Multivariable OES Data Analysis for Plasma Semiconductor Etching Process

by

Jie Yang
M.Sc., B.Sc.

A Dissertation submitted in fulfilment of the requirements for the award of Doctor of Philosophy (Ph.D.)

Dublin City University

School of Electronic Engineering

Supervisors: Dr. Stephen Daniels, Dr. Conor McArdle

September 2014

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____

ID No.: _____

Date: _____

To my dear parents and my wonderful wife

献给我敬爱的父母和美丽的妻子

List of Contents

Abstract

The semiconductor industry has played a crucial role in societal development over the past several decades. Plasma etching is a key processing step employed in Integrated Circuit (IC) fabrication. In order to improve product yield, Optical Emission Spectroscopy (OES) is widely used to monitor the etching process. OES generates high-dimensional data, which has a large information capacity but also has significant information redundancy. Based on plasma OES characteristics, two novel data analysis methods are proposed in this thesis: the Internal Information Redundancy Reduction (IIRR) method for dimension and redundancy reduction and Similarity Ratio Analysis (SRA) for fault detection. By identifying peak wavelength emissions and the correlative relationships between them, IIRR outputs a subset of the original variables. Data dimensionality is reduced significantly by IIRR with minimal information loss. The SRA method is intended for early-stage fault detection in plasma etching processes using real-time OES data as input. The SRA method can help to realise a highly precise control system by detecting abnormal etch-rate faults in real-time during an etching process, so less energy and materials will be wasted by faulty processing. Generally, previous research on OES measurements of plasma etching has largely focused on particular target applications and has used methods that rely on transforming the original data into an abstract variable space. In contrast, our approach operates directly in the original variable space allowing a more direct and easier interpretation of the dimension reduced data.

Acknowledgement

I would like to express my immense gratitude to my two supervisors, Dr. Stephen Daniels and Dr. Conor McArdle. You have given me incredible mentorship and support, which will be useful for all of my life.

My sincere gratitude goes to all people in the Energy and Design Lab, Paula Meehan, Asmaa Eltayeb, Shane Phelan, Dr. Shane Linnane, Will Ikpekha and Ahmed Almousaw. You helped me a lot with my research, as well as my life in Dublin. Especially, Paula and Shane, you are my first Irish friends. I always enjoy the drinks we had together.

Also I sincerely thank Irish Research Council and Intel Ireland for their financial support, without which I could not fully focus on my research in the past three years. Particularly, a big thank to Paul Sheehy and Dr. Niall MacGearailt from Intel Ireland.

Finally, I would like to thank my family: my mother Xiuyun Wang, my father Bangshe Yang and my wife Shasha Li. Your love is the most wonderful thing I have. I dedicate this thesis to you.

List of Peer Reviewed Publications

- Jie Yang, Conor McArdle, Stephen Daniels. Similarity Ratio Analysis for Early Stage Fault Detection with Optical Emission Spectrometer in Plasma Etching Process. PLoS ONE 9(4): e95679. doi:10.1371/journal.pone.0095679.
- Jie Yang, Conor McArdle, Stephen Daniels. Dimension Reduction of Multivariable Optical Emission Spectrometer Dataset in Plasma Etching Process. Sensors. 14(1), 52-67, 2014.
- Jie Yang, Shane Phelan, Paula Meehan, Stephen Daniels. A Distributed Real Time Sensor Network for Enhancing Energy Efficiency through ICT. In Symposium on ICT and Energy Efficiency and Workshop on Information Theory and Security (CIICT 2012), pages 8–14, Stevenage, UK, July 2012, doi:10.1049/cp.2012.1854. Institution of Engineering and Technology.

List of Contributions

- Jie Yang, Conor McArdle, Stephen Daniels. Similarity Analysis of OES Data for Chamber Matching and Fault Detection. Abstract & Presentation in Intel Ireland Research Conference. 2013.
- Jie Yang, Conor McArdle, Stephen Daniels. Design and Development of Integrated MatLab Developing Environment. Research Project Cooperated with Intel Ireland. 2013.
- Jie Yang, Conor McArdle, Stephen Daniels. A Virtual Metrology Model with SPCA for Etch Rate Estimation. Abstract in Intel Ireland Research Conference. 2012.
- Jie Yang, Conor McArdle, Stephen Daniels. Optical Emission Spectrum Monitoring & Etch Rate Prediction in a Plasma Etch Process. Poster in Intel Ireland Research Conference. 2012.
- Jie Yang, Shane Phelan, Paula Meehan, Stephen Daniels. R&D of a Real Time Wireless Sensor Network for Wind Monitoring. Research Project cooperated with DCU Estates. 2011-2012.
- Jie Yang, Conor McArdle, Stephen Daniels. R&D of an Energy Monitoring & Management System with ICT Architecture. Abstract & Poster in Intel Ireland Research Conference. 2011.

List of Figures

List of Tables

List of Abbreviations

APS	Absolute Peak Selection
BDER	Bidirectional Elimination Regression
BER	Backward Elimination Regression
BLW	Bottom Linear Width
CCD	Charged Coupled Device
CFA	Confirmatory Factor Analysis
DC	Direct Current
EFA	Exploratory Factor Analysis
EM	Expectation Maximization
FA	Factor Analysis
FD	Fault Detection
FSCA	Forward Selection Component Analysis
FSR	Forward Selection Regression
IC	Integrated Circuit
ICA	Independent Component Analysis
IIRR	Internal Information Redundancy Reduction
IRP	Iterative Ranking Process
ITRS	International Technology Roadmap for Semiconductor
MAPE	Mean Absolute Percentage Error
MDR	Mean Determination Ratio
MF	Microwave Frequency
MLR	Multiple Linear Regression
MS	Mass Spectrometer
NTS	Normalized Time Scale
OES	Optical Emission Spectrometer
OPS	Optimized Peak Selection
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PIM	Plasma Impedance Monitor
PLS	Partial Least Square
PMF	Probability Mass Function
RF	Radio Frequency
SA	Similarity Analysis
SEM	Scanning Electron Microscopy
SPCA	Sparse Principal Component Analysis
SR	Similarity Ratio
SRA	Similarity Ratio Analysis
SVM	Support Vector Machine
VM	Virtual Metrology

CHAPTER 1. Introduction

1.1 Research Motivation

The semiconductor industry has played a crucial role in economic and societal development over the past several decades, since the invention of the Integrated Circuit (IC). The plasma etching process is one of the most important processes for IC fabrication. In order to understand and properly control the plasma etching process, plasma monitoring methods are essential. Optical Emission Spectroscopy (OES) is one of the most popular and powerful plasma monitoring technologies and is widely used in IC manufacturing, however, OES data is normally very high in dimensionality and volume, and data analysis and processing methods are required. More effective OES data processing methods could provide significant benefits for prediction of crucial system parameters, system condition monitoring, amongst other etch process control applications. Two particular OES data related topics are studied in this thesis: dimensionality reduction and Fault Detection (FD).

1.2 Plasma Etching

The plasma etching procedure is one of the most important steps in IC Manufacturing. A typical chamber for plasma etching in IC fabrication is presented in Figure 1-1. Silicon wafers are placed in the chamber. A chemical gas is pumped into the chamber, and then plasma is generated from the gas by supplying microwave energy. After that, plasma is accelerated toward wafers by the Radio Frequency (RF) source. Finally, the surfaces of wafers are etched by interaction with the accelerated plasma. Of course, a real plasma etching process is much more complex than this abstract model. Typically,

a multistep etching process is used in a practical IC manufacturing procedure, in which a multiple-layer IC structure is realized.

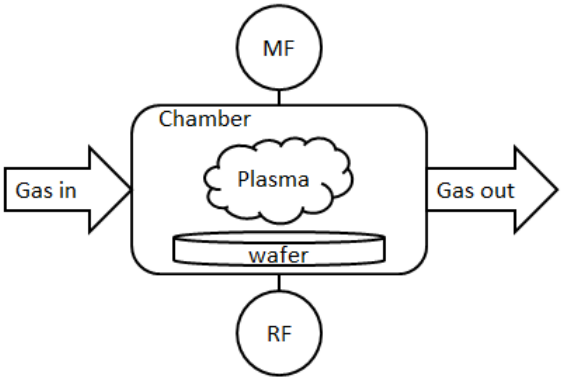

Figure 1-1. Structure of a plasma etching chamber. MF =Microwave Frequency; RF = Radio Frequency.

The ideal etching process is demonstrated in Figure 1-2(b), while Figure 1-2(a) shows the wafer before etching. The ideal etch should be anisotropic (vertical) whereby the wafer surface under the mask remains and all of other surface is removed by the etching. But practically, the plasma etching process suffers certain problems. Figure 1-2(c) and Figure 1-2(d) show two undesirable etching results. These scenarios are examples of problems which restrict the development of IC devices towards much smaller critical dimensions with high aspect ratios. Christophe Cardinaud, Marie-Claude Peignon, and Pierre-Yves Tessier summarised all of these problems from four aspects [1]: selection of the mask and substrate, profile control of the pattern, damage to the material, and etching rate control. They also concluded that the choice of the plasma chemistry and chemical species is an important factor for anisotropic etching, as well as surface temperature, pressure and so on. In other research [2], Thomas F. Edgar and his colleagues mentioned that the etch process is often operated empirically, while there are hundreds of individual operations which impact on the final result. Unfortunately, there is only little understanding of the underling physics and chemistry

in industrial IC fabrication and all of these problems are big challenges for research in plasma etching.

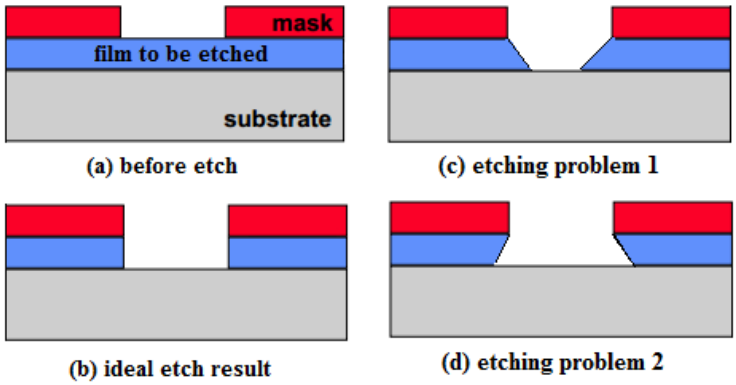

Figure 1-2. Ideal etching process & two etching problems

1.3 Optical Emission Spectrometer

OES is one of the commonly used plasma diagnostic technologies. Physically, a chemical element or chemical compound demonstrates optical emission, when it makes a transition from a high energy state to a lower one. It is an important physical characteristic that different chemical species have different emission spectra. By observing the optical emission spectrum, OES can help to identify chemical species and monitor the changes of their ratios compared to the entire plasma in the chamber. Figure 1-3 shows OES sensor deployment in a plasma etching chamber. It measures the optical emission of plasma through a window in the chamber. OES provides an indirect measurement of the plasma process.

A typical set of time-series OES measurements for a complete etching process is presented in Figure 1-4. It provides intensity measurements of 2048 wavelengths from 200 nm to 1100 nm. High dimensionality and information richness can be observed here.

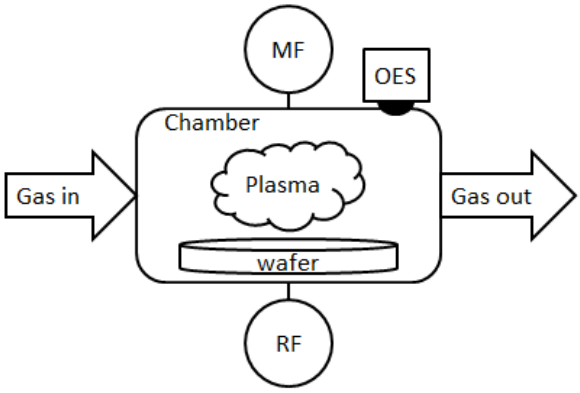

Figure 1-3. Deployment of OES

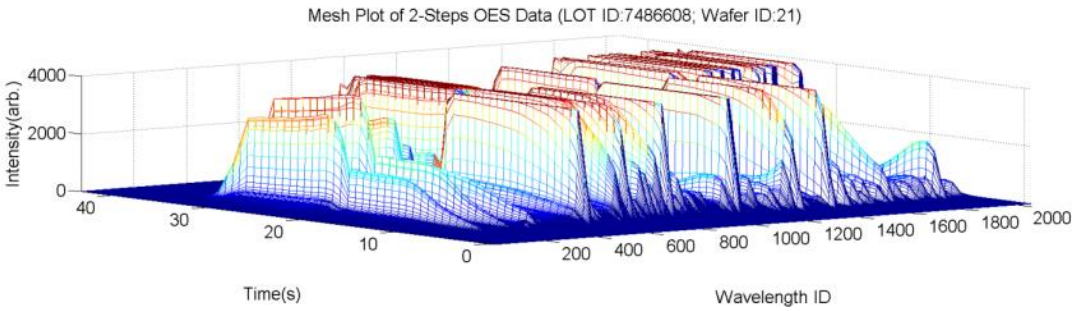

Figure 1-4. A sample of OES measurement in time series for an etching process

1.4 Research Topics

Due to the abundant information of OES and the immediate relationship with the etching process, OES is widely applied and studied in IC fabrication. Two research topics are addressed in this thesis: dimensionality reduction and fault detection.

1.4.1 Dimensionality Reduction

High dimensionality is the main research challenge for most OES-based data analysis. Such high dimensionality provides a huge information capacity, which is the main strength compared with the other sensor data. In the meanwhile, problems are also raised [3], such as useful information being sparse in a high dimension space. This can weaken the statistical significance of the information. From a physical perspective, this problem is caused by the fact that many different chemical species are included in plasma etching and each chemical species has multiple possible spectral emissions.

The performance of any application based on statistical inference from spectral data can be compromised consequently. By reducing dimensionality, a number of benefits can be achieved. Firstly, the computational efficiency of applications which take OES data as input is increased. Calculation accuracy is also improved with smaller dimension by excluding irrelevant information (e.g. noise). Secondly, important variables can be more easily identified from a smaller dimension dataset. It helps to understand interesting phenomena behind these variables [4]. In the context of an OES monitoring system in plasma etching, important variables can be specific wavelengths. It is valuable to find underling information on the chemical species, which causes system variance. This important usage is described as feature extraction [5].

1.4.2 Fault Detection

IC manufacturing has two major development trends. Firstly, more and more transistors can be built per wafer [6]. Secondly, larger diameter wafers will be used to increase the IC yield. Compared with current 300 mm diameter wafers, 450 mm diameter wafer technology is proposed as a main-stream product in the future [7]. This development requires that IC fabrication needs more precise process control mechanisms year by year, especially in fault detection. As one of the most important steps in IC fabrication, plasma etching impacts the quality of the final product output significantly. By detecting system faults in the etching process, four major advantages for the overall IC fabrication process are realised [8]: improvement of process quality, decrease of equipment downtime, improvement of wafer quality, and less usage of testing wafers.

1.5 Research Contribution

In this thesis, two methods are proposed for dimensionality reduction and fault detection (FD), evaluated separately with OES dataset from two IC manufactories. Discussion of the proposed methods and the corresponding contributions are presented in the remainder of this section.

1.5.1 Dimensionality Reduction

Dimension reduction has matured as a research area in the last several decades. It is a common topic across multiple disciplines, such as statistics, visualisation and information theory [9]. However, most of the research on the topic is based on general mathematical models, which ignore particular features of optical emission, such as the significance of peak wavelengths. They also have difficulty in finding important variables. Therefore, the Internal Information Redundancy Reduction (IIRR) method is proposed for OES dimension and redundancy reduction. IIRR operates directly in the original variable space, identifying peak wavelength emissions and the correlative relationships between them. Instead of creating new variables like the other methods (e.g. Principal Component Analysis (PCA)), IIRR picks up a subset of the original variables. So it is easier to find important variables in the original data. This set of variables has a much smaller size than the original data, but there is no significant information loss in the new variable set. This is proved by uncompromised prediction accuracy of etch rate using the IIRR output variable subset. In the future, IIRR can be easily extended to satisfy more detailed requirements. For example, further data dimension reduction can target certain system variables after processing data with IIRR, such as etch rate or etch line width.

1.5.2 Similarity Ratio Analysis for Fault Detection

Traditional FD technologies have two common problems: high computational cost and long-time delay. Because of requirements for real-time monitoring and non-intrusive sensing, the OES dataset is widely used for FD as a potential replacement for traditional FD technologies. However, most current research can only realize FD after the etching process completes. Some existing methodologies are able to detect faults at an early-stage with a single spectral scan, but they are limited to using only a small fixed set of time points.

In order to solve these research gaps, the Similarity Ratio Analysis (SRA) for Fault Detection method is proposed. Compared with previous methods, SRA can detect faults at an early stage of the etching process. In this way, less energy and materials will be wasted by faulty processing. The spectrum scans are also not limited to certain time points, so the FD system can be more flexible because the timetable for the OES sampling can be adjusted based on real-time system feedback. According to the result of the demonstrating dataset, this method gives an alarm for a system fault at about the 6^{th} second in a 45 second faulty etching process.

1.6 Thesis Structure

Chapter 2 provides a background survey on our research topics. Firstly, the plasma etching process is introduced. Secondly, popular monitoring technologies are presented, including their basic mechanisms and underlining theories. Thirdly, comprehensive literature reviews are grouped and presented based on two different classifications of strategies. Relevant papers are classified and discussed based on their algorithm types initially and then based on their research targets. Finally, two real

experimental industrial datasets are discussed. These two manufacturing datasets are used as demonstrating datasets for our proposed methods in the following chapters.

Chapter 3 presents the IIRR method for dimension and redundancy reduction. The method is described in detail for a general data model at first. Then customized designs are provided based on specific features of two different demonstration datasets. Besides the main high dimension problem, wavelength saturation and non-uniformed time scale problems are described, as well as corresponding solutions. Finally, the model output is validated by the VM method with both demonstration datasets.

Chapter 4 presents the novel SRA method for early-stage fault detection in the plasma etching processes. First of all, the SRA method is introduced generally. Then detailed descriptions are provided, followed by results on the real IC manufacturing dataset. Finally, a conclusion is given at the end of the chapter to summarize the SRA method.

Chapter 5 gives a summary on all proposed methods. Potential future work is also discussed.

CHAPTER 2. Background

The development of IC technology has been in line with Moore's law [6], which predicts the growth over time in both the number of transistors per chip and the expected cost of production. The International Technology Roadmap for Semiconductor (ITRS) is a well-known document set, which is published every year and identifies key challenges, roadblocks and industry standards [7]. As indicated in recent ITRS (2012) updates, a transition to 450 mm diameter wafer improves productivity significantly in IC manufacturing. The manufacturing cost of each $mm^2$of IC would be decreased by use of these larger diameter wafers. However more dedicated and precise process control mechanisms are required to preserve yield integrity. Hence the design of control mechanisms has been a popular research topic in IC manufacturing, driven by this practical manufacturing requirement.

Typically, ICs are constructed through three major processes: imaging, deposition, and etching. A photo of the desired circuit configuration is transferred onto the surface of silicon wafers in the steps of imaging and deposition. Then the IC chips are created on these silicon wafers in the step of etching. Our research focuses on the etching procedure. It is one of the most important processes enabling IC fabrication on wafers. A more detailed description of the etching process is given in the following section of this chapter.

2.1 Plasma Etching

The etching procedure is described in detail in this section. There are two types of etching technologies: wet etching and dry etching. In wet etching, the wafer is immersed in a liquid solution. In dry etching, plasma or a vapour phase etchant are

sputtered onto the wafer. Normally, wet etching is the simplest etching technology. A suitable combination of etchant and mask material leads to a good quality of etching, and it costs less than dry etching. However, dry etching can achieve higher resolution. Plasma etching is one of the most developed dry etching technologies in current IC manufacturing.

Plasma is a gas consisting of atoms, molecules and charged particles. The charged particles include positive ions and negative electrons, which are created by a collision between atoms (or molecules) and a charged particle (ion or electron). This procedure is termed ionization. It is normally achieved by using an RF power source in a vacuum system. In Figure 2-1, this procedure is demonstrated with an incoming electron e^- and a neutral particle including an atom i^+ and an electron e^-. The collision between them frees the positive atom and negative electron from the neutral particle. A plasma gas is generated at the end, including two electrons and one ion. In IC manufacturing, this plasma gas is used to etch the targeted parts of material surface to build an electronic circuit.

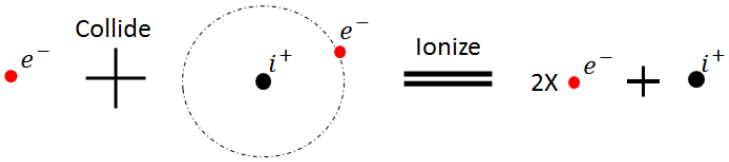

Figure 2-1. Demonstration of plasma generation.

When the energy used for the ionization is not large enough to break the atom, the atom will be activated to a higher energy state. However, the high energy state is not stable, and it will return back to a lower energy state. In this energy decay procedure, optical emissions will be observed [10]. There is an important physical theory that different chemical elements and compounds have different emission spectra. By observing the emission spectrum, chemical elements or compounds can be identified.

A combined spectrum of Ar and O is illustrated in Figure 2-2. Existence of these two particles is proved by high intensity values at three particular wavelengths in this figure. Relevant ratios between particles can also be uncovered. The OES monitoring technology which will be described later is based on this phenomenon.

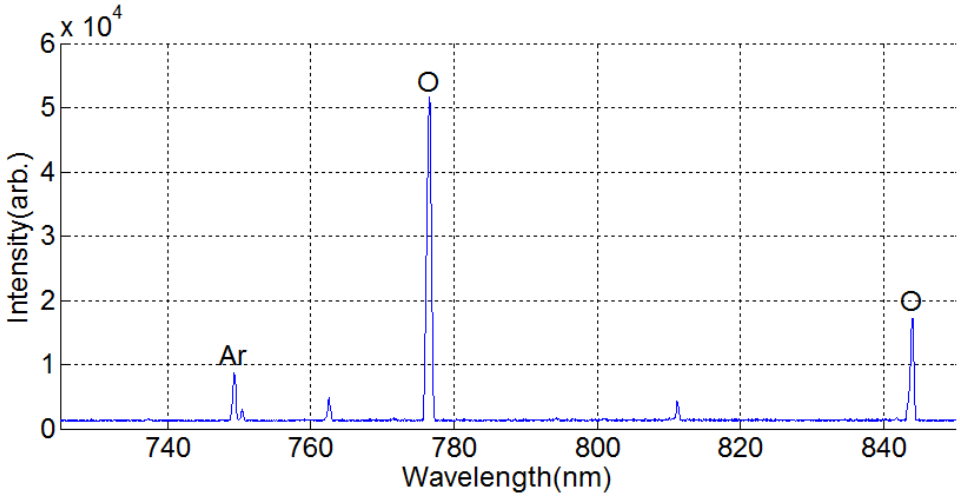

Figure 2-2. Optical emission spectrum of Ar and O.

2.2 Plasma Monitoring Technology

In order to understand the plasma etching process, effective plasma monitoring methods are necessary for data collection. The etching process is a very complex function of multiple process parameters, and these key parameters are understood poorly. Different monitoring strategies are based on various characteristics of the etching process, such as changes in radio frequency or temperature in the chamber. External sensors are normally deployed for data collection. Generally, there are two types of plasma diagnostic sensors that can be differentiated based on their impact on the etching process: intrusive sensors and non-intrusive sensors. Intrusive sensors normally give direct measurements on targeted parameters, but they also affect the process. For example, an electric current meter can reduce the current slightly during measurement. On the other hand, non-intrusive sensors do not have this weakness. Acoustic sensors and microwave sensors are two examples of this approach. However,

non-intrusive sensors introduce more background noise compared to intrusive sensors. In the rest of this section, OES which produces the data for our research is first introduced, and then the other widely used plasma monitoring technologies are discussed.

2.2.1 Optical Emission Spectroscopy

OES is a non-intrusive plasma diagnostic technology. All of our research is based on OES data. In physics, a chemical element or chemical compound produces optical emission, when it makes a transition from a high energy state to a lower one due to relaxation processes. It is an important physical characteristic that different chemical species have different emission spectra. By observing the optical emission spectra, OES helps to identify chemical species and monitor the changes of their ratios compared with the entire plasma in the chamber. Figure 2-3 outlines the OES sensor structure showing the two most important components: the diffraction grating and the Charged Coupled Device (CCD). When the emitted light hits the grating from the light source, the grating redirects the light based on their wavelengths. Light with different wavelengths will diffract to different parts of the CCD detector, where the intensity of each wavelength is converted to digital signals for further processing.

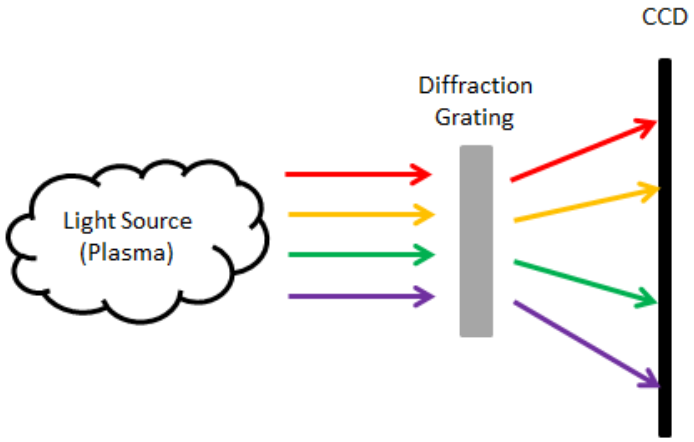

Figure 2-3. Illustration of a typical OES sensor with diffraction grating and CCD.

OES measurements are normally high in dimension and redundancy, due to two reasons. Firstly, one chemical element has more than one possibility of optical emission, because electrons can be in different states which contain different energy amounts. Secondly, not every chemical element contributes to the OES output is highly relevant or important to the output of the etching process. Aside from these problems, OES has two major advantages: large capacity for abundant information and the immediate relationship with the etching process. These advantages are not easily achieved by the other methods, so OES is widely applied to IC fabrication [11]. Much research has focussed on the OES measurements for multiple purposes, such as estimation of crucial system parameters [12, 13] and endpoint detection [14, 15]. Also there are other research purposes such as [16], where a 2048-dimensition OES dataset was condensed into principal components. Then system variations of the plasma etch process were tracked by the changes of principal component directions. So research on these OES problems is crucial for plasma monitoring [5]. More detailed description on the state of the art in this area is presented in section 2.3.

2.2.2 Langmuir Probe

The use of a Langmuir probe is an intrusive plasma diagnostic method. The probe is immersed into plasma as in Figure 2-4, and it determines the electron temperature, electron density, and electric potential of the plasma [17]. The most significant limitation of a Langmuir probe is the effect on the chamber environment. The inserted probe changes the temperature, density, and potential of the plasma, so measurement bias exists. Also, these three plasma characteristics are highly relevant to an IC fabrication, so the probe will affect the etching process consequently. The Langmuir probe is still used and studied widely [18, 19], because it provides the most direct measurements of plasma physical characteristics.

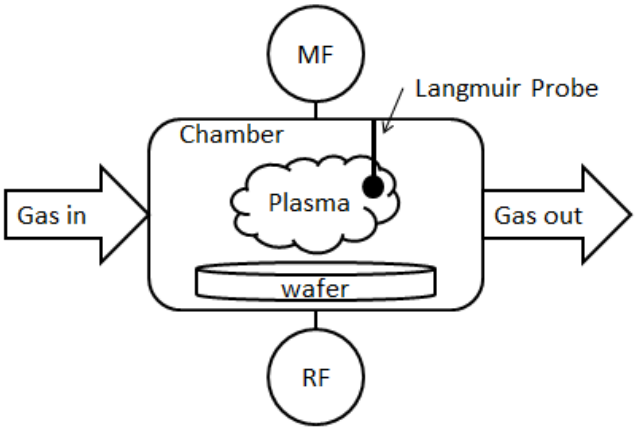

Figure 2-4. Deployment of Langmuir probe

2.2.3 Plasma Impedance Monitor

Plasma behaves like an electronic device with impedance in an RF circuit, which is connected for acceleration of plasma. Figure 2-5 illustrates a typical plasma chamber with Plasma Impedance Monitor (PIM) sensors. The measured plasma impedance variance results from chemical and physical changes in the plasma chamber [20]. It is a non-intrusive diagnostic method for plasma monitoring. The PIM monitor reads current and voltage, and calculates impedance from the circuit. These electrical parameters have been shown to relate to the outputs of the etching process. In [21]**,** a real-time model was demonstrated using PIM signals. It is used to maintain a consistent plasma electron density, in order to improve the result of plasma etching. In [22], Pascal Dubreuil and Djaffar Belharet also used PIM to develop a real time endpoint detection model, during plasma etching of structured bulk materials.

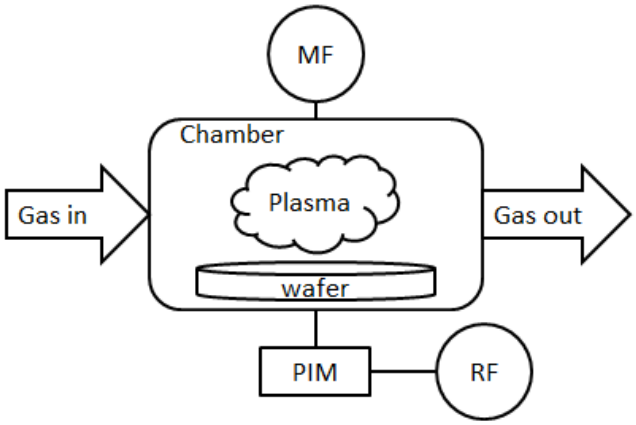

Figure 2-5. Deployment of PIM

2.2.4 Mass Spectrometer

A mass spectrometer is designed to identify the elemental composition of a material sample. The mass spectrometer generates charged molecules or molecule fragments by ionizing chemical compounds, and then mass-to-charge ratios of these chemical compounds are measured. Finally, the chemical species are determined by comparing with known masses. Mass spectrometry is widely used in multiple disciplines, like chemistry and biology. For example, protein mass spectrometry data was used for automatic extraction of biomarkers in [23]. The mass spectrometer does not give direct observation to plasma unlike OES, because the ionizing process can affect the etching process. So the mass spectrometer is normally deployed to monitor exhaust gas, to give indirect information about the plasma, Figure 2-6. The mass spectrometer also has its own usage that helps to improve understanding of the etching process [24]. One benefit of the mass spectrometer is its relatively easier identification of chemical species than OES. According to a United States Patent [25], mass spectrometry is successful in endpoint detection of etching during semiconductor fabrication.

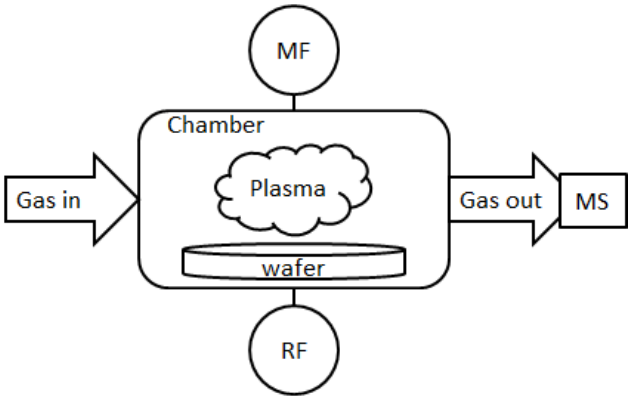

Figure 2-6. Deployment of mass spectrometer

2.2.5 Other Monitoring Methods

Besides multivariable plasma monitoring technologies such as PIM and OES, single process-state monitoring technologies are also used in some cases, and they are relatively easy to implement compared to multivariable techniques. In [26], the radio-frequency reflected power was used as an efficient method for end-point detection in etching process. In [27], changes of direct-current (DC) voltage were used in the detection of etching end-point. In [28], the use of DC current measurement was proven to be a significant improvement in signal-to-noise ratio for etching process monitoring compared with OES. In [29], the temperature of the wafer pad was used to monitor etching variations. In [30], the etch end-point was determined by monitoring the chamber pressure.

2.3 Survey of Existing Work

The state of the art of our research area is discussed comprehensively in this section. Literature relevant to our two research objectives (dimensionality reduction, and FD) is reviewed respectively.

2.3.1 Approaches to Dimensionality Reduction

In this thesis, the first reviewed research topic is dimensionality reduction. Dimension reduction is a popular research area across multiple disciplines. In [9], dimension reduction techniques were described from various viewpoints. From the statistical aspect, it involves regression methods. From the visualisation aspect, dimension reduction is necessary to transfer high dimension dataset to 2D/3D plot. From the information theory aspect, it relates to the efficiency of data transmission and storage. Principal Component Analysis (PCA), factor analysis (FA), stepwise regression, and cluster analysis are the commonly used techniques in this research area.

2.3.1.1 Principal Component Analysis

PCA is one of the oldest and best known multivariate analysis techniques. The key in PCA is to decrease the dimension of a dataset when there are large numbers of interrelated variables [31]. This is achieved by an orthogonal linear transformation from the original variables into new uncorrelated variables. These new variables are named principal components (PCs).

A mathematical description about PCA is presented below.

Given a $n \times m$ data matrix X (n observations with m variables), a $n \times m$ principal component matrix C is calculated by Equation (2-1) where L is a $m \times m$ loading matrix.

$$C = XL \tag{2-1}$$

In order to find suitable a loading matrix for PCA, singular value decomposition is commonly used, as in Equation (2-2), where S is a rectangular $n \times m$ diagonal matrix and both the $n \times n$ matrix U and the $m \times m$ matrix V are orthogonal unit vectors.

$$X = USL^T \tag{2-2}$$

Then covariance matrix X^TX can be written as:

$$X^TX = LS^2L^T \ . \tag{2-3}$$

Eigenvector factorisation of the square matrix X^TX can be expressed as Equation (2-4), where v is the eigenvector and γ is the eigenvalue which can be calculated easily. By comparing to Equation (2-3), the columns of L can be identified as eigenvectors of X^TX and the diagonal elements of S can be identified as square roots of the eigenvalues.

$$X^TXv = \gamma v \tag{2-4}$$

In [32], a diagnostic method was proposed to estimate wafer state line-width reduction with OES measurements. PCA was used for dimensionality reduction. This approach was verified on an experimental aluminium etch data set from a Lam 9600 plasma etcher. PLS was also included in a comparison with PCA. The authors concluded that both of them could offer simplicity and good accuracy under certain circumstances.

In [33], a chamber state model was developed to quantify and predict etching performance. OES was used as a real-time sensor and several etch characteristics were examined. These characteristics were etching rate, within-wafer uniformity, and aspect-ratio dependent etching. Three data reduction technologies were used in this study: ordinary least-squares regression, PCA, and PLS. The ordinary least-square

helped to select certain wavelengths based on species identification. Then PCA and PLS were compared for elimination of the correlation among selected wavelengths. PLS used fewer parameters to achieve the same result as PCA. The results also indicated that about 87% variation of etching rate and 95% variation of within-wafer uniformity could be predicted successfully with this model, but only 65% variation of aspect-ratio dependent etching could be explained.

In [34], PCA was used for dimensionality reduction as a pre-process for fault detection in a plasma etching process. By examining variable loadings to each PC, a subset of wavelengths was determined which were the most relevant to the monitoring objective. After this selection, only three wavelengths were retained.

In [35], PCA was used for endpoint detection in a plasma etching process, as well as an expanded hidden Markov model. PCA helped reduce the dimension of the OES dataset. The output of PCA increased the gap between classes in fault detection.

SPCA

A significant characteristic of standard PCA is that every PC is usually linear combinations of all original variables. In some scenarios, reasonable explanations of PCs are useful to understand the targeted process. For example, each variable could correspond to a specific gene in biology. So the explanations of the PCs would be much easier if these components are only constituted by fewer variables [36]. The same applies to research of plasma etching with OES. Each wavelength represents a certain chemical species. Identifying fewer important wavelengths can contribute to a better understanding of the chemical reactions in a plasma chamber. SPCA was developed to solve this problem. SPCA uses sets of sparse vectors as weights in the linear combinations, which still represent most of the variance in the original dataset.

There are multiple ways to implement SPCA mathematically, such as SCoTLASS [37], EN-SPCA [38], and sPCA-rSVD [39]. The application of SPCA is relatively new in the research of plasma etching.

In [40], the application of SPCA to clustering and feature selection problems was studied. Compared with standard PCA, SPCA was able to isolate relevant variables more efficiently with less computational cost.

In [41], a subset of key wavelengths was determined by SPCA from 2046 variable OES datasets. EN-SPCA was the implementation method used in this research. A trade-off between explained variance and a sparse representation was described as a key consideration in SPCA. One weakness of SPCA was also highlighted based on a grouping effect that exists in the selection of variables. The grouping effect means that EN-SPCA would give equal weights to highly correlated variables. This phenomenon, which was not ideal for variable selection, could be helpful in identification of related variables.

2.3.1.2 Factor Analysis

Factory Analysis (FA) is another statistical method used to transform an observed dataset into new coordinates, similarly to PCA. While PCA tries to combine variables into a small number of PCs, FA tries to identify the structure underlying the original variables, where latent factors are the elements building up the structure. The fundamental of FA is explained as follows. Suppose there are l observable random variables, $x_1, x_2, \ldots, x_l$ with means $\mu_1, \mu_2, \ldots, \mu_l$ and n observations for each variable. Suppose β_{ij} is constant and F_j are unobserved random variables, where $i \in 1, \ldots, l$ and $j \in 1, \ldots, m$, where $m < l$. Suppose we have

$$x_i - \mu_i = \beta_{i1}F_1 + \cdots + \beta_{im}F_m + \varepsilon_i \tag{2-5}$$

where ε_i are independently distributed error terms with zero mean and finite variance, $E(\varepsilon_i) = 0$ and $var(\varepsilon_i) = \sigma_i^2$. For each variable x_i, we could get the sample mean and sample variance, denoted by $\bar{\mu_i}$ and s_i. Standardization gives

$$z_{ik} = \frac{x_{ik} - \bar{\mu_i}}{s_i} \tag{2-6}$$

By minimizing the mean square error E, factors F_{ij} and loadings β_{ij} will be obtained by giving a best fit to the data.

$$E = \sum_{ip, i \neq p} \left[\sum_k z_{ik} z_{jk} - \sum_j \beta_{ij} \beta_{ij} \right]^2 \tag{2-7}$$

One of the biggest differences between PCA and FA is that all observed variance is analysed in PCA, but only the shared variances are analysed in FA. Generally speaking, there are two types of FA based on the assumptions made: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA seeks to determine the number of latent factors and the relationship between variables and factors. On the other hand, CFA seeks to validate the presumed factor structure and find the relationship between factors. Multiple model estimation methods are available, like the principle component method, principal factor method, and maximum likelihood [42].

In [43], a comparison of fault detection in a semiconductor etching process was provided between PCA and FA. Several sensor systems were used. Measured data included gas flow rates, power, and OES measurements. Strengths and weaknesses of these methods were presented on a set of benchmark fault detection problems.

According to the results, FA worked better than the other methods, but PCA might be the best choice in practice, because of its simplicity.

In [44], the application and misapplication of FA were described in marketing research. In order to verify the feasibility of FA on certain datasets, a number of examination methods were listed. These methods included examination of correlation matrix, plotting the latent roots obtained from matrix decomposition, and so on. A number of problems with this approach were also revealed, like the effects of extracting too few or too many factors on the factor stability after rotation. The identification of interpretable factors and simple structure were defined as the fundamental aims of a FA strategy.

In [45], FA was applied to 28 groundwater samples in the coastal black-foot disease area of Taiwan. A two-factor model was constructed which contained about 80% of the total groundwater quality variation. The first factor was interpreted as seawater salinization, while the second one was interpreted as arsenic pollution. Geographical distribution of these factors was also presented via maps. It was found that over-pumping was the major cause of groundwater salinization and arsenic pollution in that area.

In [46], an instrument development research cycle was completed by gathering new data to test the validity and reliability of an end-user computing satisfaction (EUCS) instrument. A CFA method was used. The results provided confirmation on the ability of the EUCS instrument for measuring and explaining user satisfaction. Five factors of EUCS were listed: content, accuracy, format, ease of use, and timelines.

2.3.1.3 Stepwise Regression

Stepwise regression is a way to select a best subset of independent variables for prediction of dependent variables. An iterative strategy is implemented in an automatic procedure. Normally, multiple runs are necessary, and each run has a criterion (e.g. achieve a certain level of fit goodness of the model) to check when the whole process should stop. The criterion is chosen to ensure a minimum selection of independent variables with acceptable regression result. F-test and t-test are two options for the criterion.

There are three main approaches for stepwise regression based on how regression starts: Forward Selection Regression (FSR), Backward Elimination Regression (BER), and Bidirectional Elimination Regression (BDER). In FSR, no variable is in the model at the beginning. Addition of each variable is tested using a model comparison criterion such as selecting the variable which improves the model the most. The variable which improves the model the most is added. This process will repeat until a certain threshold is reached, for example when there is no improvement from every additional variable. A general workflow of FSR is presented in Figure 2-7. In BER, all candidate variables are included at the beginning. The variable which has the least effect on regression is removed from the process for the next run. The BER process will stop having reached a certain threshold, similarly to FSR. In BDER, the ideas of FSR and BER are combined together. Variables are added in sequence to the model. A variable will be retained if it improves the regression result, and then all variables already in the model will be retested. The variables which do not contribute significantly are removed from the model [13].

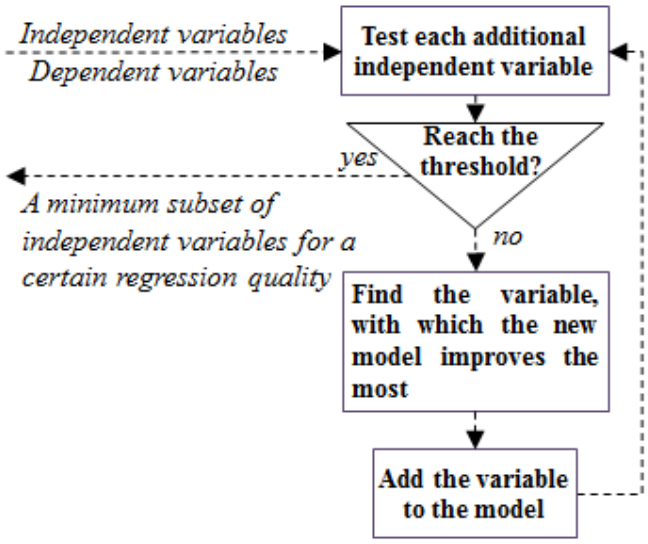

Figure 2-7. Workflow of a general FSR

In [47], FSR and forward selection component analysis (FSCA) were introduced as well as PCR and PLS in the context of semiconductor manufacturing research. All of them were used to identify key process characteristics and predict etch rate from high-dimension datasets. FSCA was based on FSA and is similar to the PCA theory. Instead of finding a minimum subset of variables for prediction, FSCA returns a set of orthogonal components from a minimum subset of variables, which represent the variance of original data. The OES dataset was used for analysis in this research. Six statistical moments were used to reduce the high dimensionality of the dataset. These statistics were kurtosis, mean, skewness, variance, maximum, and minimum. According to comparative results, FSCA and FSR were more effective than PCA and PLS for feature selection.

In [48], FSCA was used to determine a minimum set of wafer sites for plasma etching process monitoring. A VM approach was also proposed to reconstruct a complete wafer profile from the optimal wafer sites. There were 316 wafer samples for analysis. These samples were randomly grouped into two parts for training and testing respectively. This random organization was designed to reduce the impact of temporal variation between successive samples. Initially, PCA was used to determine the level

of redundancy in the historical dataset. Further FSCA was applied if high information redundancy was confirmed by PCA. Then a VM model was created on the optimal wafer sites for prediction of unmeasured sites. An R^2-fit of over 97% was achieved for prediction of all sites. Finally, a 3D wafer profile could be constructed by combining the values of the actual and the VM sites.

In [49], an input-clustering based FSR methodology was provided for VM model building in highly correlated input spaces. Max Separation Clustering (MSC) was used to pre-process data. A reduced set of representative variables were selected after the MSC process. Comparable modelling results were gained from three other modelling technologies: ridge regression, LASSO, and Forward Selection Ridge Regression (FSRR). This VM model was tested on a dataset of semiconductor plasma etch. Over 2000 wafer samples were included, and each sample consisted of 4 statistics of the OES measurements. The statistics were mean, variance, skewness, and kurtosis. 50% of the samples were used for training; 25% for validation; and the last 25% for testing. According to results, the best performance was achieved by the model with FSR.

In [50], stepwise regression was used to identify tools with large contributions to the overall variance in a high yield plasma process. An indicator variable was used to mark whether a tool was involved at a certain step of the process. For example, the variable was assigned a “1” if a tool was used in the current step. Otherwise, the variable was “0”. The indicator variable was taken as the input of the model. High information redundancy was contained in the indicator variable, so PCR was applied to remove redundancy before stepwise regression. Compared with analysis of variance, stepwise regression had a better result. It provided a holistic view of the process with a consideration of the relationship between steps. This relationship was normally

ignored by transitional analysis of variance. But stepwise regression is relatively more difficult to implement and less efficient in computational calculations.

2.3.1.4 Cluster Analysis

In terms of dimension reduction, the goal of cluster analysis is to group variables based on their features, and then use the groups to represent the original dataset. A large amount of works have been done in this area, and famous clustering methods include Expectation Maximization (EM), K-means [51] and so on. The EM approach tries to find a set of Gaussian distributions which can represent the given data set the best. The K-means approach separates samples into k groups which give the minimum sum of distances from the samples to their own group centres. Given a set of observations ($x_1, x_2, \ldots, x_n$), the observations will be divided into k groups $G = \{g_1, g_2, \ldots, g_k\}$, where $k \leq n$. The K-means objective is to find the following partition G

$$arg\ min_G \sum_{i=1}^{k} \sum_{x_j \in g_i} \|x_j - \mu_i\| \tag{2-8}$$

where μ_i is the mean of the ith cluster.

Both of these two well-known methods suffer from the same local minima problem [52] in which the clustering procedure halts on finding a local minima close to the initial starting point, and so does not find the best global solution. In terms of the other clustering methods, most cluster analysis methods share the following common problems [53]. Firstly, users should manually decide the number of groups for most methods, but this number is always difficult to determine. Secondly, there are a huge amount of different clustering methods and each of them almost gives substantially

different results. Sometimes even the same method does not give consistent results (e.g. a K-mean approach which starts with different initial points). Hence, stability is an issue of most clustering methods.

In [54], a comparison between cluster analysis and PCA was provided in the synoptic meteorology research area, in which PCA gave better results or performance than clustering methods in most cases.

In [55], the K-means clustering method was used along with a linear discriminant analysis technique for dimension reduction. K-means clustering was used to find different clusters, and then linear discriminant analysis helped for cluster selection. By comparing with other common-used clustering methods, the proposed method was proved to give the best clustering accuracy, according to the experimental results.

In [56], K-means clustering was combined with PCA. A strong relationship between K-means and PCA was disclosed. By testing them together with different combinations, PCA was proved to be a suitable solution to determine cluster indicators for K-means clustering.

2.3.1.5 Summary

There is relatively little research focusing on dimension and redundancy reduction of OES datasets. Among the few examples, dimension reduction techniques are used to a lesser or greater extent in the works. Most of these reviewed methods are based on general mathematical models, which ignore particular features of the OES data in plasma etching, such as the remarkable role of peak wavelengths. They also have difficulty in finding important variables. SPCA could be a potential solution, but the grouping effect is its weakness, as mentioned above. The proposed IIRR method of

this thesis is created to address these research gaps. IIRR focuses on dimensionality reduction of the OES dataset. Instead of building new variables (e.g. using PCA), IIRR works directly on the original variable domain, which will benefit key variable identification for further analysis.

2.3.2 Approaches to Fault Detection

Traditional FD techniques have two common problems: high implementation cost and a long-time delay between fault occurrence and fault detection. For example, the Scanning Electron Microscopy (SEM) is used to measure etch depth, and then etch rate is calculated by the depth divided by the total etch time. The etch rate is a popular statistic to assess the process and wafer quality. However, this method impacts the testing wafer for SEM measurements, which introduce a large cost. It also needs to wait for the end of the etching process, so a long-time delay is involved. The typical time delay to get the etching result with traditional metrologies was demonstrated in [57]. It often takes hours or even days to get the result. During that time period, thousands of wafers can be damaged if the detected fault persists in the process. Due to these problems, the OES dataset has been widely studied for FD in the past decades. OES has two important features: real-time monitoring and non-intrusion. However, it has its own problems too. Like other research with OES data, the OES FD strategy suffers from the problem caused by OES high dimensionality. Time series OES measurements from a healthy and a faulty etching process are demonstrated respectively in Figure 2-8. Panel (a) represent time series OES of a healthy etching process, while panel (b) represents time series OES of a faulty etching process. It is difficult to tell the difference between them by a simple observation, so data processing is always required to dig useful information from the raw OES data. Additionally, there are other problems associated with the OES approach to FD.

In [58], the author mentioned that etch quality could be affected significantly by small variations in equipment and process step phase. In other words, small variations in OES could be important information which should be extracted out. Considering the OES background noise with similar small variation, it is difficult to separate them from each other for FD.

In [59], process shift was discussed as another common problem. A typical scenario of this problem is presented and discussed in the result and discussion section of this chapter based on a manufacturing dataset, which we name D1. The process shift can lead to multiple monitoring problems. For example, more false alarms could be triggered.

Due to these challenges, a large number of research interests are attracted to FD with OES. However, most of them share some common problems which are discussed in the following paragraphs. Popular methods include PCA, support vector machine (SVM), pattern recognition, ICA, and ANN.

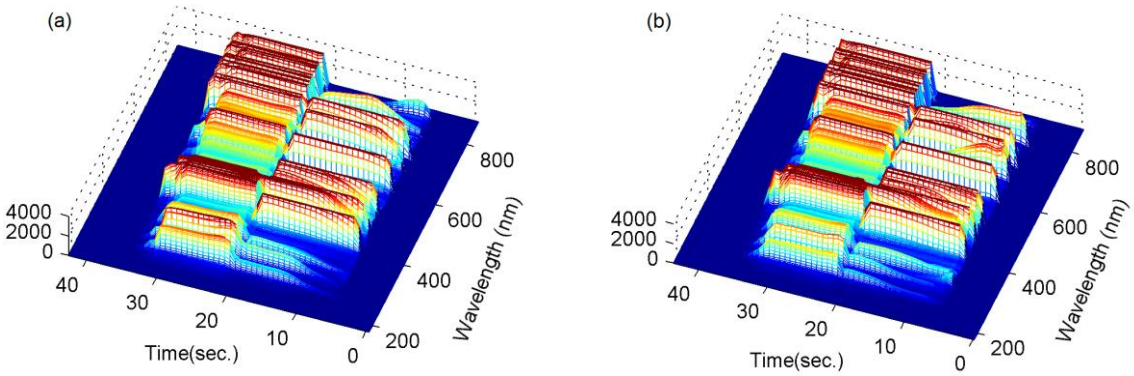

Figure 2-8. Comparison of time series OES between healthy and faulty samples. Panel (a) represents the time series OES measurements of a healthy etching process. Panel (b) represents the time series OES measurements of a faulty etching process. There is no easily observable significant difference.

2.3.2.1 Principal Component Analysis

PCA is commonly used in multiple research areas including dimensionality reduction which has been discussed in a previous section. PCA is a well-known tool for fault detection, too. In PCA, most variance is accumulated in the first few principal components. By examining these principal components, faulty samples could be easily discriminated.

In [60], an U.S. patent describes a method of monitoring the status of plasma in a chamber with OES. Real-time OES measurements were compared with reference data via PCA. The presence of foreign material faults could be determined by this comparison and the cause of the failures could also be uncovered. The method was also successful in the detection of other etching faults, such as faulty equipment, chamber wall conditioning, and so on.

In [34], a PCA model was built for fault detection in a plasma etching process. First PCA was used to reduce data dimension. Then PCA was used again for fault detection on remaining wavelengths. According to the result, this model was easy to implement and it could also help with fault classification as well as FD.

In [43], a comparison of fault detection in a semiconductor etching process was provided between PCA and FA. Several sensor systems were used. Measured data included gas flow rates, power, and OES measurements. The strengths and weaknesses of these methods were presented for a set of benchmark fault detection problems. According to the results, FA worked better than the other methods, but PCA might be the best choice in practice, because of its simplicity.

In [8], two PCA problems were discussed for FD. Firstly, the nonlinearity characteristics of semiconductor plasma processes make the traditional PCA approach difficult to implement. Secondly, PCA normally needs two control charts for the fault detection: T2 and SPE. This would increase operation cost significantly, compared with one single chart.

2.3.2.2 Independent Component Analysis

Independent Component Analysis (ICA) has a similar information transformation strategy as PCA and FA. While PCA and FA are based on second order statistics, ICA involves higher order statistics. ICA normally has more meaningful results than PCA and FA, but the ICA result is even better if the data is pre-processed by PCA first [61]. ICA assumes a statistical model that the multivariate dataset is a mixture of independent components, and ICA is the analysis to find out these hidden components. ICA is widely used for feature extraction and signal separation by capturing the essential structure of the data.

In [62], detailed algorithms for ICA were described. ICA is rarely mentioned in plasma etching research, but it is a common technique in other multivariable data analysis research areas.

In [63], ICA was applied on electroencephalographic data collected from sensors attached to the human scalp. According to the results, several conclusions were made. First of all, the ICA training was not sensitive to seed selection. Secondly, ICA might be used to separate obvious artificial variables from the others. Thirdly, overlapping data could also be segregated by ICA.

In [64], a modified ICA method was proposed. The method was demonstrated using two different datasets from a wastewater treatment system and semiconductor etching system respectively. Like most of the other research, faults can only be detected after the whole process completes. Problems with ICA were also described, for example, it was always difficult to pick the component number and order in practice for ICA.

2.3.2.3 Artificial Neural Network

Artificial Neural Network (ANN) is a mathematical model which is inspired by biological neural networks. ANN includes a group of artificial neurons and also interconnections between them. Normally, an adaptive system changes its structure during a learning phase. This learning phase aims at providing the most matched outputs according to desirable values. ANN is studied as a solution to a variety of problems, such as pattern recognition, prediction and control. There are two ANN types according to two different connection patterns: feed-forward ANN and feedback ANN. Feed-forward ANN has no loops, while feedback ANN has loops. Applications of ANN are present in many research areas and it works successfully for different purposes.

In [65], a FD method was proposed using the ANN. Two sensors were involved: the OES and the Residual Gas Analysis (RGA) sensor. The OES measured optical emission at 2048 different wavelengths. The RGA gained 200 atomic masses for each measurement. PCA was used to reduce the OES dimension as a pre-processing step. This method could only detect faults after collecting data for the whole duration of the process.

In [34], a single OES scan was used for fault detection and classification at eight fixed time points. PCA was used in this case, but only low FD accuracy was obtained.

Therefore, the author concluded that a single OES scan could not provide a reliable FD service.

In [66], a ANN method was designed for real-time monitoring in 300mm semiconductor manufacturing processes. A piecewise linear ANN and a fuzzy ANN were both included. The former ANN was use to approximate the drift trend, which was inevitable. The latter ANN was used to study the influence of process recipe on fabrication outcome. Random noise was discussed in this research, which had a significant effect on the accuracy of VM.

2.3.2.4 Other Methods

In [67], a Support Vector Machines (SVM) based method was used for FD in a reactive ion etching system with OES. For each OES measurement, 2048 wavelengths were measured from 186.58 nm to 746.85 nm. Instead of using a full set of wavelengths, 35 important wavelengths were selected by the PCA and then the average values of wavelength intensity were calculated as the input to the support vector machine. 100% detection accuracy was achieved with no false alarm. The problems of ANN were also discussed in this context. The authors argued mainly that it was very difficult to obtain faulty samples for the ANN training procedure.

In [58], a new fault detection and classification method was proposed for monitoring of chamber state in real time. Input data was produced by an impedance plasma sensor, which included key system parameters including voltage and electric current. In a training phase, a library was created to store patterns of faulty samples. By comparing with this library, faulty samples were identified and classified.

In [57], a pattern recognition method was used for FD with OES. By calculating a matching rate between the testing sample and a library representing healthy samples, a fault could be detected successfully. Originally, the matching rate took all the OES scans of a complete etching process as inputs, so the fault could only be detected at a post-process phase. Additionally, the authors did not give necessary discussion about some important parts of this method. For example, the authors built the library for healthy samples using average intensity plus/minus 3 standard deviations. But they did not discuss why 3 standard deviations were used.

2.3.2.5 Summary

In general, most methodologies only realise FD after finishing the whole etching process. A few methodologies are able to detect faults at an early stage with a single spectral scan, but the time when the spectra are taken is normally limited to a few specific time points. For example, the method proposed in [57] could potentially be adjusted to take individual scans as input, however, there is still the problem that it would only take OES measurements at a few specific time points as input. Due to these research gaps, the SRA method is proposed in this thesis to achieve reliable early-stage fault detection. Additionally, the SRA method is not limited to certain pre-defined time points, so the flexibility of the FD system is increased.

2.4 Experimental Industrial Dataset

All our research is based on two time-series OES datasets, denoted D1 and D2. These two datasets are collected from the plasma etching process in two different IC manufacturing factories separately. Different IC factories have different gaseous recipes and system configurations for plasma generation and control. They also use different sensors. Because of these reasons, D1 and D2 represent different plasma

etching procedures and have different data structures and features. In the discussion that follows, for either dataset, a *sample* is defined as time-resolved spectrum scans for a complete etching process on one wafer.

2.4.1 Demonstrating Dataset D1

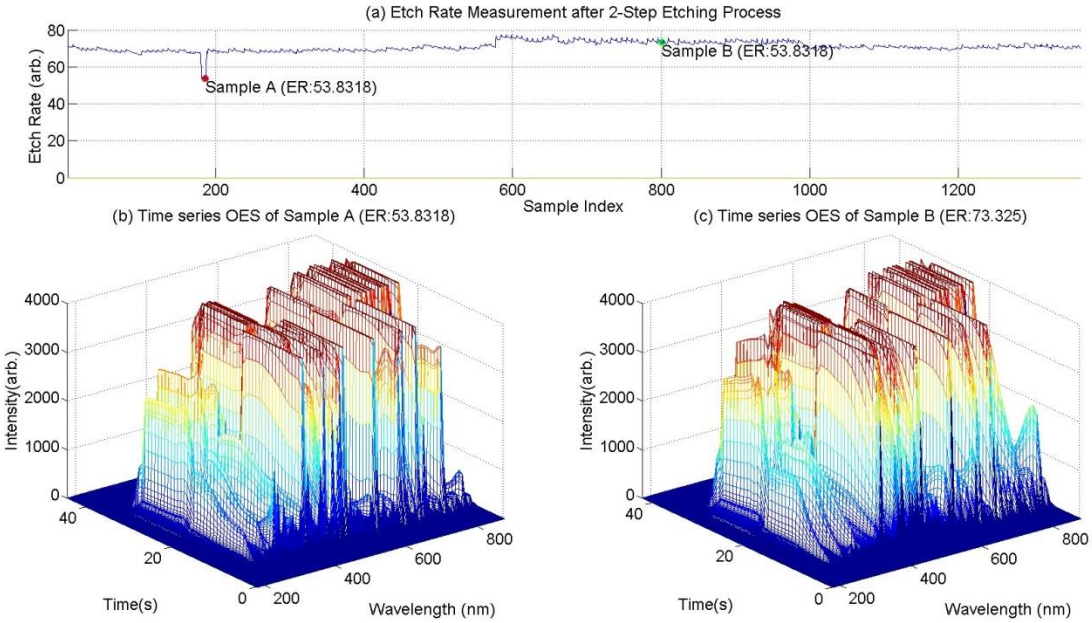

Figure 2-9. Illustration of dataset D1. Panel (a) presents etch rate measurements of all samples. Panel (b) represents OES measurements of sample A. panel (c) represents OES measurements of sample B.

The dataset D1 comprises 1750 samples, and each sample includes a set of time series OES and one etch rate measurement. In Figure 2-9, etch rates of all samples are presented in panel (a). Two samples are highlighted and relevant OES measurements are presented in panel (b) and (c) respectively. The OES measurements are acquired from a 2-step etching process, and relevant etch rate is acquired at the end of the etching process by measuring the etch depth and etch time. The USB4000 Miniature Fibre Optic Spectrometer is used to measure the optical emission. Important features of this sensor are presented in Table 2-1 along with additional parameters that are used in my research, such as the sensor signal-to-noise ratio. Every spectrum scan includes optical intensity measurements at 2048 different wavelengths from 178 nm to 874 nm.

About 50 seconds are needed to finish the whole etching process. About 25 seconds are needed for each etching step. Different gaseous recipes are used in different etching steps. Gaseous species in the two steps include noble gases, fluorine and chlorine based compounds. There is a routine cleaning procedure between these two steps. It removes residue from the step 1 before step 2. However, no cleaning is ideal, and there will always remain some by-products on the chamber wall. This practical problem is one of the major causes of variance in the process output. The etch rate measurement which is included in every sample is common metrology data used to describe the output quality of an etching process. It is difficult to measure, so it is impracticable to use it for any efficient and timely process control system. There are two major research challenges associated with this dataset: a wavelength saturation problem and variance in the sampling time points between process runs. Corresponding solutions are described with the proposed IIRR method for dimension reduction in Chapter 3.

Wavelength range:	Grating dependent
Optical resolution:	~0.1-10.0 nm FWHM (grating dependent)
Signal-to-noise ratio:	300:1 (at full signal)
A/D resolution:	16 bit
Dark noise:	50 RMS counts
Integration time:	3.8 ms - 10 seconds
Dynamic range:	3.4 x 106 (system), 1300:1 for a single acquisition
Stray light:	<0.05% at 600 nm; 0.10% at 435 nm

Table 2-1. USB 4000 Spectrometer Specifications [68].

- **Wavelength Saturation Problem**

  Saturated wavelengths can be observed as dark red flats on the top of Figure 2-10. This is caused by the limitation of the spectrometer when optical intensity is larger than the maximum measurement value of the spectrometer. Compared with other intensities below the limitation, the measurements on the dark red flat does not provide the true optical intensity value.

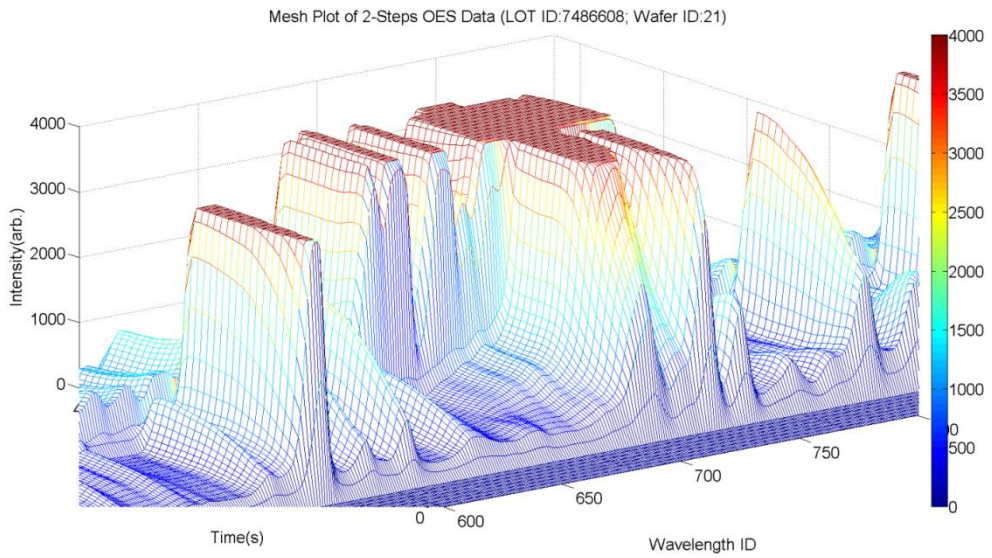

Figure 2-10. Illustration of saturation problem in OES measurements.

- **Variance of Sampling Time**

All samples represent a similar etching process, but the time when each spectrum scan is taken and the sampling frequency varies from sample to sample. The estimated Probability Mass Function (PMF) plot of time intervals of all samples is presented to explain this problem in Figure 2-11. The time interval is calculated by the time stamp of one spectrum scan minus the time stamp of its previous consecutive spectrum scan in the same sample. Based on the calculation result, about 85% of time intervals are between 0.76 second and 0.77 second, and the rest intervals are between 0.776 second to 0.789 second. Additionally, the time when OES measurements start is also different across samples. Therefore, different samples have their spectrum scans at slightly different time points. In this case, particular data processing is necessary to normalize all samples in order to have the same time scale.

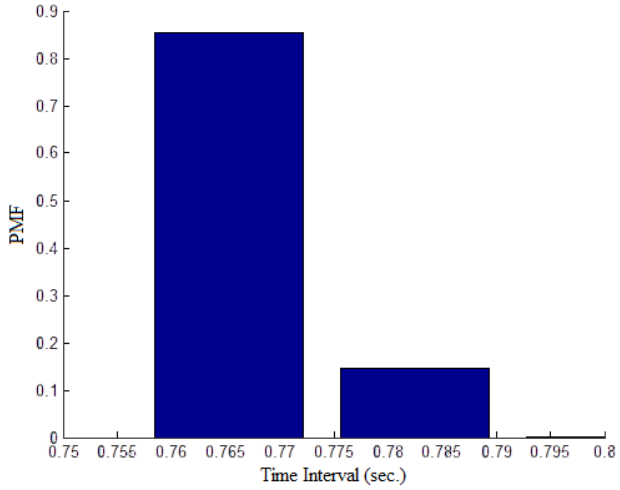

Figure 2-11. PMF of time intervals between consecutive spectra for all samples.

2.4.2 Demonstration Dataset D2

Dataset D2 is obtained from a different plasma etching process. This etching process is used for formation of a write pole mask. Three etching steps are included, as in Figure 2-12. In etching step 1 as in Figure 2-12(a), layer 1 is etched by CF4 plasma, Ar and O_2. In etching step 2 and 3 shown in Figure 2-12(b), the gaseous recipe includes O_2 plasma, Ar and SO_2. After all three etching steps, the mask for write pole formation is built. Then, an ion milling process is conducted to etch into the write pole material based on the mask. A write pole with designed shape will be gained at the end as in Figure 2-12(c), after removing the mask by a lift-off process.

There are two separate sets of consecutive OES measurements for each sample. The first OES measurement set OES1 comes from etching step 1, while the second OES measurement set OES2 comes from etching step 2 and 3. Comparing with the spectroscopy for D1, a different OES sensor is used for data collection, Spectrometer SD1024D [69]. 1201 wavelength measurements are obtained for each spectrum scan from 200 nm to 800 nm. Unlike D1, all samples of D2 start OES measure at 0.05 second, and time intervals of OES sampling is always 0.10 seconds, consistently.

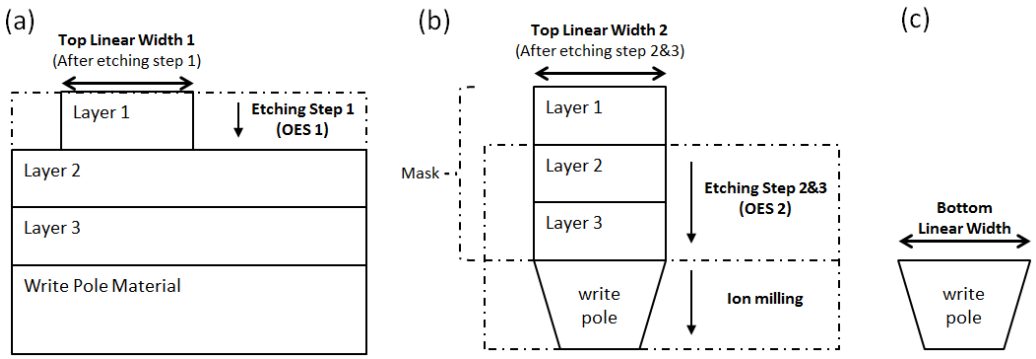

Figure 2-12. Illustration of etching process for dataset D2. Panel (a) represents the etching step 1. Panel (b) represents etching 2 and 3. Panel (c) represents the final write pole after all etching steps.

Instead of being interested in vertical etch rate, D2 focuses on two line width (LW) measurements of the mask after finishing all etching steps. LW is another important etch parameter which is highly relevant to the IC chip quality. Ideal etching should be anisotropic (vertical), but actual processing also etches into the layer wall horizontally, which impacts LW of resist layers. Therefore, it is valuable to figure out the influence of plasma etching on LW. The first LW value included in this dataset is the Top LW Bias (TLWB), which is calculated from Top LW measurement 2 (TLW2) minus Top LW measurement 1 (TLW1), as in Equation (2-9). TLW1 is acquired after etching step 1 and before step 2. TLW2 is acquired after etching step 3. The second LW measurement included in the dataset is Bottom LW (BLW). Based on two different target values of BLW (0.112 μm and 0.114 μm), all 20 samples can be separated into two groups. The first group with 14 samples has target BLW value 0.112 μm, and the second group with the remaining 6 samples has target BLW value 0.114 μm. Relevant analysis in my research is processed on each group separately. Like etch rate in D1, LW is also difficult to measure. In Figure 2-13, dataset D2 is illustrated with all LW values and OES of two demonstration samples. Panel (a) presents TLWB values of all 20 samples. Panel (b) presents BLW values of all 20 samples, as well as target values. Panel (c) represents OES measurements of Sample A in etching step 1, while Panel (d)

represents OES measurements of Sample A in etching step 2 and 3. Panel (e) represents OES measurements of Sample B in etching step 1, while Panel (f) represents OES measurements of Sample B in etching step 2 and 3.

$$TLWB = TLW2 - TLW1 \tag{2-9}$$

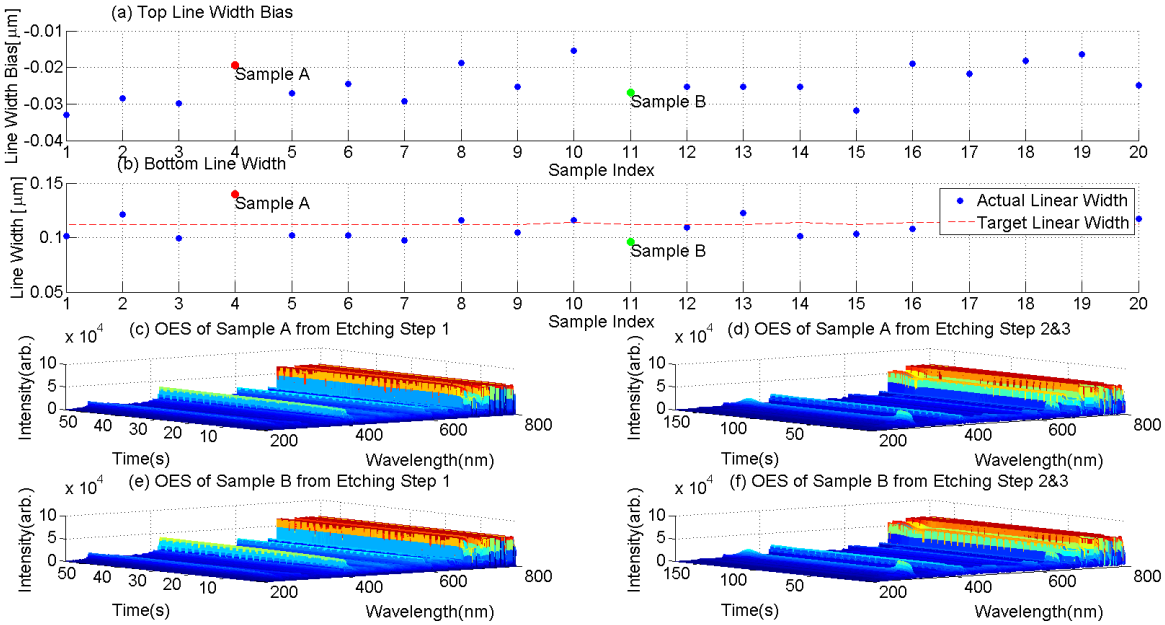

Figure 2-13. Illustration of dataset D2. Panel (a) presents TLWB of all 20 samples. Panel (b) presents BLW of all 20 samples. Panel (c) represents OES measurements of Sample A in etching step 1, while Panel (d) represents OES measurements of Sample A in etching step 2&3. Panel (e) represents OES measurements of Sample B in etching step 1, while Panel (f) represents OES measurements of Sample B in etching step 2&3.

Like D1, saturated wavelength measurements can be observed as dark red flats on the top of Figure 2-14. This is also caused by the spectrometer limitation that optical intensity cannot be measured when the value is bigger than the maximum measurement value of spectrometer. Compared with other intensities below the maximum, the measurements on the dark red flat cannot provide true optical intensity values.

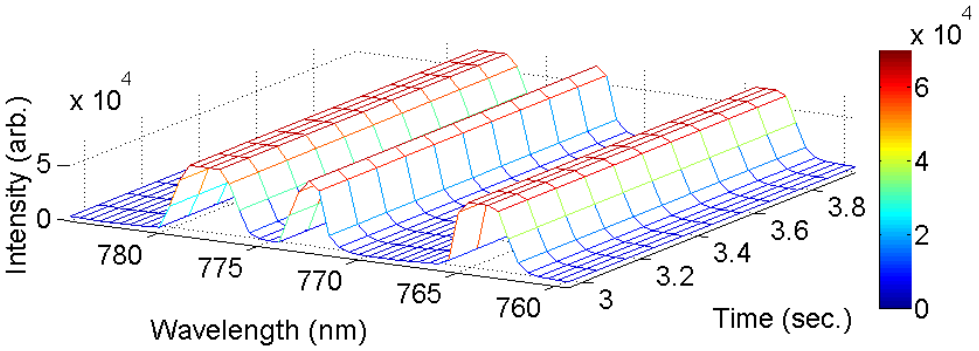

Figure 2-14. Illustration of saturation problem in OES measurements. It caused by the spectrometer reaching its maximum measurement value.

CHAPTER 3. Internal Information Redundancy Reduction

3.1 Introduction

A novel dimension and redundancy reduction method, which we call Internal Information Redundancy Reduction (IIRR), is proposed for multivariable datasets. These datasets are based on information obtained from OES in a plasma etching process. OES provides real-time information on the spectral content of light emitted from processing plasma. This sensor type is normally high in information redundancy. By reducing the data dimension, expected benefits include improvement of computational speed and rapid identification of important variables. Compared with existing methods in the literature, IIRR operates directly in the original variable space, so it is easier to identify important variables. Additionally, the method could be reused and extended easily to other application domains.

The core IIRR model comprises three sub-steps: *Absolute Peak Selection* (APS), the *Iterative Ranking Process* (IRP) and *Optimized Peak Selection* (OPS). APS filters variables based on peak wavelength characteristics. Then IRP and OPS reduce the number of wavelengths in the data further based on the linear relationship between variables. A new statistic, Mean Determination Ratio (MDR), is created to quantify the information loss after dimension reduction. IIRR is demonstrated on two real manufacturing datasets, D1 and D2. Practical problems such as wavelength saturation and non-uniform time scales are also discussed and addressed here. In order to quantify the information loss in the IIRR dimension reduction procedure, the original OES data and dimension-reduced IIRR output are used separately to predict the etch rate, and their prediction accuracies are compared.

3.2 Core Method

The proposed method is a dimension reduction method with 3 sub-steps: APS, IRP, and OPS, as shown in Figure 3-1.

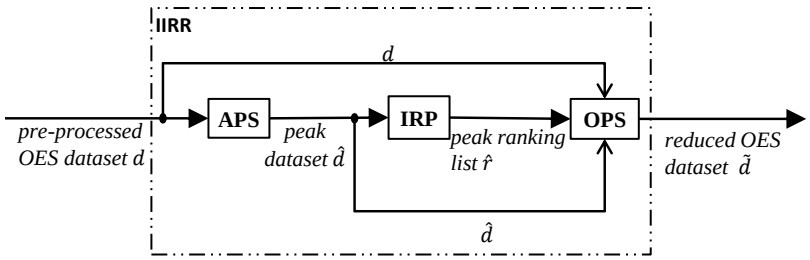

Figure 3-1. Workflow of IIRR.

At the beginning of IIRR, APS identifies peak wavelength variables in the input OES dataset $\boldsymbol{d}$. The output of APS for a time point t_j is the set of wavelength indices for which peaks in wavelength intensity occur, at time point t_j, during any of the N process runs. Formally, the output of APS is the collection of sets of peak wavelength indices (Equation (3-1))

$$\begin{aligned}\widehat{\boldsymbol{d}} &= \langle \hat{d}_{t_1}, \hat{d}_{t_2}, \dots, \hat{d}_{t_M} \rangle, where\ \hat{d}_{t_j} \\ &= \{k \in \{1,2,\dots,K\} : I_k\}\end{aligned} \tag{3-1}$$

I_k is a peak wavelength intensity at time t_j in at least one process run. Next, the IRP algorithm takes each set of peak wavelengths $\hat{d}_{t_j}$ and ranks each peak wavelength variable in the set based on how well it can be predicted from other peak intensity values in $\hat{d}_{t_j}$. Peaks that are poorly predicted by others are assumed to hold more significant information and are ranked more strongly. Repeating the procedure for each time point, the output of IRP (Equation (3-2)) is a dataset of the same dimensions as $\widehat{\boldsymbol{d}}$ containing wavelength indices in order of ranking,

$$\begin{aligned}\hat{\boldsymbol{r}} &= \langle \hat{r}_{t_1}, \hat{r}_{t_2}, \dots, \hat{r}_{t_M} \rangle \\ \hat{r}_{t_j} &= \text{the sequnece } \langle k_1, k_2, k_3, \dots \rangle,\end{aligned} \tag{3-2}$$

$$\text{where } rank(k_{l-1}) < rank(k_l), \forall k_l, k_{l-1} \in \hat{d}_{t_j}$$

Finally, at each time point t_j, the OPS algorithm calculates a measure of how well the first r top-ranked peak wavelength variables can predict the full set of original wavelength variables. An aggregate statistical measure *Mean Determination Ratio* (MDR) is used to summarize the prediction quality for different r. A minimal value of r is determined under a constraint on the MDR value. The procedure is repeated at each time point and the final output of OPS is a time series of the minimized peak wavelength sets, denoted $\widetilde{\boldsymbol{d}}$. The details of these three stages of the IIRR procedure are given in the subsections below.

3.2.1 Step 1 - Absolute Peak Selection

In the context of physics, optical emission is triggered by energy transition from high electronic states to low electronic states in atoms and molecules. The same atoms can even have various frequency combinations of light based on multiple states the electrons can have, so a plasma etching process will have a very complex optical emission output with high information redundancy. Any OES sensor will also introduce errors and noises into samples. However, intensities of emissions are concentrated on certain more pronounced wavelengths than the others, and these peak wavelengths stand out from the others.

Absolute Peak Selection (APS) operates independently at each time point t_j and, for each process sample, identifies a wavelength variable as a peak wavelength if its intensity is greater than the intensities of neighbouring wavelengths plus the bias value B. B is calculated based on Equation (3-3), where I_{max} is the maximum optical intensity of the sensor and SNR is the signal-to-noise ratio of the sensor. Let d_{i,t_j} be

the set of wavelength intensity values $\{I_k : k = 1,2, \dots, K\}$ measured at time point t_j during process run i. Then peak wavelengths within this set are identified by Equation (3-4). Having found $\hat{d}_{i,t_j}$ for all process runs $i = 1,2, \dots, N$, all peak wavelength variables are then aggregated as explained in Equation (3-5).

$$B = \frac{I_{max}}{SNR} \tag{3-3}$$

$$\hat{d}_{i,t_j} = \{k : I_{k-1} + B < I_k < I_{k+1} + B, 1 < k < K\} \tag{3-4}$$

$$\hat{d}_{t_j} = \bigcup_{i}^{N} \hat{d}_{i,t_j} \tag{3-5}$$

Finally, iterating this procedure at each time point, the complete output of APS is $\hat{\boldsymbol{d}} = \langle \hat{d}_{t_1}, \hat{d}_{t_2}, \dots, \hat{d}_{t_M} \rangle$. The rationale for this aggregation in the above equation is that the data gathered from each process run is an independent sample of an (ostensibly) fixed etching procedure, thus the process state should be similar for all samples at the same time point, but different samples could have different peak selection results because of measurement variance. Hence, aggregation can achieve data reduction without significant loss of information content by retaining all potential peaks, even if some of them only exist in a few samples.

Peak wavelength detection methods have been studied and used widely in biomarkers identification for pathology analysis, including mass spectrometry [70] and mass-to-charge ration measurements [23], but they are rarely applied to plasma process data. PCA, SPCA and FA are common methods used in dimension reduction of plasma OES, but all of them focus only on mathematical characteristics of OES [35, 41]. Compared with these methods, APS is designed to focus particularly on physical features of the OES data. Another remarkable advantage is that data dimension is reduced by filtering

wavelengths, instead of by combining original wavelengths. It is helpful for important wavelength identification in the further processes after IIRR.

The pseudo code for this algorithm is given below.

```
Procedure APS
FOR every training sample
    FOR spectrum scan at every time point
        FOR every wavelength
            IF its intensity is larger than intensity of wavelengths on the left and the
            right sides plus spectrometer bias
                mark the current wavelength as a peak;
            ENDIF
        ENDFOR
    ENDFOR
ENDFOR
```

3.2.2 Step 2 - Iterative Ranking Process

The Iterative Ranking Process (IRP) process is designed to reduce the number of possible peak combinations. Without IRP, a huge amount of peak combinations would slow down the subsequent OPS process.

Each set of peak wavelength intensity samples $\hat{d}_{t_j}$ is treated separately by IRP, as follows. For each wavelength intensity variable $I_k, k \in \hat{d}_{t_j}$, an ordinary least squares linear regression is performed as in Equation (3-6).

$$\tilde{I}_k = \beta_{k,1} I_1 + \beta_{k,2} I_2 + \cdots + \beta_{k,k-1} I_{k-1} + \beta_{k,k+1} I_{k+1} + \cdots + \beta_{k,K} I_K + \varepsilon_k \qquad (3\text{-}6)$$

$\beta_{k,k'}$ are the regression coefficients and ε_k the error term. IRP then calculates the coefficient of determination (the R^2 value) of the prediction $\tilde{I}_k$ of I_k, denoted by R_k.

The lowest ranked wavelength in $\hat{d}_{t_j}$ is then identified as having the largest R^2 value (Equation (3-7)).

$$rank(k) = \left|\hat{d}_{t_j}\right| \quad \text{if} \quad \max_{l \in \hat{d}_{t_j}} R_l = R_k \tag{3-7}$$

The ranking number assigned to wavelength k is equal to the total number of peaks in $\hat{d}_{t_j}$. This wavelength k is then removed from the pool of peak wavelengths and the process is repeated on the new set $\hat{d}_{t_j} - k$ to yield the next ranked wavelength (rank number $\left|\hat{d}_{t_j}\right| - 1$). The process repeats until only one wavelength remains in the pool, which is assigned the highest ranking (rank number 1). The complete output of the IRP process (for the time point t_j) is then the ordered set of wavelength indices (Equation (3-8)). This IRP process is repeated for each time point to yield the final output as Equation (3-9).

$$\begin{gathered} \hat{r}_{t_j} = \text{the sequnece } \langle k_1, k_2, k_3, \dots \rangle \\ \text{where } rank(k_{l-1}) < rank(k_l), \forall k_l, k_{l-1} \in \hat{d}_{t_j} \end{gathered} \tag{3-8}$$

$$\hat{\boldsymbol{r}} = \langle \hat{r}_{t_1}, \hat{r}_{t_2}, \dots, \hat{r}_{t_M} \rangle \tag{3-9}$$

The workflow of IRP is presented in Figure 3-2. Initially, all peaks are put into a pool. In each round, the peak which has the strongest relationship with the remaining peaks in the pool is removed from the next round. The relationship is quantified by the R^2 value of linear regression of the given peak on the others. The whole process will end, if there is no peak left in the pool. Ranks of peaks are determined by the sequence in which peaks are removed from the pool. The later a peak is removed from the process, the higher rank it will have.

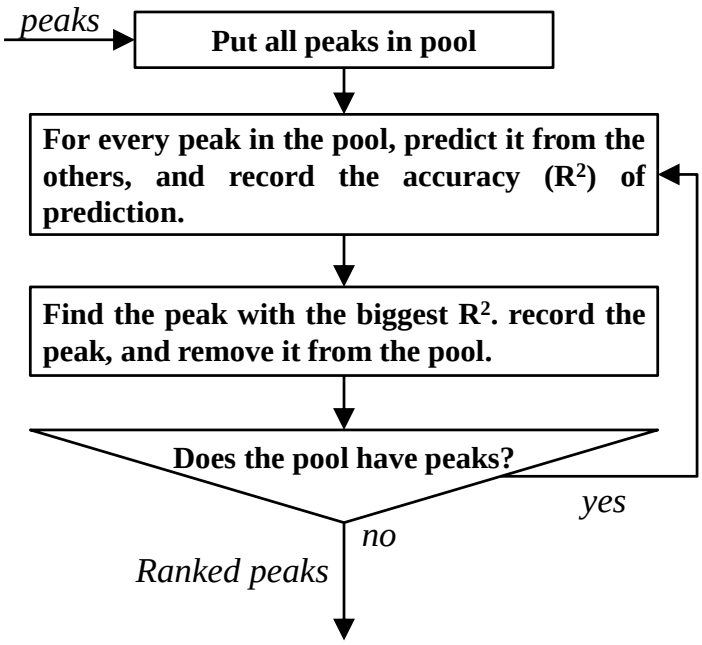

Figure 3-2. Workflow of IRP.

Another potential ranking strategy was also tested, in which a peak is not removed from the pool. So there are no multiple iterative rounds and all peaks can be ranked only based on R^2 values in one round. This was found to lead to an unreliable result. In a highly redundant dataset, most variables are very likely to having strong relationships with the rest when the number of variables is large. In that case, a lot of R^2 values could be very close to each other. It is even possible to have a lot of perfect R^2 with a value of 1. This phenomenon can be observed in the result of IRP in demonstration sections. The first few peaks removed have R^2 value 1. So it is risky to rank all variables only based on the very similar R^2. The multi-round IRP also has this problem in the first few rounds, but R^2 values will be distinguished from each other in the last few rounds after highly redundant variables are removed. Therefore, IRP could provide more reliable ranking results for high-ranking variables at least, compared with the other strategies.

Pseudo code of IRP is presented below.

Procedure IRP
Put all peaks in a pool
WHILE there are peaks in the pool
  FOR every peak in the pool
    Predict it from the other peaks in the pool;
    Record prediction accuracy (R^2);
  ENDFOR
  Record the peak with the largest R^2;
  Remove this peak from the pool;
ENDWHILE
Peaks are ranked based on the sequence in which peaks is removed from the pool;

3.2.3 Step 3 - Optimal Peak Selection

Having ranked peak wavelength variables using IRP, the OPS procedure selects a top-ranked subset of the peak variables. The number of peaks in this subset is minimized under the constraint that the prediction of all wavelength variables by the subset of peak variables meets a specified prediction quality target. Formally, let $\hat{r}_{t_j}$ be the ranked set of peak wavelength indices $\langle k_1, k_2, k_3, \dots \rangle$, for time point t_j, and let K' be the size of this set. For each $r \in \{1,2, \dots, K'\}$, the OPS procedure regresses wavelength intensity variable $I_k, k \in \{1,2, \dots, K\}$ on the set of peak wavelength variables $\langle I_{k_1}, I_{k_2}, \dots, I_{k_r} \rangle$, to yield prediction $\tilde{I}_k$, and calculates how well $\tilde{I}_k$ predicts I_k by way of the R^2 value denoted $R_{r,k}$. (Similarly to the IRP procedure, an ordinary least squares linear regression is used). An aggregate measure of how well the set of peak variables $\langle I_{k_1}, I_{k_2}, \dots, I_{k_r} \rangle$ predicts the full set of wavelength variables $\{I_k : k = 1,2, \dots, K\}$ is calculated, by way of the *Mean Determination Ratio* (MDR) metric, defined by Equation (3-10). MDR is created to quantify the information loss after dimension reduction, so different combinations of variables can be compared with each other via corresponding MDR values.

$$MDR_r \triangleq \frac{\sum_{k=1}^{K} R_{r,k}}{K} \tag{3-10}$$

Finally, an optimal value of r, denoted $\breve{r}_j$, is determined (as explained below), which selects the final reduced set of peak wavelength variables $\tilde{d}_{t_j} = \langle I_{k_1}, I_{k_2}, \ldots, I_{k_{\breve{r}_j}} \rangle$ for this time point t_j.

Empirically, I have found that as r is initially decreased from its maximum value of K', the prediction quality (MDR value) remains at a high value and decreases very slowly. Eventually, as r approaches 0, the MDR begins to drop off quickly. This pattern will be demonstrated with two test datasets D1 and D2 in the following sections. Thus, we have chosen to use a threshold on the slope of MDR_r to determine the optimal value $\breve{r}_j$ that gives a small variable set but with still high MDR value, that is, in the OPS procedure r is decreased from its maximum value until the following condition satisfied (Equation (3-11)).

$$(MDR_r - MDR_{r-1}) > m_{threshold} \tag{3-11}$$

$m_{threshold}$ is chosen to achieve a desired trade-off between prediction quality and the number of remaining peak variables. The above process is repeated for each time point to yield the final output of the IIRR procedure as $\tilde{\boldsymbol{d}} = \langle \tilde{d}_{t_1}, \tilde{d}_{t_2}, \ldots, \tilde{d}_{t_M} \rangle$.

Pseudo code of OPS is presented below.

Procedure OPS

Put all peaks in a pool.

WHILE MDR difference compared with previous one is smaller than the threshold
AND there are peaks in the pool

Predict every wavelengths intensity of original dataset from all peaks in the pool;

Calculate MDR value;

Remove the peak with the lowest rank from the pool;

ENDWHILE

IF there is no peak in the pool

Output the last peak in the pool;

ELSE

Output the peaks remaining in the pool;

ENDIF

3.3 Demonstration on Dataset D1

In this section, the IIRR method is demonstrated on the manufacturing OES dataset D1. A detailed description about D1 can be found in Section 2.4.1. A sample is defined as an OES time-series across a complete etching process. An etch rate measurement is also associated with every sample. The OES measurements are the only input of the model, while the etch rate is used for model validation. The validation assesses whether the model output keeps necessary information. The etch rate is a useful metric for monitoring etching process, but it is difficult to measure directly. Conversely, the OES measurements are relatively easier to attain from the etching process. By finding the relationship between the OES variables and the etch rate, multiple target applications can be accomodated, such as virtual metrology and fault detection of etch rate.

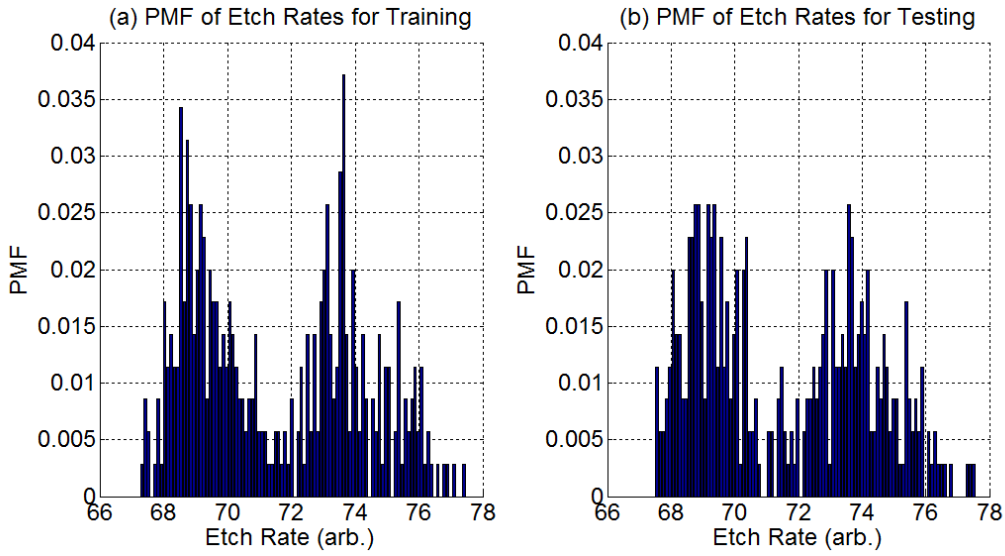

Figure 3-3. PMF of etch rates for (a) training group; (b) testing group for demonstrating dataset D1. Training and testing groups present similar distributions here which means that they includes similar types of samples.

All samples are divided into a training group and a testing group. Both groups have the same sample size, 450 samples. These two groups have similar etch rate distribution as shown in Figure 3-3. This training and testing organization strategy could make the two groups share similar OES features relevant to etch rate. Otherwise, missing features from testing to training would lead to failed modelling. Two dominant values of etch rates can be observed in the figure. Smaller etch rates are around 69 arb. units, while higher ones are around 74 arb. units. This difference is caused by a regular process of preventative maintenance in the plasma chamber during collection of the sample set.

3.3.1 Pre-processes

The previous description of the IIRR method is based on a generalized purpose of dimension reduction, but each particular dataset has its own unique features, and additional data preparation and post-processing of the data are required. In this section, two pre-processes are described. They are wavelength desaturation and time series normalization. These two pre-processes are relevant to two particular problems of

experimental industrial dataset D1 respectively: the wavelength saturation problem and variance problem of sampling time. Detailed description of these two problems can be found in the experimental data description section in Chapter 2.

3.3.1.1 Pre-process 1: Wavelength Desaturation

The wavelength desaturation process is designed to solve a problem caused by the limited range of measurement of the spectrometer. If wavelength intensity exceeds the maximum measurable value, the spectrometer only reads the maximum value. So this type of incorrect measurement is removed from the data.

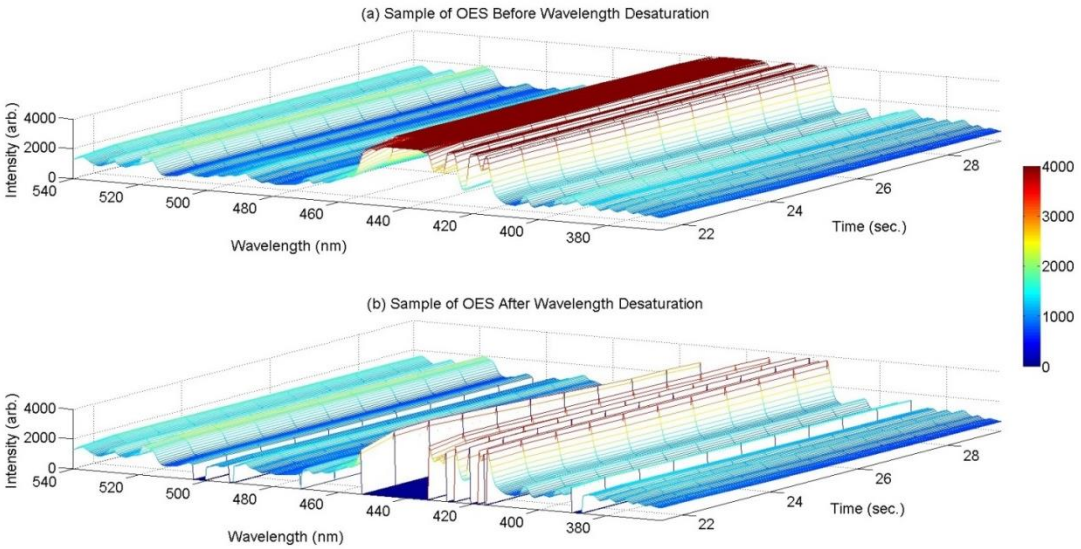

Figure 3-4. Illustration of OES measurements (a) before and (b) after desaturation for demonstrating dataset D1.

The result of wavelength desaturation is demonstrated on a partial OES data set in Figure 3-4. Saturated measurements can be observed as dark red flats on the top of Figure 3-4(a). It is risky to remove such measurements, because they also carry information. For example, such values could tell that actual intensity is too large to measure at least. However, it is proved by the final validation result (the high prediction accuracy of system parameters after dimension reduction) that such information is not important. The most reasonable explanation is that saturated

wavelengths are redundant to other wavelengths. This explanation is quite consistent with the fact that a chemical element can emit multiple wavelengths.

The proposed approach is simply to remove all wavelength variables at each time point that exhibit saturation, where the test for saturation is Equation (3-12). For each wavelength k and any sample, I_k is the intensity value under test, I_{max} is the maximum optical intensity of the sensor and B is the sensor bias which is estimated as Equation (*3-3*) with an *SNR* of 300:1 at full signal for the spectrometer, from [68].

$$I_k > I_{max} - B \tag{3-12}$$

Figure 3-4(b) shows the result of removal of the saturated intensity values. Over the full dataset, 241 of the 2048 wavelengths are removed.

3.3.1.2 Pre-process 2: Time Series Normalization

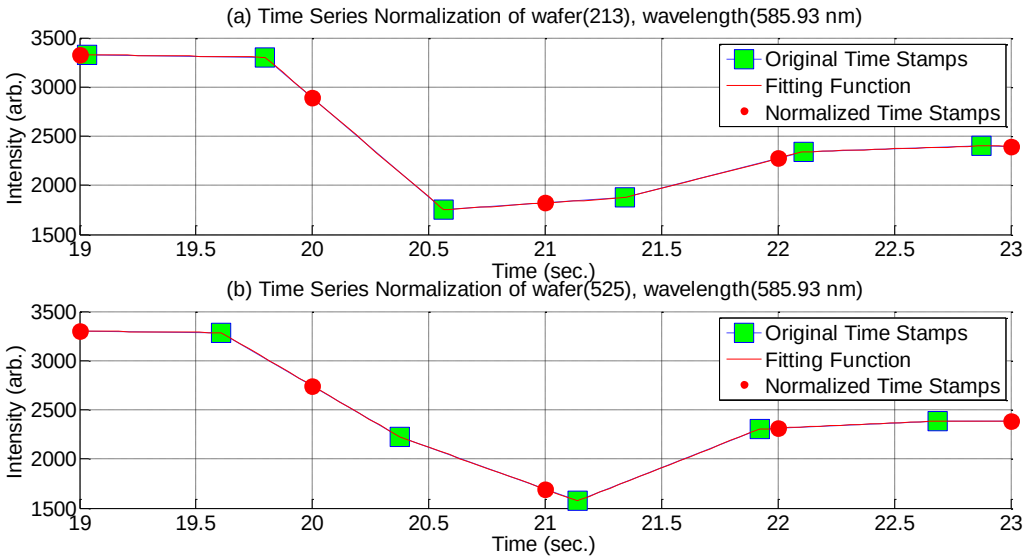

Figure 3-5. Original and normalized time series of wavelength (585.93 nm) from (a) sample 213 and (b) sample 525 for demonstrating dataset D1.

Each etching process run outputs a time series of spectral intensity scans, however, the sequence of timestamps from one process run to another is not necessarily identical. As the IIRR method needs to group all samples at a given time point during its data

processing stages, the timestamps need to be aligned to a normalized time scale. The time between samples averages approximately 0.7 seconds and, over all process samples, the minimum final time stamp is 40.14 seconds. We set the normalized time scale to have 1 second intervals, with the final timestamp at 40 seconds. Having set the time scale, the values in each time series (process run) are transformed by linearly interpolating the wavelength intensity values between the points either side of exact 1 second intervals. The process is illustrated in Figure 3-5 below with two representative samples from the data.

3.3.2 IIRR Output

3.3.2.1 Step 1 - Absolute Peak Selection

A peak selection result is demonstrated on one spectrum scan at one time point at 6 seconds in Figure 3-6. Originally, there are 2048 wavelengths from 178 nm to 874 nm. After the APS process, 48 outstanding peak wavelengths are marked in this figure. A similar APS process is applied independently on every spectrum scan at each normalized time point. Different time points have different peak sets. Physically, it is consistent with the etching process feature that the process is changing over time.

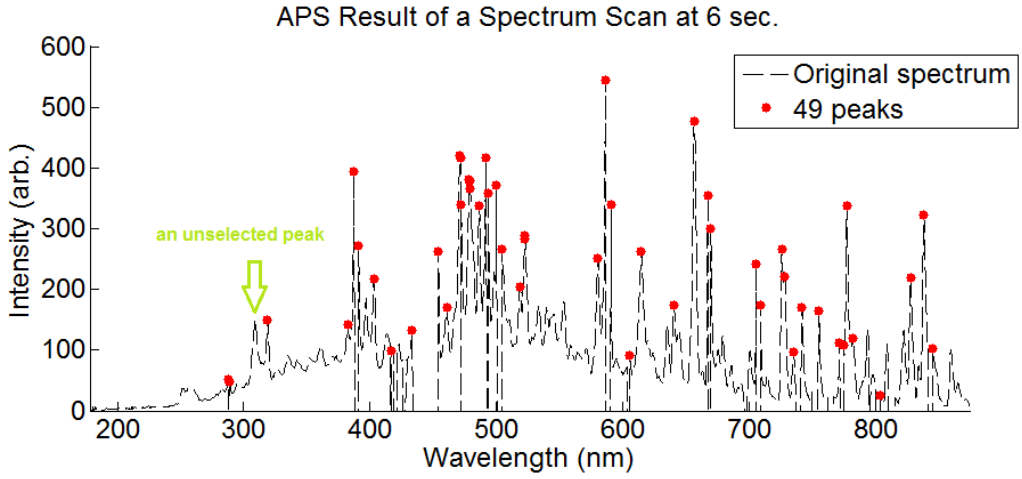

Figure 3-6. Illustration of the APS output for demonstrating dataset D1.

In Figure 3-6, some peaks which are apparent to human eyes are not picked out by the APS algorithm. For example, an unselected peak is identified by a green arrow in the plot. Due to the criterion (Equation (3-4)) for peak selection in APS, such a peak is not selected out because intensities of its neighbour wavelengths are not significantly lower than its own intensity. In other words, such a peak is not discriminative enough compared with nearby wavelengths. Potentially, this phenomenon could imply a potential limitation of APS that APS cannot pick up all peaks. Hence, the current APS method will be studied and optimized further in future. However, based on the successful dimensionality reduction result which will be discussed later, it is still safe to say that the current version of APS is sufficient for the task, at least for demonstrating dataset.

Time (sec.)	Wavelength Nu.		Time (sec.)	Wavelength Nu.
0 - 4	0		21	113
5	22		22	84
6	49		23	77
7	49		24	76
8	52		25	76
9	52		26	77
10	51		27	76
11	52		28	76
12	53		29	78
13	47		30	79
14	48		31	84
15	49		32	83
16	45		33	77
17	44		34	62
18	45		35	49
19	64		36	38
20	111		37 - 40	0

Table 3-1. Remaining peak wavelength numbers at each time point after the APS process for D1.

Generally speaking, APS reduces the number of wavelengths from 2048 to less than 100 at each single time point, as in Table 3-1. There are no peaks from 0 seconds to 4 seconds, nor from 37 seconds to 40 seconds. It is also consistent with a plasma etching

feature that the physical reaction is weak at the beginning and the end of the process. At the other time points, the minimum non-zero number of peaks is 22 and maximum number is 113.

3.3.2.2 Step 2 - Iterative Ranking Process

Peaks are ranked by IRP for each normalized time point respectively. IRP is an iterative process. The peak which has the strongest relationship with the remaining peaks is removed from the next round. Ranks of peaks are based on the sequence in which each peak is removed. Figure 3-7 represents the relationship between removed peaks in each round and the relevant remaining peaks in the pool. Results are demonstrated at four normalized time points: 6 seconds, 18 seconds, 20 seconds and 34 seconds respectively. This relationship is quantified by the R^2 value of linear regression between variables.

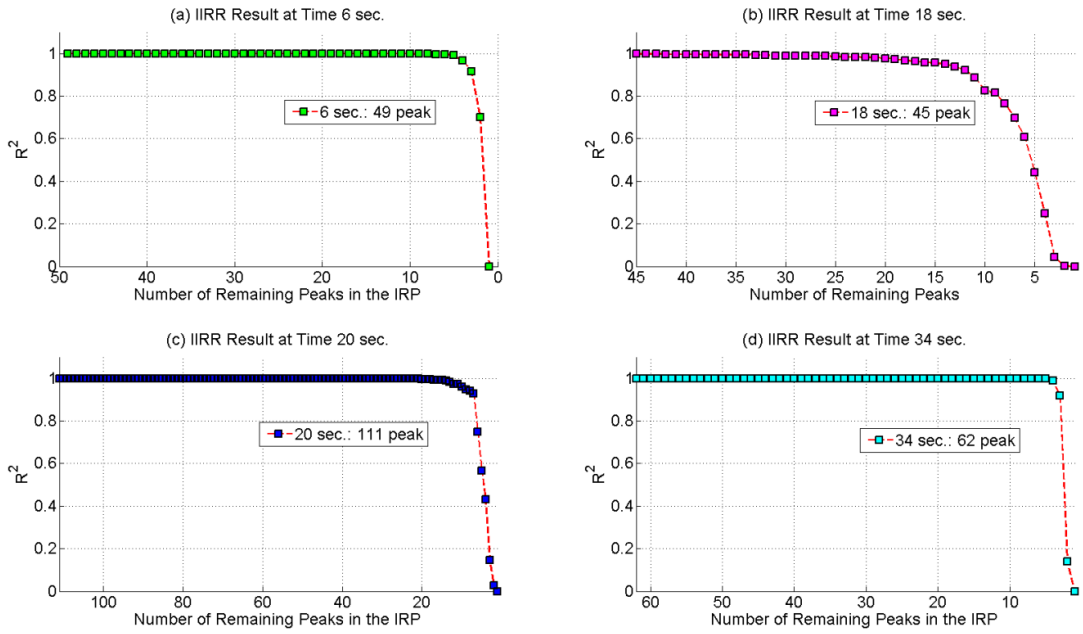

Figure 3-7. R^2 of regression of the remaining peaks on the removed peaks in each IRP iteration round (a) at time 6 seconds, (b) at time 18 seconds, (c) at time 20 seconds, and (d) at time 34 seconds, as a function of remaining peak numbers for demonstrating dataset D1.

All the other IRP results share similar shapes to those shown in Figure 3-8. Generally, for all results, the R^2 values are close to value 1 and do not change significantly, as the number of remaining peaks decreases at the beginning. This implies that peaks that are removed at the beginning is highly correlated with the rest of peaks. It also proves the existence of mathematical redundancy between wavelengths. Eventually, the R^2 value will drop quickly when the remaining peaks number is small. This suggests that peaks with high ranks have more independent information than the others.

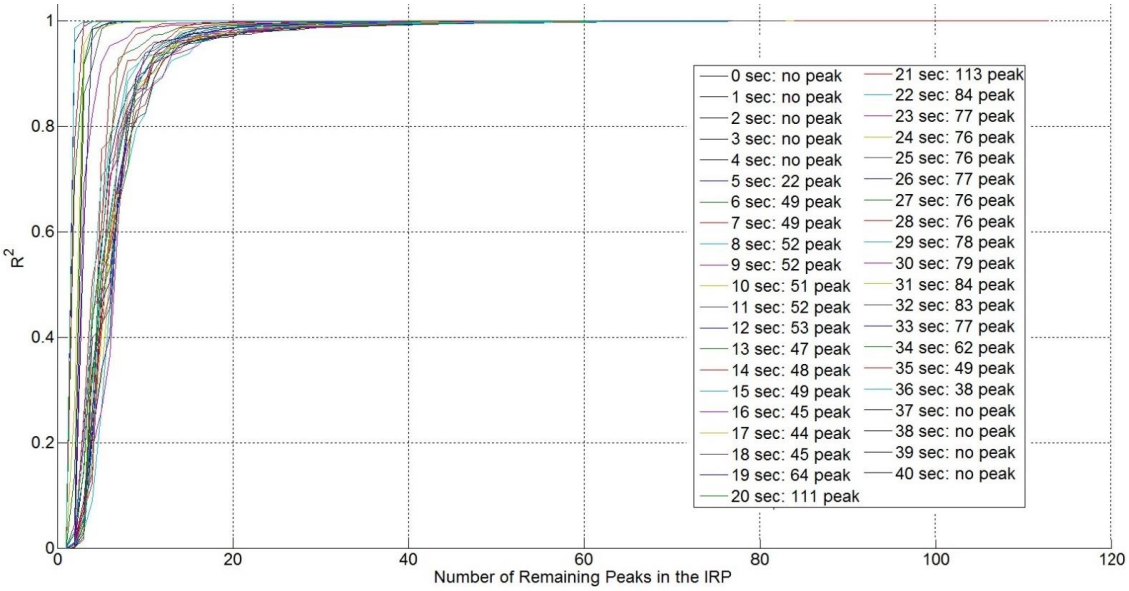

Figure 3-8. R^2 plot of removed peaks in IRP as a function of remaining peak numbers for demonstrating dataset D1.

3.3.2.3 Step 3 - Optimal Peak Selection

The MDR values and relevant MDR slopes with different sizes of candidate peak sets are presented in Figure 3-9 for all time points, based on the peak ranking list from IRP. The MDR remains at a high level and does not change significantly, when the number of peaks is relatively large. With the last few peaks, MDR drops significantly, as in Figure 3-9(a). The change in MDR is quantified by MDR slopes in Figure 3-9(b). By setting a slope threshold with a small value, peaks which do not have a significant impact on MDR can be removed. In order to give a clearer view, detailed results are

demonstrated at four normalized time points in Figure 3-10: 6 seconds, 18 seconds, 20 seconds and 34 seconds respectively. By setting the MDR slope threshold to 0.01, the first peak set whose MDR value is larger than the threshold is marked as a star. The relevant slope value is also drawn as a straight line in each subplot. The MDR slope threshold value of 0.01 is chosen empirically. If this value is too large, the model would contain too much irrelevant information. If this value is too small, the model would potentially lose some of important information.

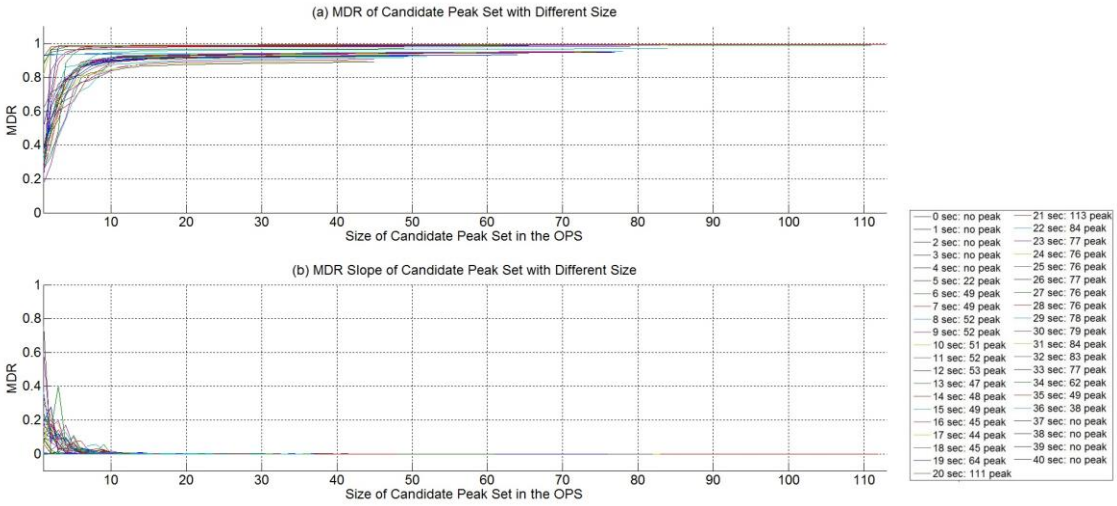

Figure 3-9. (a) MDR and (b) MDR slope plot as a function of remaining peak numbers for demonstrating dataset D1.

The number of remaining peaks at each time point is given in Table 3-2. 224 wavelengths are retained in total, of which there are only 64 distinguishable wavelengths. There is no peak outputted from the previous APS process from 0 to 4 seconds and from 37 to 40 seconds, so there are also no peaks outputted in this process for these time periods. In the other time periods, the number of peaks is reduced further by APS. The maximum number of peaks is 10. This maximum occurs at multiple time points: 10-11 seconds, 13-15 seconds, 23 seconds and 28 seconds. Minimum non-zero peak number is 1. This occurs at time points: 5 seconds and 36 seconds.

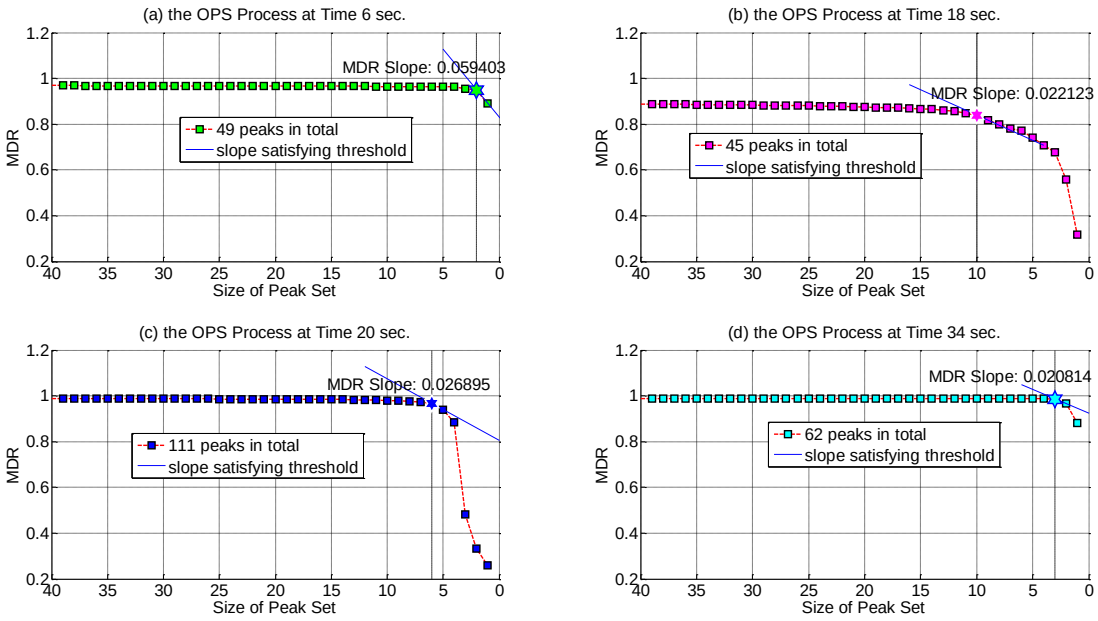

Figure 3-10. The MDR values of different sizes of candidate peak sets (a) at time 6 seconds, (b) at time 18 seconds, (c) at time 20 seconds, and (d) at time 34 seconds for dataset D1.

Time (sec.)	Number of Wavelengths		Time (sec.)	Number of Wavelengths
0 - 4	0		21	6
5	1		22	8
6	2		23	10
7	7		24	7
8	8		25	10
9	8		26	8
10	10		27	8
11	10		28	10
12	9		29	7
13	10		30	4
14	10		31	2
15	10		32	4
16	11		33	3
17	9		34	3
18	10		35	2
19	10		36	1
20	6		37 - 40	0
			Total	224

Table 3-2. Remaining numbers of peak wavelengths at each time point after the OPS process, for demonstration dataset D1.

In Figure 3-11, the number of wavelengths in the original dataset and resulting number of remaining wavelengths after the IIRR process are illustrated with spectrum scans at

four time points: 2 seconds, 8 seconds, 14 seconds, and 36 seconds. After the APS process of the IIRR procedure, all wavelength numbers decrease significantly from the original 2048 wavelengths to about 60 wavelengths for these four samples. After the process of IRP and OPS as the last two steps, the wavelength numbers decreased further to less than 10 wavelengths. Concerning measurements over all time points, APS only retains on average 47.7073 peaks of the 2048 total wavelengths. After further IRP and OPS processes, the number of peaks is reduced to 5.4634 on average.

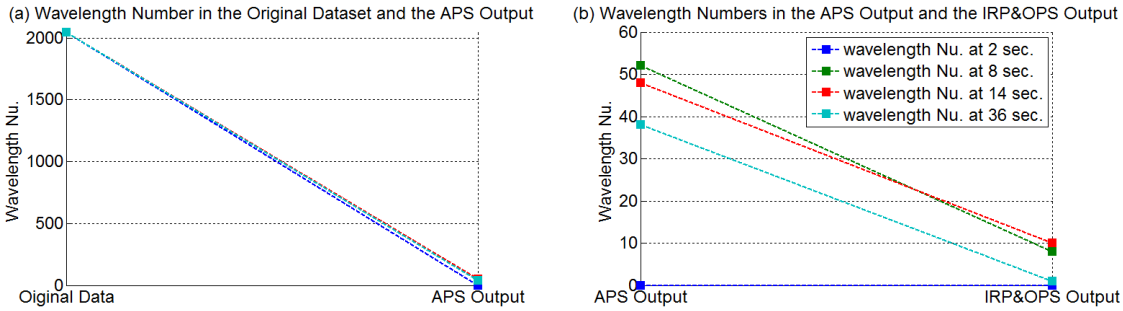

Figure 3-11. Number of wavelengths of original data, the APS output, and IRP+OPS output for dataset D1.

3.3.2.4 Final Output

Information redundancy is reduced by IIRR for each time point. Finally all important peaks from different time points are combined together. This is the final output of the whole IIRR process. A sample of OES measurements before and after the IIRR process is presented in Figure 3-12. Dimension is reduced significantly. Based on dimension reducing results at all time points, only about 0.2668% wavelength measurements are retained in the end. A list of these important wavelengths could potentially help identify underling chemical species which play the most important roles at each time point during the etching process.

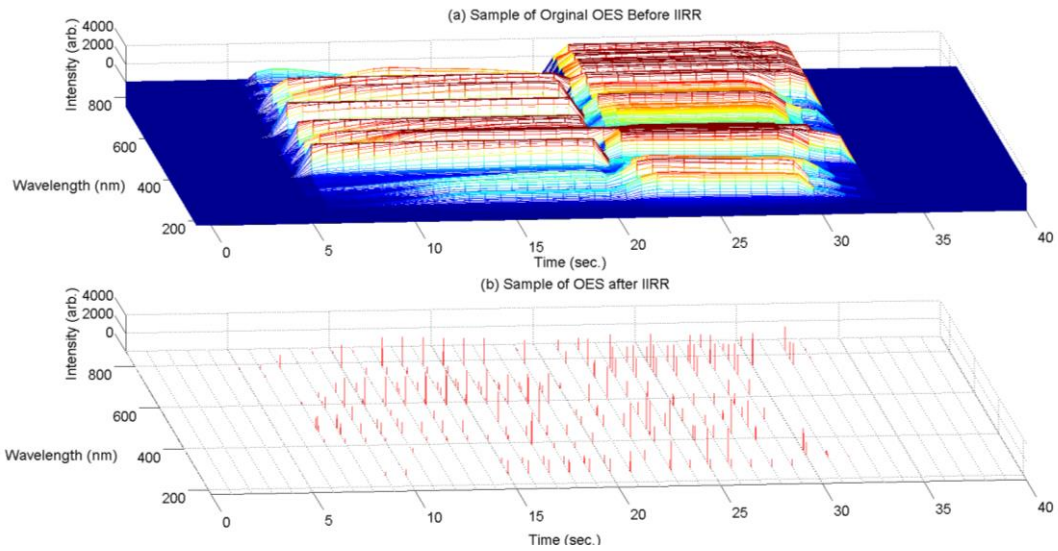

Figure 3-12. (a) Sample of original OES, (b) sample of peaks after IIRR for a complete etching process in dataset D1.

3.3.3 Validation

In order to assess whether IIRR retains the most useful information from the original dataset, etch rate is predicted from the original full OES data and the IIRR reduced output data separately. If similar prediction accuracy can be achieved, it would be safe to conclude that IIRR keeps enough information for etch rate prediction, at least. Wavelengths (or peaks) from different time points are treated as independent parameters for prediction. Three popular prediction methods are tested to provide a regression-independent result. These methods are MLR, PLS, and ANN.

Relevant statistical results are presented in Table 3-3. The MAPE values are also provided, as well as R^2 values. MAPE is quoted as a percentage, which is more direct and easier to understand than R^2. As shown by Table 3.3, all three measures provide very high prediction accuracy using IIRR reduced data. The prediction accuracy is in fact improved by IIRR, compared with the original dataset, as measured by MLR and PLS. With MLR, the IIRR prediction accuracy achieves an R^2 value 0.9439 on the testing dataset, higher than the R^2 value of 0.9329 achieved using the original dataset. With PLS, the IIRR prediction accuracy achieves an R^2 value of 0.9705 on the testing

dataset, compared to a lower R^2 value of 0.9676 when using the original dataset. This is better than the result of MLR, which is most likely due to the fact that PLS is generally more suitable than MLR when the data dimension is large and there is high redundancy [71].

Regression Method	With/out IIRR	Train		Test	
		R^2	MAPE	R^2	MAPE
MLR	IIRR	0.9930	0.0024	0.9430	0.0070
	Original	0.9944	0.0021	0.9329	0.0074
PLS	IIRR	0.9802	0.0041	0.9705	0.0051
	Original	0.9805	0.0041	0.9676	0.0053
ANN	IIRR	0.9710	0.0042	0.9049	0.0084
	Original	Too many input variables to calculate			

Table 3-3. Etch rate prediction accuracy comparison between original OES and IIRR output with a minimum MDR threshold, for dataset D1.

The PLS prediction result is presented for the training dataset and testing dataset separately in Figure 3-13. With ANN, the IIRR prediction accuracy also achieves a high R^2 value of 0.9049. There is no prediction result given for ANN with the original full OES, because the number of input parameters is too large to process with the ANN algorithm.

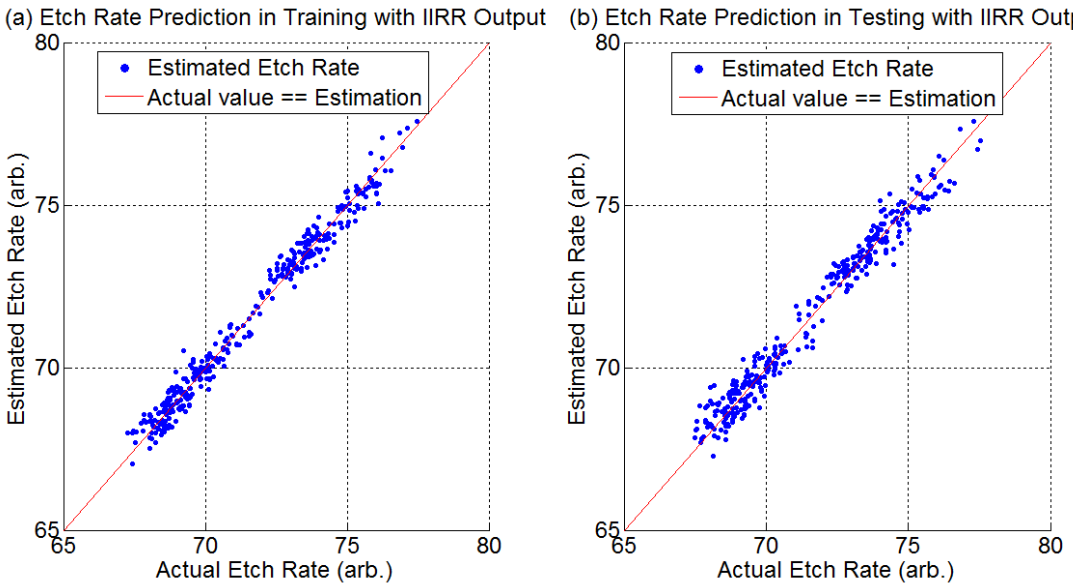

Figure 3-13. Etch rate prediction with PLS for dataset D1.

3.4 Demonstration of Methods for Dataset D2

In this section, the IIRR method is demonstrated for the second manufacturing OES dataset D2. A detailed description about D2 can be found in Section 2.4.2.

In dataset D2, instead of using etch rate prediction to validate the IIRR output, BLW measurements of write pole are used. BLW is another important feature which affects the IC chip quality significantly. Ideal etching should be anisotropic (vertical), but the actual process also etches into the layer wall horizontally, which impacts BLW of resist layers. Therefore, BLW is a highly relevant and important effect in plasma etching processes. Based on two different target BLW values (0.112 μm and 0.114 μm), 14 samples with the same target BLW value of 0.112 μm are used here for demonstration. All 14 samples are divided equally into a training group and a testing group. Both groups have 7 samples. Corresponding BLW values are presented in Figure 3-14 for training and testing groups respectively. Similarly to etch rate in dataset D1, the BLW data is used only for the IIRR validation in dataset D2, but is not part of the input dataset for the IIRR reduction procedure. A more detailed description of dataset D2 can be found in the background Chapter of this thesis.

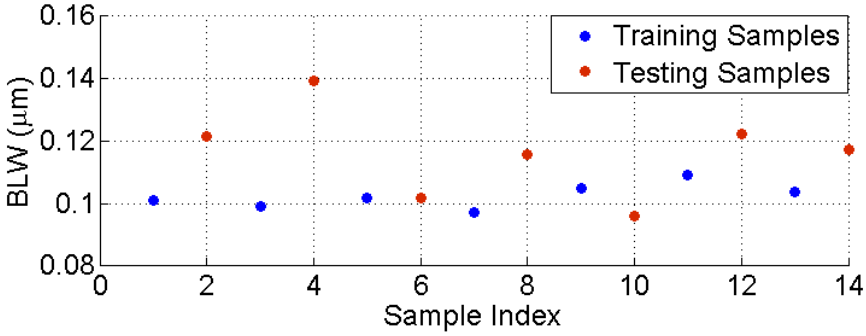

Figure 3-14. BLW measurements of write pole for training group and testing group for dataset D2.

3.4.1 Pre-process

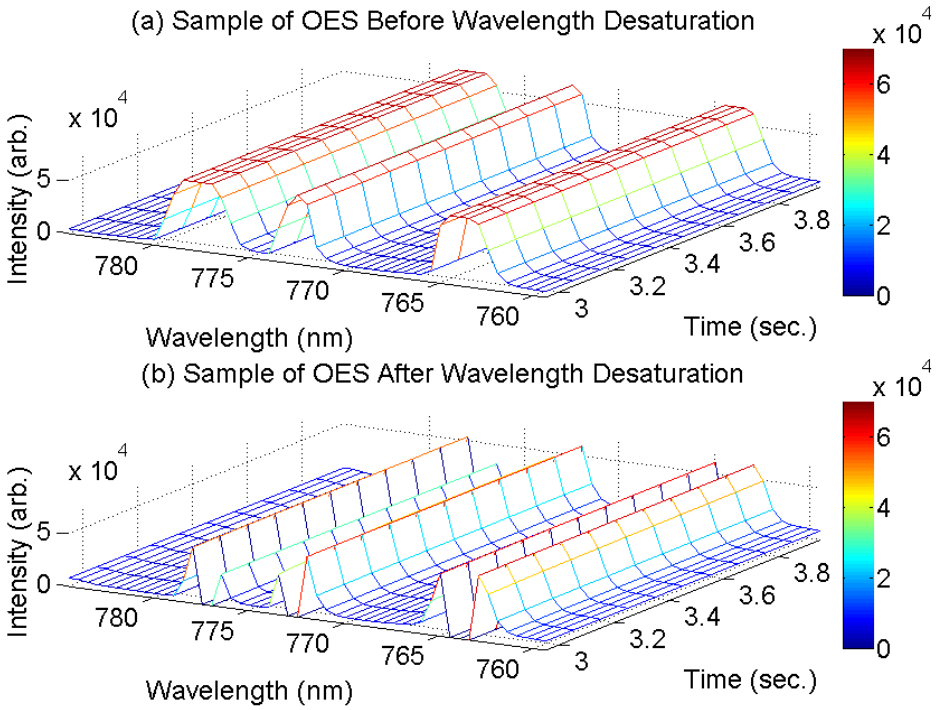

Figure 3-15. Illustration of OES measurements (a) before and (b) after desaturation for dataset D2.

Similarly to dataset D1, the wavelength saturation problem also exists in this dataset. However, D2 does not require time series normalization, as the spectrum scan of all samples starts at 0.05 seconds and has a fixed 0.10 second time-interval.

The result of pre-process desaturation step is illustrated in. Partial time series OES measurements without the wavelength desaturation process are presented in Panel (a) of Figure 3-15. In Panel (b) of Figure 3-15, saturated wavelengths are removed from the raw data.

3.4.2 IIRR Output

3.4.2.1 Step 1 - Absolute Peak Selection

The result of peak selection is demonstrated for one spectrum scan at time point 0.05 seconds in Figure 3-16. Originally, there are 1201 wavelengths spanning 200 nm to

800 nm. After the APS process, 38 outstanding peak wavelengths are retained. Similarly, the APS process is applied independently to every spectrum scan at each time point for both OES1 (the first etching phase in process) and OES2 (the combined second and third etching phases in process).

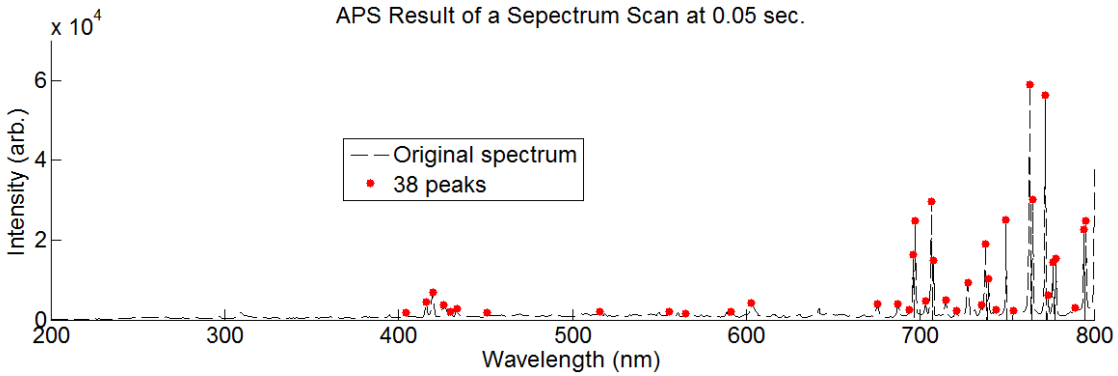

Figure 3-16. Illustration of the APS output on one spectrum scan for dataset D2.

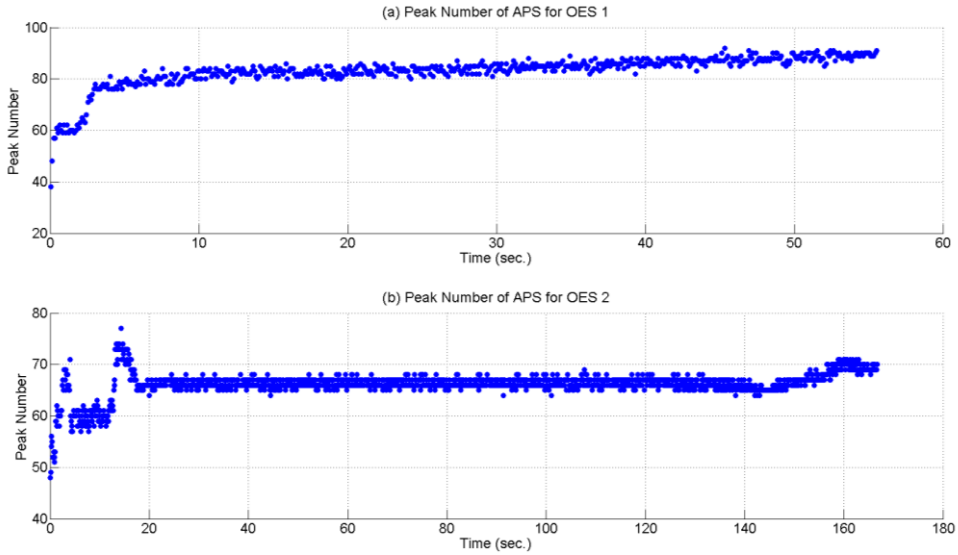

Figure 3-17. Number of peaks at all time points after APS process for (a) OES1 and (b) OES2, for dataset D2.

Considering all of the other time points, APS reduces the number of wavelengths from 1201 to less than 100 at a single time point for etching step 1 (OES1), from 1201 to less than 80 peaks for etching step 2+3 (OES2). Relevant results are presented in Figure 3-17 for all time points from the training dataset. Within these peaks, only 117 different wavelengths could be found in OES1 and 99 different peaks wavelengths in

OES2. Combining results from both OES1 and OES2, 140 different wavelengths are identified by the APS process.

3.4.2.2 Step 2 - Iterative Ranking Process

Similarly to the demonstration for dataset D1, Figure 3-18 represents the relationship between removed peaks in each round and the remaining peaks in the pool.

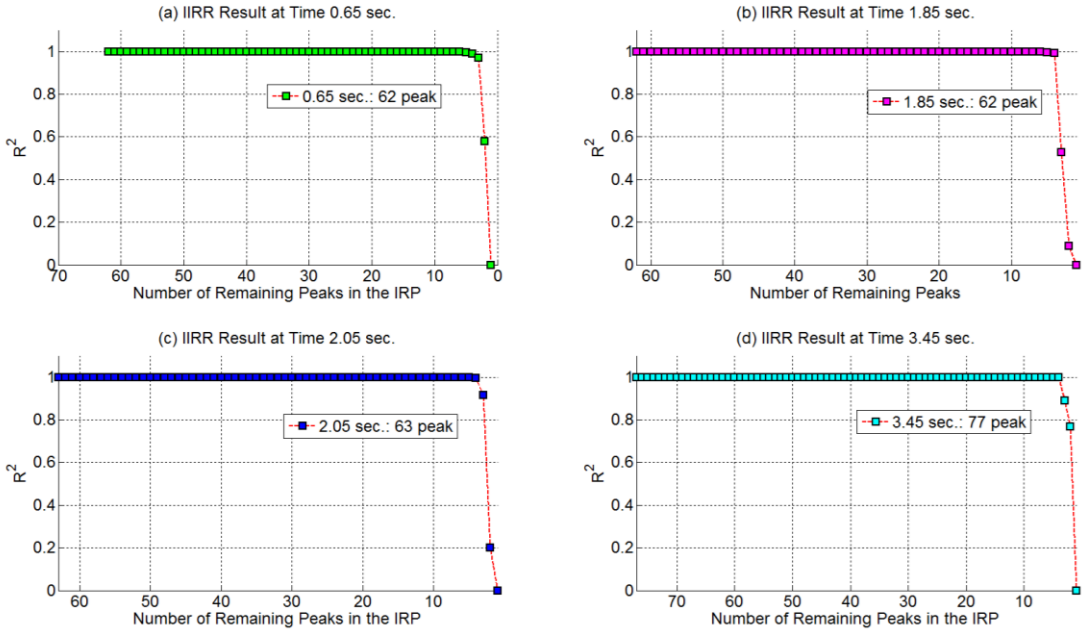

Figure 3-18. R^2 of regression of the remaining peaks on the removed peaks in each IRP iteration round (a) at time 0.65 second, (b) at time 1.85 seconds, (c) at time 2.05 seconds, and (d) at time 3.45 seconds, as a function of remaining peak numbers for demonstrating dataset D2.

Results are demonstrated at four time points: 0.65 seconds, 1.85 seconds, 2.05 seconds, and 3.45 seconds from OES1 respectively. All the other IRP results share similar shapes, shown for OES1 and OES2 in Figure 3-19. Each curve represents a set of spectrum scans at the same time point. All curves share similar shapes. R^2 values are almost equal to value 1 when there are more than about 5 remaining wavelengths in the pool. This implies a very high information redundancy between removed wavelengths and remaining wavelengths. When the number of remaining wavelengths is less than about 5, there is a significant decrease in the R^2 value for all curves. This implies a low

information redundancy between removed wavelength and remaining wavelengths at this stage. Therefore, it is reasonable to rank wavelengths which are removed earlier with lower scores, and give high scores to the last few wavelengths. This result is consistent with the conclusion for dataset D1.

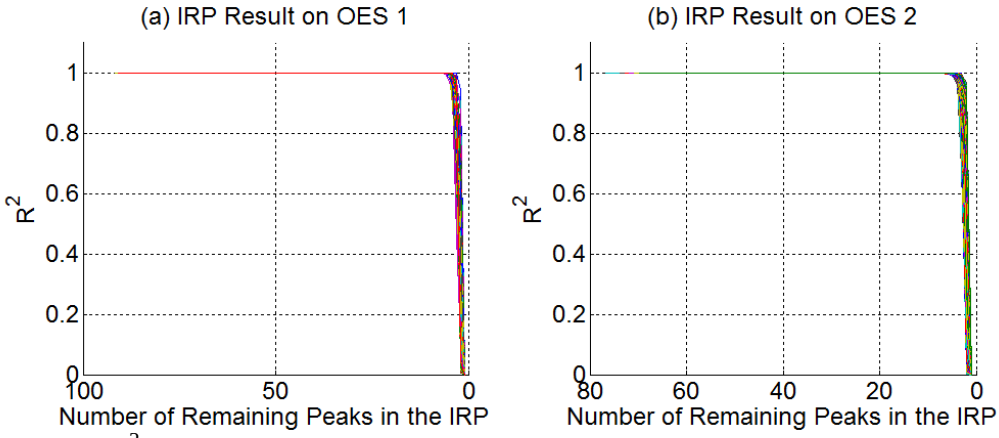

Figure 3-19. R^2 plot of removed peaks in IRP as a function of remaining peak numbers for demonstrating dataset D2 on (a) OES1 from etching step 1 and (b) OES2 from etching step 2&3.

3.4.2.3 Step 3 - Optimal Peak Selection

The MDR values and relevant MDR slopes with different sizes of candidate peak sets are presented in Figure 3-20, for all time points from OES1 and OES2, based on the peak ranking list from IRP. MDR remains at a high level and does not change significantly, when the number of peaks is relatively large. With the last few peaks, the MDR value drops significantly, as in Figure 3-20(a) for OES1 and (c) for OES2. The trend in MDR indicated by the MDR slope, in Figure 3-20(b) for OES1 and (d) for OES2. By setting a suitable threshold with a small value on the slope, peaks which do not have significant impact on MDR can be removed.

In order to give a clearer view, detailed results are demonstrated at four individual time points, in Figure 3-21: 0.65 second, 1.85 seconds, 2.05 seconds, and 3.45 seconds from OES1. By setting the MDR slope threshold to 1.00e-5, the first peak set whose MDR value is larger than the threshold is found and marked as a star in the plot. This peak

set is the OPS output, as well as the IIRR final output. The relevant slope value is also drawn as a straight line in each subplot. According to the particular results at these four time points, all of them have five peaks selected out. Due to this small number of peak wavelengths, any further analysis will not be difficult, such as interpretation of wavelengths with respect to chemical species active in the plasma.

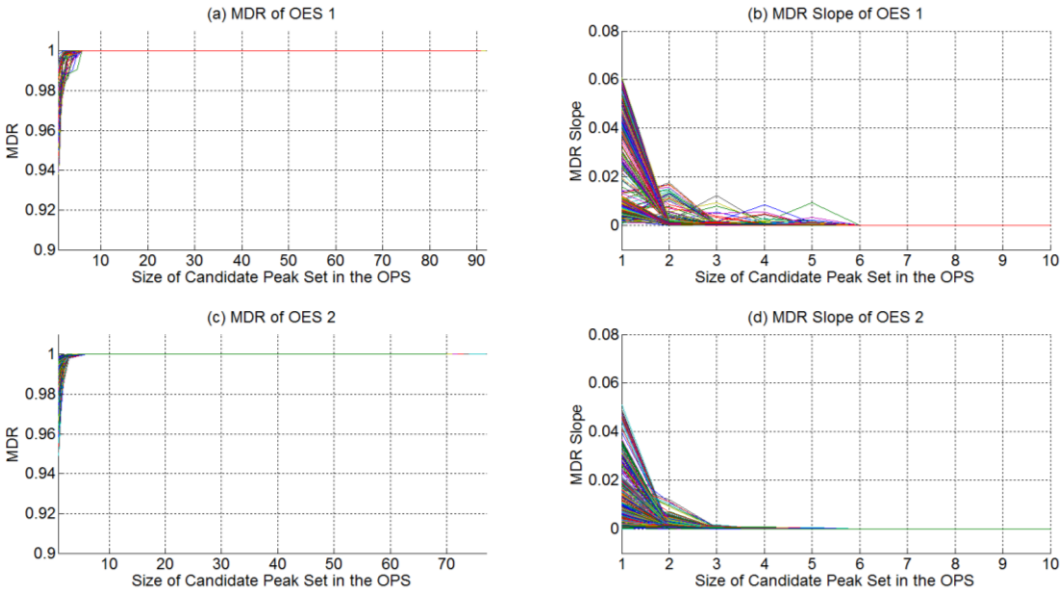

Figure 3-20. (a) MDR and (b) MDR slope plot as a function of remaining peak numbers for dataset D2.

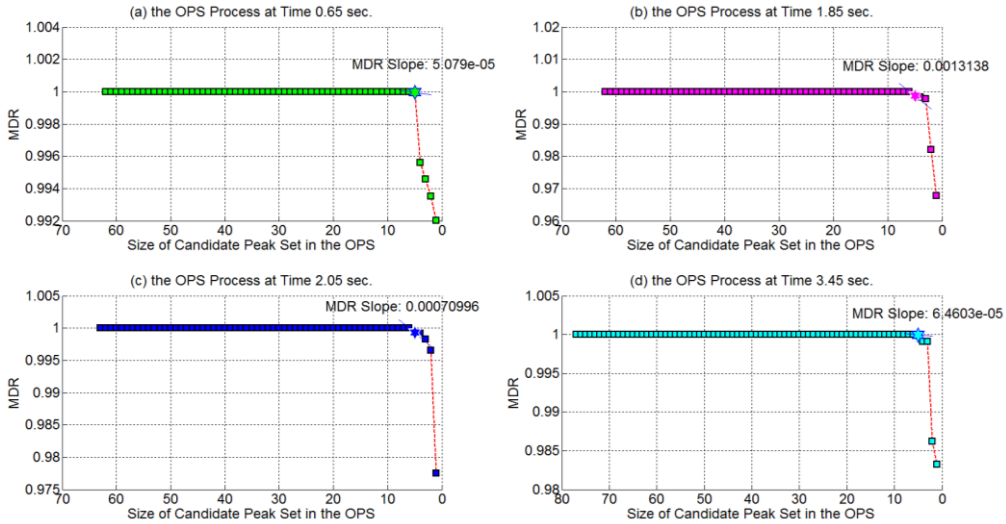

Figure 3-21. The MDR values of different size of candidate peak sets (a) at time 0.65 seconds, (b) at time 1.85 seconds, (c) at time 2.05 seconds, and (d) at time 3.45 seconds, as a function of remaining peak numbers for dataset D2.

3.4.2.4 Final Output

Information redundancy is reduced by IIRR for each time point respectively. Finally the outputted peaks from different time points are combined together, the same as the IIRR process for D1. A sample of OES measurements before and after the IIRR process is presented for OES1 (etching step 1) and OES2 (etching step 2&3) separately in Figure 3-22. Dimension is reduced significantly. Based on dimension reduction results at all different time points, only about 9 of 2048 wavelength measurements (0.4200%) remain at the end for both OES1 and OES2.

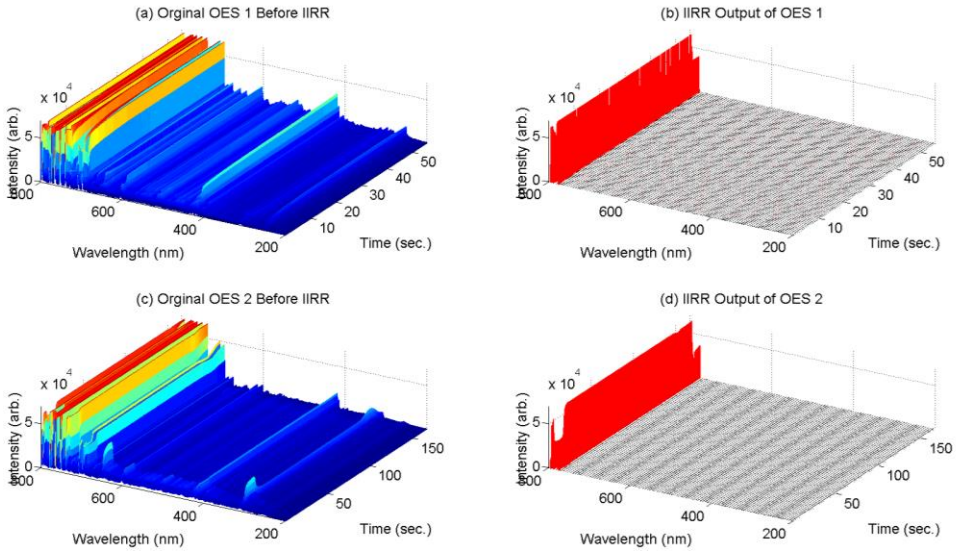

Figure 3-22. OES1 from etching step 1 illustration (a) before IIRR and (b) after IIRR. OES 2 from etching step 2&3 illustration (c) before IIRR and (d) after IIRR for demonstrating dataset D2.

All remaining wavelengths are listed in Table 3-4. This wavelength list could potentially help identify underling chemical species which play the most important roles at different etching steps.

	OES1	OES2
Wavelength List (nm)	773.50	772.00
	776.50	773.50
	778.00	776.00
	780.00	779.50
	786.50	786.50

Table 3-4. Wavelength list of the final IIRR output for demonstrating dataset D2.

3.4.3 Validation

For validation of the IIRR output, BLW is predicted from the original, full D2 OES dataset and the IIRR reduced output separately. As for dataset D1, wavelengths (or peaks) from different time points are treated as independent parameters for prediction for dataset D2. Unlike the demonstration on dataset D1, only two prediction methods (MLR and PLS) are tested, while ANN is excluded from this validation because the number of spectrum scans in each sample is too large an input for ANN. There are 556 spectrum scans from OES1 and 1668 spectrum scans from OES2, and each spectrum scan has intensity measurements at 1201 different wavelengths.

Relevant statistical results are presented in Table 3-5. MAPE and R^2 values are provided. Using MLR and PLS, high prediction accuracy is achieved for the training dataset with or without the IIRR processing of data, but low prediction accuracy is obtained from the testing dataset. The average MAPE is about 0.1489 for all cases with or without IIRR process. Compared with D1 which has 900 samples, D2 only has 14 samples in total, and 7 samples are used for training while the other 7 samples are used for testing. Each sample has intensity measurements on 1201 different wavelengths. Therefore, the number of variables for prediction is much larger than the number of samples. The over-fitting problem can occure in this situation, which could be a reasonable explanation for this poor prediction result. However, there is still no significant loss on prediction accuracy by using IIRR, even the absolute accuracy is not good. With MLR, the IIRR prediction accuracy achieves R^2 a value of -1.1094 on the testing dataset, which is very close to the R^2 value of -1.0346 achieved by the original dataset. With PLS, the IIRR prediction accuracy achieves an R^2 value of -2.4909 on the testing dataset, also very close to the R^2 value of -2.4960 achieved by the original dataset.

Regression Method	With/out IIRR	Train		Test	
		R^2	MAPE	R^2	MAPE
MLR	IIRR	0.9330	0.0081	-1.1094	0.1342
	Original	0.9806	0.0043	-1.0346	0.1292
PLS	IIRR	1.0000	5.4490e-15	-2.4909	0.1657
	Original	1.0000	7.9355e-14	-2.4960	0.1665

Table 3-5. BLW prediction accuracy comparison between original OES and the model output with a minimum MDR threshold for demonstrating dataset D2.

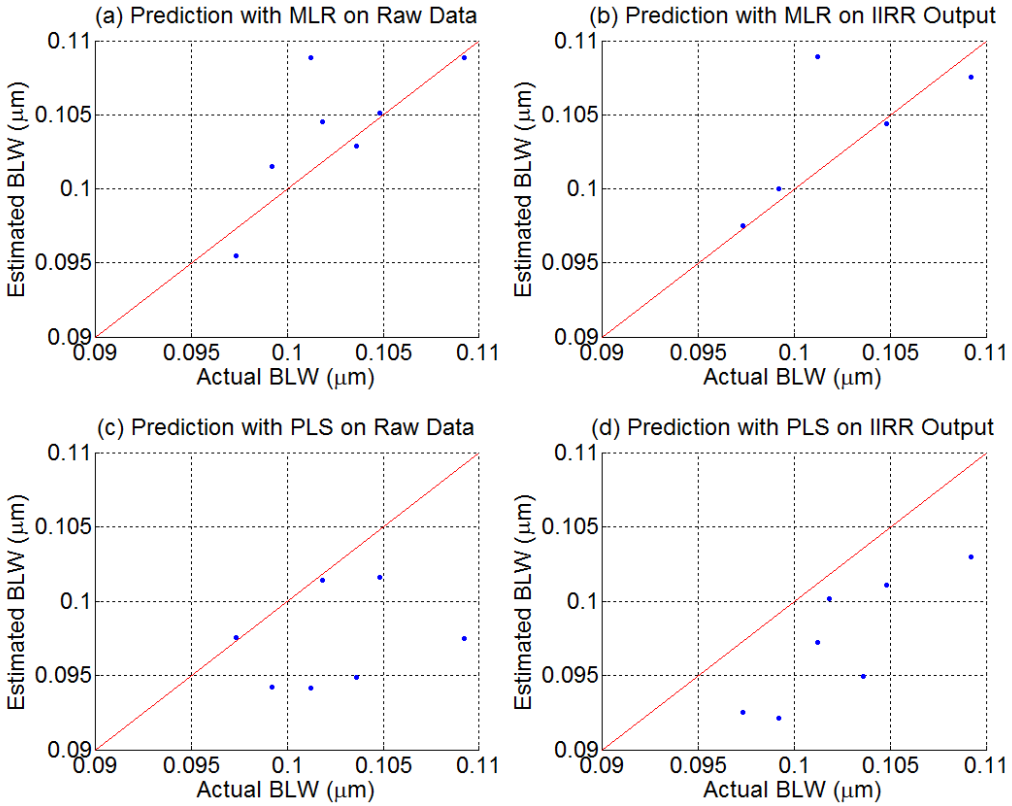

Figure 3-23. BLW prediction for testing dataset. Panel (a) represents MLR result on raw data without IIRR process. Panel (b) represents MLR result on IIRR output. Panel (c) represents PLS result on raw data without IIRR process. Panel (d) represents PLS result on IIRR output for demonstrating dataset D2.

The MLR and PLS prediction results are also presented on testing dataset separately in Figure 3-23. Distribution of all predicted values is similar between the output of the original dataset and the output of IIRR. Therefore, it can be concluded that the IIRR does not have information loss for BLW prediction, at least on dataset D2.

3.5 Discussion & Conclusion

The new dimension reduction method IIRR is presented and evaluated in this chapter. The method targets multivariable OES datasets in the IC fabrication etch process and it is demonstrated using two different manufacturing datasets D1 and D2.

In order to assess whether IIRR retains useful information after dimension reduction, the original OES and the IIRR outputs are used separately to predict etch rates for dataset D1 and predict BLW for dataset D2 separately. Three different prediction methods are used to provide a regression-independent result, MLR, PLS, and ANN. For D1, all three of them provided very high prediction accuracy using the IIRR reduced dataset. Prediction R^2 values are larger than 0.90 in all cases. IIRR also improved the MLR and PLS prediction accuracy, compared with the original dataset. So it can be concluded that useful information is kept and noise is removed by dimension reduction in IIRR. This conclusion is reliable at least for etch rate prediction in this case. For D2, there is still no significant loss on prediction accuracy by using IIRR, even though the absolute prediction accuracy is not good with either MLR or PLS. The poor prediction accuracy is expected, because there are only 14 samples in this dataset, too few to build a reliable prediction model. ANN is not tested with D1, because the number of input parameters is too large to process using a standard PC. It also can be concluded safely that IIRR does not have information loss for BLW prediction at least on dataset D2. As for results from both D1 and D2, there is no information loss from IIRR for etch rate prediction in D1 and BLW prediction in D2.

Compared with the other research work (e.g. based on PCA, ICA, etc.), IIRR operates directly in the original variable space, identifying peak wavelength emissions and the correlative relationships between them. Instead of creating new variables, as is usual in

other methods in the literature, IIRR picks up a subset of the original variables. This facilitates identification of important variables. Additionally, the result could be extended easily to different applications. For example, the dimension of OES measurements could be reduced further after IIRR, if the relationship with etch rate is considered as part of the input dataset. The new low-dimensional output would then be more tailored to system monitoring of etch rate.

In order to provide a more comprehensive evaluation of IIRR, it is compared with two popular dimensionality reduction methods: PCA and FSCA. All of them are implemented for dataset D1. In the PCA implementation, Kaiser's stopping rule (components whose latent variables are larger than 1 are retained) is used to identify up principal components. In FSCA, components are selected based on a cumulative variation explained [47] value of 0.99, that is the first few components whose cumulative variance exceed 0.99 are retained.

Dimensionality reduction results, and the corresponding validation results (etching rate prediction with MLR), are presented in Table 3-6. IIRR retains the smallest number of variables, and gives the best etching rate prediction result (R^2 value of 0.9430) on the testing data.

Method	Remaining Ratio after Dimension Reduction	Train		Test	
		R^2	MAPE	R^2	MAPE
Original	100.0000%	0.9944	0.0021	0.9329	0.0074
PCA	14.1935%	0.9954	0.0017	0.9410	0.0067
FSCA	24.8011%	0.9956	0.0018	0.9260	0.0079
IIRR	0.2668%	0.9930	0.0024	0.9430	0.0070

Table 3-6. Comparison of etching rate prediction models with MLR based on original OES and different dimensionally reduction methods respectively for demonstrating dataset D1.

It would always be better if IIRR could be tested on more datasets, in order to prove its stability and consistency. However, the datasets used in this thesis come from real

manufacturing processes, which are normally difficult to obtain because of a considerations of commercial confidentiality, so this thesis can only focus on these datasets. However, since IIRR works successfully on two different datasets which come from different etching processes, different sensor models, and different companies, it is still relatively safe to conclude that IIRR provides a stable and consistent solution for the dimensionality problem in OES datasets.

CHAPTER 4. Similarity Ratio Analysis

4.1 Introduction

Fault detection (FD) is a very important research objective highly relevant to the performance of the IC fabrication control systems. A successful FD mechanism can improve the IC fabrication process quality and decrease equipment downtime with small cost [8]. However, traditional FD technologies have two common problems: potentially high cost and long-time delay before a fault is detected. Because of real-time monitoring and non-intrusion features, OES monitoring is widely used for FD as a potential replacement for traditional FD technologies. However, most current research only realizes FD after finishing the whole etching process. Few methodologies are able to detect faults at an early-stage with a single spectrum scan. However, there is some common limitation associated with these methods. For example, existing early-stage FD methods [57] can only take spectrum scans at certain pre-defined time points, which impacts the flexibility of the methods negatively.

In order to address these problems, a novel Similarity Ratio Analysis (SRA) method is proposed in this thesis. Compared with previous research, the SRA method can detect faults at an early stage of the etching process. The OES data is also not limited to certain time points by the SRA, so the fault detection system can be more flexible because the timetable for the OES sampling can be adjusted based on real-time system feedback. System faults are detected by checking a new statistic Similarity Ratio (SR) produced with certain conditions. A confidence level is also provided as a reference for the system alarm triggered. According to the results using the example dataset (the real manufacturing OES dataset D1), this method can give an alarm for a system fault at about the 6^{th} second in a 45 second faulty etching process.

4.2 Method

The SRA method is based on a supervised training framework. A training dataset is used to build a library for healthy samples. In the context of plasma etching, the sample is defined as a time series of the OES spectrum scan for a complete etching process on one wafer. In real time SRA fault detection, a single testing spectrum scan is compared with the library at relevant time points and corresponding SR values are calculated. Then a fault detection mechanism is implemented based on the SR values.

In the rest of this section, the basic SRA framework is discussed first, followed by the description of the SRA method for fault detection.

4.2.1 Library Calculation

The SRA library is created to capture common features shared by all healthy samples. Each OES spectrum has intensity values at the same set of wavelengths, so a boundary function is computed for each wavelength individually. A detailed workflow of the SRA library building is presented in Figure 4-1.

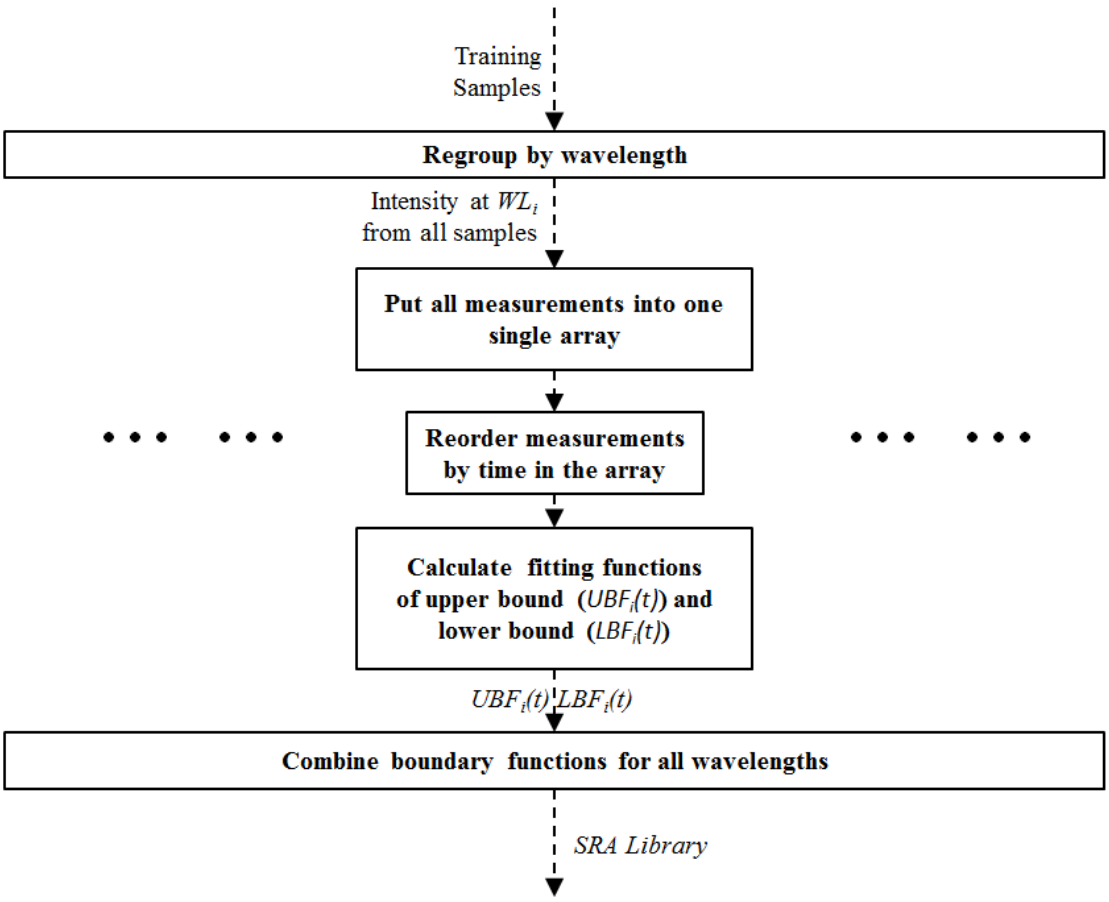

Figure 4-1. Workflow of building the SRA library in SA framework.

The boundary function describes the intensity range of healthy samples at certain wavelengths against time. Each wavelength has a pair of fitting functions to describe its upper boundary and lower boundary. All these functions of different wavelengths comprise the SRA library, as Equation (4-1). $UBF_i(t)$ and $LBF_i(t)$ represent the upper boundary function and the lower boundary function over time t for the i^{th} wavelength, respectively. The total number of wavelengths is n.

$$SA\_Library = \begin{bmatrix} UBF_1(t) & LBF_1(t) \\ UBF_2(t) & LBF_2(t) \\ \cdots & \cdots \\ UBF_n(t) & LBF_n(t) \end{bmatrix} \quad (4\text{-}1)$$

Detailed calculation of boundary functions UBF_i and LBF_i for a single wavelength is illustrated in Figure 4-2. In our example dataset D1, each sample has a unique

timetable to take spectrum scans, but all training samples derive from the same type of healthy etching process. Artificial data from only two samples is presented in the figure to provide a clear description of the method. Spectrums from different samples are re-ordered by time. A moving window is applied and average intensity μ and standard deviation σ, for each wavelength, are obtained within each window. Upper and lower boundary values are calculated at $\mu \pm m\sigma$. m is the standard score used which is determined by using a SR threshold (*FD_Threshold*) for fault detection from a given training dataset (more detailed discussion of this is in given in the results section).

It is worth noting that there is no any pre-assumption on wavelength intensity distribution type here. Actually, it is difficult to find a given distribution which can be applied to intensity measurements for each wavelength and at each time point. However, according to Chebyshev's inequality theory, formula $\mu \pm m\sigma$ can be applied, regardless of distribution type. For example, 89% of values are in the range of $\mu \pm 3\sigma$ if they fit to a normal distribution. However, if the distribution type is arbitrary, we can still say that 75% of values are in the range $\mu \pm 3\sigma$ according to Chebyshev's inequality. Hence, the boundary function can be used generally for any type of unknown distribution.

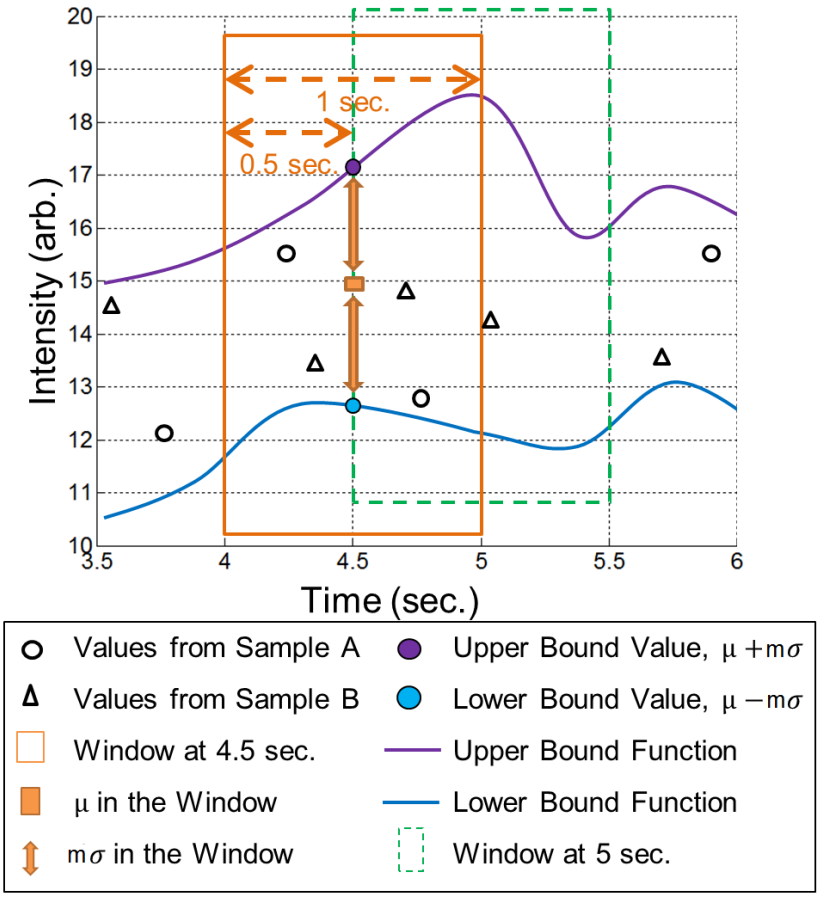

Figure 4-2. Illustration of boundary calculation for a single wavelength. Artificial data of two samples is demonstrated. A moving window is used to compute mean μ and standard deviation σ of intensity values inside the window. Upper and lower boundary values are calculated by $\mu \pm m\sigma$. Corresponding fitting functions are calculated with the piecewise linear interpolation.

Corresponding fitting functions are calculated from these distributed boundary values using a piecewise linear interpolation. The window size is set to 1.0 second. The window is moving at a distance of half of its length, 0.5 second in this case, such that the left half of every window overlaps with the previous window and the other half overlaps with the next window. According to experimental results, such a design gives a smooth fitting function. This fits with the generic feature of plasma etching that wavelength intensity changes slowly throughout the whole process. A similar moving window method was also used to address the process shift problem in [59]. Detailed discussion of the window size and sigma number is given in the section Results and Discussion of this chapter. If spectrums scans from all samples are already well

synchronised, the moving window is not needed. In that case, intensity values of the same wavelengths are grouped by the timestamps and corresponding boundary values are calculated in the same way.

Pseudo code for library calculation is presented below.

Procedure libraryCalculation

FOR every wavelength

Put intensity measurements at different time points into a single array;

Sort the measurements in time sequence;

FOR every time period of moving window

Calculate mean μ *and standard deviation* σ *of all intensity measurements falling in the window;*

Calculate upper bound value with $\mu + m\sigma$*;*

Calculate lower bound value with $\mu - m\sigma$*;*

ENDFOR

Calculate upper boundary function based on all upper bound values at different time points;

Calculate lower boundary function based on all lower bound values at different time points;

ENDFOR

4.2.2 Similarity Ratio Calculation

The Similarity Ratio (SR) calculation is based on a comparison between the boundaries in the SRA library and individual spectrum scans. Upper bound values and lower bound values are calculated for each wavelength, using boundary functions in the library as described previously. $UBF_i(t)$ and $LBF_i(t)$ are respectively the upper bound value and lower bound value for the i^{th} wavelength at time t. $WL_{i,t}$ is the actual intensity value of this wavelength at time t for the process being monitored. A similarity indicator $SI_{i,t}$ is used to record whether this wavelength at time t is between the boundaries, as per Equation (4-2). The same procedure is repeated for all

wavelengths. Dividing by the total number of wavelengths n, the SR_t value at time t for a full spectrum scan is outputted as per Equation (4-3). The potential range of SR is from 0% to 100%. For example, an SR value of 90% means that 90% of the wavelengths of the sample spectrum are similar to the healthy library.

$$SI_{i,t} = \begin{cases} 1 & , LBF_i(t) \le WL_{i,t} \le UBF_i(t) \\ 0 & , \ Otherwise \end{cases} \tag{4-2}$$

$$SR_t = \frac{\sum_{i=1}^{n} SI_{i.t}}{n} \times 100\% \tag{4-3}$$

Pseudo code for similarity ratio calculation is presented below.

Procedure SR Calculation
Set SI to 0;
FOR each wavelength
    Calculate upper bound value based on given time point;
    Calculate lower bound value based on given time point;
    IF wavelength intensity is in the range of lower bound value and upper bound value
        SI = SI+ 1;
    ENDIF
ENDFOR
Calculate SR: is equal with SI divided by wavelength total number;

4.2.3 Fault Detection Mechanism

The workflow of the proposed SRA method is shown in Figure 4-3. The fault detection mechanism is implemented by comparing the SR_t value with a SR threshold (*FD_Threshold*). If the SR value is below the threshold, an alarm will be triggered at time t. The alarm will be sent to external applications, which would reply with necessary actions, such as stopping the gas supply to the processing chamber. For the

demonstration, the *FD_Threshold* is set to 90%. Detailed experimental results are presented in the next section. However, this threshold can also be adjusted to customise the FD service with different sensitivities.

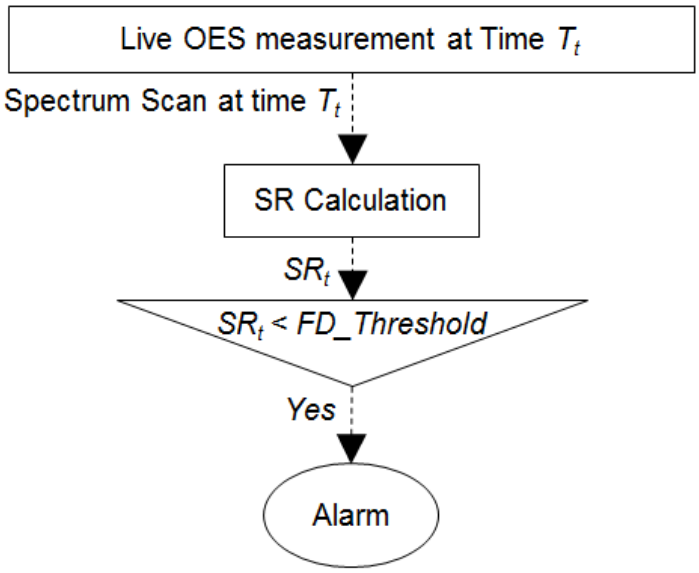

Figure 4-3. Workflow of Real-time FD with SRA. A SR value is computed for each real-time spectrum scan. A SR warning is triggered if the SR value is below SR threshold (*FD _Threshold*). The alarm will be sent to external applications, which would reply with necessary actions.

Pseudo code for fault detection mechanism is presented below.

Procedure FD

IF SR is smaller than the threshold

Trigger an alarm;

ENDIF

4.3 Results and Discussion

The SRA method is demonstrated on dataset D1. D1 is a real manufacturing dataset from an IC fabrication company. A 2-step etching process is monitored by an USB4000 Miniature Fibre Optic Spectrometer. A sample is defined as time-resolved spectrum scans for a complete etching process on one wafer. Every spectrum scan includes optical intensity measurements at 2048 different wavelengths from 178 nm to

874 nm. An etch rate measurement is also included with every sample. It is a common metrology dataset used to describe the output quality of an etching process. It is difficult to measure, so it is impracticable to use it for real time fault detection. In the demonstration, the etch rate is only used as a reference for selection of healthy samples and faulty samples for model validation. It is not needed after the SRA library is built. Actually, the SA model does not need such a precise etch rate measurement for training. Any brief information which can identify whether a sample is good or bad is sufficient, such as knowledge based on some practical experiments of the operators. This feature gives fewer constraints and more flexibility to the usage of the model.

For the model training, 200 healthy samples with similar etch rate outputs are selected from dataset D1. The SRA library is built on these samples for a healthy etching process. For model testing, another 118 healthy samples are used, as well as 7 faulty samples. All healthy samples share similar etch rates around 69 arb. units. All faulty samples have etch rates around 54 arb. units. These faulty samples are determined from detailed post-processing defect analysis of the product. The distribution of etch rates is presented in Figure 4-4 as a PMF histogram. Healthy samples from training and testing datasets have similar distributions, while the faulty samples are distinguished from the healthy ones. In the following model examination, this data selection helps to validate whether the model triggers a positive alarm for the faulty samples, and does not trigger negative alarms for the healthy samples.

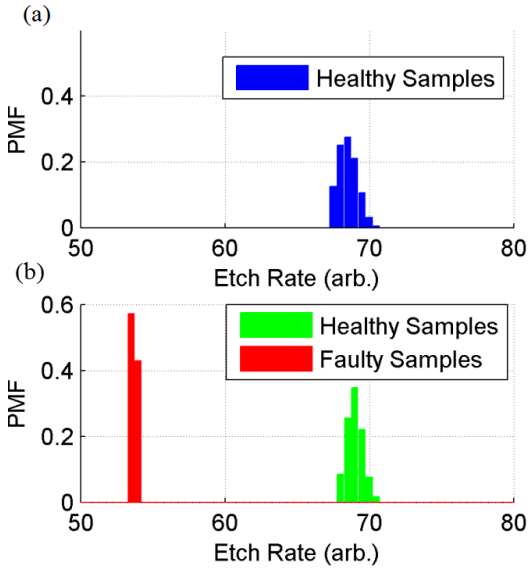

Figure 4-4. The PMF plot of etch rate values in training samples and testing samples respectively. Panel (a) represents the PMF of training samples, which only include healthy samples. Panel (b) represents the PMF of testing samples, which include healthy samples similar to the training samples and faulty samples.

It is worth noting that first wafer effect is not taken into account for the current SRA, which might impact sensitivity of the SRA library. The first wafer effect means that the first wafer in each lot through the chamber normally gives relatively different etching results compared with the subsequent wafers, because the first wafer is normally processed in a transient chamber environment before the chamber reaches equilibrium. Hence, ideally, the first wafers should be processed separately from the others. However, due to the limited sample resource, this potential problem is left as future work, which will require more data.

In the following section, the SRA library calculation result is presented at first. Corresponding fault detection results are presented at the end.

4.3.1 Library Calculation Results

In this section, the resulting SRA library derived from dataset D1 is presented. The process shift problem and two system control parameters (moving window size and sigma number) are first discussed. Finally a detailed view of the library is provided.

4.3.1.1 Survey of Process Shift & Moving Window Size

Process shift is a very common practical problem in modelling of the plasma etching process. It is usually caused by small changes in a system variable [59]. It also can be caused by the sensor readings being taken at a different phases for different samples. It normally does not affect the overall properties of the output data, but it can lead to slightly different measurements at the same time point in different process samples. In Figure 4-5, process shift is illustrated at wavelength 253.29 nm for two samples from our data. These two samples have very similar etch rate values, about 67.548 arb. units for both of them. They also share very similar curve shapes, which imply similar etching features. But a relative shift in time can be observed between them. A significant modelling bias could be introduced by this shift, if the model assumed synchronised measurement time points in all samples.

In the previous section on method description, a moving window is introduced to calculate boundary values based on the mean and standard deviation of the values falling within a window. This produces a smooth boundary function by focusing on only major features of the data. This is consistent with an etching feature that the whole chemical environment is changing slowly across the whole process. Considering the process shift problem, the moving window can also include all shifted values whose shifting distances are less than the window size. So the outputted boundary functions are capable of resolving the shift problem. Abnormal outliers are also excluded by this method, because boundary values are based on the most common sample scenarios represented by mean and standard deviation. Theoretically, the window size should be large enough to include all values at the same process phase, but too large a window size will lead to a reduction in detail in the time series information.

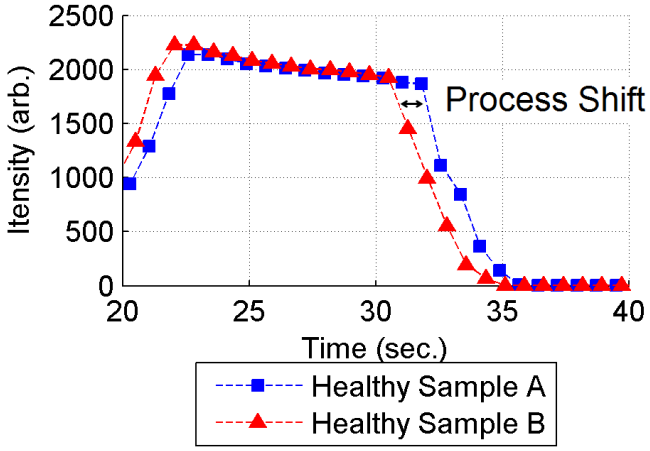

Figure 4-5. Illustration of process shift between two healthy samples with similar etch rate at wavelength 253.29 nm. Etch rate of both sample A and B are the same, 67.55 arb. unit, which implies that a similar etching output quality is shared by them. Sharps of the two curves are similar to each other, but there are differences in time series. A clear time shift can be observed.

In order to find a suitable window size, a comprehensive survey is conducted for the relevant time shift of all training sample time series. One sample is selected out as a reference. The other samples are compared against the reference sample with a cross-correlation analysis, and the relevant shift time is calculated. Statistical results are presented in Figure 4-6. According to this plot and corresponding statistical results, 86.43% of process shifts concentrate on the range from about -0.5 second to 0.5 second. Hence, most shift distances in the example dataset would be covered by using a window size of 1.0 second. In that case boundary values will be computed every second. Other methods to solve this misalignment problem (e.g. Dynamic time warping, covariance optimized warping [72] and linear time scaling) normally need to wait for complete time series data and are computationally expensive [73], so they cannot be easily used for early-stage FD, unlike the proposed moving window method.

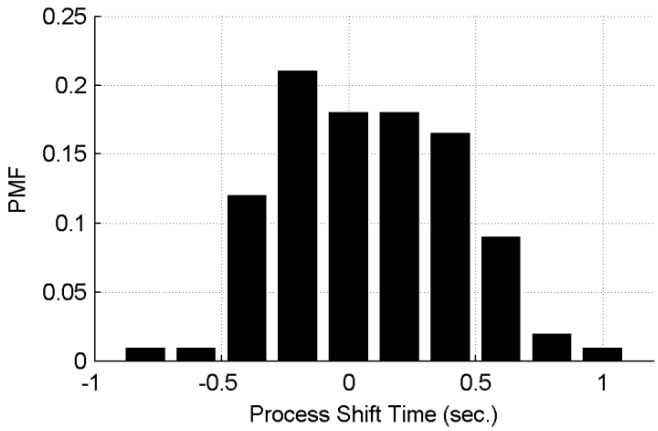

Figure 4-6. The PMF plot of process shift time in training samples. 10.55% samples has shift distance less than 0.1 second, 51.26% shift distances are less than 0.5 second, 86.43% shift distances are less than 1 second, 96.48% shift distances are less than 1.5 seconds, and 99.50% shift distances are less than 2 seconds.

4.3.1.2 Sigma Number Selection

The SRA library is built with $\mu \pm m\sigma$ boundaries for each wavelength from healthy training samples. For each wavelength, μ is the average value of all intensities falling in the moving window, and σ is the standard deviation. By choosing different sigma numbers m, different SRA libraries are tested and the corresponding minimum SR values of training samples are presented in Figure 4-7.

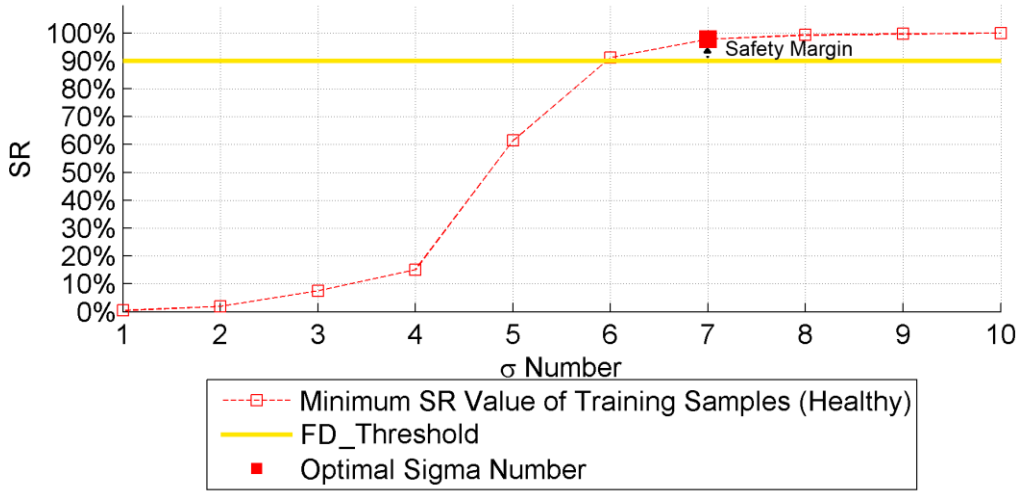

Figure 4-7. Plot of minimum SR values for training samples with different sigma numbers in the boundary function. By setting a selection threshold (threshold 90% plus a safety margin of 5%), sigma number 7 is selected as the smallest sigma number whose corresponding minimum SR value of training samples is higher than 95%.

Because the SRA library is based on healthy training samples, the minimum SR values of these samples should be close to an SR value of 100% as far as possible. In that case the SRA library would more accurately represent healthy samples. On the other hand, the sigma number should not be too large as the library would then be too general to reliably distinguish between faulty and healthy samples. Hence, Figure 4-7 is used to decide a suitable sigma number based on selection of the threshold (*FD_Threshold*) for fault detection. This threshold is also used in the fault detection mechanim, as mentioned previously. In the following demonstration, the *FD _Threshold* is set to 90% plus a safety margin of 5% to increase the reliability of fault detection. The optimal sigma number is chosen to satisfy the following condition: *choose the smallest integer sigma number whose minimum SR value is higher than FD_SRW_Threshold plus the safety margin.* In that case, sigma number 7 is chosen with a minimum SR value of 99.32% in the training data set. A range of other threshold values have also been tested, and none triggers a false alarm. Multiple true alarms are triggered with most threshold values, except for thresholds below 20% or above 95%, due to either a too specific or too general SRA library.

4.3.1.3 Library View

For each wavelength in the dataset D1, a pair of boundary functions (an upper boundary and a lower boundary) is built independently. 2048 pairs of boundary functions are created in total, and three of them are demonstrated in Figure 4-8.

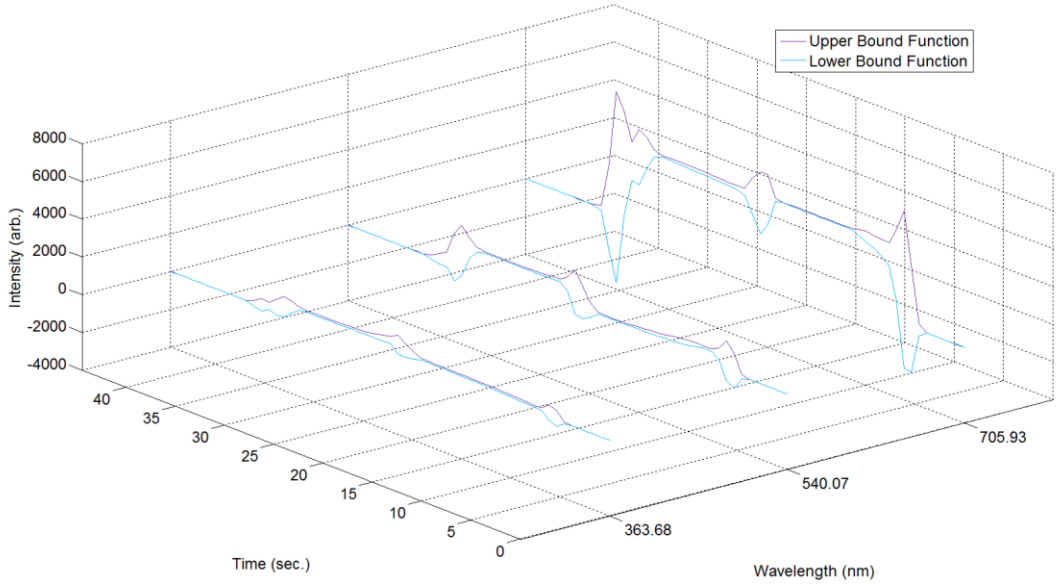

Figure 4-8. Illustration of a real SRA library at three wavelengths. The three wavelengths are 363.68 nm, 540.07 nm, and 705.93 nm. The original full spectrum includes wavelengths from about 178 nm to 874 nm.

In Figure 4-9, the SRA library function is illustrated at a single wavelength 486.46 nm, plus training and testing data. Intensity values at the same wavelength are shown in this figure from one training sample, one healthy testing sample and one faulty testing sample. The training sample and the healthy testing sample are within the boundaries, while the faulty testing sample has multiple intensity values which are outside the boundaries.

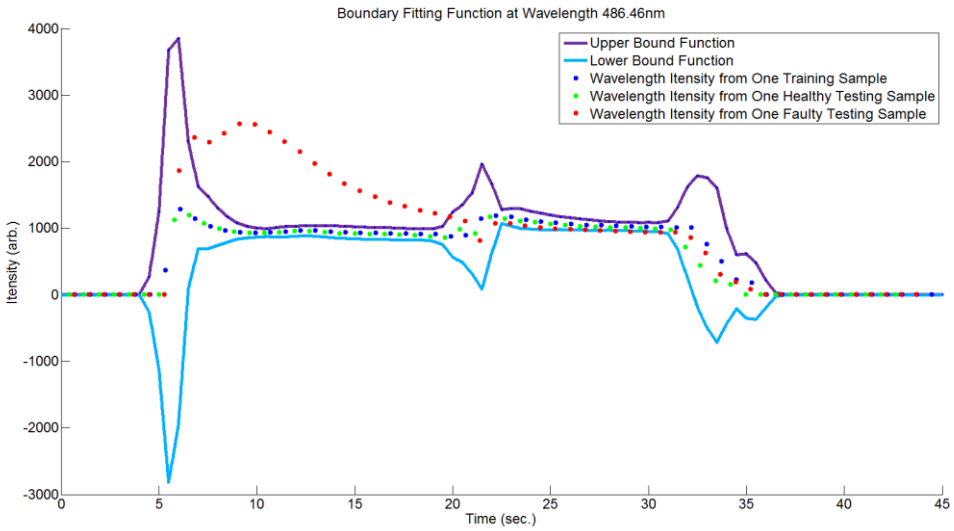

Figure 4-9. Illustration of SRA library in time series on wavelength 486.46 nm.

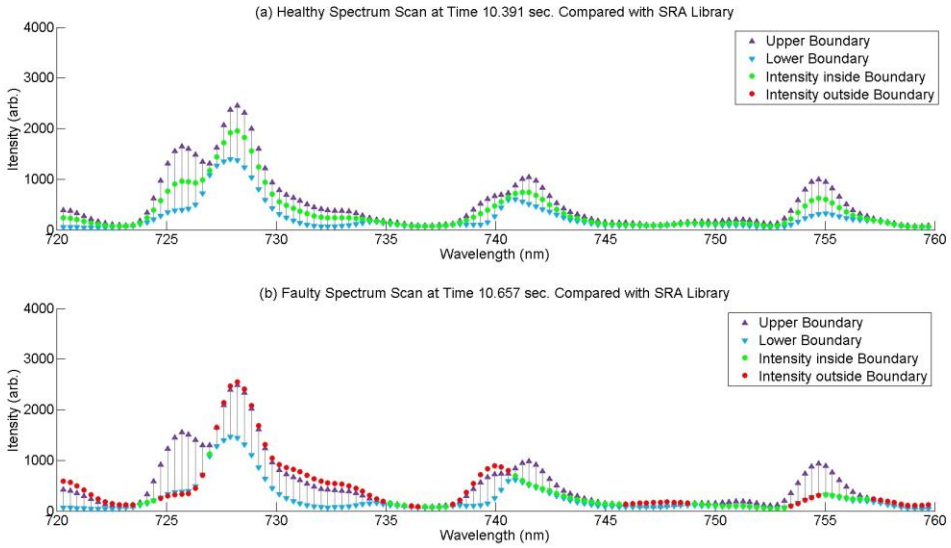

Figure 4-10. SA Comparison with a partial testing spectrum. The partial spectrum includes wavelength intensity measurements from about 720 nm to 760 nm. Panel (a) is result based on a healthy sample. Panel (b) is results based on a faulty sample.

Figure 4-10 gives another view on the comparison result between the library and testing data by looking at multiple wavelengths at a certain time point. At about the same time point of 10 seconds, one healthy testing and one faulty testing spectrum scan are compared with the SRA library separately over the wavelength range from 720 nm to 760 nm. If a wavelength intensity measurement falls within the boundaries, it is marked in green. On the other hand, if a wavelength intensity is outside of the boundaries, it is marked in red. Considering the results for a complete spectrum scan, the healthy testing spectrum scan fit well to the SRA library, while the faulty testing spectrum scan is significantly different from the SRA library. Therefore, the ability of the SRA library to describe healthy sample and separate faulty samples is validated.

4.3.2 Fault Detection Results

In the previous section, surveys of the window size and sigma number decided suitable parameter values for this particular dataset: 1.0 second for the window size and 7 for the sigma number. In this section, using these values, the proposed fault detection method is evaluated.

4.3.2.1 Time-series SR

Fault detection results are demonstrated with all healthy testing samples and a typical faulty testing sample in Figure 4-11. The other faulty samples are quite similar to this one. Each sample includes a SR time series for a complete etching process. The SR values of the healthy sample are drawn in green, and the SR values of the faulty samples are drawn in red. The fault detection threshold (*FD_Threshold*) is marked as a yellow straight line. All dots represent SR values which are above the threshold. The alarm is marked as a triangle. For the healthy samples which overlap in the figure, no alarm is triggered. According to an additional survey, 99.64% of the SR values are higher than 99.99% and the minimum SR value is 99.71%, which implies that the SRA library is carrying enough features for the entire etching process of healthy samples. For the faulty sample, multiple alarms can be observed. Beside the vertical SR axis on the left, an extra confidence level axis is also provided, calculated as 100% minus the SR value. This gives the user of the method extra information about the alarms triggered. The lower the SR value is, the higher the confidence the alarm is correct.

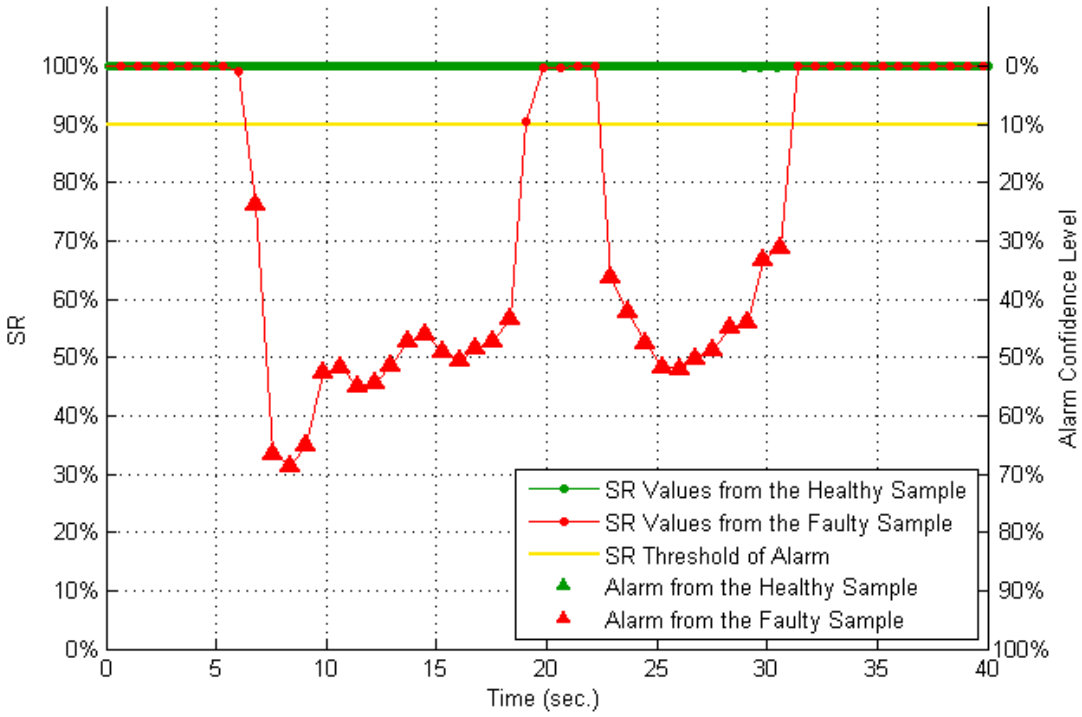

Figure 4-11. Fault detection result with optimal sigma number 7.All 118 healthy testing samples and one typical faulty testing sample are presented. For the healthy samples, no alarm is triggered. All of them overlap with each other and have similar SR values which are close to SR value 100%. For the faulty sample, multiple faulty alarms are triggered from an early-stage in the process.

It is noted that sigma number selection is an important factor which impacts the fault detecion results significantly. Accurate fault detection results presented in Figure 4-11 validate that the method for sigma number selection (*the smallest sigma number whose minimum SR value should be higher than FD_Threshold plus a safety margin.*) is successful. If *FD_Threshold* is still set to 90% but sigma number 4 is chosen, for example, instead of the optimal value of 7, corresponding fault detection results are shown in Figure 4-12. The results show that incorrect sigma number selection leads to multiple false alarms (green triangles in plot) triggered by healthy samples.

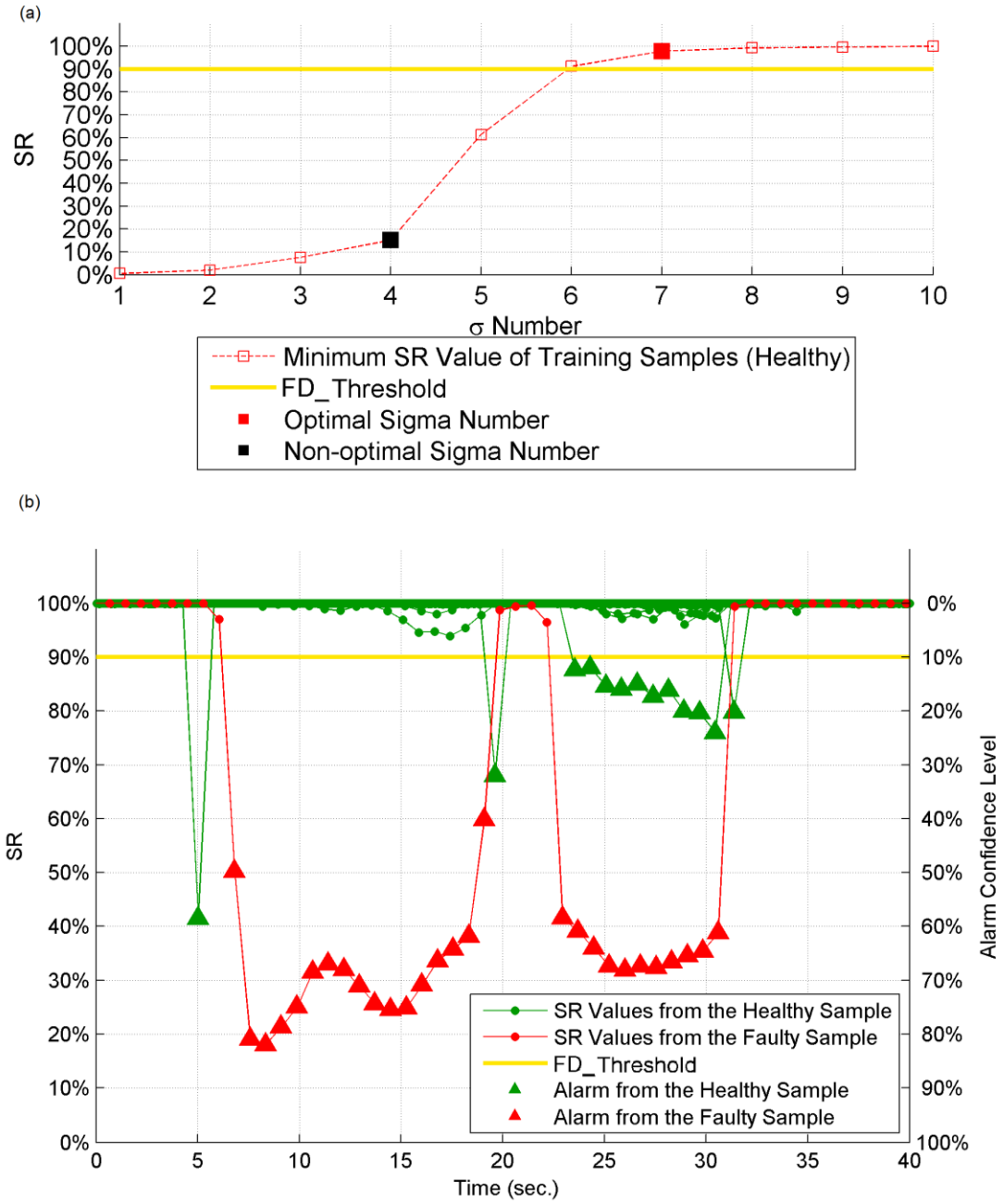

Figure 4-12. Fault detection result with non-optimal sigma number of 4. All 118 healthy testing samples and one typical faulty testing sample are presented. For the healthy samples, multiple warnings and alarms are triggered.

In Figure 4-13, the FD results of all 7 faulty testing samples are shown. All of them have a similar SR curve shape. No alarm is triggered at the beginning of etching step 1, transition from step 1 to step 2, or at the end of the step. This phenomenon is consistent with the etching feature that optical emission is weak at these three time periods, so no significant difference can be found between samples.

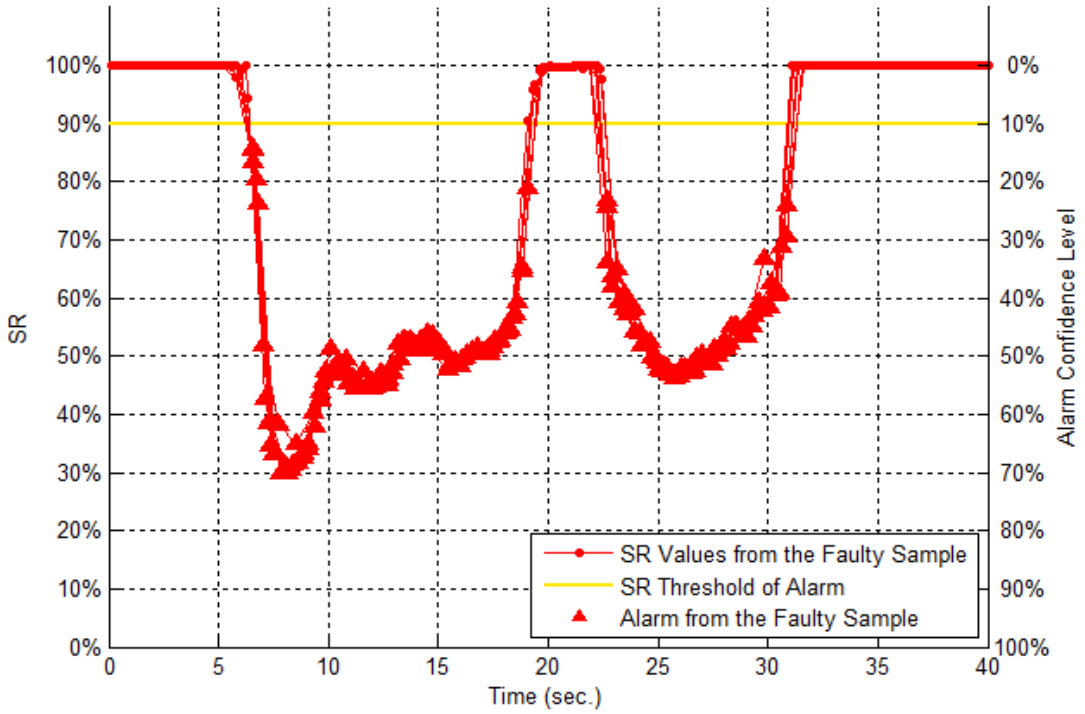

Figure 4-13. SR values of all 7 faulty testing samples.

The earliest first alarm is triggered at about 6.53 seconds in sample 4, and the latest first alarm is triggered at about 7.08 seconds in sample 3, as shown in Table 4-1. For each faulty sample, the average time of the first warning is about 6.75 seconds. Considering all faulty and healthy samples, early-stage fault detection is achieved with 100% accuracy.

Faulty Sample ID	1	2	3	4	5	6	7	Mean
Time of 1st Alarm (sec.)	6.81	6.56	7.08	6.53	7.00	6.55	6.73	6.75

Table 4-1. Time of the first alarm triggered in 7 faulty samples.

4.3.2.2 Post-process SR

The average SR values for a complete etching process are also calculated based on the full SR time-series. The result is presented for training samples, healthy testing samples and faulty testing samples separately in Figure 4-14. All training samples and healthy testing samples have average SR values of around 98%. All faulty testing samples have average SR values of around 68%. Faulty samples can be easily identified from the healthy ones by using a simple threshold between them. Compared

with the SR time-series, a bigger SR gap provides a more reliable fault detection result overall. Using an average SR may be a good option, when early-stage fault detection is not required.

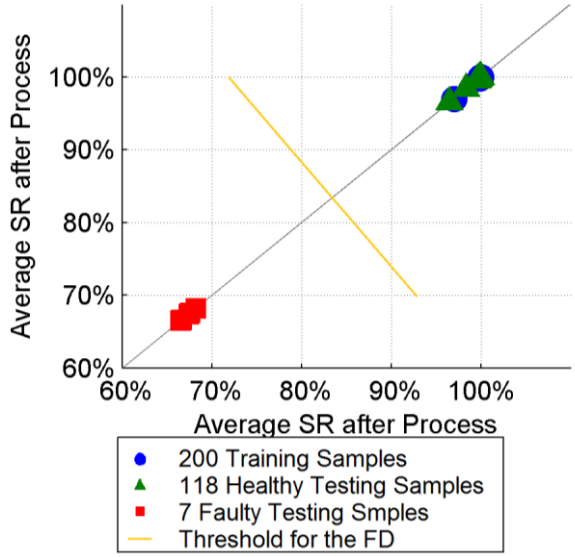

Figure 4-14. The average SR value plot of an entire etching process with training and testing dataset. All training samples and healthy testing samples have average SR values around 98%. All faulty testing samples have average SR values around 68%. Faulty samples can be easily identified from the healthy ones by using a simple threshold (yellow line) between them.

Since PCA is also commonly used for FD, a PCA-based method is implemented based on the same set of training and testing datasets for post-process FD, in order to give a comparison with the proposed SRA methods. In this approach, average wavelength intensities across an entire etching process are calculated for each wavelength and each sample. Then PCA is applied to the average wavelength intensities of training samples to get corresponding coefficients for PCA. Finally, PCs are calculated for all training, healthy testing and faulty testing samples. The first two PCs are presented in Figure 4-15. Faulty samples are easily separated from the healthy ones in the plot. Hence, the PCA method also works in this case as well as the SRA method for post-process FD. However, PCA is not easily used for the other FD scenario: early-stage FD. In order to use PCA at each time point, samples need to have OES measurements at the same time

points, but the demonstrating dataset does not have that. In contrast, SRA does not have such a problem.

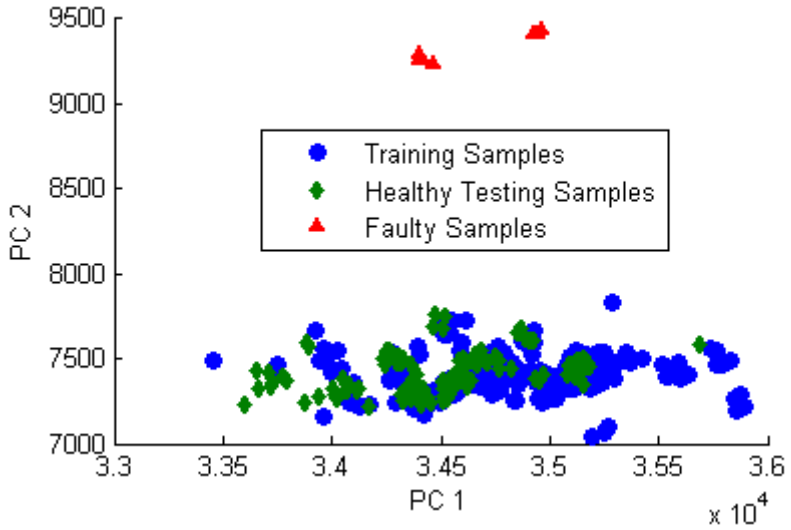

Figure 4-15. Plot of the first two PCs from an alternative PCA method for FD. This PCA method is based on the same set of training and testing samples as SRA.

4.4 SRA with IIRR

Since a dimensionality reduction method IIRR is introduced in Chapter 3 and is proposed as a pre-processing step for OES data for a range of different application, in this section the SRA method is combined with IIRR and tested for fault detection purposes. Compared with the previous stand-alone SRA, a different training dataset is used to build the SRA library, but tested on the same 7 faulty samples. First IIRR is applied to the training dataset, and key wavelengths are identified. Then SRA is applied to the key wavelengths for fault detection, instead of using the entire original OES dataset.

After the IIRR process, 64 key wavelengths are picked out from 2048 of the original wavelengths to represent the original dataset. In SRA, *FD _Threshold* is set to 90% plus a safety margin of 5%, the same as in the previous experiment. In the first step of SRA, sigma number 8 is selected to build the SRA library. Sigma number 8 is the

smallest sigma number, in this case, whose minimum SR value of training samples is higher than *FD_SRW_Threshold* plus the safety margin. The minimum SR value of all training samples is 0.9531. Hence, it is safe to say that training samples are well represented by the SRA library.

Fault detection results for SRA with IIRR are presented in Figure 4-16 for all healthy and faulty samples. No negative alarm is triggered by healthy samples, and every faulty sample triggers multiple alarms.

Compared to FD results for stand-alone SRA, SRA with IIRR provides very similar FD results. The first alarm time of faulty samples are presented in Table 4-2 for these two approaches respectively. The first alarm is triggered at almost the same time for both approaches. Hence, the IIRR-SRA method provides almost the same FD capability compared with the stand-alone SRA method.

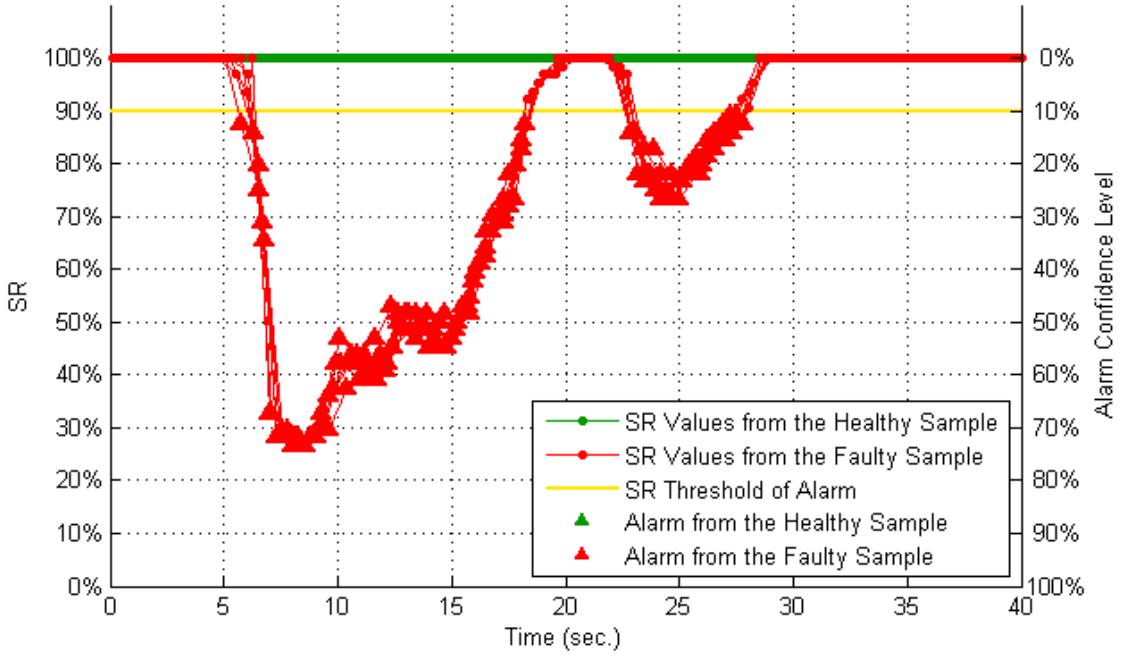

Figure 4-16. Fault detection result of IIRR-SRA. All healthy testing samples and faulty testing sample are presented. For the healthy samples, no alarm is triggered. All of them overlap with each other and have similar SR values which are close to SR value 100%. For the faulty samples, multiple faulty alarms are triggered from an early-stage in the process.

In terms of difference in computational cost, IIRR-SRA expends extra time for dimensionality reduction, but it is only run once when building the SRA library. Once the library is built, the IIRR-SRA computation takes a much smaller number of wavelengths as input (64 key wavelengths of 2048 in this experiment), so IIRR-SRA can execute faster than stand-alone SRA in real-time FD.

	Faulty Sample ID	1	2	3	4	5	6	7	Mean
Time of 1st Alarm (sec.)	Stand-alone SRA (without IIRR)	6.81	6.56	7.08	6.53	7.00	6.55	6.73	6.75
	IIRR- SRA	6.81	5.80	6.31	6.53	7.00	6.547	6.73	6.53

Table 4-2. The first alarm time comparison with stand-alone SRA and SRA combined with IIRR. Similar FD results are gained.

Additionally, since SRA is demonstrated on a different training dataset and good FD results are achieved, stability of the proposed SRA method is indicated.

4.5 Discussion & Conclusion

It has been shown that the proposed SRA method is effective and stable for early-stage fault detection with a real manufacturing OES dataset from a plasma etching process. For these types of processes, early-stage detection can help to reduce overall process cost. In our method, spectrum scans are not limited to certain time points but use all available time-domain data, where process shift is accounted for by using a windowing method. Based on a trade-off between the potential cost caused by false alarms and time delay before correct alarms are raised, users can customise the sensitivity of the model.

OES datasets from real manufacturing processes are always difficult to obtain, due to commercial confidentiality considerations. Also dataset D2 does not have faulty

samples, so the SRA method is only demonstrated for dataset D1. However, stand-alone SRA and IIRR_SRA are tested based on two different training sample sets, and both show successful FD results. Hence, it is still relatively safe to conclude that SRA is a stable and consistent solution for FD at least for OES datasets.

CHAPTER 5. Conclusions & Future Works

Plasma etching is one of the most important processes for the IC fabrication in the semiconductor industry. In order to understand and control the etching process, OES is widely used. Compared with other diagnostic technologies, OES has the significant advantages of non-intrusion, real-time data and large information capacity. On the other hand, its high dimensionality leads to some challenges. One of the biggest challenges is the high information redundancy. According to these particular features of OES data, two research objectives are the focus in this thesis: dimensionality reduction and early-stage fault detection. Chapter 1 gives a brief introduction to our research motivation and objectives. Chapter 2 outlines the background of the whole research area from the following perspectives. Firstly, it describes the underlying principles of the plasma etching process from a physical perspective. Secondly, a comparison across available plasma diagnostic technologies is provided. Thirdly, comprehensive literature reviews are presented based on different algorithms and research objectives. Finally, two manufacturing datasets which are used for evaluating the methods proposed in the thesis are discussed. After Chapter 2, each research objective is discussed separately in Chapter 3 and 4. Chapter 3 focuses on a novel dimension and redundancy reduction method, IIRR. Chapter 4 proposes an early-stage fault detection method, SRA. The summary results of each research objective are presented separately in the sections below.

5.1 IIRR Method for Dimension and Redundancy Reduction

A new dimension reduction method is presented for dimension and redundancy reduction in Chapter 3, named IIRR. The dimension of OES data is reduced

significantly by considering both peak wavelength emissions and the correlative relationships between them. The IIRR method is demonstrated for two different manufacturing datasets. Additional data pre-processing methods are also proposed to address two particular problems associated with these two datasets, namely wavelength saturation and non-normalized process time scale. After applying the IIRR process, number of measurements in the original OES dataset is reduced to about 0.2668% for the first test dataset D1 and about 0.4200% for the second dataset D2. As validation, I show that prediction of an independent output variable (etch rate for D1 and BLW for D2) can be done very effectively with the reduced set of variables. It is safe to conclude that this method can effectively reduce the number of wavelength intensity measurements required to accurately represent the data, at least for prediction of these two particular system variables. In fact for dataset D1, prediction accuracy was slightly improved, compared to prediction with the full set of input variables. It is noted that IIRR operates in the original variable space, rather than a transformed variable space, which would make the method useful for OES analysis methods whose goal relates to physical interpretation of the process, for example in virtual metrology methods. We would also expect the method to be effective for application to high-dimensional spectral data from other processes, where the dataset represents a set of time series, each of which is an independent sample from the same fundamental process. Although the APS step of the algorithm is specific to OES datasets, the core method (IRP+OPS) could be expected to be effective for other (non-OES) high-dimensional time series datasets, where multiple independent samples of the same (repeatable) underlying process behaviour are available. However, there is a caveat here. As the IRP phase of the method ranks less correlated variables highly, there is a risk of biasing noise for inclusion in the final variable set. In this case, corresponding interpretation of non-

peak data as noise and its effective reduction/removal by APS avoids this scenario. For data from other processes, some similar insight to the nature of the noise and an effective noise reduction method would be required, so that a high level of data reduction can be achieved. On the other hand, as the IRP/OPS method is 'internal' in nature, not guided/biased by a chosen output variable(s), it is conservative in terms of attempting to distinguish unexplained variation from noise. As a stand-alone method of preparing a universal reduced OES dataset, that can be applied to prediction of multiple different output variables of interest, this may be useful.

5.2 SRA Method for Early-stage Fault Detection

The Similarity Ratio Analysis (SRA) method is proposed for early-stage fault detection in Chapter 4. The SRA method is tested on a real manufactory dataset. Detailed discussion is provided on value selection of two configuration parameters based on this particular dataset, namely the moving window size and the sigma number for the SRA library building. Two SR value types are examined: time-series SR and post-process SR. The time-series SR represents SR values of each spectrum scan at individual time points, while the post-process SR is the average SR value of a complete etching process. Early-stage FD is achieved by the time-series SR, which detects faults at around 6 seconds from a 50 seconds etching process. Compared with the time series SR, the post-process SR distinguishes faulty samples from good ones more confidently. It is suitable for the scenario in which real-time FD is not needed and reliability of the FD service is emphasized.

Most previous research only realizes FD after completing the whole etching process. Few previously proposed methods are able to detect faults using a single spectrum scan, and the time when the spectrum is measured is limited to certain time points.

Process shift is another practical problem, which is normally ignored in previously proposed methods. The SRA method addresses these problems. It accepts spectrum scans at various time points. System faults will be detected at an early stage of the etching process so that less energy and materials will be wasted in faulty process runs. A confidence level associated with the alarm is also provided as a reference. The users of the method are also given a choice to customize the sensitivity of the model, by adjusting the SR threshold and consecutive warning number to trigger an alarm. Additionally, SRA is further tested, when combined with IIRR, which also gives good FD results while requiring a smaller input dataset.

5.3 Future Work

5.3.1 Future Work for Dimension Reduction

Future work relating to IIRR will investigate application of the method to other OES data sets. Additionally, it will also be considered how redundancy in the time domain can be reduced. In relation to the current OES plasma data, at least for certain periods of the process when it is less dynamic, the process is most likely over sampled and there is an opportunity for further data reduction without significant loss of important time domain information. Additionally, IIRR will be tested new datasets from different application domains in order to further test its stability and consistency.

5.3.2 Future Work based on Similarity Ratio Analysis

Future work based on SRA could include the following.

First of all, fault classification will be realized based on the features of time-series SR from faulty samples, as shown in Figure 4-13. This could help to find the underlying reasons causing the system failure quickly, reducing the time for system repair.

Secondly, the SRA method will be tested with IIRR more extensively to further reduce input data size, giving faster system response actions.

Thirdly, more system monitoring variables could be included in the SRA process, which would improve fault detection accuracy and give a more detailed profile of the etching process.

Fourthly, the SRA Method could be extended to allow feature extraction. This would help to address a common problem in plasma processing procedures where chemical elements lead to differences from chamber to chamber. Compared with the state of the art, it would be much easier to identify key wavelengths with SRA. Other methods usually use a combination or derived statistical parameters from the original data, which are difficult to trace back to identify prevalent wavelengths in the OES output.

For example, instead of counting how many wavelengths match the SRA library profile, as in Chapter 4 for fault detection, mismatching wavelengths can be outputted and interpreted as features which would help in the understanding of plasma physical/chemical changes more precisely. Such key wavelengths are illustrated in Figure 5-1(a) for one typical faulty sample in dataset D1. In Figure 5-1(b), dissimilar wavelengths from the same sample are illustrated at time point 9 seconds. Only 4 wavelengths are marked as dissimilar wavelengths from the original 2048 wavelengths. These four wavelengths are 437.9 nm, 441.8 nm, 443.2 nm, and 486.5 nm. They are potentially associated with Ar^+, O^+, SiN, and Ar^+ respectively. However, these highlighted particles are only selected based on very limited information from the etching procedure, which is protected under a consideration of commercial security, so they may not be correct, but they could none-the-less be good indicators for explaining the link between wavelength and chemical elements. Actually, this result has been

communicated to the company where the dataset D1 originated. They have confirmed the rationality of the result.

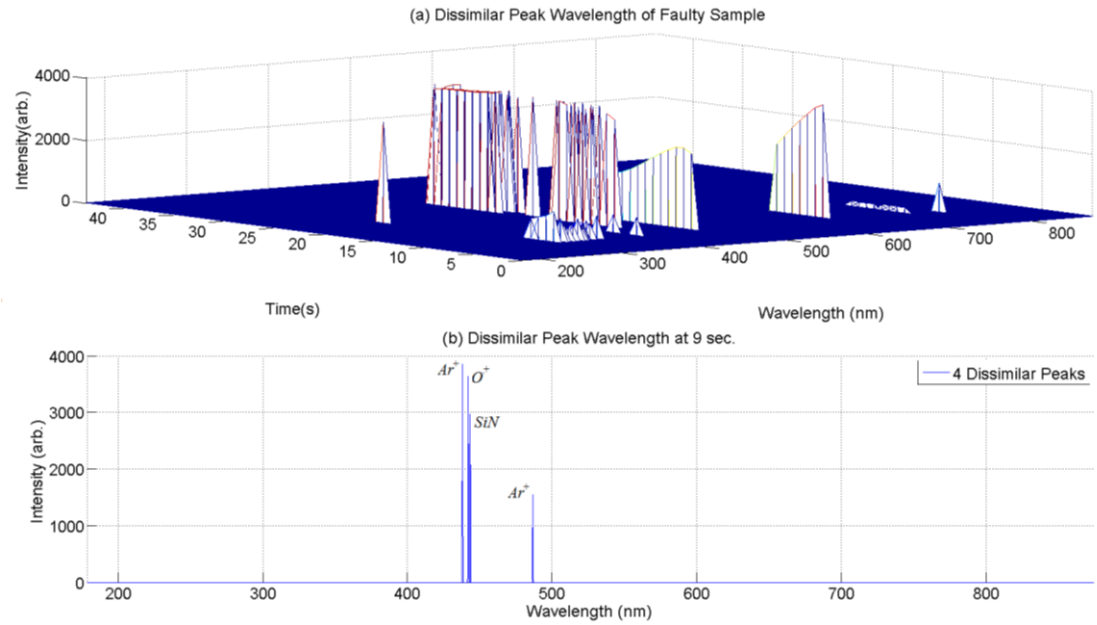

Figure 5-1. Illustration of similarity analysis for feature extraction. It is based on window size 1.0 second and sigma number 40 for one typical faulty sample. Panel (a) represents dissimilar wavelengths for a complete etching process. Panel (b) represents dissimilar wavelengths at time point 9 seconds from the same sample. Four outstanding wavelengths are four wavelengths are 437.9 nm, 441.8 nm, 443.2 nm, and 486.5 nm.

Fifthly, instead of comparing an individual sample with the SRA library for key wavelength identification, two different SA libraries based on different sets of samples can be compared with each other. This approach could help to find key wavelengths which are statistically significantly different, by removing the effects of noise and rare events. The design of a proposed method is presented in Figure 5-2. By using the Bhattacharyya coefficient, the similarity of two sample sets can be quantified, and then dissimilar wavelengths identified.

Finally, SRA could be tested with new datasets from different areas application domains, in order to further evaluate its stability and results consistency.

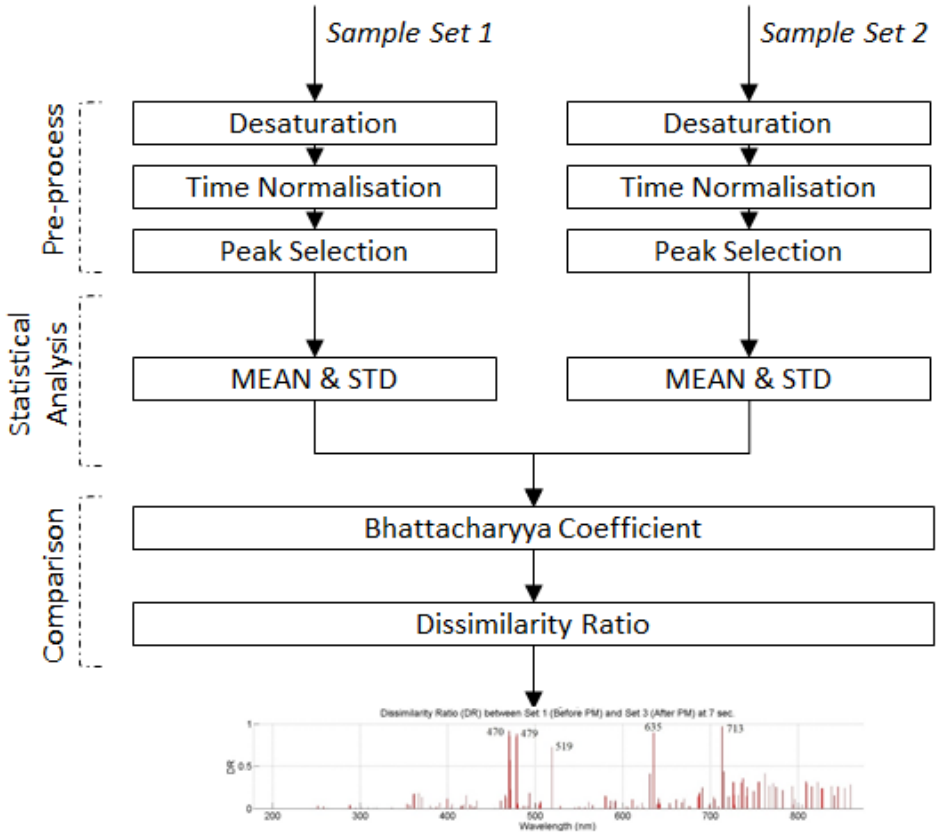

Figure 5-2. Work flow of dissimilarity analysis method for compassion of two sample sets**.** Pre-processing will help to reduce number of wavelengths in the final model output.

Reference

[1] C. Cardinaud, M.-C. Peignon, and P.-Y. Tessier, "Plasma etching: principles, mechanisms, application to micro-and nano-technologies," *Applied Surface Science,* vol. 164, pp. 72-83, 2000.

[2] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer, C. Bode, S. B. Hwang*, et al.*, "Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities," *Automatica,* vol. 36, pp. 1567-1603, 2000.

[3] R. Bellman, *Adaptive control processes: a guided tour* vol. 4: Princeton university press Princeton, 1961.

[4] I. K. Fodor, *A survey of dimension reduction techniques,* 2002.

[5] J. V. Ringwood, S. Lynn, G. Bacelli, B. Ma, E. Ragnoli, and S. McLoone, "Estimation and control in semiconductor etch: Practice and possibilities," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 23, pp. 87-98, 2010.

[6] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE,* vol. 86, pp. 82-85, 1998.

[7] (2 March 2013). *International Technology Roadmap for Semiconductors*. Available: http://www.itrs.net/about.html

[8] J. Yu, "Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 24, pp. 432-444, 2011.

[9] M. A. Carreira-Perpinan, "A review of dimension reduction techniques," *Department of Computer Science.University of Sheffield.Tech.Rep.CS-96-09,* pp. 1-69, 1997.

[10] B. N. Chapman, *Glow discharge process: sputtering and plasma etching*: Wiley, 1980.

[11] S. A. Lynn, J. Ringwood, and N. MacGearailt, "Global and local virtual metrology models for a plasma etch process," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 25, pp. 94-103, 2012.

[12] D. Zeng and C. J. Spanos, "Virtual Metrology Modeling for Plasma Etch Operations," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 22, pp. 419-431, 2009.

[13] S. Lynn, J. Ringwood, E. Ragnoli, S. McLoone, and N. MacGearailty, "Virtual metrology for plasma etch using tool variables," in *Advanced Semiconductor Manufacturing Conference*, pp. 143-148.

[14] H. H. Yue, S. J. Qin, J. Wiseman, and A. Toprac, "Plasma etching endpoint detection using multiple wavelengths for small open-area wafers," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 19, pp. 66-75, 2001.

[15] R. Westerman, D. Johnson, S. Lai, and M. Teixeira, "Endpoint detection methods for time division multiplex etch processes," 2006, pp. 61090I-61090I-11.

[16] B. Ma, S. McLoone, and J. V. Ringwood, "Tracking plasma etch process variations using Principal Component Analysis of OES data," presented at the 4th International Conference on Information in Control, Automation and Robotics, Angers, France, 2007.

[17] L. Conde, "An introduction to Langmuir probe diagnostics of plasmas," 2011.

[18] W. Z. Collison, T. Q. Ni, and M. S. Barnes, "Studies of the low-pressure inductively-coupled plasma etching for a larger area wafer using plasma modeling and Langmuir probe," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 16, pp. 100-107, 1998.

[19] C. Steinbruchel, "A new method for analyzing Langmuir probe data and the determination of ion densities and etch yields in an etching plasma," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 8, pp. 1663-1667, 1990.

[20] D. B. Ilic, "Impedance measurement as a diagnostic for plasma reactors," *Review of Scientific Instruments,* vol. 52, pp. 1542-1545, 1981.

[21] S. Lynn, N. Macgearailt, and J. Ringwood, "Real-time virtual metrology and control of plasma electron density in an industrial plasma etch chamber," in *Proceedings of 18th IFAC World Congress*, pp. 12060-12065.

[22] P. Dubreuil and D. Belharet, "Plasma impedance monitoring for real time endpoint detection of bulk materials etched in ICP tool," *Microelectronic Engineering,* vol. 87, pp. 2275-2281, 2010.

[23] A. Antoniadis, J. Bigot, and S. Lambert-Lacroix, "Peaks detection and alignment for mass spectrometry data," *Journal de la Société Française de Statistique,* vol. 151, pp. 17-37, 2010.

[24] B. A. Raby, "Mass spectrometric study of plasma etching," *Journal of Vacuum Science and Technology,* vol. 15, pp. 205-208, 1978.

[25] D. J. Bonser, "Endpoint detection utilizing ultraviolet mass spectrometry," United States Patent 5504328, 1996.

[26] G. Fortunato, "End-point determination by reflected power monitoring," *Journal of Physics E: Scientific Instruments,* vol. 20, p. 1051, 1987.
[27] J. Huang and M. Yang, "Endpoint control for small open area by RF source parameter VDC," United States Patent US 6930049B2, 2005.
[28] G. Chang, J. P. McVittie, J. T. Walker, and R. W. Dutton, "Electrical endpoint detection of VLSI contact plasma etching," *Electron Device Letters, IEEE,* vol. 5, pp. 514-517, 1984.
[29] D. A. White, "Multivariable analysis of spectral measurements for the characterization of semiconductor processes," Dissertation/Thesis, Massachusetts Institute of Technology, 2001.
[30] J. P. Roland, P. Marcoux, G. W. Ray, and G. H. Rankin, "Endpoint detection in plasma etching," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 3, pp. 631-636, 1985.
[31] I. T. Jolliffe, *Principal component analysis*: Springer verlag, 2002.
[32] D. A. White, D. Boning, S. W. Butler, and G. G. Barna, "Spatial characterization of wafer state using principal component analysis of optical emission spectra in plasma etch," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 10, pp. 52-61, 1997.
[33] R. Chen, H. Huang, C. J. Spanos, and M. Gatto, "Plasma etch modeling using optical emission spectroscopy," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 14, pp. 1901-1906, 1996.
[34] H. H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto, "Fault detection of plasma etchers using optical emission spectra," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 13, pp. 374-385, 2000.
[35] M.-W. Kim, S.-G. Kim, S. Zhao, S. J. Hong, and S.-S. Han, "Endpoint Detection in Plasma Etching Using Principal Component Analysis and Expanded Hidden Markov Model," *ECS Transactions,* vol. 34, pp. 943-948, 2011.
[36] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review,* vol. 49, pp. 434-448, 2007.
[37] I. T. Jolliffe, N. T. Trendafilov, and M. Uddin, "A modified principal component technique based on the LASSO," *Journal of Computational and Graphical Statistics,* vol. 12, pp. 531-547, 2003.

[38] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of computational and graphical statistics,* vol. 15, pp. 265-286, 2006.

[39] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation," *Journal of multivariate analysis,* vol. 99, pp. 1015-1034, 2008.

[40] R. Luss and A. d'Aspremont, "Clustering and feature selection using sparse principal component analysis," *Optimization and Engineering,* vol. 11, pp. 145-157, 2010.

[41] B. Ma, S. McLoone, J. Ringwood, and N. Macgearailt, "Selecting signature optical emission spectroscopy variables using sparse principal component analysis," in *Computer and Information Technology*, pp. 14-19.

[42] D. N. Lawley and A. E. Maxwell, "Factor analysis as a statistical method," *Journal of the Royal Statistical Society.Series D (The Statistician),* vol. 12, pp. 209-229, 1962.

[43] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White, and G. G. Barna, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *Journal of Chemometrics,* vol. 13, pp. 379-396, 1999.

[44] D. W. Stewart, "The application and misapplication of factor analysis in marketing research," *Journal of Marketing Research,* pp. 51-62, 1981.

[45] C.-W. Liu, K.-H. Lin, and Y.-M. Kuo, "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan," *Science of the Total Environment,* vol. 313, pp. 77-89, 2003.

[46] W. J. Doll, W. Xia, and G. Torkzadeh, "A confirmatory factor analysis of the end-user computing satisfaction instrument," *Mis Quarterly,* pp. 453-461, 1994.

[47] E. Ragnoli, S. McLoone, S. Lynn, J. Ringwood, and N. Macgearailt, "Identifying key process characteristics and predicting etch rate from high-dimension datasets," in *Advanced Semiconductor Manufacturing Conference, 2009. ASMC'09. IEEE/SEMI*, pp. 106-111.

[48] P. Prakash, A. Johnston, B. Honari, and S. F. McLoone, "Optimal wafer site selection using forward selection component analysis," in *Advanced Semiconductor Manufacturing Conference*, pp. 91-96.

[49] B. Flynn and S. McLoone, "Max Separation Clustering for Feature Extraction From Optical Emission Spectroscopy Data," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 24, pp. 480-488, 2011.

[50] A. T. McCray, J. McNames, and D. Abercrombie, "Stepwise regression for identifying sources of variation in a semiconductor manufacturing process," in *Advanced Semiconductor Manufacturing, 2004. ASMC'04. IEEE Conference and Workshop*, pp. 448-452.

[51] N. Alldrin, A. Smith, and D. Turnbull, "Clustering with EM and K-means," *University of San Diego, California, Tech Report,* 2003.

[52] C. Ding, X. He, H. Zha, and H. D. Simon, "Adaptive dimension reduction for clustering high dimensional data," in *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 147-154.

[53] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters,* vol. 31, pp. 651-666, 2010.

[54] U. Demšar, P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone, "Principal component analysis on spatial data: an overview," *Annals of the Association of American Geographers,* vol. 103, pp. 106-128, 2013.

[55] C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and k-means clustering," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 521-528.

[56] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 29.

[57] R. Yang and R. Chen, "Real-time plasma process condition sensing and abnormal process detection," *Sensors,* vol. 10, pp. 5703-5723, 2010.

[58] M.-S. Chen, T. Yen, and B. Coonan, "Real-time fault detection and classification for manufacturing etch tools," in *Semiconductor Manufacturing Technology Workshop Proceedings, 2004*, 2004, pp. 103-106.

[59] H. H. Yue and M. Tomoyasu, "Weighted principal component analysis and its applications to improve FDC performance," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, 2004, pp. 4262-4267.

[60] D. Angell, P. B.-L. Chou, A. R. Lee, and M. C. Sturzenbecker, "Monitoring and controlling plasma processes via optical

emission using principal component analysis," United States Patent 5658423, 1997.

[61] S. Vaseghi and H. Jetelová, "Principal and independent component analysis in image processing," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking (MOBICOM'06)*, pp. 1-5.

[62] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks,* vol. 13, pp. 411-430, 2000.

[63] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems,* pp. 145-151, 1996.

[64] J. M. Lee, S. J. Qin, and I. B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE Journal,* vol. 52, pp. 3501-3514, 2006.

[65] S. J. Hong and G. S. May, "Neural-network-based sensor fusion of optical emission and mass spectroscopy data for real-time fault detection in reactive ion etching," *Industrial Electronics, IEEE Transactions on,* vol. 52, pp. 1063-1072, 2005.

[66] Y.-J. Chang, Y. Kang, C.-L. Hsu, C.-T. Chang, and T. Y. Chan, "Virtual metrology technique for semiconductor manufacturing," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pp. 5289-5293.

[67] T. Sarmiento, S. J. Hong, and G. S. May, "Fault detection in reactive ion etching systems using one-class support vector machines," in *Advanced Semiconductor Manufacturing Conference and Workshop, 2005 IEEE/SEMI*, 2005, pp. 139-142.

[68] O. Ocean. (22 Nov 2012). *USB4000 Miniature Fiber Optic Spectrometer*. Available: http://www.oceanoptics.com/Products/usb4000.asp

[69] V. I. INC. (9 Oct). *Specification of Spectrometer SD1024D*. Available: http://www.verityinst.com/SD1024D.html

[70] S. E. Van Bramer, "An introduction to mass spectrometry," *Lecture Notes,* 1997.

[71] R. D. Tobias, "An introduction to partial least squares regression," in *Proc. Ann. SAS Users Group Int. Conf., 20th, Orlando, FL*, pp. 2-5.

[72] G. Tomasi, F. van den Berg, and C. Andersson, "Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data," *Journal of Chemometrics,* vol. 18, pp. 231-241, 2004.

[73] Y. Zhang, B. Lu, and T. F. Edgar, "Batch Trajectory Synchronization with Robust Derivative Dynamic Time Warping," *Industrial & Engineering Chemistry Research,* 2013.