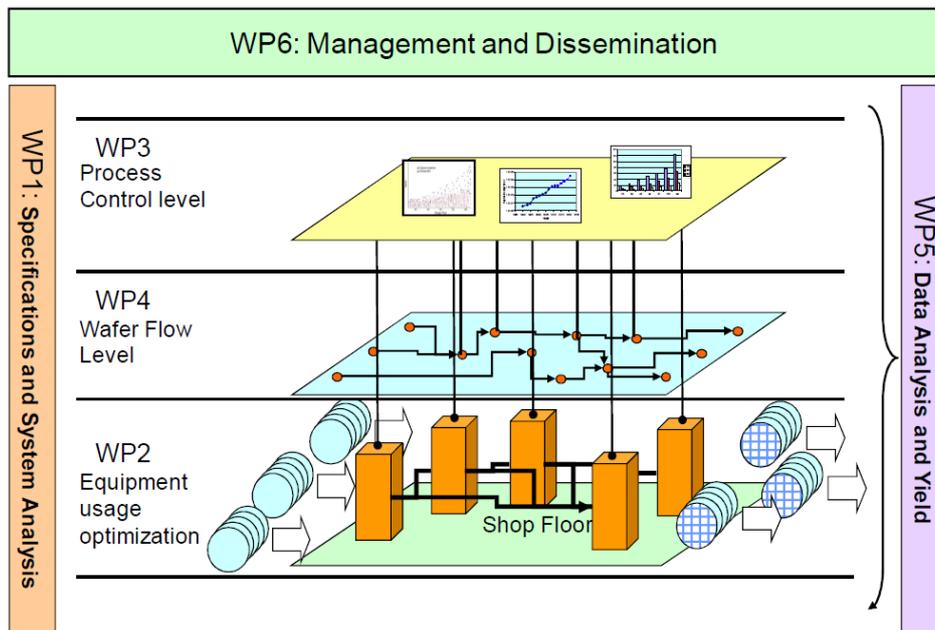


Figure B.1: The workpackages.

In this framework, our works is positioned as the provision of a methodology to extract signature to failure before its occurrence aiming to reduce unavailability of equipment due to corrective maintenance and provide the solution to implement predictive maintenance for optimization of the equip-

ment usage, thus in the WP2. WP2 in particular deals with the three possible efficiency losses mechanism on production equipment, with each of them received its own proper focus: (i) Recipe Management to eliminate waste time in setting up and/or changing recipes, (ii) extra sensor information aggregated in an equipment health index (EHI) to avoid and reduce unscheduled downtime, and (iii) wait-time-waste (WTW) to detect if equipment is not running at golden speed settings. The works done in WP2 allowed to lay down the industrialization and deployment to production of the various smart manufacturing solutions by bringing together industrial mathematics (data analysis and model adaptation), big data IT-solutions (increase data collection, CPU, database storage) and visual manufacturing (GUI-screens, tablets, improved information everywhere). Integrating these new solutions depends on collaboration to share methods and strategies. Our works have been performed with close collaboration in particular with manufacturer STMicroelectronics Crolles.

B.1.2 STMicroelectronics Crolles

STMicroelectronics (ST) is a global semiconductor manufacturer. The company designs, develops, manufactures and market a wide range of IC's and components used in many of the microelectronics and industrial applications. ST's strives to be the leader for delivering solutions key to smart driving and the internet of things. As far as turnover and profit are concerned, the situation in December 1998 differs markedly from what had been, ST is now up to tenth in the ranking of world's leading manufacturers of semiconductors. ST has over 75 sales and marketing offices in over 35 countries and a worldwide network of front-end (wafer fabrication) and back-end (assembly and test and packaging) plants. ST's principal wafer fabs are located in Agrate Brianza and Catania (Italy), Crolles, Rousset, and Tours (France), and in Singapore. These are complemented by assembly-and-test facilities located in China, Malaysia, Malta, Morocco, the Philippines, and Singapore. STMicroelectronics in Crolles has two manufacturing sites, the 200mm and 300mm (names in relation to the size of the fabricated wafers) and it has an installed manufacturing capacity of 10,700 wafers per week.

This dissertation takes place within the wafer processing lines of 300mm production site. Several case studies has been proposed within the framework of this project, and we are particularly working with equipments from thermal treatment (TT) and deposition of dielectric material (DIEL) workshops. For presentation purpose, only application and results from the TT workshop, equipment TASMI01 precisely are presented in this dissertation.

Bibliography

- Abu-Samah, A., Shahzad, M., and Zamai, E. (2016). Bayesian based methodology for the extraction and validation of time bound failure signatures for online failure prediction. *Reliability Engineering & System Safety*.
- Abu-Samah, A., Shahzad, M., Zamai, E., and Said, A. B. (2015a). Failure prediction methodology for improved proactive maintenance using bayesian approach. *IFAC-PapersOnLine*, 48(21):844–851.
- Abu-Samah, A., Shahzad, M. K., Zamai, É., and Hubac, S. (2014). Methodology for integrated failure-cause diagnosis with bayesian approach: application to semiconductor manufacturing equipment. In *Second European Conference of the Prognostics and Health Management Society*.
- Abu-Samah, A., Shahzad, M. K., Zamai, E., and Hubac, S. (2015b). Effective maintenance by reducing failure-cause misdiagnosis in semiconductor industry (si). *PHM Journal*, 6(1)009:18.
- Acid, S. and de Campos, L. M. (2003). Searching for bayesian network structures in the space of restricted acyclic partially directed graphs. *Journal of Artificial Intelligence Research*, pages 445–490.
- Al-Najjar, B. (1996). Total quality maintenance: an approach for continuous reduction in costs of quality products. *Journal of Quality in Maintenance Engineering*, 2(3):4–20.
- Al-Najjar, B. and Alsayouf, I. (2003). Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. *International journal of production economics*, 84(1):85–100.
- Atoui, M. A., Verron, S., and Kobi, A. (2015). Fault detection and diagnosis in a bayesian network classifier incorporating probabilistic boundary 1. *IFAC-PapersOnLine*, 48(21):670–675.
- Blue, J., Roussy, A., Thieullen, A., and Pinaton, J. (2012). Efficient fdc based on hierarchical tool condition monitoring scheme. In *Advanced semiconductor manufacturing conference (asmc), 2012 23rd annual semi*, pages 359–364. IEEE.
- Bouaziz, M. F. (2012). *Contribution à la modélisation Bayésienne de l'état de santé d'un système complexe: application à l'industrie du semi-conducteur*. PhD thesis, Université de Grenoble.

- Bouaziz, M. F., Zamaï, E., and Duvivier, F. (2013). Towards bayesian network methodology for predicting the equipment health factor of complex semiconductor systems. *International Journal of Production Research*, 51(15):4597–4617.
- Castanier, B. (2001). *Modélisation stochastique et optimisation de la maintenance conditionnelle des systèmes à dégradation graduelle*. PhD thesis.
- Chang, H. J., Song, D. S., Kim, P. J., and Choi, J. Y. (2012). Spatiotemporal pattern modeling for fault detection and classification in semiconductor manufacturing. *Semiconductor Manufacturing, IEEE Transactions on*, 25(1):72–82.
- Chen, A. and Blue, J. (2009). Recipe-independent indicator for tool health diagnosis and predictive maintenance. *Semiconductor Manufacturing, IEEE Transactions on*, 22(4):522–535.
- Chiang, L. H., Kotanchek, M. E., and Kordon, A. K. (2004). Fault diagnosis based on fisher discriminant analysis and support vector machines. *Computers & chemical engineering*, 28(8):1389–1401.
- Chiang, L. H., Russell, E. L., and Braatz, R. D. (2000). Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and intelligent laboratory systems*, 50(2):243–252.
- Chickering, D. M. (2002). Learning equivalence classes of bayesian-network structures. *The Journal of Machine Learning Research*, 2:445–498.
- Chickering, D. M., Geiger, D., Heckerman, D., et al. (1994). Learning bayesian networks is np-hard. Technical report, Citeseer.
- Correa, M., Bielza, C., and Pamies-Teixeira, J. (2009). Comparison of bayesian networks and artificial neural networks for quality detection in a machining process. *Expert Systems with Applications*, 36(3):7270–7279.
- Daly, R., Shen, Q., and Aitken, S. (2011). Learning bayesian networks: approaches and issues. *The knowledge engineering review*, 26(02):99–157.
- das Chagas Moura, M., Zio, E., Lins, I. D., and Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(11):1527–1534.

- de Souza e Silva, E. and Ochoa, P. M. (1992). State space exploration in markov models. In *ACM SIGMETRICS Performance Evaluation Review*, volume 20, pages 152–166. ACM.
- Denson, W. (1998). The history of reliability prediction. *Reliability, IEEE Transactions on*, 47(3):SP321–SP328.
- Doty, L. A. (1996). *Statistical process control*. Industrial Press Inc.
- El-Koujok, M., Gouriveau, R., and Zerhouni, N. (2014). Development of a prognostic tool to perform reliability analysis. In *Proc. of the ESREL-17th SRA-Europe Conf., Valencia, Spain, sept. 22*, volume 25, pages 191–199.
- Flinn, M. W. (1974). Trends in real wages, 1750–18501. *The Economic History Review*, 27(3):395–413.
- Friedman, N., Getoor, L., Koller, D., and Pfeffer, A. (1999). Learning probabilistic relational models. In *IJCAI*, volume 99, pages 1300–1309.
- Fronza, I., Sillitti, A., Succi, G., Terho, M., and Vlasenko, J. (2013). Failure prediction based on log files using random indexing and support vector machines. *Journal of Systems and Software*, 86(1):2–11.
- Fulp, E. W., Fink, G. A., and Haack, J. N. (2008). Predicting computer system failures using support vector machines. *WASL*, 8:5–5.
- García, F. P., Pedregal, D. J., and Roberts, C. (2010). Time series methods applied to failure prediction and detection. *Reliability Engineering & System Safety*, 95(6):698–703.
- Gertler, J. (1998). *Fault detection and diagnosis in engineering systems*. CRC press.
- Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & operations research*, 13(5):533–549.
- Gu, X., Papadimitriou, S., Yu, P. S., and Chang, S.-P. (2008). Online failure forecast for fault-tolerant data stream processing. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 1388–1390. IEEE.
- Guo, H. and Hsu, W. (2002). A survey of algorithms for real-time bayesian network inference. In *AAAI/KDD/UAI02 Joint Workshop on Real-Time Decision Support and Diagnosis Systems*. Edmonton, Canada.

- He, Q. P. and Wang, J. (2007). Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *Semiconductor manufacturing, IEEE transactions on*, 20(4):345–354.
- Heckerman, D. (2008). A tutorial on learning with bayesian networks. In *Innovations in Bayesian networks*, pages 33–82. Springer.
- Heng, A., Zhang, S., Tan, A. C., and Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical systems and signal processing*, 23(3):724–739.
- Hohmann, C. (2011). *Techniques de productivité: Comment gagner des points de performance-Pour les managers et les encadrants*. Editions Eyrolles.
- Iravani, S. M., Van Oyen, M. P., and Sims, K. T. (2005). Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Science*, 51(2):151–166.
- Ishikawa, K. (1990). *Introduction to quality control*. Productivity Press.
- Jardine, A. K., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510.
- Kjærulff, U. B. and Madsen, A. L. (2006). Probabilistic networks for practitioners—a guide to construction and analysis of bayesian networks and influence diagrams. *Department of Computer Science, Aalborg University, HUGIN Expert A/S*.
- Kobbacy, K. A., Vadera, S., McNaught, K., and Chan, A. (2011). Bayesian networks in manufacturing. *Journal of Manufacturing Technology Management*, 22(6):734–747.
- Kothamasu, R., Huang, S. H., and VerDuin, W. H. (2009). System health monitoring and prognostics—a review of current paradigms and practices. In *Handbook of Maintenance Management and Engineering*, pages 337–362. Springer.
- Kumar, R. (2008). *Fabless semiconductor implementation*. McGraw-Hill New York.
- Lacaille, J. and Zagrebnoy, M. (2007). An unsupervised diagnosis for process tool fault detection: the flexible golden pattern. *Semiconductor Manufacturing, IEEE Transactions on*, 20(4):355–363.

- Lam, W. and Bacchus, F. (1994). Learning bayesian belief networks: An approach based on the mdl principle. *Computational intelligence*, 10(3):269–293.
- Liang, Y., Zhang, Y., Xiong, H., and Sahoo, R. (2007). Failure prediction in ibm bluegene/l event logs. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 583–588. IEEE.
- Lin, D., Makis, V., et al. (2003). Recursive filters for a partially observable system subject to random failure. *Advances in Applied Probability*, 35(1):207–227.
- Lu, S., Tu, Y.-C., and Lu, H. (2007). Predictive condition-based maintenance for continuously deteriorating systems. *Quality and Reliability Engineering International*, 23(1):71–81.
- Margaritis, D. (2003). *Learning Bayesian network model structure from data*. PhD thesis, US Army.
- Mendez, H., Zamai, E., and Descotes-Genon, B. (2005). A methodology to synthesize monitoring laws by using quality, security, ecology, and productivity criteria. *Quality Engineering*, 17(4):659–668.
- Mili, A. (2009). *Vers des méthodes fiables de contrôle des procédés par la maîtrise du risque: Contribution à la fiabilisation des méthodes de process control d’une unité de Recherche et de Production de circuits semi-conducteurs*. PhD thesis, Institut National Polytechnique de Grenoble-INPG.
- Mobley, R. K. (2002). *An introduction to predictive maintenance*. Butterworth-Heinemann.
- Munteanu, P. and Bendou, M. (2001). The eq framework for learning equivalence classes of bayesian networks. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 417–424. IEEE.
- Naïm, P., Wuillemain, P.-H., Leray, P., Pourret, O., and Becker, A. (2011). *Réseaux bayésiens*. Editions Eyrolles.
- Nguyen, D.-T. (2015). *Diagnostic en ligne des systèmes à événements discrets complexes: approche mixte logique/probabiliste*. PhD thesis, Université Grenoble Alpes.
- Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

- Pehrsson, A. and Al-Najjar, B. (2005). Creation of industrial competitiveness. *Acta Wexionensia*, (69).
- Pérès, F. and Noyes, D. (2003). Evaluation of a maintenance strategy by the analysis of the rate of repair. *Quality and Reliability Engineering International*, 19(2):pp–129.
- Pintelon, L. and Parodi-Herz, A. (2008). Maintenance: an evolutionary perspective. In *Complex system maintenance handbook*, pages 21–48. Springer.
- Qin, S. J., Valle, S., and Piovoso, M. J. (2001). On unifying multiblock analysis with application to decentralized process monitoring. *Journal of chemometrics*, 15(9):715–742.
- Roeder, G., Mattes, A., Pfeffer, M., Schellenberger, M., Pfitzner, L., Knapp, A., Muhlberger, H., Kyek, A., Lenz, B., Frisch, M., et al. (2012). Framework for integration of virtual metrology and predictive maintenance. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2012 23rd Annual SEMI*, pages 288–293. IEEE.
- Said, A. B. (2016). *Gestion dynamique des connaissances de maintenance pour des environnements de production de haute technologie à fort mix produit*. PhD thesis, Université de Grenoble.
- Said, A. B., Shahzad, M. K., Zamai, E., Hubac, S., and Tollenaere, M. (2016). Experts’ knowledge renewal and maintenance actions effectiveness in high-mix low-volume industries, using bayesian approach. *Cognition, Technology & Work*, 18(1):193–213.
- Salfner, F., Lenk, M., and Malek, M. (2010). A survey of online failure prediction methods. *ACM Computing Surveys (CSUR)*, 42(3):10.
- Salfner, F. and Malek, M. (2007). Using hidden semi-markov models for effective online failure prediction. In *Reliable Distributed Systems, 2007. SRDS 2007. 26th IEEE International Symposium on*, pages 161–174. IEEE.
- Salfner, F., Schieschke, M., and Malek, M. (2006). Predicting failures of computer systems: A case study for a telecommunication system. In *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*, pages 8–pp. IEEE.
- Samanta, B. and Al-Balushi, K. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, 17(2):317–328.

- Sarkar, P. (2004). Clustering of event sequences for failure root cause analysis. *Quality engineering*, 16(3):451–460.
- Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, 1(1):4–23.
- Shahzad, M. K. (2012). *Exploitation dynamique des données de production pour améliorer les méthodes DFM dans l’industrie microélectronique*. PhD thesis, GSCOP Laboratory.
- Sikorska, J., Hodkiewicz, M., and Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5):1803–1836.
- Sipos, R., Fradkin, D., Moerchen, F., and Wang, Z. (2014). Log-based predictive maintenance. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1867–1876. ACM.
- Smith, G. M. (1998). *Statistical process control and quality improvement*. Prentice Hall.
- Stamatis, D. H. (2003). *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality Press.
- Susto, G. A., Pampuri, S., Schirru, A., De Nicolao, G., McLoone, S., and Beghi, A. (2012). Automatic control and machine learning for semiconductor manufacturing: Review and challenges. In *Proceedings of the 10th European Workshop on Advanced Control and Diagnosis (ACD 2012)*.
- Teyssier, M. and Koller, D. (2012). Ordering-based search: A simple and effective algorithm for learning bayesian networks. *UAI*, pages 584–590.
- Vasile, O. (2008). Contribution au pronostic de défaillances par réseau neuro-flou: maîtrise de l’erreur de prédiction. *en France*.
- Venkatasubramanian, V., Rengaswamy, R., and Kavuri, S. N. (2003a). A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies. *Computers & Chemical Engineering*, 27(3):313–326.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003b). A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & chemical engineering*, 27(3):327–346.

- Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S. N. (2003c). A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):293–311.
- Verron, S., Li, J., and Tiplica, T. (2010). Fault detection and isolation of faults in a multivariate process with bayesian network. *Journal of Process Control*, 20(8):902–911.
- Vilalta, R. and Ma, S. (2002). Predicting rare events in temporal domains. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 474–481. IEEE.
- Vrignat, P., Avila, M., Duculty, F., and Kratz, F. (2015). Failure event prediction using hidden markov model approaches.
- Wang, W. (2002). A model to predict the residual life of rolling element bearings given monitored condition information to date. *IMA Journal of Management Mathematics*, 13(1):3–16.
- Weber, P., Medina-Oliva, G., Simon, C., and Iung, B. (2012). Overview on bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4):671–682.
- Weidl, G., Madsen, A., and Israelson, S. (2005). Applications of object-oriented bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes. *Computers & chemical engineering*, 29(9):1996–2009.
- Widodo, A. and Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, 21(6):2560–2574.
- Yang, L. and Lee, J. (2012). Bayesian belief network-based approach for diagnostics and prognostics of semiconductor manufacturing systems. *Robotics and Computer-Integrated Manufacturing*, 28(1):66–74.
- Yang, Z. M., Djurdjanovic, D., and Ni, J. (2008). Maintenance scheduling in manufacturing systems based on predicted machine degradation. *Journal of Intelligent Manufacturing*, 19(1):87–98.
- Yoon, M. and Malerba, F. (2010). Technological interrelatedness, knowledge generality and economies of scale in the evolution of firm boundaries: A

- history-friendly model of the fabless ecosystem. In *International Conference on Opening Up Innovation: Strategy, Organization and Technology*.
- Yue, H. H. and Qin, S. J. (2001). Reconstruction-based fault identification using a combined index. *Industrial & engineering chemistry research*, 40(20):4403–4414.
- Yue, H. H. and Tomoyasu, M. (2004). Weighted principal component analysis and its applications to improve fdc performance. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 4, pages 4262–4267. IEEE.
- Yugma, C., Blue, J., Dauzère-Pérès, S., and Obeid, A. (2015). Integration of scheduling and advanced process control in semiconductor manufacturing: review and outlook. *Journal of Scheduling*, 18(2):195–205.
- Zhou, X., Xi, L., and Lee, J. (2007). Reliability-centered predictive maintenance scheduling for a continuously monitored system subject to degradation. *Reliability Engineering & System Safety*, 92(4):530–534.
- Zhou, Z.-J., Hu, C.-H., Xu, D.-L., Chen, M.-Y., and Zhou, D.-H. (2010). A model for real-time failure prognosis based on hidden markov model and belief rule base. *European Journal of Operational Research*, 207(1):269–283.

Résumé français

Introduction

L'ère de l'après-révolution industrielle a débouché sur l'émergence environnement de production de haute mix et faible volume (High-mix low-volume, HMLV) qui est défié par l'augmentation de la diversité de demande. Par conséquent, les cycles de vie de produit s'est raccourci. Dans ce contexte de production fragile, complexe et extrêmement compétitif, la part de marché peut seulement être augmentée et soutenue en développant de nouvelles technologies et une utilisation améliorée de capacités de production existantes.

Cependant, les changements fréquents dans les recettes (de processus) ont mené les conduites dérivées d'équipement et l'augmentation des pannes imprévus qui perturbent et réduisent les capacités de production. Là exister des solutions multiples de ce problème. Par exemple, nous pouvons investir pour acheter l'équipement de production supplémentaire et neutraliser les défis de HMLV. Ce sera une transformation artificielle aux productions avec bas volume et bas mixtes, mais à forte capacité de capital. Par ailleurs, la part de marché durable exige de nouvelles technologies dans les intervalles courts mais réguliers; donc, nous avons un risque d'incompatibilité technologique avec l'équipement de production, ceux qui sont nouvellement employé. Ces défis peuvent également être résolus en intégrant le système d'automatisation et la gestion informatisée de la maintenance. Cependant, le comportement de l'équipement hostile nécessite des systèmes d'automatisation plus intelligents que les systèmes d'automatisation programmables pour améliorer les capacités de production. La panne d'équipement imprévu augmentant s'ensuit dans l'augmentation de la maintenance corrective et préventive (déclenchement action préventives due aux correctives maintenance). Cela non seulement déstabilise, mais réduit aussi les capacités de production. En fait, les cycles de vie de produit raccourcis s'ensuivent aussi dans le cycle de vie de technologie raccourci; donc, la part de marché augmentée et durable ne peut pas être garantis en incorporant seulement nouvelle technologie. Le problème de base est dans les capacités de production instables et réduites. Dans cette thèse de recherche, nous sommes objectivement concentrés à les stabiliser et les améliorer.

Nous pensons que cela peut être réalisé en améliorant la maintenance corrective (CM) et la maintenance préventive (PM). Cet argument a servi de base à une enquête initiale à l'aide de données recueillies sur la chaîne de production HMLV d'une ligne de production mondialement réputée. L'analyse

(voir le chapitre 1, sous-sections 1.5.1, 1.5.2) l'a accentué la raison d'augmentation de la CM. Il s'agit du misdiagnosis dans l'identification de la source de dérives de qualité de produit et de défaillance(s)/de cause(s). Ce fait est prouvé dans les durées de pannes, les nombres de pannes et les nombres d'actions de réparation augmentés. Cela a un impact direct sur le PM. Il est à cause du fait que pendant l'exécution d'actions de CM, misdiagnosis s'ensuivent souvent dans l'exécution d'actions de PM par les techniciens, dans l'anticipation pour remettre la machine dans la production le plus tôt possible.

Dans les cadres de production HMLV, les pratiques de maintenance généralement employées sont le CM, systématique PM et la maintenance conditionnelle (Condition Based Maintenance, CBM). Le CM intervient une fois que le problème est survenu, alors que systématique PM vise à empêcher tout incident technique au moyen de la prévention, autrement dit, par un entretien régulier du bien concerné fait allusion au temps ou d'autre unité écoulée a basé la maintenance avec un objectif de garder l'équipement dans sa condition fonctionnelle. Ces stratégies de maintenance pourraient facilement s'ensuivre dans sous ou sur l'ingénierie et à la base de l'émergence de CBM où les actions de maintenance sont déclenchées basées sur des informations disponible (du passée au temps de décision) de la condition d'équipement. Dans l'industrie de semiconducteur, le CBM est exécuté avec FDC (Fault Detection and Classification) et est utilisé pour contrôler l'arrêt d'équipement en cas de la conduite anormale découverte par le supervision continu de signaux des capteurs d'équipement. Il est accompli dans deux pas, d'abord la détection de la faute et ensuite par l'identification de causes respectives pour la faute.

De plus, le succès de CBM basé sur FDC compte sur des fautes/défaillances répétables d'équipement. Il peut être conclu que cette stratégie de maintenance n'est pas très utile dans l'environnement HMLV à cause des conduites d'équipement qui se change fréquemment. En comparaison de ces stratégies de maintenance, PdM (la maintenance prédictive) a émergé comme une solution prometteuse où efforts sont faits pour prédire des échecs avant leurs occurrences avec des critères complémentaires à ceux du CBM. Il y a un haut degré d'incertitude associée aux approches de PdM, et dans cette thèse de recherche, nous sommes subjectivement concentrés d'adresser en premier temps les misdiagnosis dans la source de dérive de qualité de produit et la défaillance(s)/la cause(s). Dans la revue de littérature (voir le chapitre 2), nous avons constaté que l'on croit que l'équipement est la seule source de dérives de qualité de produit. Pourtant, dans les cadres de production de HMLV, la source de dérive de qualité de produit peut varier de l'équipement, le produit, le processus (e.g. changement de recette) et la maintenance. Donc, nous soutenons qu'en absence d'un model diagnostic multi-source

pour les dérives de qualité de produit, il n'est pas possible de réduire des pannes d'équipement imprévus augmentantes et donc respectif CM et les actions de PM due à la CM. Nous avons aussi constaté que la plupart des approches de diagnostic de défaillance(s)/de causes(s) proposées est fondé sur les données de capteurs d'équipement. De plus, les approches concernées sont développées et testés au niveau d'équipement. Nous soutenons que les problèmes de fiabilité de capteur pourraient facilement s'additionner à misdiagnosis plutôt qu'aider dans la réduction des durées de pannes et des nombres de réparation exécutées pendant une panne d'équipement imprévue. Nous soutenons aussi que l'équipement est composé des modules (les assemblages qui sont reliés l'un à l'autre dans la relation parent-fils et a l'effet significatif sur le fonctionnement l'un de l'autre).

Au delà de là, les approches émergentes et prometteuse basées sur le PdM s'appuient sur le calcul de la durée de vie résiduelle (Remaining Useful life, RUL) ou l'identification des signatures de pannes. Les approches basées sur RUL utilisent des modèles de dégradation physique pour prédire la défaillance potentielle et sont comparées au niveau des composants alors que les approches basées sur la signature de défaillance sont basées sur les données et s'appuient sur l'apprentissage automatique et les techniques d'intelligence artificielle. Le désavantage associé aux approches basées de la signature consiste en ce que ces signatures de défaillance ne garantissent pas l'alerte bien avant le temps pour que les mesures de proactive/corrective puissent être exécutées/planifiées pour éviter des occurrences de défaillance ou pour améliorer l'utilisation de ressources.

Basé sur ces arguments, nous avons fixé la portée de cette thèse dans la recherche des réponses à (i) comment adresser le misdiagnosis à la dérive de qualité de produit source?, (ii) comment améliorer le diagnostic de défaillance(s) et ou de cause(s) et (iii) comment extraire le temps bornées aux signatures de défaillance pour améliorer non seulement le CM, mais aussi le PM. Ces réponses nous permettront de déplacer vers l'implémentation de maintenance proactive pour améliorer les capacités de production dans l'environnement de production de HMLV. De plus pour éviter le problème de fiabilité de capteurs d'équipement pour la validation de nos approches, on se restreint aux données de type informations contextuelles d'un cas d'étude d'une ligne de production mondiale.

Choix de modèles et les approches proposées

Les approches utilisées pour le diagnostic et les prédictions de défaillance(s) sont basées sur les données (voir le chapitre 4). Ceux-ci peuvent être da-

vantage généralement classifiés comme statistiques (ex: l'analyse en composantes principale, l'analyse discriminante, régression des moindres carrés partiels etc.) et les techniques de machine learning (ex: machine à vecteurs de support, arbres de défaillances, modèle de Markov, réseau de neurone, le réseau de Bayesian etc.). Le choix de BN comme modèle dans cette thèse de recherche est fondé sur le fait que nous utilisons des informations contextuelles basées d'événement au lieu des données temporelles des capteurs. Cela limite l'usage direct d'approches qui sont fondées sur les données purement temporelles. Cependant, puisque les données temporelles basées sur les capteurs peuvent être discrétisées, nous pensons que cela se traduira par l'agrégation des biais du capteur conduisant à de mauvaises décisions

De plus, notre objectif n'est pas d'automatiser la détection de source de dérive de qualité de produit, le diagnostic automatique de défaillance/cause et la prédiction de défaillance mais d'aider les équipes de maintenance avec des mécanismes de prise de décision. Naturellement, l'incertitude est aussi associée à l'occurrence de défaillance aussi bien que diagnostic de défaillance/de cause en raison de la complexité d'opérations de chaîne de fabrication (contraint HMLV) et l'équipement de production à usages multiples complètement automatisé. Le BN est un graphe acyclique dirigé composé de nœuds et d'arcs. Les nœuds représentent des variables tandis que les arcs présentent l'influence d'une variable sur l'autre en fonction de leur direction de cause et effet. Par conséquent, nous avons choisi BN comme principale approche de modélisation puisqu'elle donne des résultats plus précis avec des informations contextuelles basées sur des événements et une inférence probabiliste. De plus, il fournit une visualisation graphique des variables d'intérêt avec leur structure d'influence causale.

L'inconvénient de BN est sa limitation dans le traitement d'un grand nombre de variables car il résulte en problème NP-dur lorsque des probabilités conditionnelles sont calculées à chaque nœud. Cela peut être contrôlé en sélectionnant les variables les plus importantes et les plus critiques. Il existe des approches statistiques par exemple l'information mutuelle et l'entropie, pour classer les variables les plus pertinentes dans un ensemble de données d'étude. Dans cette thèse, nous avons basé la sélection de variables en premier temps sur l'opinion des experts. De plus, le temps de calcul pour la construction du BN et le calcul des probabilités conditionnelles a été encore réduit par l'application des règles d'apprentissage de la structure.

Au meilleur de nos connaissances, il n'existe pas de littérature qui se concentre sur l'identification de la source de la dérive de la qualité du produit en prenant l'hypothèse que l'équipement n'est pas la seule source de la qualité des produits dérive. Par conséquent, nous avons proposé, développé et validé un Réseau Bayésien (BN) pour réduire les erreurs de diagnostic dans la source

de la dérive de la qualité du produit comme première ligne de défense contre l'augmentation des pannes imprévues de l'équipement. Le diagnostic de défaillance/cause(s) avec BN est développé en utilisant des informations contextuelles basées sur des événements collectées à partir de sources de données de produit, de processus, d'équipement et de maintenance, à la fois au niveau de l'équipement et du module (assemblage). Il s'agit de la deuxième ligne de défense contre l'augmentation des pannes. Les résultats du cas d'étude démontrent que BN identifie de manière significative la source de dérive de la qualité du produit autre que l'équipement et que le modèle BN au niveau de module (assemblage) donne des prédictions plus précises que le niveau de l'équipement.

De plus, nous avons proposé, développé et validé une méthodologie d'extraction de signature de défaillance au temps bornés basée sur BN et des informations contextuelles. Le rôle de BN dans cette méthodologie est de tracer les probabilités de défaillance par rapport à la ligne de temps discrétisée et agrégée d'information contextuelle. Cette méthodologie offre la souplesse nécessaire pour extraire plusieurs règles avec des délais de réalisation associés variables afin que des mesures proactives ou correctives puissent être exécutées pour éviter les incidents de défaillance ou pour planifier les ressources de maintenance. Le calcul du délai d'exécution associé aux actions de maintenance requises est subjectif et dépend de la compétence de l'équipe de maintenance. Par conséquent, cela est laissé à la discrétion des experts en équipement. Cette méthodologie offre la flexibilité aux équipes de maintenance pour détendre ou resserrer les alertes de défaillance correspondantes en fonction des indicateurs de dérive de ligne de production. Cela est important parce qu'un choix inapproprié pourrait entraîner un sur ou sous l'ingénierie.

Organisation de la thèse

Cette thèse est divisée en trois parties. La partie I présente l'introduction au contexte de problème, le but de recherche et ses objectifs, la revue de la littérature et positionnement scientifiques. La partie II est concentrée sur le choix du BN pour modeler nos approches. Dans la partie III, nous présentons le cas d'étude et les résultats de validation des méthodologies proposées comme présentées dans la partie II. L'organisation de thèse schématisés et détaillée est présentée dans la partie I, le chapitre 3.

