SENSOR BASED REAL-TIME PROCESS MONITORING FOR ULTRA-PRECISION MANUFACTURING PROCESSES WITH NON-LINEARITY AND NON-STATIONARITY

By

OMER FARUK BEYCA

Bachelor of Science in Industrial Engineering
Fatih University
Istanbul, Turkey
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July 2013

SENSOR BASED REAL-TIME PROCESS

MONITORING FOR ULTRA-PRECISION

MANUFACTURING PROCESSES WITH NON-

LINEARITY AND NON-STATIONARITY

Dissertation Approved:

Dr. Zhenyu (James) Kong

Dissertation Adviser

Dr. Satish T.S. Bukkapatnam

Dissertation Co-Adviser

Dr. Joe Cecil

Dr. Qi Cheng

External Member

ACKNOWLEDGEMENTS

I would like to thank my adviser, Dr. Zhenyu Kong, for his mentorship and guidance during my PhD study. The knowledge and skills I learned from him were key to the completion of this work. I would also like to thank my co-adviser, Dr. Satish Bukkapatnam, and committee members, Dr. Joe Cecil and Dr. Qi Cheng for their invaluable input.

I would also like to express my gratitude for Dr. Ranga Komanduri (1942-2011), with whom this project started. Finally, thanks are due to my colleagues, Dr. A. Oztekin, Mr. P. Rao, Mr. C. Cheng, Mr. M. Mistarihi, Mr. K. Bastani, Mr. Z. Wang, Mr. A. Arulmozhi, and Mr. R. Ravichandran for their collaboration and friendship.

I dedicate this work to my parents, Hilmi and Ayten Beyca, and my sisters Hicran, Betul, Tuba and Cemre, for supporting me throughout my studies. I am eternally indebted to them.

I thank Allah for helping me complete my studies successfully and for the patience He bestowed on me while working on this project.

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: OMER FARUK BEYCA

Date of Degree: JULY 2013

Title of Study: SENSOR BASED REAL-TIME PROCESS MONITORING FOR ULTRA-PRECISION MANUFACTURING PROCESSES WITH NON-LINEARITY AND NON-STATIONARITY

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: This research investigates methodologies for real-time process monitoring in ultra-precision manufacturing processes, specifically, chemical mechanical planarization (CMP) and ultra-precision machining (UPM), are investigated in this dissertation.

The three main components of this research are as follows: (1) developing a predictive modeling approaches for early detection of process anomalies/change points, (2) devising approaches that can capture the non-Gaussian and non-stationary characteristics of CMP and UPM processes, and (3) integrating multiple sensor data to make more reliable process related decisions in real-time.

In the first part, we establish a quantitative relationship between CMP process performance, such as material removal rate (MRR) and data acquired from wireless vibration sensors. Subsequently, a non-linear sequential Bayesian analysis is integrated with decision theoretic concepts for detection of CMP process end-point for blanket copper wafers. Using this approach, CMP polishing end-point was detected within a 5% error rate.

Next, a non-parametric Bayesian analytical approach is utilized to capture the inherently complex, non-Gaussian, and non-stationary sensor signal patterns observed in CMP process. An evolutionary clustering analysis, called Recurrent Nested Dirichlet Process (RNDP) approach is developed for monitoring CMP process changes using MEMS vibration signals. Using this novel signal analysis approach, process drifts are detected within 20 milliseconds and is assessed to be 3-7 times faster than traditional SPC charts. This is very beneficial to the industry from an application standpoint, because, wafer yield losses will be mitigated to a great extent, if the onset of CMP process drifts can be detected timely and accurately.

Lastly, a non-parametric Bayesian modeling approach, termed Dirichlet Process (DP) is combined with a multi-level hierarchical information fusion technique for monitoring of surface finish in UPM process. Using this approach, signal patterns from six different sensors (three axis vibration and force) are integrated based on information fusion theory. It was observed that using experimental UPM sensor data that process decisions based on the multiple sensor information fusion approach were 15%-30% more accurate than the decisions from individual sensors. This will enable more accurate and reliable estimation of process conditions in ultra-precision manufacturing applications.

Table of Contents

Chapter I Introduction 1

1.1 Research Motivation 1

1.2 Research Challenges and Objectives 4

1.3 Major Contributions of the Dissertation 5

1.4 Organization of the Dissertation 7

Chapter II Background and Literature Review 9

2.1 Sensor Based Process Monitoring of CMP 9

2.1.1 CMP end point detection 10

2.1.2 CMP process monitoring 10

2.1.3 Defect detection in CMP process 12

2.2 Sensor-based UPM Process Modeling 13

2.3 Online State Estimation of Dynamic Processes 13

2.4 Decision Making in Design and Manufacturing 15

2.4.1 Application of utility theory in decision making 15

2.4.2 Clustering methods for change point detection 16

2.5 Sensor Fusion in Manufacturing Process 17

Chapter III Overall Research Methodology 20

3.1 Nonlinear Sequential Bayesian Analysis-Based Decision Making for Chemical Mechanical Planarization Process (Chapter 4) 20

3.2 Process Monitoring by Using Evolutionary Clustering Analysis for Chemical Mechanical Planarization (CMP) (Chapter 5) 22

3.3 Multi-sensor Fusion Based Process Monitoring of Ultra Precision Machining (Chapter 6) 23

Chapter IV Nonlinear Sequential Bayesian Analysis-Based Decision Making for High Precision Machining 25

4.1 Introduction 25

4.2 Research Methodology for End-point Detection 27

4.2.1 Online predictive modeling for process state 29

4.2.2 Performance measure (MRR) prediction based on predicted process state 35

4.2.3 Decision making approach for CMP endpoint detection 36

4.3 Case Studies 39

4.3.1 Offline regression analysis connecting process state with process performance 40

4.3.2 Predictive model for process state variables 43

4.3.3 Predictive model for MRR 46

4.3.4 Decision making for CMP endpoint detection 48

4.4 Summary 50

Chapter V Process Monitoring by Using Evolutionary Clustering Analysis for Chemical Mechanical Planarization (CMP) 52

5.1 Introduction 53

5.2 Proposed Research Methodology for Change Point Detection 56

5.2.1 Experimental setup 57

5.2.2 Dirichlet process (DP) based modeling of CMP vibration sensor signals 58

5.2.3 Recurrent Dirichlet process (RDP) model 65

5.2.4 Recurrent nested Dirichlet process (RNDP) model 68

5.2.5 Formulation of RNDP-based control charts for process monitoring applications 69

5.3 Validation of the Proposed Method 74

5.3.1 Change point detection with simulated time series data 75

5.3.2 Application of RDNP method for detection of CMP process anomalies 91
5.4 Summary 100

Chapter VI Multi-sensor fusion based process monitoring of Ultra Precision MachinINg 101

6.1 Introduction 101
6.2 Research Approach for Multi-sensor Based Process Monitoring 104
6.2.1 Feature level fusion by using Dirichlet process 108
6.2.2 Decision level fusion with evidence theory 110
6.3 Validation of proposed method 111
6.3.1 Numerical case studies for feature-level fusion 112
6.3.2 UPM Case Studies 118
6.4 Summary 121

Chapter VII Conclusions and Future Work 123

References 127

APPENDIX 148

Matlab Codes 148
A.1 Nonlinear Sequential Bayesian Analysis-Based Decision Making for Chemical Mechanical Planarization Process 148
A.2 Chemical Mechanical Planarization (CMP) process monitoring by using evolutionary clustering analysis 151
A.3 Process Monitoring of Ultra Precision Machining (UPM) based on Sensor Fusion 159

List of Figures

Figure Page

Figure 1.1 Taniguchi curve prediction of machining accuracy by year [6]............ 2

Figure 1.2 Illustration of Moore's law [7] 3

Figure 3.1 Modules and techniques used in this dissertation............................... 21

Figure 3.2 Overall research methodology for CMP end-point detection.............. 22

Figure 3.3 Overall methodology for CMP monitoring by RNDP 23

Figure 3.4 Proposed overall methodology for sensor fusion for UPM process monitoring... 24

Figure 4.1 Proposed overall research methodology.. 28

Figure 4.2 Representation of multimodal distribution using samples 30

Figure 4.3 Utility function for EPD of CMP 38

Figure 4.4 Empirical distribution of material removed and estimation for probability of events ϕ_1 (under-polish) and ϕ_2 (over-polish) 38

Figure 4.5 Experimental setup using a LapMaster L 12 Bench-top lapping machine. 40

Figure 4.6 The multilayer perceptron neural network for regression analysis between sensor signals and MRR 42

Figure 4.7 Mutual information analysis for determining time lag........................ 44

Figure 4.8 Predicted result for feature 1 using particle filter with logistic function 44

Figure 4.9 Comparison MSE of predicted feature using various methods, i.e., PF(logistic), PF(polynomial), ARMA(3,3), EKF, and KF 45

Figure 4.10 Predicted MRR vs. measured MRR 47

Figure 4.11 Predicted MRR using various methods 47

Figure 4.12 Empirical distribution of material removed at the 275th second 48

Figure 4.13 Expected utility for "continue polishing" and "stop polishing" with endpoint threshold of 450 mg material removal (MR) (On the x-axis the numbers 1-11 corresponds to 254.5 to 255.5 seconds) 49

Figure 4.14 Relationship between the time of endpoint and material to be removed at endpoint 50
Figure 5.1 Overall Methodology for CMP monitoring by using RNDP 56
Figure 5.2 Buehler Automet® 250 experimental CMP polishing setup and wireless Xbee Vibration Sensor 57
Figure 5.3 Polished blanket copper wafers after 12 min of CMP 58
Figure 5.4 Time series and histogram of vibration sensor data 59
Figure 5.5 Possible cluster evolutions 65
Figure 5.6 Phase space plot (b) constructed from time series (a) by using mutual Information (c) and false nearest neighbor (d) 72
Figure 5.7 Evaluation of estimated distribution vs. true distribution at different number of iterations 73
Figure 5.8 Chi square statistics with 5% p-value threshold 73
Figure 5.9 Illustration of changes in the process 78
Figure 5.14 Non-Gaussian process noise added for fault 5 81
Figure 5.10 ARL_1 results for fault type 1 83
Figure 5.11 ARL_1 results for fault type 2 84
Figure 5.12 ARL_1 results for fault type 3 85
Figure 5.13 ARL_1 results for fault type 4 86
Figure 5.15 ARL_1 results for fault type 5 87
Figure 5.16 Generated sinusoidal signal (first 500 is normal case with 4 frequencies second 500 is change case with 5 frequencies) 90
Figure 5.17 Frequency comparison of normal and fault case for sinusoidal signal 90
Figure 5.18 Detection delay comparison for sinusoidal signal 90
Figure 5.19 Typical vibration sensor data from CMP process 92
Figure 5.20 Frequency spectrum of vibration signal data 92
Figure 5.21 Representative vibration signal patterns obtained under changing load conditions. 93
Figure 5.22 Detection delay results for changing load conditions 94
Figure 5.23 Representative vibration signal patterns obtained for slurry depletion experiments 95
Figure 5.24 Detection delay results for slurry depletion experiments. 96

Figure 5.25 Representative vibration signal patterns obtained for pad wear experiments 97

Figure 5.26 Detection delay results for pad wear 98

Figure 5.27 Vibration data time series for multiple fault experiment........ 99

Figure 6.1 UPM experimental apparatus 105

Figure 6.2 Proposed Information Fusion Mechanism........ 107

Figure 6.3 Proposed overall methodology for sensor fusion for UPM process monitoring........ 108

Figure 6.4 Simulated data generated from non-Gausssian distribution for three condition 112

Figure 6.5 Comparison of feature-level fusion algorithms along with the training data on top left. 114

Figure 6.6 Spindle speed experiments 115

Figure 6.7 Feed rate experiments........ 116

Figure 6.8 Depth of cut experiments........ 117

Figure 6.9 Correlation analysis 121

List of Tables

Table Page

Table 4.1 Pseudo code of the particle filter method with multiple time lags (N is number of particles, k is the time index, p is the time lag)................... 33

Table 4.2 Decision Table for preferences.. 37

Table 4.3 Comparison of MSE of predicted feature using different methods...... 45

Table 4.4 Decision table for CMP endpoint detection... 48

Table 5.1 Pseudo code for DP clustering... 64

Table 5.2 Summary of fault types.. 77

Table 5.3 Classification accuracy matrix for fault type 6 88

Table 5.4 Classification accuracy for multiple faults experiment 100

Table 6.1 Design of experiment parameters 106

Table 6.2 Details of the sensing system mounted on the UPM setup................. 106

Table 6.3 Summary of extracted features 109

Table 6.4 Distribution parameters used for generated data 112

Table 6.5 Classification results for different classification methods.................. 113

Table 6.6 Accuracy of individual sensors for spindle speed experiments.......... 119

Table 6.7 Accuracy of fused sensors for spindle speed experiments 119

Table 6.8 Accuracy of individual sensors for feed rate experiments.................. 119

Table 6.9 Accuracy of fused sensors for feed rate experiments 119

Table 6.10 Accuracy of individual sensors for depth of cut experiments 120

Table 6.11 Accuracy of fused sensors for depth of cut experiments.................. 120

CHAPTER I

INTRODUCTION

This dissertation investigates methodologies for real-time process monitoring in ultra-precision manufacturing processes. Specifically, this study focuses on two nano-level precision manufacturing processes, namely, chemical mechanical planarization (CMP) and ultra precision machining (UPM) [1-2]. Nonetheless, the approaches developed in this dissertation are applicable to a broad spectrum of manufacturing processes.

1.1 Research Motivation

UPM products play an important part in both our daily life and in areas of national interest, such as cosmetics, textiles, biomedical, healthcare, electronics and computer, and space and defense industry, etc. [3-4]. In the near future National Science Foundation (NSF) foresees a $1 trillion dollar market for nano-manufactured products [5].

In the early eighties, Taniguchi introduced a roadmap for manufacturing accuracy where he suggested that nano-level precision in manufacturing would be possible before the new millennium in [6]. This is illustrated with the so-called Taniguchi curve (Figure 1.1).

Another well known illustration of the exponential tightening of manufacturing is the trend in the density of transistors on integrated circuits which doubles every two years as predicted by Moore [7] (Figure 1.2) .

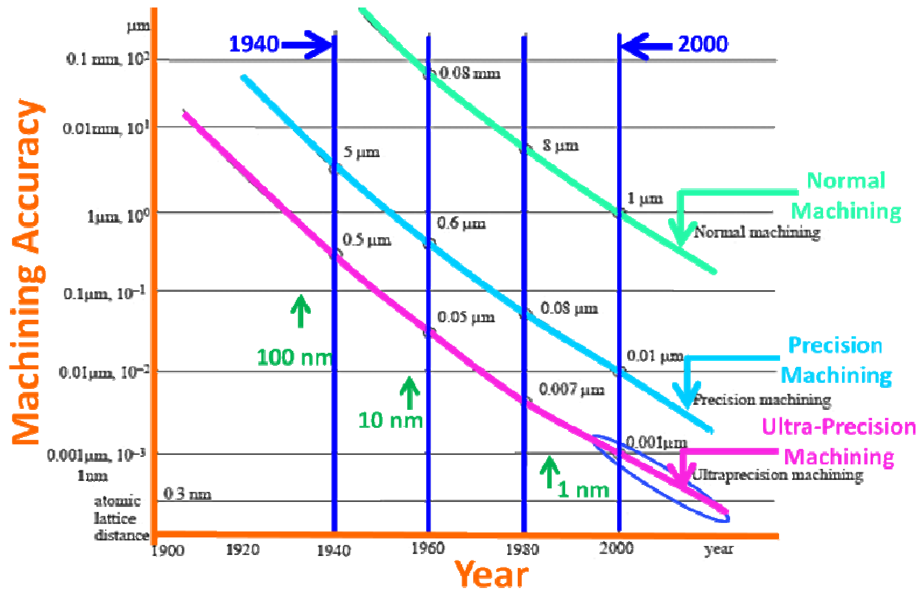


Figure 1.1 Taniguchi curve prediction of machining accuracy by year [6]

Predictions by Taniguchi and Moore suggest that due to ever-tightening dimensions, monitoring of nano-manufacturing process is vital for quality assurance. Traditional statistical methods are not sufficient for monitoring and prediction applications in this area due to the complex nature of the nano-manufacturing process which exhibits non-linear and non-stationary signal characteristics. In this research, novel monitoring and prediction techniques are developed for ultra-precision manufacturing processes based on sensor data.

Figure 1.2 Illustration of Moore's law [7]

As mentioned earlier, ultra-precision machining (UPM) and chemical mechanical planarization (CMP) are the main focuses of this research. Both UPM and CMP are processes used for crafting variety of materials such as, metals, semiconductors, and composites with surface finishing in the range of 1-50 nm. Material removed in UPM is achieved by using a single crystal diamond cutting tool on a very precise lathe with 2-50 micrometer depth of cut range [1, 8-9]. For a successful UPM processing, precise control over process conditions is needed. This includes measures to decrease the effect of external vibrations, such as granite bases with pneumatic vibrations isolators, frictionless slide, air bearing spindles, and stiff work holding arrangements.

Until recently, UPM process was primarily used in research oriented applications, such as X-ray telescope mirrors [8, 10-11]. However, newer consumer applications for UPM are beginning to emerge. While still a niche market, UPM hardware sales in 2003 were recorded ~$100 million and are still growing [12].

CMP process uses chemical and mechanical effects to planarize surfaces with a polishing action. Nano-particulate abrasives, such as alumina (Al_2O_3) and silica (SiO_2), are used to soften the material by chemical reactions, and, subsequently, material is removed by mechanical polishing [13-14].

Semiconductor industry has recorded significant growth in recent years. In 2010, market revenues in semiconductor industry were estimated to be approximately $300 billion, which is 30% higher than the revenues of previous three years [15]. Advances in CMP yield rates are a key to this growth.

1.2 Research Challenges and Objectives

Bukkapatnam, *et al.* [16] summarized recommendations and challenges in nano-manufacturing process which were outcomes from a workshop on nano-technology organized by the National Science Foundation (NSF) in 2009. These recommendations suggest means to overcome vexing challenges in nano-manufacturing research.

Some of the challenges summarized in Ref. [16] are as follows:

- accesibilty to signal source is challenging;

- in-situ sensing is almost impossible;
- signals are short, evanescent and weak;
- quantization of signals makes transduction difficult; and
- signal-to-noise ratio is low

To surmount the above challenges, this research develops innovative solutions for monitoring of nano-level manufacturing processes. The research objectives of this dissertation are as follows:

(1) To develop a predictive modeling of dynamic process in high precision machining.

(2) To estimate probability density information of sensor signal in high-precision machining which is non-linear and non-stationary.

(3) To integrate multi-sensor signals for coherent and reliable decisions.

1.3 Major Contributions of the Dissertation

This research develops methods for real-time monitoring in nano-manufacturing process by integrating multiple heterogeneous sensors such as, force, vibration and acoustic emission. The major contributions of this research are as follows:

(1) **Feature extraction for representation of sensor signals:** In nano-level manufacturing processes, signal-to-noise ratio is low, which creates high uncertainty in data acquisition and modeling. In this research, various feature extraction methods are used for better representation of sensor

data. As an example, time and frequency features are extracted, and, subsequently, principal component analysis (PCA) is used for dimension reduction. Process performance can be estimated successfully by using these features as input (Chapter 4). However, the inherently non-linear dynamics of the process causes the statistical distribution of sensor signal to be non-Gaussian. This non-Gaussian distribution is estimated using mixture of Gaussian distributions. Parameters of such mixtures of Gaussian distributions can be used as test statistics for monitoring purposes. (Chapter 5).

(2) **Process state prediction of dynamic processes**: In nano-level manufacturing applications, accurate and timely decisions for process monitoring and control are important from a quality assurance perspective. Due to its complex, non-linear and non-stationary nature, process state prediction is challenging. Linear and stationary models are not able to capture the non-linear and non-stationary nature of the process. In order to tackle this challenge we use sequential Monte Carlo method which is used for prediction of process state for early detection method (Chapter4).

(3) **Decision making for monitoring of nano-manufacturing process**: Decision making techniques are devised for monitoring manufacturing process applications, such as detecting a change in surface variation in UPM process and end-point detection in CMP. We integrated utility

theory with the predictive modeling for early detection of end-point in CMP (Chapter 4). We used evolutionary clustering method to monitor process change in CMP process effectively by monitoring the evolution of the cluster parameters in real time (Chapter 5).

(4) **Sensor Fusion of heterogonous signals:** With the development of the sensor technology, more and more data can be acquired from complex processes. In the context of process monitoring, it is important to determine how to integrate multiple sensor data to achieve more accurate results. Evidence theory is combined with non-parametric Bayesian modeling for fault detection and fault classification in UPM process (Chapter 6).

1.4 Organization of the Dissertation

In this chapter, the research objectives, challenges, and the major contributions of this research are presented. The rest of the dissertation is organized as follows:

Chapter 2 Literature Review and Background: This chapter presents a review of pertinent literature in sensor based process modeling of CMP and UPM processes in addition to techniques used in process monitoring methodologies.

Chapter 3 Overall Research Methodology: This chapter outlines the overall methodology and research framework used in this dissertation.

Chapter 4 Non-Linear Sequential Bayesian Analysis-Based Decision Making for CMP process: In this chapter, we integrated the nonlinear sequential Bayesian analysis with decision theory. We established a quantitative relationship connecting the statistical features (inputs) extracted from real-time sensor signals with the process performance. An application of end-point detection in CMP process is illustrated by using the proposed method.

Chapter 5 Process Monitoring by Using Evolutionary Clustering Analysis for Chemical Mechanical Planarization (CMP): In this chapter a non-parametric Bayesian modeling is developed to model non-Gaussian and non-stationary sensor signal captured by various sensors. Evolutionary clustering analysis is used to monitor process changes in the process.

Chapter 6 Multi-sensor Fusion Based Process Monitoring of Ultra Precision Machining (UPM): Sensor fusion techniques are investigated in order to monitor UPM process using signals collected from heterogonous sources. We integrated non-parametric Bayesian analysis with sensor fusion techniques to monitor surface variations in UPM process.

Chapter 7 Conclusions and Future Work: This chapter presents the research contributions, conclusions and future directions.

CHAPTER II

BACKGROUND AND LITERATURE REVIEW

The methodologies adopted in this dissertation are closely related to the following research areas:

- Sensor based process monitoring of Chemical Mechanical Planarization (CMP) and Ultra Precision Machining (UPM) processes [17-19]
- Online state estimation of dynamic processes [20-24]
- Decision making and sensor fusion in manufacturing processes [25]

The related research in these areas is presented in the following sections.

2.1 Sensor Based Process Monitoring of CMP

Current research efforts in CMP process monitoring mainly focus on process endpoint detection (EPD) that enables us to make decision on when to stop process monitoring and defect detection. Some of the recent studies are presented in the following subsections. The use of optical, thermal, tribological, acoustic, capacitive, and inductive sensing principals has been reported in Refs. [17-19]

2.1.1 CMP end point detection

An electrochemical approach for EPD in Cu-CMP is reported in Ref. [26] by monitoring the concentration of Cu-ion. The decrease in the Cu-ion, which is measured by using a capillary and ion-selective electrode, signals an End-Point. An AE-based *in situ* EPD technique using highly selective slurries for metal CMP is presented in Ref. [27]. The End-Point is detected by observing the change in the acoustic emission and coefficient of friction signals. The EPD approach reported in Ref. [28] leverages a highly selective slurry (HSS). The investigation shows that using HSS increases material removal rate of silicon-oxide and increases the oxide-to-nitride selectivity, which improves the detection of EPD. A vision-based method is developed for EPD [29].

Wavelet decomposition of acoustic emission and coefficient of friction signals is analyzed by using sequential probability ratio test to detect End Point online [30]. In Ref. [31], particle filtering technique is integrated with neural network and decision making theory to predict EPD accurately. An electrochemical approach for EPD in Cu-CMP is reported in Ref. [32] by monitoring the concentration of Cu-ion.

2.1.2 CMP process monitoring

CMP process was optimized by analyzing the effect of the slurry feeding position and sample holder movement on the polishing performance, such as wafer uniformity and material removal rate in Ref. [33]. Both acoustic emission (AE)

and coefficient of friction (COF) signals were also recorded during the experiments. The authors reported that AE signal magnitude drops and COF increases while the wafer surfaces become more uniform. Acoustic emission signal was used for monitoring oxidation level during the CMP process [34]. This study showed that the magnitude of AE signal increases during the oxidization process. A feed forward neural network controller was developed to monitor CMP performance such as material removal rate and wafer uniformity. The input variables included polishing time, down force, and relative speed between the polishing pad and wafer [35]. Multiple sensors including piezoelectric force sensor, hall effect sensor and acoustic emission sensor were used to monitor the CMP process [36]. These experiments showed that force sensor and hall-effect sensor can clearly detect end-point if the friction characteristics are distinct between materials.

The relation between vibration sensor data and material removal rate is investigated in Ref. [37]. In Ref. [38], non-linear dynamics of the vibration sensor signal is correlated with the material removal rate and the authors showed that non-linear dynamic features improve the predicting material removal rate in CMP by 20% comparing to conventional statistical features. Particle Filtering technique is integrated with statistical regression to predict CMP process performance such as material removal rate in Ref. [39]. Non-linear dynamics of functional process variables (FPVs), such as pad temperature, and COF between

wafer and pad, as well as their interactions are used for detecting process changes in CMP [40].

2.1.3 Defect detection in CMP process

Acoustic emission sensor signal and micro scratches correlation is investigated in Ref. [41]. Sharp peak value in acoustic emission signal was noted when scratches occur during CMP process. In Ref. [42] both COF and AE sensor data were collected to detect delamination defects in low-k Cu-CMP. Experiments show that COF signals are not sensitive to delamination effect while wavelet analysis of the AE sensor data can detect delamination defects in low-k Cu-CMP effectively. A physical scratch model is developed using pad response and slurry behavior [43] . This study showed that the scratch depth is increasing as the scratch frequency decreases. Optical methods are used for the detection of defects on a wafer surface [44].

Most of the previous efforts are mostly based on EPD detection and process monitoring. There are very few studies that have examined sensor based detection of evanescent changes in CMP process. Parametric methods which have been used in the literature are not effective for capturing these fast-changing mechanics by using sensor data. Therefore, a non-parametric data-driven model is needed to model non-stationarity caused by defects.

2.2 Sensor-based UPM Process Modeling

Tool wear mechanism is investigated by using Coefficient of Friction (COF) sensor in Ref. [45]. The authors show that coating the diamond tool with Perfluoropolyether (PFPE) polymer causes reduction of COF signal. In Ref. [46] acoustic emission (AE) signal acquired from 1045 steel finishing experiments are related with tool wear and surface roughness. The authors monitored to increase of surface roughness caused by tool wear. They correlated the extracted features from AE signal such as, zero crossing rate, mean and standard deviation with the surface roughness hence the tool wear.

In Ref. [47], AE signal is used to investigate material anisotropy ahead of the tool in UPM process for both single and polycrystalline copper. AE and micro thermo sensors are used to monitor the machine status in UPM process [48]. Artificial neural network (ANN) is used to relate AE and force signal with surface roughness in surface finishing of Stavax (S136) die steel work pieces [49].

2.3 Online State Estimation of Dynamic Processes

State-space model is widely used in data driven modeling area where the states and model are determined by using process operational data [50]. Accurate estimation of the state variables and model parameters are essential for capturing the characteristics of a dynamic process.

The Kalman filter (KF) approach was proposed as an optimal solution to the state estimation problem when the state models are linear and the posterior density is

Gaussian [50]. However, if these assumptions are not satisfied, the KF method will be ineffective. As an alternative, the extended Kalman filter (EKF) has been used to estimate the state of non-linear dynamic processes [51-52]. A Gaussian posterior density is assumed to implement EKF and a first-order Taylor series expansion is applied to provide a local approximation of state estimation. Nevertheless, the EKF may cause a large estimation error if dynamic process is highly non-linear. Under this circumstance, the Gaussian assumption may not hold. One countermeasure, named "point-mass filters" or "probability-grid filters," is utilized to approximate the posterior density by discretizing the continuous state variables into samples [53-54]. However, the computation of point-mass filters increases significantly with the increase in the state dimensions. So, they are too costly to apply in practice. This greatly limits their applications in process monitoring and control.

Particle filters can be described as an extension of the point-mass filters [51]. The fundamental principal of particle filters is that a large number of samples (particles) are generated by using sequential Monte Carlo sampling methods to approximate the posterior probability of the states. Thus, the particles will have a tendency to be dense in the regions of high probability. Particle filter method has become the most effective way for state estimation of a nonlinear dynamic process with satisfactory computational efficiency and has recently been applied to many fields, such as robotics, multimedia, and surveillance [55].

Gustafson *et al*. [56] applied particle filters approach to broad application areas, such as positioning, navigation, and target tracking. A significant number of studies have been done in robot localization area [20-24]. Particle filters have also been applied to other areas, such as acoustic source localization [57], variable resolution [58], and Eigen tracking in noisy targets [59]. From these broad application areas, it can be seen that particle filtering is a notably effective tool.

2.4 Decision Making in Design and Manufacturing

In this study in order to make decisions about the process two kind of techniques are preferred: (1) Utility theory and (2) clustering methods. Brief literature review in about these techniques are given in the following sections.

2.4.1 Application of utility theory in decision making

Decision and utility theory has been extensively used in engineering design [60-61] for the determination of optimal design alternative. There are also some applications of decision making in the manufacturing field. An interactive decision support system is implemented for on-line process control in refinery production [25]. The structure and components of a distributed decision-making system are described in Ref. [62] for complex discrete systems and processes control. The application of a process decision program chart is discussed in Ref. [63] for total quality control in the area of process planning. A methodology in handling uncertainties in multi-period energy models is developed by modeling multistage decision processes [64]. Nevertheless, very little work has been

reported in applying decision/utility theory for online monitoring and control of manufacturing processes.

2.4.2 Clustering methods for change point detection

Change point detection can be considered as unsupervised clustering problem. Accurate clustering between normal state and defected state is essential to monitor the process quality. Distance based clustering methods such as k-means [65] and k-nearest neighbor [66] has been widely used to cluster data. Distance based methods are highly dependent on pre defined parameters such as number of clusters and neighbor size. Self organizing map is a distance based clustering technique [67] in which the number of cluster does not need to be know a priori. However the clusters are purely based on the distances, and no statistical distribution is formulated. This is not preferable for decision making applications.

Dirichlet Process (DP) Mixture models are applied to broad application areas, such as bioinformatics, healthcare, document clustering and image processing [68-71]. Its self-growing clustering ability enables to model data without prior knowledge of number of clusters. However, DP assumes the data is fully exchangeable and thus cannot capture non-stationarity, this is not suitable for CMP monitoring applications since data comes in sequential way. Recurrent Dirichlet Process (RDP) is developed to overcome this challenge in Ref. [72]. RDP technique can utilize the previous time epochs information to formulate CMP data in current time epoch effectively. Higher level clustering technique is

needed to detect the change point between two epochs, in other words a detection method is need to capture the change between Gaussian mixture distributions. Nested Dirichlet Process (NDP) is developed for clustering for Gaussian mixture distributions [73].

DP method is used to model non-Gaussian distribution but cannot capture non-stationarity. RDP is an effective technique to capture non-stationarity and uses previous time epochs information. However RDP is unable to detect process changes between two consecutive time epochs. NDP can be utilized detect changes between Gaussian mixture distributions but it yet assumes the data is fully exchangeable. In this research we combine RDP and NDP to utilize a novel non parametric Recurrent Nested Dirichlet Process to monitor CMP process.

2.5 Sensor Fusion in Manufacturing Process

There are 3 levels of fusion techniques which are summarized below [74]

- Signal level fusion: Same type of signals from different sensors are collected and fused to create a new signal with improved signal to noise ratio [75-77].

- Feature level fusion: In feature based fusion features in time domain and frequency domain are extracted from various sensor signals. Similar features extracted from sensors are fused to make decision [78-82].

- Decision level fusion: Sensor signals are processed and analyzed individually. The information obtained from individual analysis are combined to apply decision rule for final decision [83-84].

Principal component analysis (PCA) transformation has been widely used to convert correlated sensor signals into uncorrelated components. Subsequently, the first principal component which explains the largest variability in sensor data is used as fused sensor to make decision [75-76]. Sparce PCA method has been used to process decomposition and decorrelation for sparse data then maximum entropy technique is used for decision fusion for fault detection [80].

Estimation techniques such as Kalman filtering has been widely used for fusion mechanism. Dynamic systems are monitored by using multiple sensors. Kalman/particle filtering is used to estimate pdf of process state for each sensor measurement. Correlated decisions fused to make more accurate decisions [83]. Kalman Filtering is used for fusion mechanism. Spectral norm of the normalized innovation matrix is used as test statistics to monitor the process [78]. Extended kalman filtering is used for both centralized and decentralized sensor fusion to detect faults in simulation continuous stirred tank reactor benchmark problem [85].

Machine learning tools are used to for signal and feature level fusion. Fuzzy clustering method is used as fusion mechanism in feature level which are monitored using artificial immune system for change detection [81]. Two staged

fuzzy-logic model is used to as a sensor fusion mechanism for monitoring tool wear condition. In the first stage feature level fusion mechanism is used where, features are extracted from sensor signals acquired from force, sound and vibrations signals. The outputs of the first stage are used for the second stage for decision level-fusion [79]. Support vector machines are used to fused sensor signals to monitor motor faults [77]. Bayesian machine learning technique is used to select minimum number of sensors and the best features of sensor data to predict process condition [82]. Artificial neural network is used as fusion mechanism for features extracted from heterogeneous sensor signals, such as force, vibration and sound for tool wear monitoring on CMC turning process [86].

Most of the decision fusion techniques are developed based on Bayesian theory. The most widely used technique is to calculate posterior probability with Bayes formula in which prior and conditional probabilities determined in advance [87-90]. As an extension to Bayesian theory evidence theory is used to fuse decisions which is based on quantifying and evaluating evidence [84]. Neural network is used to model sensor data from drilling experiments. The outputs of neural network are used for construction of mass function. Dampster-Shafer evidence theory is used for decision-level fusion [91]. From these broad applications it can be seen that sensor fusion is a notably effective tool for process monitoring by using heterogonous sensors.

CHAPTER III

OVERALL RESEARCH METHODOLOGY

This dissertation proposes new solutions for monitoring of ultra-precision manufacturing process. These solutions have mainly four parts (1) feature extraction, (2) process state prediction, (3) decision making, (4) sensor fusion, which are mentioned in the introduction chapter. Methodologies for these four modules are proposed throughout this dissertation for an effective monitoring methodology of UPM process. Figure 3.1 shows modules used for each chapter.

3.1 Nonlinear Sequential Bayesian Analysis-Based Decision Making for Chemical Mechanical Planarization Process (Chapter 4)

In CMP process end-point facilitates decision on when to stop the planarization process. In this study we use vibration sensor data to monitor CMP process for detecting end-point.

Challenges in process monitoring CMP include as: (1) the CMP process is complex, non-linear, and non-stationary [92-95], which brings significant difficulties for the process monitoring and control, and (2) the signal to noise (S/N) ratios in CMP tend to be low with conventional sensors, which causes

inherent uncertainty in the data collection and modeling [64]. In order to overcome these challenges we integrated sequential Monte Carlo analysis with decision theory. Overall methodology is summarized in Figure 3.2.

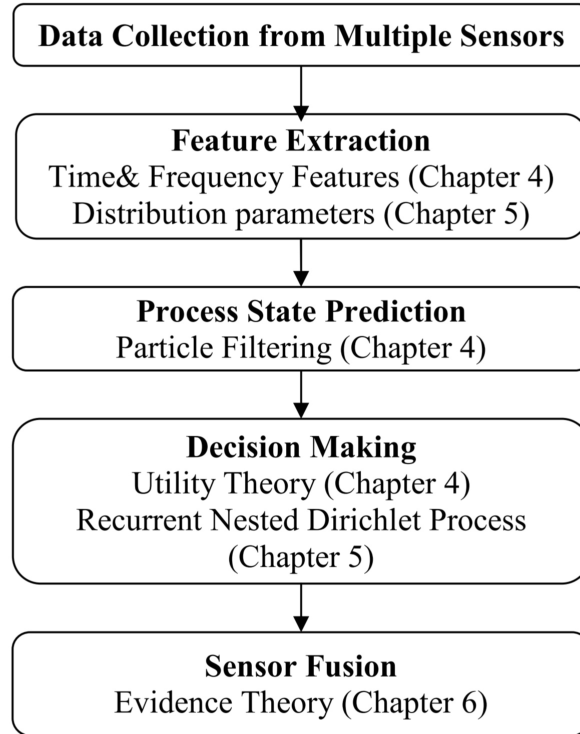


Figure 3.1 Modules and techniques used in this dissertation

Based on the sensor data captured by vibration sensor in CMP experiments, a nonlinear regression method is applied to relate sensor information with material removal rate (MRR). A non-linear sequential Bayesian analysis is used to predict process state. With integration of sequential Bayesian analysis and non-linear regression model we can predict MRR for the next step. The amount of material

removed can be calculated with MRR and decision theory can be applied to detect end-point.

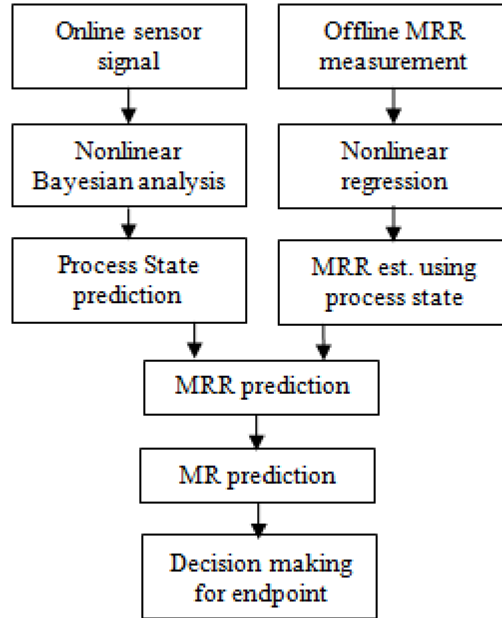


Figure 3.2 Overall research methodology for CMP end-point detection

3.2 Process Monitoring by Using Evolutionary Clustering Analysis for Chemical Mechanical Planarization (CMP) (Chapter 5)

This research is based on Cu-CMP experiments by using MEMS vibration sensors for process monitoring. The non-linear dynamics of the CMP process state causes the sensor signal collected from MEMS vibration sensors distributed as non-Gaussian. This non-Gaussian distribution can be modeled as mixture of Gaussian distributions. The overall framework of this study is demonstrated in Figure 3.3.

Sensor data is collected from MEMS vibration sensor and sliding windows is used for sensor data analysis. Dirichlet Process [96] is used to model sensor data for the initial window. Recurrent Dirichlet Process [72] is applied within each sliding window to formulate the non-Gaussian distribution of the sensor signal using a mixture of Gaussian distributions. Finally, the change in constructed mixture models for each sliding window will be detected by developing a novel Recurrent Nested Dirichlet process.

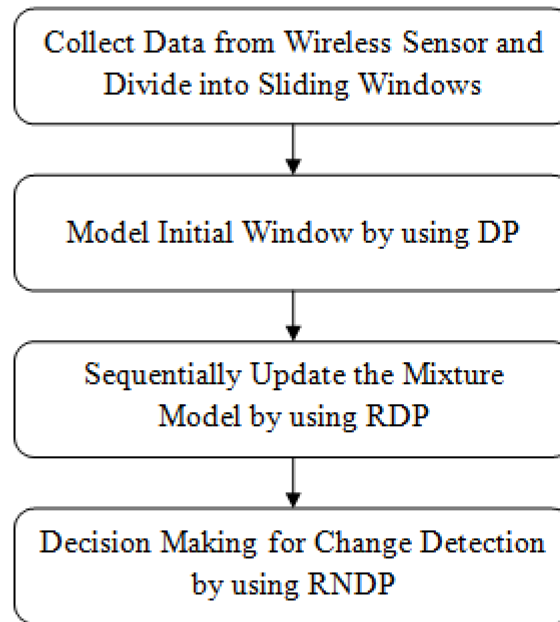


Figure 3.3 Overall methodology for CMP monitoring by RNDP

3.3 Multi-sensor Fusion Based Process Monitoring of Ultra Precision Machining (Chapter 6)

Several research efforts have been made to monitor the UPM process by using sensors. Vibration [97], acoustic emission (AE) [98], temperature [99] sensors

have been used for monitoring the UPM process. While these methods are, to various extends, effective to detect anomalies in the process, combination of sensor information may be able to further improve the accuracy of detection. However, this type of work has not been reported in the literature of UPM process. In this study we combine sensor signals from heterogonous sensors in order to get more precise and reliable decisions. Overall framework for this research is summarized in Figure 3.4

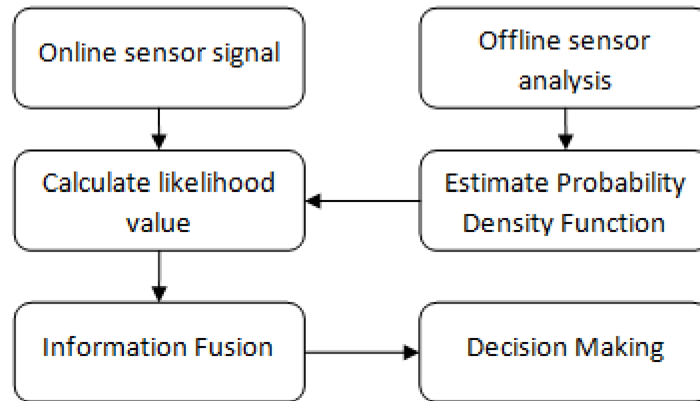


Figure 3.4 Proposed overall methodology for sensor fusion for UPM process monitoring

Sensor data are captured by heterogonous sensors such as, vibration and force, from UPM experiments with different depth of cut settings. A non-parametric Bayesian modeling is used to classify each condition for each sensor. A decision is made by each sensor signal for the testing data point. Information fusion technique is used to fuse these decision in order to detect and classify any change during the process.

CHAPTER IV

NONLINEAR SEQUENTIAL BAYESIAN ANALYSIS-BASED DECISION MAKING FOR HIGH PRECISION MACHINING

Chemical Mechanical Planarization (CMP) process has been widely used in the semiconductor manufacturing industry for realizing highly polished (surface roughness Ra ~ 1 nm) and planar (WIWNU ~ 1%, thickness variation standard deviation (SD) ~ 3 nm) surfaces of an in-process wafer. In CMP, accurate and timely decisions for endpoint detection (EPD) are extremely important to enable the process to effectively respond to demand variations and disruptions. In this paper, we apply nonlinear sequential Bayesian analysis and decision theory to establish a quantitative relationship that connects the measured sensor signal features (inputs) with the process performance measure such as material removal (outputs) for end-point detection (EPD) of the CMP process. A case study with actual CMP data is provided to demonstrate the effectiveness of the present approach.

4.1 Introduction

In the semiconductor industry, the relentless competition and customer driven

demand for functionality and versatility have led to significantly increased circuit density [100]. One of the key enablers for this incredible development is the chemical mechanical planarization (CMP) process which is one of the most critical and widely used operations in the semiconductor industry for realizing highly polished (surface roughness Ra ~ 1 nm) and planar (WIWNU ~ 1%, thickness variation standard deviation ~ 3 nm) surfaces of inter-level dielectrics and metal stud levels of wafers in the fabrication of integrated circuits (IC) [101].

In the CMP process, accurate and timely decisions for process monitoring and control, such as end-point detection (EPD) that facilitates decisions on when to stop the polishing process, and adjustment of process parameters for optimal performance are extremely critical to enable the process to effectively respond to demand variations and disruptions, and to ensure the quality and productivity requirements of the wafer yields [92, 100]. Recognizing this need, the semiconductor manufacturing enterprises have been investing in various sensor technologies used for process monitoring and control. However, accurate prediction and consistent realization of quality parts are still major challenges in industry. The reasons for this gap include: (1) the CMP process is complex, nonlinear, and nonstationary [92-95], which brings significant difficulties for the process monitoring and control, and (2) the signal to noise (S/N) ratios in CMP tend to be low with conventional sensors, which causes inherent uncertainty in the data collection and modeling [64]. Due to these obstacles, most of the current

approaches used are based on linear, largely deterministic and stationary models while the actual processes are far from that. Therefore, it is imperative to develop a methodology that is capable of tackling the nonlinearity and uncertainty of the CMP process to facilitate accurate and timely process monitoring and control, such as the EPD.

4.2 Research Methodology for End-point Detection

This investigation is based on sensor-based Cu-CMP experiments. In the experimental study, material removal rate (MRR) is measured offline. With the sensor-based CMP experiment setup, the instantaneous pad-wafer deflections are considered as process state variables which are captured by the wired and wireless vibration sensor signals during the CMP process. Features in both time and frequency domains of the recorded time series of vibration signals are extracted and further compressed using the principal component analysis (PCA). The resulting features are used to quantify the state variables in this study.

The proposed nonlinear sequential Bayesian-based decision-making approach is summarized in Figure 4.1. Based on sensor data captured in the Cu-CMP experiments, a nonlinear Bayesian analysis is proposed to sequentially predict (one step ahead) the process state, and a nonlinear regression method, such as neural network (NN) analysis is applied to relate the process state variables with process performance measures, such as material removal rate (MRR). Thereafter, by integration of the above sequential Bayesian analysis and nonlinear regression,

the MRR can also be predicted sequentially. With the predicted MRR, the amount of material that would have to be removed (MR) can be predicted conveniently. Thus, by applying the decision making theoretic approach to compare the predicted MR with the specified threshold value of MR, the endpoint of the CMP process can be properly detected. In Section 4.2.1 we present the nonlinear sequential Bayesian method for process state and performance prediction, based on which we apply decision making approach for CMP endpoint detection in Section 4.2.2.

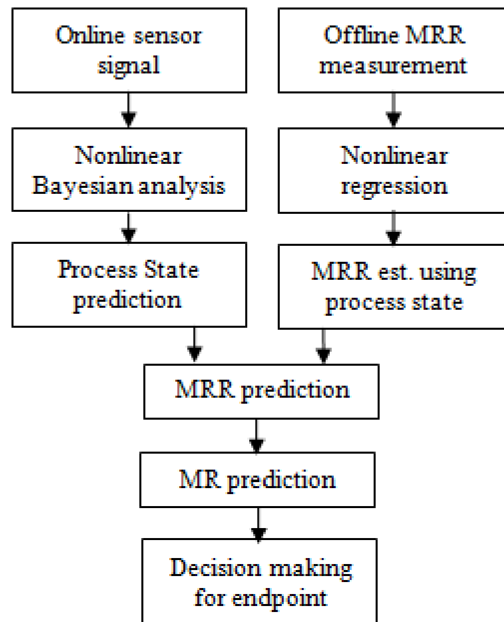


Figure 4.1 Proposed overall research methodology

4.2.1 Online predictive modeling for process state

The process state $\boldsymbol{x}_k$ (the instantaneous pad-wafer deflection in CMP process) at time k is captured using online vibration sensor data $\boldsymbol{y}_k$. The underlying relationship between the dynamic state $\boldsymbol{x}_k$ and measurement $\boldsymbol{y}_k$ is captured using the observation equations of the form,

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}, u_k, \omega_{k-1}) \tag{4.1}$$

$$\boldsymbol{y}_k = h_k(\boldsymbol{x}_k, v_k) \tag{4.2}$$

where f_k is the state transition function, u_k is the control input (can be treated as a decision or an action), ω_{k-1} is white noise, h_k is the observation function, and $v_k$is the observation noise. The Bayesian approach is used to dynamically estimate and predict the state. The purpose of state estimation is to determine the distribution of state vector ($\boldsymbol{x}_k$) given a sequence of observations at various times, $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, up to $\boldsymbol{y}_k$, i.e., to estimate $p(x_k|y_{1:k})$. State prediction is to estimate the distribution of process state one step ahead, i.e., to estimate $p(x_{k+1}|y_{1:k})$. Based on the Bayesian theory, the conditional probability density function (PDF) of the state can be estimated from the observation as,

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)}{p(\boldsymbol{y}_k|\boldsymbol{y}_{1:k-1})} p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) \tag{4.3}$$

Once $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ is estimated, the function of the future state $\boldsymbol{x}_{k+1}$ conditional on measurements $\boldsymbol{y}_{1:k}$ can be predicted as:

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})d\boldsymbol{x}_k \tag{4.4}$$

In these formulations, Eq. (4.3) is the update function which provides posterior probability in Eq. (4.4), i.e., the prediction function, which in turn will provide prior probability for the update function. This is a recursive process that evolves over time.

The recursive process represented by Eqs. (4.3) and (4.4) forms the basis for the optimal Bayesian solution [51]. However, Eqs. (4.3) and (4.4) only illustrate a conceptual solution, which cannot be solved analytically, if the state transition function f_k is nonlinear. This is because the nonlinear function f_k will cause the posterior distribution of state $\boldsymbol{x}_k$ to be non-Gaussian. It is very difficult to draw samples from a non-Gaussian posterior distribution. Thus, the integral operation in Eq. (4.4) cannot be solved analytically.

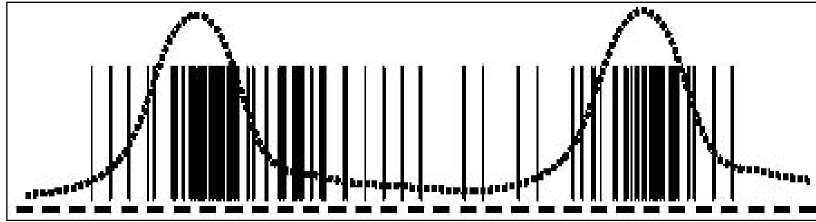


Figure 4.2 Representation of multimodal distribution using samples

In order to tackle the challenge related to the non-analytical solution for Eqs. (4.3) and (4.4), a sequential Bayesian analysis based on Monte Carlo method (also called particle filter) [51] is used in this study to recursively compute Eqs. (4.3) and (4.4). The basic concept behind particle filter method is that any probability

density function (PDF) can be approximately represented by a set of particles (samples) as shown in Figure 4.2. Although it is challenging to analytically formulate an arbitrary probability density distribution, an effective approximation for the discrete distributions is possible, such as,

$$p(x) \approx \sum_{i=1}^{N} \omega(i)\delta(x - x(i)) \tag{4.5a}$$

$$\omega(i) \propto \frac{p(x_i)}{q(x_i)} \tag{4.5b}$$

where $x(i)$ is the ith sample that approximates the distribution; set $x_i \sim q(x)$, $i = 1,\ldots, N$ be samples that can be easily drawn from another distribution $q(.)$ called importance density; and δ is the Dirac delta function; the ratio between the posterior PDF and the selected importance density function is the associated weight $\omega(i)$ with the sample $x(i)$ as shown in Eq. (4.5b).

Eq. (4.1) is actually a Markov process. In this study, in order to model the CMP process more accurately, a more general scenario, namely, a higher order Markov process is considered,

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}, \ldots, \boldsymbol{x}_{k-p}, u_k, \omega_{k-1}) \tag{4.6}$$

where p is the time lag. Based on Eq. (4.6), we can choose an important density that can be factorized as,

$$q(x_{0:k}|y_{0:k}) = q(x_k|x_{k-p:k-1}, y_{0:k})q(x_{0:k-1}|y_{0:k-1}) \tag{4.7}$$

Once the process measurement $y_{0:k}$ is available, it can be used to update the process state $x_{0:k}$. Thus, the posterior distribution of the process state can be formulated as follows,

$$p(x_{0:k}|y_{0:k}) = \frac{p(y_k|x_{0:k})p(x_{0:k}|y_{0:k-1})}{p(y_k|y_{0:k-1})}$$

$$= \frac{p(y_k|x_{0:k})p(x_k|x_{0:k-1},y_{0:k-1})p(x_{0:k-1}|y_{0:k-1})}{p(y_k|y_{0:k-1})} \tag{4.8}$$

In Eq. (4.8), based on Eq. (4.2), we used $p(y_k|x_{0:k}) = p(y_k|x_{0:k}, y_{0:k-1})$. Based on Eqs. (4.2) and (4.6), Eq. (4.8) can be further formulated as,

$$p(x_{0:k}|y_{0:k}) = \frac{p(y_k|x_k)p(x_k|x_{k-p:k-1})p(x_{0:k-1}|y_{0:k-1})}{p(y_k|y_{0:k-1})}$$

$$\propto p(y_k|x_k)p(x_k|x_{k-p:k-1})p(x_{0:k-1}|y_{0:k-1}) \tag{4.9}$$

Thus, the weights can be formulated as based on Eqs. (4.5b) and (4.7),

$$\omega_k \propto \frac{p(y_k|x_k)p(x_k|x_{k-p:k-1})p(x_{0:k-1}|y_{0:k-1})}{q(x_k|x_{k-p:k-1})q(x_{0:k-1}|y_{0:k-1})}$$

$$= \omega_{k-1}\frac{p(y_k|x_k)p(x_k|x_{k-p:k-1})}{q(x_k|x_{k-p:k-1})} \tag{4.10}$$

In this study, we choose the important density as the prior distribution that is already available, namely, $q(x_k|x_{k-p:k-1}) = p(x_k|x_{k-p:k-1})$. Thus, Eq. (4.10) becomes,

$$\omega_k \propto \omega_{k-1}p(y_k|x_k) \tag{4.11}$$

where $p(y_k|x_k)$ is the likelihood of measurement $y_k$which can be estimated by the measurement noise v_k in Eq. (4.2).

In summary, using Eq. (4.6), the process state can be predicted one step ahead; and then using Eq. (4.11) the weights of the particles of the process state can be updated by using the measurement data. Therefore, by using this sequential Monte Carlo sampling method, Eqs. (4.3) and (4.4) can be sequentially and recursively solved to estimate and predict the state vector. The pseudo-code of the particle filter method is provided in Table 4.1.

Table 4.1 Pseudo code of the particle filter method with multiple time lags (*N* is number of particles, *k* is the time index, *p* is the time lag)

$[(x_k^i, \omega_k^i)_{i=1}^N]$=PF$[(x_{k-1:k-p}^i, \omega_{k-1}^i)_{i=1}^N, y_k]$
For i = 1: N
    Generate a particle $x_k^i \sim q(x_k | x_{k-1:k-p})$
    Assign the particle with a weight ω_k^i, according to Eq. (4.11)
End

Since the state-space model (Eq. (4.1)) may not be available upfront for the CMP process, the model structure and parameters need to be determined and estimated. Because of the complex nature of the CMP process, a simple polynomial state-space model may not be able to capture the non-linearity. Logistic model is applied in this study to formulate the nonlinear state-space model, which is an effective technique to handle non-linearity due to its complex model structure [102]. Its mathematical expression is

$$x_k = \frac{\alpha_1}{1+e^{-\theta_1 x_{k-1}}} + \frac{\alpha_2}{1+e^{-\theta_2 x_{k-2}}} + \cdots + \frac{\alpha_p}{1+e^{-\theta_p x_{k-p}}} + \beta \qquad (4.12)$$

where x_k represents the process state at time k, p is the time lag, and α, θ, and β represent model parameters. With the model structure (Eq. (4.12)) determined, the problem of estimating the model parameters (α, θ, and β) can be handled by simultaneously estimating the state variables and state model parameters. Bayesian theory provides the rule of the joint distribution,

$$p(\boldsymbol{x}_k, \theta_k | \boldsymbol{y}_{1:k}) = p(\boldsymbol{y}_k | \boldsymbol{x}_k, \theta_k) p(\boldsymbol{x}_k | \theta_k, \boldsymbol{y}_{1:k-1}) p(\theta_k | \boldsymbol{y}_{1:k-1}) \qquad (4.13)$$

where θ_k is the vector of the state-space model parameters (α, θ, and β). By augmenting the state vector x_k with the parameter vector θ_k, the combined state and parameter estimation can be performed. In this investigation, we utilized the Kernel smooth approach proposed in Ref. [103] to estimate model parameter which follows the joint distribution of Eq. (4.13). For the details of the implementation Ref. [103] may be referred.

For another important state-space model parameter in Eq. (4.12), namely, the time lag p, we determined its value in prior using the CMP experimental sensor data. The mutual information is applied to determine its optimal value (Eq. (4.1) is in fact a special case with $p = 1$). The mutual information of two random variables X and Y is a quantity that measures the mutual dependence of two random variables [104],

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (4.14)$$

where $p(x, y)$ is the joint probability density function of random variables X and Y, and $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y, respectively. For random time serious (x_k, x_{k-1},…, x_{k-p}, x_{k-p-1},…, x_0), if we let X = (x_k, x_{k-1}, x_{k-2}, x_{k-3},…) and Y = (x_{k-p}, x_{k-p-1}, x_{k-p-2}, x_{k-p-3},…), the best time lag p in Eq. (4.12) is chosen as the smallest p value that results the minimal value of the mutual information measured by Eq. (4.14)

4.2.2 Performance measure (MRR) prediction based on predicted process state

The determination of endpoint followed by a decision on whether to stop the process at the current time index k can be made based on comparing the amount of material to be removed against a specified threshold value. At any time k, the decision to stop polishing (u_k=0) is made if the threshold is to be reached. Otherwise, new sensor signals ($\mathbf{y}_{k+1}$) are collected and used for further state estimation and prediction, and then a new decision will be made.

Our studies on CMP process [37] have indicated that instantaneous pad-wafer deflection of the CMP process, namely, the process state x_k is significantly related to MRR z_k. Thus, MRR can quantitatively be related to the state variables using nonlinear regression analysis (e.g., Neural Network). In Section 3.1, the online process dynamics model is generated based on the processed vibration sensor signals using nonlinear Bayesian analysis, which is an effective approach for

process state estimation and further prediction. The final step is to combine this online dynamics model and the nonlinear regression model to develop the MRR prediction model. Once MRR is predicted, the estimated total material removed at a specified time k is given by,

$$\xi_k = \sum_{j=0}^{k} z_j \Delta t_j \tag{4.15}$$

where Δt_j denotes a time interval between two successive predictions of MRR (z_j).

4.2.3 Decision making approach for CMP endpoint detection

Assume that in the CMP process, there are a list of d_1, d_2, ..., and d_m of decisions (such as stop or continue the polishing), and ϕ_1, ϕ_2, ..., and ϕ_n of events (such as the endpoint is reached or not reached) with the uncertainty that is the probability $p(\phi_j)$ of event ϕ_j ($j=1,...n$). The optimal decision among the possible ones (d_1, d_2,..., and d_m) should be chosen to determine the actions (control input u_k in Eq. (4.1)) to adjust the CMP process. This can be achieved using maximal expected utility theory [105].

As shown in

Table 4.2, the utility (preference) of the consequence corresponding to decision i on event j, c_{ij}, is determined by a utility function. For endpoint (EP) detection of CMP, there are two decisions, namely, d_1: continue polishing and d_2: stop polishing, and two events, ϕ_1 : under-polish and ϕ_2: over-polish. The two correct

decisions, i.e. to stop polishing if it is over-polish, and to continue if under-polish, have significant utility (c_1), while the two incorrect decisions have less utility (c_2 and c_3), as shown in Figure 4.3.

In order to compute the expected utilities for different decisions [105], the probabilities for each event, $p(\phi_j)$ shall be determined. They can be initially estimated based on historical data. Thereafter, once the *in situ* sensor data y is available, $p(\phi_j)$ will be updated as $p(\phi_j|y)$. Thus, the optimal decision (d_i) will be chosen based on maximal expected utility [105], namely,

$$\max_i \sum_{j=1}^{N} c_{ij} p(\phi_j \mid y) \tag{4.16}$$

Table 4.2 Decision Table for preferences

	ϕ_1	ϕ_2	…	ϕ_n
d_1	c_{11}	c_{12}	…	c_{1n}
d_2	c_{21}	c_{22}	…	c_{2n}
…	…	…	…	…
d_m	c_{n1}	c_{n2}	…	c_{mn}
$Prob.$	$p(\phi_1)$	$p(\phi_2)$	…	$p(\phi_n)$

$p(\phi_j \mid y)$ in Eq. (4.16) can be estimated using the sequential Monte Carlo method presented in Section 3.1. It can be seen from Figure 4.2 that the predicted process state has a form of empirical probability density function (PDF) approximated by the particles. Thus, we are able to compute the empirical distribution of material removed (MR) using Eq. (4.15). Figure 4.4 depicts an example of the empirical distribution of material removed at time k.

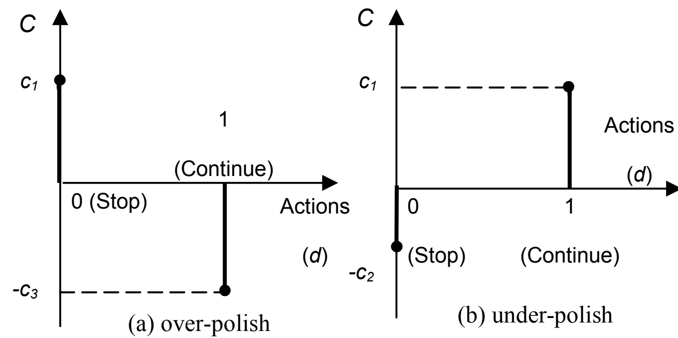


Figure 4.3 Utility function for EPD of CMP

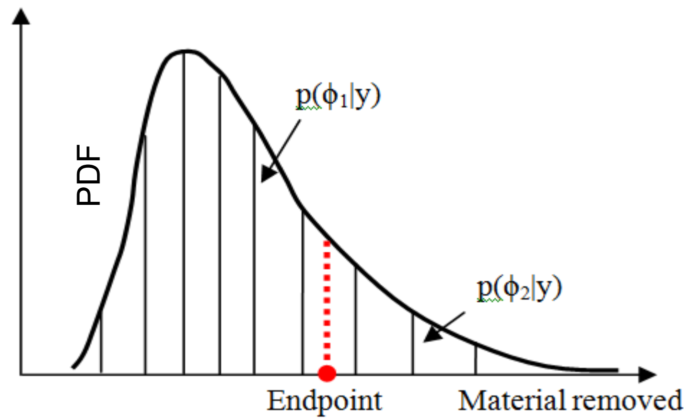


Figure 4.4 Empirical distribution of material removed and estimation for probability of events ϕ_1 (under-polish) and ϕ_2 (over-polish)

Since the specified amount of material that should be removed to reach the endpoint can be accurately calculated in advance (marked as "Endpoint" in Figure 4.4), based on the thickness of the copper layer of the wafer, $p(\phi_j \mid y)$ can be estimated accordingly, based on the empirical distribution of the MR. Assume the total number of samples used in the particle filtering method is N, and at time k the number of samples, whose value smaller than "Endpoint" (Figure 4.4) is m,

then the probability of event ϕ_1 (under-polish) and ϕ_2 (over-polish) can be easily estimated as follows,

$$p(\phi_1) = m/N \tag{4.17}$$

$$p(\phi_2) = 1 - m/N \tag{4.18}$$

Based on Eqs. (4.16) to (4.18), we can choose the decision d_i that will result in the maximal utility during the CMP process. This decision making process based on the particle filter method is carried out at every time k. Once the decision of action is switched from "continue polishing" to "stop polishing," then the endpoint of the CMP is detected.

4.3 Case Studies

In order to track the variations of various slurry chemistry parameters of Cu-CMP and predict the performance (MRR) of the CMP process, polishing experiments were carried out on a LapMaster 12 bench top lapping machine. For the experiments, the selection of proper sensors is carefully investigated. Wired and wireless sensors have their advantages and disadvantages for CMP experiments. Wired sensors have higher sampling rate. However, the existence of wires makes it difficult for sensor installation, especially, if one wants to put the sensor close to the site to be analyzed. For wireless sensors, although their sampling rate is lower, the installation could be much easier since no wires are involved. Thus, an *in-situ* sensor can be implemented with wireless sensors, which enables more accurate

data acquisition (for lower frequency signal). By considering the aforementioned factors, we use both wired and wireless sensors due to their complementary features. Two vibration sensors were mounted on the polishing machine, namely, one wired accelerometer (KISTLER) sampling at 5kHz, and one wireless accelerometer (MOTEiv) sampling at 500Hz. These two sensors are used to capture the instantaneous pad-wafer deflection during the CMP operation. The CMP setup is shown in Figure 4.5 Experimental setup using a LapMaster L 12 Bench-top lapping machine.

Figure 4.5 Experimental setup using a LapMaster L 12 Bench-top lapping machine.

4.3.1 Offline regression analysis connecting process state with process performance

In the experiment, a Taguchi L12 array was used as the design matrix. The machine settings, namely, pH of polishing slurry, flow rate of the slurry, amount of complexing agent, and amount of BTA were set at two levels, high or low.

During the CMP experiment, the MRR data is obtained using the following method,

$$MRR = \frac{\Delta W}{\Delta t} = \frac{W_i - W_f}{t_f - t_i} \tag{4.19}$$

where W_i is the initial wafer weight before polishing and W_f is the final wafer weight after polishing. Δt is the duration of a polishing cycle. The instantaneous pad-wafer deflection that determines the process state variables are collected using wired and wireless vibration sensors.

In the experiment, the wired vibration sensor data is captured at a frequency of 5k Hz, and wireless sensor data at 500 Hz. The time series sensor signals are extracted to the following eight features in time and frequency domains, i.e., peak-to-peak value (PTP), first two dominant frequencies and amplitudes, standard deviation, skewness, and kurtosis values. For wireless sensor signal, six features are extracted which are basically the same as the features extracted from wired sensors data except that only one dominant frequency and amplitude is extracted.

After the features are extracted from the original time series data, in order to reduce the impact of the noise in the data, a moving average with a range of ten is applied to smooth the features. Afterwards, the principal component analysis (PCA) is conducted to compress these features. A threshold value of 95% is used to determine the number of eigenvalue /eigenvector to be remained. Finally, five

out of the eight features from wired sensor data and four out of the six from wireless sensor data, i.e., a total of nine features are considered and treated as state variables for process-state prediction and further regression analysis.

In order to relate the nine features extracted from wired and wireless sensor data with material removal rate (MRR), a multilayer perceptron (MLP) neural network (NN) is used for regression analysis, with 2 layers, and 9 hidden neurons, as shown in Figure 4.6. 200 training points and 122 testing points are chosen for analysis. The resulting R-square value for testing data points is very satisfactory (90.14%).

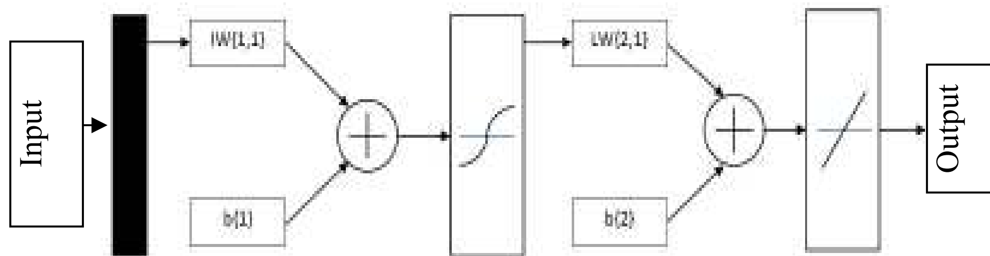


Figure 4.6 The multilayer perceptron neural network for regression analysis between sensor signals and MRR

Regarding the online state estimation and prediction, the algorithm of the particle filter (Table 4.1) using state space model of logistic function was implemented using Mablab^{TM}. In this case study, the particle filter method was implemented with 500 samples (particles). The logistic model (Eq. (4.12)) was utilized to formulate the state space model (Eq. (4.1)). The method of mutual information analysis (Eq. (4.14)) presented in Section 3.1 is applied first to determine the

value of time lag p in Eq. (4.12). As shown in Figure 4.7, the lag value is chosen as seven.

4.3.2 Predictive model for process state variables

In Section 4.3.1, nine features (state variables) are extracted from sensor signals to represent the process state, and predicted sequentially. In order to demonstrate the effectiveness of the predictive method presented in Section 4.2.1, the expected value of feature 1 based on the empirical distribution is computed at every second and plotted in Figure 4.8, in which the actually sensor data is the solid curve and the predicted data dashed. The nine features should be predicted in order to predict the MRR. The remaining eight plots for features 2 to 9 are not shown here due to limitation in space, but all of them show a close consistency between the predicted and the actual sensor data.

In order to demonstrate the advantage of the logistic function based particle filtering method over the other widely used predictive methods, such as particle filter method using polynomial function, time series (ARMA(3,3)), Kalman filtering (KF) and extended Kalman filtering (EKF), by using the same data set, the aforementioned methods are also applied to predict the sensor data. The mean square errors (MSE) between the prediction and actual sensor data are given in Table 4.3 for all nine states. Figure 4.9 shows a comparison of MSE of predicted sensor signals using various prediction methods. The superiority of the particle

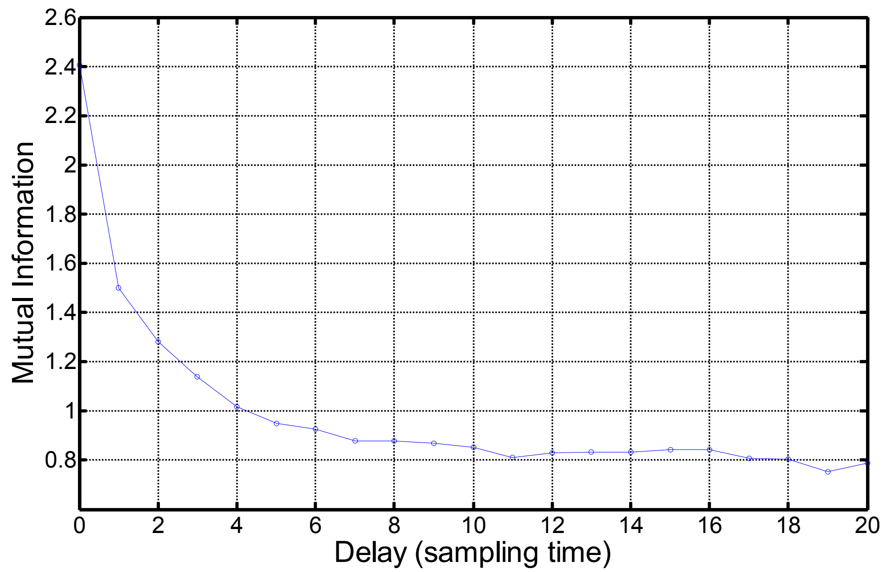


Figure 4.7 Mutual information analysis for determining time lag

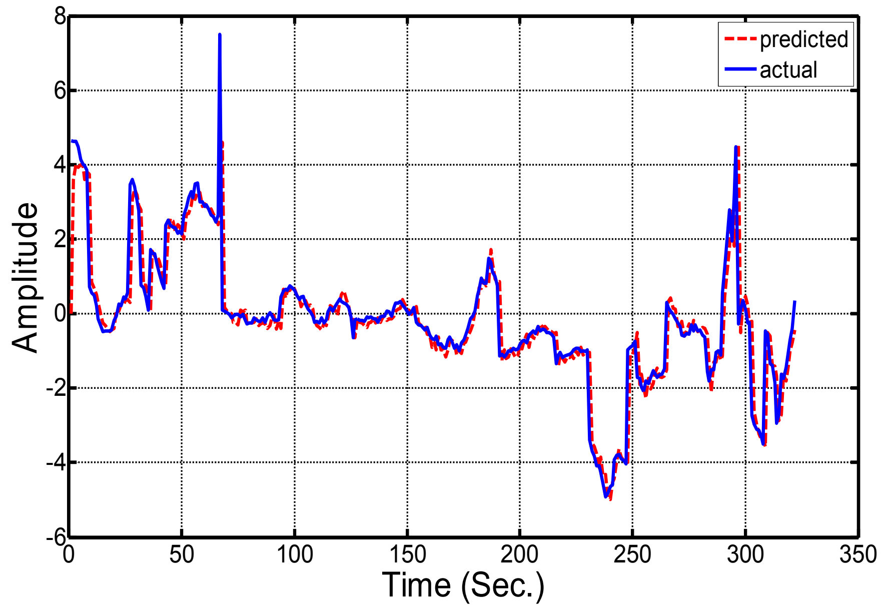


Figure 4.8 Predicted result for feature 1 using particle filter with logistic function

filtering method can be seen from the following results: among the predictions for all the 9 states, particle filtering has the smallest MSE in 6 features, EKF 3, ARMA(3,3) 0, and KF 0.

Table 4.3 Comparison of MSE of predicted feature using different methods

	PF (logistic)	PF (polynomial)	ARMA (3,3)	EKF	KF
Feature1	0.32	0.39	0.54	0.50	0.52
Feature2	0.11	0.14	0.27	0.27	0.29
Feature3	0.11	0.13	0.17	0.16	0.19
Feature4	0.06	0.08	0.34	0.28	0.31
Feature5	0.10	0.13	0.12	0.13	0.15
Feature6	0.51	0.53	0.57	0.49	0.53
Feature7	0.35	0.55	0.43	0.45	0.55
Feature8	0.25	0.36	0.24	0.22	0.26
Feature9	0.25	0.26	0.28	0.20	0.23

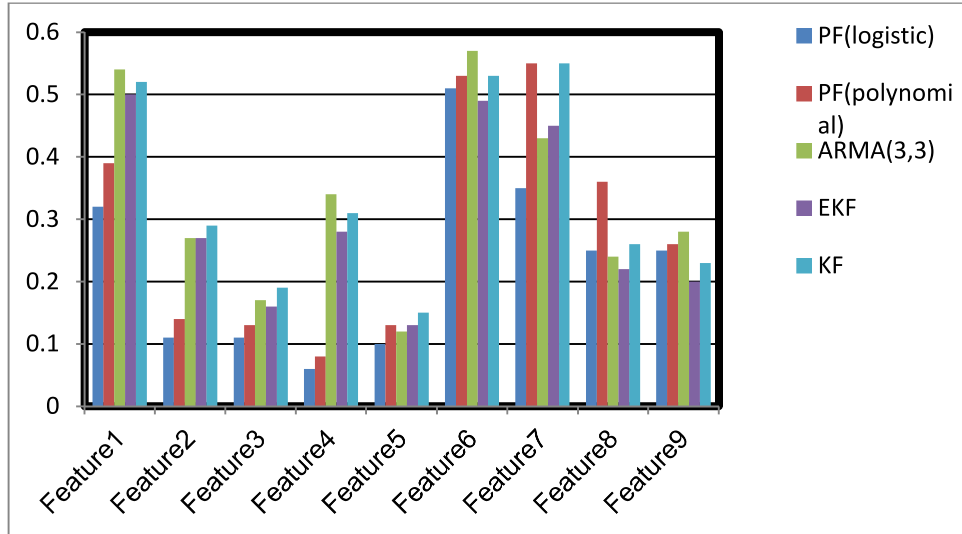


Figure 4.9 Comparison MSE of predicted feature using various methods, i.e., PF(logistic), PF(polynomial), ARMA(3,3), EKF, and KF

4.3.3 Predictive model for MRR

Once the process state represented by the nine sensor features are predicted, the neural network model presented in Section 4.3.1 is used to predict the corresponding MRR. In a similar vein, as the sensor features are predicted, the expected value for MRR is computed for every second. Figure 4.10 depicts the measured MRR versus predicted MRR. For comparison purpose, besides the MRR prediction using the particle filtering method with logistic function, the MRRs predicted by other methods, i.e., particle filter with polynomial function, ARMA(3,3), KF and EKF also are also obtained. The resulting R-square values using these methods for MRR prediction are illustrated in Figure 4.11, which shows that the particle filtering method with logistic function achieves more satisfactory results than other predictive methods.

For example, Figure 4.12 shows the empirical PDF of MR at the 275th second. Thus, by using Eqs. (4.17) and (4.18), at every second, the probability of the events (under-polishing and over polishing) can be estimated accordingly.

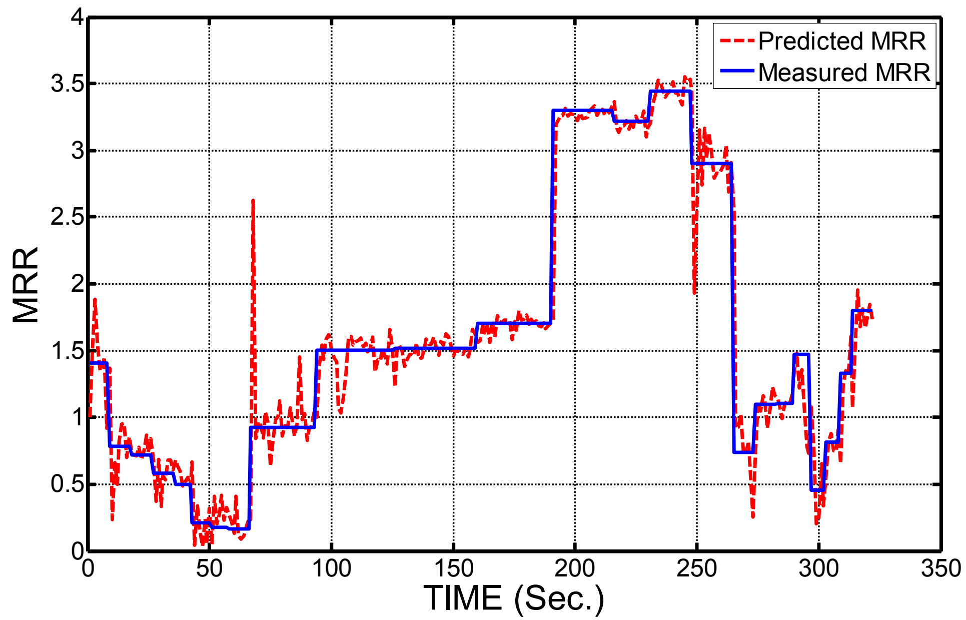


Figure 4.10 Predicted MRR vs. measured MRR

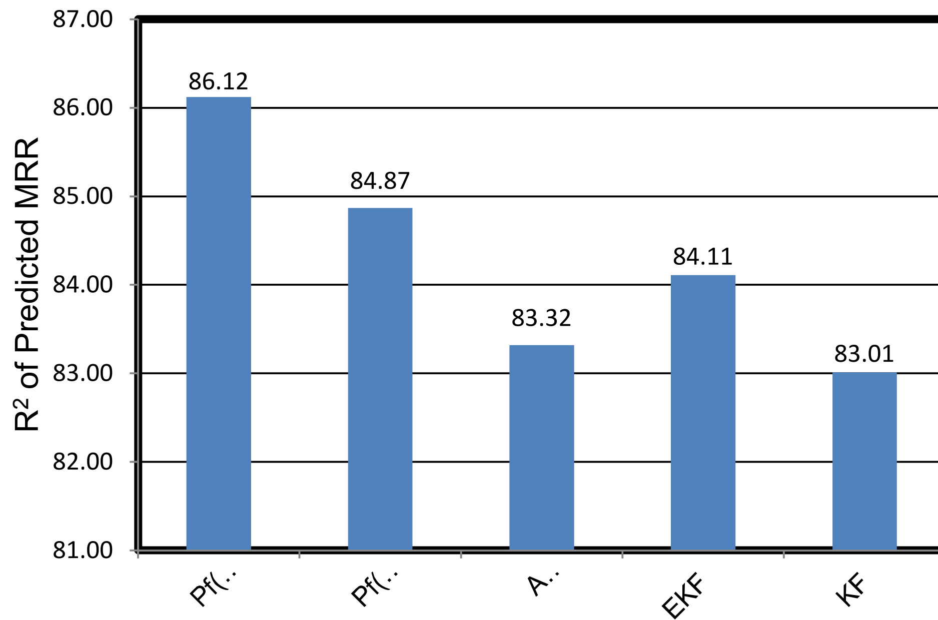


Figure 4.11 Predicted MRR using various methods

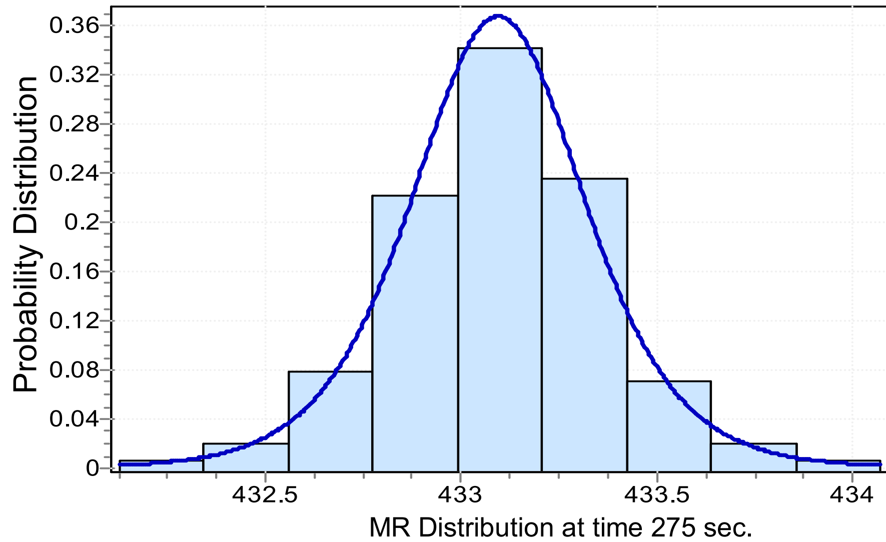


Figure 4.12 Empirical distribution of material removed at the 275th second

4.3.4 Decision making for CMP endpoint detection

In this case study, we assume that the threshold value of endpoint is 450 mg, i.e., the polishing should be stopped when material with weight of 450 mg is removed.

Table 4.4 Decision table for CMP endpoint detection

	ϕ_1 : Under-polish	ϕ_2 : Over-polish
d_1: Continue polishing	c_{11}: 1	c_{12}: -1
d_2: Stop polishing	c_{21}: -0.8	c_{22}: 1
Probability	$p(\phi_1)$	$p(\phi_2)$

Table 4.4 gives the utility values used in this CMP endpoint detection.It may be noted that $c_{12} < c_{21}$ because over-polishing is less desirable than under-polishing. Eq. (4.16) is applied to calculate the expected utility for both decisions d_1 (continue polishing) and d_2 (stop polishing) at every second. Figure 4.13 shows

the two expected utility curves for d_1 and d_2 with a threshold value of 450 mg (material removal) for endpoint. Based on the maximal expected utility criteria, we should take the decision with larger utility value between d_1 and d_2. Thus, the endpoint, namely, when to stop polishing can be identified by the intersection between the two utility curves in Figure 4.13.

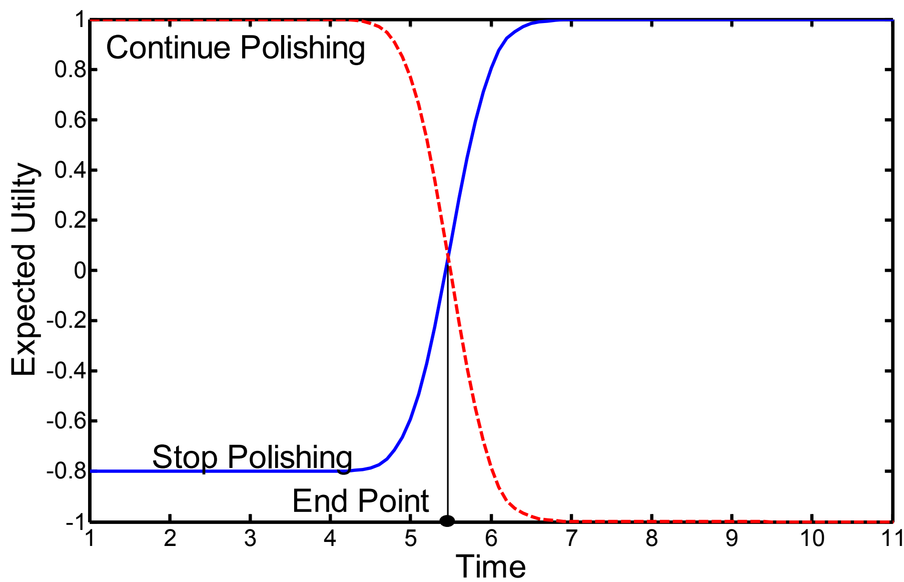


Figure 4.13 Expected utility for "continue polishing" and "stop polishing" with endpoint threshold of 450 mg material removal (MR) (On the x-axis the numbers 1-11 corresponds to 254.5 to 255.5 seconds)

Figure 4.14 shows the relationship between the time of endpoint and the material to be removed at the endpoint. The blue dashed curve was generated based on the actual measured MRR and the red sold curve was generated based on the predicted MRR. The consistency between these two curves demonstrates the effectiveness of the proposed method.

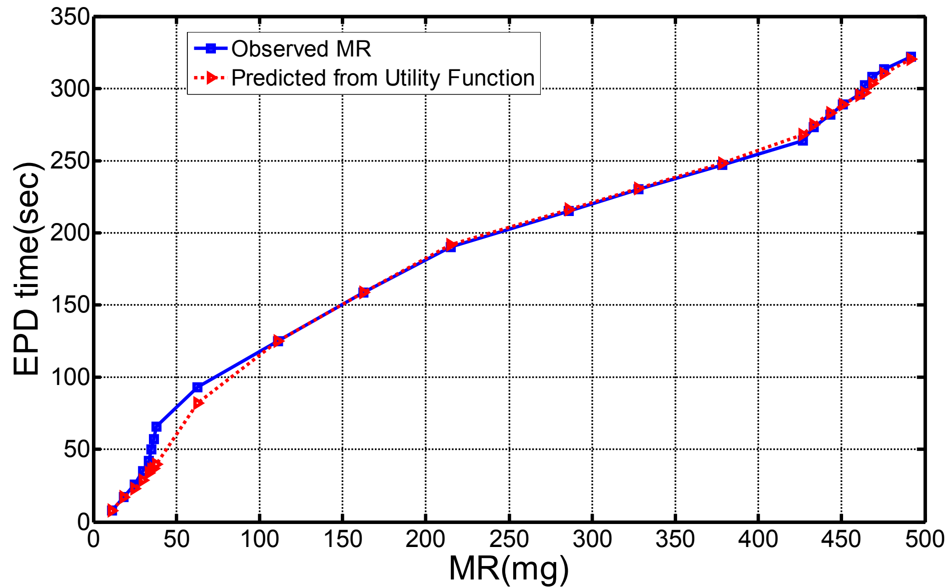


Figure 4.14 Relationship between the time of endpoint and material to be removed at endpoint

4.4 Summary

The main purpose of the methodology proposed in this investigation is to integrate nonlinear Bayesian method and decision theory to establish an online prediction model for accurate endpoint detection of the CMP process. The sequential Monte Carlo method was utilized for state prediction of a nonlinear dynamic process. During the CMP process, the raw vibration signals captured by the sensors are processed and a number of features are extracted. Thereafter, the neural network model that relates MRR with sensor signals is created with a satisfactory R-square value (90.14%) which indicates the high accuracy of the model. Based on the state prediction model and neural network nonlinear regression model developed, the MRR can be predicted accordingly. Thus, the

material removed at any time during the CMP process can be estimated and predicted as well. By combining maximum expected utility decision theory, the endpoint of the CMP process can be accurately detected.

CHAPTER V

PROCESS MONITORING BY USING EVOLUTIONARY CLUSTERING ANALYSIS FOR CHEMICAL MECHANICAL PLANARIZATION (CMP)

Conventional parametric methods are predominantly used for process monitoring in Chemical Mechanical Planarization (CMP) process. These linear-Gaussian statistical methods are reticent, and do not adequately capture evanescent CMP nano-defects owing to the inherent nonlinear and nonstationary nature of the process. We present a non-parametric Recurrent Nested Dirichlet Process (RNDP) approach which invokes data gathered from real-time *in-situ* MEMS wireless sensors. The MEMS vibration signal patterns are approximated using mixtures of Gaussian distributions. The resulting multi-modal distribution can capture the underlying complex spatio-temporal dynamics of the CMP process. We monitor these Gaussian mixture distributions to detect incipient process deviations characterized by the fast changing nature of mixture components. The ability of the RNDP-generated mixture of Gaussian distributions to detect fast-changing process deviations is demonstrated using several case studies. Simulated

and experimental case studies were conducted to illustrate the effectiveness of the proposed method by comparing with the widely used process monitoring techniques.

5.1 Introduction

Semiconductor wafer characteristics, such as local and global planarity, roughness, defects, etc., have a significant effect on the functional integrity of semiconductor devices and are therefore tightly controlled during manufacture [106]. Chemical Mechanical Planarization technique is used in semiconductor industry to achieve these necessary local and global scale wafer properties [14, 107].

Since CMP is often the final step before wafer test and packaging stages, wafer anomalies and defects resulting from CMP can lead to high revenue and yield losses [108]. The industry has identified CMP-induced wafer defects among to be the top five reasons inhibiting device yield. It is, therefore, desirable to detect the onset of CMP process anomalies at an early stage so that prompt corrective action can be taken.

In the last decade, semiconductor wafer sizes have increased from 200 mm (ca. 2000) to the current 300 mm (ca. 2010), with 450 mm wafer size production projected in 2017 – 2019 [1, 2]. Concurrently, device feature sizes have reduced from 130 nm (Intel Pentium III Tualatin, ca. 2000 – 2001) to 22 nm (Intel Ivy Bridge, ca. 2012) [3]. In addition, copper (Cu) has recently emerged as a viable

interconnect alternative to tungsten (W) and aluminum (Al) owing to the low specific electrical resistance of Cu [109-110], enabling faster processor speeds due to reduction in RC delay and crosstalk [109-110].

However, CMP of copper interconnects is challenging due to the relative softness of Cu and the low selectivity of Cu with Tantalum-Tantalum Nitride (Ta/TaN) barrier layers [110]. Therefore, in contrast to SiO_2 semiconductor wafers, Cu wafers are more liable to CMP induced defects and yield losses.

Conventional statistical process control (SPC) techniques are widely used for CMP process monitoring. SPC techniques based on Gaussian parametric statistics are of limited utility for the detection of incipient process anomalies and are typically found to be reticent in capturing subtle inherent process drifts (i.e., high Type II, failing to detect errors), such as gradual pad wear, slurry coagulation, slurry contamination, etc., from process noise. Also, SPC-based process monitoring entails offline inspection using preconditioned test wafers leading to as much as 35% reduction in yield [111]. Accordingly, SPC approaches do not afford real-time monitoring of CMP process anomalies.

Additionally, most of the existing sensor-based monitoring applications in CMP have been relegated to the objective of end-point detection as opposed to monitoring of process anomalies [18]. Efforts are seldom made to extend sensor-based monitoring in CMP to defect detection. Typically, in endpoint detection, the transition between inter-metallic to dielectric is readily apparent when using high

resolution AE sensors [47, 112-114]. In contrast, defects in CMP are mainly caused by evanescent physio-chemical phenomena at the nano-scale [115-116]. Therefore, while parametric methods may suffice for endpoint detection, the sensor signals arising from temporal fast-changing mechanics are characteristically nonstationary and nonlinear, and may not be readily detected by using statistical parametric models.

We present an approach that invokes Bayesian Dirichlet Process (DP) mixture models in order to detect incipient CMP process anomalies from vibration signal patterns acquired from close proximity MEMS sensor arrays. The main contribution of this work stems from the application of Bayesian DP models to represent sensor signal patterns and, consequently, detect subtle variations in signal patterns. Such an application of DP models for the analysis of sensor signal patterns in order to capture process anomalies has not been explored thus far. We demonstrate using experimentally acquired signal patterns that this novel DP-based monitoring approach can be valuable from a quality assurance perspective in CMP, as it can detect various CMP process anomalies, such as slurry depletion, gradual pad wear, and change in polishing load. The rest of this chapter is structured as follows: Section 5.2 presents the research methodology. Section 5.3 demonstrates the results of the case studies. Section 5.4 summarizes the results of this research methodology.

5.2 Proposed Research Methodology for Change Point Detection

We have conducted experimental investigations of Cu-CMP process; MEMS wireless vibration sensor data were during these studies. As the process changes in CMP process (such as dishing, erosion and defect occurring), the dynamics of vibration between the wafer and pad may change as well. It is possible that this change of dynamics can be captured by the MEMS vibration sensor employed in CMP experiment. Thus, the proposed methodology in this research is based on a hypothesis as follows: the process changes in the CMP polishing will be reflected in the change of vibration sensor signals; therefore, by capturing the change of sensor signals, the process changes in CMP can be identified consequently. The overall framework of the proposed methodology is summarized in Figure 5.1.

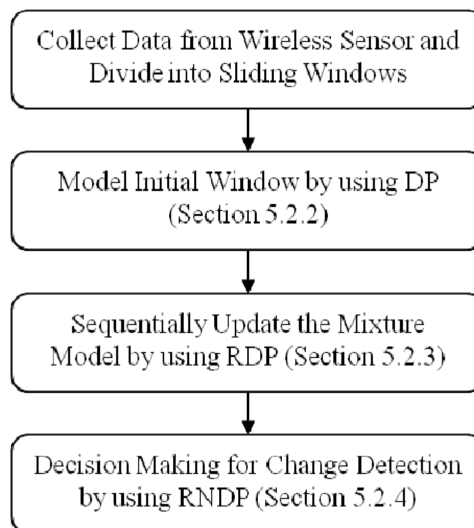


Figure 5.1 Overall Methodology for CMP monitoring by using RNDP

First, wireless MEMS vibration sensor data acquired and a sliding window technique is used for all subsequent analyses (Sec. 5.2.1). A recurrent Dirichlet process (RDP) model is constructed within each sliding window to estimate the non-Gaussian distribution of the sensor signal using a mixture of Gaussian distributions (Sec 5.2.2). Finally, the change in constructed mixture models for each sliding window will be detected by using Recurrent Nested Dirichlet Process (Sec. 5.2.3).

5.2.1 Experimental setup

A Buehler Automet 250 bench top CMP apparatus is used for planarization experiments. A tri-axis MEMS vibration sensor (ADXL 335) manufactured by Anolog Devices Inc. is mounted on the apparatus to collect sensor data. The sensor signals are sampled at 700 Hz and transmitted wirelessly to desktop computer using a wireless transmitter unit. The CMP setup and Xbee sensor network is shown in Figure 5.2.

Figure 5.2 Buehler Automet® 250 experimental CMP polishing setup and wireless Xbee Vibration Sensor

Blanket copper wafer disks of Φ1.625 inch (40.625 mm) are polished in KOH-based alkaline colloidal silica slurry medium, which has a constant flow rate of 20ml/min. Near-optical (Ra~ 5nm) quality surface finish blanket copper wafers are obtained on polishing with *a priori* identified optimal processing conditions wafers (Figure 5.3).

Figure 5.3 Polished blanket copper wafers after 12 min of CMP

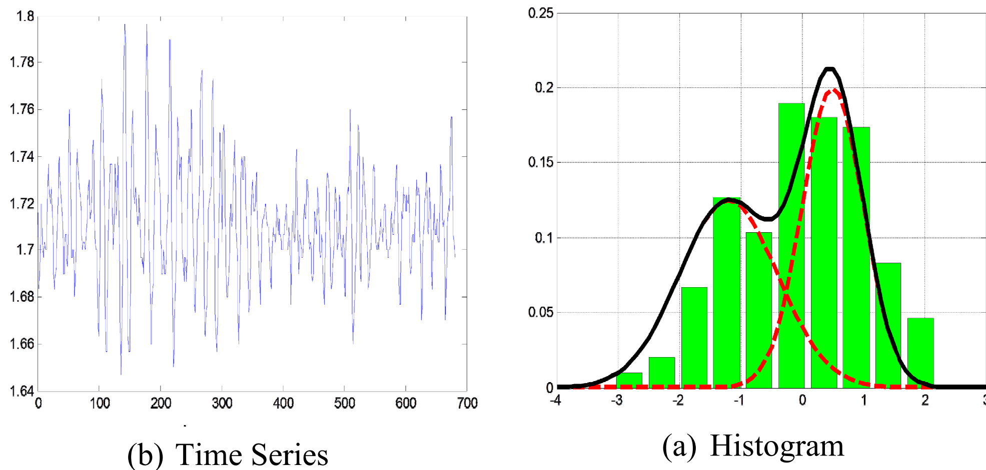
5.2.2 Dirichlet process (DP) based modeling of CMP vibration sensor signals

Any non-Gaussian distribution can be modeled by using mixture of Gaussian distributions as in Eq. 5.1.

$$p(x_i) = \sum_{j=1}^{K} \pi_j N(x_i|\theta_j) \tag{5.1}$$

where x_i is the data point collected by wireless vibration sensor from CMP process as shown in Figure 5.4. A linear combination of K Gaussian distributions with weight π_j and parameters θ_j (mean μ_j and variance $\sigma_j{}^2$) is used to

approximate non-Gaussian distribution. If the number of mixture components K is known *a priori*, the expectation maximization (EM) algorithm [117] can be used to estimate π_j and θ_j but in practice K is usually unknown. Dirichlet Process [118] can be utilized to approximate non-Gaussian distribution without prior knowledge of K. In Figure 5.4; non-Gaussian CMP data is approximated by two mixture Gaussian components by using Dirichlet Process in an adaptive way, where the number of mixture components is obtained using a data driven technique.



(b) Time Series (a) Histogram

Figure 5.4 Time series and histogram of vibration sensor data

Given a set of observation $\boldsymbol{x} = (x_1, x_1, \ldots, x_N)$, we can estimate the posterior distribution of parameters by combining the likelihood of the set of observations given parameters with the prior distribution of the parameters using a Bayesian representation:

$$p(\boldsymbol{\theta}, \boldsymbol{\pi}|\boldsymbol{x}) \propto p(\boldsymbol{\theta}, \boldsymbol{\pi})p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\pi}) \tag{5.1}$$

A Gaussian distribution with hyper parameters m and γ: $p(\mu_j|m,\gamma) \sim N(m, \gamma^{-1})$ is chosen as the prior for the mixture means. The posterior distribution of μ_j is computed by multiplication of prior and the likelihood function, which results in the following Gaussian distribution:

$$p(\mu_j|\boldsymbol{x}, \boldsymbol{c}, s_j, m, \gamma) \sim N\left(\frac{\bar{x}_j n_j s_j + m\gamma}{n_j s_j + \gamma}, \frac{1}{n_j s_j + \gamma}\right) \tag{5.1}$$

where, $\bar{x}_j$ is the sample mean of data points of mixture component *j*, n_j is the number of data points belonging to component *j*, and $\boldsymbol{c}$ is the indicator vector for class labels.

The gamma distribution is selected as prior for component precision (inverse of variance) with hyper parameters β and ω: $p(s_j|\beta, \omega) \sim Ga(\beta, \omega^{-1})$. Subsequently, on multiplying the likelihood function $p(\boldsymbol{x}|s_j)$ with the prior for the component precision $p(s_j|\beta, \omega)$ we obtain the following posterior distribution for component precision which is a gamma distribution:

$$p(s_j|\boldsymbol{x}, \boldsymbol{c}, \mu_j, \beta, \omega) \sim Ga\left(\beta + n_j, \left\{\frac{\omega\beta + \sum_{i:c_i=i}(x_i - \mu_j)^2}{n_j s_j + \gamma}\right\}^{-1}\right) \tag{5.2}$$

A Dirichlet distribution with a parameter of α/k, is chosen as the prior for mixing weights of the components as follows:

$$p(\pi_1, \ldots, \pi_k|\alpha) \sim Dir(\alpha/k, \ldots, \alpha/k) \quad (5.3)$$

where, α is the concentration parameter and k is the number of mixture components. Given this Dirichlet prior for mixing weights, the prior for the number of data points for each component n_j becomes a multinomial distribution; subsequently the joint distribution of indicator vector is expressed as follows:

$$p(c_1, \ldots, c_N|\pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} {\pi_j}^{n_j} \quad (5.4)$$

where N is the number of data points.

Using the property of Dirichlet integral, we can integrate out the mixing weights and get the prior for joint distribution of indicator vector only dependent on concentration parameter α:

$$p(c_1, \ldots, c_N|\alpha) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \prod_{j=1}^{k} \frac{\Gamma(n_j + \alpha/k)}{\Gamma(\alpha/k)} \quad (5.5)$$

where, $\Gamma(.)$ is the Gamma function. As a result, the conditional prior for the single indicator of data point i can be written as:

$$p(c_i = j|c_{-i}, \alpha) = \frac{n_{-i,j} + \alpha/k}{N - 1 + \alpha} \quad (5.6)$$

where, $-i$ indicates the all indices except i. The posterior distribution for the single indicator c_i can be calculated by multiplying the prior for indicator $p(c_i = j|c_{-i}, \alpha)$ with the likelihood function $p(x_i|\mu_j, s_j)$ as given by:

$$p\left(c_i = j|c_{-i}, \mu_j, s_j, \alpha\right) \propto n_j p(x_i|\mu_j, s_j) \tag{5.7}$$

Consequently, the finite Gaussian mixture model can be represented hierarchically using the following form:

$$\begin{gathered} x_i \sim N\left(\theta_j\right) \\ \theta_j \sim G_0 \\ \pi_1, \ldots, \pi_k|\alpha \sim Dir(\alpha/k, \ldots, \alpha/k) \\ c|\pi_1, \ldots, \pi_k \sim Discrete(\pi_1, \ldots, \pi_k) \end{gathered} \tag{5.8}$$

where, $\boldsymbol{x_i}$ is sampled from Gaussian distribution with parameters $\boldsymbol{\theta_j}$ which is sampled from prior distribution $\boldsymbol{G_0}$. Mixture weights $\boldsymbol{\pi_j}$ is sampled from Dirichlet distribution, whereas indicators $\boldsymbol{c}$ are proportional to mixture weights.

Infinite Gaussian mixture model with Dirichlet process

So far, our discussion has focused on for finite number of components under the assumption that the number of mixture components is known *a priori*. When we assume that the limit for number of clusters *k* goes to infinity, the resulting is *Dirichlet process* [119].

The Chinese restaurant process is metaphorically used to explain this idea [120-121]. Assuming that a Chinese restaurant has infinite numbers of tables and each table can be seated by infinite number of customers. The process begins when the first customer sits on any available table. Thereafter, the second customer either sits on the table where the first customer is sitting or on an unoccupied table. Continuing with this reasoning, the ith customer can sit at an already occupied table with a probability proportional to the number of customers at that particular table. Alternatively, the same customer can sit at an unoccupied table with probability proportional to concentration parameter α. Accordingly, the prior distribution for the component indicators is formulated as follows:

$$p(c_i = k|c_{-i}, \alpha) \propto \begin{cases} \frac{n_{-i,k}}{N-1+\alpha} & if\ k\ is\ active \\ \frac{\alpha}{N-1+\alpha} & if\ k\ is\ not\ active \end{cases} \tag{5.9}$$

This result signifies a Dirichlet process mixture (DPM) which can be used to model a set of observations $(x_1, x_2, \dots,\ x_n)$, with latent variables of $(\theta_1, \theta_2, \dots,\ \theta_n)$ as follows:

$$\begin{gathered} G \sim DP(\alpha, G_0) \\ \theta_i \sim G \\ x_i \sim F(.\,|\theta_i) \end{gathered} \tag{5.10}$$

In the above equation, consider that data points x_i (e.g., features from sensor data) are drawn from a distribution *F* with parameters θ_i. Additionally, each θ_i is drawn from unknown distribution G. We note, because the distribution G is

discrete, the same values can be assigned to multiple θ_i. Data points x_i which have the same θ_i value belongs to the same component [118, 122-123]. Furthermore, on integrating out G, the following conditional distribution for θ_n is obtained [124]:

$$\theta_n|\theta_{n-1},\dots,\theta_1,G_0,\alpha \sim \frac{\alpha}{\alpha+n-1}G_0 + \sum_{k=1}^{n-1}\frac{1}{\alpha+n-1}\delta_{\theta_k} \tag{5.11}$$

where, δ_{θ_k} is the Dirac function peaked on θ_k. Subsequently, combining the likelihood function (Eq. (5.1)) with the prior distribution for single indicator of data *i*, results in the following posterior distribution for an indicator:

$$p(c_i = k|c_{-i},\boldsymbol{x}) \sim \begin{cases} \dfrac{n_{-i,k}}{N-1+\alpha} F(x_i|\theta_k), & if\ k\ is\ active \\ \dfrac{\alpha}{N-1+\alpha} \displaystyle\int F(x_i|\theta)dG_0(\theta), & if\ k\ is\ not\ active \end{cases} \tag{5.12}$$

Table 5.1 Pseudo code for DP clustering

```
➢ Construct DP mixture with no data items to get priors
➢ Initialize DP mixture with data
➢ Gibbs Sampling
     For #iteration
          For each data element
               • remove data item i from cluster j and its sufficient
                 statistics
               • check if the cluster is empty, if so delete it
               • compute conditional probabilities of data item i for each
                 clusters
               • choose cluster for data item i
               • create new cluster if needed
               • add data item and its statistics to cluster
          end
     end
end
```

5.2.3 Recurrent Dirichlet process (RDP) model

The DPM assumes data points $(x_1, x_2, \dots, x_n)$ are fully exchangeable, although in many cases, such as process monitoring applications, data points are acquired sequentially. During the physical process, new mixture components may emerge (Figure 5.5 (c)), die (Figure 5.5(d)) or parameters of mixture components may change overtime (Figure 5.5 (b)). Therefore, to monitor the process, we need to track the parameters of mixture components for making physically tenable and accurate decisions. We posit that accurate mixture model updates can be obtained by utilizing the information from prior cluster estimates.

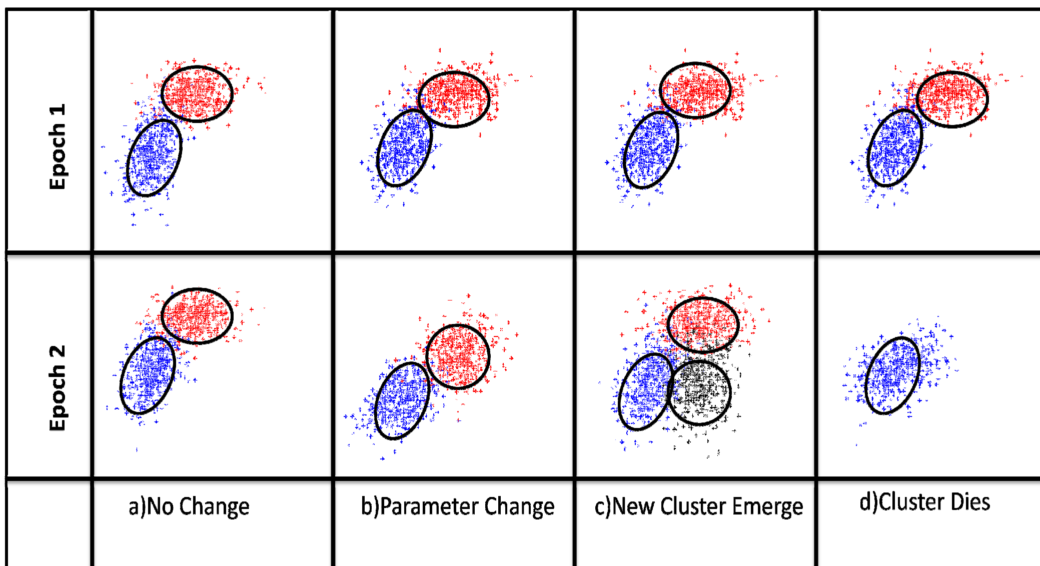


Figure 5.5 Possible cluster evolutions

Initially, the incoming sensor data is divided into sliding windows. For instance, if epoch 1 consists of data points $(x_1, x_2, \dots, x_{n-1})$, and epoch 2 consists of data points $(x_2, x_3, \dots, x_n)$ with the latent variables $(\theta_1, \theta_2, \dots, \theta_{n-1})$ and $(\theta_2, \theta_3, \dots,$

θ_n), respectively; the conditional distribution of the latent variable θ_n at time epoch t can be formulated by using previous time epoch's information as follows [72]:

$$\theta_{t,i}|\theta_{t-1,i},\theta_{t,1:i-1},G_0,\alpha\sim\frac{1}{N_{t-1}+i+\alpha-1}\left[\sum_{k\in I_{t-1}\cup I_t}(n_{k,t-1}+n_{k,t})\,\delta(\theta_{k,t})+\alpha G_0\right] \quad (5.13)$$

where, N_t denotes the number of data points in the tth epoch, $n_{k,t}$ denotes the number of data points associated with mixture component k at time t, and I_t denotes the mixture of Gaussian distributions at time t.

Once again, the Chinese restaurant process metaphor can be used to illustrate this situation; the conditions are articulated as follows:

- Suppose that the owner of the restaurant keeps an accurate log of the dishes served on each table, also the number of people who share that particular dish
- Additionally, the owner assumes that a particularly popular dish remains popular the next day, so that he can buy the required ingredients.
- Lastly, the information about the previous day's consumption is displayed on all tables, so that customers can make a choice based on popularity of various dishes

Accordingly with these conditions, on a given day t, a customer can choose a table (labeled as k) proportional to the number of customers who occupied table k

on the previous day, plus the number of customers that are sitting at this table (table k) at the particular instant this customer enters the restaurant. Using Eq. 5.13 and letting $K \to \infty$, the cluster assignment is mathematically represented as follows:

$$p(c_i = k|c_{-i}) = \begin{cases} \frac{n_{k,t-1} + n_{-i,k,t}}{N_{t-1} + i - 1 + \alpha}, & if\ k\ is\ active \\ \frac{\alpha}{N_{t-1} + i - 1 + \alpha}, & if\ k\ is\ inactive \end{cases} \tag{5.14}$$

As a physical analogy to the Chinese restaurant metaphor, wireless MEMS vibration sensor signals obtained from the CMP process can be divided into epochs. Subsequently, the cluster parameters of particular epoch can be estimated invoking the RDP model. However, in the case of CMP data we note that the parameters for the mixture components are sampled in a Markovian fashion using the information from previous time epochs:

$$\theta_t \sim P(.|\theta_{t-1}) \tag{5.15}$$

Inference for component assignments by using Gibbs sampling algorithm will be as follows:

$$p(c_{i,t} = k|c_{-i,t}) \sim \begin{cases} \frac{n_{k,t-1} + n_{-i,k,t}}{N_{t-1} + N_{t-1} + \alpha} F(x_i|\theta_k), & if\ k\ is\ active \\ \frac{\alpha}{N_{t-1} + N_{t-1} + \alpha} \int F(x_i|\theta) dG_0(\theta), & if\ k\ is\ inactive \end{cases} \tag{5.16}$$

5.2.4 Recurrent nested Dirichlet process (RNDP) model

Although we have estimated the parameters for the clusters at each time epoch in a sequential manner, we have not yet considered the clusters *between* time epochs in order to detect incipient changes in the process. In this context the nested Dirichlet process is useful when there are multiple groups of data, and additionally within each group there are also multiple clusters [73]. Therefore, in order to detect changes between two consecutive time epochs we propose a novel approach which integrates recurrent Dirichlet process (RDP, Section 5.2.3) with nested Dirichlet process (NDP, Section 5.2.4) and consequently term the resulting approach as **R**ecurrent **N**ested **D**irichlet **P**rocess (RNDP), which is formulated as follows:

$$\begin{aligned} G_t|\alpha,\gamma,G_0 &\sim DP(\alpha,\gamma,G_0) \\ \theta_{ti}|G_t &\sim G_t \\ x_{ti}|\theta_{ti} &\sim F_{\theta_{ti}} \end{aligned} \tag{5.17}$$

where, x_{ti} is the ith data point in sliding window t, G_0 is the prior distribution over groups, $F_{\theta_{ji}}$ is the distribution over the data points, G_j is the distribution over the parameters within clusters, G_0 is the distribution over the clusters within groups.

On combining the NDP and RDP cluster labels for data points and time epochs as shown in Eqs (5.16) and (5.17), respectively, we obtained the following

probability distribution functions for both the higher level (Eq. (5.18)), as well as, lower level (Eq. (5.19)) indicators:

$$p\left(z_{j,t} = k \middle| .\right) \sim \begin{cases} (m_{k,t-1} + m_{-i,k,t})\gamma \, F\left(x_{ji} \middle| \theta\right), & ,if \ k \ is \ active \\ \gamma \int F\left(x_{ji} \middle| \theta\right) dG_0(\theta), & ,if \ k \ is \ inactive \end{cases} \tag{5.18}$$

$$p\left(c_{ji} = l \middle| .\right) \sim \begin{cases} (n_{j,l,t-1} + n_{-i,j,l,t})\alpha \, F\left(x_{ji} \middle| \theta_l\right) & ,if \ k \ is \ active \\ \alpha p\left(x_{ji} \middle| c_{ji} = l^{new}\right) & ,if \ k \ is \ inactive \end{cases} \tag{5.19}$$

where z_j is the cluster label for time epoch t, and c_{ji} is the cluster label for sensor data point i at time epoch j.

By tracking the evolution of mixture Gaussians parameters at consecutive epochs, the incipient changes in the CMP process can be monitored. Consequently, knowing the distribution of mixture Gaussian parameters *a priori*, control charts can be set up so that CMP process drifts can be detected at an early stage in an SPC setting. In the forthcoming section, we delineate how the RNDP method can be tractably extended to a SPC setting.

5.2.5 Formulation of RNDP-based control charts for process monitoring applications

The formulation of RNDP-based control charts entails two key aspects:

(a) Setting up the control limits for the RNDP-based control chart

(b) Comparison of RNDP control charts with conventional SPC charts, such as EWMA and CUSUM with respect to Type II error in terms of the average run length (ARL_1) statistic.

In this section, the first of the above two aspects will be elucidated. Concomitantly, for purposes of juxtaposition we will also state the well-known control limits for popular SPC charts, such as EWMA and CUSUM.

Control limits for EWMA: Exponentially weighted moving average (EWMA) charts are frequently used to monitor the small shifts in the first moment (mean) of a process [125]. The EWMA chart assumes each observation/data point is normally distributed, i.e., $x_i \sim N(\mu, \sigma^2)$ with the following test statistics z_i:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1} \tag{5.20}$$

where, $z_0 = 0$, and the parameter λ is a heuristically determined. We note a higher λ value implies a greater degree of importance has been assigned to current data point, as opposed to previous data points.

CUSUM: Cumulative Sum (CUSUM) control chart is a sequential technique for monitoring drifts in both the first and second moments (mean, standard deviation, respectively) [126]. The CUSUM assumes each observation/data point is normally distributed, i.e., $x_i \sim N(\mu, \sigma^2)$ with the following test statistics for mean and standard deviation (z_i and v_i):

$$z_i = \frac{|x_i - \mu|}{\sigma}$$

$$v_i = \frac{(\sqrt{z_i} - 0.822)}{0.349} \tag{5.21}$$

For positive (S^+) and negative (S^-) change in mean two separate CUSUM test statistics can be constructed as shown in Eq. 5.22.

$$S_i^+ = \max[0, z_i - K + S_{i-1}^+]$$

$$S_i^- = \max[0, -z_i - K + S_{i-1}^+] \tag{5.22}$$

where, initial values for S_i^+ and S_i^- are 0. In order to monitor the change in standard deviation v_i is used instead of statistics z_i.

RNDP: First, the time series data are segregated into sliding windows and subsequently, the RNDP methodology is applied for each window. Detail steps are as follows:

(a) Time series data is mapped into phase space which provides information about the dynamics of the system [127]. In order to reconstruct the time series in phase space, the appropriate embedding dimensions (delay and neighborhood dimension) are estimated. For this purpose, the auto mutual information (AMI) function and false nearest neighborhood (FNN) algorithm is used, from which the optimal delay (τ) and embedding dimension (d_E) are respectively obtained [127].

(b) Subsequently, the RNDP process applied to the time series reconstructed in phase space. Pearson's Chi-square goodness of fit (GoF) test is used for every iteration of Gibbs sampling to determine the optimal stopping point. Figure 5.7 illustrates how the estimated distribution converges to the true distribution as the number of iterations increases. Pearson's chi square GoF test is applied at each iteration with type I error set at 5% (See Figure 5.8).

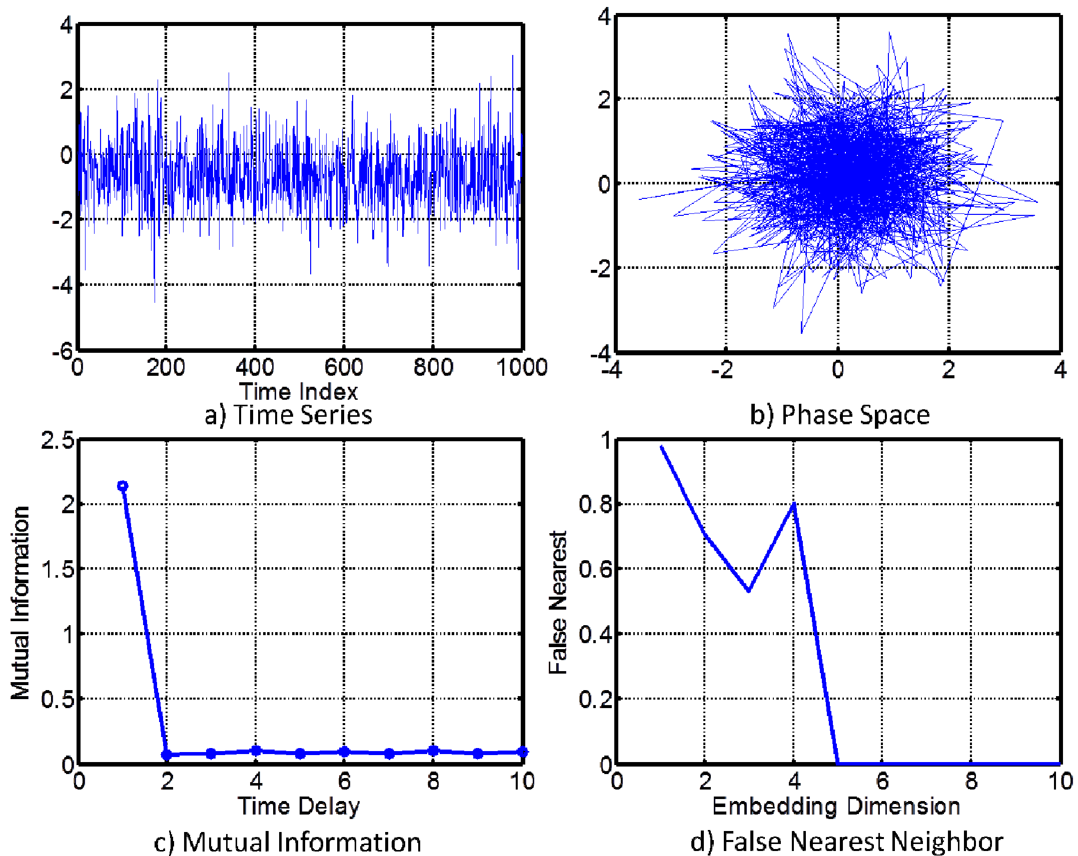


Figure 5.6 Phase space plot (b) constructed from time series (a) by using mutual Information (c) and false nearest neighbor (d)

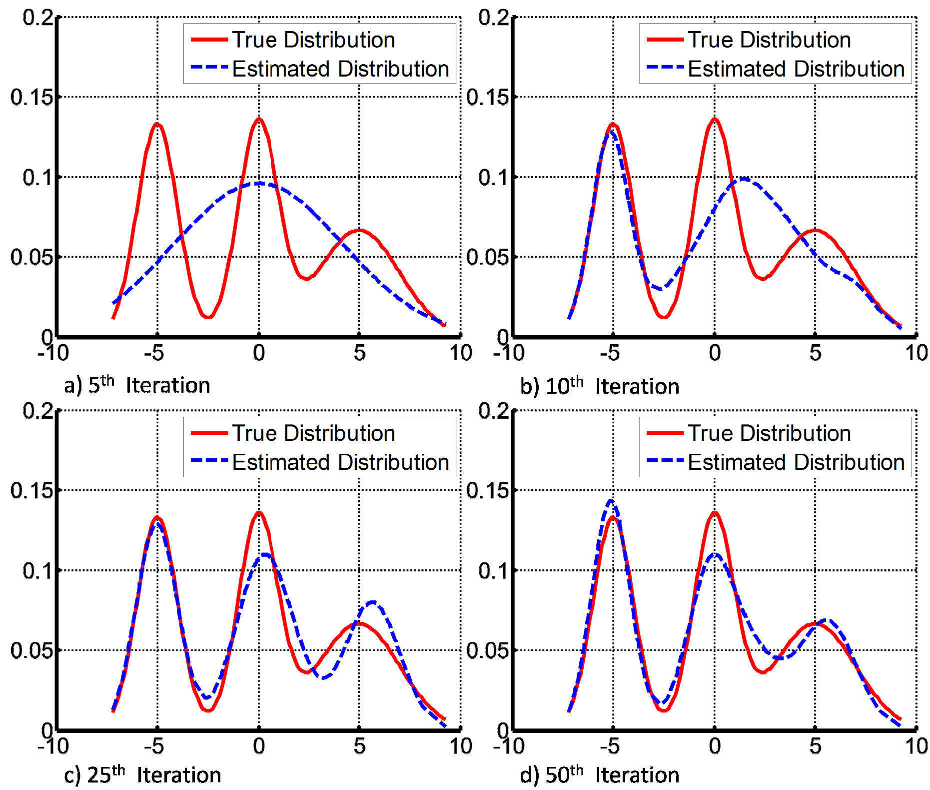


Figure 5.7 Evaluation of estimated distribution vs. true distribution at different number of iterations

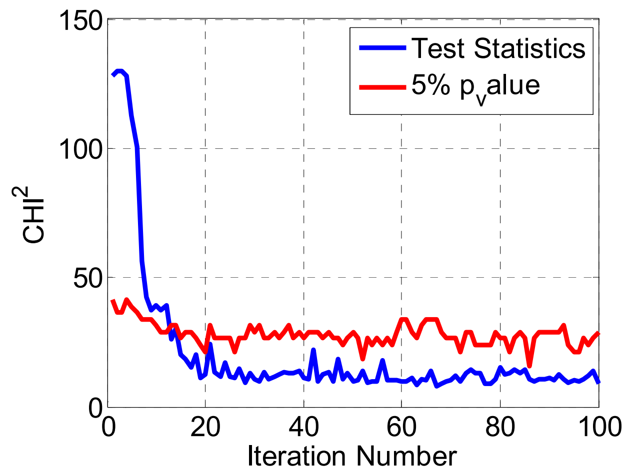


Figure 5.8 Chi square statistics with 5% p-value threshold

Gibbs sampling iterations are stopped once the GoF test statistics drops below a set p-value (5%).

Finally, the ARL_1 performances (smaller is better) of EWMA, CUSUM, and RNDP-based control charts are compared. To ensure equitable comparisons, all these charts are set up with control limits, such that, type I probability in terms of the ARL_0 is maintained at approximately 370 data points. Consequently, the reference value K is set at 0.5 for the CUSUM control chart, the λ value for the EWMA chart set at 0.2, and the scale parameter α for the RNDP-based control chart set at 1.

5.3 Validation of the Proposed Method

Multiple case studies are reported in order to illustrate the effectiveness of the proposed RNDP approach over conventional SPC methods. For this purpose, both numerical (computer generated time series) and experimental (wirelessly acquired MEMS vibration sensor signals from CMP process) data sets are tested.

For simulation based studies, the following two kinds of scenarios are investigated:

(a) An auto regressive moving average (ARMA) model is used to generate signals with six kinds of fault types (Table 5.2, Figure 5.9).

(b) A sinusoidal signal with a intermittent extra frequency, essentially indicating the change in the autocorrelation function.

In the simulation studies we will specifically concentrate on being able to detect changes in the signal. Pertinently, we desire a low detection delay given that a change in the signal characteristics has occurred. A smaller detection delay implies that the method is sensitive to subtle process drifts.

Finally, vibration sensor signal patterns from CMP tests conducted under conditions where certain process parameters have been deliberately changed are tested to illustrate the effectiveness of the RNDP approach. In these tests it was consistently observed that the RNDP method not only captures the change at an earlier stage, but also is capable of identifying the *type/nature* of changes. This is valuable from an engineering perspective, because by knowing both when a change has occurred and also type of the change, an operator can make expeditious adjustment to the process, which can be helpful for preventing scrap and rework.

5.3.1 Change point detection with simulated time series data

A time series indicative of normal process conditions is generated from an ARMA(m,n) process as follows [128]:

$$A(q)x_t = D(q)a_t \tag{5.23}$$

where, $A(q)$ and $D(q)$ are defined as:

$$A(q) = 1 - \sum_{i=1}^{m} \varphi_i q^{-i}$$

$$D(q) = 1 - \sum_{j=1}^{n} \theta_j q^{-i} \quad (5.24)$$

$$q^{-1} x_t = x_{t-1}$$

where, x_t is the process state variable at time t, and $a_t \sim N(0, \sigma_a^2)$ is the process noise at time t. State space model is constructed with measurement noise $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ at time t in the following manner:

$$x_t = \sum_{i=1}^{m} \varphi_i x_{t-i} - \sum_{j=1}^{n} \theta_j a_{t-j} + a_t \quad (5.25)$$

$$y_t = x_t + \varepsilon_t$$

where, y_t is the measurement data.

Specifically, the following ARMA (2, 1) models are used in this study:

Model 1: $$x_t = 0.99 x_{t-1} - 0.49 x_{t-2} + 0.7 a_{t-1} + a_t \quad (5.26)$$

Model 2: $$x_t = 0.1 x_{t-1} - 0.8 x_{t-2} + 0.7 a_{t-1} + a_t \quad (5.27)$$

The state space model of Eq. 5.25 is considered to represent the normal process condition (in control process). Various faults (process drifts) as summarized in Table 5.2, and illustrated in Figure 5.9 are induced. As a physical analogy, these faults may be considered as anomalies that cause the process to drift out of control.

Table 5.2 Summary of fault types

Fault Types	Cause of Faults
Fault 1	$a_t \sim N(0, \delta^2\sigma^2_a)$ (process variance change)
Fault 2	$e_t \sim N(0, \delta^2\sigma^2_e)$ (measurement errors variance change)
Fault 3	$a_t \sim N(\delta, \sigma^2_a)$ (process mean shift)
Fault 4	$a_t \sim N(\delta, \delta^2 \sigma^2_a)$ (mean and variance fault)
Fault 5	$a_t \sim (N(0, \delta^2\sigma^2_a)\ U\ N(\delta, \sigma^2_a))$ (non-Gaussian fault)
Fault 6	1. $a_t \sim N(0, \delta^2\sigma^2_a)$ 2. $a_t \sim N(\delta, \sigma^2_a)$ (different faults)

Comparison for different fault types

The normal condition assumes that the measurement noise is in accordance with the following null hypothesis:

$$H_0: \varepsilon_t \sim N(0, 0.5^2) \quad a_t \sim N(0, 1^2)$$

In the above ε_t is the measurement noise, while a_t is the process noise. In a physical sense, ε_t is evocative of errors due to sensing, whereas, a_t is representative of environmental noise (e.g., vibration from nearby machinery, temperature fluctuations, etc.). Note also, ε_t is significantly less volatile in comparison to a_t. The alternative hypothesis consists of a variety of changes in ε_t and/or a_t. Also, for the first three fault conditions, the result from a contemporary wavelet based cumulative sum (WCUSUM) control chart reported in Ref. [129] is used for comparison.

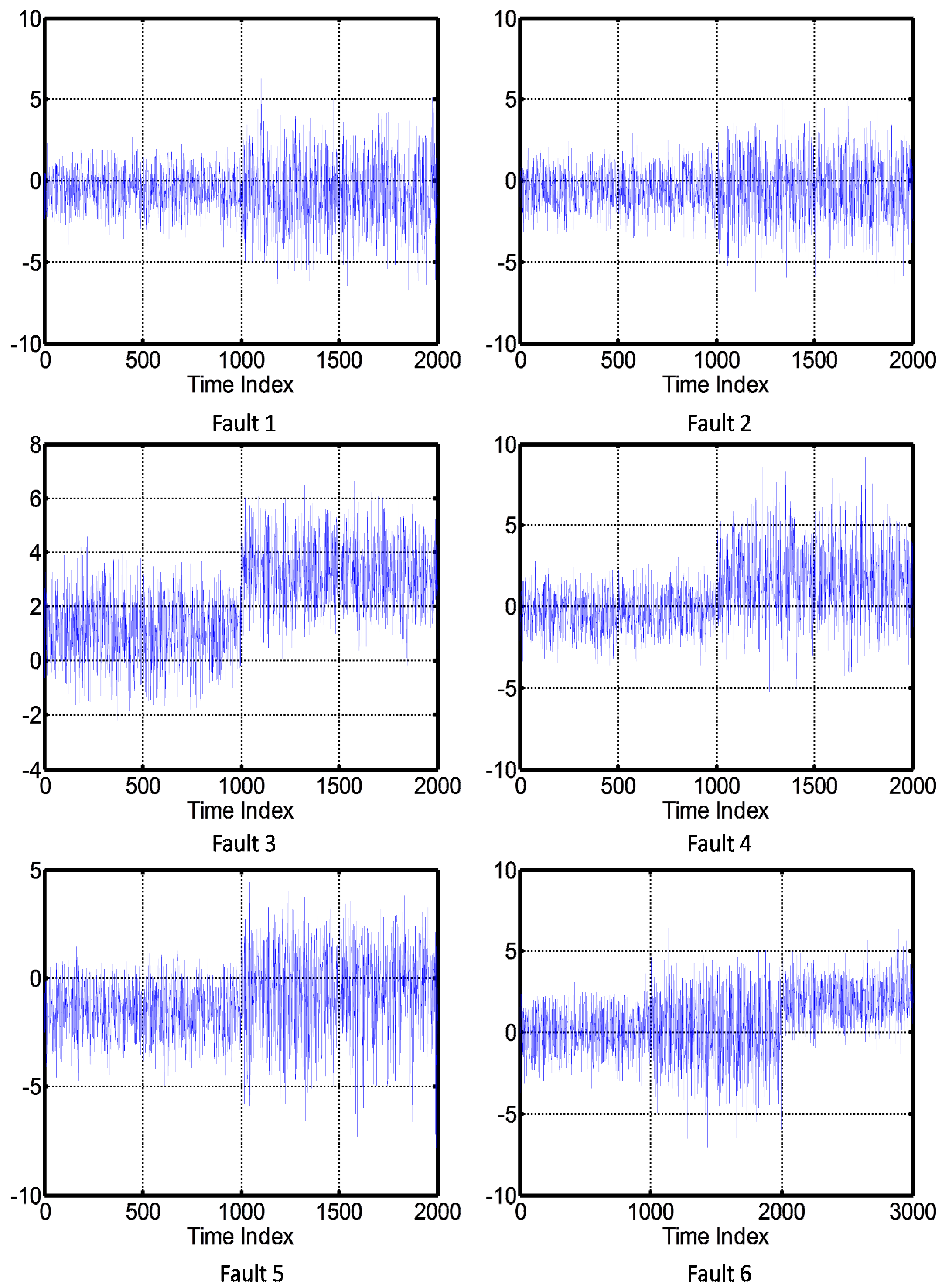


Figure 5.9 Illustration of changes in the process

In these six case studies the following standard procedure is maintained:

i. Initially, for both models 1 and 2 (Eq. 5.26 and Eq. 5.27, respectively) a 2000 data point long time series is generated. This time series is considered to be representative of normal process conditions (in control).

ii. From the normal condition data sets the upper and lower control limits for CUSUM, WCUSUM, and RNDP-based control charts are estimated using Monte Carlo simulation. These control limits are adjusted such that the Type I probability is identical for all control charts. This is achieved by maintaining the ARL_0 value at approximately 370 data points. This step is important for a fair comparison of different approaches.

   In other words, by maintaining the ARL_0 is identical we ensure that no one approach is overly susceptible to Type I error.

iii. Once the ARL_0 limits have been determined, we simulate a 2000 data point long time series with different types of fault infused into both model 1 and 2. Based on the predetermined control limits (from step ii) we can estimate the ARL_1 values for each of the control charts. The approach with smallest ARL_1 is considered more sensitive for detecting changes. Each case is replicated for 30 instances and the averaged ARL_1 values are reported.

We summarize the different simulated (see Table 5.2 and Figure 5.9) fault types herewith:

- **Fault 1:** Illustrates a situation where the process noise variance changes. Model 1 and 2 are investigated under the alternative hypothesis of H1: $a_t \sim N(0, \delta^2 1^2)$. The results are shown in Figure 5.11, from which we notice that the RNDP approach outperforms the CUSUM and WCUSUM. For instance, for a process noise variation of 1.5σ the ARL_1 (noting that the result is the average of 30 replications) is approximately 25-30 data points and is almost ~ 30% less than WCUSUM.
- **Fault 2:** Demonstrates the change of variance in measurement noise with an alternative hypothesis H_1: $\varepsilon_t \sim N(0\ \delta^2 0.5^2)$. The results are illustrated in Figure 5.12, from where it is evident that RNDP approach is comparable to WCUSUM for model 1, while significantly outperforming (ARL_1 is approximately 50% less for all levels of ε_t) WCUSUM for model 2.
- **Fault 3:** In this case, the drifts in process mean are investigated with alternative hypothesis set as H_1: $a_t \sim N(\delta, 1^2)$. The results are presented in Figure 5.13 illustrates that WCUSUM is not sensitive to drifts in the process mean, whereas RNDP successfully detects the changes within 75 data points for $\delta = 1$ (the WCUSUM, CUSUM ARL_1 values were approximately 370, 125 data points, respectively).
- **Fault 4:** We test the scenario where there is a change in both mean and variance of the process noise as represented with the alternative hypothesis

H1: $a_t \sim N(\delta, \delta^2 0.5^2)$. For both models the RNDP ARL1 values were almost 50%-75% smaller compared to the CUSUM chart.

- **Fault 5:** In complex physical processes like CMP, the process noise may have a non-Gaussian characteristic. Under such situations, traditional methods which explicitly assume process noise follows a Gaussian distribution may fail to capture subtle changes in the process or even prone to Type I error. We demonstrate the potency of the RNDP method for cases that have non-Gaussian noise by sampling from a multimodal distribution shown in Figure 5.10. Formally, the alternative hypothesis for this case is stated as H1: $(a_t \sim N(\delta, 1^2)\ U\ a_t \sim N(0, \delta^2 1^2))$.

  The results from this scenario are shown in Figure 5.15; once again it is observed that the RNDP approach detects process deviations approximately 30%-40% earlier than CUSUM.

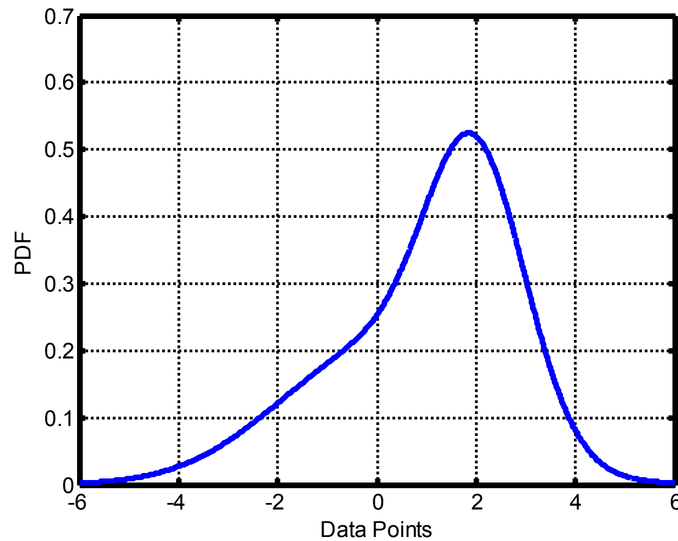


Figure 5.10 Non-Gaussian process noise added for fault 5

- **Fault 6:** It is often observed that, several types of changes may concurrently occur in a real life scenario. For example, during CMP process, due to changes in the down-force (polishing load) the pad life also changes. Conventional SPC charts often fail to consider the effect of multiple faults simultaneously. However, owing to its self-evolving ability, RNDP can detect different kinds of changes simultaneously. We demonstrate such a case in which multiple faults occur consecutively. Accordingly, for this case the alternative hypotheses are, H_1: $a_t \sim N(0, \delta^2 1^2)$ and H_2: $a_t \sim N(\delta, 1^2)$. Classification errors for detecting various kinds of changes are shown in Table 5.3. The normal condition is denoted by C1 whereas, two fault conditions are symbolically represented as C2 and C3, these correspond to H_1 and H_2, respectively.

  For almost all noise levels tested, the RNDP approach classification accuracy is > 75%.

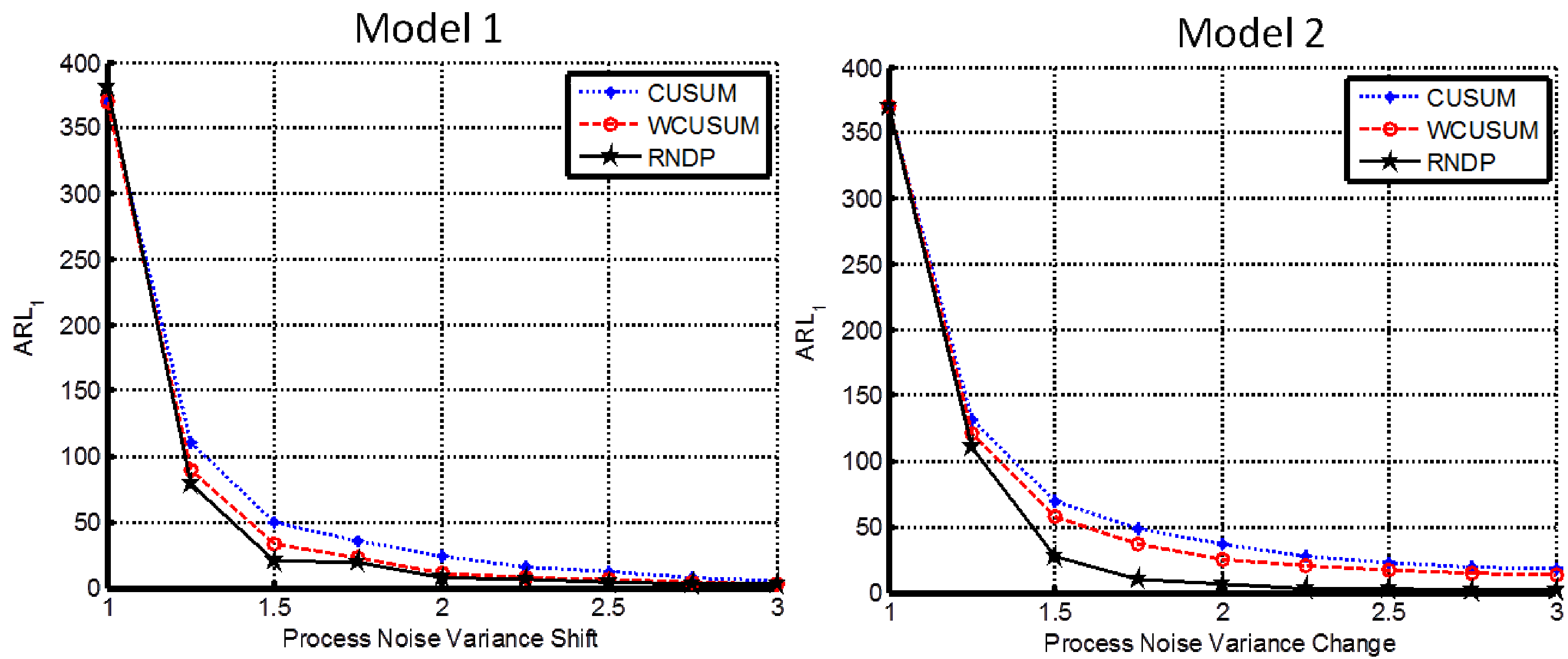
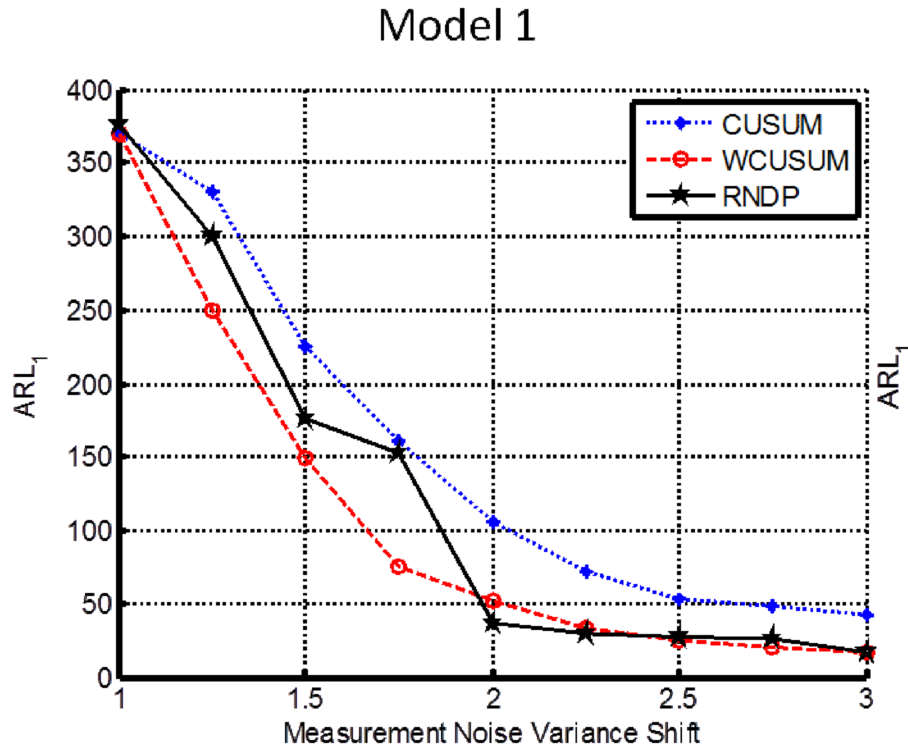


Figure 5.11 ARL_1 results for fault type 1



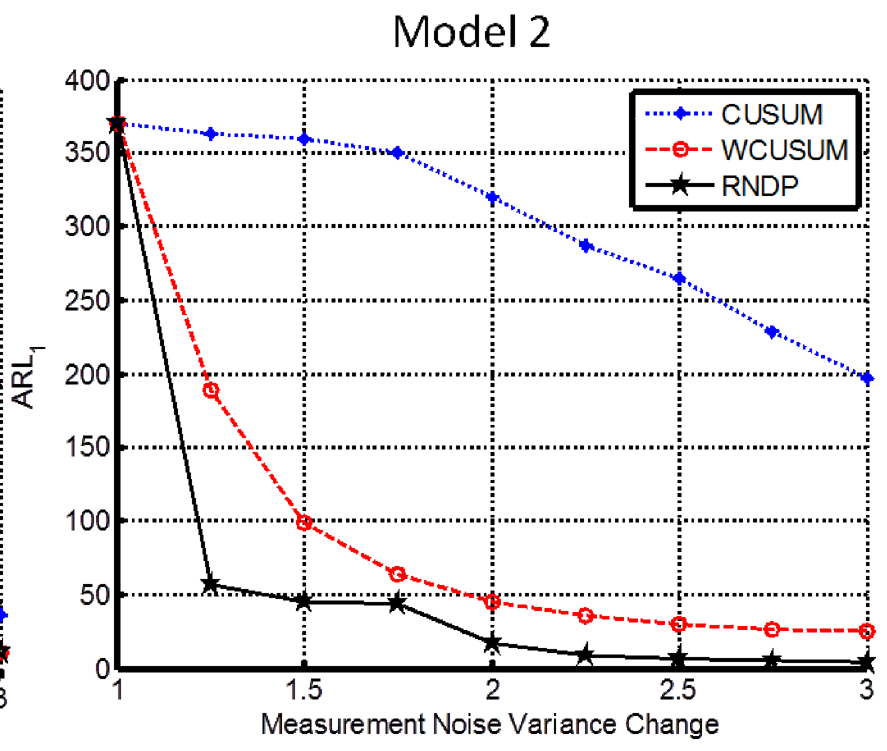


Figure 5.12 ARL_1 results for fault type 2

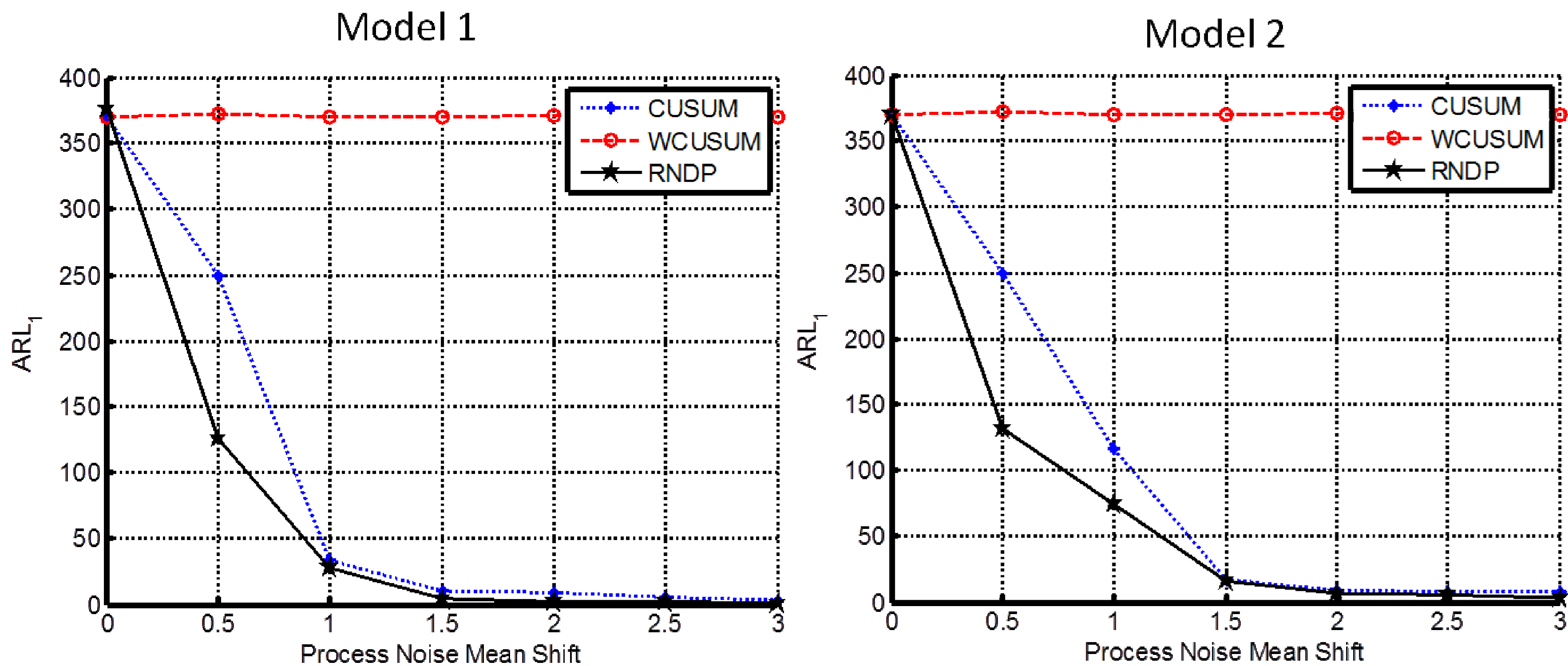
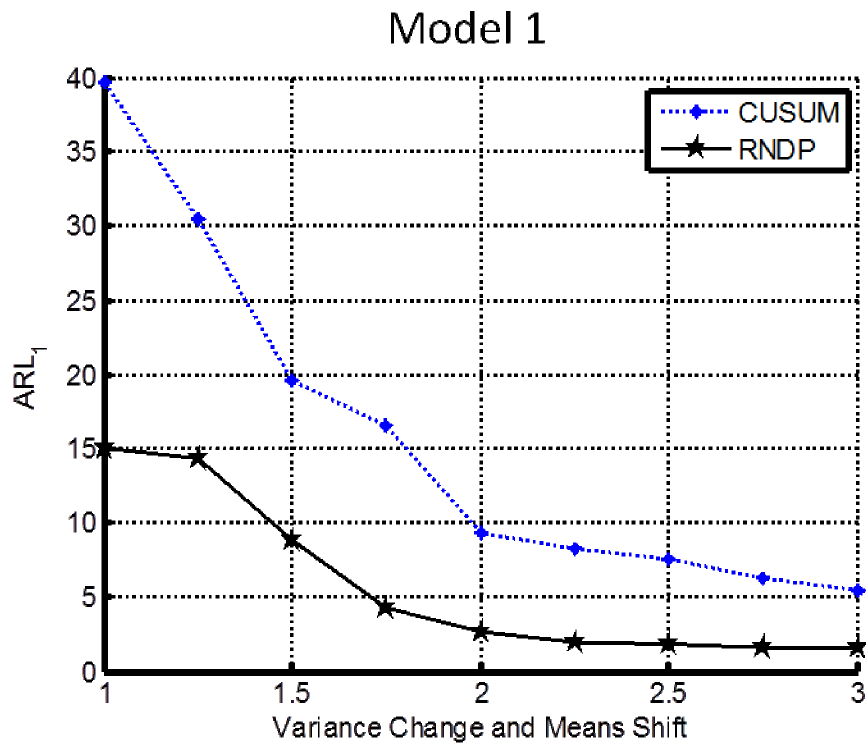


Figure 5.13 ARL_1 results for fault type 3



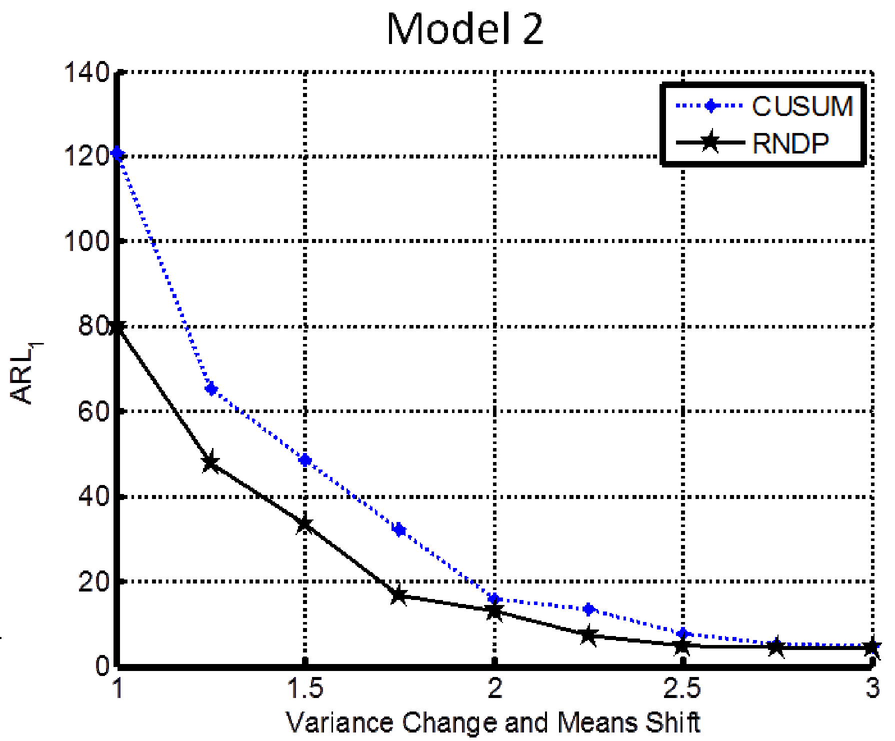
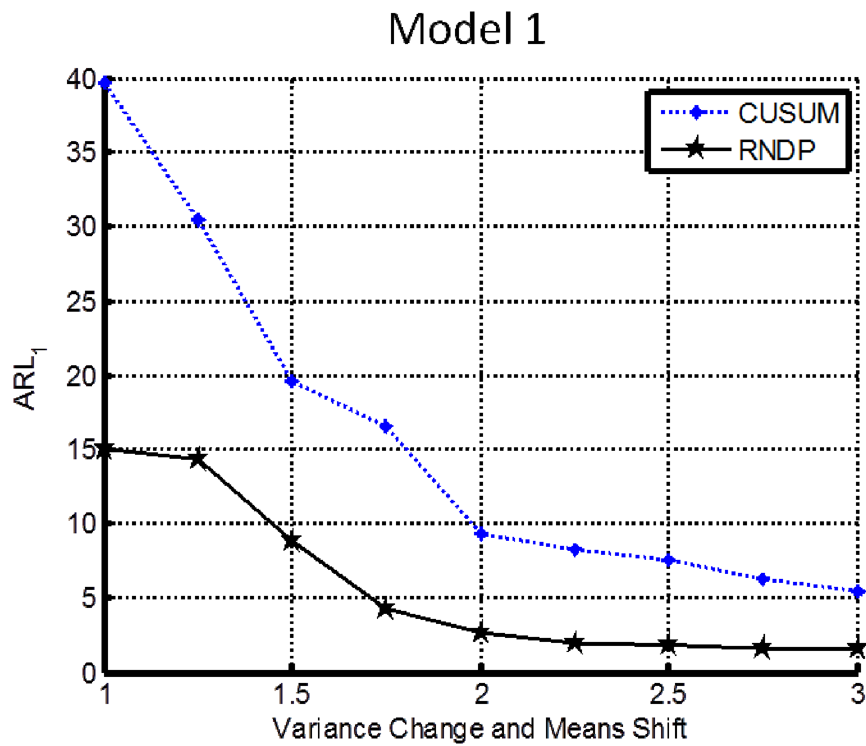


Figure 5.14 ARL_1 results for fault type 4



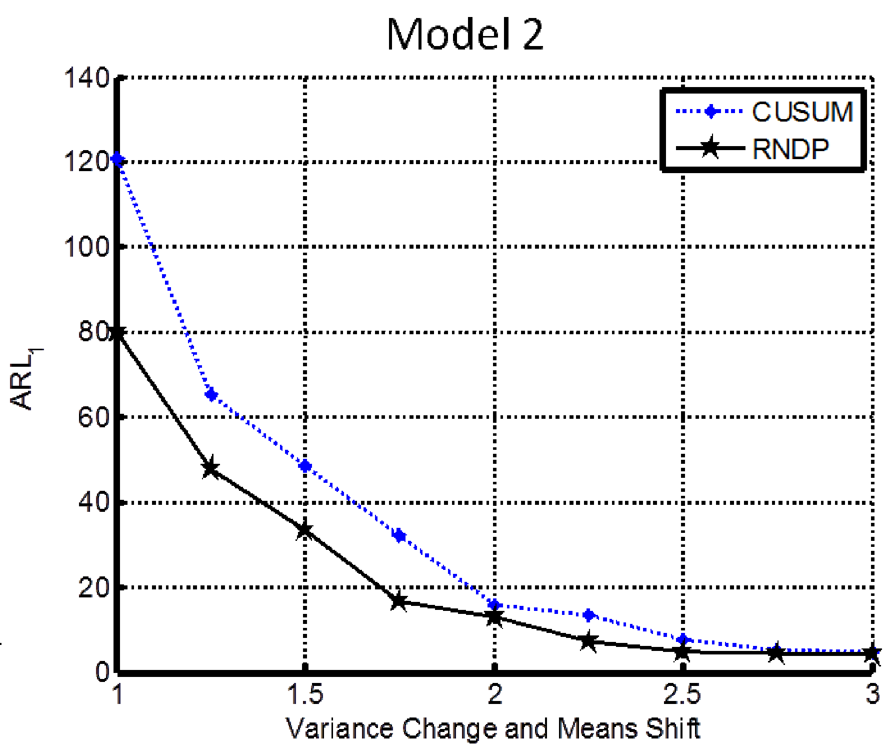


Figure 5.15 ARL_1 results for fault type 5

.

Table 5.3 Classification accuracy matrix for fault type 6

	Classification Accuracy Matrix			
Mean and Variance Shift (Multiplication of σ)		C1	C2	C3
1	C1	98.78%	-	18.54%
	C2	-	-	-
	C3	1.22%	-	81.46%
1.25	C1	79.96%	38.60%	6.48%
	C2	20.04%	60.28%	10.47%
	C3	0%	1.12%	83.05%
1.5	C1	93.53%	30.00%	2.50%
	C2	6.47%	69.55%	12.55%
	C3	0%	0.45%	84.95%
1.75	C1	99.42%	24.45%	2.15%
	C2	0.58%	75.08%	12.10%
	C3	0%	0.47%	85.75%
2	C1	99.89%	21.35%	1.62%
	C2	0.11%	77.72%	12.30%
	C3	0%	0.93%	86.08%
2.25	C1	99.91%	21.90%	0.77%
	C2	0.09%	77.17%	12.73%
	C3	0%	0.93%	86.50%
2.5	C1	100.00%	21.80%	1.57%
	C2	0%	77.70%	11.72%
	C3	0%	0.50%	86.72%
2.75	C1	100.00%	20.40%	0.73%
	C2	0%	78.57%	12.95%
	C3	0%	1.03%	86.32%
3	C1	100.00%	19.13%	0.75%
	C2	0%	80.12%	11.08%
	C3	0%	0.75%	88.17%

Effect of additional frequency

We now validate the RNDP approach using a sinusoidal signal with multiple frequencies. The normal condition for the signal is composed of 4 frequencies; one more frequency has been added to create a change in the signal. The amplitudes of the signal for both change and normal case have been attenuated to make signal energies the same for both cases, which creates a very subtle change as illustrated in Figure 5.16. The frequency comparison for normal and change case is depicted in Figure 5.17.

The detection delay results are illustrated in Figure 5.18 for the generated sinusoidal signal. Figure 5.18 shows the box-plot of detection delay results for various methods. The box-plot results are obtained by realizing 30 replications for each method. As shown in Figure 5.18 traditional SPC methods such as CUSUM and EWMA are not sensitive enough to detect such subtle changes in signal. Detection change for RNDP is almost twenty times faster than CUSUM and EWMA.

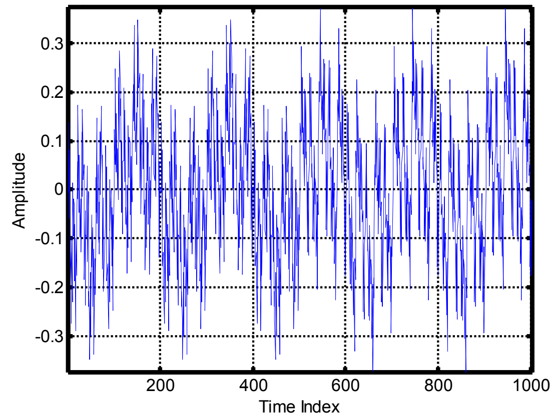


Figure 5.16 Generated sinusoidal signal (first 500 is normal case with 4 frequencies second 500 is change case with 5 frequencies)

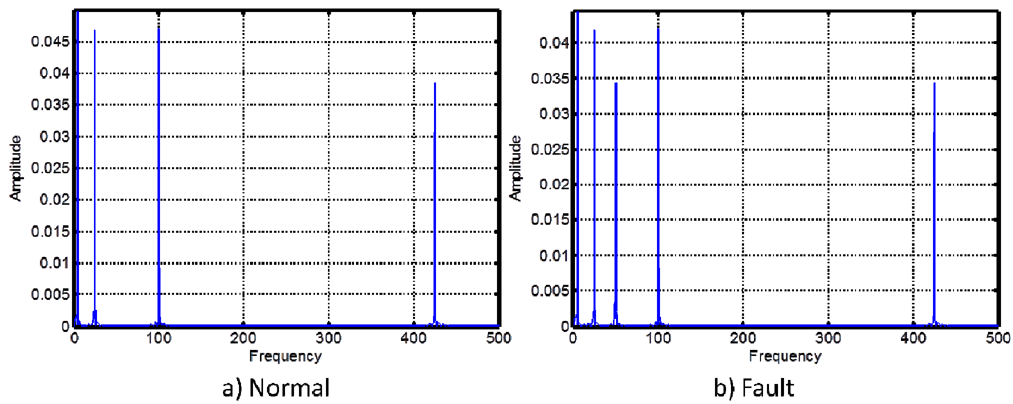


Figure 5.17 Frequency comparison of normal and fault case for sinusoidal signal

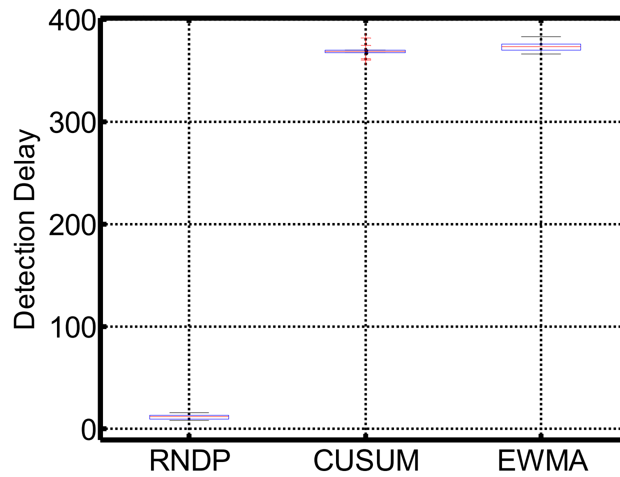


Figure 5.18 Detection delay comparison for sinusoidal signal

5.3.2 Application of RDNP method for detection of CMP process anomalies

As described earlier in Section 5.2.1, MEMS wireless vibration sensor data was acquired during CMP processing of blanket copper wafers. The sensor data is gathered at a sampling rate of approximately 700 Hz. As seen in Figure 5.19, the raw signal time series depicts complex non-linear behavior with several interlaced frequency components. For instance, the signal in Figure 5.19 shows prominent low frequency behavior, while on closer examination short-time high frequency components can also be discerned.

In the frequency domain (Figure 5.20), at least four prominent regions are evident. Our experimental investigations [130-131] have shown that the prominent 110 – 120 Hz frequency region of the fast Fourier transform (FFT) shown in Figure 5.20 is sensitive to changes in process state. However, these changes in the frequency domain are subtle and cannot be tractably quantified using traditional statistical methods. One of the main obstacles for process monitoring stems from the broadband nature of the frequency spectrum. Therefore, traditional SPC based approaches may not lend towards detection of CMP process anomalies.

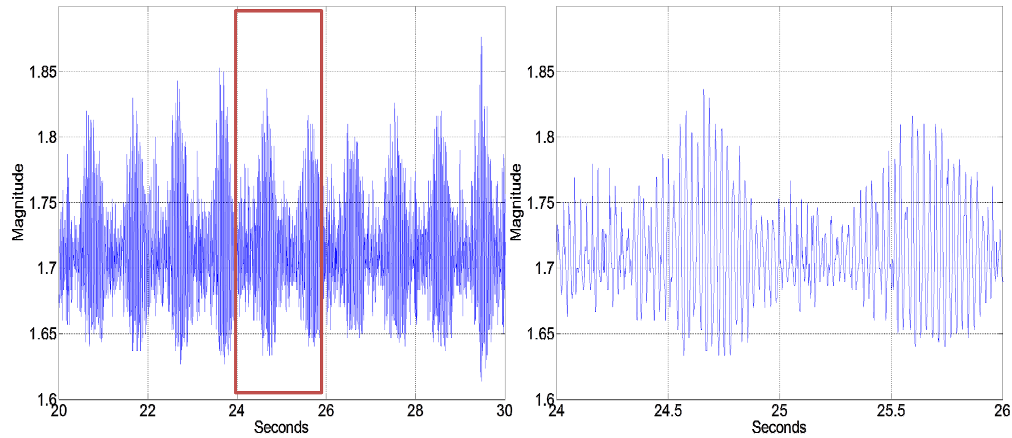


Figure 5.19 Typical vibration sensor data from CMP process

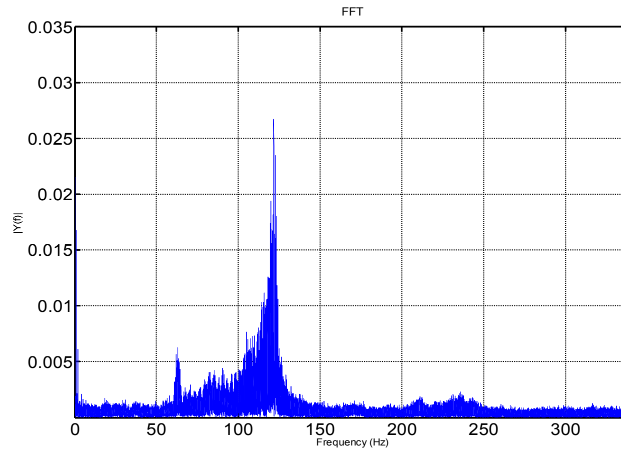


Figure 5.20 Frequency spectrum of vibration signal data

In this section, we apply the RNDP approach for detection of different types of in process CMP anomalies. These anomalies take the following forms:

i. Changes in polishing load or down-force.

ii. Sudden depletion of polishing slurry.

iii. Gradual wear of the polishing pad.

Changes in polishing load or down-force

In this experiment various polishing loads are applied and vibration sensor data is acquired. For instance, in Figure 5.21, a low load (5 lb) is active for the first half (2000 of data points, ~ 3 sec. of polishing) of the time series. The second half of the data (2000 data points onwards) belongs to an experiment conducted under high load (8 lb) conditions. All other factors, namely, head speed and base speed are maintained constant at 60 RPM and 150 RPM, respectively.

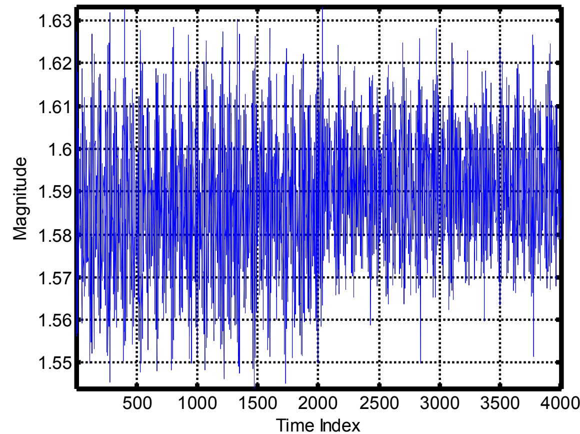


Figure 5.21 Representative vibration signal patterns obtained under changing load conditions.

RNDP, CUSUM and EWMA control charts are applied to this time series data in order to compare detection delay. As in the simulated cases, the control limits are adjusted *a priori* in order to maintain identical Type I error probabilities. The first out of control point is noted as the detection time, consequently, the results are presented in Figure 5.22 for 10 unique realizations, i.e., the analyses is repeated with 10 different data sets. As evident from Figure 5.22, the RNDP method is at

least 5 times faster compared to CUSUM and EWMA control charts. The RNDP method detects the change in polishing load within ~ 21 milliseconds (ms), whereas CUSUM and EWMA require over > 120 ms.

This is pertinent from an engineering perspective, because the polishing load is considered as one of the most significant factors in CMP and determines not only physical aspects, such as the nature of tribological contact, but also key process output variables, namely, material removal rate, within wafer non-uniformity, surface quality, etc [14].

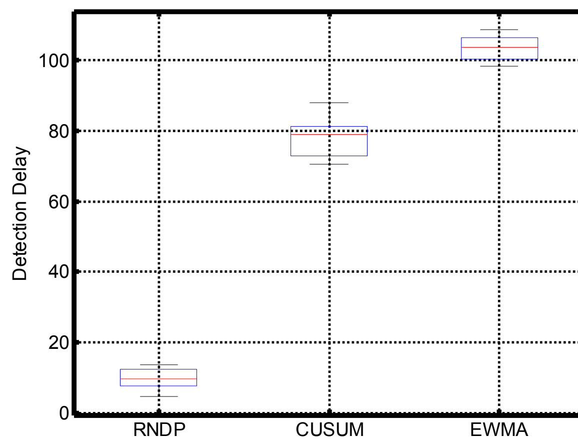


Figure 5.22 Detection delay results for changing load conditions.

Sudden depletion of polishing slurry

In this experiment, the contact load, base speed, and head speed maintained constant at 8 lb, 150 RPM, and 60 RPM, respectively. However, after 2000 data points viz. ~ 3 sec of polishing time, the slurry feed pump was choked to simulate slurry depletion. As evident from Figure 5.23, a subtle change in signal amplitude

can be discerned. We also observe that when the slurry supply is insufficient, the surface finish significantly deteriorates, although the rest of the parameters are maintained at close to optimal settings. In this particular experiment, since slurry supply was completely stopped (albeit, for a few seconds), the resulting surface was replete with scratches.

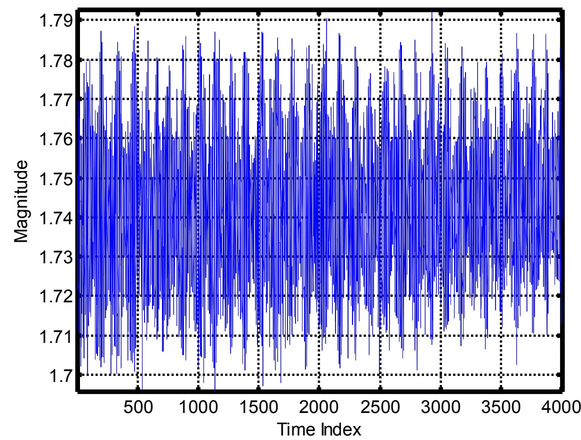


Figure 5.23 Representative vibration signal patterns obtained for slurry depletion experiments

The comparison of RNDP, CUSUM, and EWMA vis-à-vis to detection delay is presented in Figure 5.24, from which we note that the change in the vibration signal patterns as a result of slurry depletion can be detected within 50 ms using the RNDP method. In contrast, the traditional CUSUM and EWMA control charts are 2-3 times more reticent and require over 100 ms to detect the change.

In CMP operations, wafer defects, such as scratches, pitting, corrosion, etc., can cause significant revenue losses [14]. Indeed, CMP related wafer defects are considered to be among the top five reasons inhibiting semiconductor yield rates

[132]. Therefore, early detection of such anomalies can be beneficial for mitigating yield losses and rework rates in semiconductor manufacturing.

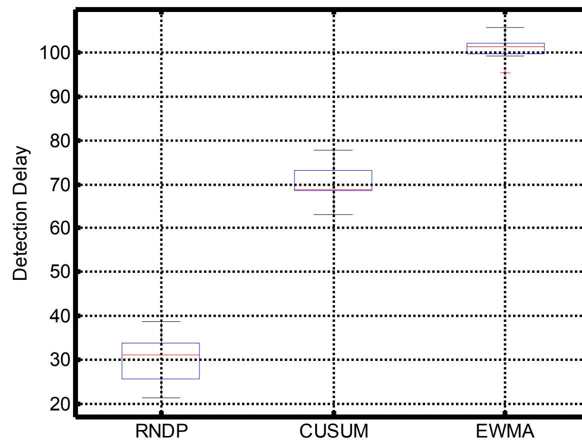


Figure 5.24 Detection delay results for slurry depletion experiments.

Gradual wear of the polishing pad

In these experiments, we deliberately degraded the performance of the polishing pad, by allowing the slurry to dry and coagulate[1]. As a result, the polishing pad is *glazed* [130-131], i.e., the fibers of the polishing pad become entangled and lose the ability to retain slurry abrasives. Also, some portions of the polishing pad may be sheared away during polishing, thus exposing the underlying layer.

Figure 5.25 depicts the vibration time series data gathered under the following CMP conditions: 8 lb contact load, 150 rpm base speed and 60 rpm head speed. The first half of the data (2000 data points, ~ 3 sec) is obtained from an

[1] For our experiments we use the Buehler Microcloth polishing pad, which is a soft synthetic polishing cloth having a fibrous structure and is preferred for gentle material removal for non-ferrous materials.

experiment where a new pad was used, while the second half of the data is gathered from an experiment conducted with a glazed pad. We can discern from Figure 5.25 that not only does the mean of the vibration signal change, but also the variance of the signal slightly increases. This is similar to the simulated case study with fault type 4.

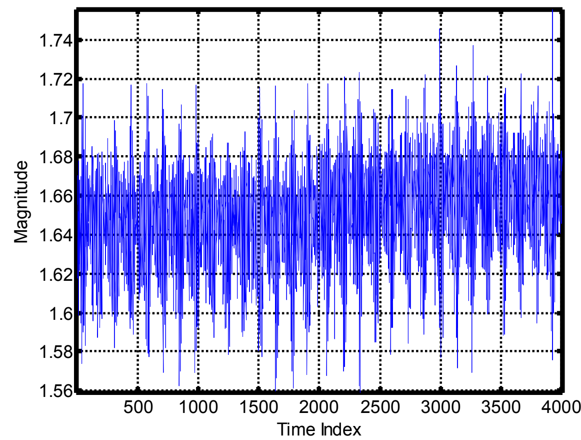


Figure 5.25 Representative vibration signal patterns obtained for pad wear experiments

The results are presented in Figure 5.26. Evidently, from Figure 5.26 the RNDP method is approximately 3 times faster compared to CUSUM and EWMA control charts. The RNDP method detects the change in pad wear within ~ 71 ms, whereas CUSUM and EWMA require over > 200 ms on average over 10 replications.

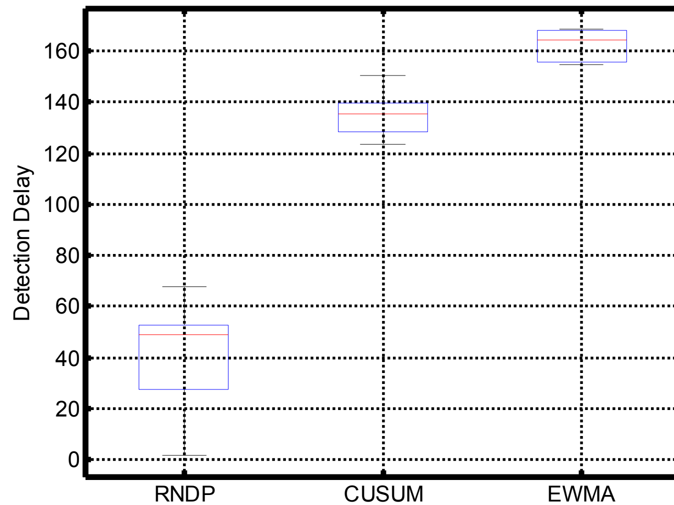


Figure 5.26 Detection delay results for pad wear

Effect of Multiple Faults

In this study, during CPM process, two different kinds of CMP process drifts were induced concurrently. The normal condition is assumed to occur under the following experimental settings: 5lb polishing load, 150 RPM base speed, and 60 RPM head speed. After 2000 data points, the polishing load was increased to 8 lb (the other settings were maintained at constant). Next, the slurry feed was cut off after 4000 data points while the polishing load was maintained at 8 lb. Vibration signal patterns acquired for this experiment is presented in Figure 5.27.

In order to assess the performance of the RNDP method for such multiple faults case, we not only detect the change, but also attempt to *classify* the type of drift. In essence, this study is analogous to simulated case study with different types faults (fault type 6).

This is important from an application stand point; because in addition to warning the operator that the process has drifted out of control, the technique can be used indentify the underlying root cause of the drift. This can enable accurate fault diagnosis, so that timely corrective action can be taken. Comparison between RNDP method with mean shift (a frequently used unsupervised clustering method [133]), is illustrated in Table 5.4.

The results indicate that the RNDP approach, apart from being able to detect the onset of process anomalies at an early stage, can also classify the type of fault with > 80% accuracy, which is significantly better than the mean shift approach. More pertinently, the Type II error probability is much lower (~ 15%) for RNDP method compare to the mean shift approach (~ 25%). This implies that, the RNDP method affords both lower Type I as well as Type II error rates.

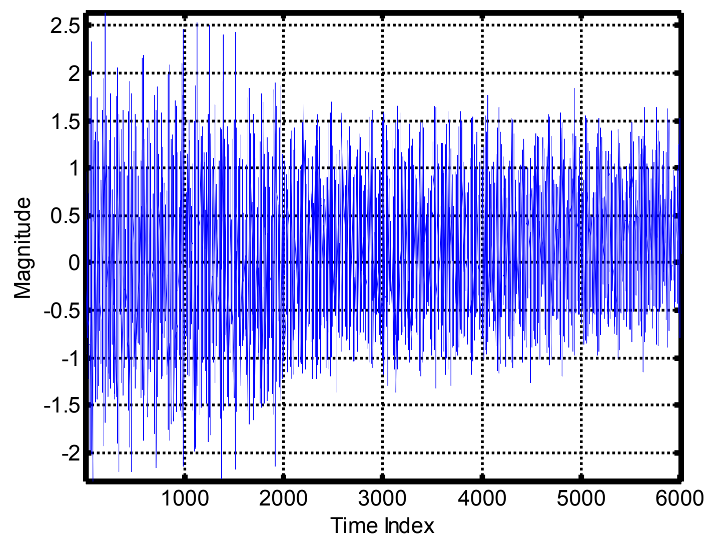


Figure 5.27 Vibration data time series for multiple fault experiment

Table 5.4 Classification accuracy for multiple faults experiment

RNDP		
Fault Detection Accuracy	94.2%	
Fault Classification	C2	C3
C2	**87.1%**	12.9%
C3	16.3%	**83.7%**
Mean Shift		
Fault Detection Accuracy	85.2%	
Fault Classification	C2	C3
C2	**75.4%**	24.6%
C3	31.4%	**68.6%**

5.4 Summary

In this chapter, we devised a non-parametric Bayesian modeling to establish an online process monitoring for change point detection in CMP process. During the CMP process, vibration signal data captured by the MEMs sensor and divided into sliding windows are modeled by using RDP in a markovian fashion. Based on the mixture models, resulting RDP for each sliding window RNDP technique is utilized to detect any change between sliding windows, hence the change in the CMP process. We showed by simulation and experimental case studies that the RNDP technique outperforms the traditional SPC methods such as CUSUM and EWMA.

CHAPTER VI

MULTI-SENSOR FUSION BASED PROCESS MONITORING OF ULTRA PRECISION MACHINING

Consistent surface finish is vital for industrial ultra-precision machining (UPM) processes. Therefore, monitoring the UPM process is critical to maintain dimensional and surface integrity of UPM substrates. A non-parametric Bayesian analysis approach is utilized to model non-stationary and non-linear sensor signal acquired from multiple heterogonous sensors. Evidence theory is used to combine the each decision obtained from individual sensors for more reliable and coherent results. Several experimental case studies are carried out to validate the proposed method.

6.1 Introduction

Surface morphology is a critical determinant of functional performance for components which are critical to optical, semiconductor, aerospace, and defense industries [1, 9]. Ultra-precision machining (UPM) is a single point turning process which is used to achieve specular surface finish (surface roughness (Ra) in the nanometer range). In UPM a single crystal diamond tool is used to remove

material at the depth of cut in micrometer range (1-50 μm) on a very precise and rigid lathe, which is specially built to minimize the effect of vibrations and extraneous noise. Since minute instabilities and process drifts can cause significant surface variations in UPM process, monitoring of UPM process is vital for achieving satisfactory surface quality.

In the past two decades, considerable research has been conducted for UPM process monitoring using various sensors, such as piezoelectric vibration [97], acoustic emission (AE) [98], and temperature [99] sensors were. While these methods are effective to some extend in detecting process anomalies, particularly in post-hoc scenarios, these methods suffer from two key lacunae:

i. They do not utilize information from multiple sensors.
ii. They are mostly offline and not designed for real-time monitoring of UPM process conditions.

From the perspective of multi-sensor based process monitoring two vital concerns need to be addressed; (1) how to use signals acquired from sensors to make decisions on the process change, and (2) how to integrate these decisions from each individual sensor to reach more coherent and reliable decisions.

In UPM process monitoring applications, the first research question is to investigate the relationship between the measured signal and process performance or process changes. Sensor based tool wear monitoring using acoustic emission

signals [134-135] and cutting force signals [136-138] have been reported. Acoustic emission (AE) [139] and vibration sensors [140] used for monitoring surface profile in UPM process. However, most of these sensor-based process monitoring techniques are based on linear and deterministic models, while it is known that the UPM process is non-linear and non-stationary [16]. In order to accurately represent the non-linear and non-stationary nature of the process, it is crucial to develop a non-parametric data-driven model.

Due to the development of different type of sensor technology, more and more data can be acquired from sensors. The second question is how to combine the data gathered from different sensors to obtain more accurate and reliable information from the complex processes such as UPM. Extensive work has been done in information fusion which includes signal-level fusion, feature-level fusion and decision-level fusion [74]. Principal component analysis [75-76, 80] and dynamic system estimation methods, such as Kalman filtering [78, 83], have been used for signal fusion. Machine learning tools [77, 81-82] have been used to fuse information in feature level. Bayes theory is reported as one of the most commonly used decision-level fusion technique [87-90].

In this research, three piezoelectric vibration sensors, a three-axis force sensor, and an acoustic emission sensor are used to collect data from the UPM process. It is possible that the outcomes of these different sensors lead to conflicting

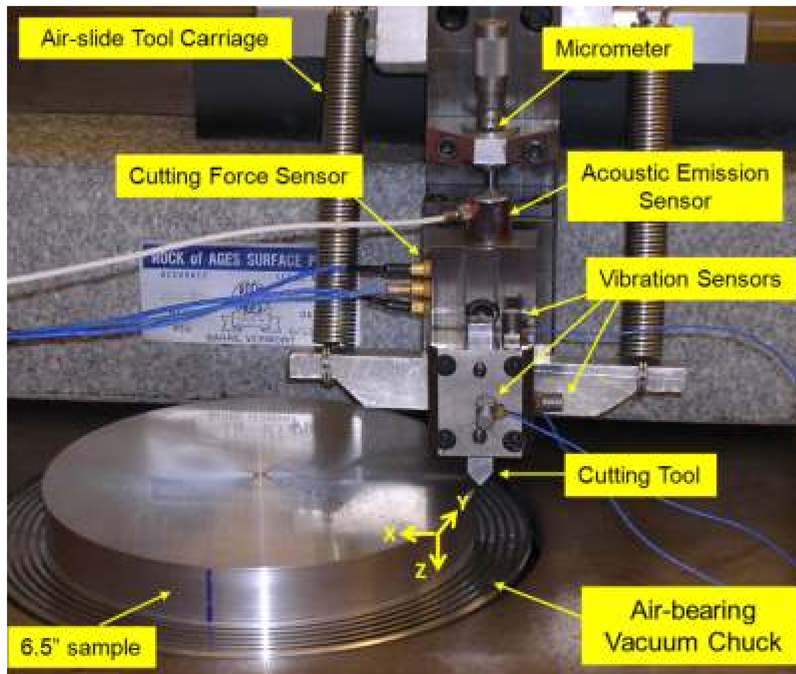
decisions. Therefore, it is vital to fuse information from different sources into a coherent decision.

This research involves integrating the non-parametric Bayesian modeling with information fusion technique. Complex non-linear and non-Gaussian patterns are modeled by using a non-parametric Bayesian model, namely, Dirichlet Process [96] for each sensor. The decisions from each sensor are fused by using information fusion techniques [84].

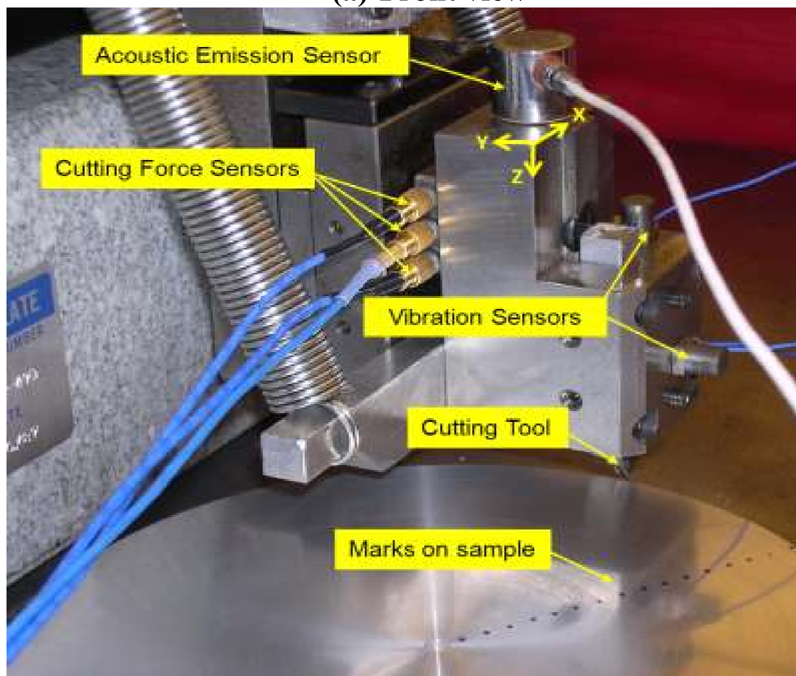
The rest of the chapter is organized as follows. The research methodology for multi-sensor based process monitoring is presented in Section 6.2. The proposed method is demonstrated with case studies in Section 6.3. Summary and the discussion of this research are provided in Section 6.4.

6.2 Research Approach for Multi-sensor Based Process Monitoring

This research is based on a single diamond tool surface cutting experiments with nano-metric surface finish (Ra~ 5 to 100 nm). The experimental setup consists of an aerostatic spindle bearing (model Block-Head® 4R) and air-slide tool carriage manufactured by Professional Instruments Inc. (Figure 6.1). Cylindrical (φ 16.25 cm × 3.75 cm) aluminum alloy (Al 6061) work-pieces are used for surface cutting on the UPM setup (Figure 6.1). Three piezoelectric vibration and a three-axis force sensors are mounted on the setup.



(a) Front view



(b) Side view

Figure 6.1 UPM experimental apparatus

As seen in Figure 6.1, vibrations sensors (10 KHz) are located on the orthogonal directions of the tool holder and three-axis force sensors (10 KHz) are located underside of toll holder to measure force signals. In addition, an AE sensor (1 MHz) is mounted on the top of the tool holder. Details of the sensor system are summarized in Table 6.2.

Full factorial design of experiments are done by setting parameters of spindle speed (500, 1000, 2000 RPM), feed rate (1.5, 3, 6 mm/min), and depth of cut (5, 10, 20, 25 μm) as shown in Table 6.1.

Table 6.1 Design of experiment parameters

<table>
<tr><th>Depth Of Cut</th><th>Feed Rate</th><th>Rotating speed</th></tr>
<tr><td>5 μm</td><td>1.5 mm</td><td>500 rpm</td></tr>
<tr><td>10 μm</td><td>3.0 mm</td><td>1000 rpm</td></tr>
<tr><td>20 μm</td><td rowspan="2">6.0 mm</td><td rowspan="2">2000 rpm</td></tr>
<tr><td>25 μm</td></tr>
</table>

Table 6.2 Details of the sensing system mounted on the UPM setup

<table>
<tr><th>Signal Type</th><th>Sensor Orientation</th><th>Symbol</th><th>Sampling Rate</th><th>Sensor Type</th></tr>
<tr><td>Vibration</td><td>Along the feed direction
Across the feed in the XY plane
Vertical</td><td>V_X
V_Y
V_Z</td><td>10 KHz</td><td>Kistler 8728A500</td></tr>
<tr><td>Force</td><td>Along the feed direction
Across the feed in the XY plane
Vertical</td><td>F_X
F_Y
F_Z</td><td>10 KHz</td><td>Kistler 9251A</td></tr>
<tr><td>AE</td><td>Vertical</td><td>AE</td><td>1 MHz</td><td>Physical Acoustics R80</td></tr>
</table>

The overall methodology for sensor fusion is summarized in Figure 6.2. The proposed research mainly consists of two stages of fusion mechanism. The first stage is feature-level fusion, where the features, such as statistical features are extracted from heterogeneous sensor signals and fused using Dirichlet process (DP). In the second stage, Depmster–Shafer evidence theory [141] is used as a decision-level fusion mechanism.

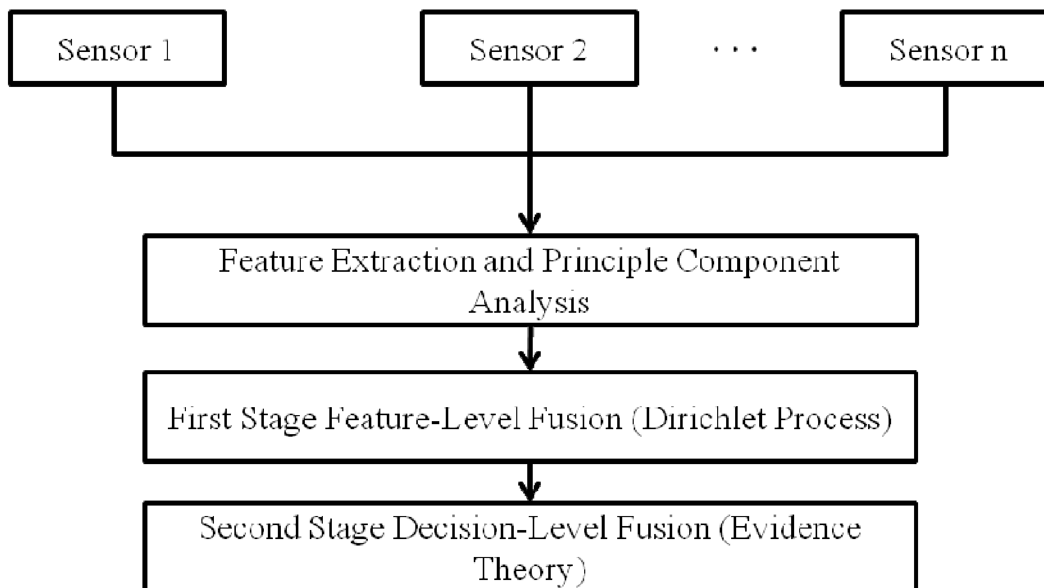


Figure 6.2 Proposed Information Fusion Mechanism

The detailed overall methodology is illustrated in Figure 6.3. In the first stage, features such as mean, standard deviation, skewness, kurtosis, etc, are extracted from force, vibration and acoustic emission sensor signals and modeled by using Dirichlet process by estimating their probability density function for each sensor. Likelihood value is calculated and probability mass function from is constructed

for each sensor. Using evidence theory, probability mass functions from multiple sensors are fused and final decision is made by choosing the configuration the maximum likelihood value.

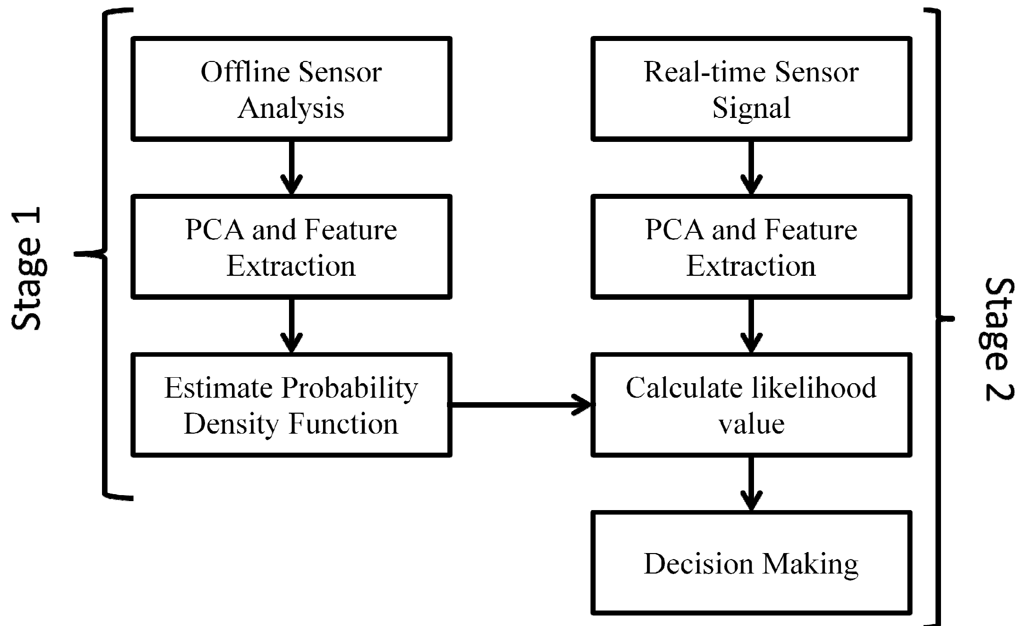


Figure 6.3 Proposed overall methodology for sensor fusion for UPM process monitoring

6.2.1 Feature level fusion by using Dirichlet process

In feature-level fusion structure, features, such as such as frequency domain and time domain features are extracted from multiple sensor data. Features are extracted from each sensor and combined to obtain higher dimensional data which will provide more information about the process. High dimensional data will be modeled which will perform as a feature-level fusion mechanism. Extracted time series features (mean, standard deviation, skewness, kurtosis, range and root mean square) which are used for feature-level fusion summarized in Table 6.3.

Table 6.3 Summary of extracted features

Features	Equation
Mean	$\bar{x} = \frac{\sum_{j=1}^{N} x_j}{N}$
Standard deviation	$\sqrt{\frac{\sum_{j=1}^{N}(x_j - \bar{x})^2}{N-1}}$
Skewness	$\frac{\frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^3}{\left(\frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^2\right)^{3/2}}$
Kurtosis	$\frac{\frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^4}{\left(\frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^2\right)^{2}} - 3$
Range	$\max(x_j) - \min(x_j), \quad j = 1, \dots, N$
Root mean square	$\sqrt{\frac{1}{N}\sum_{j=1}^{N} {x_j}^2}$

Probability density information of extracted features are estimated with mixture of Gaussian distributions as in Eq. (6.1):

$$p(x_i) = \sum_{j=1}^{K} \pi_j N(x_i|\theta_j) \tag{6.1}$$

where x_i is the feature vector extracted from sensor data in UPM process . Linear combination of K Gaussian distribution with weights π_j and parameters θ_j (mean μ_j and variance ${\sigma_j}^2$) is used to approximate non-Gaussian distribution. If the number of mixture components K is known a priori the expectation maximization (EM) algorithm [117] can be used to estimate π_j and θ_j but in practice K is usually unknown. Dirichlet Process can be utilized to approximate non-Gaussian

distribution without prior knowledge of K. For more detailed information about the procedure refer to the Section 5.2.2.

6.2.2 Decision level fusion with evidence theory

In decision-level fusion structure, sensor signals are first processed and analyzed individually. Then, the decisions obtained from individual analysis are combined for final decision. In the proposed methodology, detection and classification techniques are applied, and a decision is obtained for each individual sensor. These individual decisions are fused to get more accurate results in decision level fusion. The features from multiple sensors are extracted and Dirichlet Process technique is used for classification. Dirichlet Process is used for estimating the density function for each class, and then likelihood value is calculated by using density function to construct mass function for each test data point after normalizing as follows:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KM} \end{bmatrix} \tag{6.2}$$

where, K is the number of sensors and M is the number of conditions. Each row $p_k = [p_{k1}\, p_{k2} \ldots p_{kM}]$ is the probability mass function assigned by the kth sensor assigned to the set of M classes. All combinations of sensors are analyzed and the best combination is selected according to classification accuracy results. The set of n-sensor combinations are represented as $\binom{K}{n}$, where it defined as n-way

$(n = 1,2, \dots, K)$ combinations of K sensors. The probability mass function for fused decision can be expressed as follows:

$$m = p_1 \oplus p_2 \oplus \dots p_K \tag{6.3}$$

where $\oplus$ represents the operator for combination of probability mass function. If the types of faults are assumed independent, probability mass function for fault j can be computed as follows [142]:

$$m_j = \prod_{k=1}^{K} p_{kj} \Bigg/ \left(1 - \sum_{j=1}^{M} \prod_{k=1}^{K} p_{kj}\right) \tag{6.4}$$

where m_j represents the fused probability mass function. The term $\sum_{j=1}^{M} \prod_{k=1}^{K} p_{kj}$, is the basic probability mass associated with conflicts among the sources. The denominator $\left(1 - \sum_{j=1}^{M} \prod_{k=1}^{K} p_{kj}\right)$ is the normalization factor. Lower the value of $\sum_{j=1}^{M} \prod_{k=1}^{K} p_{kj}$, means less conflicts between the sources. After getting the probability mass function for j^{th} fault highest belief measure will be chosen for fused decision.

6.3 Validation of proposed method

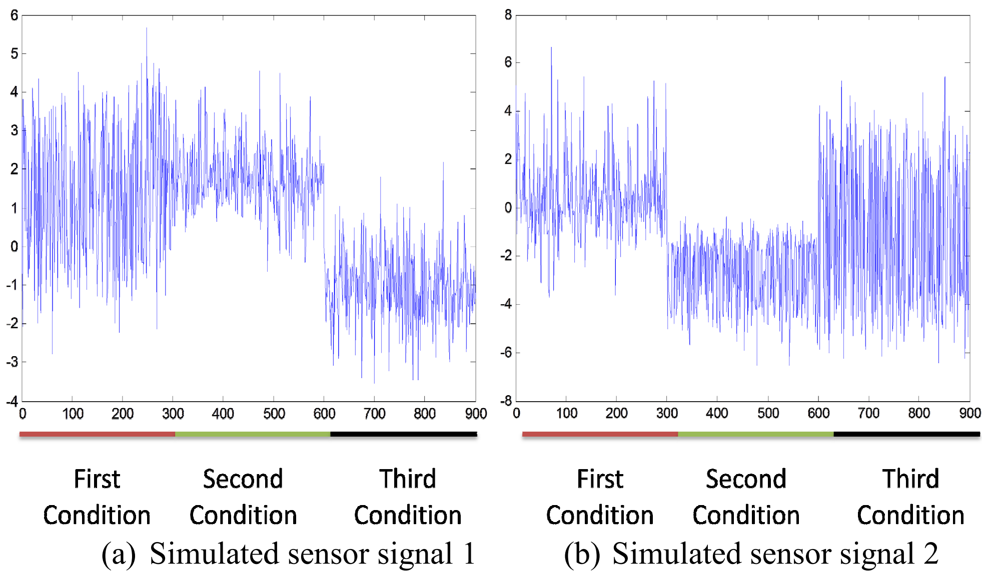
The proposed sensor fusion methodology for UPM in this study has two stages, i.e., feature-level fusion and decision-level fusion. Several case studies will be carried out to validate the proposed fusion methods in both stages.

6.3.1 Numerical case studies for feature-level fusion

Two random sensor signals are simulated each of which is generated from three normal distributions with different mean and standard deviations shown in Table 6.4. Three different conditions are created by choosing different means and variances for each condition. The time series data for two different signals are shown in Figure 6.4.

Table 6.4 Distribution parameters used for generated data

	Condition 1	Condition 2	Condition 3
Signal 1	$\mu = 1, \sigma = 1.2$	$\mu = 1.5, \sigma = 0.5$	$\mu = -1, \sigma = 0.8$
Signal 2	$\mu = 1, \sigma = 1$	$\mu = -3, \sigma = 0.5$	$\mu = 0, \sigma = 2$



(a) Simulated sensor signal 1 (b) Simulated sensor signal 2

Figure 6.4 Simulated data generated from non-Gausssian distribution for three condition

Dirichlet process is applied to generated signals individually, and the classification accuracy results for each signal are shown in first and second column of the Table 6.5. In addition, Dirichlet process is used as fusion method by modeling with mixture of multivariate normal distributions to combined signal. The results of classification accuracies and comparisons with other classification methods for fused signal are presented in the last column of Table 6.5. Table 6.5 shows the mean values and standard deviation values (in parenthesis) after realization of 10-fold cross validation. Fused data in 2D is illustrated in Figure 6.5 with the comparisons of various fusion algorithms. Both Table 6.5 and Figure 6.5 confirm that Dirichlet process is more accurate and robust.

Table 6.5 Classification results for different classification methods

Classifier	Accuracy		
	Signal 1	**Signal 2**	**Fused Signal**
Linear Discrimant	0.65 (0.012)	0.58(0.010)	0.75(0.016)
Naïve Bayes	0.71(0.024)	0.62(0.018)	0.83(0.021)
Neural Network	0.73(0.029)	0.64(0.026)	0.88(0.027)
Support Vector Machine	0.75(0.021)	0.69(0.024)	0.89(0.022)
Dirichlet Process	0.76(0.031)	0.71(0.022)	0.92(0.027)

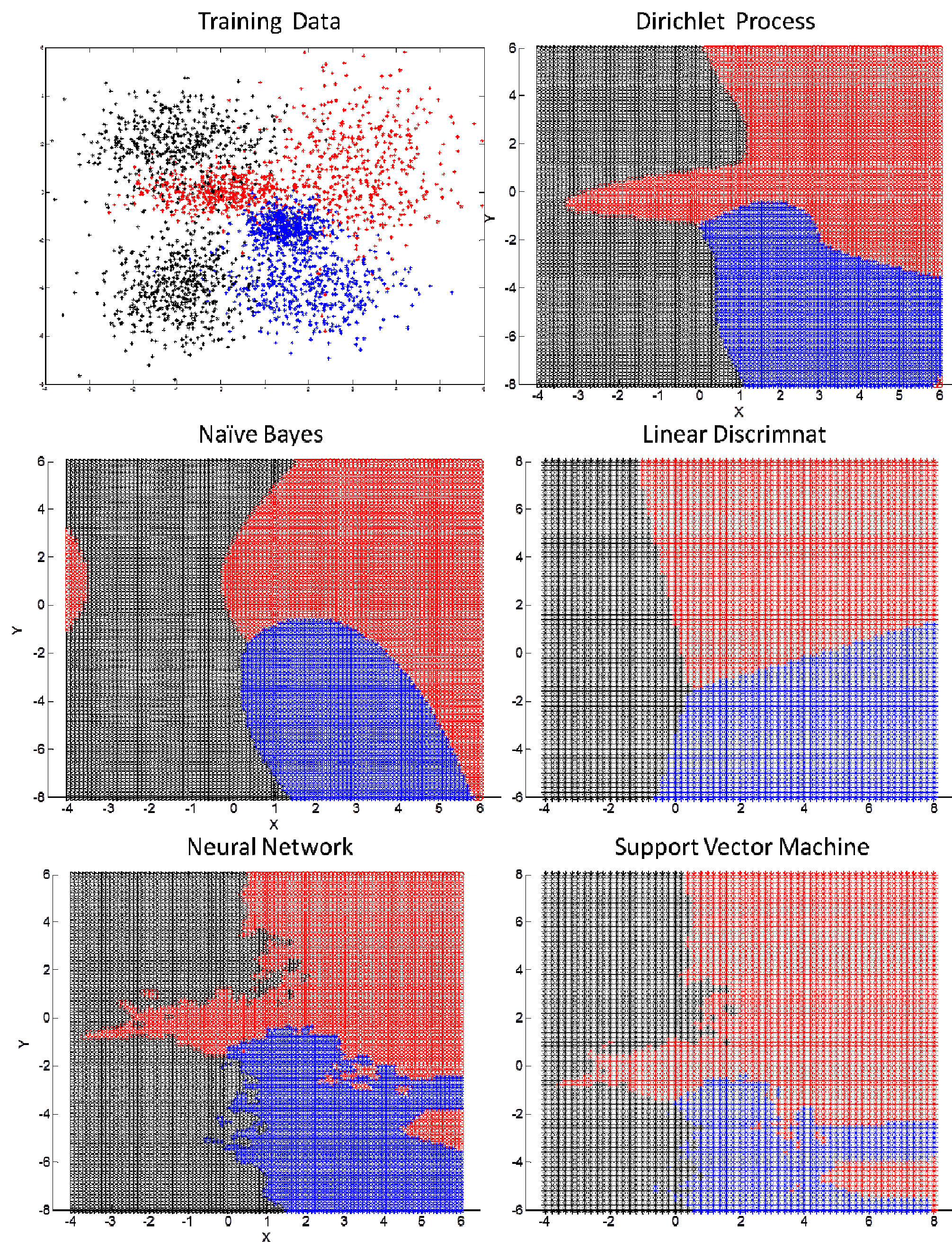


Figure 6.5 Comparison of feature-level fusion algorithms along with the training data on top left.

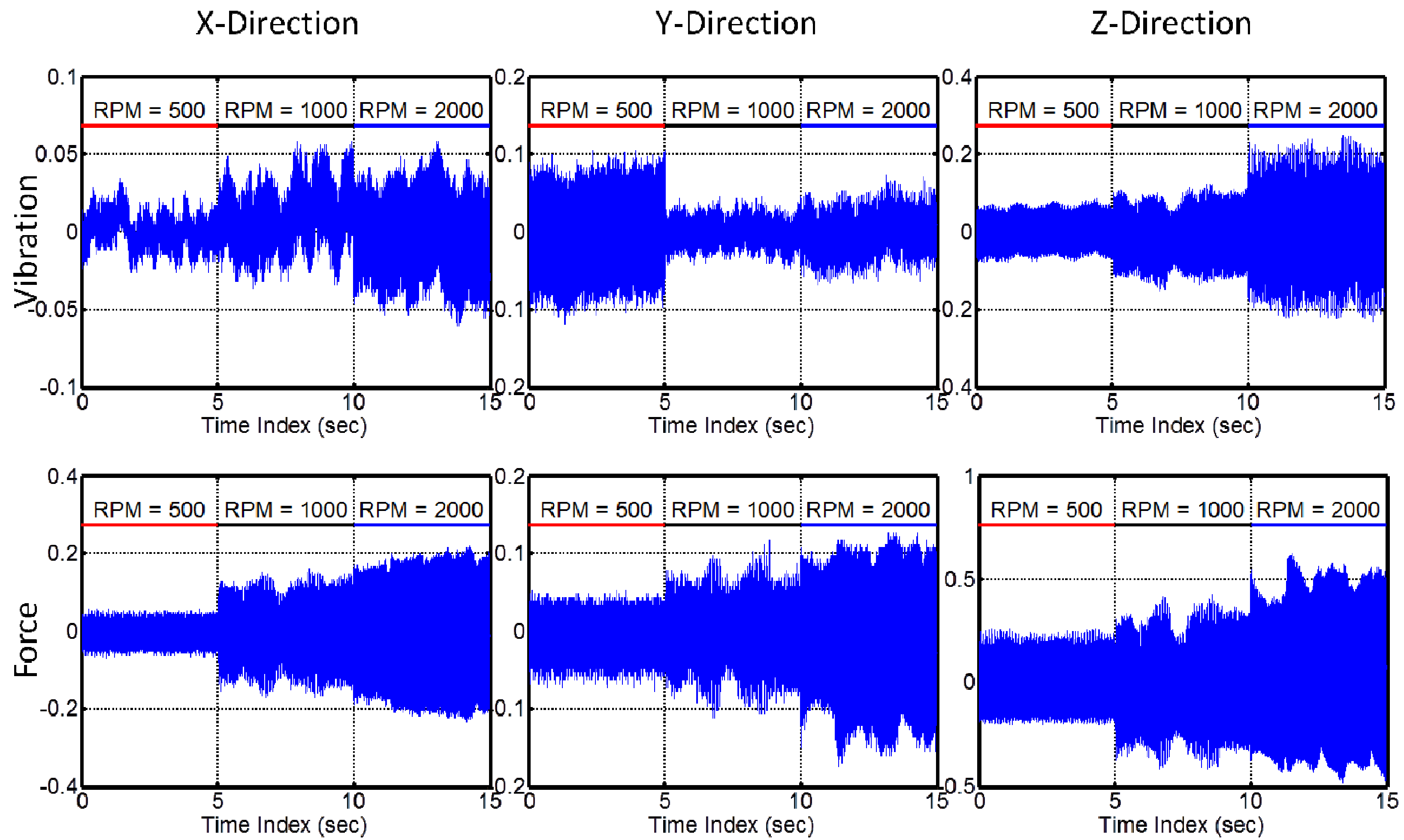


Figure 6.6 Spindle speed experiments

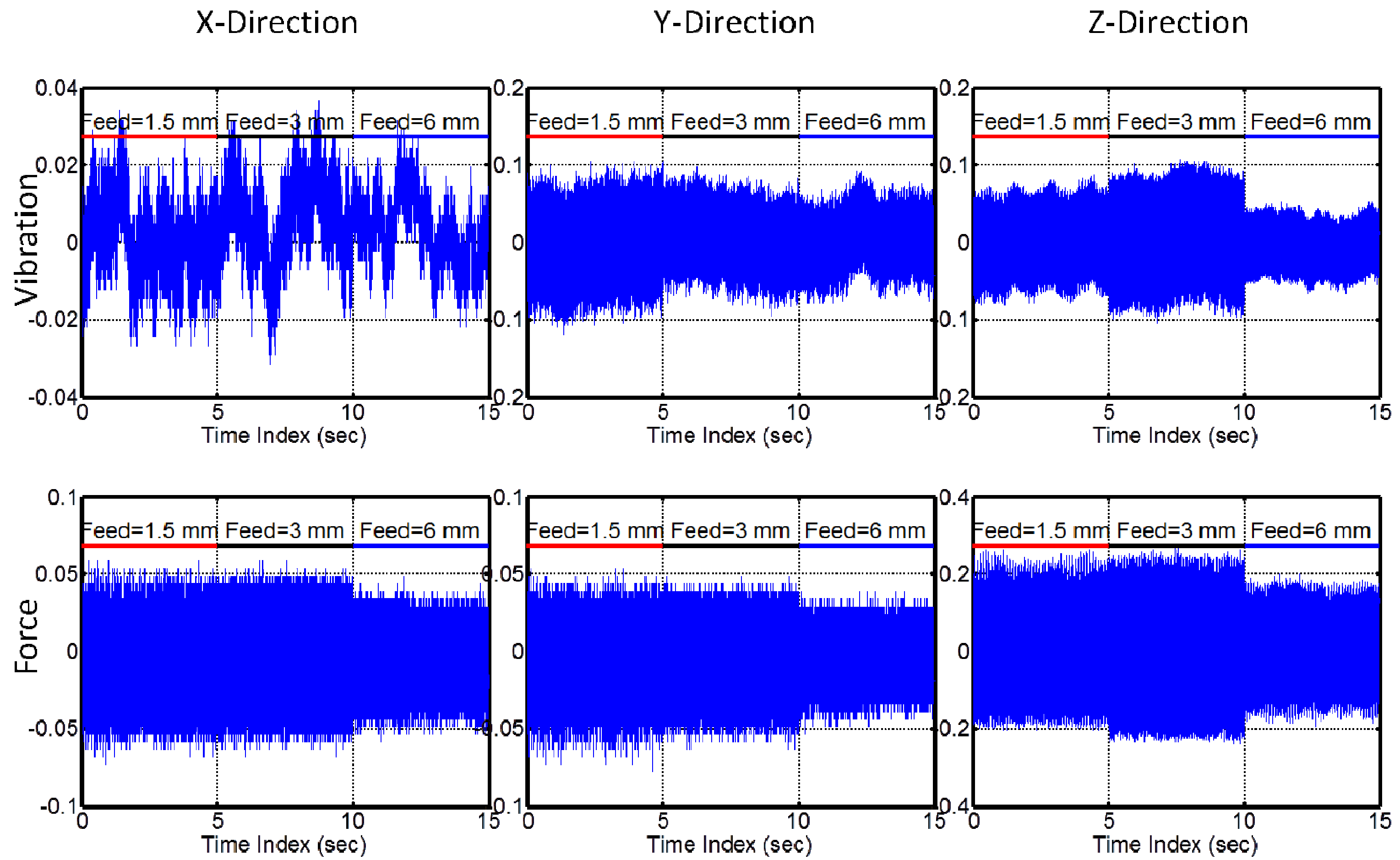


Figure 6.7 Feed rate experiments

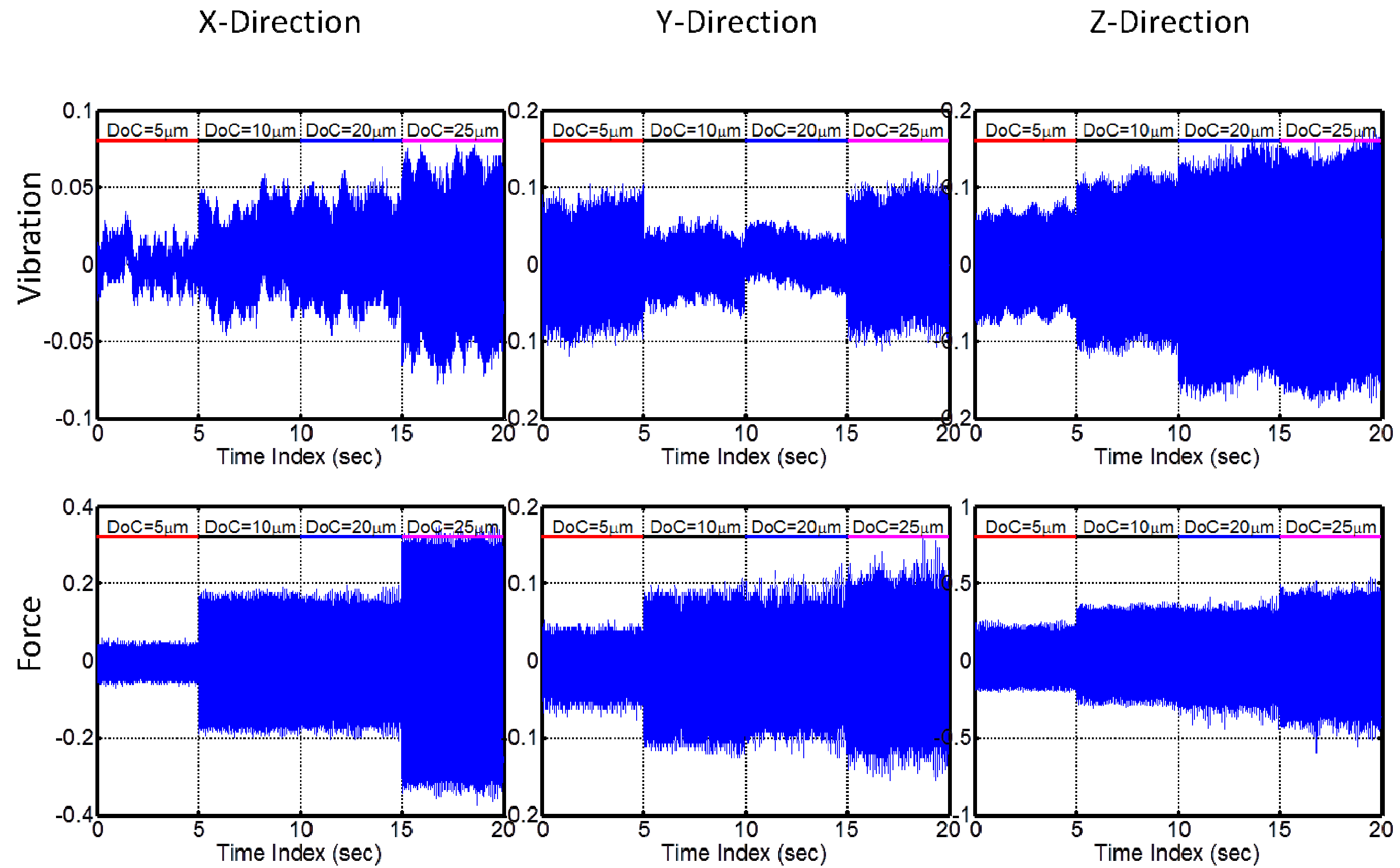


Figure 6.8 Depth of cut experiments

6.3.2 UPM Case Studies

Ultra Precision Machining (UPM) data is analyzed to classify UPM data with three conditions using three signals. Vibration and force signal with 3 channels (X-axis, Y-axis and Z-axis) are collected from experiments with different depth of cut (5 µm, 10 µm, 20 µm and 25 µm), spindle speed (500 rpm, 1000 rpm and 1500 rpm) and feed rate (1.5 mm/sec, 3 mm/sec and 6mm/sec).

Dirichlet Process is applied to estimated density information for each condition within each signal data. Probability mass function for each test data point can be constructed in Eq. (6.8) by computing the likelihood value. Then as in Eq. (6.10), probability mass functions for each sensor are combined to get fused decision according to the highest likelihood value.

In Spindle speed experiments depth of cut and feed rate values are kept at constant values of 5 µm and 1.5 mm/sec. Investigation of change in the spindle speed values (500, 1000, 2000 rpm) is done and results are shown in Table 6.6 (individual sensors) and Table 6.7 (fused sensors).

The Second experiments are investigating the effect of the feed rate. Constant values are used for spindle speed (500 rpm) and depth of cut (5 µm) while feed rate takes values of 1.5 mm/sec, 3 mm/sec and 6 mm/sec, respectively. The results are illustrated in Table 6.8 and Table 6.9.

Spindle speed and feed rate values are kept constant at 500 rpm and 1.5 mm/sec, while depth of cut has taken values of 5 µm, 10 µm, 20 µm and 25 µm in the final

experiments. Classification errors for individual and fused sensors are summarized in Table 6.10 and Table 6.11, consecutively. From Table 6.10 and Table 6.11, it is evident that fused decision has higher classification accuracy than individual sensors. Best individual sensor Fy has an accuracy rate of 60%, while the combination of sensors Vz, Fx, and Fz has an accuracy rate of 89%.

Table 6.6 Accuracy of individual sensors for spindle speed experiments

	V_x	V_y	V_z	F_x	F_y	F_z
Accuracy	0.50 (0.07)	0.41 (0.04)	0.58 (0.06)	0.68 (0.04)	0.74 (0.05)	0.64 (0.04)

Table 6.7 Accuracy of fused sensors for spindle speed experiments

	VxVy VzFy	VyVzFy	VyVzFx Fy	VzFy	VxVyVzFy Fz	VyVzFy Fz
Accuracy	0.88 (0.07)	0.87 (0.04)	0.87 (0.08)	0.86 (0.09)	0.86(0.02)	0.86 (0.02)

Table 6.8 Accuracy of individual sensors for feed rate experiments

	V_x	V_y	V_z	F_x	F_y	F_z
Accuracy	0.50 (0.07)	0.41 (0.04)	0.58 (0.06)	0.68 (0.04)	0.74 (0.05)	0.64 (0.04)

Table 6.9 Accuracy of fused sensors for feed rate experiments

	VxVy VzFy	VyVzFy	VyVzFx Fy	VzFy	VxVyVzF yFz	VyVzFy Fz
Accuracy	0.88 (0.07)	0.87 (0.04)	0.87 (0.08)	0.86 (0.09)	0.86(0.02)	0.86 (0.02)

Table 6.10 Accuracy of individual sensors for depth of cut experiments

	V_x	V_y	V_z	F_x	F_y	F_z
Accuracy	0.43 (0.02)	0.31 (0.04)	0.44 (0.05)	0.57 (0.03)	0.60 (0.08)	0.58 (0.05)

Table 6.11 Accuracy of fused sensors for depth of cut experiments

	VzFx Fz	VxVzFx Fz	VzFxFy	VxVzFx	VxVzFx Fy	VzFx
Accuracy	0.89 (0.04)	0.84 (0.05)	0.83 (0.04)	0.83 (0.06)	0.83 (0.04)	0.83 (0.07)

The results depicted in above tables indicate that certain sensor combination gives the best result. While the common sense dictates that adding more information should increase the classification accuracy, UPM case studies shows that it is not always the case. In the following we will carry out simulated case studies to investigate how correlation between sensor signals and noise level in sensor signals effect the performance of sensor fusion.

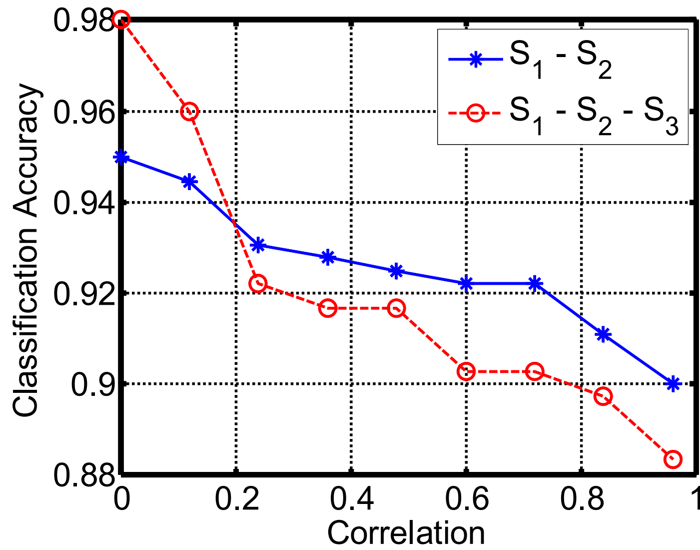


Figure 6.9 Correlation analysis

Figure 6.9 shows the how the correlation between two signals effect the classification accuracy. As a continuation of the case study in Section 6.3.1 we added one more signal which is correlated with the second signal. Correlation levels are shown in the x-axis and the corresponding classification results are plotted solid line for fusion of signal 1 and signal 2 while fusion of all signals are depicted with the dashed line. We can conclude from the graph that adding highly correlated signals does not increase the accuracy.

6.4 Summary

In this research, a two stage fusion mechanism is developed. In the first stage, features from heterogeneous sensor signals are extracted and principal component analysis is applied for further compression. Dirichlet Process modeling (DPM) is used to formulate density information for each sensor. It is shown that DPM

provides higher accuracy and more robust modeling compared to the other widely used classification methods. In the second stage, of the fusion mechanism, evidence theory is applied for decision level fusion. In UPM case studies we have shown that fusing information from different sensors increases the accuracy by 15% to 30% over each individual sensor.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

Real-time monitoring approaches in nano-level manufacturing processes, such as ultra-precision machining (UPM) and chemical mechanical planarization (CMP), are invesitgated in this study. Detailed conclusions to achieve specific research objectives are listed as follows:

(1) To establish how process state relates to sensor features:

a. In CMP process statistical features are extracted from wired and wireless sensors. Subsequently, principal component analysis (PCA) is applied for further compression of data. Compressed features are related with material removal rate by utilizing artificial neural network (ANN) with high R-square value (~90%), whereas linear regression can only achieve (~47%) R-square value.

b. In the literature, it is shown that sensor signal acquired form CMP process is non-linear and non-stationary [92-95]. Sensor signal is distributed non-Gaussian due to non-linearity. Recurrent Dirichlet

process (RDP) is applied to model sensor data which captures non-Gaussian and the non-stationary nature of the process efficiently.

c. Various depths of cut parameters are used to simulate different surface variations. Statistical features are extracted from sensor data which are acquired from heterogonous sources. A probabilistic non-parametric Bayesian modeling is used to model successfully different surface variations represented by statistical features.

(2) To monitor process state from measured signals in real time:

a. A real-time monitoring approach for end-point detection is developed. For early detection of end-point, a sequential Bayesian analysis, namely, particle filtering (PF), is used for the prediction of process state represented by sensor signal features (1a). We showed that, the PF method outperforms convential online estimation techniques, such as ARMA, Kalmana filtering and extended Kalman filtering.

b. Change point detection in CMP process is investigated. Recurrent nested Dirichlet process (RNDP) is developed to monitor sensor signal which is modeled by RDP (1b) to monitor any change in the process. Experimental studies show that RNDP technique detects changes in the process ~10 times faster than traditional statistical control charts, such as EWMA and CUSUM.

c. Non-parametric Bayesian modeling (1c) is used to model surface variations for each sensor. According to Bayesian modeling, individual decisions are made for each sensor. Decision outputs from each sensor are fused by using evidence theory to achieve a more coherent global decision. Initial studies show that fused decisions are ~20% more accurate than individual decisions.

(3) To validate the theoretical contributions in research objectives (1) and (2):

a. A LapMaster 12 bench-top lapping machine is equipped with wired and wireless sensors. Copper wafers of 4-in diameter and 16-gauge thickness were polished to test end-point detection method (1a-2a).

b. A Buehler (Automet 250) polishing machine is used for polishing blanket copper work pieces. The CMP setup is mounted with MEMS accelerometer (Analog Devices ADXL 335).Various experimental studies are used to validate change monitoring methodology (1b-2b).

c. UPM setup is used for surface finishing of aluminum alloy (AL 6061) work pieces. Sensor signals were acquired from heterogeneous sources, such as, acoustic emission, 3-axis vibration and force sensors for surface variation experiment (1c-2c).

Future work directions

The methodologies developed in this dissertation can be applied to a wide variety of areas. One future direction of this research would be applying techniques developed in this dissertation in different areas, such as, sensor based manufacturing, robotics, biostatistics, etc.

Some future directions of this research can be suggested as follows

- In this dissertation, investigations have been made on change point detection in CMP and a special application, i.e., end point detection (EPD). Although the developed methodology effectively detects the change such as end-point, it is not suitable for classifying multiple changes. During CMP process, the quality of wafers may suffer from different kind of defects, such as scratches, dishes, and erosion. In order to improve the quality of the process, these anomalies should be detected timely and classified accurately.
- In UPM process, we showed that with combination of heterogonous sensors, the surface variation prediction created by various experimental conditions can be improved. In this research, the physical interpretation of the results has not been investigated yet. Why such combination of sensors gives us better results can be a good motivation for digging deeper in this direction.

REFERENCES

[1] N. Ikawa, R. Donaldson, R. Komanduri, W. Konig, P. McKeown, T. Moriwaki, and I. Stowers, "Ultraprecision metal cutting — The past, the present and the future," *CIRP Annals - Manufacturing Technology,* vol. 40, pp. 587-594, 1991.

[2] R. Komanduri, D. A. Lucca, and Y. Tani, "Technological Advances in Fine Abrasive Processes," *CIRP Annals - Manufacturing Technology,* vol. 46, pp. 545-596, 1997.

[3] "PBS Nova, Making Stuff: Smaller," ed, 2011.

[4] J. Corbett, P. A. McKeown, G. N. Peggs, and R. Whatmore, "Nanotechnology: International Developments and Emerging Products," *CIRP Annals - Manufacturing Technology,* vol. 49, pp. 523-545, 2000.

[5] "Societal Implications of Nanoscience and Nanotechnology," National Science Foundation2001.

[6] N. Taniguchi, "Current Status in, and Future Trends of, Ultraprecision Machining and Ultrafine Materials Processing," *CIRP Annals - Manufacturing Technology,* vol. 32, pp. 573-582, 1983.

[7] G. E. Moore, "Cramming more components onto integrated circuits (Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp. 114)," *IEEE Solid-State Circuits Newsletter,* vol. 11, pp. 33-35, 2006.

[8] K. Blaedel and D. Thompson. (1998) Engineering precision into laboratory projects. *Science and Technology Review*. 12-21.

[9] I. Stowers, R. Komanduri, and E. Baird, "Review of precision surface generating processes and their potential application to the fabrication of large optical components," in *Proceedings of the SPIE*, 1988, pp. 62-73.

[10] J. B. Bryan, "Design and construction of an ultraprecision 84 inch diamond turning machine," *Precision Engineering,* vol. 1, pp. 13-17, 1979.

[11] J. Klingmann. (2001) The World's Most Accurate Lathe. *Science and Technology Review*. 12-14.

[12] G. Bylinsky. (2003, Closing In On Perfection Ultraprecision machine tools are putting manufacturers within nanometers of absolute accuracy http://money.cnn.com/magazines/fortune/fortune_archive/2003/06/23/344598/index.htm. *Fortune Magazine*.

[13] P. Van Zant, *Microchip fabrication*. New York, NY: McGraw-Hill, 2004.

[14] P. B. Zantye, A. Kumar, and A. K. Sikder, "Chemical mechanical planarization for microelectronics applications," *Materials Science and Engineering - R: Reports,* vol. 45, pp. 89-220, 2004.

[15] J. Steigerwald. (2009, *Chemical mechanical polish: the enabling technology*. Available: http://www.techdesignforums.com/eda/volumes/volume-6/march-2009/chemical-mechanical-polish-the-enabling-technology/

[16] S. Bukkapatnam, S. Kamarthi, Q. Huang, A. Zeid, and R. Komanduri, "Nanomanufacturing systems: opportunities for industrial engineers," *IIE Transactions,* vol. 44, pp. 492-495, 2012/07/01 2012.

[17] I. Inasaki and H. K. Tönshoff, "Fundamentals: Roles of Sensors in Manufacturing and Application Ranges," in *Sensors in Manufacturing*, H. K. Tönshoff and I. Inasaki, Eds., ed, 2002, pp. 1-6.

[18] T. Bibby and K. Holland, "Endpoint detection in CMP," *Journal of Electronic Materials,* vol. 27, pp. 1073-1081, 1998.

[19] R. Komanduri, "Annual Report to Micron Corporation: Review of Sensors in CMP," Oklahoma State University2005.

[20] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust monte carlo localization for mobile robots," *Artificial Intelligence,* vol. 128, pp. 91-141, 2001.

[21] N. Vlassis, B. Terwijn, and B. Krose, "Auxiliary particle filter robot localization from high-dimentional sensor observations," in *Proc. IEEE Int. Conf. Robot. Automat.*, Washington, DC, May 2002, pp. 7-12.

[22] C. Kwok, D. Fox, and M. Meila, "Adaptive real-time particle filters for robot localization," in *Proceedings of IEEE International Conference on Robotics & Automation*, Taipei-Taiwan, 2003.

[23] M. Montemerlo, S. Thrun, and W. Whittaker, "Conditional particle filters for simultaneous mobile robot localization and people-tracking Robotics and Automation," presented at the Proceedings. ICRA apos, 2002.

[24] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that Provably Converges," in *Proceedings of the International Joint Conference on Artificial intelligence (IJCAI)*, Acapulco, Mexico, August 2003.

[25] K. Rajaram and R. Jaikumar, "An interaction decision support system for on-line process control," *European Journal of Operational Research,* vol. 138, pp. 554-568, 2002.

[26] D. Zeidler, M. Plotner, and K. Drescher, "Endpoint detection method for CMP of copper," *Microelectronic Engineering,* vol. 50, pp. 411-416, 2000.

[27] P. Zantye, A. K. Sikder, N. Gulati, and A. Kumar, "Study of slurry selectivity and endpoint detection in Cu-CMP process," in *CMP-MIC Conference*, Santa Clara, CA, 2003.

[28] S.-Y. Kim, K.-J. Lee, and Y.-J. Seo, "In-situ end point detection of the STI-CMP process using a high selective slurry," *Microelectronic Engineering,* vol. 66, pp. 463-471, 2003.

[29] T. Bibby, J. A. Adams, and K. Holland, "Optical endpoint detection for chemical mechanical planarization," *Journal of Vaccum Science and Technology. B.,* vol. 17, pp. 2378-2384, 1999.

[30] T. K. Das, R. Ganesan, A. K. Sikder, and A. Kumar, "Online end point detection in CMP Using SPRT of wavelet decomposed sensor Data semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing,* vol. 18, pp. 440-447, 2005.

[31] Z. Kong, O. Beyca, R. Komanduri, and S. Bukkapatnam, "Nonlinear Sequential Bayesian Analysis-Based Decision Making for End-Point Detection of Chemical Mechanical Planarization (CMP) Processes," *Semiconductor Manufacturing, IEEE Transactions on,* pp. 1-1, 2011.

[32] D. Zeidler, M. Plotner, and K. Drescher, "Endpoint detection method for CMP of copper," *Microelectronic Engineering,* vol. 50, pp. 411-416, 2000.

[33] A. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, "Optimization of tribological properties of silicon dioxide during the chemical mechanical planarization process," *Journal of Electronic Materials,* vol. 30, pp. 1520-1526, 2001.

[34] D. E. Lee, J. Choi, and D. A. Dornfeld, "In-Situ Acoustic Emission Monitoring of Surface Chemical Reactions for Copper CMP," *UC Berkeley: Laboratory for Manufacturing and Sustainability,* 2005.

[35] G.-J. Wang, B.-S. Lin, and K. Chang, "In-situ neural network process controller for copper chemical mechanical polishing," *The International Journal of Advanced Manufacturing Technology,* vol. 32, pp. 42-54, 2007.

[36] H. Jeong, H. Kim, S. Lee, and D. Dornfeld, "Multi-Sensor Monitoring System in Chemical Mechanical Planarization (CMP) for Correlations with Process Issues," *UC Berkeley: Laboratory for Manufacturing and Sustainability,* 2006.

[37] U. Phatak, S. Bukkapatnam, Z. Kong, and R. Komanduri, "Sensor-based modeling of slurry chemistry effects on the material removal rate (MRR) in copper-CMP process," *International Journal of Machine Tools and Manufacture,* vol. 49, pp. 171-181, 2009.

[38] S. Bukkapatnam, P. Rao, and R. Komanduri, "Experimental dynamics characterization and monitoring of MRR in oxide chemical mechanical planarization (CMP) process," *International Journal of Machine Tools & Manufacture (in press),* 2008.

[39] Z. Kong, A. Oztekin, O. F. Beyca, U. Phatak, S. Bukkapatnam, and R. Komanduri, "Process Performance Prediction for Chemical Mechanical Planarization (CMP) by Integration of Nonlinear Bayesian Analysis and

Statistical Modeling," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 23, pp. 316-327, 2010.

[40] W. Hui, Z. Xi, A. Kumar, and H. Qiang, "Nonlinear Dynamics Modeling of Correlated Functional Process Variables for Condition Monitoring in Chemical-Mechanical Planarization," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 22, pp. 188-195, 2009.

[41] J. Tang, D. A. Dornfeld, S. K. Pangrle, and A. Dangca, "In-Process detection of microscratching during CMP using acoustic emission sensing technology," *Journal of Electronic Materials,* vol. 27, pp. 1099-1103, 1998.

[42] R. Ganesan, T. K. Das, A. K. Sikder, and A. Kumar, "Wavelet-Based identification of delamination defect in CMP (Cu-Low k) using nonstationary acoustic emission signal," *IEEE Transactions on Semiconductor Manufacturing,* vol. 16, pp. 677-685, 2003.

[43] A. Chandra, P. Karra, A. F. Bastawros, R. Biswas, P. J. Sherman, S. Armini, and D. A. Lucca, "Prediction of scratch generation in chemical mechanical planarization," *CIRP Annals - Manufacturing Technology,* vol. 57, pp. 559-562, 2008.

[44] C. Dennison, "Developing effective inspection systems and strategies for monitoring CMP process," *Micro magazine http://www.micromagazine.com/archive/98/02/dennison.html,* pp. 31-, 1998.

[45] A. Zareena and S. Veldhuis, "Tool wear mechanisms and tool life enhancement in ultra-precision machining of titanium," *Journal of Materials Processing Technology,* vol. 212, pp. 560-570, 2012.

[46] A. Diniz, J. Liu, and D. Dornfeld, "Correlating tool life, tool wear and surface roughness by monitoring acoustic emission in finish turning," *Wear,* vol. 152, pp. 395-407, 1992.

[47] D. E. Lee, I. Hwang, C. M. O. Valente, J. F. G. Oliveira, and D. A. Dornfeld, "Precision manufacturing process monitoring with acoustic emission," *International Journal of Machine Tools and Manufacture,* vol. 46, pp. 176-188, 2006.

[48] H. Yoshioka, M. Hayashi, and H. Shinno, "Status Monitoring of Ultraprecision Machining Using Micro Thermo Sensor and AE Sensor," *International Journal of Automation Technology,* vol. 3, pp. 422-427, 2009.

[49] J. Oh and S. Lee, "Prediction of surface roughness in magnetic abrasive finishing using acoustic emission and force sensor data fusion," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* vol. 225, pp. 853-865, 2011.

[50] A. Jazwinski, *Stochastic processes and filtering theory*. New York: Academic Press, 1970.

[51] M. S. Arulampalam, S. Maskel, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing,* vol. 50, pp. 174-188, 2002.

[52] J. Carpenter, P. Cliffird, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEEE Proceedings-Radar, Sonar Navigation,* vol. 146, pp. 1-7, 1999.

[53] R. S. Bucy and K. D. Senne, "Digital synthesis of non-linear filters," *Automatica,* vol. 7, pp. 287-298, 1971.

[54] H. W. Sorenson, "Recursive estimation for non-linear dynamic systems," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed., ed New York: Marcel Dekker, 1988, pp. 127-165.

[55] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing,* vol. 10, pp. 197-208, 2000.

[56] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. Nordlund, "Particle filters for positioning, navigation and tracking," *IEEE Trans. Signal Processing,* vol. 50, pp. 425-437, 2002.

[57] D. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech and Audio Processing,* vol. 11, pp. 826-836, 2003.

[58] V. Verma, S. Thrun, and R. Simmons, "Variable resolution particle filter," presented at the Proceedings of the International Joint Conference on Artificial Intelligence, August 2003.

[59] Z. Khan, T. Balch, and F. Dellaert, "A Rao-Blackwellized particle filter for eigentracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Washington, DC, 2004, pp. 980-986.

[60] G. A. Hazelrigg, *Systems engineering: an approach to information-based design* Prentice Hall, 1996.

[61] K. Lewis, W. Chen, and L. Schmidt, "Decision making in engineering design," *ASME,* 2006.

[62] S. Iassinovski, A. Artibab, and C. Fagnart, "A production rules-based tool for on-line simulation, decision making and discrete process control," *Engineering Application of Artificial Intelligence,* vol. 21, pp. 406-418, 2008.

[63] H. W. Chu and P. Tosirisuk, "Process decision program Chart from operations research to total quality control," *Computers industry engineering,* vol. 21, pp. 111-115, 1991.

[64] H. Gerking, "Modeling of multi-stage decision-making process in multi-period energy-models," *European Journal of Operation Research,* vol. 32, pp. 191-204, 1987.

[65] C. W. Chen, J. Luo, and K. J. Parker, "Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications," *Image Processing, IEEE Transactions on,* vol. 7, pp. 1673-1683, 1998.

[66] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?," *Database Theory—ICDT'99,* pp. 217-235, 1999.

[67] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE,* vol. 78, pp. 1464-1480, 1990.

[68] M. Medvedovic and S. Sivaganesan, "Bayesian infinite mixture model based clustering of gene expression profiles," *Bioinformatics,* vol. 18, pp. 1194-1206, 2002.

[69] A. R. Ferreira da Silva, "A Dirichlet process mixture model for brain MRI tissue classification," *Medical Image Analysis,* vol. 11, pp. 169-182, 2007.

[70] J. Zhang, Z. Ghahramani, and Y. Yang, "A probabilistic model for online document clustering with application to novelty detection," 2005.

[71] E. Sudderth, A. Torralba, W. Freeman, and A. Willsky, "Describing visual scenes using transformed dirichlet processes," *Advances in neural information processing systems,* vol. 18, p. 1297, 2006.

[72] A. Ahmed and E. P. Xing, *Dynamic non-parametric mixture models and the recurrent chinese restaurant process*: Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2007.

[73] A. Rodriguez, D. B. Dunson, and A. E. Gelfand, "The nested Dirichlet process," *Journal of the American statistical Association,* vol. 103, pp. 1131-1154, 2008.

[74] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE,* vol. 85, pp. 6-23, 1997.

[75] C. Cimander, M. Carlsson, and C. F. Mandenius, "Sensor fusion for on-line monitoring of yoghurt fermentation," *Journal of Biotechnology,* vol. 99, pp. 237-248, 2002.

[76] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters,* vol. 24, pp. 2115-2125, 2003.

[77] T. P. Banerjee and S. Das, "Multi-sensor data fusion using support vector machine for motor fault detection," *Information Sciences,* 2012.

[78] A. Okatan, C. Hajiyev, and U. Hajiyeva, "Fault detection in sensor information fusion Kalman filter," *AEU-International Journal of Electronics and Communications,* vol. 63, pp. 762-768, 2009.

[79] C. Aliustaoglu, H. M. Ertunc, and H. Ocak, "Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system," *Mechanical Systems and Signal Processing,* vol. 23, pp. 539-546, 2009.

[80] M. Grbovic, W. Li, P. Xu, A. K. Usadi, L. Song, and S. Vucetic, "Decentralized fault detection and diagnosis via sparse PCA based decomposition and Maximum Entropy decision fusion," *Journal of Process Control,* 2012.

[81] M. A. K. Jaradat and R. Langari, "A hybrid intelligent system for fault detection and sensor fusion," *Applied Soft Computing,* vol. 9, pp. 415-422, 2009.

[82] N. Subrahmanya, Y. C. Shin, and P. H. Meckl, "A Bayesian machine learning method for sensor selection and fusion with application to on-board fault diagnostics," *Mechanical Systems and Signal Processing,* vol. 24, pp. 182-192, 2010.

[83] Q. Cheng, P. K. Varshney, J. H. Michels, and C. M. Belcastro, "Distributed fault detection with correlated decision fusion," *Aerospace and Electronic Systems, IEEE Transactions on,* vol. 45, pp. 1448-1465, 2009.

[84] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory," *Information Fusion,* vol. 8, pp. 379-386, 2007.

[85] K. Salahshoor, M. Mosallaei, and M. Bayat, "Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended Kalman filter algorithm," *Measurement,* vol. 41, pp. 1059-1076, 2008.

[86] S. Tangjitsitcharoen and C. Rungruang, "In-Process Monitoring and Estimation of Tool Wear on CNC Turning by Applying Multi-Sensor with Back Propagation Technique," *Advanced Materials Research,* vol. 291, pp. 3036-3043, 2011.

[87] J. Shukla, "Predictability in the midst of chaos: A scientific basis for climate forecasting," *Science,* vol. 282, p. 728, 1998.

[88] T. Stathaki, *Image fusion: algorithms and applications*: Academic Press, 2011.

[89] S. K. Jayaweera, "Bayesian fusion performance and system optimization for distributed stochastic Gaussian signal detection under communication constraints," *Signal Processing, IEEE Transactions on,* vol. 55, pp. 1238-1250, 2007.

[90] P. J. Lucas, "Bayesian model-based diagnosis," *International Journal of Approximate Reasoning,* vol. 27, pp. 99-119, 2001.

[91] J. P. Wang, S. D. Lin, and Z. F. Bao, "Neural Network and DS Evidence Theory Based Condition Monitoring and Fault Diagnosis of Drilling," *Applied Mechanics and Materials,* vol. 249, pp. 481-486, 2013.

[92] C. J. Evans, E. Paul, D. A. Dornfeld, D. A. Lucca, G. Bryne, M. Tricard, F. Klocke, O. Dambon, and B. A. Mullany, "Material removal mechanisms in lapping and polishing," *Annals of the CIRP - Keynote Papers (G),* vol. 52, pp. 611-, 2003.

[93] R. Komanduri, "On material removal mechanisms in finishing of advanced ceramics and glasses," *Annals of the CIRP,* vol. 45, pp. 509-514, 1996.

[94] R. Komanduri, D. A. Lucca, and Y. Tani, "Technological advances in fine abrasives processes," *Annals of the CIRP,* vol. 46, pp. 545-596, 1997.

[95] M. R. Oliver, *Chemical-Mechanical Planarization of Semiconductor Materials*. New York, NY: Springer-Verlag, 2004.

[96] M. D. Escobar, "Estimating normal means with a Dirichlet process prior," *Journal of the American Statistical Association,* vol. 89, pp. 268-277, 1994.

[97] O. Abouelatta and J. Madl, "Surface roughness prediction based on cutting parameters and tool vibrations in turning operations," *Journal of Materials Processing Technology,* vol. 118, pp. 269-277, 2001.

[98] C. Beggan, M. Woulfe, P. Young, and G. Byrne, "Using acoustic emission to predict surface quality," *The International Journal of Advanced Manufacturing Technology,* vol. 15, pp. 737-742, 1999.

[99] M. Hayashi, H. Yoshioka, and H. Shinno, "An adaptive control of ultraprecision machining with an in-process micro-sensor," *Journal of Advanced Mechanical Design, Systems, and Manufacturing,* vol. 2, pp. 322-331, 2008.

[100] P. B. Zantye, A. Kumar, and A. K. Sikder, "Chemical mechanical planarization for microelectronics applications," *Materials Science and Engineering: R: Reports,* vol. 45, pp. 89-220, 2004.

[101] "Http://goliath.ecnext.com, High Tech Ceramics News," ed, 2004.

[102] M. I. Jordan, "*Why the logistic function? A tutorial discussion on probabilities and neural networks*, ," August 1995.

[103] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo in practice*, A. Doucet*, et al.*, Eds., ed New York: Springer-Verlag, 2001, pp. 197-223.

[104] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second ed. New York: McGraw-Hill, 1984.

[105] D. V. Lindley, *Making decisions*: John Wiley & Sons, 1985.

[106] R. Doering and Y. Nishi, *Handbook of semiconductor manufacturing technology*. Boca Raton, FL: CRC Press, 2007.

[107] J. M. Steigerwald, S. P. Murarka, and R. J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*: New York, John Wiley & Sons, 1997.

[108] J. Moyne, E. Del Castillo, and A. M. Hurwitz, *Run-to-run control in semiconductor manufacturing*. Boca Raton, FL: CRC Press, 2001.

[109] R. J. Gutmann, J. M. Steigerwald, L. You, D. T. Price, J. Neirynck, D. J. Duquette, and S. P. Murarka, "Chemical-mechanical polishing of copper with oxide and polymer interlevel dielectrics," *Thin Solid Films,* vol. 270, pp. 596-600, 1995.

[110] U. Paik and J. G. Park, *Nanoparticle engineering for chemical-mechanical planarization: fabrication of next-generation nanodevices*. Boca Raton, FL: CRC Press, 2009.

[111] M. Berman, T. Bibby, and A. Smith, "Review of *in-situ* & In-line Detection for chemical mechanical planarization Applications," *Semiconductor Fabtech,* vol. 8, pp. 267-274, 1998.

[112] H. Jeong, H. Kim, S. Lee, and D. Dornfeld, "Multi-Sensor Monitoring System in Chemical Mechanical Planarization for Correlations with Process Issues," *CIRP Annals - Manufacturing Technology,* vol. 55, pp. 325-328, 2006.

[113] J. Tang, D. Dornfeld, S. K. Pangrle, and A. Dangca, "In-process detection of microscratching during CMP using acoustic emission sensing technology," *Journal of Electronic Materials,* vol. 27, pp. 1099-1103, 1998.

[114] R. Ganesan, T. K. Das, A. K. Sikder, and A. Kumar, "Wavelet-based identification of delamination defect in CMP (Cu-low k) using nonstationary acoustic emission signal," *IEEE Transactions on Semiconductor Manufacturing,* vol. 16, pp. 677-685, 2003.

[115] H. Liang and D. Craven, *Tribology in Chemical-Mechanical Planarization*. Boca Raton , FL: CRC Press, Taylor and Francis Group, 2005.

[116] S. Runnels and L. M. Eyman, "Tribology analysis of chemical-mechanical polishing," *Journal of the Electrochemical Society,* vol. 14, pp. 1698-1701, 1994.

[117] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 1-38, 1977.

[118] C. E. Rasmussen, "The infinite Gaussian mixture model," *Advances in neural information processing systems,* vol. 12, p. 2, 2000.

[119] M. D. Escobar and M. West, "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association,* vol. 90, 1995.

[120] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the American statistical Association,* vol. 101, pp. 1566-1581, 2006.

[121] J. Pitman, *Combinatorial stochastic processes* vol. 1875: Springer-Verlag, 2006.

[122] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American statistical Association,* pp. 577-588, 1995.

[123] R. M. Neal, "Bayesian mixture modeling," 1992, pp. 197-211.

[124] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Pólya urn schemes," *The Annals of Statistics,* vol. 1, pp. 353-355, 1973.

[125] D. C. Montgomery, *Introduction to Statistical Quality Control*: John Wiley & Sons, 2008.

[126] D. M. Hawkins and D. H. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement*: Springer, 1998.

[127] H. Kantz and T. Schreiber, *Nonlinear time series analysis*. Cambridge: Cambridge University Press, 1997.

[128] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*: Prentice Hall, 1994.

[129] H. Guo, K. Paynabar, and J. Jin, "Multiscale monitoring of autocorrelated processes using wavelets analysis," *IIE Transactions,* vol. 44, pp. 312-326, 2012.

[130] P. Rao, B.Bhushan, S.Bukkapatnam, Z.Kong, S.Byalal, O.Beyca, A.Fields, and R.Komanduri, "Process-machine interaction (PMI) modeling and monitoring of chemical mechanical planarization (CMP) using wireless vibration sensors," *IEEE Transactions on Semiconductor Manufacturing (Accepted),* 2013.

[131] P. Rao, "Sensor-based monitoring and inspection of surface morphology in ultraprecision manufacturing processes," PhD Dissertation, Industrial Engineering and Management, Oklahoma State University, Stillwater, 2013.

[132] J. D. Morillo, T. Houghton, J. M. Bauer, R. Smith, and R. Shay, "Edge and bevel automated defect inspection for 300mm production wafers in manufacturing," in *Advanced Semiconductor Manufacturing Conference and Workshop, 2005 IEEE/SEMI*, 2005, pp. 49-52.

[133] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 603-619, 2002.

[134] J. Liu and D. Dornfeld, "Modeling and analysis of acoustic emission in diamond turning," *ASME-PUBLICATIONS-PED,* vol. 58, pp. 43-43, 1993.

[135] X. Chen, J. Tang, and D. Dornfeld, "Monitoring and analysis of ultraprecision metal cutting with acoustic emission," in *Proc. Int'l. Mech. Eng. Congress and Exposition, ASME*, 1996, pp. 387-393.

[136] T. J. Ko, D. W. Cho, and J. M. Lee, "Fuzzy pattern recognition for tool wear monitoring in diamond turning," *CIRP Annals-Manufacturing Technology,* vol. 41, pp. 125-128, 1992.

[137] J. Yan, K. Syoji, and J. i. Tamaki, "Some observations on the wear of diamond tools in ultra-precision cutting of single-crystal silicon," *Wear,* vol. 255, pp. 1380-1387, 2003.

[138] I.-H. Choi and J.-D. Kim, "Development of monitoring system on the diamond tool wear," *International Journal of Machine Tools and Manufacture,* vol. 39, pp. 505-515, 1999.

[139] Y. Lee, A. K. Chang, and D. A. Dornfeld, "Acoustic emission monitoring for the diamond machining of oxygen-free high-conductivity copper," *Journal of Materials Processing Technology,* vol. 127, pp. 199-205, 2002.

[140] C. Cheung and W. Lee, "A theoretical and experimental investigation of surface roughness formation in ultra-precision diamond turning,"

International Journal of Machine Tools and Manufacture, vol. 40, pp. 979-1002, 2000.

[141] G. Shafer, *A mathematical theory of evidence* vol. 1: Princeton university press Princeton, 1976.

[142] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics,* vol. 5, pp. 115-133, 1943.

APPENDIX

Matlab Codes

Matlab® Codes for each chapter are presented in the followings

Nonlinear Sequential Bayesian Analysis-Based Decision Making for Chemical Mechanical Planarization Process

Main function for particle filtering

```
function [x_hatarr
pararr]=pfmain1(data,N,state_space_model,order,h)
% x_hatarr is the predicted state
% pararr is the parameters updated for each time step

if nargin<5 , h=0.01; end %learning rate for parameter update
if nargin<4 , order=3; end % for logistic state space it is
number of lags and for polynomial state space it is order of the
polynomial
if nargin<3 , state_space_model='poly'; end %state space model
if nargin<2 , N=100; end %number of particles
R=0.01; %measurement noise
switch lower(state_space_model)
    case('poly')
        [x_hatarr pararr]=poly1(data,order,N,h,R);
    case('logistic')
        [x_hatarr pararr]=logistic(data,order,N,h,R);
end
```

Particle filtering with polynomial state space function

```
function [x_hatarr pararr]=poly1(data,order,N,h,R)

%this program runs particle filter with polynomial state space
%represantation

order=order+1;
h=0.05;
L=length(data);
x_hat=ones(N,1)*randn*sqrt(R);
x_hatarr=zeros(L,1);
```

```
par=randn(N,order);
pararr=zeros(L,order);
X=zeros(N,order);
for k=1:L-1
   for i=1:order
   X(:,i)=x_hat.^(i-1);
   end
   xpartmin=(sum((X.*par),2));
   ypart=xpartmin+sqrt(R)*randn(N,1);
   vhat=ones(N,1)*data(k)-ypart;
   q = (1 / sqrt(R) / sqrt(2*pi)) * exp(-vhat.^2 / 2 / R);
   %weight update
   qsum = sum(q);
   for i = 1 : N
        q(i) = q(i) / qsum;
   end
   %parameter update
   m=zeros(order,N);

   for i=1:order
   m(i,:)=sqrt(1-h^2)*par(:,i)+(1-sqrt(1-h^2))*sum(par(:,i).*q);
   end

   for i=1:N
    for j=1:order
        par(i,j)=randn(1)*h^2*var(par(:,j).*q*N)+m(j,i);
    end
   end
       % Resample.
    for i = 1 : N
        u = rand; % uniform random number between 0 and 1
        qtempsum = 0;
        for j = 1 : N
            qtempsum = qtempsum + q(j);
            if qtempsum >= u
                x_hat(i) = xpartmin(j);
                break;
            end
        end
    end
    %update
    x_hatarr(k+1)=mean(x_hat);
    for i=1:order
        pararr(k+1,i)=mean(par(:,i));
    end
end
```

Particle filtering with logistic state space function

```
function [x_hatarr pararr]=logistic(data,order,N,h,R)
%this program runs particle filter with logistic regression state
space
```

```
%represantation
order=order+1;
L=length(data);
x_hat=ones(N,order-1)*randn*sqrt(R);
x_hatarr=zeros(L,1);
par=randn(N,order);
pararr=zeros(L,order);
X=zeros(N,order);
for k=1:L-1
   for i=1:order-1
       X(:,i)=[1./(1+exp(-x_hat(:,i)))];
   end
   X(:,order)=ones(N,1);
   xpartmin=(sum((X.*par),2));
   ypart=xpartmin+sqrt(R)*randn(N,1);
   vhat=ones(N,1)*data(k)-ypart;
   q = (1 / sqrt(R) / sqrt(2*pi)) * exp(-vhat.^2 / 2 / R);
   %weight update
   qsum = sum(q);
   for i = 1 : N
        q(i) = q(i) / qsum;
   end
   %parameter update
   m=zeros(order,N);

   for i=1:order
   m(i,:)=sqrt(1-h^2)*par(:,i)+(1-sqrt(1-h^2))*sum(par(:,i).*q);
   end

   for i=1:N
    for j=1:order
        par(i,j)=randn(1)*h^2*var(par(:,j).*q*N)+m(j,i);
    end
   end
   if order-1>1
        for i=1:order-1
            x_hat(:,i+1)=x_hat(:,i);
        end
   end

       % Resample.
    for i = 1 : N
        u = rand; % uniform random number between 0 and 1
        qtempsum = 0;
        for j = 1 : N
            qtempsum = qtempsum + q(j);
            if qtempsum >= u
                x_hat(i,1) = xpartmin(j);
                break;
            end
        end
    end
    %update
```

```
    x_hatarr(k+1)=mean(x_hat(:,1));
    for i=1:order
        pararr(k+1,i)=mean(par(:,i));
    end
    k
end
```

Chemical Mechanical Planarization (CMP) process monitoring by using evolutionary clustering analysis

Recurrent nested Dirichlet process modeling demo for 1D

```
% demo for cluster change
clear all
clc
close all

initpath        %initializing the file path
%constructing training data
training_data = [randn(1,150) 3+randn(1,150)];
ii=randperm(length(training_data));
training_data=training_data(ii);

%constructing testing data
testing_data = [randn(1,100) 3+randn(1,100) -3+randn(1,100)];
ii=randperm(length(testing_data));
testing_data=testing_data(ii);

%parameters
dat=[training_data testing_data];
dd=1;                     %dimension of data set
width=100;                %length of sliding windows
overlap=10;               %length of sliding data set

number_of_windows_training=length(1:overlap:length(training_data)
-width+1);

%extracting features by RDP
[features dpmcell]=dpm_rdp(dat,dd,width,overlap);

%Applying RNDP
dpm=dpm_mixture(features(:,1:number_of_windows_training),size(fea
tures,1),1);

%Calculating Likelihood function for monitoring
ml=zeros(size(features,2),1);
mpi=dirrnd(dpm.nn);
for i=1:dpm.KK
    [mu sigma] = map(dpm.qq{i}
);
    ml=ml+mpi(i)* mvnpdf(features',mu',sigma);
```

```
end

plot(ml,'LineWidth',3)
hold on
plot([number_of_windows_training number_of_windows_training],[0
1.5*max(ml)],'r','LineWidth',3)
grid on
xlabel('time epoch')
ylabel('likelihood value')
%%
% enter time epoch to draw mixture of gaussian
t = 30;
DPM=dpmcell{t};

mpi= dirrnd(DPM.nn);
tt=min(DPM.xx)-1:0.05:max(DPM.xx)+1;
pp=zeros(length(tt),1);
figure;
hold on
for i=1:DPM.KK
    [mu sigma] = map(DPM.qq{i});
    pc=mpi(i)*mvnpdf(tt',mu,sigma);
    pp=pp+pc;
    plot(tt,pc,'k','LineWidth',2)
end

plot(tt,pp,'r','LineWidth',3)
plot(DPM.xx,zeros(width,1),'o','MarkerSize',12)
```

Recurrent nested Dirichlet process demo for 2D

```
% demo for cluster change
clear all
clc
close all
%constructing training data
initpath        %initializing the file path

training_data = [randn(2,150) 3+randn(2,150)];
ii=randperm(length(training_data));
training_data=training_data(:,ii);

%constructing testing data
testing_data = [randn(2,100) 3+randn(2,100) -3+randn(2,100)];
ii=randperm(length(testing_data));
testing_data=testing_data(:,ii);

%parameters
dat=[training_data testing_data];
dd=2;
```

```
width=100;
overlap=10;

number_of_windows_training=length(1:overlap:length(training_data)
-width+1);

%extracting features by RDP
[features dpmcell]=dpm_rdp(dat,dd,width,overlap);

%Applying RNDP
dpm=dpm_mixture(features(:,1:number_of_windows_training),size(fea
tures,1),1);

%Calculating Likelihood function for monitoring
ml=zeros(size(features,2),1);
mpi=dirrnd(dpm.nn);
for i=1:dpm.KK
    [mu sigma] = map(dpm.qq{i});
    ml=ml+mpi(i)* mvnpdf(features',mu',sigma);
end

plot(ml,'LineWidth',3)
hold on

plot([number_of_windows_training number_of_windows_training],[0
1.5*max(ml)],'r','LineWidth',3)
grid on
xlabel('time epoch')
ylabel('likelihood value')
%%
% enter time epoch to draw mixture of gaussian
t = 30;
DPM=dpmcell{t};

figure;
hold on
C={'b','r','g','k','m','c','y'};
for i=1:width
    plot(DPM.xx(1,i),DPM.xx(2,i),'*','Color',C{DPM.zz(i)})
end
```

Recurrent Dirichlet Process

```
function [ss, dpmcell]=dpm_rdp(dat,dd,width,overlap)
% Dividing data into sliding windows and applying RDP for each
sliding
% window
% inputs:
% dat : data set of row vectors
```

```
% dd  : dimension of data set
% width : window length
% overlap : length of sliding data points
% outputs:
% ss: feature set

%sliding windows
t=1:overlap:length(dat)-width+1;

%number of sliding windows
number_of_windows=length(t);

%initial window for DP
data=dat(:,1:width);

% initiliaze DP mixture
dpm=dpm_mixture(data,dd,1);

%clearing empty clusters
dpm=clearempty(dpm);

%clearing small clusters
dpm=mclean(dpm,data);

%parameters will be kept in this cell
MM=cell(number_of_windows,1);
dpmcell=cell(number_of_windows,1);
%extracting parameters (mean and variance)
mm=[];
for k=1:length(dpm.nn)
    [mu,sigma] = map(dpm.qq{k});
    sigma=uptri(sigma);
    mm=[mm;mu;sigma];
end
MM{1}=mm;

for i=1:number_of_windows
    data=dat(:,overlap*(i-1)+1:overlap*(i-1)+width);

    %keeping cluster label information for overlapped data
    dpm.zz(1:end-overlap)=dpm.zz(overlap+1:end);

    %randomly assigning new data points to existing clusters
    dpm.zz(end-overlap+1:end) =
ceil(length(dpm.nn)*rand(1,overlap));
    dpm.nn=zeros(1,length(dpm.nn));

    dpm.xx=data;

    %updating cluster information
    for j=1:width
```

```
        kk=dpm.zz(j);
        dpm.qq{kk} = additem(dpm.qq{kk},data(:,j)); % add
sufficient stats of data item
        dpm.nn(1,kk)=dpm.nn(1,kk)+1;
    end

    %gibbs sampling for rdp
    for iter=1:10
        dpm=dpm_gibbs2(dpm,1,data);

        %clearing empty clusters
        dpm=clearempty(dpm);
    end
    %clearing small clusters
    dpm=mclean(dpm,data);

    dpmcell{i}=dpm;

    mm=[];
    for tt=1:length(dpm.nn)
        [mu,sigma] = map(dpm.qq{tt});
        sigma=uptri(sigma);
        mm=[mm;mu;sigma];
    end
    %matching the similar clusters
    mm=similarity(mm,MM{1},dd);

    MM{i}=mm;
    fprintf('%d/%d segments has
completed\n',i,number_of_windows);
end

% finding the length of features for each sliding window
len=zeros(length(MM),1);
for i=1:length(MM)
    len(i)=length(MM{i});
end

% length of feature for each cluster
l=dd+(dd*(dd+1)/2);
% max number of cluster
cnum=max(len)/l;

% initializing feature set with mean 0 and std 1
mix=zeros(l,1);
mix(dd+1:dd+dd)=1;

% assigning feature set
ss=repmat(mix,cnum,length(MM));
for i=1:length(MM)
    ss(1:length(MM{i}),i)=MM{i};
```

```
end
```

Dirichlet Process

```
function dpm=dpm_mixture(data,dd,aa)
% demo of DP mixture model
% inputs:
% data: row vector of data set to cluster
% dd: dimension of data set
% aa: concentration parameter

KK = 1;
[MM NN] = size(data);

s0 = 3;
ss = 1;
numiter = 50;

hh.dd = dd;
hh.ss = s0^2/ss^2;
hh.vv = 30;
hh.VV = ss^2*eye(dd);
hh.uu = zeros(dd,1);



yy=data;

xx = num2cell(yy,1);

% initialize component assignment
zz = ceil(rand(1,NN)*KK);

% initialize DP mixture
dpm = dpm_init(KK,aa,Gaussian(hh),xx,zz);

% initialize records
record.KK = zeros(1,numiter);

%gibbs sampling for inference
for iter = 1:numiter

    dpm = dpm_gibbs(dpm,1);
%     fprintf('iter number %d\n',iter);
end
```

Initializing DP mixture

```
function dpm = dpm_init(KK,aa,q0,xx,zz);
% initialize DP mixture model, with
% KK active mixture components,
```

```
% aa concentration parameter,
% q0 empty component with hh prior,
% xx data, x_i=xx{i}
% zz initial cluster assignments (between 1 and KK).

dpm.KK = KK;
dpm.NN = length(xx);
dpm.aa = aa;
dpm.qq = cell(1,KK+1);
dpm.xx = xx;
dpm.zz = zz;
dpm.nn = zeros(1,KK);

% initialize mixture components
% component KK+1 takes care of all inactive components
for kk = 1:KK+1,
  dpm.qq{kk} = q0;
end

% add data items into mixture components
for ii = 1:dpm.NN
  kk = zz(ii);
  dpm.qq{kk} = additem(dpm.qq{kk},xx{ii});
  dpm.nn(kk) = dpm.nn(kk) + 1;
end
```

Gibbs sampling for estimating parameters of DP mixture

```
function dpm = dpm_gibbs(dpm,numiter);
% run numiter number of iterations of gibbs sampling in the DP
mixture

KK = dpm.KK; % number of active clusters
NN = dpm.NN; % number of data items
aa = dpm.aa; % alpha parameter
qq = dpm.qq; % row cell vector of mixture components
xx = dpm.xx; % row cell vector of data items
zz = dpm.zz; % row vector of cluster indicator variables
nn = dpm.nn; % row vector of number of data items per cluster




for iter = 1:numiter
  % in each iteration, remove each data item from model, then add
it back in
  % according to the conditional probabilities.

  for ii = 1:NN % iterate over data items ii

    % remove data item xx{ii} from component qq{kk}
```

```
    kk = zz(ii); % kk is current component that data item ii
belongs to
    nn(kk) = nn(kk) - 1; % subtract from number of data items in
component kk
    qq{kk} = delitem(qq{kk},xx{ii}); % subtract data item
sufficient statistics

    % delete active component if it has become empty
    if nn(kk) == 0,
      %fprintf(1,'del component %3d. K=%3d\n',find(nn==0),KK-
sum(nn==0));
      KK = KK - 1;
      qq(kk) = [];
      nn(kk) = [];
      idx = find(zz>kk);
      zz(idx) = zz(idx) - 1;
    end

    % compute conditional probabilities pp(kk) of data item ii
    % belonging to each component kk
    % compute probabilities in log domain, then exponential
    pp = log([nn aa]);
    for kk = 1:KK+1
      pp(kk) = pp(kk) + logpredictive(qq{kk},xx{ii});
    end
    pp = exp(pp - max(pp)); % -max(p) for numerical stability
    pp = pp / sum(pp);

    % choose component kk by sampling from conditional
probabilities
    uu = rand;
    kk = 1+sum(uu>cumsum(pp));

    % instantiates a new active component if needed
    if kk == KK+1
      %fprintf(1,'add component %3d. K=%3d\n',kk,KK+1);
      KK = KK + 1;
      nn(kk) = 0;
      qq(kk+1) = qq(kk);
    end

    % add data item xx{ii} back into model (component qq{kk})
    zz(1,ii) = kk;
    nn(1,kk) = nn(1,kk) + 1; % increment number of data items in
component kk
    qq{1,kk} = additem(qq{1,kk},xx{ii}); % add sufficient stats
of data item

  end
end

% save variables into dpm struct
```

```
dpm.qq = qq;
dpm.zz = zz;
dpm.nn = nn;
dpm.KK = KK;
```

Process Monitoring of Ultra Precision Machining (UPM) based on Sensor Fusion

Sensor fusion for UPM experiments

```
close all
clear all
clc
initpath

load upm_all_conditions

num_class=36;
sensor_id=[1:6];
sensor_name={'Vx','Vy','Vz','Fx','Fy','Fz'};
num_sensor=length(sensor_id);
[comb n_comb] = combnts(num_sensor);


for i=1:num_class
    data{i}=data{i}(:,sensor_id);
end

features=cell(num_class,num_sensor);
ar_dim=3;
dim=4+ar_dim;
w=100;
t=1:w:length(data{1});

N=length(t);

for j=1:num_class
    for k=1:num_sensor
        features{j,k}=zeros(N,dim);
        temp_data=data{j}(:,k);
        for i=1:N
            ar=arburg(temp_data((i-1)*w+1:i*w,1),ar_dim);
            features{j,k}(i,:)=[mean(temp_data((i-1)*w+1:i*w,1))
std(temp_data((i-1)*w+1:i*w,1)) skewness(temp_data((i-
1)*w+1:i*w,1)) kurtosis(temp_data((i-1)*w+1:i*w,1)) ar(2:end)];
        end
    end
end
%%
acc_mat=zeros(n_comb,10);
group=[];
```

```
temp_data=zeros(num_class*N,6*dim);
for j=1:num_class
    temp_data((j-1)*N+1:j*N,:)=[features{j,1} features{j,2}
features{j,3} features{j,4} features{j,5} features{j,6}];
end

stdr = std(temp_data);
sr = temp_data./repmat(stdr,length(temp_data),1);
[coefs,scores,variances,t2] = princomp(sr);

num_sensor=5;
dim=1;
[comb n_comb] = combnts(num_sensor);

y = linspace(1,num_class*N,11);
y=ceil(y);

for iter=1:1
    for i=1:num_class
        group=[group;i*ones(N,1)];
    end
    ii=randperm(num_class*N);
    group=group(ii);


%      tes= y(iter)
    tra=ceil(num_class*N*0.7);
    tes=num_class*N-tra;

    training_group = group(1:tra,:);
    testing_group = group(tra+1:end,:);

    class=zeros(tes,num_sensor);
    p=cell(num_sensor,1);
    acc=zeros(num_sensor,1);
    for i=1:num_sensor
        temp_data=scores(:,i);
        temp_data=temp_data(ii,:);
        training_data = temp_data(1:tra,:);
        testing_data = temp_data(tra+1:end,:);
        [class(:,i),
p{i}]=dir_class(training_data,testing_data,dim,num_class,training
_group);
    end


    for i=1:num_sensor
        p{i}=p{i}./repmat(sum(p{i},2),1,num_class);
    end
```

```
    accu=zeros(n_comb,1);
    names=cell(n_comb,1);
    count=1;

    for i=1:num_sensor
        for j=1:size(comb{i},1)
            pp=1;
            for k=1:i
                pp=pp.*p{comb{i}(j,k)};

names{count}=horzcat(names{count},sensor_name{comb{i}(j,k)});
            end
            [C, classf] = max(pp,[],2);
            accu(count)=sum((classf-testing_group)==0)/tes;
            count=count+1;

        end
    end
    acc_mat(:,iter)=accu;
    iter
end

[C ii]=sort(accu,'descend');
for i=1:5
fprintf('%-10s\t%.2f\n',names{ii(i)},accu(ii(i)))
end
```

Dirichlet process modeling for classification

```
function [class,
p]=dir_class(training_data,testing_data,dim,num_class,group)
%{
class : class assignment
training_data  :  Training data for classification (row vector)
testing_data  :  Testing data for validation (row vector)
dim  :  Dimension of data
num_class  :  Number of classes
%}

dpm_cell=cell(num_class,1);

for i=1:num_class
    ii=find(group==i);
    data=training_data(ii,:);
    dpm_cell{i} = dpm_mixture(data',dim,1);
end

p=zeros(length(testing_data),num_class);

for i=1:num_class
    dpm=dpm_cell{i};
    mpi= dirrnd(dpm.nn);
    for j=1:dpm.KK
```

```
        [mu sigma]=map(dpm.qq{j});
        p(:,i)=p(:,i)+mpi(j)*mvnpdf(testing_data,mu',sigma);
    end
end

[C class] = max(p,[],2);
```

VITA

Omer Faruk Beyca

Candidate for the Degree of

Doctor of Philosophy

Thesis: SENSOR BASED REAL-TIME PROCESS MONITORING FOR ULTRA-PRECISION MANUFACTURING PROCESSES WITH NON-LINEARITY AND NON-STATIONARITY

Major Field: Industrial Engineering and Management

Biographical: Born on 21st February 1984, Batman, Turkey

Education:
Completed the requirements for the Doctor of Philosophy in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in July, 2013.

Completed the requirements for the Bachelor of Science in Industrial Engineering at Fatih University, Istanbul, Turkey in 2007.

Experience:
Research Assistant, Oklahoma State University, 2007-2013
Teaching Assistant, Oklahoma State University, 2011, 2013

Professional Memberships:
American Society for Quality (ASQ)
Institute for Operations Research and Management Science (INFORMS)