Copyright © 2001, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

EQUIPMENT AND PROCESS MODELING AND DIAGNOSTICS IN SEMICONDUCTOR MANUFACTURING

by

Jiangxin Wang

Memorandum No. UCB/ERL M01/40

20 December 2001

EQUIPMENT AND PROCESS MODELING AND DIAGNOSTICS IN SEMICONDUCTOR MANUFACTURING

by

Jiangxin Wang

Memorandum No. UCB/ERL M01/40

20 December 2001

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Equipment and Process Modeling and Diagnostics in Semiconductor Manufacturing

by

Jiangxin Wang

B.E. (Tsinghua University) 1996
M.S. (University of California, Berkeley) 1998
M.S. (University of California, Berkeley) 2001

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering-Mechanical Engineering

in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor Costas J. Spanos, Cochair
Professor Alice M. Agogino, Cochair
Professor Kameshwar Poolla
Professor Joseph M. Kahn

Fall 2001

The dissertation of Jiangxin Wang is approved:

Cochair Date

Cochair Date

Date

Date

University of California, Berkeley

Fall 2001

Equipment and Process Modeling and Diagnostics in Semiconductor Manufacturing

Copyright © 2001
by
Jiangxin Wang

Abstract

Equipment and Process Modeling and Diagnostics in Semiconductor Manufacturing

by

Jiangxin Wang

Doctor of Philosophy in Engineering-Mechanical Engineering

University of California at Berkeley

Professor Costas J. Spanos, Professor Alice M. Agogino, Cochairs

The modern semiconductor industry is currently advancing into the world of sub 0.1-micron technology. Extremely fine feature sizes, dozens of masking levels, and very high product yields are now the norm. As device sizes diminish, final product quality becomes increasingly sensitive to variations at every single process step. This thesis aims at developing advanced methodologies for efficient and accurate diagnostics of process and equipment variations. Demonstrated on two critical systems of semiconductor manufacturing, the techniques developed in this thesis can also be adapted to a variety of process and equipment diagnostics.

Furnace systems have tremendous importance in semiconductor manufacturing since many heating steps are involved in a semiconductor process and it is crucial to properly monitor and control the temperature at each step. A furnace system is a typical dynamic system whose physical model can be constructed based on our

understanding of the thermal systems. In this thesis we describe how to choose a linear dynamic model to approximate the real system, how to estimate model parameters when system information is only partially available, and how to detect variations of system parameters using a statistical model-classification approach. This approach is evaluated using experimental data collected from a five-zone Tylan furnace at the Berkeley Microfabrication Laboratory. We also discuss the application of sensor fusion techniques for enhancing the reliability of diagnostics. This general approach can be extended to other systems that have relatively simple mathematical-approximation models based on known physics.

Compared to furnace systems, photolithography processes are much more complex. While complicated physical-mathematical models of lithography are available in certain commercial packages, the prohibitively large computation required makes them infeasible for real-time applications. In this thesis, we explore replacing the complex first principle models with simpler empirical models. Due to the high dimensionality and high non-linearity of the problem, a simple mapping usually does not exist between the Critical Dimension (CD) profiles and input recipe parameters, which is desired for diagnosing input parameter variations. Two different approaches are proposed to solve this problem and they may be used complementarily in different situations. The first approach is to construct an explicit inverse model using statistical modeling techniques. The second is to build a library of input-output data pairs and, during diagnostics, search for a candidate-solution set whose statistics are used to calculate the final solution. The two approaches are evaluated and compared on computer simulation results. Finally, time series models are considered to enhance the diagnostics of input parameter variations. The key contribution of this work is that it provides a computationally tractable

modeling and diagnostic framework for lithography processes. In addition, our approaches also provide a principled way to evaluate the utilities of parameter measurements and the required precision and accuracies. This knowledge can be invaluable to engineers for designing effective metrology techniques. Although the application is to semiconductor process diagnostics, it is possible to apply the concepts and techniques developed to other complex processes and systems.

To summarize, this thesis focuses on the development of statistical diagnostic strategies for monitoring and controlling processes and equipment in modern semiconductor manufacturing. Using experimental and simulated data, we have demonstrated the effectiveness of the proposed approaches for a system with a relatively simple physical model (the furnace system) as well as for an extremely complex (photolithography) process.

Professor Costas J. Spanos
Dissertation Committee Cochair

Professor Alice M. Agogino
Dissertation Committee Cochair

To Love, Survival and Peace

锦瑟无端五十弦， 一弦一柱思华年；

庄生晓梦迷蝴蝶，望帝春心托杜鹃。

沧海月明珠有泪，蓝田日暖玉生烟；

此情可待成追忆，只是当时已惘然。

— <锦瑟>，唐 李商隐

Contents

List of Figures vi

List of Tables xi

1 Introduction 1

1.1 Motivation 1

1.2 Thesis Objective 2

1.3 Thesis Organization 4

2 Modeling 5

2.1 Physical Models 6

2.2 Empirical Modeling 8

2.2.1 Linear Regression 9

2.2.2 Non-linear Regression 10

2.2.3 Artificial Neural Networks 10

2.2.4 Probabilistic Graphical Models 12

2.2.5 Fuzzy Logic 14

2.3 Parameter Estimation 15

2.3.1 Design of Experiments (DOE) 15

2.3.2 Cost Function 16

2.3.3 Optimization 17

2.3.4 Model Overfitting 17

3 Diagnostics **19**

3.1 Sources of Errors 20

3.1.1 Types of errors 20

3.1.2 Error Modeling 21

3.1.3 Variations in Semiconductor Manufacturing Systems 22

3.2 Fault Detection and Classification 23

3.2.1 Fault Detection for Linear Dynamic Systems 23

3.2.2 Fault Classification Methods 24

3.2.3 Fault Detection and Process Control in Semiconductor Manufacturing 26

3.3 Sensor Fusion 28

4 Diagnosing Malfunctions in Thermal Systems **30**

4.1 Introduction 30

4.2 Furnace Modeling 32

4.2.1 System Modeling 33

4.2.2 Parameter Extraction 38

4.3 Furnace Diagnostics 47

4.4 Results Analysis 54

4.4.1 Modeling Results 54

4.4.2 Diagnostic Results 55

4.4.3 Data Fusion Applied in Diagnostics 60

4.5 Summary 62

5 Lithography Process Diagnostics **64**

5.1 Lithography Process Diagnostics 66

5.1.1 Process Overview 66

5.1.2 Input and Output Parameters 69

5.1.3 Critical Dimension Metrology 71

5.1.4 Model-based Diagnostics 74

5.2 Empirical Modeling of Lithography Process 80

5.2.1 Modeling Data from PROLITH 80

5.2.2 Linear Regression 84

5.2.3 Neural Network Modeling 85

5.3 Inverse Modeling for Process Diagnostics 90

5.3.1 Application of Explicit Inverse Model 91

5.3.2 Space Segmentation 93

5.4 Library Searching Approach 113

5.5 Estimation of the Time-Series Model of Input Variations 127

5.5.1 Prediction and Estimation 129

5.5.2 Time Series Model Estimation 130

5.5.3 Model Mismatch Problem 133

5.6 Summary 142

6 Conclusions and Future Work **144**

6.1 Summary and Conclusions 144

6.2 Future Work 145

Bibliography **148**

A List of Symbols **163**

List of Figures

2.1 An artificial neuron. 13

2.2 A multi-layer perceptron (MLP) network. 13

3.1 Error type according to time-dependency. 21

4.1 Schematic of furnace system. 34

4.2 Electrical equivalent of the thermal system. 35

4.3 Overall system structure. 39

4.4 A typical data set for a dry oxidation furnace. 41

4.5 Data set used for parameter extraction. 42

4.6 Steady state data set. 44

4.7 Cooling down data set. 45

4.8 Heating up data set. 46

4.9 Model classification method. 49

4.10 Selection criterion. 50

4.11 Comparison of experimental data and linear model simulation results (temperature profile). 56

4.12 Comparison of experimental data and linear model simulation results (power profile). 57

4.13 Sensitivity test of diagnostic algorithm. 60

4.14 Data fusion to enhance diagnostic reliability. 61

5.1 Flow chart of the lithography process. 68

5.2 Layer structure that we use in this work. 68

5.3 CD profile features. 72

5.4 Automatic feedback parameter adjustment. 75

5.5 Forward Model. 76

5.6 Inverse Model. 77

5.7 Overview of diagnostics system using a hierarchical structure. . . . 79

5.8 CD vs. focus for different exposure, when $coh = 0.55$, $t_{PEB} = 60s$, and $T_{PEB} = 132^oC$. 82

5.9 CD vs. PEB temperature for different PEB time, when $exp = 12mJ/cm^2$, $foc = 0.1\mu m$, and $coh = 0.55$. 83

5.10 Forward MLP model. 86

5.11 Inverse MLP model. 87

5.12 Inverse MLP model for exposure dose. 88

5.13 Explicit inverse model approach case I #1~#4 results: the estimated parameter is the only unknown. 94

5.14 Explicit inverse model approach case II #1 results: exposure, focus unknown; PEB time and temperature known. 95

5.15 Explicit inverse model approach case II #2 results: PEB time and temperature unknown; exposure, focus known. 95

5.16 Explicit inverse model approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known. 96

5.17 Explicit inverse model approach case II #4 results: PEB temperature, exposure unknown; PEB time, focus known. 96

5.18 Explicit inverse model approach case II #5 results: PEB time, exposure unknown; PEB temperature, focus known. 97

5.19 Explicit inverse model approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known. 97

5.20 Explicit inverse model approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known. 98

5.21 Explicit inverse model approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known. 99

5.22 Explicit inverse model approach case III #3 results: exposure, PEB time and temperature unknown; focus known. 100

5.23 Explicit inverse model approach case III #4 results: focus, PEB time and temperature unknown; exposure known. 101

5.24 Explicit inverse model approach case IV results: exposure, focus, PEB time and temperature all unknown. 102

5.25 Space segmentation to find inverse models. 105

5.26 Example of space segmentation boundaries. 106

5.27 Segment inverse model approach case I #2 results: the estimated parameter is the only unknown. 107

5.28 Segment inverse model approach case II #1 results: exposure, focus unknown; PEB time and temperature known. 107

5.29 Segment inverse model approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known. 108

5.30 Segment inverse model approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known. 108

5.31 Segment inverse model approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known. 109

5.32 Segment inverse model approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known. 110

5.33 Segment inverse model approach case III #4 results: focus, PEB time and temperature unknown; exposure known. 111

5.34 Segment inverse model approach case IV results: exposure, focus, PEB time and temperature all unknown. 112

5.35 Determining weights using a Gaussian curve. 115

5.36 Diagnosis using a library searching approach. 116

5.37 Library searching approach case I #1~#4 results: the estimated parameter is the only unknown. 118

5.38 Library searching approach case II #1 results: exposure, focus unknown; PEB time and temperature known. 119

5.39 Library searching approach case II #2 results: PEB time and temperature unknown; exposure, focus known. 119

5.40 Library searching approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known. 120

5.41 Library searching approach case II #4 results: PEB temperature, exposure unknown; PEB time, focus known. 120

5.42 Library searching approach case II #5 results: PEB time, exposure unknown; PEB temperature, focus known. 121

5.43 Library searching approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known. 121

5.44 Library searching approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known. 122

5.45 Library searching approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known. 123

5.46 Library searching approach case III #3 results: exposure, PEB time and temperature unknown; focus known. 124

5.47 Library searching approach case III #4 results: focus, PEB time and temperature unknown; exposure known. 125

5.48 Library searching approach case IV results: exposure, focus, PEB time and temperature all unknown. 126

5.49 Time series simulation: estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$. 131

5.50 Time series simulation: estimation and prediction using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model for focus: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$; actual model for exposure: ARIMA(1,1,1), $a = 0.6$, $b = 0.3$. 132

5.51 Time series simulation: model parameter estimation, estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$. 134

5.52 Time series simulation: model parameter estimation, estimation and prediction of focus using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$. 135

5.53 Time series simulation: model parameter estimation, estimation and prediction of exposure using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(1,1,1), $a = 0.6$, $b = 0.3$. 136

5.54 Time series simulation: mismatched model approximation, estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(3,1,2), $a = [0, 0.12, 0.016]$, $b = [0, 0.01]$. Approximate model: ARIMA(1,1,1). 138

5.55 Time series simulation: mismatched model approximation, estimation and prediction of focus using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(3,1,2), $a = [0, 0.12, 0.016]$, $b = [0, 0.01]$. Approximate model: ARIMA(1,1,1). 139

5.56 Time series simulation: mismatched model approximation, estimation and prediction of exposure using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(3,1,2), $a = [0, 0.07, 0.006]$, $b = [0.1, 0.02]$. Approximate model: ARIMA(1,1,1). 140

List of Tables

4.1 Electrical equivalents of thermal parameters. 35

4.2 Performance of Least Square Method. N/A denotes "Not Applicable" and N/F denotes "Not Feasible". $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9$: nine thermal resistance faults, $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$: five power inefficiency faults, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$: five temperature sensor bias faults. Number in parenthesis indicates number of data dimensions, not shown when it equals or exceeds the number of parameters. 58

5.1 Comparison of CD metrology tools. 73

5.2 Input parameters and their ranges. 81

5.3 Output measurement noise range. 81

5.4 First order linear regression. 84

5.5 Second order response surface model linear regression. 84

5.6 Forward MLP modeling results. 86

5.7 Inverse model from four outputs to five inputs. 88

5.8 Inverse model from four outputs and four inputs to one input. . . . 89

5.9 Simulation cases for all diagnostic approaches. Y: known; N: Unknown 92

5.10 Library searching approach results summary. Y: known; N: Unknown 117

5.11 Time series simulation summary when focus is the only unknown parameter. 137

5.12 Time series simulation summary for focus when focus and exposure are the unknown parameters. 141

5.13 Time series simulation summary for exposure when focus and exposure are the unknown parameters. 141

Acknowledgements

I would like to express my deepest gratitude to my research advisors, Professor Costas J. Spanos and Professor Alice M. Agogino. Without the excellent guidance and full support from Professor Spanos, this work would not have been possible. I thank him for his consistent help on every detail of this research and I thank him for his understanding, trust and friendship. I thank Professor Agogino for her continuous encouragement and support on everything throughout my entire graduate school life. I appreciate the time she spent and her painstaking care on reading this dissertation. Without her picking on every tiny error, this work would not have been what it is now.

I thank Professor Kameshwar Poolla for his insightful input to this work and for reviewing my dissertation. Special thanks go to Professor Joseph M. Kahn for his serving both on my Qualifying Exam Committee and on my Dissertation Committee, and for his guidance on my second Master degree in communication systems. I am also grateful to my other Qualifying Exam Committee members: Professor Lotfi A. Zadeh, Professor Karl J. Hedrick and Professor Andrew Packard.

I would like to extend my gratitude to all current and former BEST labbers (Berkeley Expert System Technology Laboratory), for making the first three years of my graduate school life at Berkeley a wonderful experience, and making BEST LAB a warm home forever. Especially I would like to thank Ya Wen, Ningning Zhou, and Youhao Jing for the precious time we spent together, Susan Chao and Kai Goebel for their numerous supports during the beginning years of my graduate school, and Jialong Wu and Shuang Song for their friendship.

I thank members of the BCAM (Berkeley Computer-Aided Manufacturing) group, Junwei Bao, Runzi Chang, Haolin Zhang, Mason Freed, Michiel Krugger, Nickhil Jakatdar, Jason Cain, and DongWu Zhao, for their help and friendship. And in particular, I appreciate the help from Junwei Bao and Haolin Zhang on using PROLITH and the interesting discussions with them on other technical issues. Also I would like to thank Tim Duncan, Thanh-Xuan Nguyen, Huy Nguyen and T K Chen for their support.

I thank the administrative assistants of both the Mechanical Engineering and the Electrical Engineering and Computer Science departments who provided numerous help during the past five years. Special thanks go to Rui Neves, Pat Giddings, Ruth Gjerde, Charlotte Jones and Diane Chang. I also thank the staff members of the Berkeley Microfabrication Laboratory, Bob Hamilton, Sia Parsa, and Patrick Wehrly, who helped me conduct experiments on the Tylan furnaces.

I am grateful to my Berkeley buddies, Ye Sheng, Xunlei Wu, Yanhong Jin and Jingang Yi, who had made my Berkeley life more colorful and memorable.

I would like to thank my big brother, Shanxing Wang, for his support on every aspect. I thank him for his life-long inspiration, enlightenment and love. Thank my sister Hexin Wang and her husband Bao Mi for their always being there for me. Thank my parents, my sister Haixin Wang and my brother Chuanxin Wang for their love and encouragement.

Finally and above all, I am indebted to my husband, Shuangyu Chang. I thank God for sending him to me. His companionship makes everything I do become more enjoyable and more meaningful. What he did for me and for this work is just

uncountable. This work is for him.

This work was funded by the State of California SMART program under research contract SM97-01, and by the following participating companies: Advanced Energy, ASML, Atmel Corp., Advanced Micro Devices, Applied Materials, Asyst Technologies Inc., BOC Edwards, Cymer, Etec Systems Inc., Intel Corporation, KLA-TENCOR, Lam Research Corp., Nanometrics Inc., Nikon Research Corp., Novellus Systems Inc., Silicon Valley Group, Schlumberger, Texas Instruments Inc., Tokyo Electron.

Chapter 1

Introduction

1.1 Motivation

After more than three decades of rapid growth, the modern semiconductor industry is moving into a new frontier, the world of sub 0.1-micron technology. Extremely fine feature sizes, dozens of masking levels, and very high product yields are the norm of today's semiconductor manufacturing. There are strong economic incentives for having very high small feature reproducibility, process reliability and efficiency [1]. Research and development efforts are made in a number of areas by different groups of people. Physicists and chemists are looking into better understanding of the physics and chemistry underlying semiconductor processes; material scientists are trying to discover and adopt new materials; electrical engineers are creating new sensors to improve metrology to meet *in-situ* and high-precision requirements; and we, statistical control engineers, are designing better methodologies to monitor process and equipment health based on available experimental observations.

Modern semiconductor manufacturing is full of variations as processes and equipment are affected by environmental parameters, material properties, machine variables, as well as human factors. When the features become smaller, these variations are increasingly significant in causing degraded yield, slower parts, and higher production cost. Thus, it is imperative to have advanced process control methods to reduce these variations. Many studies are under way in this area [32]. Some simple statistical control tools are currently used in modern semiconductor manufacturing. Manufacturing engineers have applied techniques such as statistical process control to detect abnormal observations [30, 112]. However, more advanced application of statistical control is needed to solve complex diagnostic problems. One important, yet difficult goal, is to identify the precise sources of variations and understand their effects. Using statistical and probabilistic methods, we can build models that capture the essentials of the variations through empirical experiments, we can perform diagnostics that identify possible deviations from the desired norm, and we can develop control methodologies that bring about optimal performance.

1.2 Thesis Objective

This thesis explores applications of some general modeling and diagnostic approaches for monitoring and control of modern semiconductor equipments and processes. It discusses appropriate modeling approaches and fault diagnostic strategies under different conditions. In particular, it describes methodologies for two different problems, a typical furnace system and the photolithography process. The fault diagnostics for the furnace system is based on physical modeling and

the lithography process is based on empirical modeling. While the study of both systems are important, the modeling and diagnostic methodologies can be easily transferred to other equipments or processes (e.g., Plasma Etching, CMP, etc.).

A furnace is a dynamic system whose physical model can be constructed based on our understanding of thermal systems. We use a simple linear dynamic model to accurately approximate the real system. Abnormal changes in some system parameters may indicate malfunctions or potential failures of some system components. Statistical model classification methods and Minimum Mean Square Error (MMSE) estimation approaches are adopted to detect variations of three different types of system parameters that are important to the furnace system. Our analysis is based on experimental data collected from a five-zone Tylan furnace in the Berkeley Microfabrication Laboratory. This is a relatively simple system and the diagnostics perform well under normal conditions when not too many different faults happen at the same time. Application of sensor fusion techniques [2, 4, 68, 120] is also discussed to enhance the reliability of the diagnostics.

In contrast, the photolithography process is extremely complex involving many subsystems and processes [100]. Although there are physically based simulation models of the entire photolithography process that are feasible when parameters are well-tuned [52, 78, 95], such models are often too complicated and their computation requirements are prohibitively high for real-time control and diagnostic applications. Thus, a simple statistical model that enables fast simulation is needed. This thesis proposes a novel approach for modeling the photolithography process based on statistical analyses of the relationships between input and output variables. This empirical model is developed for a set of parameters over their ranges

of interest. High complexity, non-linearity, and uncertainties of the lithography process make diagnostics non-trivial. In this work we are particularly concerned with the variations of the Critical Dimension (CD) profiles of small layout features. Errors in several factors (and their combinations) such as light source, photoresist parameters, and Post Exposure Bake (PEB) process parameters, may all potentially cause variations in CD profiles. Some parameters may be measured and adjusted easily, but most of them cannot, due to difficulties and high metrology cost. Successful diagnostic strategies are developed based on the empirical modeling approach where input drifts are predicted by CD profile measurements. In most cases, input parameter drifts can be modeled as time series sequences. We discuss how to estimate the time series model parameters and make use of the time series model in diagnostics. The performance of the diagnostics is evaluated based on computer simulation results. We do not have any actual experimental results for verification at this stage.

1.3 Thesis Organization

Chapter 2 reviews major modeling techniques for different situations with emphasis on approaches that are utilized in the two sample problems. Chapter 3 discusses fault detection and diagnostic methodologies in general. Then, the two sample problems, a furnace system and a photolithography process, will be discussed in Chapter 4 and Chapter 5, respectively. Modeling, diagnostic and experimental results of each problem will be presented. Finally Chapter 6 summarizes important contributions of this thesis and discusses some interesting future directions.

Chapter 2

Modeling

A model is an abstract, often simplified, mathematical description of a physical system or a process. A good model does not have to be a complete description of reality, but it should capture certain properties and characteristics germane to a particular goal, such as diagnostics, prediction, classification, etc. Accuracy is often desired in a good model, but depending on the exact nature of the problem at hand, compromises among accuracy, simplicity, efficiency, and robustness have to be considered. With a model, one can achieve better understanding of the system or process in order to control its state, monitor its behavior, and generalize its principles to other domains. Almost all serious scientific endeavors require building some kind of model. In today's engineering applications, having a good model often leads to low-cost and safe experimentation, easy manipulation and exploration of a system or process, and accelerated research and development cycles.

Modeling is the course of creating a model for a system or process. Different modeling techniques are required depending on the detailed nature of a model, such as a physical model vs. an empirical model, a parametric model vs. a non-

parametric model, etc. There are also some common practices in all modeling, such as observation and experimentation, model selection and parameter estimation, and model verification and assessment [13]. Modeling often faces a number of challenges, including complexity, non-linearity, and dynamics of a system or process, limitations in metrology and numerical computation, uncertainties such as systematic and stochastic errors, insufficient time and data, hidden system variables, and simply poor understanding of the true nature of the system in question.

Modeling is especially important for modern semiconductor manufacturing. Due to the extremely high complexity and economic risk, good modeling plays a pivotal role in equipment diagnostics and process control [1, 122]. Single errors in a single subsystem may cause the loss of an entire line of wafers; many system parameters are difficult to measure with confidence and recipe drifts are common; early detection and real-time monitoring are essential to prevent serious problems from happening; and development of control strategies are often too expensive on real hardware, while it is producing real product. This chapter focuses on general modeling issues relevant to modern semiconductor manufacturing, in particular, physical modeling and empirical modeling, and techniques for model parameter estimation.

2.1 Physical Models

Here let us make a rough distinction between physical and empirical models. Physical models are mostly based on knowledge of physical properties of a system or process, and empirical models are derived mostly from observed inputs and outputs. For example, a furnace may be viewed as a thermal system and a physical

model can be built based on our knowledge of thermodynamics and heat transfer; on the other hand, if we do not have a sufficient physical understanding of a furnace, we can design experiments to collect input data such as supplied power, temperature settings, baking time, and output data such as temperature measurements and wafer quality to infer an empirical, statistical relationship between inputs and outputs. Obviously, physical models are intuitive and convenient, and logical deduction can be made to predict the state of a system or a process unambiguously. On the other hand, realistic systems and processes are often extremely complex and not entirely understood. It is often very difficult or even impossible to build an accurate physical model. For example, a plasma etching system is still too complex to be understood completely [110]. In this situation, we often have to resort to empirical models or consider a simplified physical model with many necessary assumptions. Another drawback of a physical model is its lack of robustness especially when a compromise between model complexity and reality is made. But nevertheless, whenever an accurate and efficient physical model is available, we should consider it before resorting to empirical models.

Most physical systems are inherently non-linear and involve dynamics. Linear approximation models are often used for simplicity [106]. The study of linear dynamic systems is mature, with powerful analysis techniques and a long history of successful industrial applications [26, 61, 103, 107]. However, the non-linearity of a system has to be taken into account when linear approximation cannot meet the requirements of control accuracy or modeling needs [106]. Typically, a linear dynamic system is described by a set of linear differential equations (difference equations in discrete time). Non-linear systems, however, have a range of possible formats.

Many efforts have been made in the semiconductor manufacturing industry to understand the physics behind the processes and utilize the physical knowledge to perform controls [75, 100, 110]. A good example of a process suitable for physical modeling is the thermal process. There are numerous heating processes involved in the product line. Modeling of a furnace system is essential for accurate and stable temperature control. Thermal systems are among those that are well studied. Later in this thesis, we describe how to build a furnace model based on heat transfer theory.

2.2 Empirical Modeling

Empirical models refer to models that are derived from input/output observations of a system or process without explicit use of underlying theory. As discussed previously, the distinction between empirical models and physical models is very blurred, and many problems involve both. For example, knowledge of physical properties is often used to design data collection experiments for empirical modeling. Empirical models are needed when the complexity of a system or process is extremely high and the physics are not well understood, or when simple statistical relationships potentially exist between input/output variables of interest. Empirical models enable fast and inexpensive experimentation and can be chosen from a variety of different types according to the practical consideration of the problem at hand. Statistics and other methods used in empirical modeling can deal with uncertainty in a principled way [33]. On the other hand, empirical models usually lack intuitive physical interpretation and must be based on large amounts of data.

Empirical models can be categorized in several ways, such as parametric vs.

non-parametric, forward vs. inverse, linear vs. nonlinear, deterministic vs. probabilistic, etc. *Parametric models* refer to those whose model complexities (e.g., number of active parameters) do not increase as the number of observation data points increases [11]. Some of the commonly used parametric models include linear regression and various non-linear regression methods [7, 10, 13, 71]. *Non-parametric* models do not assume a particular functional form for the model, but increase model complexity as more data are available. Examples are kernel density estimation [11] and nearest neighbor methods [25]. Yet some other methods have characteristics of both parametric and non-parametric models, such as artificial neural networks and Gaussian mixture models, and sometimes are referred to as *semi-parametric models* [11]. *Forward models* are those that map user-defined inputs to outputs and *inverse models* map outputs to inputs. In many realistic situations, models do not have one-to-one mapping. Next, we describe a few commonly used empirical modeling methods and some relevant applications in semiconductor manufacturing.

2.2.1 Linear Regression

Linear regression (LR) is perhaps the most widely used empirical modeling method in various engineering and scientific applications. A linear model assumes that output variables are linearly dependent on some fixed transformations of input variables [7, 13, 71, 109]. Since the fixed transformation of input variables (forming a design matrix) can be non-linear, the model is more general than a purely linear relationship between inputs and outputs. For example, let (X, Y) be input-output pair, and

$$Y = aX + bX^2 + \epsilon \tag{2.1}$$

is a linear regression model where a and b are coefficients to be estimated and ϵ is additive noise with certain assumed probabilistic distribution (often Gaussian). Linear regression has been intensely studied and well understood [13, 109]. Although exact linearity rarely exists in the natural world, linear regression models offer good approximation for many practical applications. Furthermore, linear regression has some nice properties such as simplicity and transparency (input-output relationship can be expressed by a simple formula). It can help data analysis such as sensitivity tests, principal component analysis, etc. LR is often used first, since it can help assess the complexity of the problem.

2.2.2 Non-linear Regression

Non-linear regression is a natural extension to linear regression where arbitrary non-linear relationships may exist between the model parameters and output variables [10]. They are obviously much more powerful than simple linear models to capture realistic systems and processes. However, non-linear models are much more difficult to determine. They may require some knowledge of the physical system to decide what types of non-linear functions are appropriate. Finally, even when the form of the model is determined, the methods for estimating the parameter values are much more complex and time consuming than that for linear models.

2.2.3 Artificial Neural Networks

Artificial neural networks were originally motivated by conjectures of how neurons in animal brain process information [11, 105]. A typical feed-forward neural network consists of a number of nodes (sometimes called neurons) and links be-

tween nodes. Each node has an output activation potential, which is propagated through the links in one direction. At each node (cf. figure 2.1), a weighted sum of input activation potentials from all incoming links is taken and a transformation (typically a sigmoid function) and thresholding are performed to obtain an output activation potential for that node.

Artificial neural networks have been shown to be good universal function approximators. With a sufficient number of nodes and links, a neural network can model very complex non-linear relationships between input and output variables. Furthermore, neural networks can be trained efficiently using the back-propagation algorithm [11, 104, 105] from a given set of input and output training pairs of data. After each input data point is presented to the network, link weights are adjusted to minimize an error function E (e.g., root mean squared error) measuring dissimilarity between network outputs and desired target outputs. The objective is to assess the source of the error and divide the blame among the contributing weights. In back-propagation, one first finds the partial derivatives of E with respect to top-layer node activation potentials, and then finds that with respect to link weights coming into the top-layer nodes by applying chain rule of differentiation; this process is repeated until we reach the earliest layer of nodes so that we obtain the partial derivatives of E with respect to each link weight in the network[1]. This process takes only $O(W)$ steps where W is the number of tunable link weights. The link weights are then adjusted following the negative direction of the gradients.

One commonly used neural network model is the multilayer perceptron (MLP) [9, 104, 105], which has its nodes arranged in layers and links between nodes are

[1]Thresholding on each node activation potential can be implemented as having an incoming link from a common bias node emitting a constant unit activation potential. Thus, thresholds can be treated in the same fashion as link weights.

only allowed in a single forward direction. A typical three-layer[2] MLP is shown in figure 2.2 (of course, actual number of nodes may vary). It was proven that with a sufficient number of hidden nodes and links, an MLP can approximate any linear or non-linear (even discontinuous) function to arbitrary precision. One of the major drawbacks of using neural networks is their lack of transparency in that even a well-trained neural network offers little intuition about the process it presents.

In recent years, neural networks have been used widely, and successfully, in many engineering applications [6, 16, 19, 53, 59], and have emerged as a potentially powerful technique for many complex semiconductor manufacturing problems. There are some successful examples of applying neural networks in plasma etching process [80, 108] and other processes [59, 89, 90]. Substantial amount of work has been done by Gary May and his colleagues [15, 46, 80] on applying Neural networks to plasma etching processes, Plasma-Enhanced Chemical Vapor Deposition (PECVD) processes and some other processes. Nadi [90] used neural networks in his PhD work for modeling two processes, including a dry-oxidation process of silicon and a Low Pressure Chemical Vapor Deposition (LPCVD) process of polysilicon.

2.2.4 Probabilistic Graphical Models

Graphical modeling is a technique that combines probability and graph theory [62, 72]. This method is becoming increasingly popular during the past decade. Graphical model families unify a wide variety of known models, such as Bayesian Networks (BN) [48, 49, 98], Hidden Markov Models (HMM) [62, 99], Kalman

[2]Here "three-layer" refers to three layers of nodes. Some authors prefer naming an MLP by the number of link layers, in which case we would call our example a "two-layer" MLP.

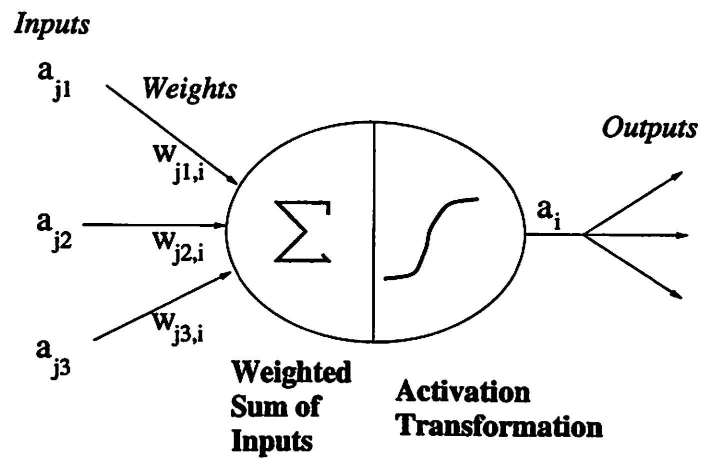

Figure 2.1: An artificial neuron.

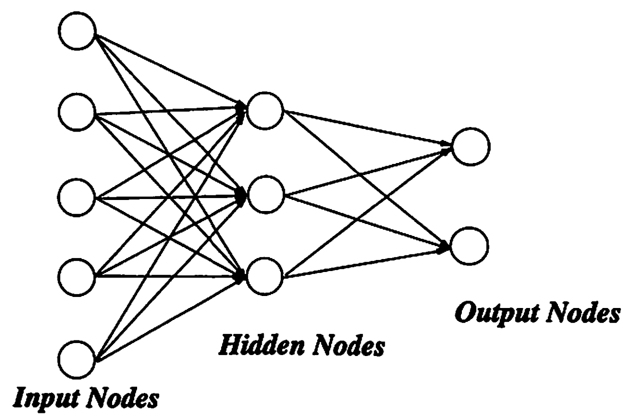

Figure 2.2: A multi-layer perceptron (MLP) network.

Filters (KF) [47, 62], Factor Analysis (FA) [33], Gaussian Mixtures (GM) [11], etc. Graphical models capture uncertainty and probabilistic relationships among observable and hidden variables of a static or dynamic system, and can be constructed using domain knowledge or learned from data in a principled way. The "inference" (joint probability estimation) of arbitrary sets of observable or hidden variables are theoretically possible for a graphical model, although in general it involves exponentially large amounts of computation. Many approximation algorithms for inference in graphical models have been proposed such as variational methods [63], mean-field methods [62], and Monte Carlo sampling [31, 105].

2.2.5 Fuzzy Logic

Fuzzy logic [124, 125, 126] came about when researchers became unsatisfied with the rigidity of classical logic, and observed a gap between the fuzzy nature of real world entities and the crispness of existing logical and mathematical tools. Fuzzy logic assumes statements are true only to a certain degree and variables take on a "fuzzy" set of values with different degrees. Fuzzy logic is a fundamentally different way of dealing with imprecision and uncertainty than other approaches, such as probability theory. Many successful engineering applications, including semiconductor manufacturing, have been successfully modeled by fuzzy logic and fuzzy set theory, for applications such as fuzzy automatic control and fuzzy pattern recognition [8, 45, 77, 81].

2.3 Parameter Estimation

Once the structure of a model is determined for a given problem, an essential task is parameter estimation. For both physical and empirical models, there are some common approaches to estimating an optimal set of parameters. First, experiments need to be implemented (except for some known physical models) to collect data used for parameter estimation. An appropriate cost metric has to be selected to evaluate the optimality of a parameter estimate. Optimization algorithms [65] are employed to find optimal parameters for the selected metric. Parameter estimation is usually performed on a different set of data (training set) than that to be used for testing (test set). One issue that has to be considered is the possibility of overfitting [11]. In this section, we describe general parameter estimation techniques and issues related to some of the above models in more detail.

2.3.1 Design of Experiments (DOE)

When the structure of the model is chosen, either based on physical considerations or empirical assumptions, experimental data are needed to extract parameters and evaluate model accuracy. The task of design of Experiments (DOE) involves defining experiment goals and objectives, planning and conducting experiments and collecting data in desired formats [38]. A successful DOE requires a good understanding of every component of the system and equipment utilized in the experiments. The data used for modeling need to be fault-free, which means that random noise is allowed for processes or measurements but no systematic error should be present.

2.3.2 Cost Function

There are a number of different cost functions commonly used for parameter estimation, and each is more appropriate under certain conditions and inappropriate under others. Minimum Mean Squared Error (MMSE) [11] is often used in regression problems where the expected value of the squared difference between model output and desired output is minimized. However, MMSE is usually not appropriate for a classification problem (whose desired targets are discrete class labels) where Mean Cross Entropy (MCE) [11, 62] may be more suitable. For problems endowed with probability distributions, Maximum Likelihood (ML) [11] is often used where estimation is based on maximizing the conditional probability of the observed data given specific parameter estimates. For certain problems, such as those having data with Gaussian distributions, ML is equivalent to MMSE. Bayesian statistics suggest the use of Maximum a Posteriori (MAP) [11] estimation of parameters, where the posterior probability of a model (M), given observations (D), can be written as

$$p(M|D) = \frac{p(D|M)p(M)}{p(D)} \tag{2.2}$$

according to Bayes' rule [11]. Then the MAP estimate is the model M^* that maximizes the $p(M|D)$. Notice that if the prior probabilities of different models $p(M)$'s are equal, since the marginal probability of data $p(D)$ is the same for all models, MAP is equivalent to ML. Many other commonly used cost metrics are based on information theory, for example, Maximum Mutual Information (MMI) [22].

2.3.3 Optimization

For a given cost metric, one would like to find a global optimal set of parameters for a given model. Some optimization problems can be solved analytically. For example, solving a system of normal equations gives the optimal parameters for linear regression [7, 13, 109]. Yet in most other problems, a gradient descent (or ascent) is employed. This involves initialization of parameters, calculation of gradients of the cost function with respect to parameters to be optimized, and testing of convergence. Many algorithms exist to speed up the gradient search, many based on higher order gradient estimation (e.g. Hessian matrix) such as the Quasi-Newton family of methods [11]. One common problem with gradient search for non-convex problems is convergence to local minima. Remedies include random restart gradient search [105], simulated annealing [21, 105], etc. Although these remedies may guarantee finding the global optimum theoretically, they often take too long to be cost-effective. The success of gradient-based methods depends on the structure of the cost function space (e.g. convexity, smoothness and the number of local minima) as well as the accurate estimation of gradients. For example, the success of neural networks can be largely attributed to the discovery of the efficient back-propagation algorithm for gradient estimation [104].

2.3.4 Model Overfitting

Parameter estimation is often performed on a separate set of data (training set) from the set used for evaluating the model (test set) [11]. Although we often assume both data sets come from the same underlying distribution, a finite, often noisy, training set cannot guarantee that the optimality of parameters carries over

to the test set. This is the problem of overfitting where error rates are much lower on the training set than on the test set since a model may be fitted on the idiosyncrasies and noises of training data instead of on the general patterns of the underlying system behavior. This problem is especially severe for very complex models and for problems with limited training data [11].

There are several methods to alleviate the problem of overfitting. One way is to limit the size of the potential parameter space such that a restricted model is actually used. For example, for a linear regression problem, one can limit the transformation of input variables to some simple form; for neural network, one can reduce the number of hidden nodes and links, which leads to fewer active parameters. One can also effectively limit the parameter space by adding a regularization term [116]. For example, Minimum Description Length (MDL) principle [102], Akaike Information Criterion (AIC) and Takayuki Information Criterion (TIC) [76] are such regularization methods (they are often referred to as model selection methods as well). Another popular method, especially useful for early stopping in gradient search optimization, is to evaluate the cost function periodically on a cross-validation data set that is separate from the training set [105]. The training process should be terminated when the cost function stops decreasing on the cross-validation set.

Chapter 3

Diagnostics

Any realistic physical system must face the possibility of noise and failure. The task of diagnostics is to monitor the behavior of a system and infer the values of system variables so that if anything abnormal occurs, we would be able to locate the source of error and propose preventive or corrective action.

In order to achieve high reproducibility, in a semiconductor manufacturing line, it is desired to keep all system parameters relatively stable while a recipe is being run. Parameter drifts may bring significantly adverse effects if not detected quickly and accurately. Persistent parameter drifts not only cause loss of wafers, but may also be precursors to severe system damage. Most semiconductor manufacturing lines are themselves very expensive, and downtime is costly. Unfortunately, real-time measurement of many system parameters of interest is often very difficult, expensive and inaccurate. Therefore, it is important to be able to estimate parameter drifts through diagnostic methods.

In this chapter, we give a short review of some popular techniques for detecting and isolating component faults [37, 55, 96, 97, 101]. These techniques involve many

different fields, such as control theory, statistics, signal processing, filter theory, etc. In practice, they may be applied individually or combined with each other.

3.1 Sources of Errors

Before describing techniques for error detections, we first give a synopsis of different types of errors, discuss common error models, and introduce potential error sources in semiconductor manufacturing.

3.1.1 Types of errors

We may define an error as a deviation between a measured or computed value of a system variable and the desired value. It is important to distinguish between two error types, the systematic error and the random error, since different solutions are required to deal with each of them. A systematic error is caused by any factors that systematically affect the value of the variable across samples. For example, if a thermostat is malfunctioning in a furnace and always sets the baking temperature to be three degrees higher than the specified value, we may call the resulting deviation a systematic error. Thus, a systematic error tends to affect variables in a fixed way and is often referred to as *bias.* On the other hand, a random error is caused by any factors that randomly affect measurement of the variable across the sample. For example, the random fluctuation of temperature measurements made on a furnace in a steady state may be called a random error. Such random error adds variability to individual measurements but does not affect the average across many samples.

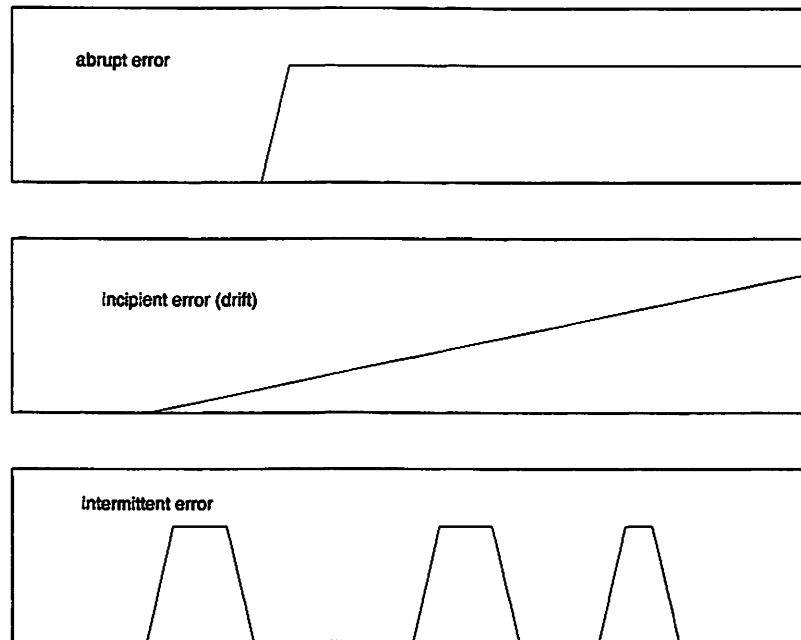

Figure 3.1: Error type according to time-dependency.

Errors can also be categorized by their time-dependencies (cf. figure 3.1). An *abrupt* error has a sharp onset; an *incipient* error has a smooth onset and is also called a *drift*; an *intermittent* error has alternating on-and-off patterns [56]. Among these three time-dependencies, *drifts* are the most difficult to estimate and are also the most common type of errors in semiconductor manufacturing.

3.1.2 Error Modeling

To eliminate or reduce a systematic error, we need to locate the source of the error and make appropriate adjustments. However, this approach does not work for

random errors due to their nature of uncertainty. What we can do is to study the statistical properties of random errors and minimize their effect using statistical approaches. Random errors can be characterized statistically by their probability distribution functions or probability density functions. The Gaussian distribution is often a good approximation of the probability distribution of a random error, and in such cases, estimating the mean and variance of the error is sufficient. For other cases where the probability distribution of a random error is very different from the Gaussian, we have to resort to other techniques for estimating the probability density function [117, 118]. Commonly available techniques include the histogram method [11], kernel density estimation method [11], etc. For many physical systems, we may be interested in modeling the random dynamic behavior of system parameters over time. Time series models can be applied in such cases and they are often very useful for estimation and prediction [14, 39].

3.1.3 Variations in Semiconductor Manufacturing Systems

Modern semiconductor manufacturing relies on highly complex processes and very expensive equipment, which involve hundreds, and even thousands, of active parameters that have to be controlled properly in order to achieve reliable system behavior. Variations in such a complex setting are abundant and may be caused by changing environmental conditions, uncertain material properties, machine instabilities, as well as many human factors. Many of these variables are difficult to measure accurately. Metrology tools are only available for a limited number of variables and are often expensive and time-consuming, which makes real-time diagnostics particularly difficult to realize. So we have to rely on those parameters that can be measured to estimate those that cannot be measured. Parameter

estimation (cf. Chapter 2) is therefore very important for detecting variations. Unfortunately, no general solution exists that can deal with all situations. Although the general statistical approach remains similar, specific solutions are system- or process-dependent. Therefore, in this thesis, we illustrate our statistical diagnostic approaches on two different problems, a furnace system and a photolithography process.

3.2 Fault Detection and Classification

A fault is an unpermitted deviation of at least one characteristic property or variable of the system from acceptable behavior. Fault detection is thus the determination of faults present in a system. Fault classification is to assign a detected fault into one of the prescribed fault classes according to its location, size and time-variant behavior.

3.2.1 Fault Detection for Linear Dynamic Systems

Fault detection for linear dynamic systems has been studied extensively, based on well-established linear system theory. It had been successfully applied to many different areas, such as vehicle dynamic systems, aircraft control systems, robotics, etc. Several major techniques for model-based fault detection have been introduced in the literature [44, 50, 55, 56, 97, 117].

Based on modern control theory, when the model of a linear dynamic system is known, state observers can be developed for state estimation, which is then used for fault detection. Jones [61] and Willsky [123] developed fault detection filters

for state estimation. Mehra and Peschon [84] suggested using Kalman Filters [47] to set up dedicated observers for fault detection. Willsky [123], Clark [20] and Frank [36] used bank of observers for similar purposes.

Parameter estimation is also one of the most important fault detection methods. We discussed parameter estimation techniques in detail in Chapter 2 (cf. Section 2.3). Examples of applying parameter estimation to fault detection have been reported in [35, 55, 101].

For linear dynamic systems, the parity equation approach is also a popular method of fault detection. Parity equations based on both state space equations [18, 96] and input-output equations [41, 51] have been studied.

While most of the techniques above were developed based on linear systems, some of them have also been adopted for fault detection of nonlinear dynamic systems under appropriate assumptions [36, 40].

3.2.2 Fault Classification Methods

When a fault has been detected in a system, classification techniques are needed to identify what type of fault it is. If we have the knowledge of exactly what possible fault types there may be and have access to a sufficient number of labeled data samples of each fault type, we can apply a variety of standard pattern recognition techniques to perform fault classification. In each case, supervised learning algorithms can be adopted to train classifiers on the labeled training data [86, 105].

Statistically, the optimal classification performance is given by Bayes classifiers [86], which give the most probable classification of a new instance given the

training data. The Bayes classifier combines the predictions of all possible hypotheses, weighted by their posterior probabilities given the training data. For problems whose hypotheses are equivalent to classes, the Bayes classifier amounts to choosing the class with the maximum *a posteriori* (MAP) probability [86].

The k-Nearest-Neighbor (kNN) [25] is an instance-based learning technique that is intuitive and can be readily used for classification. It operates on the geometric space defined by the variables of interest with some distance metric (e.g., the L_2 norm). For each new data vector to be classified, its distances to each of the saved data samples are computed and the k data samples that are nearest to the new data vector are identified. Then, the new data vector is said to belong to the class that has the most instances in the k nearest data samples.

Given labeled training data vectors, it is often possible to approximate decision rules by performing linear or non-linear regression for each class. If the regression is based on polynomials of the data vector, it is often called the Polynomial Classifier (PC) [73].

Artificial Neural Networks (ANN) include a family of popular classification tools, such as Multi-layer Perceptrons (MLPs) and Radial Basis Function (RBF) networks [11]. Since ANNs are capable of learning arbitrary nonlinear functions, they can form highly complex decision boundaries for very difficult classification problems.

Classifiers based on fuzzy set theory [125] and fuzzy logic [126] are also very popular, especially for problems with fuzziness and uncertainty. They address the problem of rigidity in classical logic and can be constructed from knowledge as well

as in a data-driven manner. Fuzzy inferences [23, 69, 124] employ a number of fuzzy if-then rules for approximate reasoning and include fuzzification and defuzzification steps to interface with conventionally defined problems. Fuzzy classifiers have also been combined with ANNs to take advantage of both frameworks, and the resulting hybrid systems are often called Neuro-fuzzy classifiers [8, 45, 81].

There also exist other popular classifiers that can be used for fault classification, such as statistical decision trees [86, 105] and kernel-based classifiers including Support-Vector Machines (SVMs) [24].

3.2.3 Fault Detection and Process Control in Semiconductor Manufacturing

Currently, fault detection and classification in semiconductor manufacturing are rather primitive, and simple forms of Statistical Process Control (SPC) [30, 112] are dominant in industry. Directly measurable parameters are monitored and alarms are set if some parameter exceeds a predefined tolerance level, and sometimes when necessary, machines are shutdown accordingly. Control charts are used to identify alien data points and the control limits are usually defined based on the statistics of parameters in normal system operation. The control rules are defined according to the needs of specific tasks. Standardized rules, such as the Western Electric Rules [66], are also widely used in industry.

A number of advantages of such a simple process control mechanism may justify its popularity in the semiconductor industry today. For example, SPC is easy to implement, easy to understand for field process engineers (with relatively little training), and reliable under familiar situations. However, several serious disadvan-

tages of SPC suggest we seek more sophisticated process control strategies. First of all, setting control limits is only suitable for steady state operations and may be quite problematic for dynamic operations. Secondly, SPC only detects changes of directly measurable parameters, but system states also depend on many important parameters that are not directly observable. Thirdly, SPC can only find statistically significant excursions without giving explanations of the real source of the problem or how to fix it. Finally, SPC usually does not give early detection of small faults, which often leads to more serious faults and expensive machine shutdowns and repairs.

It is therefore worth the effort to develop Advanced Process Control (APC) strategies incorporating adaptable fault diagnostics that provide early detection of parameter drifts, give analysis of true error sources, and suggest possible fixes of problems. Although the APC strategies are becoming the current trend in the semiconductor industry, no unified standard exists and even the scope is not well defined. Due to the complex nature of semiconductor manufacturing processes, the current APC focus is on multivariate analysis, principal component analysis (PCA), factor analysis, data modeling, etc.

Regarding control techniques, Run-to-Run (R2R) control has been receiving increasing attention in recent years in the semiconductor manufacturing community. In a typical R2R control, feedback control techniques are utilized to adjust recipe settings automatically. This, however, requires accurate process modeling, high-performance metrology and adjustable actuators, which are all difficult to obtain for most processes. Much work has been done in this area [32, 88].

In this work, we focus on applying parameter estimation-based fault detection

methodologies to different semiconductor equipment and processes.

3.3 Sensor Fusion

Any sensor is subject to failures, noise, and other uncertainties. Therefore, we cannot put complete faith on the results of any single sensor. Sensor redundancies, both physical and functional, are often exploited to improve measurement reliabilities. Sensor fusion is a technique that integrates information from different sensors, and uses the integrated information to improve and interpret sensor readings, as well as to monitor sensor performances. Much work has been done in this area [2, 4, 44, 68, 117, 119].

Sensor fusion generally involves two major steps, a validation step and and an information-fusion step. The function of the validation step is to identify erroneous sensors. This amounts to making decisions on whether a sensor reading should be discounted, removed or adjusted. One example of validation is to perform smoothing, which removes or reduces noise (in this case, noise is the identified error) using filtering techniques such as a Kalman filter. A validation range (also called a validation gate) can be defined (based on the prediction of the filters in the case of Kalman filter). A sensor reading within the validation range is assumed to be valid and used in the subsequent steps. Each valid sensor measurement is assigned a confidence "value" as a measure of how confident we are in that particular measurement. Confidence values can be defined in a number of different ways [44, 64, 67]. For example, we may define it as a function of the distance between the measurement and the prediction [2]. Based on the results of sensor validation, there exist a variety of methods to perform information fusion. One

simple example is to calculate a weighted average of all valid sensor measurements where the weights are the (normalized) confidence values.

Substantial research work on sensor fusion has been done in the Berkeley BEST Lab (Berkeley Expert System Technology laboratory) led by Professor Alice M. Agogino [2, 3, 42, 43, 117]. Alag suggested a probabilistic framework in his PhD dissertation [4]. Goebel developed a fuzzy sensor validation and fusion in his PhD research [44]. The two approaches were compared [43] and both of them had been proven to be successful for different applications [5, 118, 119].

In this work, we apply a simple, easy-to-implement sensor validation and fusion technique on the furnace problem to enhance diagnostic reliability in Chapter 4.

Chapter 4

Diagnosing Malfunctions in Thermal Systems

4.1 Introduction

Precise control of process temperature becomes important in today's semiconductor industry due to the high accuracy needed in each process step. Multi-zone batch furnaces are widely used, and high reliability of furnace systems is a crucial factor in achieving high product yield. Temperature control itself is not very difficult. A PID controller will do a good job as long as the furnace system is working well. However, uncertainty caused by system or sensor failures may degrade reliability. Therefore monitoring the health of the thermal system becomes an important task. Many researches have been focused on controlling cross-wafer temperature uniformity [34]. But little work has been done on diagnosing the health of the thermal system itself.

In this chapter, we study a furnace system as an example of diagnostics based on physical modeling. We show how to find the structure of the system model

based on the physics of the thermal system. We design experiments to extract the system parameters. We develop an approach to detect failures including temperature sensor failures, power supply failures, and system faults for multi-zone furnace systems. For a certain fault pattern, multiple data sets may be collected. Therefore, we can treat these data sets as different sensors that can measure faults. A sensor fusion technique is then used to integrate these detection results to find a more reliable estimate of the faults.

A simple linear dynamic model with a closed-loop PID-based temperature controller is adopted to describe the overall thermal system based on physical considerations. The parameters are estimated using the data collected in experiments conducted on an actual five-zone furnace in the Berkeley Microfabrication Laboratory.

There are three major fault sources of concern in this work. The first is the failure of the temperature sensors (thermocouples). We are particularly interested in the parameterized bias of these sensors. The furnace we used in the experiments has five thermocouples that monitor the temperature of the five zones respectively. Each of these sensors might develop a bias or its random noise might increase.

The second fault source is power supply inefficiency, i.e., the heaters fail to supply enough power requested by the controller. In our furnace, there are five independent heaters corresponding to the five zones, so the power inefficiencies of the five heaters are considered independently.

Finally, the insulation failure of the tube is another fault source, which corresponds to the change of thermal resistance parameters in the system model. In this

work we define fault parameters to evaluate the faults and develop a generalized system model that includes all three types of faults.

Our method depends on real-time collection of temperature readings and of the power delivered to the heating elements. This information is processed in conjunction with the thermal model of the system. A statistical model classification approach [120] is utilized in this problem. Different models are created corresponding to the fault-free system and to systems with single or multiple failures. Analysis of real-time temperature settings, temperature sensor readings, and requested power outputs, determines the model that fits the best. Thus, the most likely failure model is identified. For convenient fault parameter estimation, we use a least square matching algorithm. Different data types have been used for different fault detection purposes. Finally, a sensor fusion technique [2, 64] is used to enhance diagnostic reliability.

Experiments were performed on a five-zone furnace system used for dry oxidation. Simulations of a fault-free furnace, a variety of failure models, sensor fusion, and fault detection algorithms were performed in MATLAB. All models were based on both physical constraints and empirical data fitting parameter estimation is discussed. In the next section, we will describe the system model of the furnace.

4.2 Furnace Modeling

In this work, we use a typical furnace system as an example to suggest a general approach to detect the furnace system failures that could be the primary factors of

temperature inaccuracy. The system we chose to model is a five zone, 6-inch tube made by Tylan and controlled by the Tystar controller. This system is used in the Berkeley Microfabrication Laboratory for routine dry oxidation of 6-inch wafers. We start with setting up the system model. First, we study the structure of the model based on physical consideration of the thermal system. Then, we perform a series of experimental runs with the actual furnace in order to extract the model parameters.

4.2.1 System Modeling

Figure 4.1 is a schematic of a five-zone furnace system (not to scale.) Five heater elements supply power to each of the five different zones, respectively. Five thermocouples located in the five zones provide readings for monitoring and control.

The behavior of the heat flow and temperature in a thermal system can be described mathematically by expressions similar to those used in an electrical network [54, 85]. The advantage is that we can then use electrical network theory (e.g., Miller theorem) [87] to gain insight on the significance of each parameter. This insight is very important for fault parameter estimation. We take the temperatures at the thermocouple locations as the temperatures of the five zones. For each zone, heat is supplied by the heating element, heat exchange happens between zones, and between the furnace tube and the environment. All three modes of heat transfer, conduction, convection and radiation, happen in such a furnace system. But since this furnace tube runs at very low pressure, convection by air can then be ignored. Also, most heat exchange transferred by radiation

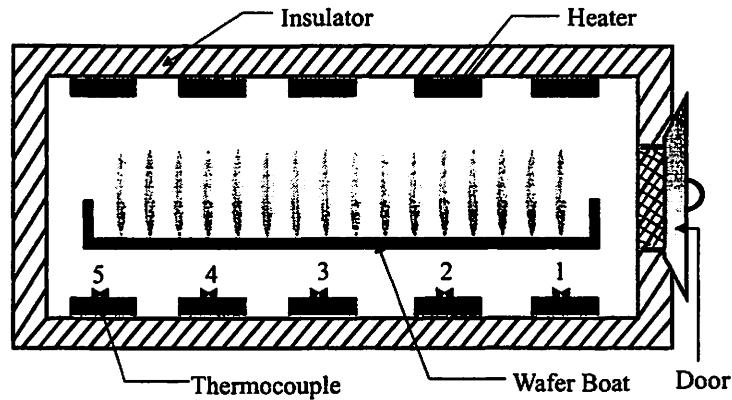

Figure 4.1: Schematic of furnace system.

happens within each zone, so we assume the effect of radiation across zones is not significant. Therefore, in this work, we consider conduction as the major mode of heat transfer. We use thermal resistance to capture the relationship between heat supply and temperature change, so that the analogy between the heat flow system and the DC electric circuits can be utilized to simplify the analysis. Here we assume that for a specific furnace system, the values of the thermal resistances and capacitances are fixed (not a function of temperature), since conduction is the major mode of heat transfer here [85]. This may not be exact, but it can make the problem easier and it is proven to be a valid assumption for the system we study. Figure 4.2 depicts the electrical equivalent of the thermal system, where the heaters $\{P_1, P_2, P_3, P_4, P_5\}$ are equivalent to current sources, and temperatures $\{T_1, T_2, T_3, T_4, T_5\}$ are equivalent to potentials. The correspondence of the thermal system to its electrical equivalent is found in Table 4.1.

Since this is a linear circuit, we can easily develop the following continuous

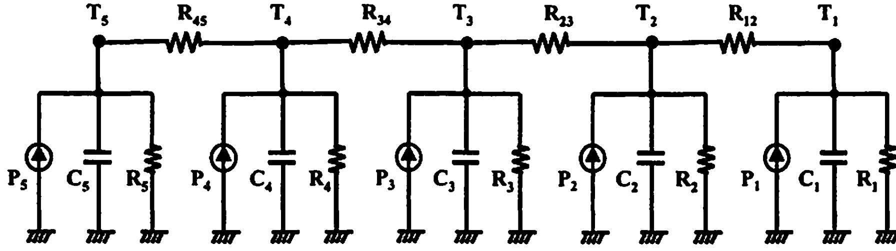

Figure 4.2: Electrical equivalent of the thermal system.

Thermal Parameter	Electrical Parameter
Temperature: T(K)	Voltage: V(V)
Heat Flow, Power: P(W)	Current: I(A)
Resistance: R(K/W)	Resistance: R(Ω=V/A)
Capacity: C(J/K)	Capacitance: C(F=A· s/V)
Ambient Temperature	Electrical Ground

Table 4.1: Electrical equivalents of thermal parameters.

time open loop system equations.

$$C_1\frac{dT_1}{dt}+\frac{T_1-T_a}{R_1}+\frac{T_1-T_2}{R_{12}}=P_1; \tag{4.1}$$

$$C_2\frac{dT_2}{dt}+\frac{T_2-T_a}{R_2}+\frac{T_2-T_1}{R_{12}}+\frac{T_2-T_3}{R_{23}}=P_2; \tag{4.2}$$

$$C_3\frac{dT_3}{dt}+\frac{T_3-T_a}{R_3}+\frac{T_3-T_2}{R_{23}}+\frac{T_3-T_4}{R_{34}}=P_3; \tag{4.3}$$

$$C_4\frac{dT_4}{dt}+\frac{T_4-T_a}{R_4}+\frac{T_4-T_3}{R_{34}}+\frac{T_4-T_5}{R_{45}}=P_4; \tag{4.4}$$

$$C_5\frac{dT_5}{dt}+\frac{T_5-T_a}{R_5}+\frac{T_5-T_4}{R_{45}}=P_5. \tag{4.5}$$

where T_a is the ambient temperature (ground of the circuit network in figure 4.2).

For simulation convenience, The discrete time system equation (Eq 4.6) can be converted from the continuous time system equation (Eq 4.1~Eq 4.5).

$$\frac{dT_i}{dt} \approx \frac{T_i(k+1) - T_i(k)}{t_{k+1} - t_k}, \quad i = 1, \ldots, 5 \tag{4.6}$$

where $t_{k+1} - t_k = \Delta t = 1, k = 0, 1, 2, \ldots$.

One-minute time step is used, which is sufficient, since thermal systems are relatively slow. Let

$$\begin{aligned} \mathbf{T}(k) &= [T_1(k)\ T_2(k)\ T_3(k)\ T_4(k)\ T_5(k)]^T; \\ \mathbf{P}(k) &= [P_1(k)\ P_2(k)\ P_3(k)\ P_4(k)\ P_5(k)]^T. \end{aligned} \tag{4.7}$$

where bold face letters are used to represent data matrices.

The discrete time furnace system equation then becomes Eq 4.8:

$$\mathbf{T(k+1) = AT(k) + BP(k)} \tag{4.8}$$

where

$$\mathbf{A} = \begin{bmatrix} 1-\frac{1}{R_1C_1}-\frac{1}{R_{12}C_1} & \frac{1}{R_{12}C_1} & 0 & 0 & 0 \\ \frac{1}{R_{12}C_2} & 1-\frac{1}{R_2C_2}-\frac{1}{R_{12}C_2}-\frac{1}{R_{23}C_2} & \frac{1}{R_{23}C_2} & 0 & 0 \\ 0 & \frac{1}{R_{23}C_3} & 1-\frac{1}{R_3C_3}-\frac{1}{R_{23}C_3}-\frac{1}{R_{34}C_3} & \frac{1}{R_{34}C_3} & 0 \\ 0 & 0 & \frac{1}{R_{34}C_4} & 1-\frac{1}{R_4C_4}-\frac{1}{R_{34}C_4}-\frac{1}{R_{45}C_4} & \frac{1}{R_{45}C_4} \\ 0 & 0 & 0 & \frac{1}{R_{45}C_5} & 1-\frac{1}{R_5C_5}-\frac{1}{R_{45}C_5} \end{bmatrix} \tag{4.9}$$

$$\mathbf{B} = \begin{pmatrix} \frac{1}{C_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{C_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{C_3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{C_4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{C_5} \end{pmatrix}. \tag{4.10}$$

All the thermal resistances and capacitances in the above equations are unknown and must be estimated using experimental data.

While the open loop thermal system is considered equivalent to a linear circuit, the actual behavior of the temperature controller must be taken into account. In our furnace system, a heuristically enhanced PID controller is used for temperature control. The details of the built-in controller are not completely known to us, so we have to assume a likely structure and extract the parameters by observing experimental data. Note that knowing the exact controller behavior is helpful but not necessary. This is an advantage of our method, since it takes significant effort to decode the vendor's design, and we want our methodology to be able to compensate for this type of uncertainty.

The overall structure of the furnace system is shown in figure 4.3. If we consider the thermal system and the controller together as our plant, the only input to the plant are the temperature setting trajectories, which are designed by the user according to the needs of a particular process. The noise to the sensors is assumed to be additive, independently, identically, normally distributed (IIND) for each zone of the furnace.

4.2.2 Parameter Extraction

Once the system structure is fixed, the free parameters in the system equations have to be extracted from experimental data. These parameters include all thermal resistances and capacities. For parameter accuracy, several sets of experimental data need to be collected, which are generated when the system is working properly. Figure 4.4 shows a typical run of the furnace system. The curves in the figure are

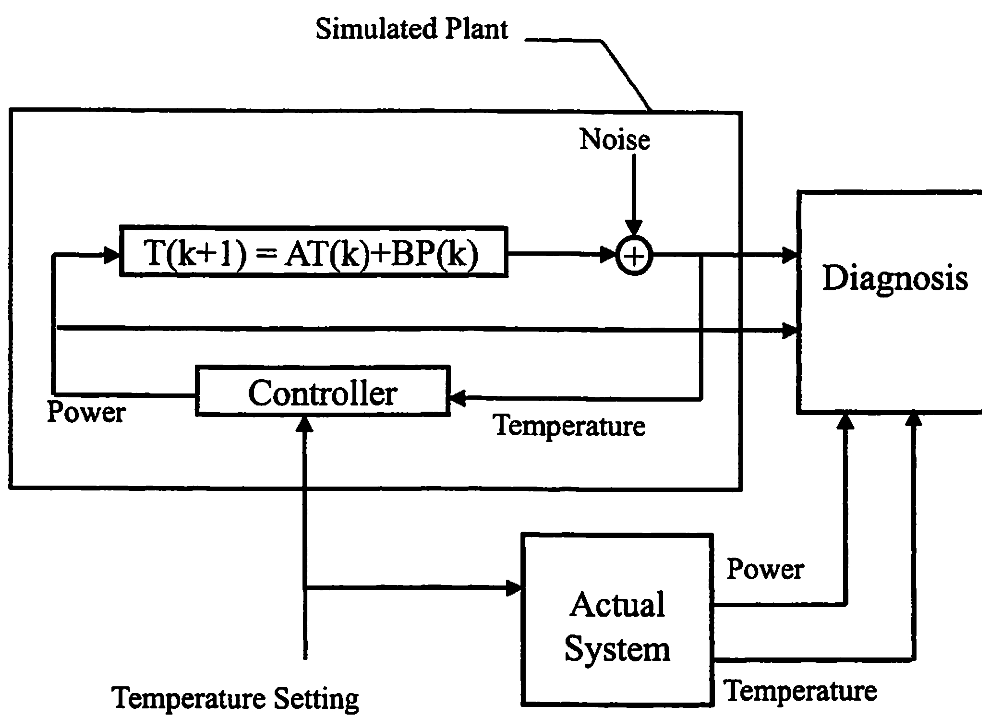

Figure 4.3: Overall system structure.

the temperature and power profiles of a dry oxidation process.

Initially the system was at 750^oC (standby temperature). The desired final temperatures were set to 1000^oC for all five zones. It takes about 30 minutes for the temperatures to stabilize. The oxidation lasts for 3 hours and then the temperatures are set back to the standby level of 750^oC. Note that zone #1, which is next to the furnace door, needs more power than other zones to compensate for the additional heat loss.

Most processes require uniform temperatures inside the furnace tube. However, special temperature settings are necessary in the experimental design for the purpose of system parameter extraction. In this work, we set a temperature gradient across the five zones with a difference of 10^oC between adjacent zones (e.g., zone #1: 930^oC, zone #2: 940^oC, zone #3: 950^oC, zone #4: 960^oC, zone #5: 970^oC) as shown in figure 4.5. If this difference is too small, we would not be able to extract some of the parameters accurately, and if it is too large, the control system of the furnace might not be able to achieve and sustain it.

From the system equations (Eq 4.8) we obtain the steady state expression depicted in Eq 4.11 $\sim$ Eq 4.15.

$$\frac{T_1 - T_a}{R_1} + \frac{T_1 - T_2}{R_{12}} = P_1; \tag{4.11}$$

$$\frac{T_2 - T_a}{R_2} + \frac{T_2 - T_1}{R_{12}} + \frac{T_2 - T_3}{R_{23}} = P_2; \tag{4.12}$$

$$\frac{T_3 - T_a}{R_3} + \frac{T_3 - T_2}{R_{23}} + \frac{T_3 - T_4}{R_{34}} = P_3; \tag{4.13}$$

$$\frac{T_4 - T_a}{R_4} + \frac{T_4 - T_3}{R_{34}} + \frac{T_4 - T_5}{R_{45}} = P_4; \tag{4.14}$$

$$\frac{T_5 - T_a}{R_5} + \frac{T_5 - T_4}{R_{45}} = P_5. \tag{4.15}$$

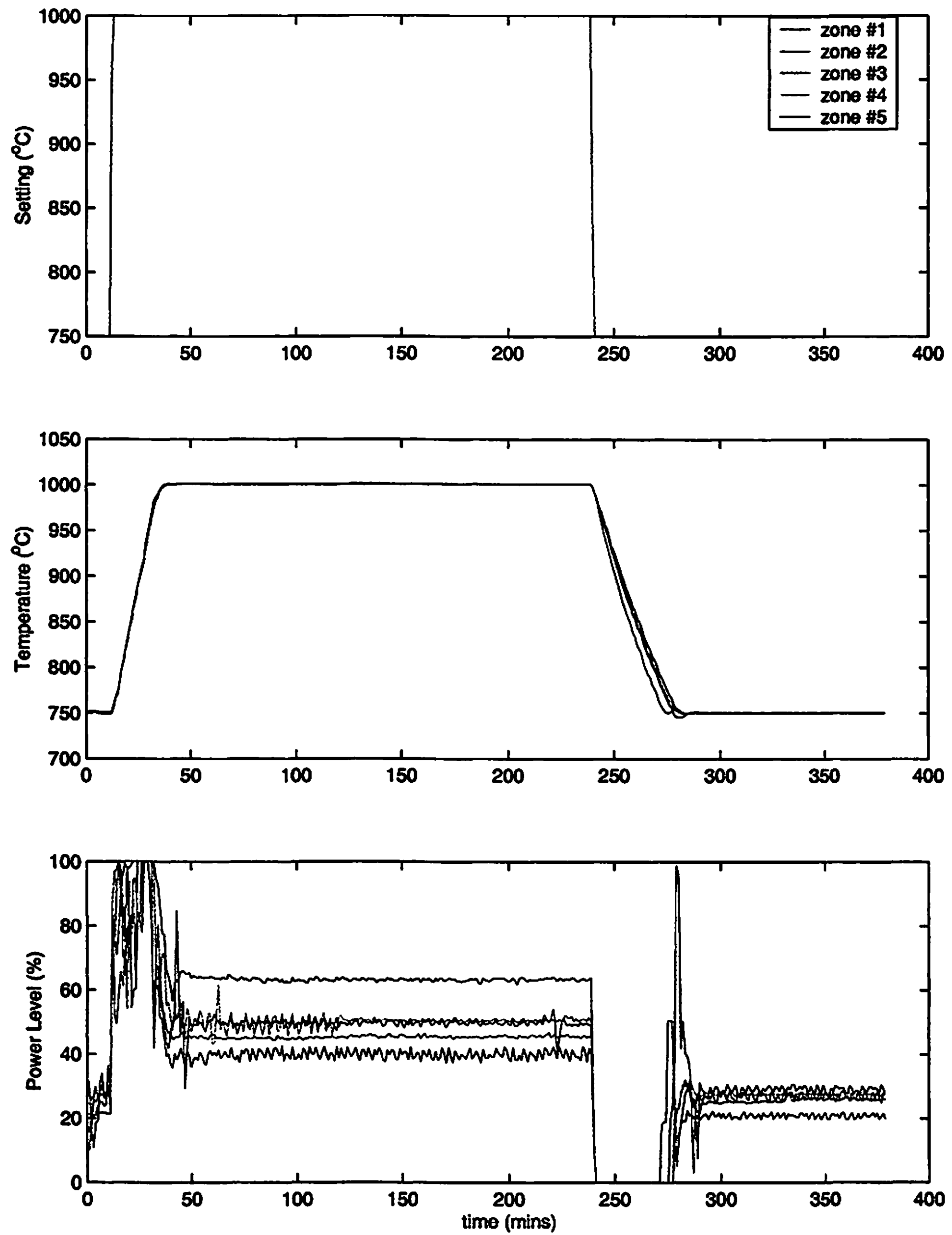

Figure 4.4: A typical data set for a dry oxidation furnace.

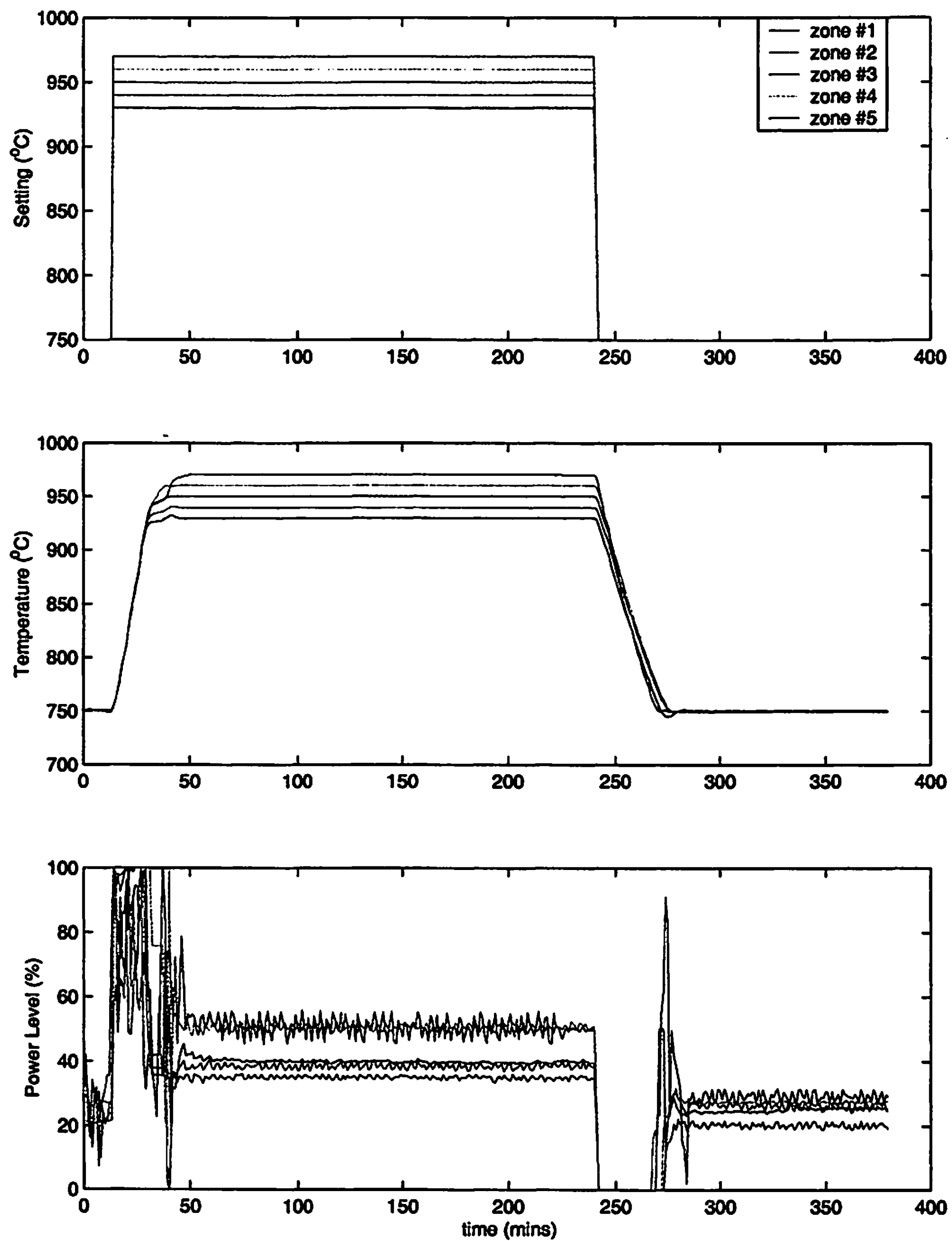

Figure 4.5: Data set used for parameter extraction.

From the above steady state equations we find that the thermal capacitance parameters do not play a role at steady state. By measuring the temperature and power delivery of the five zones at steady state and at various temperatures, we are able to find the thermal resistances using the steady state equations. Figure 4.6 shows an example of two sets of steady state data of zone #1 at different temperatures. All other zones have similar data trajectories.

It is also observed that when the measured temperature is higher than the temperature setting by more than 10^oC, the built-in controller would simply set the power delivery to zero. With the power completely off, the system becomes:

$$\mathbf{T}(k+1) = \mathbf{A}\mathbf{T}(k) \tag{4.16}$$

Although the exact control algorithm is unknown, the transient part of the data (cooling down) could be used to estimate the thermal capacitances (assuming the thermal resistances are already obtained from the steady state data) since the power output is exactly known. Figure 4.7 depicts an actual cooling transient for zone #1.

After all the open loop system parameters are determined, we adjust the parameters of the simulated PID-based controller. The heating up part (cf. figure 4.8) of the data is used for this purpose by adjusting controller parameters of the simulated plant such that the simulated temperature and power profile best fit the experimental profile.

In this way all the parameters needed by the thermal model and its simulated controller are identified. Note that this is done under the assumption that the data we used for estimation are all fault-free. In the next section, fault diagnostic

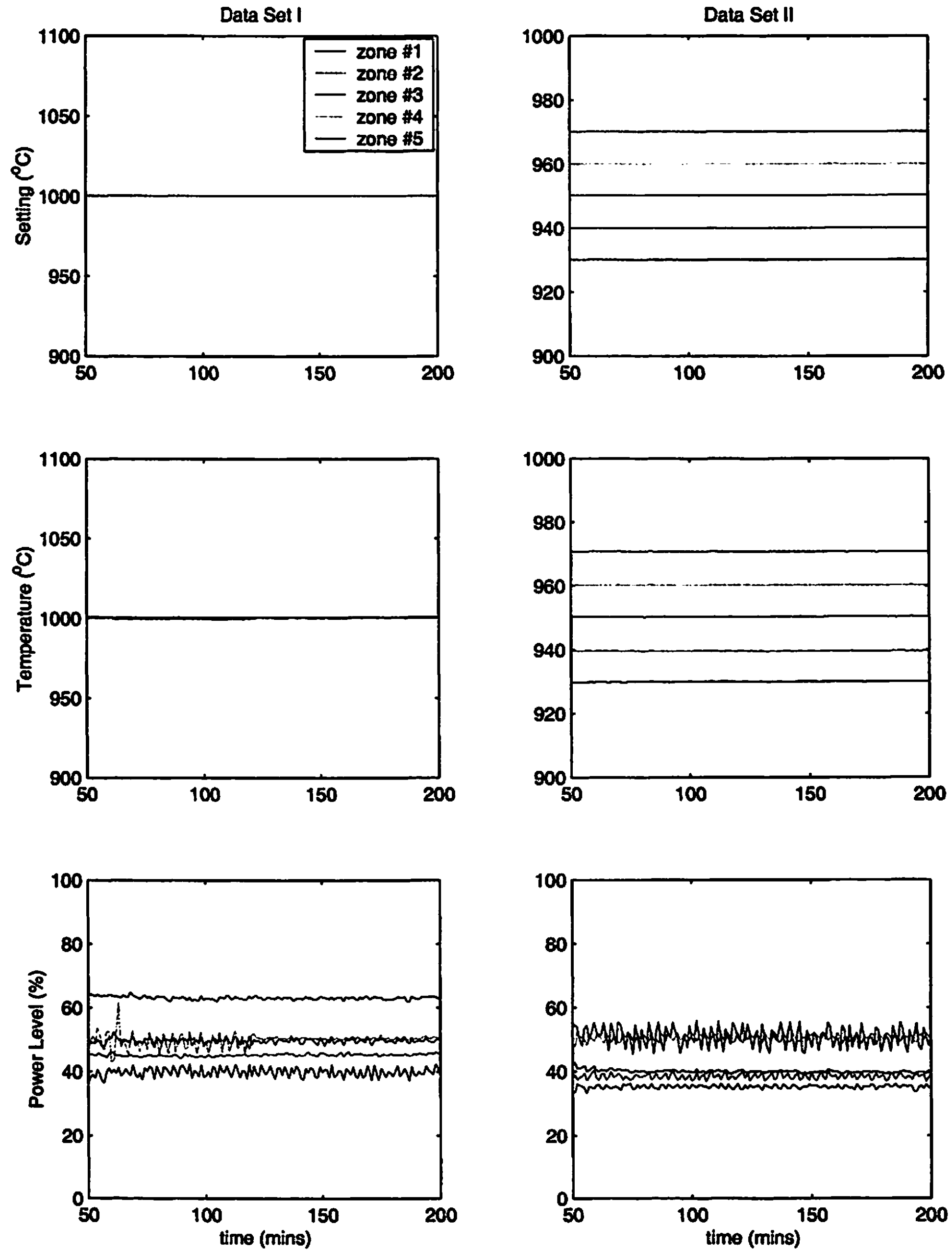

Figure 4.6: Steady state data set.

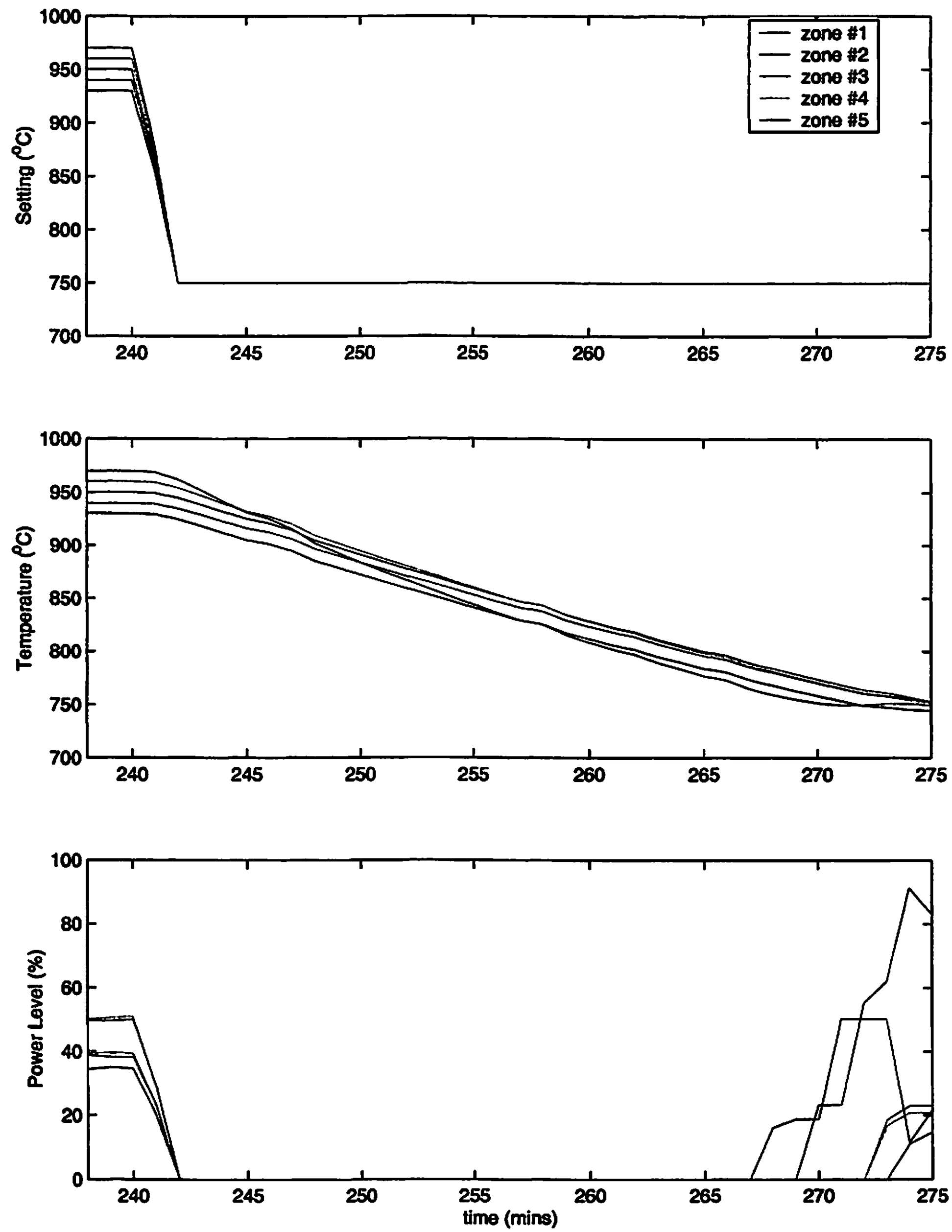

Figure 4.7: Cooling down data set.

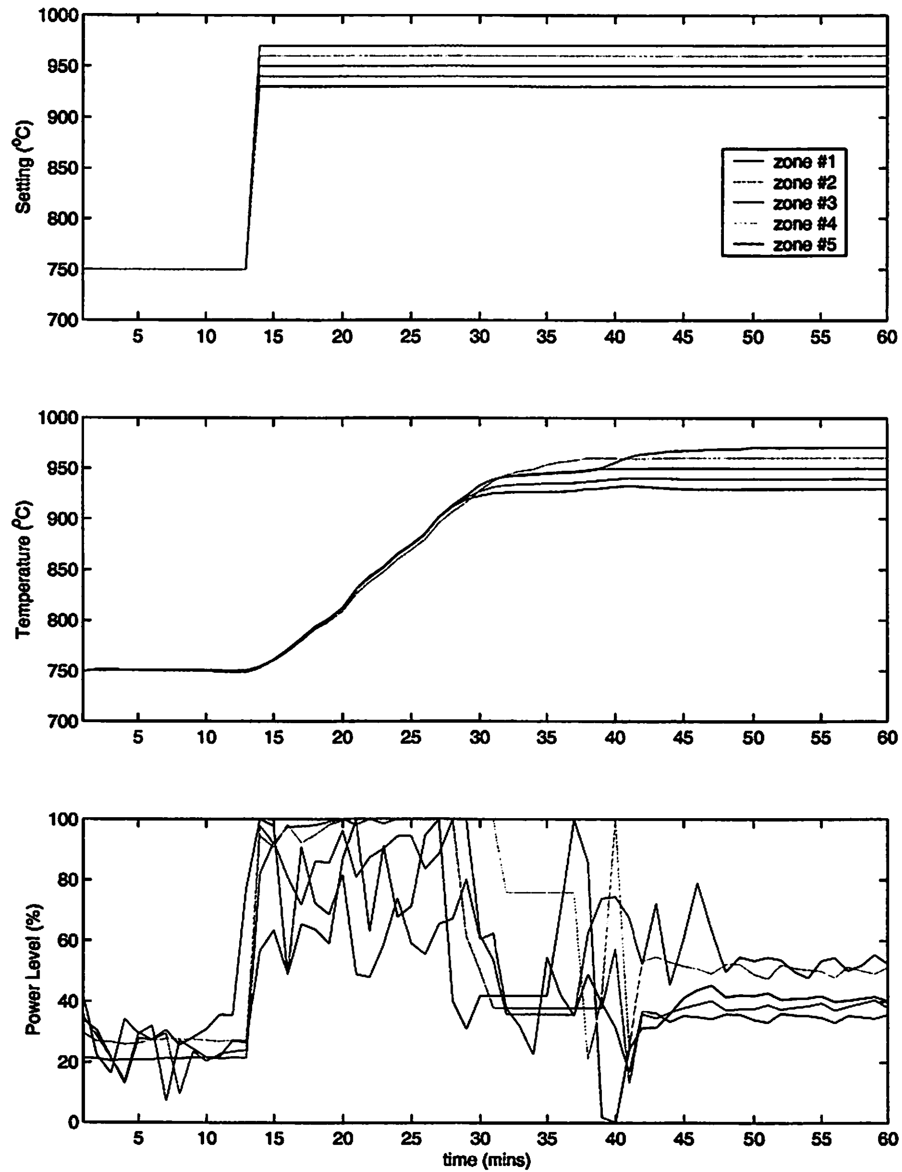

Figure 4.8: Heating up data set.

methodologies will be discussed.

4.3 Furnace Diagnostics

In this work, the goal of fault diagnostics is to use all available information (temperature settings, temperature sensor readings, power delivery readings) to answer the following questions.

1. Is there any fault in the system?
2. What type of fault is it?
3. Which sensor or which zone is involved?
4. How serious is the fault?

To answer these questions, especially the last one, we need to model and parameterize the faults. So, in the first part of this section we set up a model for the faults and in the second part we discuss fault detection.

When the additive white Gaussian noise is taken into account, we have the fault-free system model as

$$\mathbf{T}(\mathbf{k}+1) = \mathbf{A_{FF}}\mathbf{T}(\mathbf{k}) + \mathbf{B}\mathbf{P}(\mathbf{k}) + \mathbf{e}(\mathbf{k}+1) \tag{4.17}$$

where $\mathbf{A_{FF}}$, $\mathbf{B}$ are the same as $\mathbf{A}$, $\mathbf{B}$ in the last section (assuming all the parameters are estimated correctly), and $\{\mathbf{e}(\mathbf{k})\} \sim \mathcal{N}(0, \sigma^2\mathbf{I_5})$ is IIND. Here subscript FF is used for the fault-free system to distinguish it from faulty models.

There are three fault types of interest for this system. The first is the fault

caused by bad insulation of the furnace system, which affects the system by changing the values of the thermal resistance. We assume the errors are constant biases ($\boldsymbol{\theta} = [\theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5\ \theta_6\ \theta_7\ \theta_8\ \theta_9]^T$) of the inverse of the thermal resistances ($[1/R_1\ 1/R_{12}\ 1/R_2\ 1/R_{23}\ 1/R_3\ 1/R_{34}\ 1/R_4\ 1/R_{45}\ 1/R_5]$).

The insulation errors can be represented easily by introducing an additive matrix $\boldsymbol{A_\theta}$ to the system matrix $\mathbf{A_{FF}}$. The elements of $\boldsymbol{A_\theta}$ correspond to the errors of the elements of the fault-free system matrix caused by $\boldsymbol{\theta}$. The second type of fault is the temperature sensor (thermocouple) bias. A 5-d vector ($\boldsymbol{\beta} = [\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5]^T$) is used to denote the temperature biases (in ^oC) of the five zones. The third type of fault is caused by power delivery inefficiency, which means that the power system does not actually deliver the amount of power requested by the controller. A 5-d vector ($\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5]^T$) is used to represent power efficiency (in %) of the five zones. We call the parameters $\boldsymbol{A_\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ error parameters. Based on this formulation, the task of fault diagnosis is equivalent to the estimation of the values of the error parameters.

When all the three types of errors of the five zones are included, we can get the generalized system equation as follows:

$$\mathbf{T}(k+1) = (\mathbf{A_{FF}} + \boldsymbol{A_\theta})[\mathbf{T}(k) - \boldsymbol{\beta} - \mathrm{e}(k)] + \mathbf{B}diag(\boldsymbol{\alpha})\mathbf{P}(k) + \boldsymbol{\beta} + \mathrm{e}(k+1) \quad (4.18)$$

Note that the fault-free system corresponds to the case when all the elements of $\boldsymbol{A_\theta}$ are zero, $\boldsymbol{\alpha}$ is all at 100%, and $\boldsymbol{\beta}$ is all zero, $diag(\boldsymbol{\alpha})$ is the matrix whose diagonal elements are elements of $\boldsymbol{\alpha}$ and all other elements are zeros.

So far the system fault model has been established. The next step is to investigate the fault detection methodologies. One approach that can be used for

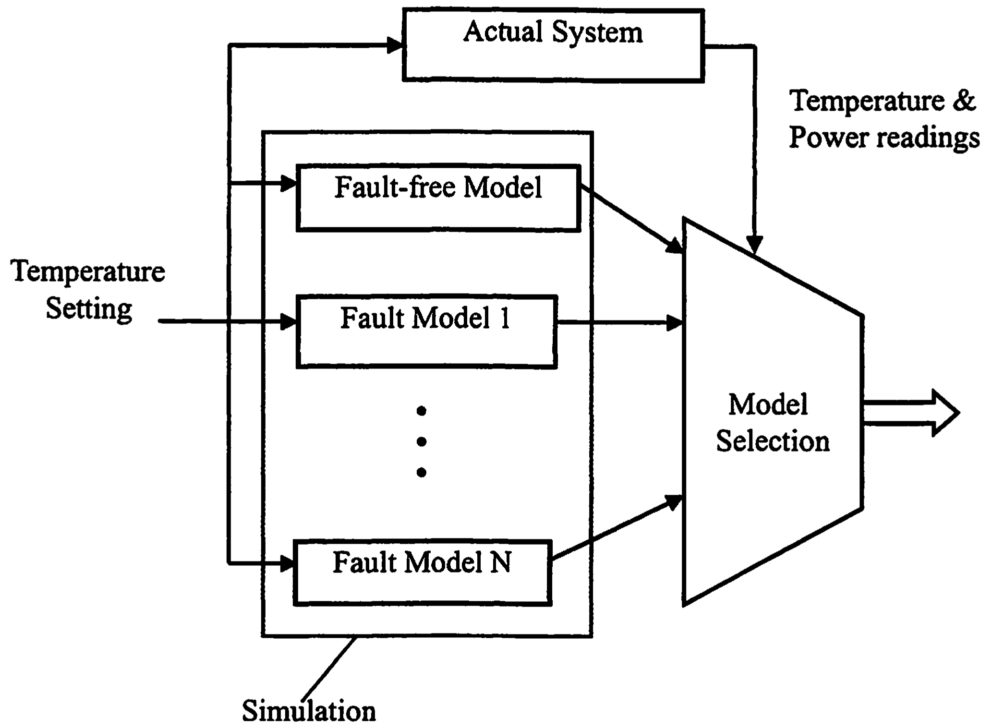

Figure 4.9: Model classification method.

this problem is that of model classification, as illustrated in figure 4.9. For each possible system fault, a corresponding model is set up and simulated. When an actual experiment is run, the temperature setting of the experiment is the input to all the simulated fault models along with the fault-free model. The output (measurement) of the actual experiment is compared to the outputs of all the simulated models. The model whose output is closest to the actual system is then selected.

If the fault-free model is chosen, we may conclude that the actual system is functioning properly. When a certain fault model is chosen, we would conclude that the actual system is suffering from the corresponding fault. To measure the closeness of the outputs, different techniques can be used, as discussed in Chapter 2.

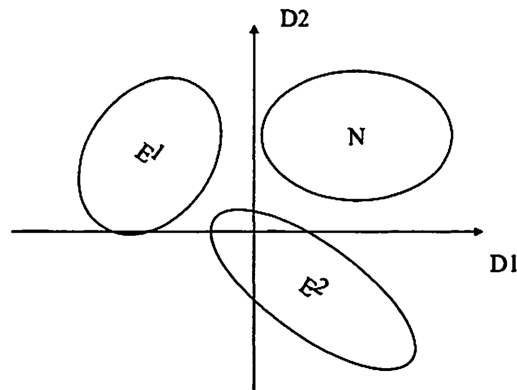

Figure 4.10: Selection criterion.

The least square matching approach [13] is commonly used. The general idea is illustrated in figure 4.10, assuming that the system output features fall in a two dimensional space, and that the fault models are separable. D_1 and D_2 could be linear combinations of important system parameters or any other significant output features.

One drawback of the direct application of this approach is that it is difficult to parameterize faults. For each value of a certain fault parameter an additional fault model needs to be set up. In our case, all fault parameters take continuous values, which makes this approach impractical. Also, this approach is not ideal for the case of multiple faults. If all different combinations of faults are considered, the total number of fault models that have to be considered is prohibitive.

Having these considerations in mind we developed a least square fault parameter estimation approach instead of using the model classification method directly. This approach has the advantages that all different types of faults and different

values of fault parameters could be represented in a general model.

$$\begin{aligned}\mathbf{T}(k+1) &= (\boldsymbol{A_{FF}}+\boldsymbol{A_\theta})\mathbf{T}(k)+\mathbf{B}diag(\boldsymbol{\alpha})\mathbf{P}(k) \\ &\quad +(\mathbf{I_5}-\mathbf{A_{FF}})\boldsymbol{\beta}+\mathbf{e}(k+1)-\mathbf{A_{FF}}\mathbf{e}(k)\end{aligned} \tag{4.19}$$

The generalized system (cf. Eq 4.18) could be approximated by Eq 4.19, which is linear in the fault parameters. Moving all the terms without fault parameters on the right-hand side of Eq 4.19 to the left and rearranging the remaining terms we get:

$$\mathbf{Y_k} = \mathbf{M_k}\boldsymbol{\Phi}+\boldsymbol{\varepsilon_k}, \tag{4.20}$$

where

$$\mathbf{Y_k} = \mathbf{T}(k+1) - \mathbf{A_{FF}}\mathbf{T}(k), \tag{4.21}$$

$$\boldsymbol{\Phi} = [\boldsymbol{\theta}^{\mathbf{T}}\ \boldsymbol{\alpha}^{\mathbf{T}}\ \boldsymbol{\beta}^{\mathbf{T}}]^{\mathbf{T}}, \tag{4.22}$$

$$\boldsymbol{\varepsilon_k} = \mathbf{e}(k+1) - \mathbf{A_{FF}}\mathbf{e}(k), \tag{4.23}$$

$$\mathbf{M_k} = [\mathbf{M}_{\boldsymbol{\theta}\mathbf{k}}\ \mathbf{M}_{\boldsymbol{\alpha}\mathbf{k}}\ \mathbf{M}_{\boldsymbol{\beta}\mathbf{k}}] \tag{4.24}$$

where

$$
\mathbf{M}_{\boldsymbol{\theta}\mathbf{k}} = \begin{bmatrix}
-\frac{T_1(k)-T_a}{C_1} & -\frac{T_1(k)-T_2(k)}{C_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{T_2(k)-T_1(k)}{C_2} & -\frac{T_2(k)-T_a}{C_2} & -\frac{T_2(k)-T_3(k)}{C_2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{T_3(k)-T_2(k)}{C_3} & -\frac{T_3(k)-T_a}{C_3} & -\frac{T_3(k)-T_4(k)}{C_3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{T_4(k)-T_3(k)}{C_4} & -\frac{T_4(k)-T_a}{C_4} & -\frac{T_4(k)-T_5(k)}{C_4} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{T_5(k)-T_4(k)}{C_5} & -\frac{T_5(k)-T_a}{C_5}
\end{bmatrix} \tag{4.25}
$$

and $\mathbf{M}_{\boldsymbol{\alpha}\ \mathrm{k}}$ is a 5×5 matrix with $\mathbf{M}_{\boldsymbol{\alpha}\mathbf{k}}(i,i) = \frac{P_i(k)}{C_i}$ for $i = 1, \ldots, 5$ and all other elements are zeros; $\mathbf{M}_{\boldsymbol{\beta}\mathbf{k}} = \mathbf{I_5} - \mathbf{A}$. $\mathbf{I_5}$ is the 5×5 identity matrix.

Eq 4.20 is equivalent to Eq 4.19. In the above equations, all the three types of faults of the five zones are taken into account. Elements of $[\mathbf{M}_{\theta\mathbf{k}}\ \mathbf{M}_{\alpha\mathbf{k}}\ \mathbf{M}_{\beta\mathbf{k}}]$ correspond to the regressors of resistance, power and temperature respectively. $\mathbf{M_k}$ can be a combination of any of subsets of them.

If we have a collection of n data points and all the data points are concatenated into a single vector, we have

$$\mathbf{Y} = \mathbf{M}\boldsymbol{\Phi} + \boldsymbol{\varepsilon} \tag{4.26}$$

where $\mathbf{Y} = [\mathbf{Y_1^T}\ \mathbf{Y_2^T}\ \cdots \mathbf{Y_n^T}]^\mathbf{T}$, $\mathbf{M} = [\mathbf{M_1^T}\ \mathbf{M_2^T}\ \cdots \mathbf{M_n^T}]^\mathbf{T}$, and $\boldsymbol{\varepsilon} = [\varepsilon_1^T\ \varepsilon_2^T\ \cdots \varepsilon_n^T]^T$. Note that $\mathbf{Y}$ is $5n$ by 1, $\mathbf{M}$ is $5n$ by 19, $\boldsymbol{\Phi}$ is 19 by 1 and $\boldsymbol{\varepsilon}$ is $5n$ by 1.

If $\boldsymbol{\varepsilon}$ is white Gaussian, the least square (LS) matching can be used to get the optimal estimation of the fault parameters:

$$\hat{\boldsymbol{\Phi}} = (\mathbf{M^T M})^{-1}\mathbf{M^T Y} \tag{4.27}$$

and the variance-covariance matrix of the estimate is

$$\mathbf{V}(\hat{\boldsymbol{\Phi}}) = (\mathbf{M^T M})^{-1}\sigma_{\boldsymbol{\varepsilon}}^2 \tag{4.28}$$

where $\sigma_{\boldsymbol{\varepsilon}}^2$ is the variance of $\boldsymbol{\varepsilon}$.

However, $\boldsymbol{\varepsilon}$ is not white (cf. Eq 4.23). Its variance-convariance matrix ($\boldsymbol{cov\varepsilon}$ can still be calculated since $\{\mathbf{e(k)}\}$ are samples of IIND random variable with known variance. We can then whiten $\boldsymbol{\varepsilon}$ using the following approach. First calculate $\boldsymbol{cov\varepsilon}$. Then find the Cholesky factorization [91] of $\boldsymbol{cov\varepsilon}$ such that:

$$\mathbf{R^T R} = \boldsymbol{cov\varepsilon} \tag{4.29}$$

$(\mathbf{R^T})^{-1}\boldsymbol{\varepsilon}$ is white since its variance-covariance matrix is an identity matrix.

By multiplying $(\mathbf{R^T})^{-1}$ at both sides of Eq 4.26, we get

$$(\mathbf{R^T})^{-1}\mathbf{Y} = (\mathbf{R^T})^{-1}\mathbf{M}\mathbf{\Phi} + (\mathbf{R^T})^{-1}\varepsilon \tag{4.30}$$

By applying the above least square solution to Equation 4.30, we obtain the new LS solution:

$$\begin{aligned} \hat{\mathbf{\Phi}} &= [((\mathbf{R^T})^{-1}\mathbf{M})^\mathbf{T}((\mathbf{R^T})^{-1}\mathbf{M})]^{-1}((\mathbf{R^T})^{-1}\mathbf{M})^\mathbf{T}((\mathbf{R^T})^{-1}\mathbf{Y}) \\ &= (\mathbf{M^T}\mathbf{R_{-1}}(\mathbf{R^{-1}})^\mathbf{T}\mathbf{M})^{-1}\mathbf{M^T}\mathbf{R^{-1}}(\mathbf{R^{-1}})^\mathbf{T}\mathbf{Y} \end{aligned} \tag{4.31}$$

$$\mathbf{V}(\hat{\mathbf{\Phi}}) = (\mathbf{M^T}\mathbf{R^{-1}}(\mathbf{R^{-1}})^\mathbf{T}\mathbf{M})^{-1}\sigma^2 \tag{4.32}$$

This concludes the basic idea of the LS approach. Theoretically, as long as our model is accurate and the data used for diagnosis are rich enough, all those fault parameters could be estimated accurately. But in practice we are dealing with a real physical system. Our linear model is not exact, since there are some minor non-linearities due to some known and unknown reasons. The fact that we cannot collect arbitrary large data sets also makes the diagnosis more difficult.

In the next section, we discuss the experiments, the simulation and the strategies of choosing different data sets for different diagnostic purposes, as well as a sensor fusion approach to enhance diagnostic reliability.

4.4 Results Analysis

4.4.1 Modeling Results

In previous sections we set up the structure of the linear model and discussed how to extract the system parameters using different types of experimental data.

Some typical experimental data sets were shown in figure 4.4. These normally consist of three parts: heating up, steady state and cooling down. As we mentioned earlier, steady state data are used to solve for thermal resistance parameters $\{R_1, R_{12}, R_2, R_{23}, R_3, R_{34}, R_4, R_{45}, R_5\}$, cooling down data are used to estimate thermal capacity parameters $\{C_1, C_2, C_3, C_4, C_5\}$, and heating up data are used for adjusting controller parameters. Figure 4.11 and figure 4.12 show the model estimation results. After fixing the model structure, an experiment was run on the tube and the furnace model was simulated with the same recipe. Comparing the results, it can be seen that the estimated model is very accurate for steady state and cooling down periods. However, for the heating up period, there are obvious differences between the actual furnace and its models. In this work, we decide not to put too much effort on decoding the vendor's controller design. The only knowledge that has been used is that it is a modified PID type controller. There is actually a tradeoff between model complexity and accuracy. A more complex controller (with more parameters to adjust) may be adopted, but it would require significantly more good experimental data to estimate the model. Therefore we suggest using only the steady state and cooling down data to perform diagnostics.

4.4.2 Diagnostic Results

Table 4.2 summarizes the diagnostic results. In this table, the results of five cases of fault combinations are listed as examples to show the diagnostic performance. The second column lists the number of fault parameters being considered (any combinations of those parameters are possible) for each case. The third column gives the specific fault parameters to be estimated for each case. The rest of the table shows the types of the data sets used for estimation and performance of

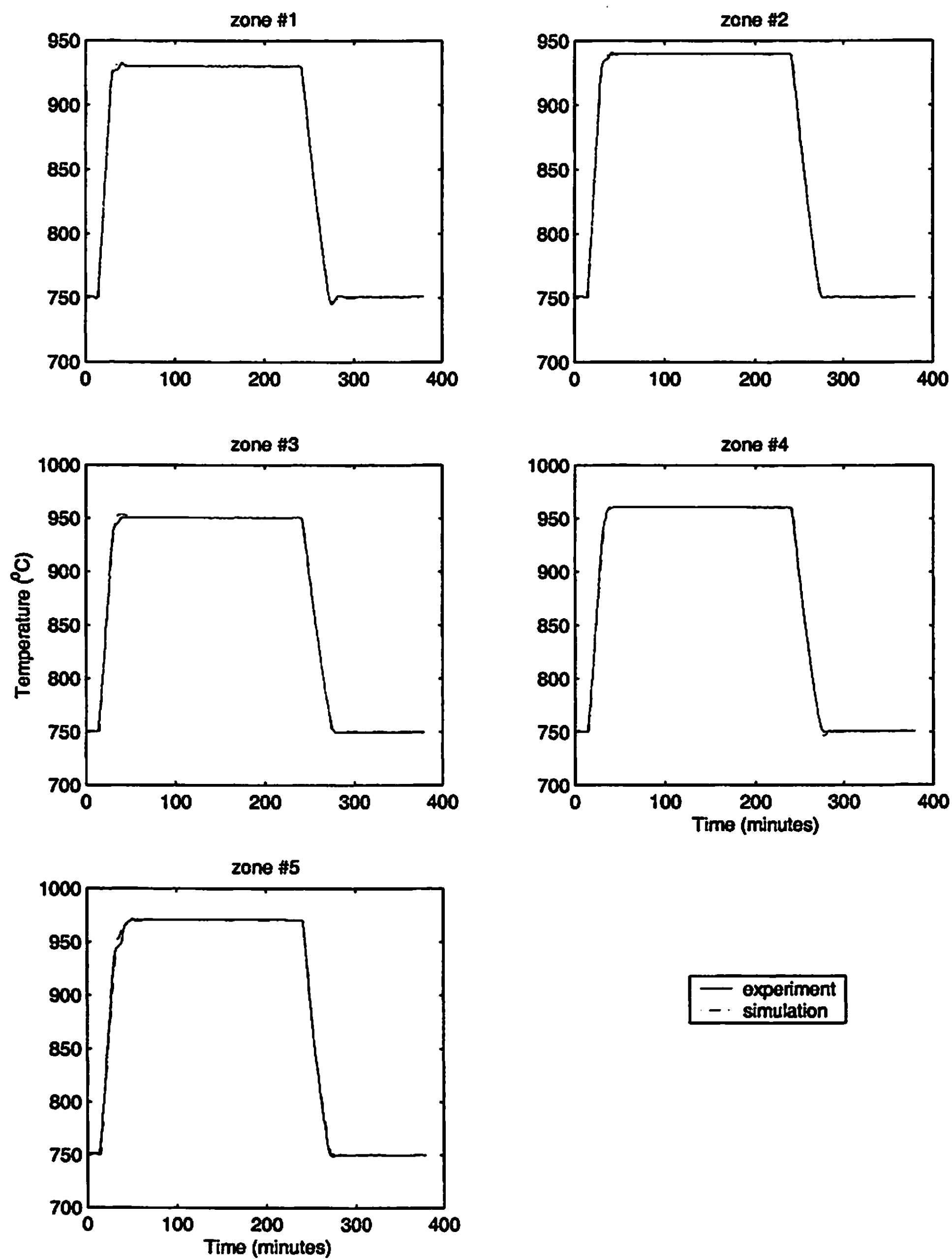

Figure 4.11: Comparison of experimental data and linear model simulation results (temperature profile).

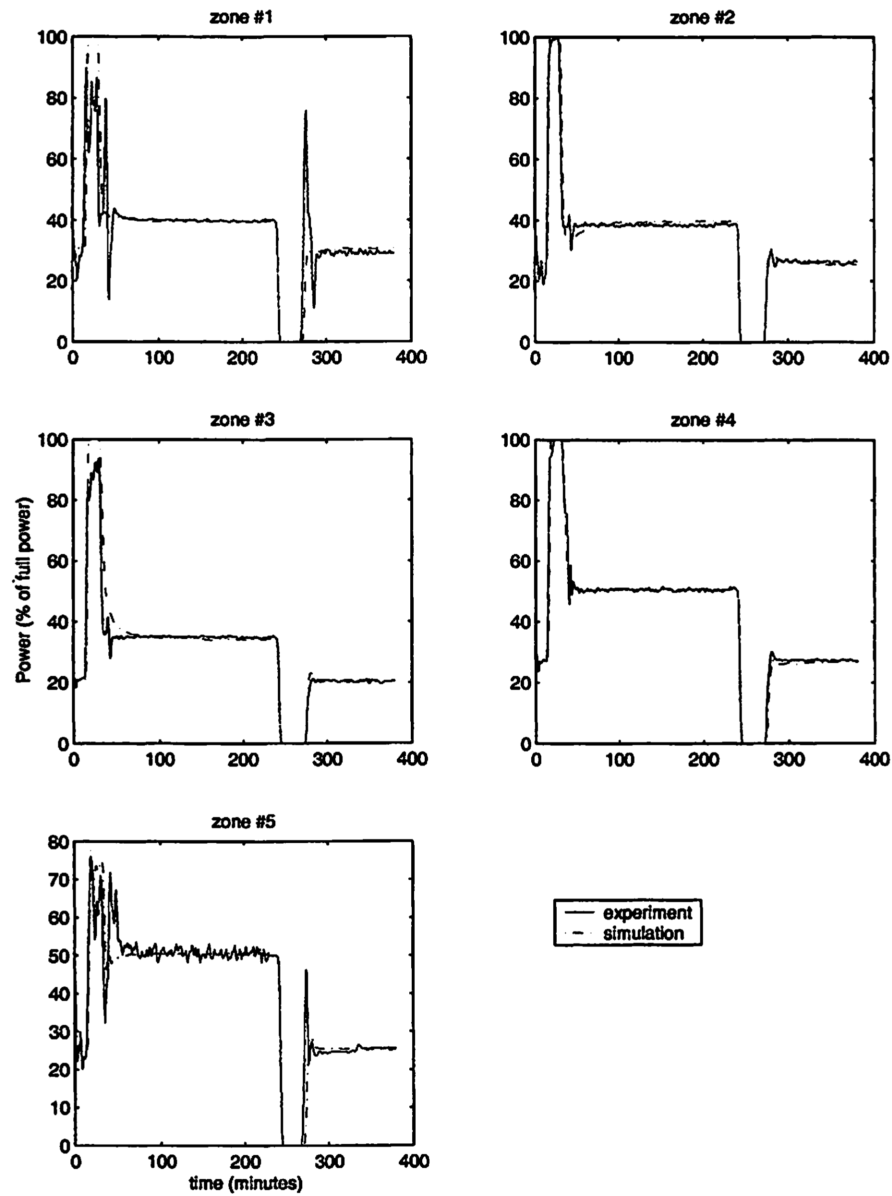

Figure 4.12: Comparison of experimental data and linear model simulation results (power profile).

Case #		Number of Faults (Parameters to estimate)	Faults Parameters to estimate	Data Used & LS Performance		
				Steady state data	Cooling down data	Steady state+ Cooling down data
1		19	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5,$ $\theta_6, \theta_7, \theta_8, \theta_9$ $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5,$ $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$	N/F (5)	N/F (11)	N/F (11)
2		10	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5,$ $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$	N/F (5)	N/F (7)	N/F (7)
3	A	5	$\theta_1, \theta_3, \theta_5, \theta_7, \theta_9$	Good	Good if $\theta_2, \theta_4, \theta_6, \theta_8$ are known	Good
	B	4	$\theta_2, \theta_4, \theta_6, \theta_8$	N/A	Good	Good
4		5	$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$	Good	N/A	Good
5		5	$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$	Good	Good	Good

Table 4.2: Performance of Least Square Method. N/A denotes "Not Applicable" and N/F denotes "Not Feasible". $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9$: nine thermal resistance faults, $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$: five power inefficiency faults, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$: five temperature sensor bias faults. Number in parenthesis indicates number of data dimensions, not shown when it equals or exceeds the number of parameters.

the LS estimation.

"Not feasible (N/F)" means the least square problem is not solvable due to the limited degrees of freedom in the available data. The number in the parentheses gives the number of dimensions of the used data (number of non-zero eigenvalues of the data matrix). The number of data dimension can help us determine whether the least square problem is solvable for different cases. If the available number of data dimensions is less than the number of the fault parameters, the problem cannot be solved. "Good" means the data are rich enough to solve for the fault parameters. Note that in case #3, both steady state and cooling down data are needed for the estimation of the nine fault parameters. First, the steady state data is used for

estimating fault parameters corresponding to the five self-resistances (resistances linking each zone to ground in figure 4.2). Then the cooling down data is used for estimating fault parameters corresponding to the four inter-resistances (resistances linking the five zones in figure 4.2). Therefore, case #3A and case #3B need to be done in sequence. When the system is at steady state, since normally all runs need uniform temperature inside the tube, the influence of the fault of inter-resistances is insignificant (these can be viewed as Miller resistances [87], that only play a role if the system gain is other than unity). So we can fix the value of inter-resistances at nominal values and estimate the faults of the five self-resistances. Then we can fix the values of self-resistances and use the cooling down data to estimate faults of inter-resistances. This is possible because the cooling down process would introduce non-uniformity to the temperature profile and make the role of inter-resistances significant. For case #4, cooling down data cannot be used to detect power inefficiency since power is shut down. During actual experiments, we observed voltage fluctuations (up to about 5%) on the power line. This could be a source of the power errors. A power monitor was used to track the fluctuation and compensate the simulation for the power delivery fluctuations.

For case #5, both types of the data can be used for estimating temperature sensor faults. Use of the cooling down data only in this case is not recommended because the cooling down process is brief and its data set is limited.

The precision of the fault parameter estimation can be evaluated by the following procedure (cf. figure 4.13). We first input the same temperature settings to both the actual system and the simulated system. We then calculate modeling error (including noise) by taking their difference. When simulating the detection

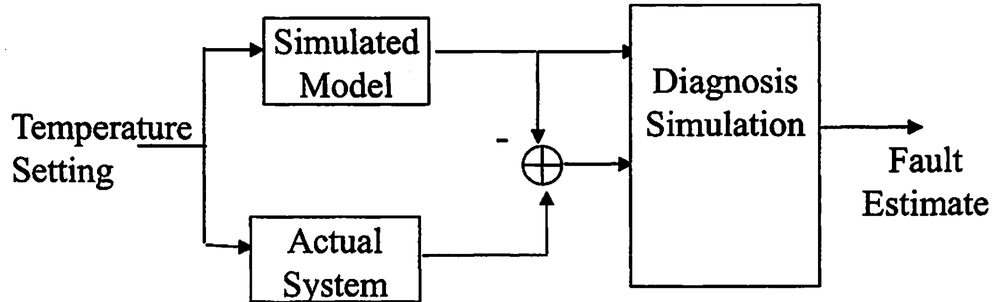

Figure 4.13: Sensitivity test of diagnostic algorithm.

algorithm, we input the simulated fault-free system output added by the modeling error and see the fault parameter estimation. The value (mean and variance) of the estimates would determine the precision of the detection. For example, when we do this for case #5, we obtained means of the five estimates around zero and their standard deviation less than 1.15^oC. The 95% confidence interval is $(-2.3^oC, 2.3^oC)$. This means that if the actual temperature sensor bias is within this interval, we would not be able to distinguish it from the modeling error.

4.4.3 Data Fusion Applied in Diagnostics

While precision is important for fault detection, reliability is also crucial. We can never rely on the results from one data set due to uncertainty in real operating conditions. To improve the reliability of the diagnosis, we integrate information from multiple data sets based on the concept of sensor fusion to make the final decision about the fault estimation. Repetitive or complementary experiments may be done to get multiple data sets for the purpose of fault detection reliability. Each of them will provide a set of estimated fault parameter values. These data sets can be treated as different "sensors" measuring fault parameters. Sensor fusion

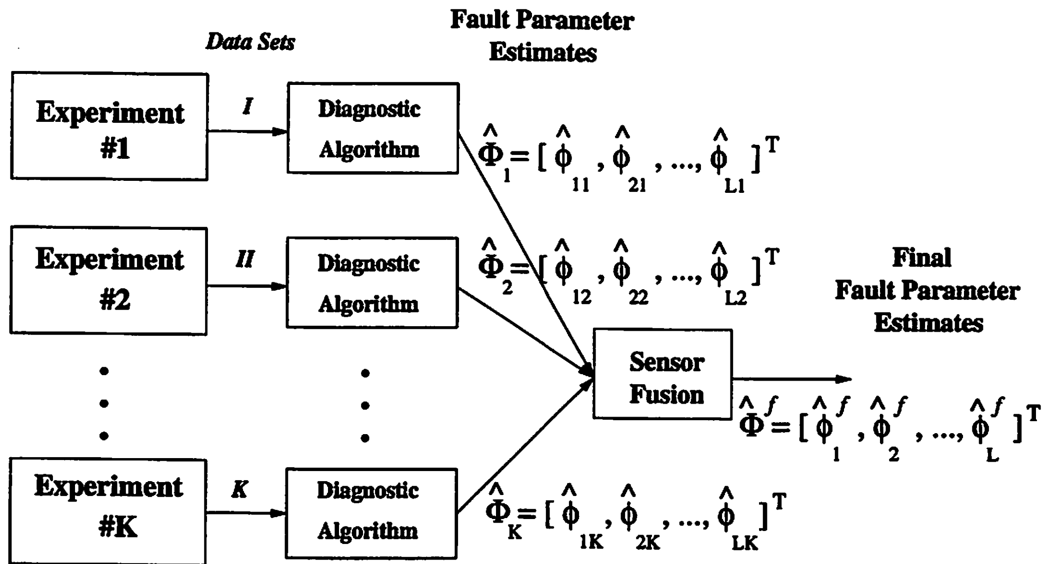

Figure 4.14: Data fusion to enhance diagnostic reliability.

techniques are then utilized to integrate the information into the best estimation of the concerned fault parameters. The idea is shown in figure 4.14.

There are a variety of ways to perform the fusion [2, 44, 64]. The method we are using here is based on assigning a confidence value to each "sensor" and calculate their weighted average using the confidence values as weights. The confidence value depends on the variance of each estimated fault parameter, which is a by-product of the LS estimation algorithm. Estimates with smaller variances would deserve larger confidence values. In this way, if one set of the data were corrupted, a very small or zero confidence value would be assigned to it. It will then not adversely affect the final estimation. In summary, sensor fusion is performed as in Eq 4.33.

$$\hat{\phi}_i^f = \sum_{j=1}^{K} w_{ij}\hat{\phi}_{ij} \tag{4.33}$$

where $i = 1, 2, \ldots, L$, and L is the total number of parameters to be estimated, $(\hat{\Phi}_j = [\hat{\phi}_{1j}, \hat{\phi}_{2j}, \ldots, \hat{\phi}_{Lj}]^T$, $\hat{\Phi}_j$ is the estimation of parameters based on the jth

set of data). K is the number of data sets, $\hat{\phi}_{ij}$ is the LS estimation of the ith parameter based on the jth set of data, w_{ij} is the weight for $\hat{\phi}_{ij}$ and is a function of estimated variance of $\hat{\phi}_{ij}$ with $\sum_{j=1}^{K} w_{ij} = 1$, and $\hat{\phi}_i^f$ is the fused estimation of the ith parameter. Different methods can be used to calculate the confidence values [44, 67]. Various distance measures have been suggested to measure the closeness between a sampled probability density function and the reference density function [114]. There reference can be obtained by baselining good historical data or predicting based on stochastic system model [44]. One example for calculation of w_{ij} is to use the inverse of the estimated variance of $\hat{\phi}_{ij}$ and then normalize it so that w_{ij} sum (over j) to one.

One thing that should be pointed out is that, when we implement sensor fusion as we described in this section on a specific furnace system, it cannot be real-time, since we need multiple data sets to perform fusion, which cannot be collected at the same time. But slow dynamic faults can be diagnosed more reliably. Sensor fusion does not necessarily increase the precision of the estimation, but it does improve the reliability of the diagnostics effectively.

4.5 Summary

This chapter demonstrates diagnostic strategies for systems with physical models, using a thermal dynamic system as an example. We show some experimental and simulation results on a real furnace system in the Berkeley Micorfabrication Laboratory. A Least Square (LS) estimation approach is used to extract model parameters and to estimate fault parameters, based on a model classification diagnostic framework.

If we are able to accurately estimate the model parameters of a fault-free system, all different fault combinations can be diagnosed successfully (e.g., a 1^oC temperature sensor bias can be detected) through simulation. However, in reality, system modeling is often inaccurate and the types of experimental data are limited. Consequently, diagnostic performance is often sub-optimal. Single failures can be evaluated if the fault type is known, and the accuracy of the evaluation depends on the accuracy of the system parameter estimation. Some important failure combinations can be effectively diagnosed by our proposed approach using steady-state data and/or cooling-down data sets, while the diagnosis of other failure combinations remains difficult given the currently available data sets. It is likely that more accurate system modeling through the studying of the system non-linearity may provide better diagnostic performance. Finally, sensor fusion techniques may also be used to enhance diagnostic reliability.

Chapter 5

Lithography Process Diagnostics

As one of the most crucial processes in semiconductor manufacturing, photolithography is used at almost every stage of production. Nearly one third of the total cost of semiconductor manufacturing is spent on lithography processing. Holding the chip size constant, a decrease in feature size by a factor of two would result in an increase of the number of devices by a factor of four as well as a significant performance increase. Naturally, manufacturers seek to reduce feature size as much as possible. Only a couple of years ago, "submicron" was still an industry buzzword, but now it has been replaced by "sub 0.1-micron". In the meantime, increasing wafer size has been another major course of productivity improvement. Currently, the industry is quickly transferring from 200mm technology to 300mm technology. The ever decreasing device sizes and increasing wafer sizes result in many more devices on each wafer than previously possible. Consequently, the wafer cost has been rising. This economic risk has put great pressure on lithography engineers and researchers to enhance small feature reproducibility and reduce defects. Some improvements are evident: process equipment precision is higher, fabs are cleaner, and the accumulated process experience has advanced. However,

the problem of process variation remains. A process feature produced in the morning may become quite different in the afternoon even without anyone changing the experiment setup. The environment (e.g., the temperature and humidity) could have changed, the tool condition could have become different, and the production settings could have drifted. Many things are subject to change due to the dynamic nature of the world that we live in.

Scientists and engineers have worked very hard in trying to understand the physics of every part of a photolithography process [58, 107]. Many advanced metrology tools have been developed to collect process information timely and accurately [60, 94]. All of these efforts aim to enable process control. The purpose of process control is to reduce process variation by monitoring critical process parameters. Traditional Statistical Process Control (SPC) has been successfully implemented in the semiconductor industry. Typically, SPC is used for monitoring process parameter trends, setting control limits, and generating alarms when appropriate. Many fabs are shutting down equipment according to SPC results. Current SPC methods have the advantage of implementation simplicity but, since there is no feedback loop, they cannot make automatic adjustment to a process. Much more is expected from process control and thus, Advanced Process Control (APC) methods are receiving increasing attention [32], but the scope is not yet well defined.

In this chapter, we propose a novel APC approach for diagnosing lithography parameter drifts based on empirical modeling and statistical analysis of Critical Dimension (CD) profile data.

5.1 Lithography Process Diagnostics

5.1.1 Process Overview

Lithography is a process of transferring the desired pattern from the mask to the surface of a wafer [57]. The process determines the line-width on the wafer and therefore the overall size of the chips. In order to enhance product yield and device speed, there is a continuous effort to scale down the pattern size. As the image size becomes smaller, linewidth variations plot a critical role.

Photolithography consists of a number of different physical and chemical processes and involves many equipment and materials. In such a complex sequence, there are many sources of variation. Recipe drifts, equipment instability, and environmental changes are all important sources for Critical Dimension (CD) profile variations. In this study, we focus on the detection of recipe parameter drifts.

Many studies on understanding the physics of photolithography have been presented previously [78, 100, 107, 111]. In the current work, we propose a novel approach for modeling and diagnosing of the photolithography processes, based on statistical analysis of relationships between relevant input and output parameters.

A basic Deep-Ultra-Violet (DUV) lithography process consists of thin film preparation, spin-coat of resist, pre-bake (soft bake), exposure, post-exposure bake (PEB), and development as shown in figure 5.1. In this work, we are concerned with controlling variations in the Critical Dimension (CD) profiles. A CD profile is a two-dimensional representation of the CD, which we will describe in more detail in section 5.1.3. Factors (and their combinations) that may cause variations in CD

profiles include the properties of the light source (such as the exposure dose, focus position, and partial coherence of light), photoresist parameters (such as Dill's A, B, C parameters and other resist chemical parameters [107]), resist thickness, and PEB process parameters (such as the PEB time and temperature). In real manufacturing lines, these input parameters may drift over time, which, in turn, may result in CD profile variations. Some parameters can be measured and adjusted easily, but many cannot. Direct measurement of parameters such as exposure dose and focus position can be very costly and even impractical. Thus, the detection of input parameter drifts based on CD profile measurements is of great potential interest.

Figure 5.2 shows a typical layer structure of the silicon wafer that we use as example. The Anti-Reflection Coating (ARC) layer is utilized to reduce reflectivity in order to mitigate the effect of standing waves in the side wall of the CD profile. In a lithography process, the wavelength of light, in conjunction with the numerical aperture of the lens, determines the image resolution. Deep Ultraviolet (DUV) light is widely used nowadays to achieve small feature sizes.

In the following sub-sections, we will describe some input and output parameters that have been considered in this work. We will also describe Critical Dimension metrology and the diagnostics structure. Although our modeling is purely data-driven, it is essential to understand intuitively how each input parameter affects the CD profile. Such intuition will help us design efficient experiments and decide what types of data are appropriate for modeling.

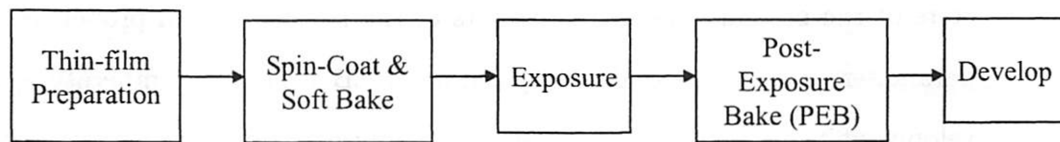

Figure 5.1: Flow chart of the lithography process.

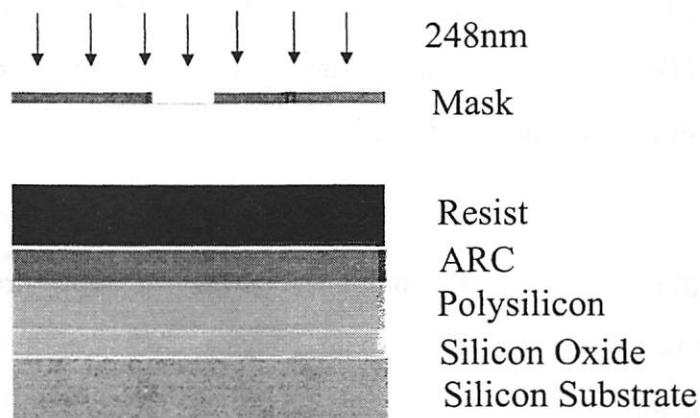

Figure 5.2: Layer structure that we use in this work.

5.1.2 Input and Output Parameters

Input parameters of a lithography process include at least the following: parameters of the optical system; parameters of the resist material properties; process parameters such as pre-bake temperature, PEB time and temperature, and development time; material thickness; and many other parameters associated with various subprocesses. We therefore have a complex system where many parameters can individually or in combination contribute to linewidth variations.

For most modeling and diagnosing efforts, computation complexity grows exponentially with the number of input parameters. To make the problem more tractable, and without loss of generality, we have chosen to consider only five input parameters to demonstrate the effectiveness of our approach to lithography process diagnostics. These five parameters are: 1. Exposure dose (exp), 2. Focus position (foc), 3. Partial coherence (coh), 4. PEB time (t_{PEB}), and 5. PEB temperature (T_{PEB}). All other input parameters are assumed fixed and known, but a similar approach may be adopted when those parameters are also considered as variables.

The optical system is a key component of the lithography process. The exposure system projects an image, transferred through a mask with the desired pattern onto the silicon wafer that is being processed. Many levels of such masking are often required. For example, an advanced memory chip may undergo more than twenty masking levels. A thin layer of resist material transfers the image in such a way that the layers underneath the resist can form a permanent device after several subprocesses, such as baking, development, and etching. The accuracy of this image transformation determines the feature size that the process can achieve. It

is thus essential to have very high stability in the optical system. Drifts in a number of parameters associated with optics and lights can bring profile defects. Exposure energy dose and focus position are the primary factors that have to be considered. The exposure system is a highly complex optical system consisting of a projection lens, the illumination subsystem, and the wafer positioning subsystem. Exposure dose is the amount of light incident on wafers coated with resist on stepper, and it is set according to the sensitivity of photoresist and the time allowed for exposure. Variations in exposure dose can drastically affect the shape of features projected on a wafer. Unfortunately, direct measurement of exposure dose remains very difficult and costly. Focal position is another key parameter that has to be kept constant. Images become blurred should focus get out of position. When high accuracy is desired, it is often very difficult to directly measure the focus position.

Partial coherence (*coh*) is a figure that describes the range of angles used for illumination. Although partial coherence is inherently stable, it is still useful to be included in the modeling, since different coherence values determine a wide range of optical effects. In our modeling and diagnostic activities, partial coherence is assumed to be known.

Post-exposure bake (PEB) is an important thermal subprocess in lithography. It "activates" the photoactive compounds, it reduces the standing wave effect and thus increases linewidth control and resolution. Drifts in PEB temperature and duration may significantly affect the shape of the resist CD profile.

We have also considered a few other parameters such as resist thickness and the thickness of the ARC layer. However, empirical results suggest that CD profiles are not very sensitive to variations in these parameters at least for the parameter

ranges that we are concerned with. Hence, we focus on the aforementioned five input parameters: Exposure Dose (exp), Focus (foc), Partial Coherence (coh), PEB Time (t_{PEB}) and PEB Temperature (T_{PEB}).

Critical Dimension (CD) profile is a two-dimensional representation of the cross section of a resist profile after lithography process(cf. figure 5.3). CD profile is the most important measurement that lithographers rely upon. Although an entire two-dimensional CD profile may be used for the purpose of process control, it is often more convenient to use only certain features extracted from the profile. In this study, we use features such as the 10%, 50% and 90% CD measurements and the Side Wall Angle (SWA), as system outputs. The Side Wall Angle can actually be calculated from the 10% and 50% CDs. But we choose to use this redundant information because they may have different noise characteristics and looking at both of them would provide us more stable information.

Resist thickness loss could have also been considered as an output parameter. However, it is very difficult to measure resist thickness loss with high precision (on the order of several nanometers in our simulations). Hence, since resist thickness loss does not seem to be very sensitive to any of the five input parameters that we have selected, we decided not to include it as an output parameter.

5.1.3 Critical Dimension Metrology

The rapidly decreasing linewidth has resulted in a corresponding decrease in linewidth uncertainty tolerance, which, in turn, places a greater demand on linewidth measurement tools. While the current generation device gate length is on the order of 100 nanometers, only a few nanometers uncertainty is allowed for

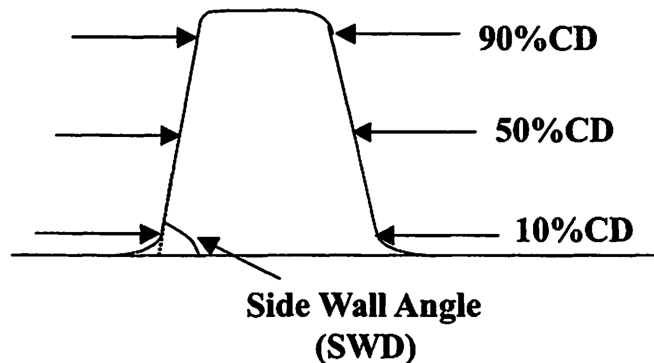

Figure 5.3: CD profile features.

a CD metrology tool. There has been a number of different CD metrology tools applied to modern photolithography processes [100], e.g., optical linewidth metrology, electrical linewidth metrology [115], CD SEM (CD Scanning Electron Microscopy) [82], and CD AFM (Atomic Force Microscopy) [79]. We are not directly concerned with the physical principles of these tools since no physical experiments are involved here. However, to understand their limitations, we do need to get some sense of the accuracy and speed that current metrology tools can provide.

Optical tools had been popular historically due to their availability and low costs. However, the accuracy of conventional (far-field) optical techniques is limited by the wavelength of light, which makes them not applicable when the linewidth dimension goes under 1.0 μm. As a consequence, lithographers have been looking for alternative techniques such optical scatterometry [83]. For example, the Specular Spectroscopic Scatterometry (SSS), investigated by Xinhui Niu in his PhD research [94, 92], is a promising technique that provides reliable cross-sectional CD profile information for current generation technology.

Electrical linewidth metrology can provide very precise measurements of the

Metrology Tools	Accuracy	Cost	Speed	Comment
Traditional Optical	unsuitable for submicron	low	high	does not meet current resolution requirement
SSS	high	low	high	under development
Electrical	very high	low	high	only for some conductive films
CD SEM	high	high	slow	has to cut wafer
CD AFM	high	high	very slow	speed is a major problem

Table 5.1: Comparison of CD metrology tools.

average linewidth of a conductive film through electrical tests. It has been an invaluable tool to many semiconductor applications. However, it can only be used on conductive film and it does not provide cross-sectional CD profile information. Also it can only be used after all processes are finished.

CD SEM technology is currently the most popular metrology tool for CD profile measurements. Two types of CD SEM are available, one for cross-sectional profile and the other for top-down measurements. High resolution may be achieved by CD SEM even though it is often time-consuming and expensive. The cross-sectional CD SEM requires physically cutting the wafer, which is obviously undesired.

CD AFM technology has been receiving increasing attention in the lithography community. CD AFM can also provide very accurate cross-sectional profile information without cutting the wafer, but currently it is relatively slower than CD SEM.

Table 5.1 summarizes the comparison among the major CD metrology tools.

5.1.4 Model-based Diagnostics

Traditionally, test wafers have often been used to adjust process parameters in fabs while there are many problems associated with them. Due to the extremely small device size, ever increasing wafer size and high complexity of modern processes, test wafers can be very expensive. This means that engineers have to rely upon the sparse information obtained from a small number of test wafers. Naturally, the laws of statistics leave us in doubt about the reliability of control decisions made under those conditions.

Recently, people have started looking at process parameter adjustment methods that adjust parameters automatically by monitoring process parameter drifts [12, 74]. In this study, we perform diagnostics of the lithography processes by monitoring the drifts in the input parameters through observations made on CD profile measurements. This is not an easy task due to the non-linearity and high dimensionality of lithography processes. Figure 5.4 shows the basic idea of automatic feedback parameter adjustment. Similar efforts have been made in the area of detecting some lithography process parameter drifts using empirical methods [16], but they do not provide a general diagnostics framework. They only focus on a couple of important parameters, for example, exposure dose and focus. The approach proposed in this work is not limited to specific parameters.

Why is it important to diagnose each of the input parameters? To answer this question, let us first take a look at the alternatives. Recall that the ultimate goal of diagnostics is to ensure the system's ability to reproduce the desired features over a sustained period. When one or more input parameters start drifting, the output feature is likely to deviate from the desired values. In many cases, it may

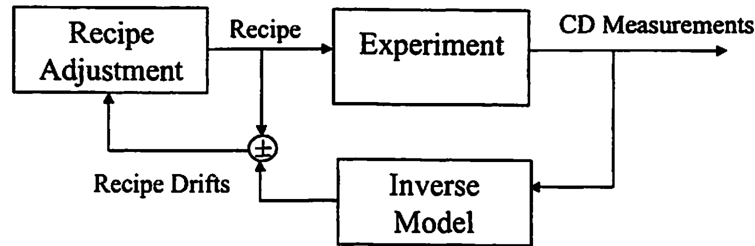

Figure 5.4: Automatic feedback parameter adjustment.

be possible to adjust some input parameters (not necessarily the drifting ones) to bring the system output closer to the desired features, where the parameters and amount to adjust are chosen through experience or trial-and-error. This is in fact the most common engineering practice in the semiconductor industry today, although there are several fundamental flaws with this approach. First of all, adjusting one or several input parameters that are not exactly the drifting ones can compensate for output deviation only to a limited degree since the highly complex parameter space makes it very difficult to find alternative mappings from the input to the desired output features. Secondly, since each system parameter has its own range, adjusting certain parameters to their extreme values for compensating other parameter drifts may bring adverse and often catastrophic effect to the overall system state. In such cases, even if the system were able to compensate for some parameter drifts and produce desired output features in a short term, performance often cannot be sustained and more serious problems may arise due to the undesired system state. For example, if only the focus position is adjusted while both exposure dose and PEB temperature are drifting, we may easily place focus position over its physical limit and still not fully obtain the desired features. Therefore, a truly principled solution must be one that monitors every parameter and has it under control whenever a drift from its specified setting occurs. This is

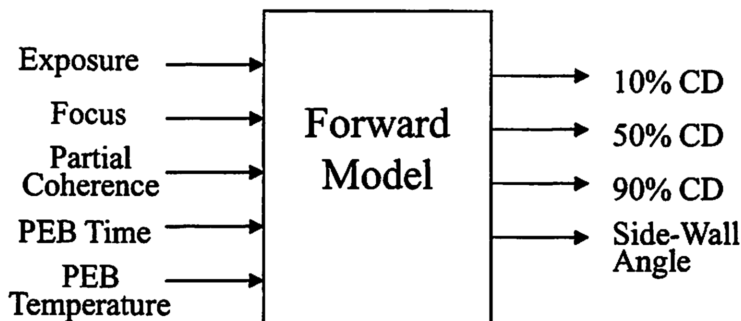

Figure 5.5: Forward Model.

precisely what we hope to achieve with statistical diagnostics. Of course, for each particular statistical model we develop, we have to make the assumption that certain states of the system, such as the recipe, equipment and environment, remain unchanged. Although it is possible to also include those intrinsic system states in a statistical model, it is beyond the scope of our current study.

In this work, we propose two different approaches for diagnosing input parameter drifts. We define *forward model* as a model from input recipe parameters to output CD measurements (cf. figure 5.5) and *inverse model* as a model from CD measurements to recipe parameters (cf. figure 5.6). Each of the two basic diagnostic approaches that we propose in this study employs a different type of model. We call the diagnostics based on an explicit inverse model (a function that maps CD measurements to input recipe parameters) the *inverse model approach*. In cases where explicit inverse models do not exist, we rely on forward-model simulation to build a library of input-output data pairs. Estimation of input parameters can then be accomplished by searching in the library for the matching output. We call this methodology the *library searching approach* [121].

Regardless of the diagnostic approach we take, the success of the overall frame-

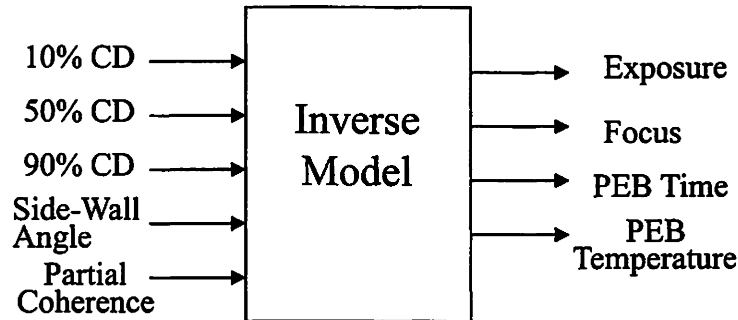

Figure 5.6: Inverse Model.

work largely depends on appropriate system modeling. As we pointed out at several occasions, lithography is an extremely complex process with many interacting subcomponents.

The field of optical lithography modeling was pioneered by Rick Dill of IBM in the early 1970s. In a series of papers around 1975, now known as the "Dill Papers" [27, 28, 29], Dill and colleagues presented a simple model for image formation with incoherent illumination (the first order kinetic "Dill model" of exposure) and an empirical model for development with a photoresist profile calculation program. In the late 1970s, Andy Neureuther and Bill Oldham advanced state-of-the-art lithography modeling by including partial coherence and more sophisticated dissolution calculation algorithms in their SAMPLE package [95], the first widely accessible lithography modeling program. Since 1985, first introduced by Chris Mack of FINLE Technologies [78], PROLITH was the first lithography simulation package running on personal computers and its subsequent commercial versions, PROLITH/2 and PROLITH/3D, became widely used in the industry. The PROLITH family not only includes comprehensive physical modeling of lithography processes based on the most current understanding but also provides a very user-

friendly graphical interface.

Although these packages are able to accurately perform fairly large lithography process simulation, the prohibitively expensive computational complexity prevents their application in many important tasks, such as real-time fault detection. This has been one of the major motivations for us to develop statistical models for process diagnostics.

Figure 5.7 summarizes the entire diagnostic framework in a cascade fashion. Empirical data collection is essential for any modeling effort. A large amount of training data are required for accurate process modeling. However, physical experiments are often too time-consuming and not cost-effective for data collection purposes. Fortunately, physical-model based mathematical simulators are now able to simulate experiments at a small fraction of the real cost. Nevertheless, in practice, physical experiments are still necessary for tuning certain parameters in simulators (e.g., the resist chemical parameters in PROLITH) to account for a particular fab setup [93, 113]. Since our primary purpose is to develop diagnostic approaches through statistical modeling, in this study, we may safely assume that the PROLITH simulator is well tuned without performing any expensive physical experiments, which will only be necessary when the approach is to be used on real applications.

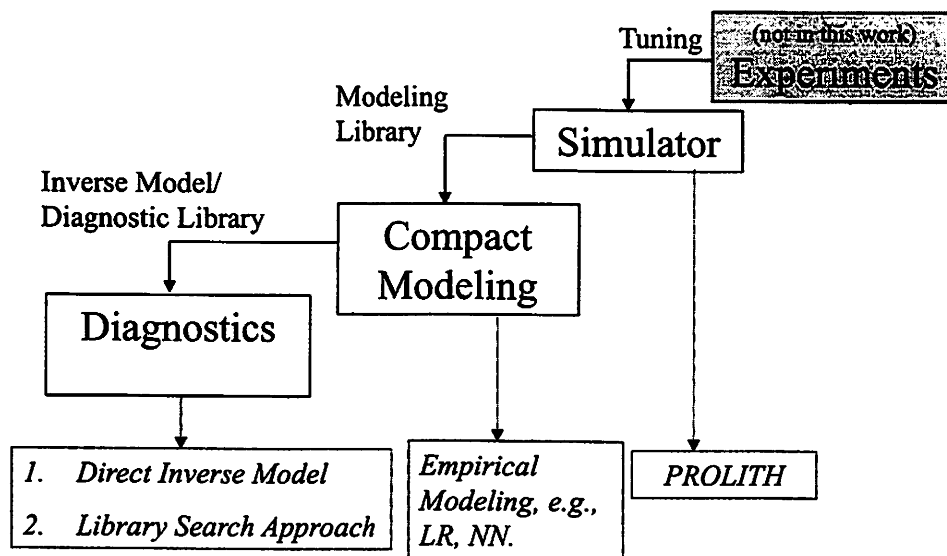

Figure 5.7: Overview of diagnostics system using a hierarchical structure.

5.2 Empirical Modeling of Lithography Process

In this section, we first describe the procedure for generating modeling data with PROLITH. Then, we show modeling results by linear regression (LR) and neural networks (NN) for forward/inverse models of different cases. Although we did not use linear regression models ultimately in diagnostics, we found that linear regression was a good starting point for data analysis due to its simplicity and transparency. Applications of forward/inverse models in diagnostics will be discussed in subsequent sections.

5.2.1 Modeling Data from PROLITH

The training and testing data that we use for modeling the lithography process were obtained through PROLITH simulations. The simulation runs were designed in such a way that the five input parameters were distributed evenly in the parameter space. Each of the five parameters took on between five and nine values and a total of 11,340 data points were generated, which we believe is sufficient for our purpose. The simulation runs took about five hours on a 500MHz Pentium III PC.

Table 5.2 shows the ranges of input parameters and their numbers of values that we chose to simulate. Two thirds of all generated data were randomly selected as the training set and the remaining as the test set to ensure no overlap between the two sets. For comparability, the same training and test data sets were used for all modeling cases and all performance results that we report are based on the test data set unless specified otherwise. To accommodate measurement noise,

Input Parameters	Input Parameter Ranges	# of Values taken in simulation
Exposure Dose	$10 \sim 15\ mJ/cm^2$	6
Focus Position	$-0.3 \sim 0.5\ \mu m$	9
Partial Coherence	$0.4 \sim 0.7$	7
PEB Time	$54 \sim 66\ sec$	5
PEB Temperature	$130 \sim 136\ ^oC$	6

Table 5.2: Input parameters and their ranges.

CD Profile Features	Measurement Noise Range
10%CD	$1\ nm$
50%CD	$1\ nm$
90%CD	$1\ nm$
SideWall Angle	0.1^o

Table 5.3: Output measurement noise range.

commonly observed in realistic environments, we have introduced Gaussian noises in the data we used for diagnostics. Table 5.3 shows the 95% ranges (two folds of the standard deviations) of the white Gaussian noises we added to the simulation output data.

The high dimensionality of the data set makes it very difficult to visualize. To get a sense of the relationship among various input and output variables, we plot a subset of all variables at a time while keeping other variables constant. Figure 5.8 shows CDs versus focus for different exposure values while partial coherence, PEB time and temperature are fixed. Figure 5.9 shows CDs versus PEB temperature for different PEB time values while exposure, focus and partial coherence are fixed.

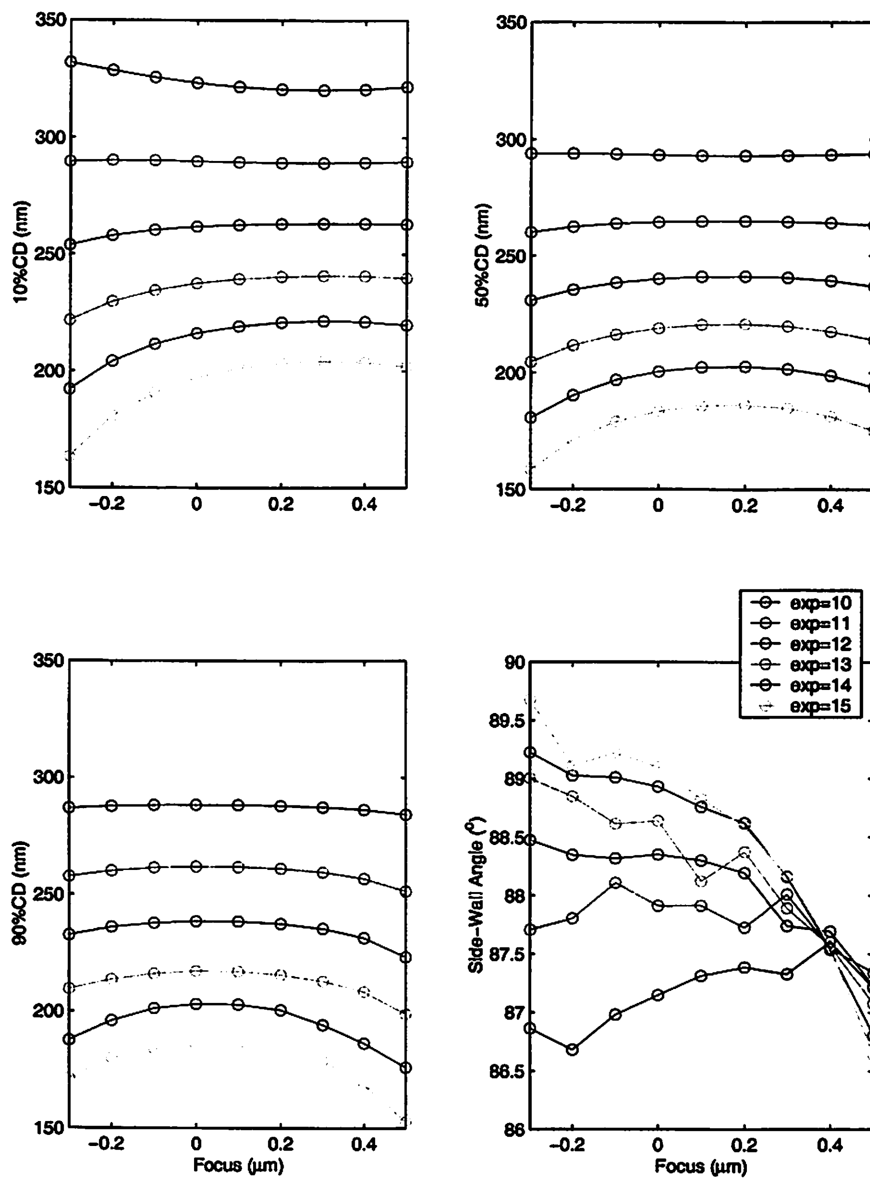

Figure 5.8: CD vs. focus for different exposure, when $coh = 0.55$, $t_{PEB} = 60s$, and $T_{PEB} = 132^oC$.

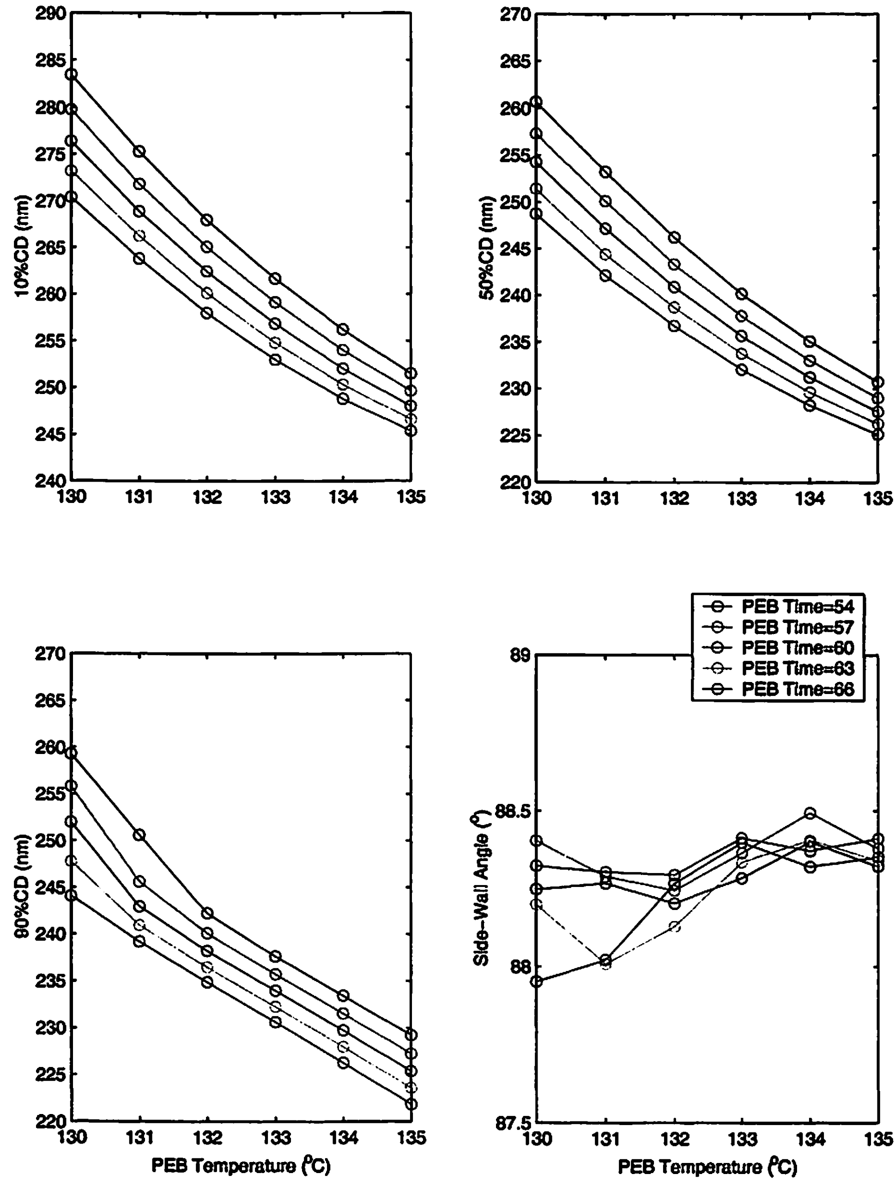

Figure 5.9: CD vs. PEB temperature for different PEB time, when $exp = 12mJ/cm^2$, $foc = 0.1\mu m$, and $coh = 0.55$.

Response	R^2 Adj.	RMSE	Mean of Response
$CD_{0.1}$	0.964185	9.165574 nm	248.0956 nm
$CD_{0.5}$	0.972881	7.108201 nm	225.1709 nm
$CD_{0.9}$	0.964005	7.952275 nm	220.6018 nm
SWA	0.476337	0.564782^o	87.96233^o

Table 5.4: First order linear regression.

Response	R^2 Adj.	RMSE	Mean of Response
$CD_{0.1}$	0.992834	4.099669 nm	248.0956 nm
$CD_{0.5}$	0.994200	3.287385 nm	225.1709 nm
$CD_{0.9}$	0.991919	3.768042 nm	220.6018 nm
SWA	0.938786	0.193099^o	87.96233^o

Table 5.5: Second order response surface model linear regression.

5.2.2 Linear Regression

Linear regression is perhaps the most popular statistical modeling method and is widely used in the semiconductor industry. Although linear regression may not be as powerful as some non-linear modeling methods, it is certainly helpful in gaining insights from data with relatively little computational overhead. The simple and transparent nature of linear models can often make good suggestions of what more complicated models to try next.

Table 5.4 gives the fitting results of a first order linear regression model. The Root Mean Squared Errors (RMSE) of CDs range between 7.11 and 9.17nm. A second order response surface model fits the data much better then first order model, but the RMSEs of CDs could still be as high as 4.10nm. Neither the first order nor the second order linear models fit the data very accurately.

5.2.3 Neural Network Modeling

Compared to linear regression, neural networks are much more flexible and powerful in modeling highly non-linear mappings. The number of active parameters (which is usually the number of link weights) determines the function space a neural network can approximate as well as its modeling accuracy. For the two-layer Multilayer Perceptron (MLP) networks that we have used in this study (cf. section 2.2.3), the number of hidden nodes essentially determines the number of active link weights. More hidden units an MLP has leads to higher modeling accuracy at the cost of longer training time. When there are only a limited amount of training data, too many hidden units may also lead to overfitting the network parameters, in which case, network performance seems extremely good on training data but much poorer on unseen test data. Therefore, some strategy of regularization, such as early-stopping using cross-validation, may be necessary. All MLPs in this study were trained using the Scaled Conjugate Gradient algorithm [11], a commonly used second-order quasi-Newton method, where the gradients were estimated via the standard back-propagation procedure.

The first case is forward modeling from the five input parameters to the four output CD measurements. The architecture of the MLP is shown in figure 5.10. Table 5.6 shows the results of MLP forward modeling where the RMSEs of CDs range between 1.11 and 1.62nm when an MLP with 10 hidden nodes were trained for 5,000 iterations. Clearly, when sufficient number of hidden nodes were used, the MLP model fits the data much more accurately than the second order linear regression model.

In the second case, we tried inverse modeling using MLP as a precursor for

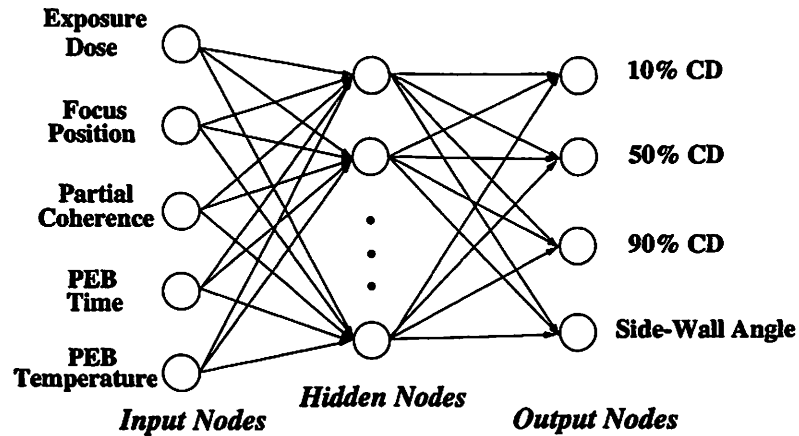

Figure 5.10: Forward MLP model.

MLP Simulation		Root Mean Square Error				CPU Time
# of Hidden Nodes	# of Iterations	$CD_{0.1}$ (nm)	$CD_{0.5}$ (nm)	$CD_{0.9}$ (nm)	SWA (^{o})	($minutes$)
5	1000	3.2679	2.4374	2.8299	0.1795	6.8
5	5000	4.1680	3.5641	3.5147	0.1458	33.0
10	1000	1.4818	1.2408	1.7034	0.1230	11.7
10	5000	1.1828	1.1071	1.6219	0.1190	56.5

Table 5.6: Forward MLP modeling results.

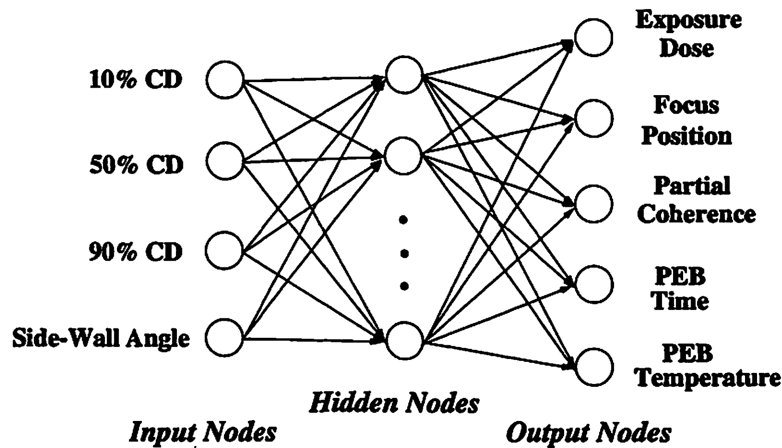

Figure 5.11: Inverse MLP model.

the diagnostic approach using explicit inverse model. Figure 5.11 shows the network architecture and table 5.7 gives the modeling results. Unfortunately, the poor results indicate that the model was not sufficiently accurate for the purpose of process diagnostics. This is not surprising given the intrinsic ambiguity of the inverse mapping from output variables to input parameters. Further analysis suggests that much better modeling performance may be expected if we would limit the estimation to only one input parameter (instead of all five), assuming others were known, especially if the output parameters were monotonic with respect to that input parameter. As a proof of concept, we show in figure 5.12 and table 5.8 the network architecture and modeling performance when only exposure dose is unknown. In next section, we describe the explicit inverse model approach with more details.

MLP Simulation		Root Mean Square Error					CPU Time ($mins$)
# of Hidden Nodes	# of Iterations	Exposure (mJ/cm^2)	Focus (μm)	Partial Coherence	PEB Time (s)	PEB Temp (^oC)	
5	1000	0.5254	0.0804	0.0789	4.1968	1.6431	7.0
5	5000	0.5017	0.0843	0.0714	4.1955	1.6308	33.9
10	1000	0.5086	0.0809	0.0761	4.1902	1.6009	12.0
10	5000	0.4940	0.0771	0.0699	4.1805	1.5892	57.3

Table 5.7: Inverse model from four outputs to five inputs.

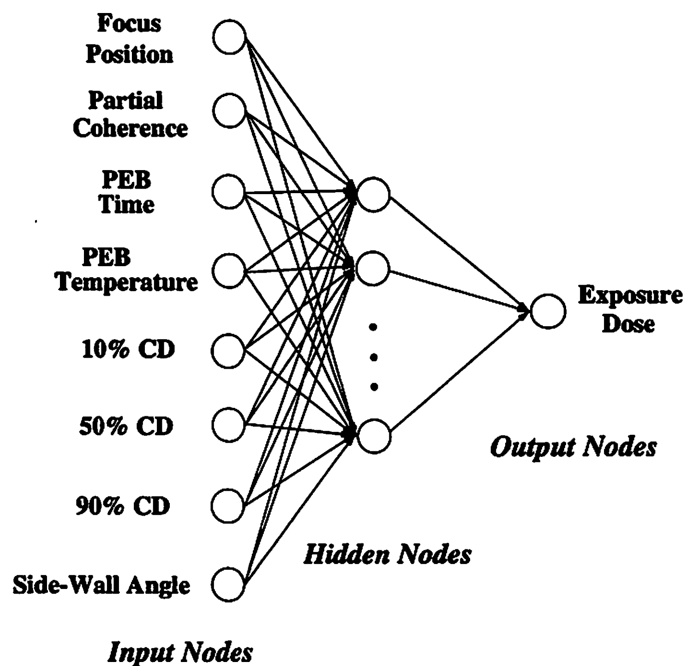

Figure 5.12: Inverse MLP model for exposure dose.

MLP Simulation		Root Mean Square Error	CPU Time
# of Hidden Nodes	# of Iteration	Exposure (mJ/cm^2)	(minutes)
5	1000	0.0689	5.7
5	5000	0.0654	28.3
10	1000	0.0242	10.5
10	5000	0.0205	51.1

Table 5.8: Inverse model from four outputs and four inputs to one input.

5.3 Inverse Modeling for Process Diagnostics

If a unique, accurate inverse model exists for a process, the diagnostics will be simple. In such cases, for each set of CD measurements, we can estimate the corresponding inputs directly from the inverse model.

A unique inverse model does not exist in cases where outputs are non-monotonic functions of certain input parameters (e.g., focus position as shown in figure 5.8) or the effects of some parameters are highly coupled (e.g., PEB time and temperature as shown in figure 5.9). In the highly non-linear lithography process that we are studying, a one-to-one mapping from the outputs to the inputs often does not exist. This means that each set of input parameters uniquely determines a CD profile, but the same CD profile may be generated by different combinations of input parameters, and diagnostics are therefore non-trivial.

In many cases where an explicit inverse model is difficult to obtain or simply does not exist, we have to rely on forward models to simulate input-output pairs and seek alternative strategies to identify possible inputs for a given output (e.g., the library-search approach that we will introduce in the next section). In other cases, although a universal inverse model does not exist, some local ones do. That is, if we are able to properly divide the parameter space into subspace segmentations, an inverse model would exist for each segmentation. For a given output measurement, if we have the knowledge of which segmentation the input parameters fall in, an estimation of the input parameters can be made using local inverse models. Clearly, an important task for this approach is the space segmentation, which may be quite challenging due to the non-linearity and high-dimensionality

of the problem.

For this five-input, four-output process that we are studying, the CD measurements are non-monotonic functions of the five input parameters due to, specifically, the focus position parameter and also effects of some input parameters to CD are highly coupled, such as PEB time and temperature. In the following subsections, we will first show that when the focus position and some of the interacting parameters are known, we can perform diagnostics reasonably well using the explicit inverse model approach. We will then consider how to perform space segmentation when the focus position is unknown. Decoupling PEB time and temperature is a very difficult problem, for which we do not have a viable solution yet.

5.3.1 Application of Explicit Inverse Model

If we know the precise mapping from outputs to input parameters, we can apply the inverse model directly to perform diagnostics as shown in figure 5.4. Unfortunately, in our problem, a unique analytical solution for the inverse model does not exist due to the non-linearity of the process. We can build statistical inverse models using the same techniques as in forward modeling. However, because the outputs and inputs are not related in a one-to-one fashion, we cannot expect optimal performance from the explicit inverse modeling. In the following series of figures (figure 5.13 through figure 5.24), the estimation results are shown for all possible modeling cases (cf. Tabel 5.9). Each of the models uses an MLP$(m, 10, n)$ where m is the number of model inputs (5 to 8) and n is the number of model outputs (1 to 4). For all the models, ten hidden nodes are used. The data we use for inverse modeling are the same as that for the forward modeling. For all the

Case #		Exposure	Focus	PEB Time	PEB Temperature
I (1 unknown, 3 known)	1	N	Y	Y	Y
	2	Y	N	Y	Y
	3	Y	Y	N	Y
	4	Y	Y	Y	N
II (2 unknown, 2 known)	1	N	N	Y	Y
	2	Y	Y	N	N
	3	Y	N	N	Y
	4	N	Y	Y	N
	5	N	Y	N	Y
	6	Y	N	Y	N
III (3 unknown, 1 known)	1	N	N	N	Y
	2	N	N	Y	N
	3	N	Y	N	N
	4	Y	N	N	N
IV (4 unknown)	1	N	N	N	N

Table 5.9: Simulation cases for all diagnostic approaches. Y: known; N: Unknown

results, an estimate of each parameter is calculated based on the statistics of 1000 simulated data samples normally distributed with standard deviation 0.5nm for all CDs (so that 95% confidence interval is $\pm$1nm) and 0.05^o (so that 95% confidence interval is $\pm 0.1^o$) for Side Wall Angle.

Figure 5.13 shows the parameter estimation performances of the inverse-model approach when only one parameter is unknown. Under such conditions, the exposure can be estimated accurately while the other three parameters have large estimation variances. This is largely due to the inaccuracy of the inverse models. When more hidden nodes are included in the MLP, the estimation results improve slightly. However, as the number of hidden nodes increases, the problem of overfitting worsens.

Figure 5.14 through figure 5.19 show the results when different combinations

of two parameters are unknown. Figure 5.20 through figure 5.23 show the cases when different combinations of three parameters are unknown. Finally, figure 5.24 shows the results when all four parameters are unknown. These results are generally unsatisfactory and the performance degrades as the number of unknown parameters increases. Next, we show how space segmentation improves estimation performances when the universal explicit inverse models do not yield good solutions.

5.3.2 Space Segmentation

From section 5.2.1 (cf. figure 5.8) we notice that the CDs are non-monotonic functions of focus. However, it may be possible to divide the parameter space into segments such that the CD outputs within each segment are monotonic with respect to all input parameters. If a separate model is developed for each of these segments, it should give a better fit to the data than a single model for the entire parameter space. Since the non-monotonicity of focus is an important cause of the inaccuracy of the inverse modeling, addressing this problem is likely to improve the parameter estimation performance.

Of course, the high dimensionality and non-linearity of the lithography process lead to a very difficult segmentation problem if we were to find an optimal segmentation involving all input parameters. So as a first step, we consider a simplified situation where the parameter space is only subdivided along the dimension of the focus position, and propose the following approach as a rough input parameter space segmentation. We take the sign change of the first derivative of output function with respect to input as an indication of non-monotonicity. For certain

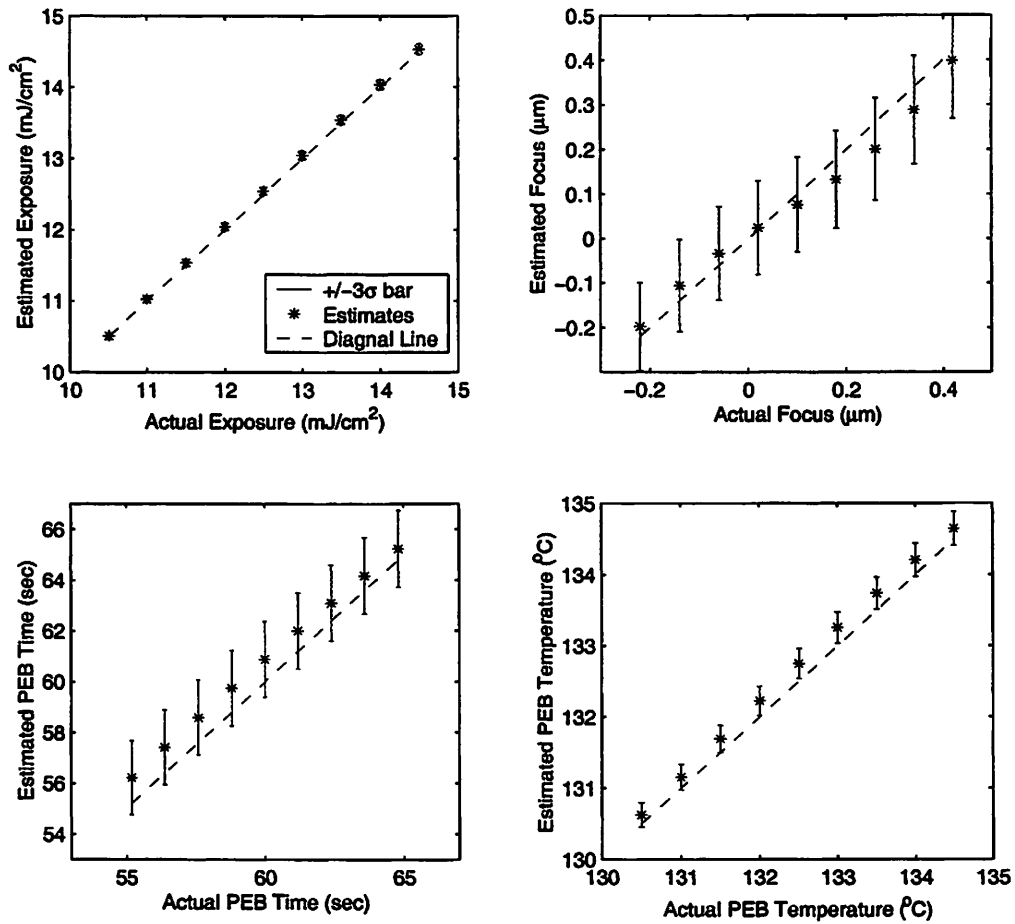

Figure 5.13: Explicit inverse model approach case I #1~#4 results: the estimated parameter is the only unknown.

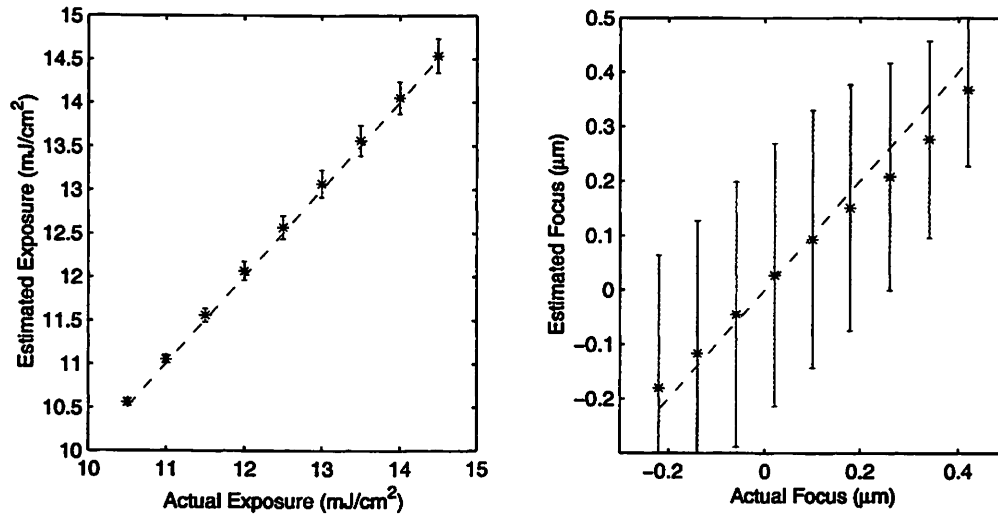

Figure 5.14: Explicit inverse model approach case II #1 results: exposure, focus unknown; PEB time and temperature known.

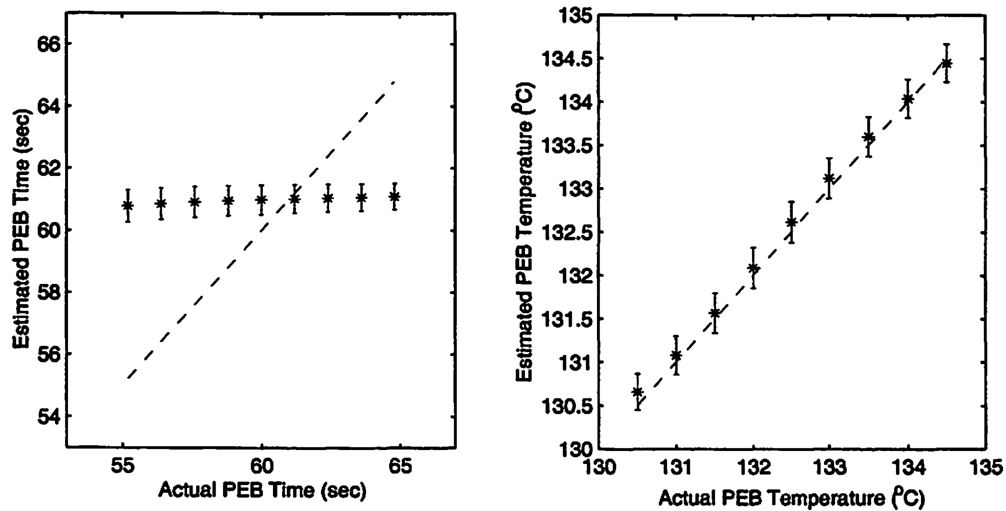

Figure 5.15: Explicit inverse model approach case II #2 results: PEB time and temperature unknown; exposure, focus known.

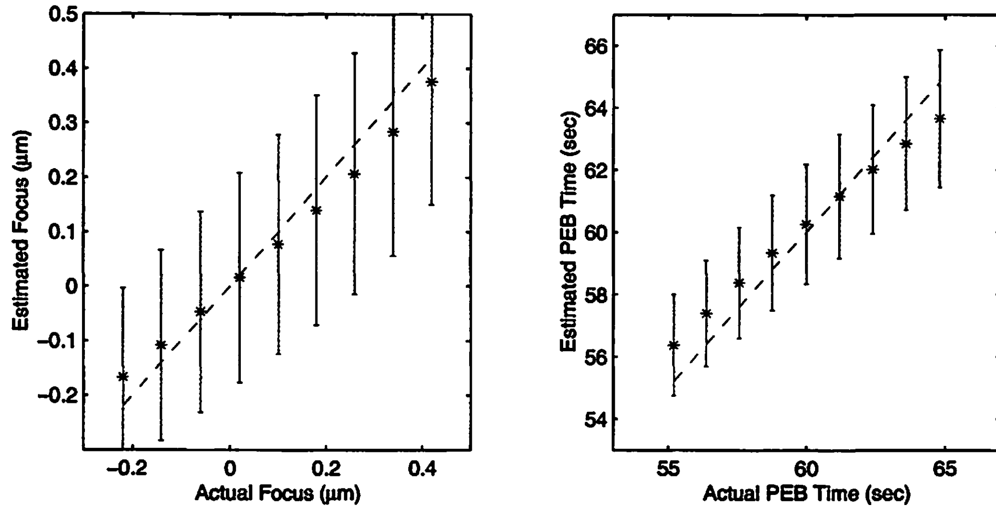

Figure 5.16: Explicit inverse model approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known.

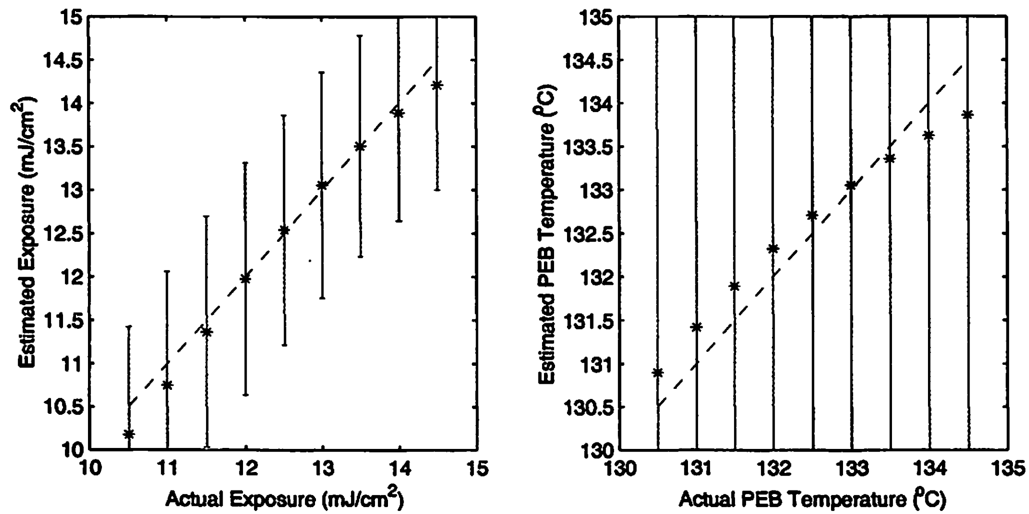

Figure 5.17: Explicit inverse model approach case II #4 results: PEB temperature, exposure unknown; PEB time, focus known.

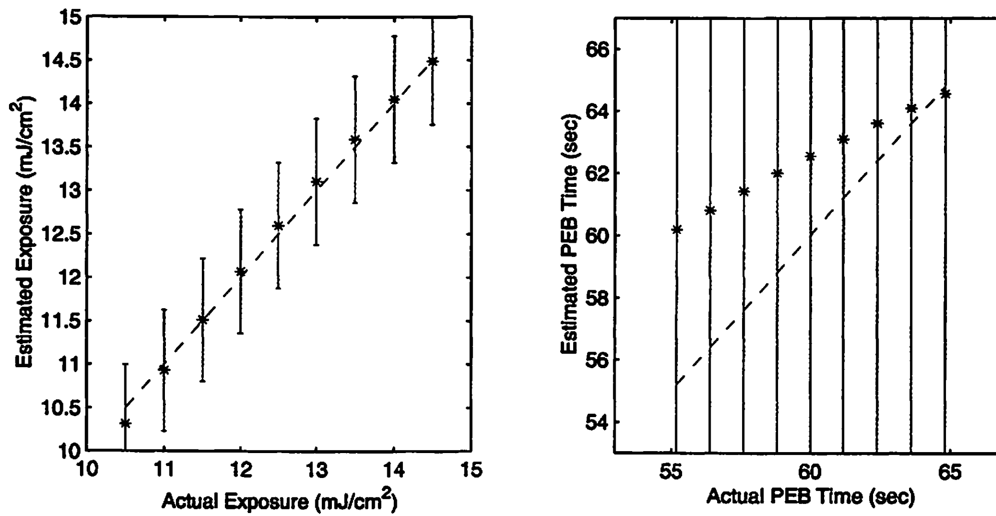

Figure 5.18: Explicit inverse model approach case II #5 results: PEB time, exposure unknown; PEB temperature, focus known.

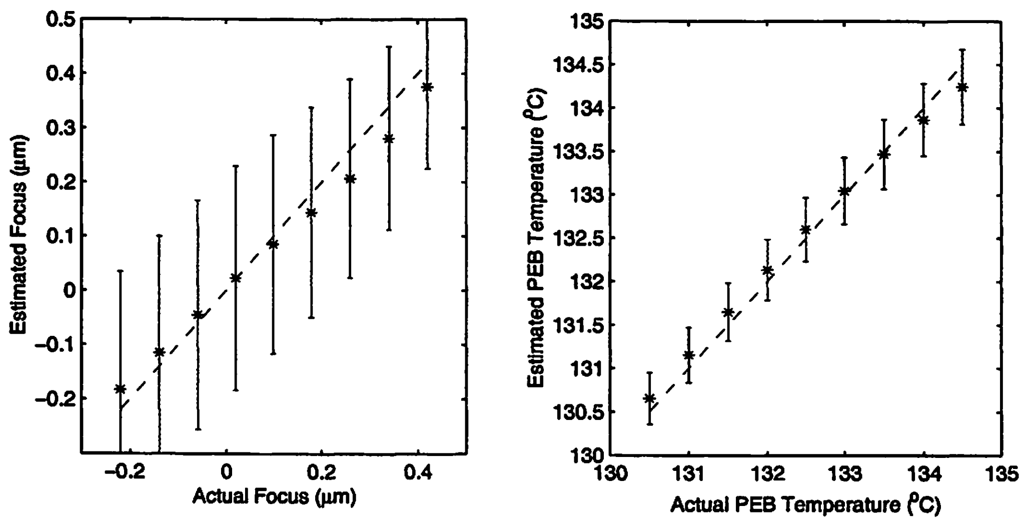

Figure 5.19: Explicit inverse model approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known.

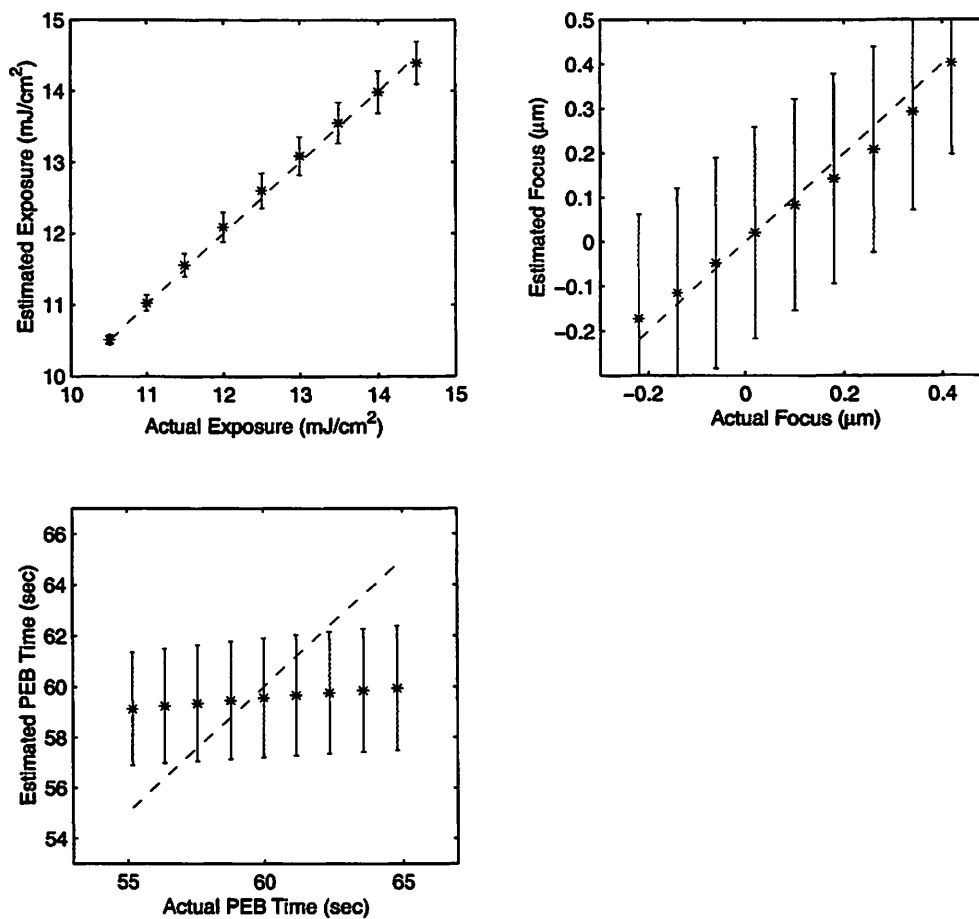

Figure 5.20: Explicit inverse model approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known.

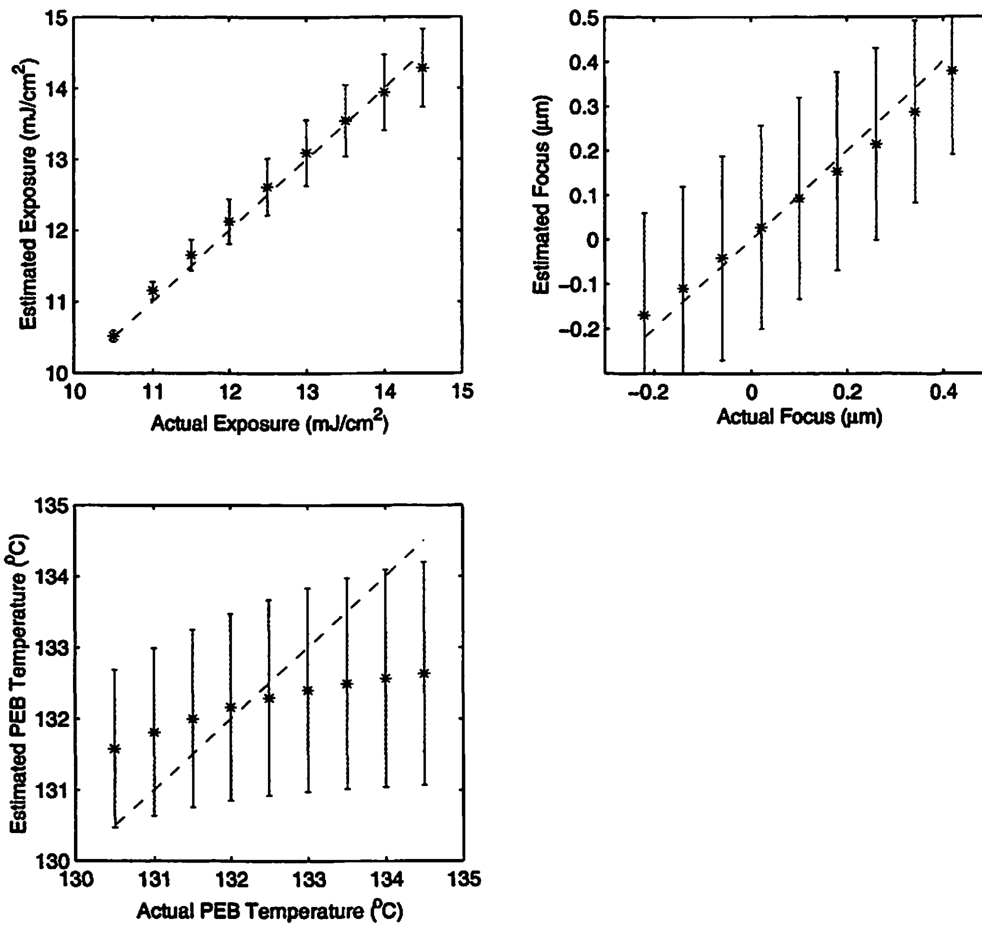

Figure 5.21: Explicit inverse model approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known.

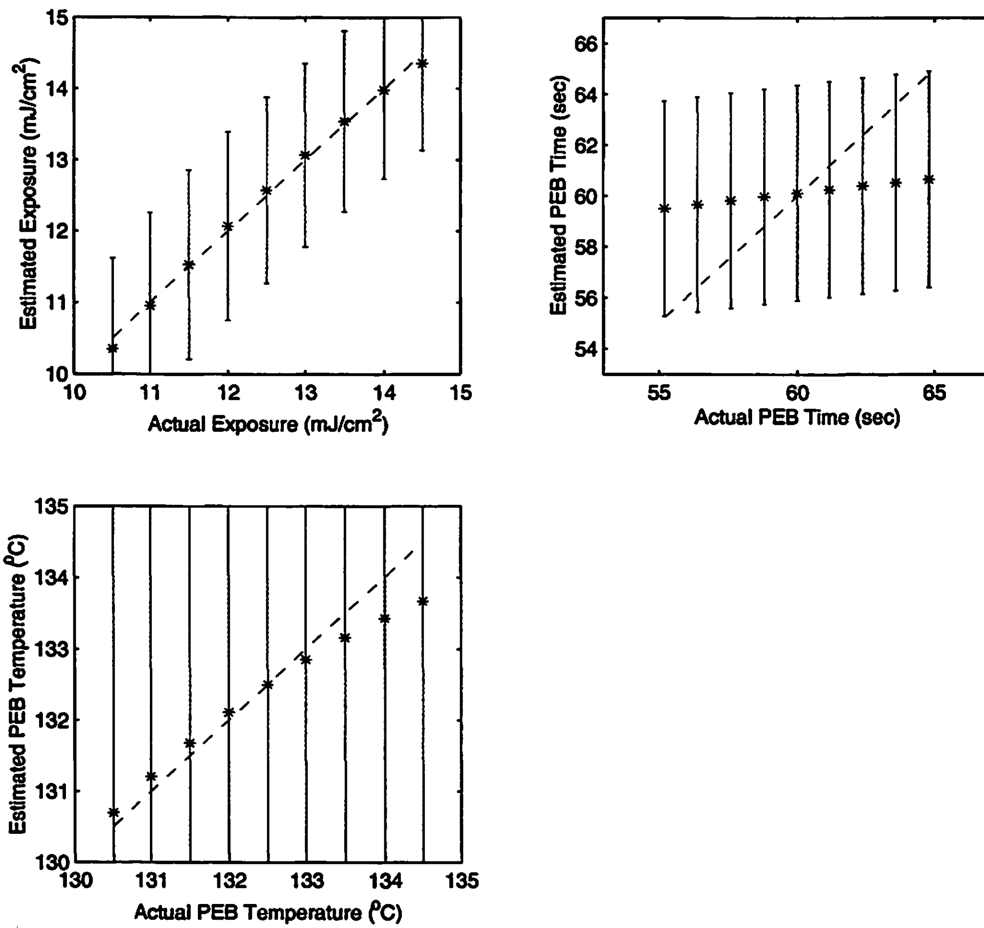

Figure 5.22: Explicit inverse model approach case III #3 results: exposure, PEB time and temperature unknown; focus known.

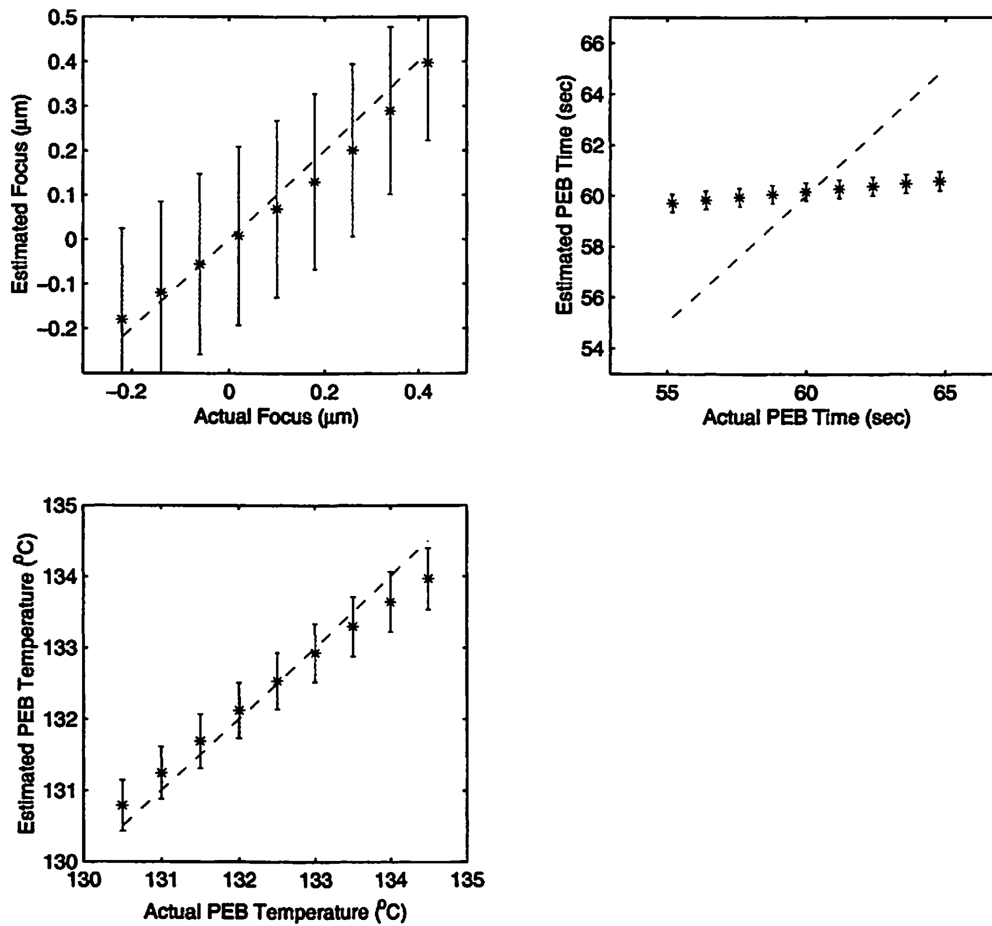

Figure 5.23: Explicit inverse model approach case III #4 results: focus, PEB time and temperature unknown; exposure known.

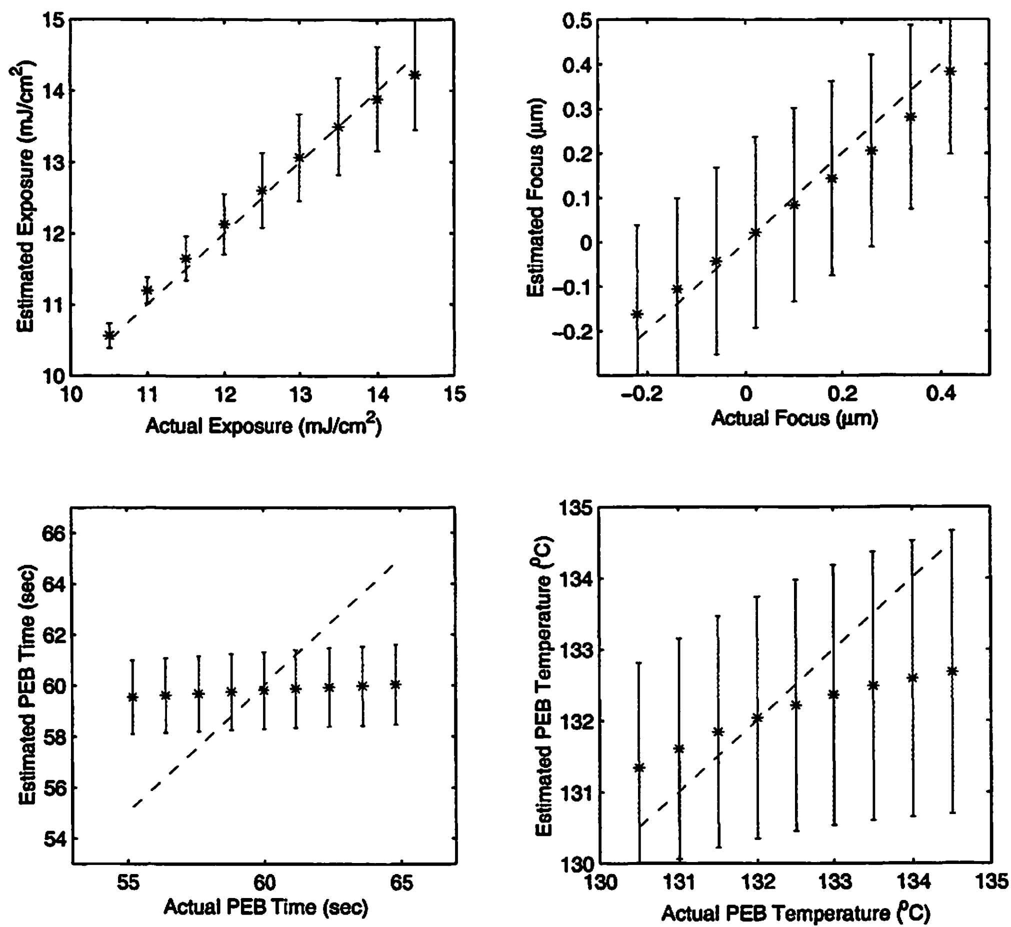

Figure 5.24: Explicit inverse model approach case IV results: exposure, focus, PEB time and temperature all unknown.

modeling methods, such as the neural network modeling, it is very difficult to calculate the first derivative since an analytical expression between input and output is extremely complex. Recall however, that the second order response surface linear regression model gives reasonable performance while offering an analytical expression for calculating the first derivative of output with respect to input parameters. Thus, we may take the second order response surface linear regression model as a compromise between accurate segmentation and computation plausibility.

By observing the data collected from PROLITH, we realize that due to the different values other parameters take, the location of the first derivative sign change is not necessarily a single point but rather a small range, over which the sign of the first derivative of focus position also depends on other parameters. We define the boundary range so that within that range, at least one output is not a monotonic function (i.e., we require that the first derivative of at least one output over an input changes sign) of the parameter to be examined (i.e., focus). In general, the dimension of this type of boundary ranges depends on the number of parameters involved in the non-monotonicity. In our case, focus is the only parameter involved, therefore the boundary ranges are just one-dimensional intervals of focus. When the focus position has a value that is far from any of the boundary ranges, we should expect to find a good inverse model. However, when the focus position is in the vicinity of a boundary range, it becomes considerably more difficult to find an accurate inverse model. One possible remedy is to generate more data close to the boundary ranges in order to refine modeling; another is to use a more complex modeling structure (e.g., to use more hidden nodes in the MLP network). Figure 5.26 is an example of how this boundary range of focus position is determined. This figure gives the scatterplots of all data points in 2-d planes,

where the horizontal axis is always the focus and the vertical axes are the first order partial derivatives of the three CDS over focus respectively. The "vertical bar" for each focus value is made up of a large number of data points that share the common focus value when all other variables take different values. The black arrows on the plots in figure 5.26 show the range over which the change of sign of the first derivatives takes place for each output CDs. [-0.1μm, 0.2μm] is the common subset, within which zero derivative lines of all three outputs over focus lie. This range is thus the boundary range that we are seeking.

This boundary range [-0.1μm, 0.2μm] for focus is used in simulations in this work. Three MLP models are developed for the three segments (according to the range of focus: [-0.3μm, -0.1μm], the left segment; [-0.1μm, 0.2μm], the boundary segment; [0.2μm, 0.5μm], the right segment). For the left and right segments, we use MLPs with 10 hidden nodes, just as we did for the universal inverse models. For the boundary segment, we use MLPs with 15 hidden nodes. Figure 5.27 through figure 5.34 are the simulation results for all unknown parameter combinations for the case when focus is unknown. Compared to the results for the universal inverse models, there are some obvious improvements in these results. In most of cases, we observe clear decreases in estimation variances.

However, the results are still not completely satisfactory as we still observe some high estimation variances. However, we have demonstrated that space segmentation is potentially useful if a reliable method is found to perform the segmentation systematically. In the next section, we investigate a different approach – the library searching approach, which is shown to perform much better.

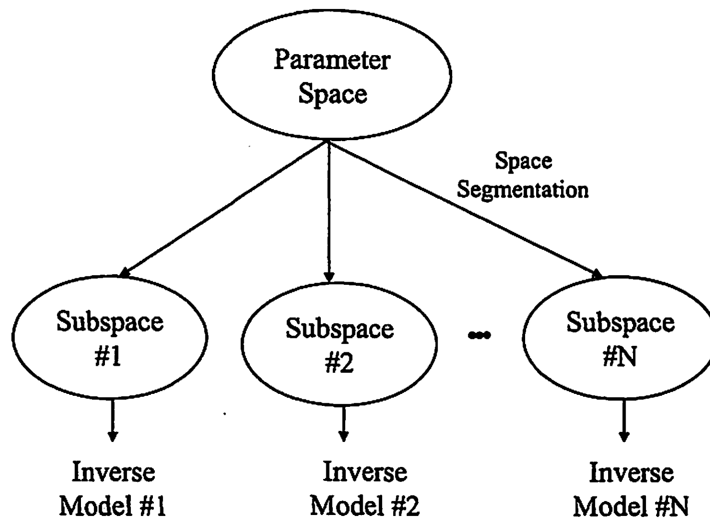

Figure 5.25: Space segmentation to find inverse models.

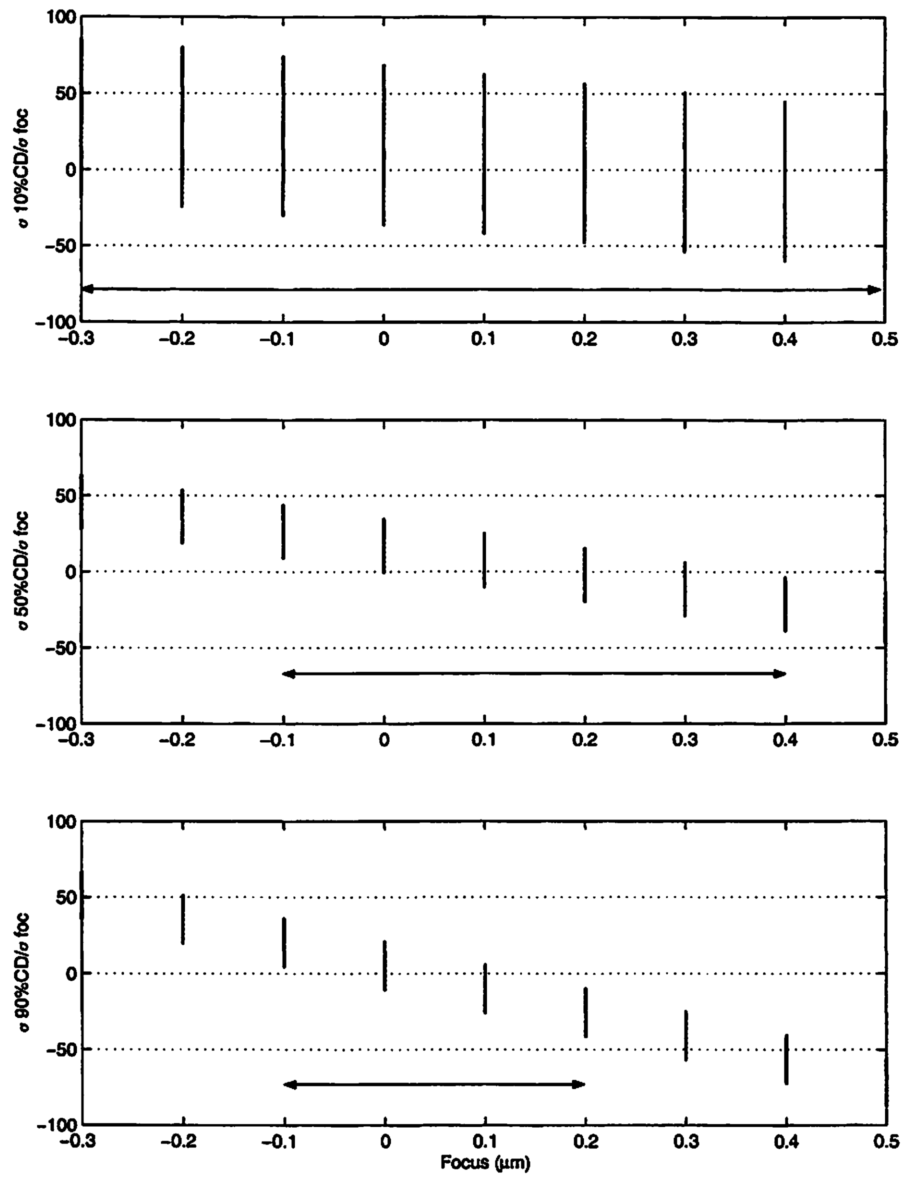

Figure 5.26: Example of space segmentation boundaries.

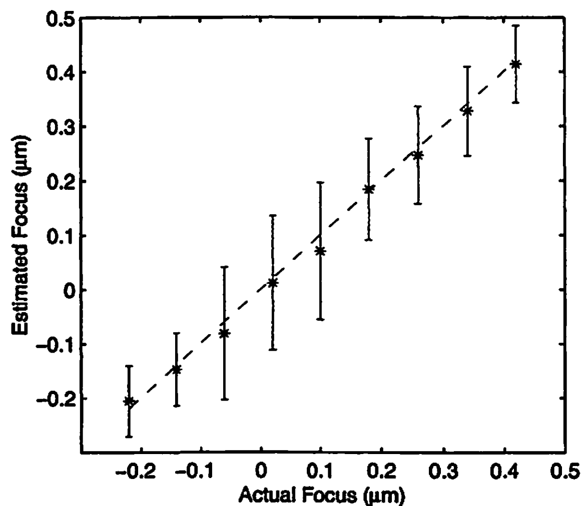

Figure 5.27: Segment inverse model approach case I #2 results: the estimated parameter is the only unknown.

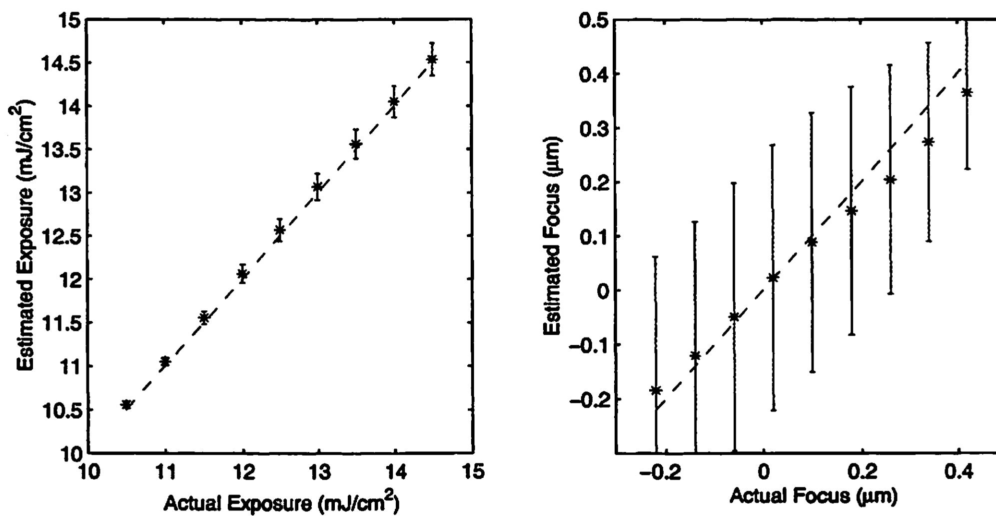

Figure 5.28: Segment inverse model approach case II #1 results: exposure, focus unknown; PEB time and temperature known.

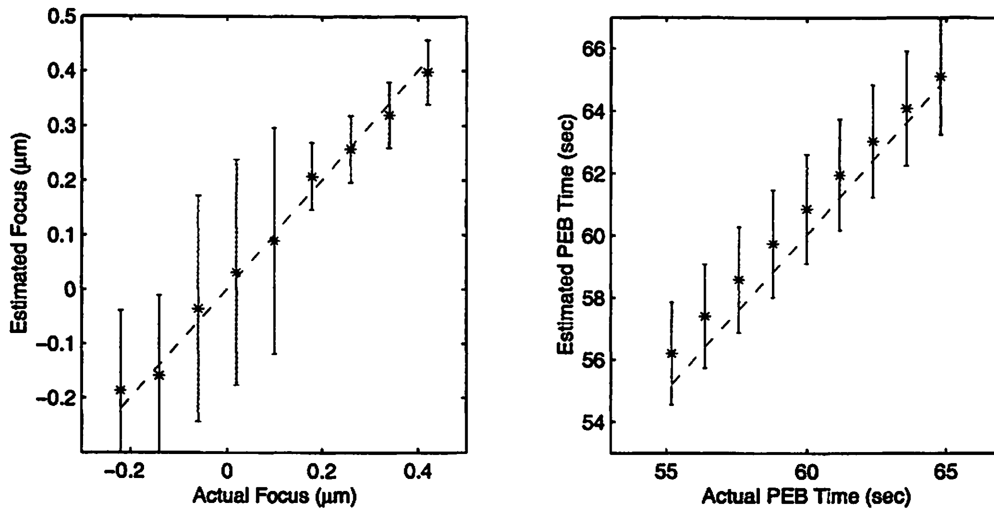

Figure 5.29: Segment inverse model approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known.

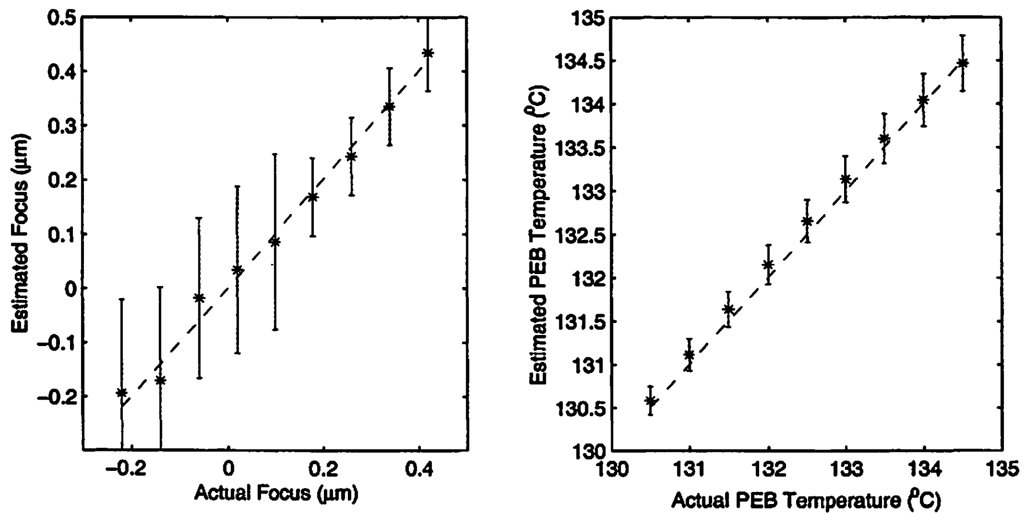

Figure 5.30: Segment inverse model approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known.

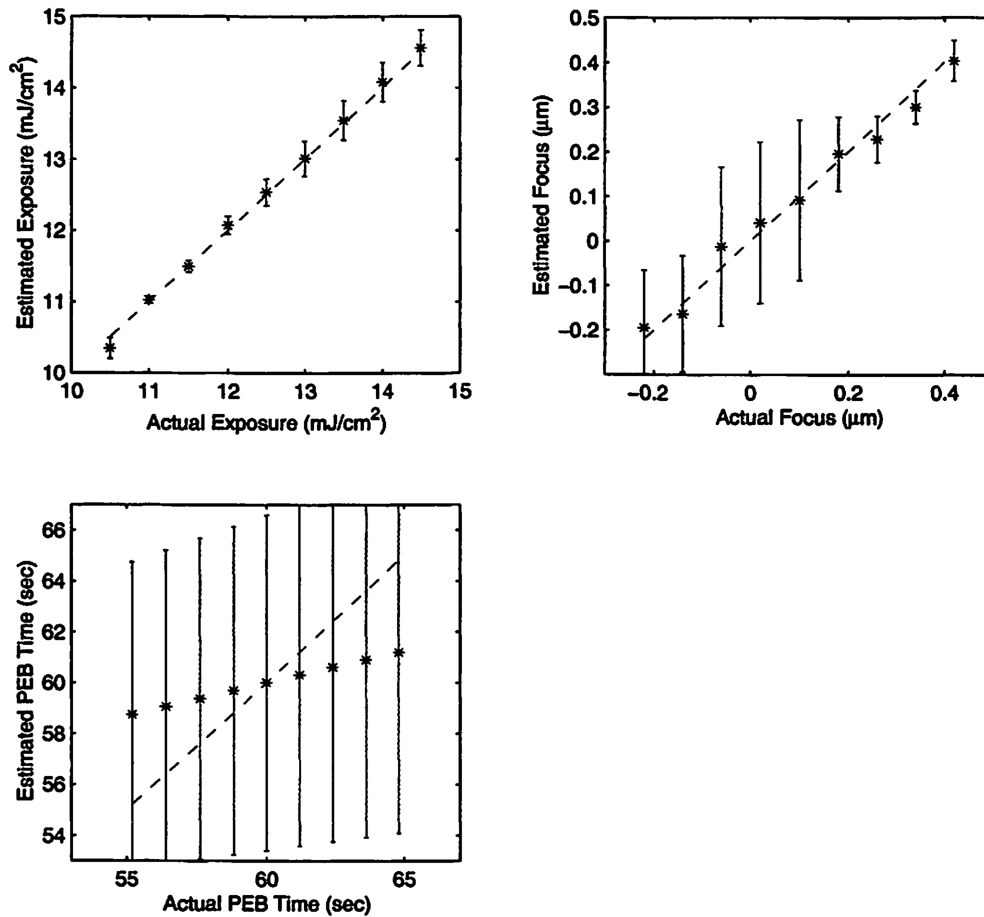

Figure 5.31: Segment inverse model approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known.

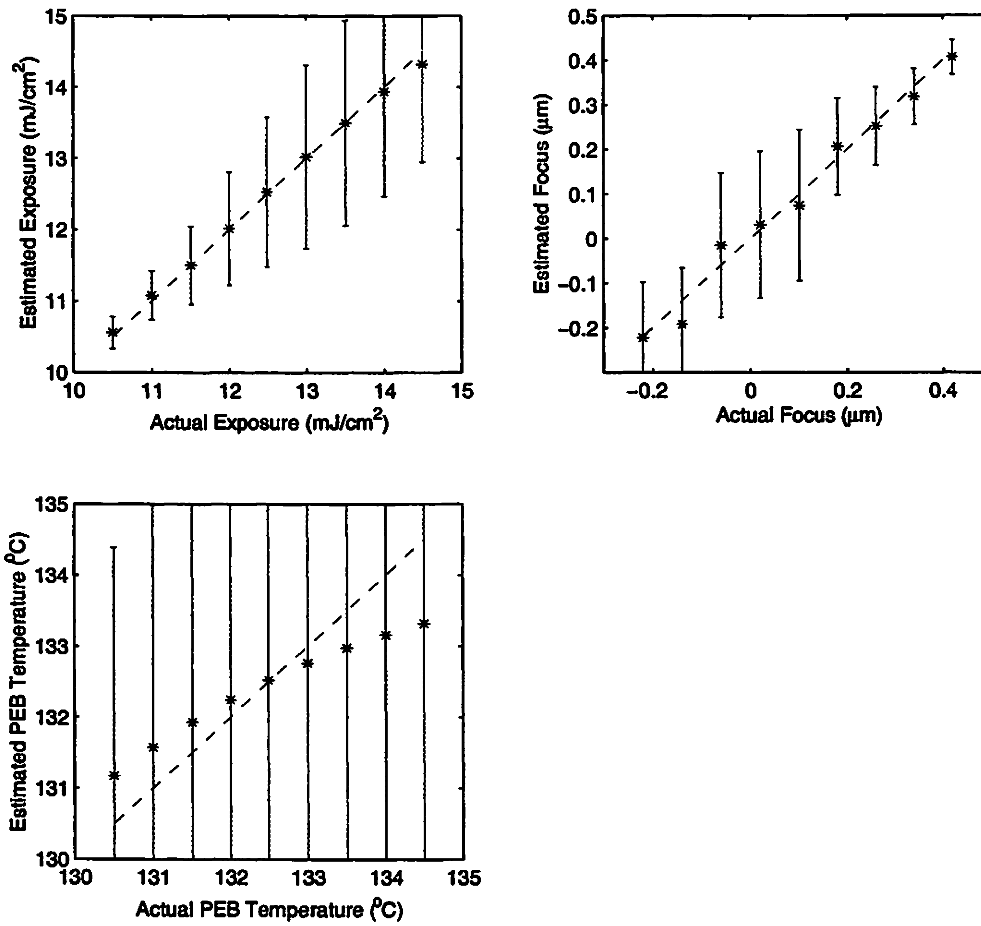

Figure 5.32: Segment inverse model approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known.

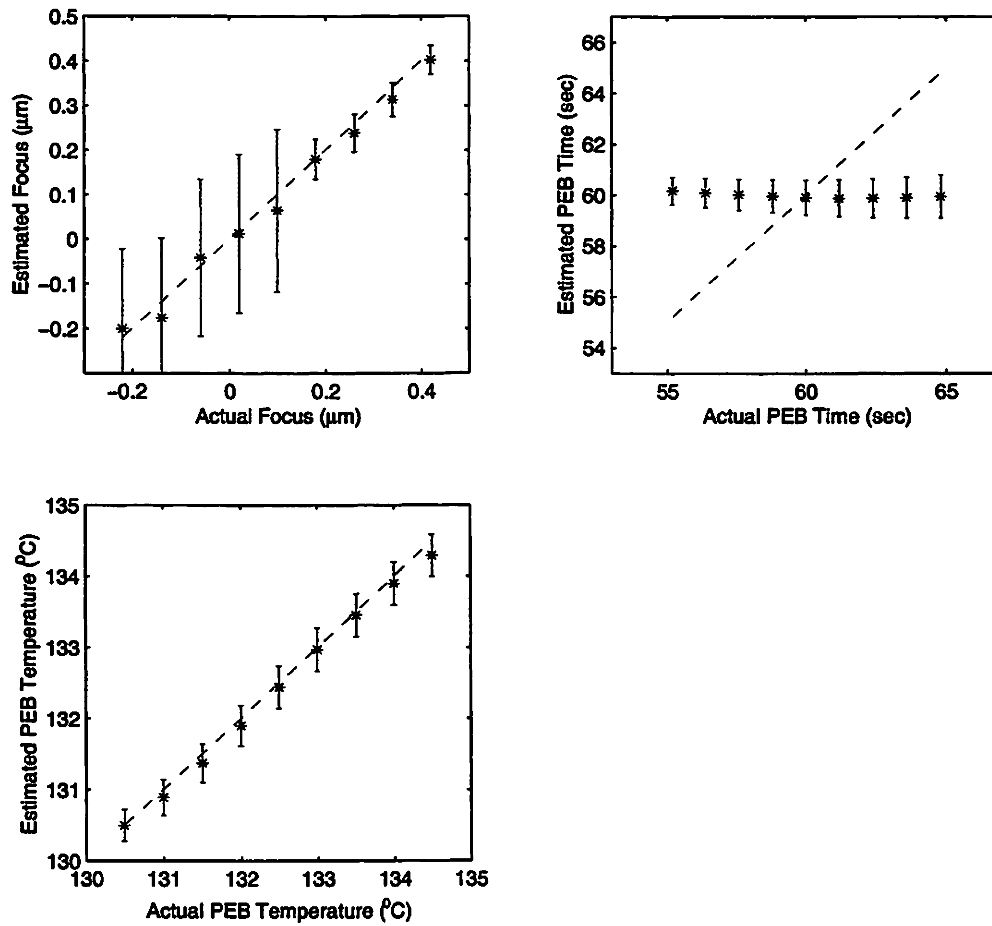

Figure 5.33: Segment inverse model approach case III #4 results: focus, PEB time and temperature unknown; exposure known.

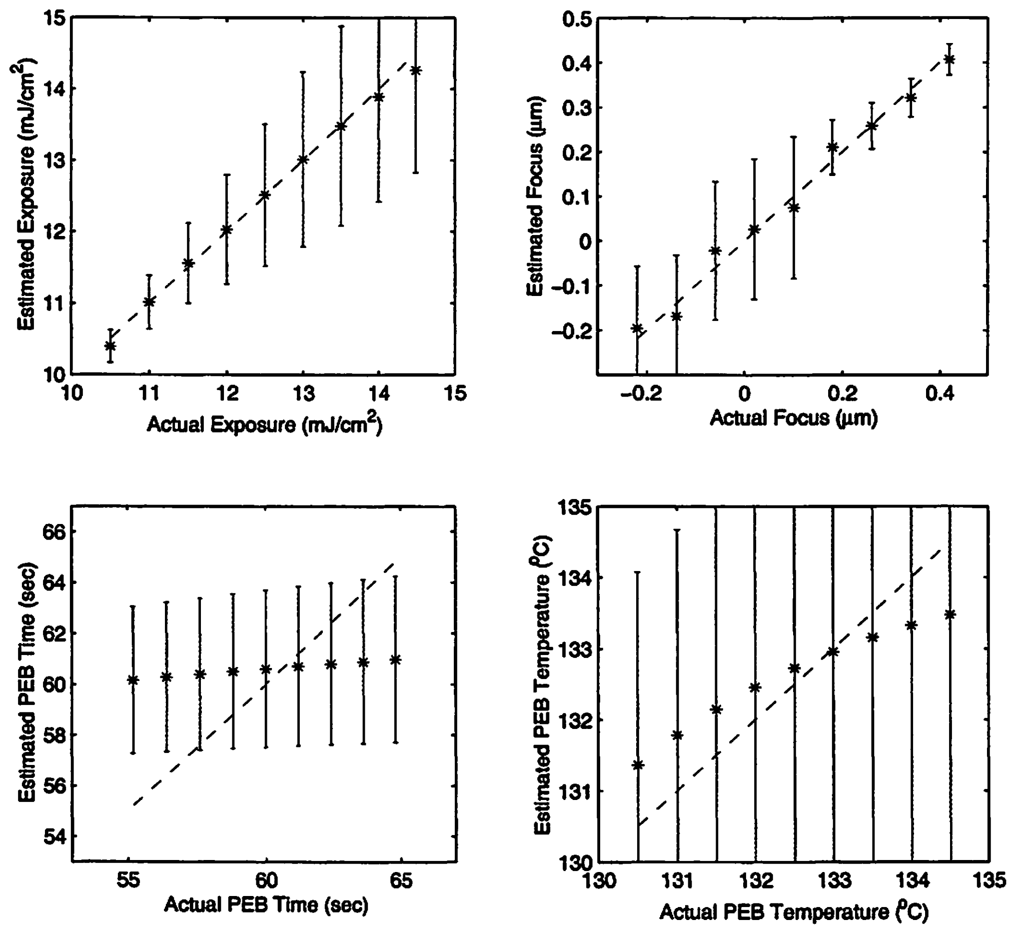

Figure 5.34: Segment inverse model approach case IV results: exposure, focus, PEB time and temperature all unknown.

5.4 Library Searching Approach

This section describes an alternative diagnostic approach that may be used when obtaining an explicit inverse model is not practical. In this approach, we build a library of input-output pairs by running the neural-network-based forward system model. The elements in the library have inputs uniformly distributed in the parameter ranges that we are interested in and the size of the library is determined by the number of the input parameters to detect. In theory, when a set of new CD measurements become available, we search in the library to find a match to the outputs and record the corresponding inputs. However, reality is not always so simple. In many cases, there may not exist an exact match since the library elements are generated randomly and we allow a certain amount of measurement error on the outputs. A cost function has to be defined to evaluate the closeness of the new measurements to the library elements, and a threshold is imposed for deciding which elements match the new CD measurements. In most cases, we get a set of close matches instead of a single unique solution. In this work, we use the following weighted sum of squared-error cost functions to evaluate the match between new measurements and library elements:

$$\begin{aligned} Cost(i) &= w_1[CD_{0.1} - CD_{0.1}^{lib}(i)]^2 + w_2[CD_{0.5} - CD_{0.5}^{lib}(i)]^2 \\ &\quad + w_3[CD_{0.9} - CD_{0.9}^{lib}(i)]^2 + w_4[SWA - SWA^{lib}i]^2 \end{aligned} \tag{5.1}$$

where $CD_{0.1}$, $CD_{0.5}$, $CD_{0.9}$ and SWA are the new CD measurements and $CD_{0.1}^{lib}(i)$, $CD_{0.5}^{lib}(i)$, $CD_{0.9}^{lib}(i)$, and $SWA^{lib}(i)$ are the output values of the corresponding library elements. The weights can be determined according to the desired matching accuracy for CD profiles and noise level of each output variable measurements.

To find candidate solutions with matching costs below the threshold, we use an exhaustive search method that evaluates the closeness of each element in the library to the new measurements. Other search algorithms may also be applicable, but exhaustive search provides an upper bound on search performance and remains computational feasible when the size of the library is not too large. After obtaining the candidate solution set, we calculate the input parameter estimate by a weighted sum of all solutions in the set. The weight for each solution is determined primarily by its cost value and a lower cost value corresponds to a greater weight. More specifically, we use a Gaussian weight curve as shown in figure 5.35 as the mapping from cost values to weights for each solution. The parameter σ of the Gaussian function is determined by the error ranges of parameters. Any bell-shaped curve can serve as the mapping for the weights, the Gaussian curve is just one example. In practice, different curves can be tried and compared in order to find a curve suitable for a specific problem.

Figures 5.37~5.48 give the simulation results of the library-searching approach for all different unknown parameter combinations. We can see that this approach is suitable for accurate estimation of each individual parameter. For the combinations of two unknown parameters, it does well for estimating exposure and focus, focus and PEB time, as well as focus and PEB temperature. For the combinations of three and four unknown parameters, it can estimate some parameters extremely accurately but less well for others. Table 5.10 gives a summary of the library searching results, which shows how well the approach fares on each of the parameter combinations. A "Y" indicates that the parameter is known; an "N" indicates that the parameter is unknown and needs to be estimated; a "Good" or a "Bad" after an "N" tells whether the estimation result is good or bad, re-

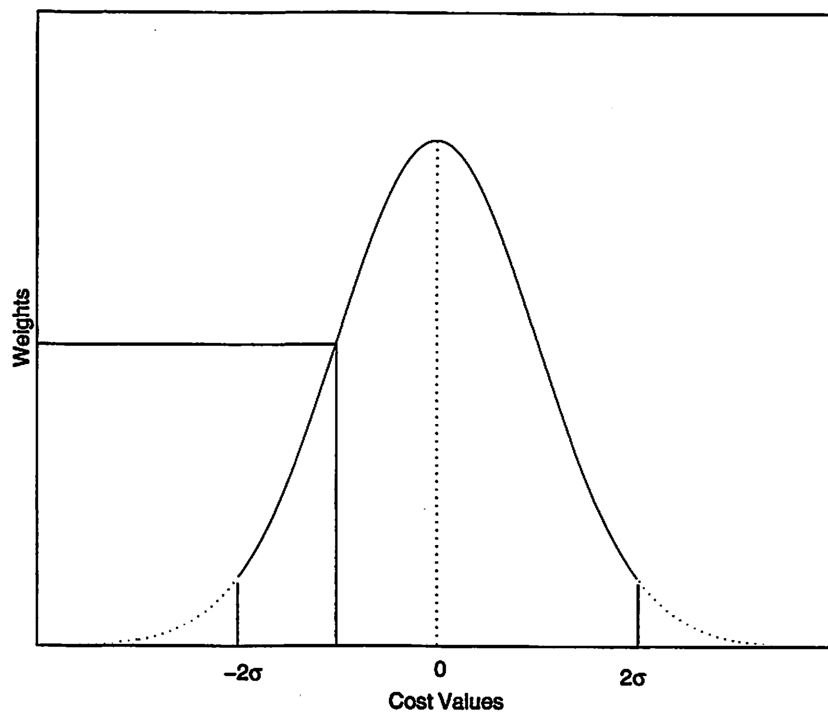

Figure 5.35: Determining weights using a Gaussian curve.

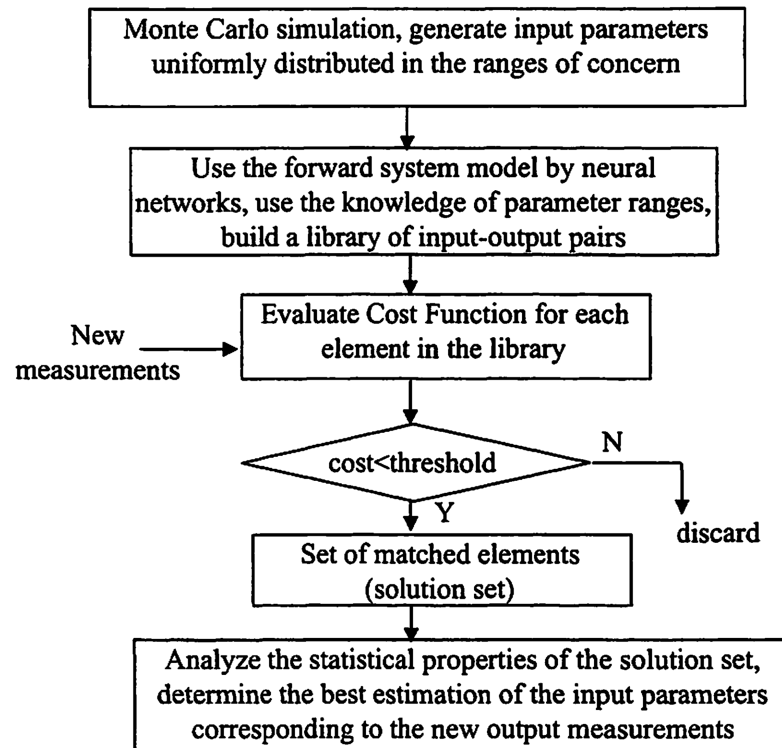

Figure 5.36: Diagnosis using a library searching approach.

Case #		Exposure	Focus	PEB Time	PEB Temperature
I (1 unknown, 3 known)	1	N (Good)	Y	Y	Y
	2	Y	N (Good)	Y	Y
	3	Y	Y	N (Good)	Y
	4	Y	Y	Y	N (Good)
II (2 unknown, 2 known)	1	N (Good)	N (Good)	Y	Y
	2	Y	Y	N (Bad)	N (Bad)
	3	Y	N (Good)	N (Good)	Y
	4	N (Good)	Y	Y	N (Bad)
	5	N (Good)	Y	N (Bad)	Y
	6	Y	N (Good)	Y	N (Good)
III (3 unknown, 1 known)	1	N (Good)	N (Good)	N (Bad)	Y
	2	N (Bad)	N (Good)	Y	N (Bad)
	3	N (Good)	Y	N (Bad)	N (Bad)
	4	Y	N (Good)	N (Bad)	N (Bad)
IV (4 unknown)	1	N (Bad)	N (Good)	N (Bad)	N (Bad)

Table 5.10: Library searching approach results summary. Y: known; N: Unknown

spectively. In this problem, the cases with mean square root estimation errors less than $0.1mJ/cm^2$ for exposure, $0.02\mu m$ for focus, $0.2sec$ for PEB time and 0.1^oC for PEB temperature are "good". Otherwise, the results are considered "bad".

These results serve dual purposes. On one hand, they illustrate the performance of the estimation approach. On the other hand, they provide us the important information that in order to use this approach to estimate certain parameter(s), what other parameters need to be measured. For example, according to our results, in order to estimate PEB time accurately, we need to be able to measure PEB temperature and exposure dose.

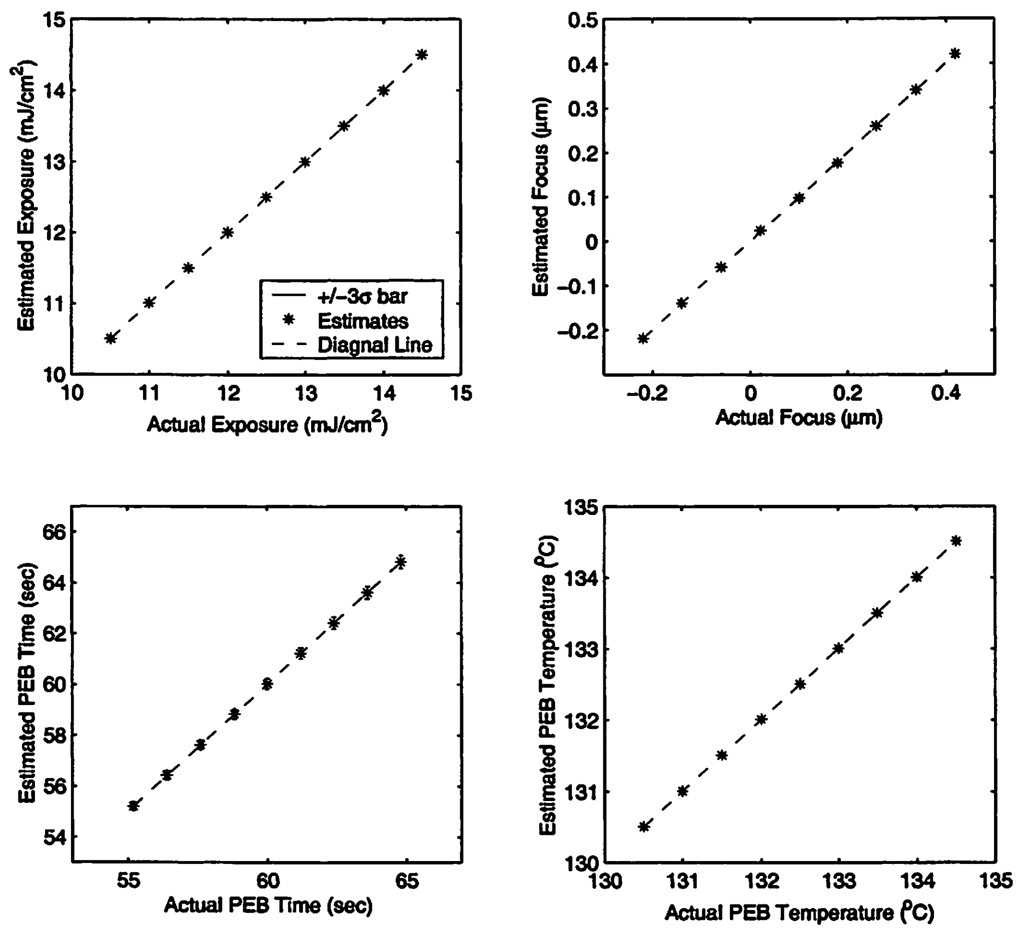

Figure 5.37: Library searching approach case I #1~#4 results: the estimated parameter is the only unknown.

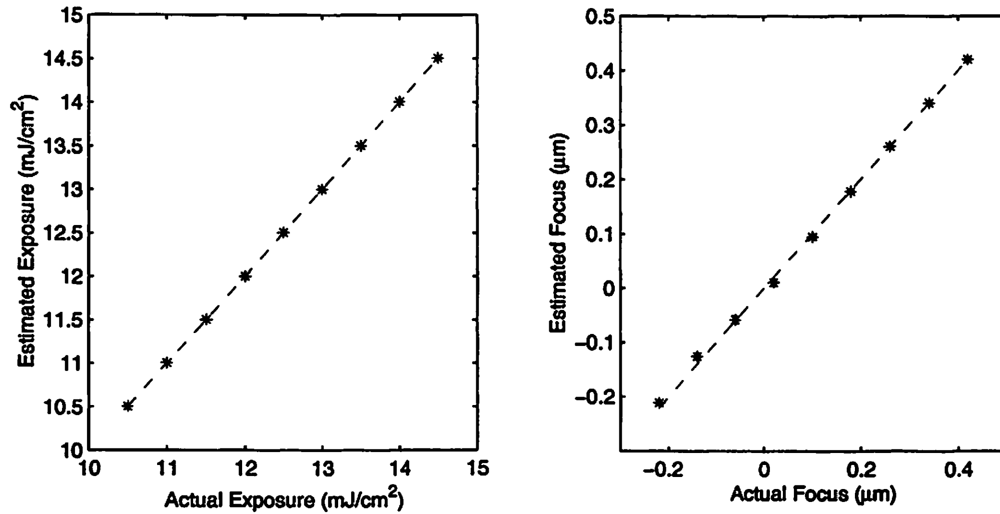

Figure 5.38: Library searching approach case II #1 results: exposure, focus unknown; PEB time and temperature known.

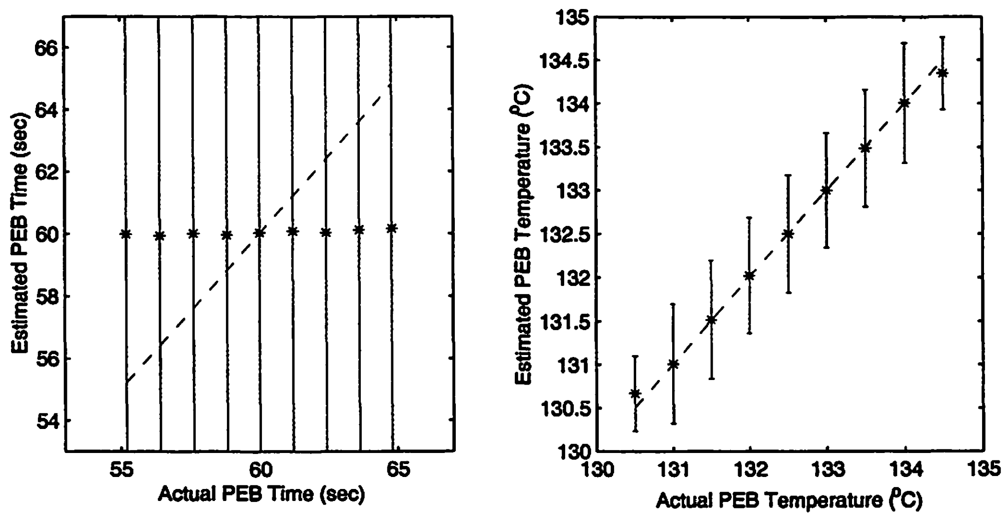

Figure 5.39: Library searching approach case II #2 results: PEB time and temperature unknown; exposure, focus known.

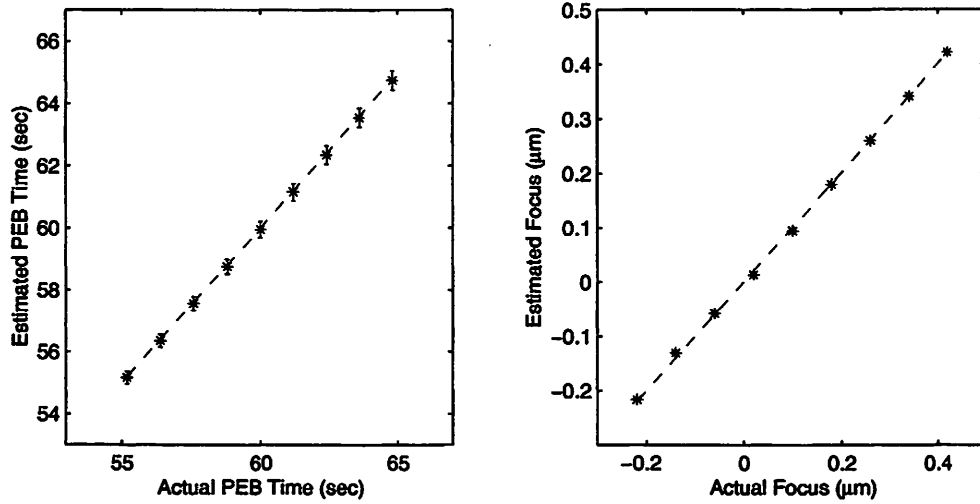

Figure 5.40: Library searching approach case II #3 results: PEB time, focus unknown; PEB temperature, exposure known.

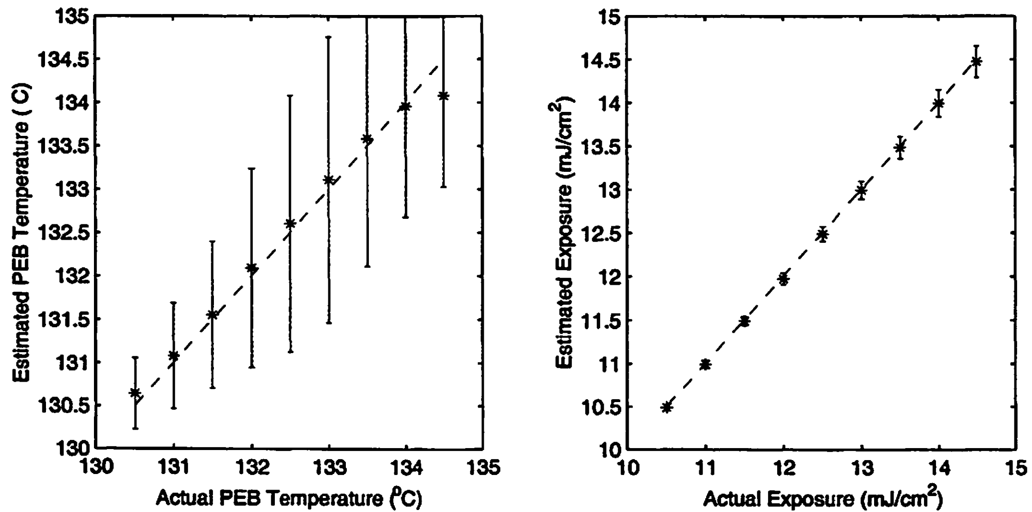

Figure 5.41: Library searching approach case II #4 results: PEB temperature, exposure unknown; PEB time, focus known.

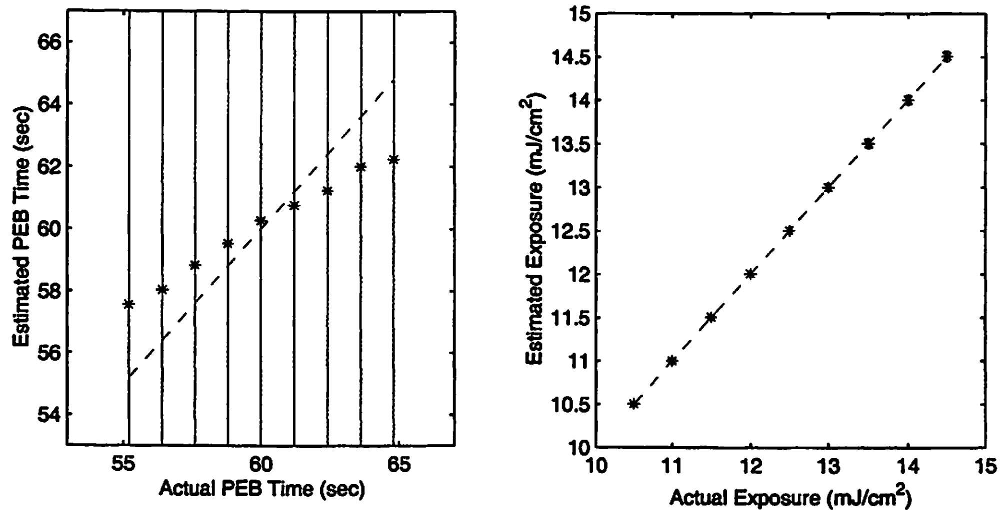

Figure 5.42: Library searching approach case II #5 results: PEB time, exposure unknown; PEB temperature, focus known.

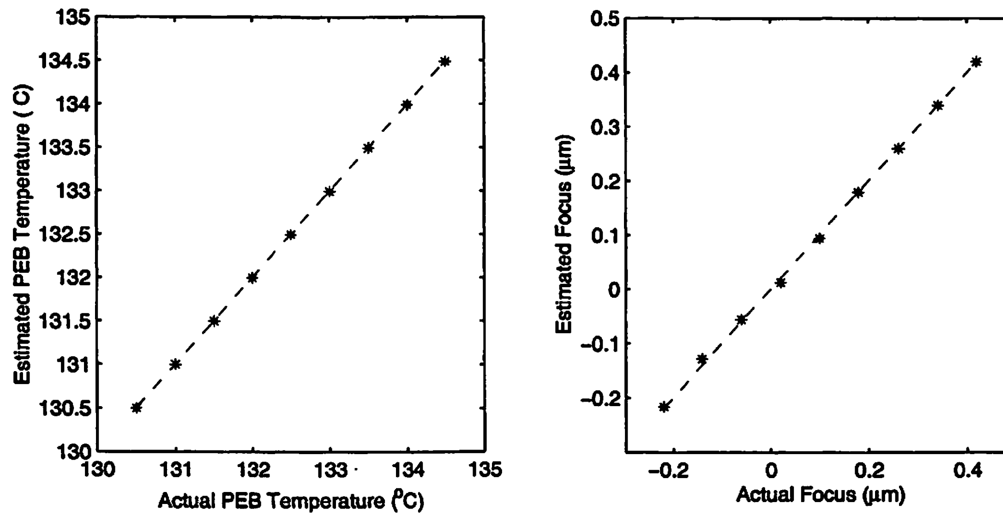

Figure 5.43: Library searching approach case II #6 results: PEB temperature, focus unknown; PEB time, exposure known.

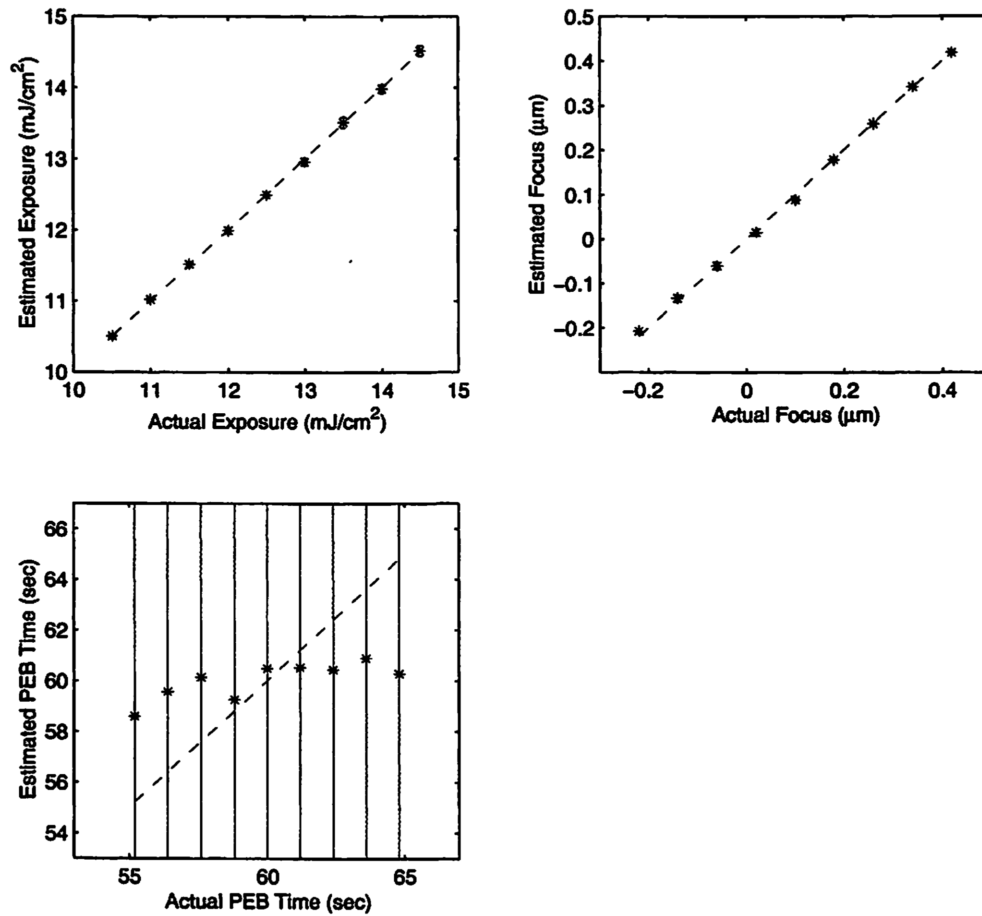

Figure 5.44: Library searching approach case III #1 results: exposure, focus, PEB time unknown; PEB temperature known.

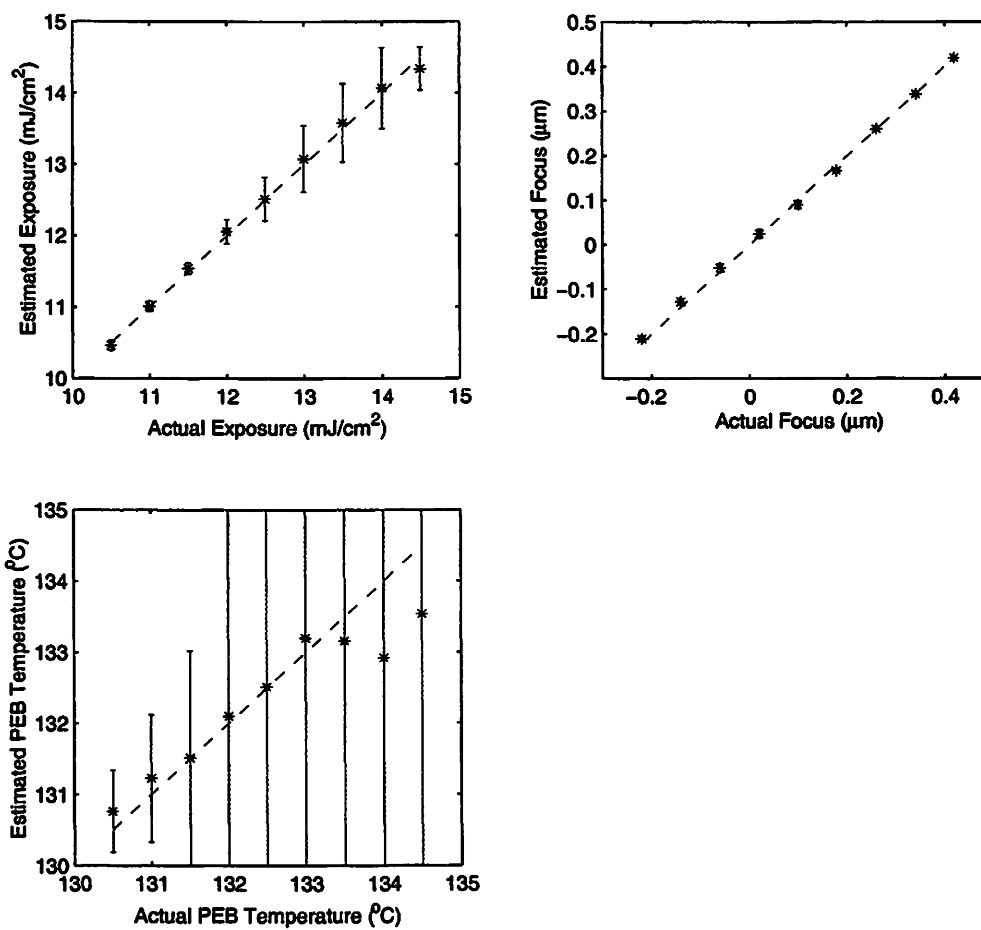

Figure 5.45: Library searching approach case III #2 results: exposure, focus, PEB temperature unknown; PEB time known.

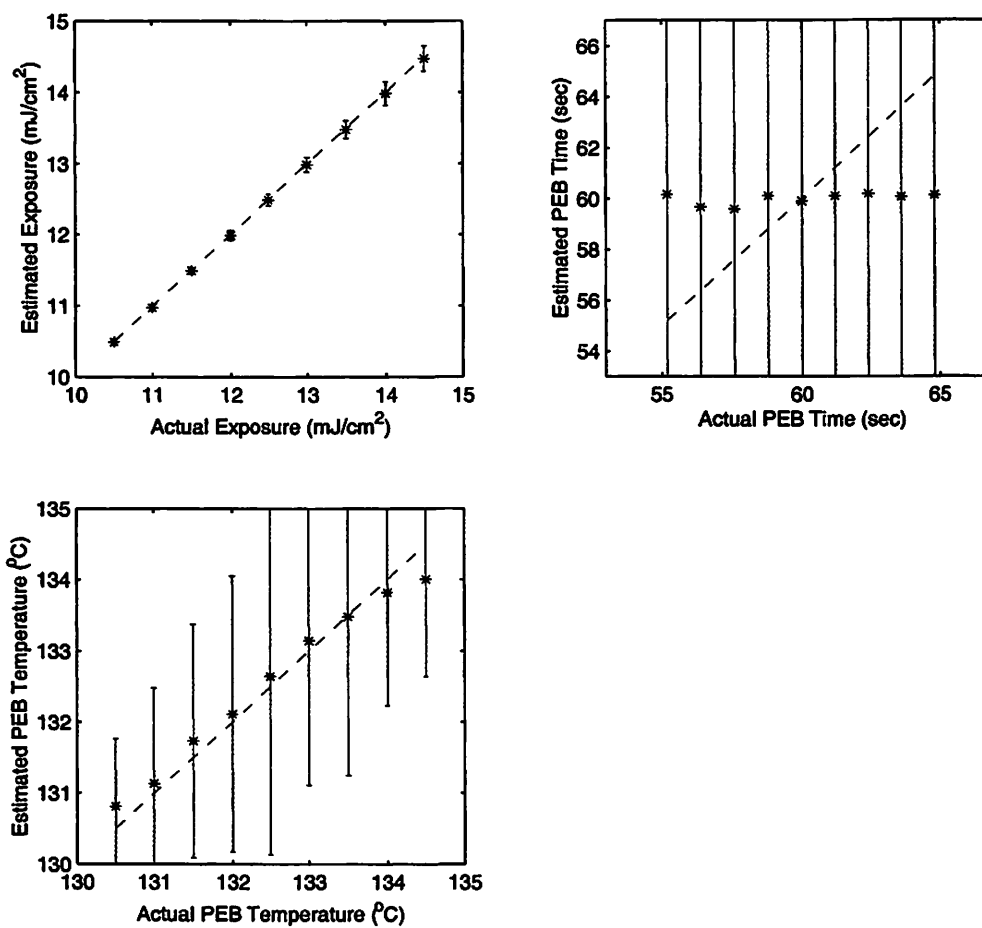

Figure 5.46: Library searching approach case III #3 results: exposure, PEB time and temperature unknown; focus known.

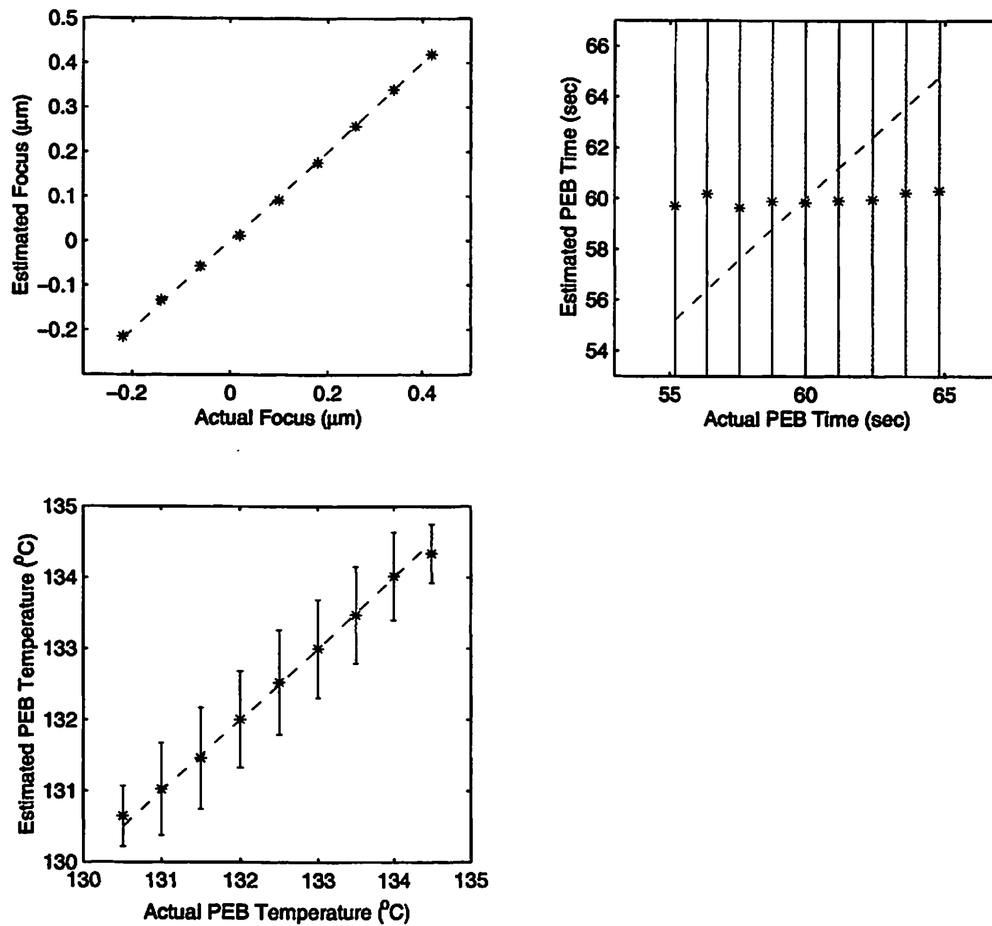

Figure 5.47: Library searching approach case III #4 results: focus, PEB time and temperature unknown; exposure known.

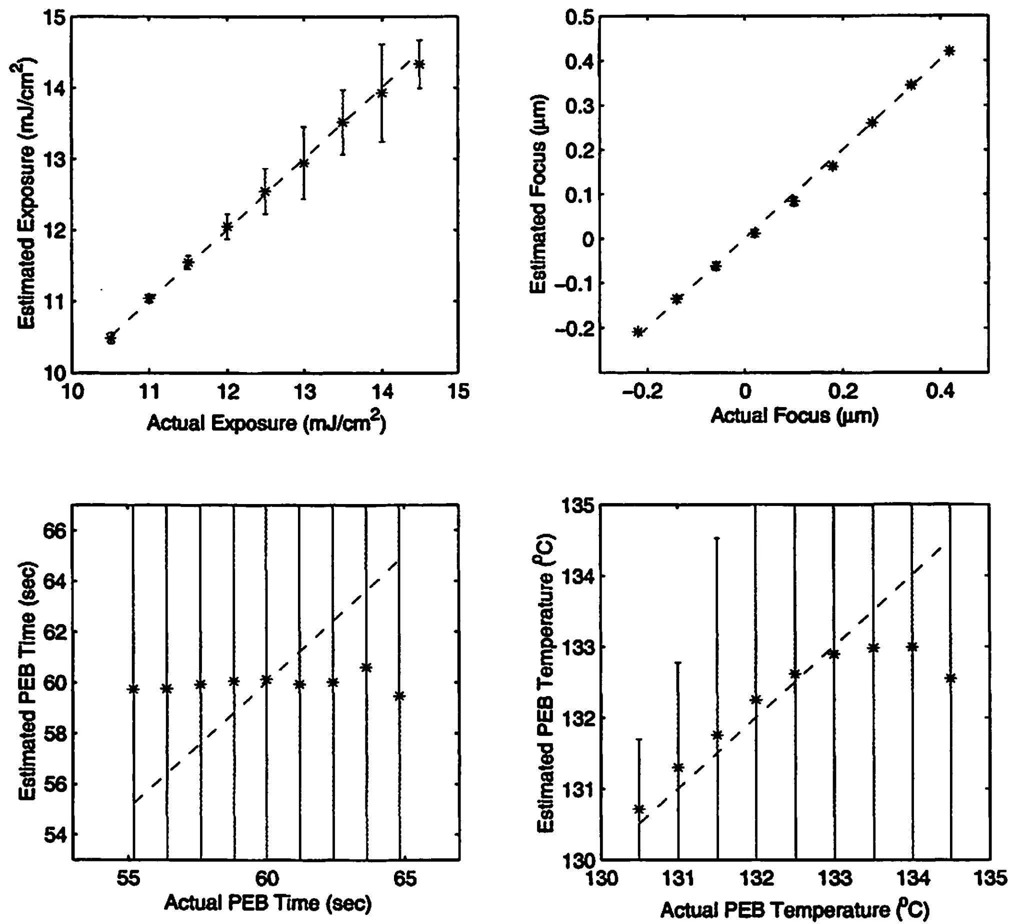

Figure 5.48: Library searching approach case IV results: exposure, focus, PEB time and temperature all unknown.

5.5 Estimation of the Time-Series Model of Input Variations

Estimates of input variables at any particular time, whether measured directly or inferred from output variables through the diagnostic techniques that we have been discussing, can almost never be noise-free. One common approach to alleviate the adverse effect of noise is to postulate a stochastic model of the variables of interest and noises endowed with appropriate probability distributions. Computing the expectation of the true parameters based on their estimates and the stochastic model is likely to improve the estimation accuracy. In a photolithography process, the estimates of input parameters, when taken sequentially in time, can be considered as time series. In this section we consider the application of the stochastic time-series modeling techniques [14, 17] to photolithography processes and how this relates to our statistical process modeling.

First, we describe in the simple cases where the input variables are following a known time-series model, how we could make use of the diagnostic results to obtain more accurate estimation and prediction of the input parameters. Next, we consider the cases where the time-series model structure is known but the model parameters have to be estimated. Finally, we consider the most difficult cases where both the model structure and parameters are unknown. In particular, we make assumptions on only the general family of model structures and study how inaccurate model structures affect estimation and prediction. In all these cases, we use the following two conditions as examples: first, the focus is the only unknown drifting parameter; second, both the exposure and focus are unknown and drifting. Under both conditions, the diagnostic results from the library searching technique

are used as the noisy measurements of the input parameters.

Autoregressive and Moving Average models [14, 39, 70] are two commonly used basic time-series models. Their mixture, the Autoregressive Moving Average model, is a typical time-series model used to represent linear stationary stochastic processes, denoted by ARMA(p,q) where p is the order of autoregression and q is the order of moving average. In practice, many time series do not have fixed means. This non-stationary behavior can be approximated by assuming that some difference of the process is stationary, which can be captured by an ARMA model. The original process can then be represented by an Autoregressive Integrated Moving Average model ARIMA(p, d, q) where d is the order of difference (cf. Eq 5.3).

$$y(t) = \sum_{i=1}^{p} a_i y(t-i) + e(t) + \sum_{j=1}^{q} b_j e(t-j) \tag{5.2}$$

$$y(t) = \nabla^d z(t) \tag{5.3}$$

where ∇ is the difference operator and ∇^d means the dth order difference. For example, when $d = 1$, $y(t) = z(t) - z(t-1)$; when $d = 2$, $y(t) = [z(t) - z(t-1)] - [z(t-1) - z(t-2)]$; and so on. It is typically assumed that $\{e(t)\} \sim \mathcal{N}(0, \sigma_e^2)$ for all $t > 1$ (IIND).

In this section, we focus on the ARIMA models since they are appropriate for modeling input parameter drifts in lithography processes, and for simplicity, we choose ARIMA(1,1,1) for all our examples, which has the form as in Eq 5.5.

$$y(t) = ay(t-1) + e(t) - be(t-1) \tag{5.4}$$

$$y(t) = z(t) - z(t-1) \tag{5.5}$$

where $\{e(t)\} \sim \mathcal{N}(0, \sigma_e^2)$ for all $t > 1$ (IIND). $z(t)$ could be either focus or exposure dose.

In this section, our chief concern is using time series models to do prediction and model estimation, but not control. Thus, we omit terms related to controls in the model. Interested readers can refer to the book by Kumar and Varaiya [70] for control algorithms.

5.5.1 Prediction and Estimation

At each time step (i.e., at each time when a new observation of CD outputs is obtained), the input parameters estimated by either of the diagnostic approaches are treated as a measurement for the ARIMA model. Since the library-searching approach gives more accurate results, we use it in all our simulations in this section. When the time series model of a drifting parameter is known, we can use the model and the current-step measurement to perform current-step estimation and next-step prediction. This can be performed using a Kalman filter approach [47]. To use the Kalman filter, we first need to transfer Eq 5.5 to its state space formulation as shown in Eq 5.7.

$$x(t+1) = Ax(t) + Gv(t) \tag{5.6}$$

$$y(t) = Cx(t) + w(t) \tag{5.7}$$

where $x(t) = [x_1(t), x_2(t)]^T$, $A = \begin{bmatrix} 0 & 1 \\ 0 & a \end{bmatrix}$, $G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $v(t) = [e(t), (a-b)e(t)]^T$, $C = [1, 0]$, $\{w(t)\} \sim \mathcal{N}(0, \sigma_w^2)$ for all $t > 1$ (IIND).

The Kalman filter equations are as follow [47]:

$$\begin{aligned} \hat{x}(t \mid t) &= \hat{x}(t-1 \mid t) \\ &\quad + P(t \mid t-1)C^T[CP(t \mid t-1)C^T + \sigma_w^2]^{-1}[y(t) - C\hat{x}(t \mid t-1)] \end{aligned} \tag{5.8}$$

$$P(t \mid t) = P(t \mid t-1)$$

$$-P(t \mid t-1)C^T[CP(t \mid t-1)C^T + \sigma_w^2]^{-1}CP(t \mid t-1) \quad (5.9)$$

$$\hat{x}(t+1 \mid t) = A\hat{x}(t \mid t) \quad (5.10)$$

$$P(t+1 \mid t) = AP(t \mid t)A^T + GQG^T \quad (5.11)$$

where Q is the covariance matrix of $v(t)$. The initial state of the recursions are $\hat{x}(0) = x_0$ and $P(0 \mid -1) = P_0$. In our simulation, we assume that the initial state is known. $\hat{x}(t \mid t)$ (Eq 5.8) is the estimation and $\hat{x}(t+1 \mid t)$ (Eq 5.10) is the one-step prediction. $P(t \mid t)$ and $P(t+1 \mid t)$ are the covariance estimates.

Figure 5.49 shows the Kalman filter estimation and one-step prediction results for focus when it is the only unknown parameter. Figure 5.50 shows the results for the case when both focus and exposure are unknown. We can see that the Kalman filter estimation and prediction are good when the measurements (estimated by the library-searching approaches) are accurate.

5.5.2 Time Series Model Estimation

When the time series model structure is known but with some unknown parameters, we need to estimate the model parameters from measurements, which are in turn estimated by the library-searching approaches presented earlier. Standard time series model parameter estimation methods can be found in many time series textbooks [14, 39, 70]. In this section, we assume that parameters a and b in Eq 5.5 are both unknown. Their estimations can be calculated from Eq 5.12 and Eq 5.13 [39].

$$\hat{a}(t) = \frac{\sum_{k=1}^{t-1}[y(k) - \bar{y}(t)][y(k+1) - \bar{y}(t)]}{\sum_{k=1}^{t}[y(k) - \bar{y}(t)]^2} \quad (5.12)$$

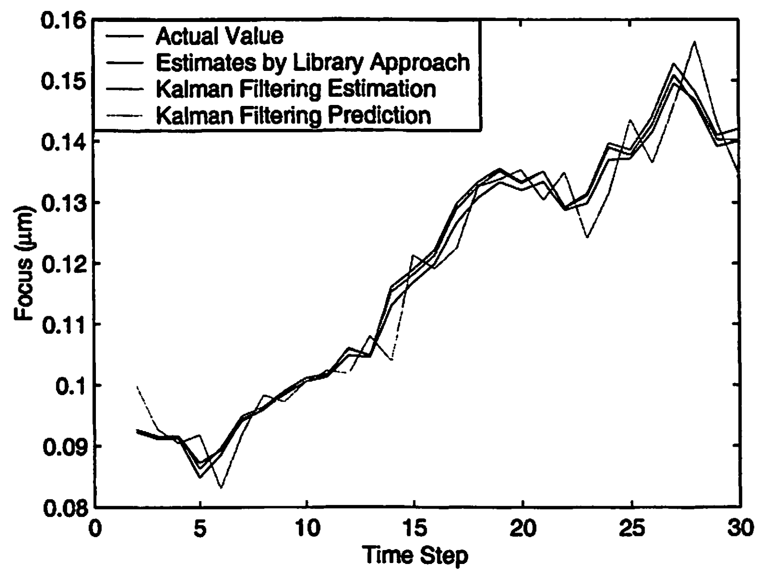

Figure 5.49: Time series simulation: estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$.

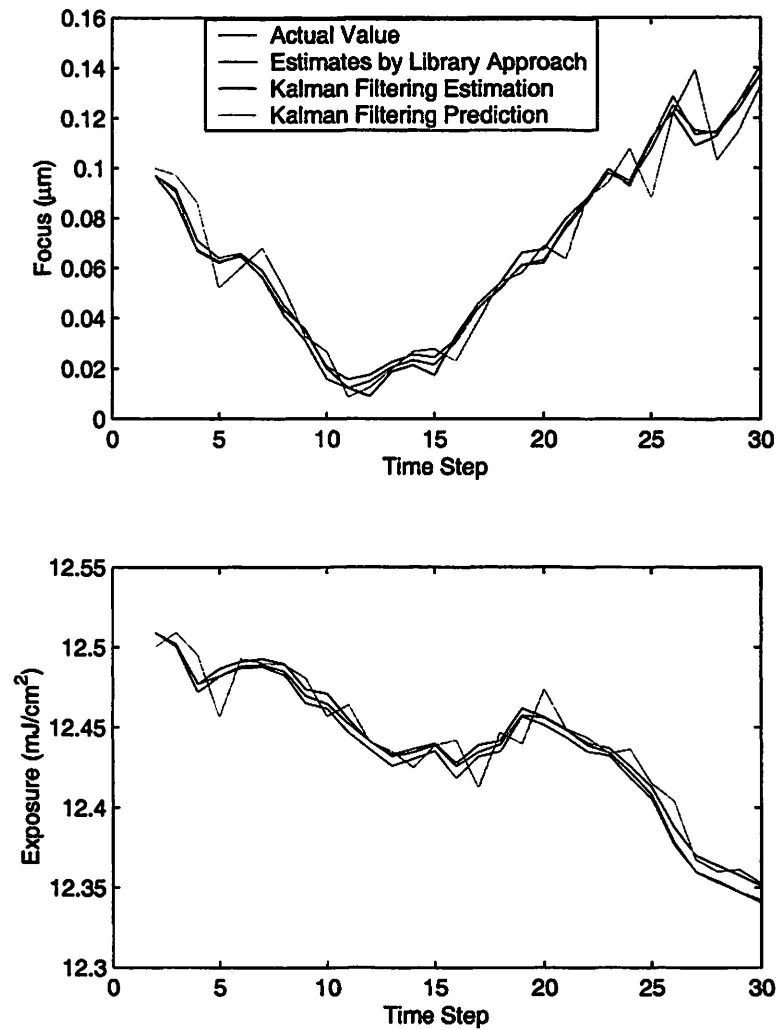

Figure 5.50: Time series simulation: estimation and prediction using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model for focus: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$; actual model for exposure: ARIMA(1,1,1), $a = 0.6$, $b = 0.3$.

$$\hat{b}(t) = \begin{cases} [2\hat{a}(t)]^{-1}[1 - \sqrt{1 - 4\hat{a}^2(t)}], & 0 <\mid \hat{a}(t) \mid \leq 0.5 \\ -1, & \hat{a}(t) < -0.5 \\ 1, & \hat{a}(t) > 0.5 \\ 0, & \hat{a}(t) = 0 \end{cases} \tag{5.13}$$

where $\bar{y}(t) = (1/t)\sum_{i=1}^{t} y(i)$.

Figure 5.51 shows the model parameter estimation at each time step, as well as the estimation and one-step prediction using Kalman filtering based on the estimated time series model, when focus is the only unknown parameter. Figure 5.52 and figure 5.53 show the results of focus and exposure when both are unknown. From these figures, we can see that it takes some time for the estimates of parameters to converge to the true values. It is also shown that even if the model parameters a and b have not yet fully converged, a Kalman filter can still produce reasonable accuracy on estimation and prediction.

5.5.3 Model Mismatch Problem

In practice, we often face the problem that some parameters are drifting but it is unclear what a correct time series model structure should be. In other words, the model is completely unknown. In these cases, we have to make assumptions on the model structure. In most occasions, simple models are adopted for computational convenience. However, if we make incorrect assumptions on the model structure, would it be a big problem? In this section, we run simulations using a higher-order ARIMA model (ARIMA(3,1,2)) in conjunction with the library-searching approach to obtain a set of data, which is then used to estimate model parameters of a lower-order ARIMA model (ARIMA(1,1,1)). Figure 5.54 gives the simulation results for the case when focus is the only drifting parameter. Figure 5.55 and figure 5.56

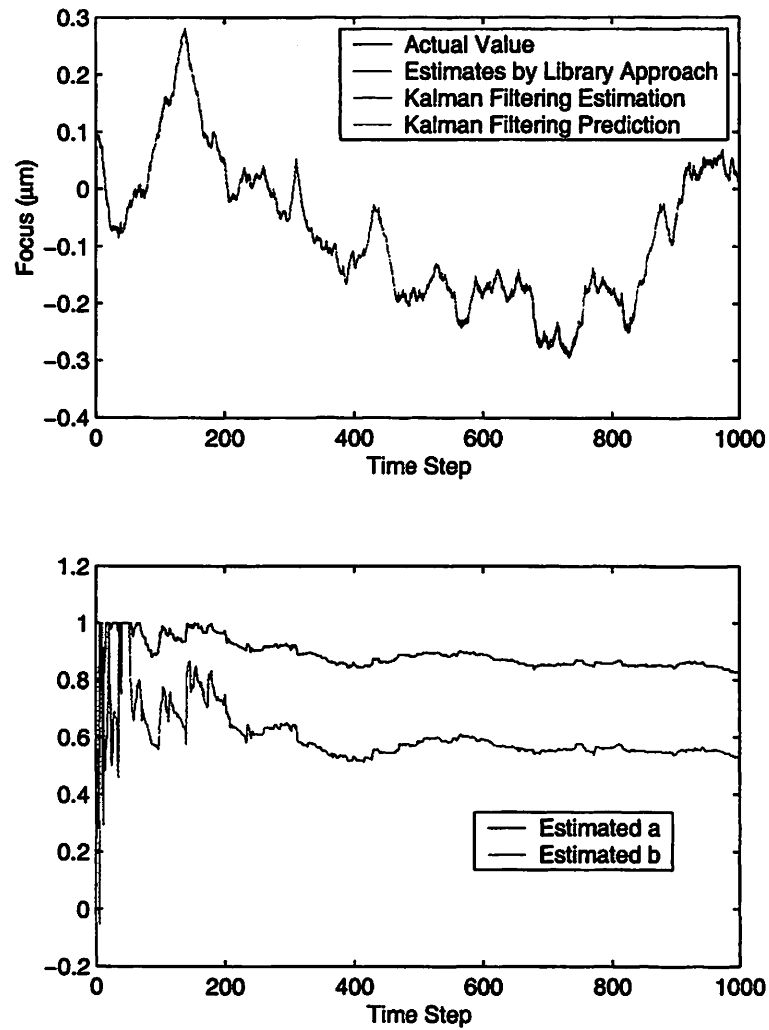

Figure 5.51: Time series simulation: model parameter estimation, estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$.

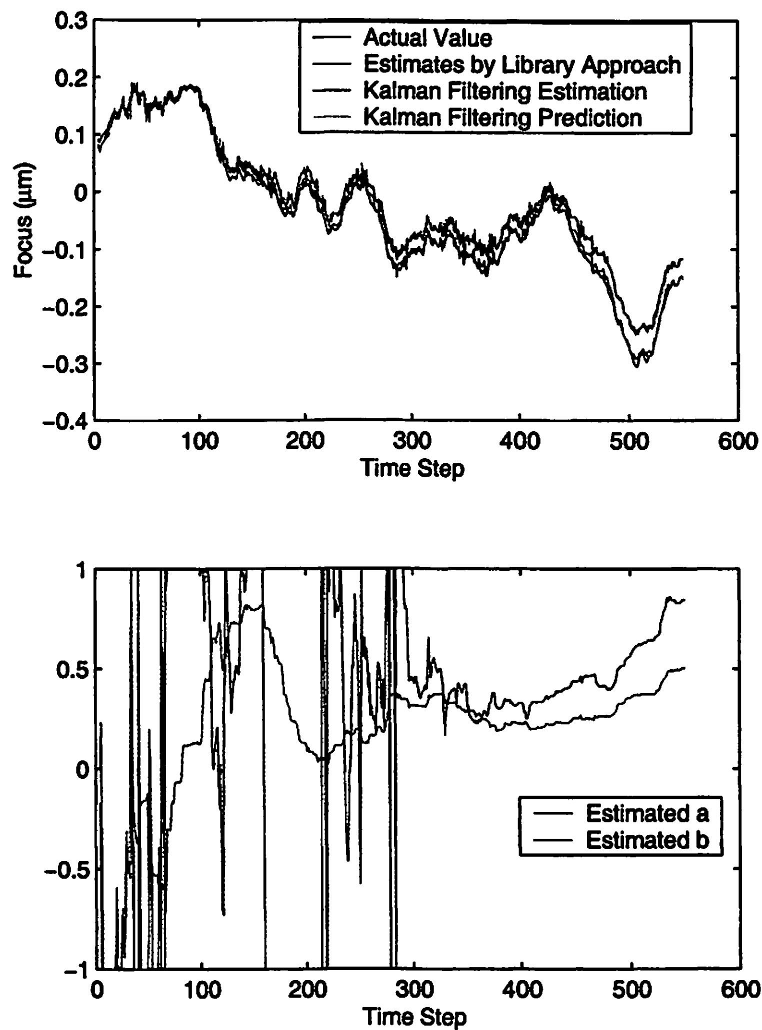

Figure 5.52: Time series simulation: model parameter estimation, estimation and prediction of focus using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(1,1,1), $a = 0.8$, $b = 0.5$.

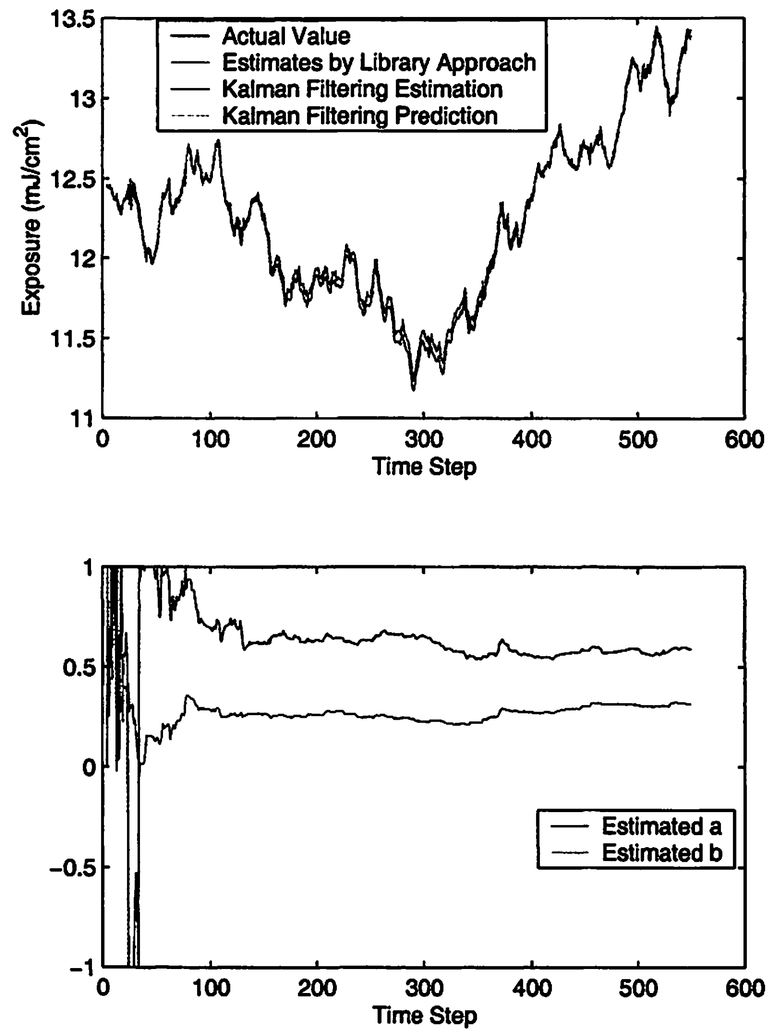

Figure 5.53: Time series simulation: model parameter estimation, estimation and prediction of exposure using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(1,1,1), $a = 0.6$, $b = 0.3$.

Cases	Actual Model	Estimated Model	Root Mean Square Error (μm) Measurement	KF Estimation	KF Prediction
1. Time Series Model is known	ARIMA (1,1,1) $a = 0.8$ $b = 0.5$	NA	0.0007	0.0008	0.0013
2. Time Series Model parameters are unknown	ARIMA (1,1,1) $a = 0.8$ $b = 0.5$	ARIMA (1,1,1)	0.0007	0.0013	0.0021
3. Time Series Model is unknown	ARIMA (3,1,2) $a = [0, 0.12, -0.016]$ $b = [0, 0.01]$	ARIMA (1,1,1)	0.0008	0.0015	0.0019

Table 5.11: Time series simulation summary when focus is the only unknown parameter.

show the case that both focus and exposure are drifting. For each of the simulation results, the relationship between the estimated model parameters a and b and the actual higher-order model parameters is not obvious. The fact that the estimation of a and b do not converge very well as shown in these figures could be an indication that the approximation does not fit the actual model.Fortunately, this does not fundamentally affect the performance of the Kalman filter. Therefore we may conclude that model mismatch is not necessarily a big problem for estimation and prediction by the Kalman filter approach. Of course, it remains to be investigated whether this is still true when feedback control is considered, but it is not our focus in this thesis.

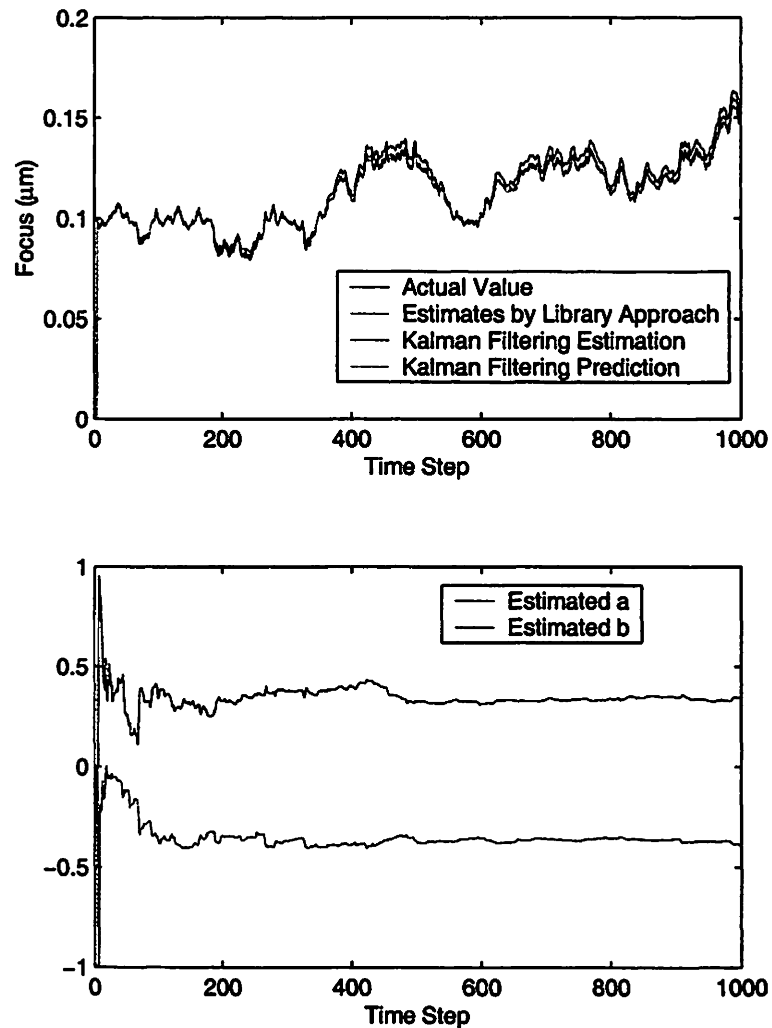

Figure 5.54: Time series simulation: mismatched model approximation, estimation and prediction using Kalman filtering, when focus is the only unknown parameter. Actual model: ARIMA(3,1,2), $a = [0, 0.12, 0.016]$, $b = [0, 0.01]$. Approximate model: ARIMA(1,1,1).

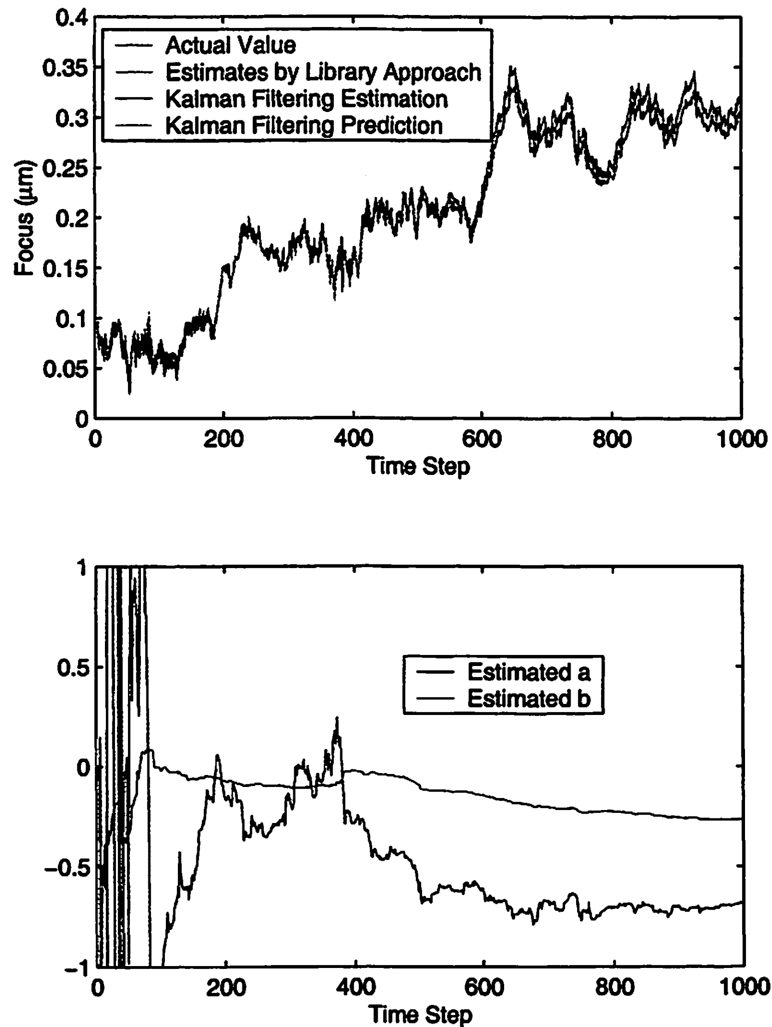

Figure 5.55: Time series simulation: mismatched model approximation, estimation and prediction of focus using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(3,1,2), $a = [0, 0.12, 0.016]$, $b = [0, 0.01]$. Approximate model: ARIMA(1,1,1).

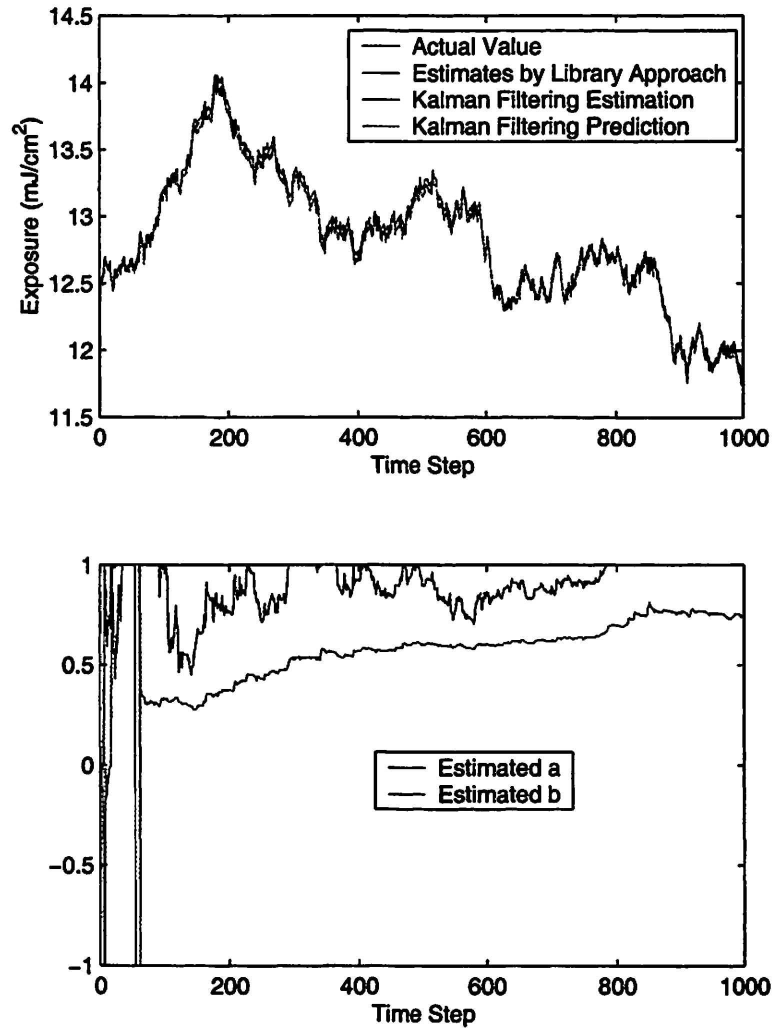

Figure 5.56: Time series simulation: mismatched model approximation, estimation and prediction of exposure using Kalman filtering, when focus and exposure are the only unknown parameters. Actual model: ARIMA(3,1,2), $a = [0, 0.07, 0.006]$, $b = [0.1, 0.02]$. Approximate model: ARIMA(1,1,1).

Cases	Actual Model	Estimated Model	Root Mean Square Error (μm)		
			Measure-ment	KF Estimation	KF Prediction
1. Time Series Model is known	ARIMA (1,1,1) $a = 0.8$ $b = 0.5$	NA	0.0021	0.0019	0.0028
2. Time Series Model parameters are unknown	ARIMA (1,1,1) $a = 0.8$ $b = 0.5$	ARIMA (1,1,1)	0.0023	0.0039	0.0044
3. Time Series Model is unknown	ARIMA (3,1,2) $a = [0, 0.12, -0.016]$ $b = [0, 0.01]$	ARIMA (1,1,1)	0.0018	0.0062	0.0068

Table 5.12: Time series simulation summary for focus when focus and exposure are the unknown parameters.

Cases	Actual Model	Estimated Model	Root Mean Square Error (mJ/cm^2)		
			Measure-ment	KF Estimation	KF Prediction
1. Time Series Model is known	ARIMA (1,1,1) $a = 0.6$ $b = 0.3$	NA	0.0010	0.0011	0.0025
2. Time Series Model parameters are unknown	ARIMA (1,1,1) $a = 0.6$ $b = 0.3$	ARIMA (1,1,1)	0.0012	0.0024	0.0034
3. Time Series Model is unknown	ARIMA (3,1,2) $a = [0, 0.07, 0.006]$ $b = [0.1, 0.02]$	ARIMA (1,1,1)	0.0009	0.0031	0.0038

Table 5.13: Time series simulation summary for exposure when focus and exposure are the unknown parameters.

5.6 Summary

Using a photolithography process as an example, this chapter shows two empirical approaches for process input parameter estimation. The first approach is to develop either a universal explicit inverse model or a set of inverse models, each for a segment of the input parameter space. Multi-layer Perceptron Neural Networks are adopted for all inverse models. An accurate universal inverse model is often difficult to obtain due to the high complexity of the problem. Using separate models in different segments shows potential improvement, and the performance of this approach largely depends on whether the parameter space can be segmented so that in each segment, parameters are effectively decoupled. The method described in this chapter considers focus as the only dimension for segmentation. More complex segmentation may be required to obtain better estimation results.

The second approach is to generate large samples of input-output pairs and perform input-output matching based on accurate forward models using neural networks. This library-searching approach can be applied to any parameter sets. In general, this approach performs better than the explicit inverse approach on problems of high complexity. Results of this approach can also be used to delineate what parameters must be measured in order to obtain accurate estimations of some other parameters.

Finally, we also demonstrate how to improve parameter estimation by considering parameter drifts as time series. Integrated autoregressive moving average models are simulated for focus and exposure. We use Kalman filter techniques to perform estimation and prediction when parameter estimates obtained from the

library searching approach are taken as the time-series measurements. We also discuss time series model parameter estimation and the potential problem of model structure mismatching.

Chapter 6

Conclusions and Future Work

6.1 Summary and Conclusions

Semiconductor manufacturing is stepping into the new era of extremely fine features and the control of variations has become more critical than it was ever before. Due to the highly complex nature of most semiconductor manufacturing equipment and processes, variations can come from a large variety of sources and the diagnostics of the variations are often very difficult. In this thesis, we have developed advanced statistical techniques to perform efficient and accurate diagnostics for semiconductor manufacturing, demonstrated on two important systems with different characteristics.

For a furnace system, we developed an approximate physical model. System parameters were estimated by a least-squares approach using experimental data. A model classification approach was investigated for estimating faulty parameters and a sensor fusion technique was discussed for enhancing diagnostic reliability.

For a lithography process, a more sophisticated diagnostic framework was pro-

posed. Instead of using experimental data, we suggested to use simulation data (obtained from a fine-tuned software simulator) for process modeling. Different parameter estimation methods were investigated, including the explicit inverse model approach and the library searching approach based on forward modeling. Their performances were compared empirically on the simulation results. The library searching approach performed very well in general and showed many advantages over the explicit inverse model approach. In particular, the results of the library searching approach could give us useful recommendations on what parameters to measure, what to estimate, and the required precision. The results of the explicit inverse model approach were less ideal in this work, mainly due to the lack of an effective method for space segmentation. Time series models were considered for modeling parameter drifts. We showed that accurate estimation and prediction could be obtained with Kalman filtering using an ARIMA model, in conjunction with the library searching approach for generating measurements. We also studied the estimation of the ARIMA model parameters and discussed the problem of model mismatch.

6.2 Future Work

For the furnace system, an interesting future research direction is to refine the model. We used a linear dynamic model in our analysis, but observed some non-linearity in the experiments. For example, we observed a slight variation of the thermal insulation parameters as a function of the temperature. If the non-linearity is taken into account in system modeling, the diagnostics may become more accurate. In addition, some fault detection methods for nonlinear dynamic

systems [36, 106] can be considered.

In testing the diagnostics methodologies on lithography processes, one weakness of the current work is that only simulated data but not true experimental data were used. If access to real lithography processes were available, it is desired to verify the effectiveness of the proposed approach on true experimental data, with the following procedure:

1. Design and run experiments, and collect data;
2. Divide the data into a training set and a test set;
3. Tune PROLITH using only the training set data;
4. Run simulation in PROLITH to collect data for process modeling;
5. Train neural network models for forward modeling using simulated data;
6. Run Monte Carlo simulation using trained models to set up a data library;
7. Perform diagnostics (i.e., estimate input parameters) on the test set data;
8. Compare estimates with actual experimental input parameter values for the test set data.

Research can also be expanded in a number of other directions for the lithography processes. For the explicit inverse model approach, we described finding local inverse models by performing space segmentation with an ad-hoc method. However, we recognize that more systematic methods for space segmentation need to be developed. For the library searching approach, the size of the library may become very large as the number of parameters grows, and, thus, an efficient and

accurate search algorithm needs to be developed. When parameter drifts are considered as time series, control can be performed using existing stochastic control algorithms [70]. Finally, besides neural networks, other modeling methods, such as those described in Chapter 2, can be applied.

In general, the diagnostic approaches developed in this thesis can be applied to other processes. For example, the model estimation and fault classification methodologies for the furnace problem can be applied to any system that has a relative simple mathematical model based on known physics. The techniques that we have developed for lithography processes can also be applied to other processes, such as plasma etching and the CMP processes. For plasma etching, no mathematical model based on physics has been fully developed, so we are not able to collect data through a simulator. Fortunately, experimental data of plasma etching is much easier to obtain than that of lithography processes so that a simulator may not even be necessary. However, since plasma etching processes are even more complex, techniques of dimension reduction might be needed to simplify diagnostics.

Bibliography

[1] The national technology roadmap for semiconductors: Technology needs. Semiconductor Industry Association, 2000 Update.

[2] A. M. Agogino, S. Alag, and K. Goebel. A framework for intelligent sensor validation, sensor fusion, and supervisory control of automated vehicles in IVHS. *Proceedings of the ITS America Annual Meeting, pp. 77-87*, 1995.

[3] A. M. Agogino, K. Goebel, and S. Alag. Intelligent sensor validation and sensor fusion for reliability and safety enhancement in vehicle control. Technical report, MOU132, California PATH Research Report, 1995.

[4] S. Alag. *A Bayesian Decision-Theoretic Framework for Real-time Monitoring and Diagnosis of Complex Systems: Theory and Application.* PhD thesis, University of California, Berkeley, 1996.

[5] S. Alag, K. Goebel, and A. M. Agogino. A methodology for intelligent sensor validation and fusion used in tracking and avoidance of objects for automated vehicles. *Proceedings of the ACC, pp. 3647-3653*, 1995.

[6] L. B. Almeida and C. J. Wellekens, editors. *Neural Networks*, volume 412 of *Lecture Indexs in Computer Science.* Springer Verlag, February 1990.

[7] D. R. Anderson, D. R. Sweeney, and T. A. Williams. *Statistics: Concepts and Applications.* West Publishing Company, 1986.

[8] M. Ayoubi. Fuzzy systems design based on a hybrid neural structure and application to the fault diagnosis of technical processes. *Control Engineering Practice, Vol. 4, No. 1, pp. 35-42*, 1996.

[9] D. H. Ballard. *An Introduction to Natural Computation.* The MIT Press, 1997.

[10] D. M. Bates and D. G. Watts. *Nonlinear Regression Analysis and its Applications.* John Wiley & Sons, 1988.

[11] Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[12] George Box and Alberto Luceno. *Statistical Control, By Monitoring and Feedback Adjustment.* John Wiley & Sons, 1997.

[13] George E. P. Box, William G. Hunter, and J. Stuart Hunter. *Statistics of Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley & Sons, 1978.

[14] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control, Revised Edition.* Holden-Day, 1976.

[15] L. Cai, A. Rohatgi, S. Han, and G. May. Investigation of the properties of plasma-enhanced chemical vapor deposited silicon nitride and its effect on silicon surface passivation. *Journal of Applied Physics, Vol. 83, No. 11, pp. 5885-5889*, 1998.

[16] Gino Cardarelli, Mario Palumbo, and Pacifico M. Pelagagge. Use of neural networks in modeling relations between exposure energy and pattern dimension in photolithography process. *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Vol. 19, No. 4, pp. 290-299*, October 1996.

[17] Chris Chatfield. *Time-Series Forecasting.* Chapman and Hall/CRC, 2001.

[18] E. Y. Chow and A.S. Willsky. Analytical redundancy and the design of robust detection systems. *IEEE Transactions on Automatic Control, AC-29, 7, pp. 603-614*, 1984.

[19] A. Cichoki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing.* John Wiley and Sons, Chichester, 1993.

[20] R. N. Clark. Instrument fault detection. *IEEE Transactions on Aerospace and Electronic Systems, pp. 456-465, May*, 1978.

[21] N. E. Collins, R. W. Eglese, and B. L. Golden. Simulated annealing - an annotated bibliography. *American Journal of Math. And Management Sci., Vol. 8, No. 3-4, pp. 209-307*, 1988.

[22] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications, 1991.

[23] E. Cox. *The Fuzzy Systems Handbook.* Academic Press, Inc., Cambridge, MA, USA, 1994.

[24] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines.* Cambridge University Press, Cambridge, UK, 2000.

[25] B. Dasarthy, editor. *Nearest Neighbor Pattern Classification Techniques.* IEEE Computer Society Press, 1990.

[26] John J. D'Azzo. *Linear Control System Analysis and Design: Conventional and Modern.* McGraw-Hill, 1988.

[27] F. H. Dill. Optical lithography. *IEEE Trans. Electron Devices, Vol. ED-22, No. 7, pp. 440-444*, 1975.

[28] F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw. Characterization of positive photoresist. *IEEE Trans. Electron Devices, Vol. ED-22, No. 7, pp. 445-452*, 1975.

[29] F. H. Dill, A. R. Neureuther, J. A. Tuttle, and E. J. Walker. Modeling projection printing of positive photoresists. *IEEE Trans. Electron Devices, Vol. ED-22, No. 7, pp. 456-464*, 1975.

[30] R. J. M. M. Does, K. C. B. Roes, and A. Trip. *Statistical process control in industry: implementation and assurance of SPC.* Kluwer Academic, 1999.

[31] A. Doucet and N. de Freitas, editors. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, 2001.

[32] Brad V. Eck, editor. *Advanced Equipment Control/Advanced Process Control Symposium XIII, Banff, Canada*, 2001.

[33] Brian S. Everitt. *An Introduction to Latent Variable Models.* Chapman and Hall, 1984.

[34] Yong-Hui Fan and Taiqing Qiu. Transient heat transfer in batch thermal

reactors for silicon wafer processing. *International Journal on Heat and Mass Transfer, Vol. 41, No. 11, pp. 1549-1557*, 1998.

[35] D. Filbert and K. Metzger. Quality test of systems by parameter estimation. *Proceedings of the 9th IMEKO-Congress, pp. 403-411*, 1982.

[36] P. M. Frank. Advanced fault detection and isolation schemes using nonlinear and robust observers. *Automatic Control - World Congress, 1987. Selected Papers from the 10th Triennial World Congress of the International Federation of Automatic Control, Vol. 3, pp. 63-68*, 1987.

[37] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica, No. 26, pp. 459-474*, 1990.

[38] Normand L. Frigon and David Mathews. *Practical Guide to Experimental Design.* John Wiley & Sons, 1997.

[39] W. A. Fuller. *Introduction to Statistical Time Series, 2nd Edition.* Wiley Series in Probability and Statistics, 1996.

[40] Vivek Garg. *Fault Detection in Nonlinear Systems: An Application to Automated Highway Systems.* PhD thesis, University of California, Berkeley, 1995.

[41] J. Gertler. Survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine, Vol. 8, No. 6, pp. 3-11*, 1988.

[42] K. Goebel and A. M. Agogino. An architecture for fuzzy sensor validation and fusion for vehicle following in automated highways. *Proceedings of the 29th*

International Conference of Automative Technology & Automation (ISATA), pp. 203-209, 1996.

[43] K. Goebel, S. Alag, and A. M. Agogino. Probabilistic and fuzzy methods for sensor validation and fusion in vehicle guidance: A comparison. *Proceedings of the 30th International Conference of Automative Technology & Automation (ISATA), Vol. 1, pp. 711-719*, 1997.

[44] Kai F. Goebel. *Management of Uncertainty in Sensor Validation, Sensor Fusion, and Diagnostics of Mechanical Systems Using Soft Computing Techniques.* PhD thesis, University of California, Berkeley, 1996.

[45] S. Halgamuge and M. Glesner. Neural networks in designing fuzzy systems for real world applications. *Fuzzy Sets and Systems, Vol.65, No. 1, pp. 1-12, July*, 1994.

[46] Seung-Soo Han and Gary S. May. Using neural network process models to perform PECVD silicon dioxide recipe synthesis via genetic algorithms. *IEEE Transactions on Semiconductor Manufacturing, Vol. 10, No. 2, pp. 279-287*, 1997.

[47] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons, Inc., 1996.

[48] D. Heckerman. A tutorial on learning with Bayesian networks. *Microsoft Research Tech. Report, MSR-TR-95-06*, 1996.

[49] D. Heckerman. Bayesian networks for data mining. *Data Mining and Knowledge Discovery, pp. 79-119*, 1997.

[50] D. M. Himmelblau. *Fault Detection and Diagnosis in Chemical and Petrochemical Processes.* Elzevier, Amsterdam, 1978.

[51] T. Hofling. Detection of parameter variations by continuous-time parity equations. *Proceedings of the 12th IFAC World Congress, Sydney, Australia, pp. 513-518*, 1993.

[52] Chang Hsu, R. Young, J. Cheng, P. Chien, V. Wen, and A. R. Neureuther. LAVA: Lithography analysis using virtual access. *Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 3334, pp. 197-201*, 1998.

[53] W. Y. Huang and R. P. Lippmann. Neural net and conventional classifiers. *IEEE Conf. on Neural Information Processing Systems (NIPS), Denver, CO, pp. 42-55*, 1987.

[54] F. P. Incropera and D. P. DeWitt. *Fundamentals of Heat and Mass Transfer.* John Wiley & Sons, Inc., fourth edition, 1997.

[55] R. Isermann. Process fault detection based on modeling and estimation methods: A survey. *Automatica, No. 20, pp. 387-404*, 1984.

[56] R. Isermann. Integration of fault detection and diagnosis methods. *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes - SAFEPROCESS'94, Espoo, Finland, pp. 597-612*, 1994.

[57] Richard C. Jaeger. *Introduction to Microelectronic Fabrication: Modular Series on Solid State Devices*, volume 5. Addison-Wesley, 1988.

[58] N. Jakatdar, J. Bao, and C. J. Spanos. Physical modeling of deprotection

induced thickness loss. *Microlithography, Proceedings of the SPIE, pp. 275-282*, 1999.

[59] N. Jakatdar, X. Niu, and C. J. Spanos. A neural network approach to rapid thin film characterization. *Proceedings Of the SPIE, Vol. 3275, pp. 163-171*, 1998.

[60] Nickhil H. Jakatdar. *Deep Sub-Micro Photolithography Control through In-Line Metrology.* PhD thesis, University of California, Berkeley, 2000.

[61] H. L. Jones. *Failure Detection in Linear Systems.* Department of Aeronautics and Astronautics, M.I.T., Cambridge, MA, USA, 1973.

[62] M. I. Jordan, editor. *Learning in Graphical Models.* Kluwer Academic Press, 1998.

[63] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning, Vol. 37, No. 2, pp. 183-233*, 1999.

[64] R. Joshi and A. C. Sanderson. *Multisensor Fusion: A Minimal Representation Framework.* World Scientific, 1999.

[65] E. W. Kamen and J. K. Su. *Introduction to Optimal Estimation.* Springer, 1999.

[66] J. B. Keats and N. F. Hubele, editors. *Statistical Process Control in Automated Manufacturing.* Dekker, 1989.

[67] Young-Jin Kim. *Uncertainty Propagation in Intelligent Sensor Validation.* PhD thesis, University of California, Berkeley, 1991.

[68] Lawrence A. Klein. *Sensor and Data Fusion Concepts and Applications.* SPIE Press, second edition, 1999.

[69] R. Kruse, J. Gebhardt, and F. Klawonn. *Foundations of Fuzzy Systems.* Wiley and Sons, Chichester, 1994.

[70] P. R. Kumar and Pravin Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control.* Prentice Hall, 1986.

[71] R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and its Applications.* Prentice Hall, Inc., 1981.

[72] S. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[73] S. Leonhardt and M. Ayoubi. Methods of fault diagnosis. *Control Engineering Practice, Vol. 5, No. 5, pp. 683-692*, 1997.

[74] Harry J. Levinson. *Lithography Process Control.* SPIE Press, 1999.

[75] Shin H. Li and Robert O. Miller. *Chemical Mechanical Polishing in Silicon Processing.* Academic Press, 2000.

[76] A. P. Liavas and P. A. Regalia. On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing, Vol. 49, No. 8, pp. 1689-1695*, 2001.

[77] Z. C. Lin and C.-Y. Liu. Application of an adaptive neuro-fuzzy inference system for the optimal analysis of chemical-mechanical polishing process parameters. *International Journal of Advanced Manufacturing Technology, Vol. 18, No. 1, pp. 20-28*, 2001.

[78] Chris A. Mack. *Inside PROLITH: A Comprehensive Guide to Optical Lithography Simulation.* FINLE Technologies, 1997.

[79] H. M. Marchman and N. Dunham. AFM: a valid reference tool? *Proceedings Of the SPIE, Vol. 3332, pp. 2-9,* 1998.

[80] Gary S. May. Manufacturing ICs the neural way. *IEEE Spectrum, pp. 47-51, September,* 1994.

[81] A. Mayazaki, K. Kown, H. Ishibuchi, and H. Tanaka. Fuzzy regression analysis by fuzzy neural networks and its application. *Proceedings of the 3rd IEEE Conf. On Fuzzy Systems, Orlando, FL, pp. 52-57,* 1994.

[82] J. M. McIntosh, B. C. Kane, J. B. Bindell, and C. B. Vartuli. Approach to CD SEM metrology utilizing the full waveform signal. *Proceedings of the SPIE, Vol. 3332, pp. 51-60,* 1998.

[83] J. R. McNeil et al. Scatterometry applied to microelectronics processing. *Solid State Technology, Vol. 36, No. 3, pp. 29-30,* 1993.

[84] R. K. Mehra and J. Peschon. An innovation approach of fault detection and diagnosis in dynamic systems. *Automatica, No. 7, pp. 637-640,* 1971.

[85] A. F. Mills. *Basic Heat and Mass Transfer.* IRWIN, 1995.

[86] T. M. Mitchell. *Machine Learning.* The McGraw-Hill Companies, Inc., 1997.

[87] G. S. Moschytz. *Linear Integrated Networks: Design.* Van Nostrand Reinhold, 1974.

[88] John Musacchio. Run to run control in semiconductor manufacturing. Master's thesis, University of California, Berkeley, 1999.

[89] F. Nadi, A. M. Agogino, and D. A. Hodges. Use of influence diagrams and neural networks in modeling semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing, Vol. 4, No. 1, pp. 52-58,* 1995.

[90] Fariborz Nadi. *Modeling Complex Manufacturing Processes via Integration of Influence Diagrams and Neural Networks.* PhD thesis, University of California, Berkeley, 1989.

[91] W. Keith Nicholson. *Linear Algebra with Applications, Third Edition.* PWS Publishing Company, 1995.

[92] X. Niu, N. Jakatdar, J. Bao, and C. J. Spanos. Specular spectroscopic scatterometry in DUV lithography. *Microlithography, Proceedings of the SPIE, pp. 159-168*, 1999.

[93] X. Niu, N. Jakatdar, and C. J. Spanos. Deep ultraviolet lithography simulator tuning by resist profile matching. *International Symposium on Microelectronic Manufacturing Technologies, pp. 245-252*, 1999.

[94] Xinhui Niu. *An Integrated System of Optical Metrology for Deep Sub-Micron Lithography.* PhD thesis, University of California, Berkeley, 1999.

[95] W. G. Oldham, S. N. Nandgaonkar, A. R. Neureuther, and M. M. O'Toole. A general simulation for VLSI lithography and etching processes. Part I - Application to projection lithography. *IEEE Transaction on Electronic Devices, Vol. ED-26, pp. 717-722*, 1979.

[96] R. J. Patton and J. Chen. Review of parity space approaches to fault diag-

nosis for aerospace systems. *Journal of Guidance, Control, and Dynamics, Vol. 17, No. 2, pp. 278-85, March-April*, 1994.

[97] R. J. Patton, P. M. Frank, and R. N. Clark. *Fault Diagnosis in Dynamic Systems - Theory and Applications.* Prentice Hall (Control Engineering Series), London, 1989.

[98] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inferences.* Morgan Kaufmann, San Mateo, CA, 1988.

[99] L. R. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition.* Prentice-Hall, 1993.

[100] P. Rai-Choudhury, editor. *Handbook of Microlithography, Micromachining and Microfabrication, Volume 1: Microlithography.* SPIE Press, 1997.

[101] A. Rault, D. Jaume, and M. Verge. Online identification and fault detection of industrial processes. *Proceedings of the Sixth IFAC Symposium, pp. 1299-1305*, 1983.

[102] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific Series in Computer Science, Singapore: World Scientific, 1989.

[103] Wilson J. Rugh. *Linear System Theory.* Prentice Hall, 1993.

[104] David E. Rumelhart, G. E. Hinton, and R. J. Williams. *Parallel Distributed Processing.* The MIT Press, 1986.

[105] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach.* Prentice Hall, 1995.

[106] Shankar Sastry. *Nonlinear systems: analysis, stability, and control.* Springer, 1999.

[107] James R. Sheats and Bruce W. Smith, editors. *Microlithography: Science & Technology.* Marcel Dekker, New York, 1998.

[108] D. Stokes and G. S. May. Real-time control of reactive ion etching using neural networks. *IEEE Transactions on Semiconductor Manufacturing, Vol. 13, No. 4, pp. 469-480*, 2000.

[109] Charles J. Stone. *A Course in Probability and Statistics.* Duxbury Press, 1996.

[110] M. Sugawara. *Plasma etching: fundamentals and applications.* Oxford University Press, 1998.

[111] K. Tadros, A.R. Neureuther, and R. Guerrieri. Understanding metrology of polysilicon gates through reflectance measurements and simulation. *Proceedings of the SPIE, Vol. 1464, pp. 177-186*, 1991.

[112] Geoff Tennant. *Six Sigma: SPC and TQM in manufacturing and services.* Gower, 2001.

[113] S. Thornton and C. Mack. Lithography model tuning: Matching simulation to experiment. *Proceedings of the SPIE, Vol. 2726, pp. 223-234*, 1996.

[114] G. Toussaint. Some inequalities between distance measures for feature evaluation. *IEEE Transactions on Computers, pp. 409-410*, 1972.

[115] T. Turner. Electrical measurement of IC device CDs and alignment. *Solid State Technology, Vol. 41, No. 6, pp. 115-116*, 1998.

[116] V. N. Vapnik. *Statistical Learning Theory.* Wiley Interscience, 1998.

[117] Jiangxin Wang. Sensor validation and fusion of GPS aided longitudinal positioning system for IVHS. Master's thesis, University of California, Berkeley, 1998.

[118] Jiangxin Wang, Susan Y. Chao, and Alice M. Agogino. Sensor noise model development of a longitudinal positioning system for AVCS. *Proceedings of the American Control Conference (ACC'1999), San Diego, Session FM11-1, pp. 3760-3764*, 1999.

[119] Jiangxin Wang, Susan Y. Chao, and Alice M. Agogino. Validation and fusion of longitudinal positioning sensors in AVCS. *Proceedings of the American Control Conference (ACC'1999), San Diego, Session TM08-2, pp. 2178-2182*, 1999.

[120] Jiangxin Wang and Costas J. Spanos. A sensor fusion based methodology for real time furnace diagnostics. *Proceedings of Advanced Semiconductor Manufacturing Conference (ASMC'2000), Boston, pp. 397-406*, 2000.

[121] Jiangxin Wang and Costas J. Spanos. A novel approach for modeling and diagnosing the lithography sequence. *Proceedings of Advanced Equipment Control/Advance Process Control (AEC/APC XIII), Session 6, Banff, Canada, pp. 1652-1668*, 2001.

[122] Xue Z. Wang. *Data mining and knowledge discovery for process monitoring and control.* Springer, 1999.

[123] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica, No. 12, pp. 601-611*, 1976.

[124] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man and Cybernetics, Vol. 3, pp. 28-44*, 1973.

[125] L. A. Zadeh. Fuzzy sets as a basis for a theory of probability. *Fuzzy Sets System, Vol. 1, No. 1, pp. 3-28*, 1978.

[126] L. A. Zadeh. Fuzzy logic. *IEEE Computer Magazine, Vol. 1, No. 4, pp. 83-93*, 1988.

Appendix A

List of Symbols

Symbol	Description	Units
$(\cdot)^T$	matrix transpose	-
$(\cdot)_k$ or $\cdot(k)$	value at time step k	-
$diag(\cdot)$	diagonal matrix with diagonal elements contained in vector $\cdot$	-
$\mathcal{N}(\cdot,\cdot)$	normal distribution	-
∇^d	the dth order difference operator	-
α	fault parameters for furnace power elements	percentage
β	fault parameters for furnace temperature measurements	^oC
ϵ	general system error	-
ε	noise for the least square fault detection model of the furnace system	-
θ	biases of the inverse of the thermal resistance parameters	$\%power/^oC$

σ^2	variance of **e** in Chapter 4	-
σ^2_{ε}	variance of ε in Chapter 4	-
σ^2_e	variance of $e(t)$ in Chapter 5	-
σ^2_w	variance of $w(t)$ in Chapter 5	-
$\hat{\phi}_{ij}$	LS estimation of the ith parameter base on the jth data set	-
$\hat{\phi}^f_i$	fused estimation of the ith parameter	-
$\mathbf{\Phi}$	fault parameter matrix for the least square fault detection model of the furnace system	-
$\hat{\mathbf{\Phi}}$	fault parameters estimation by least square approach for the furnace system	-
a	model parameter	-
$\hat{a}(t)$	estimation of a at time step t	-
$a_1 \sim a_p$	time series model parameters for autoregressive terms	-
A, B	state space matrices of the furnace system model	-
A, G, C	state space matrices of the ARIMA(1,1,1) model	-
$\mathbf{A}_{\boldsymbol{\theta}}$	errors in state space matrix caused by $\boldsymbol{\theta}$	-
A_{FF}	state space matrice of fault-free furnace system model	-
ARIMA(p, d, q)	integrated autoregressive moving average time series model with p^{th} order autoregressive terms, d^{th} order integration terms and q^{th} order moving average terms	-
b	model parameter	-

$\hat{b}(t)$	estimation of b at time step t	-
$b_1 \sim b_q$	time series model parameters for moving average terms	-
coh	Partial Coherence	-
$\boldsymbol{cov\varepsilon}$	variance-covariance matrix of $\boldsymbol{\varepsilon}$	-
$C_1 \sim C_5$	thermal capacity for each zone of the furnace system	$minute/^oC$
$CD_{0.1}$	10% level Critical Dimension measurement	nm
$CD_{0.5}$	50% level Critical Dimension measurement	nm
$CD_{0.9}$	90% level Critical Dimension measurement	nm
$CD_{0.1}^{lib}$	10% level Critical Dimension measurement stored in data library	nm
$CD_{0.5}^{lib}$	50% level Critical Dimension measurement stored in data library	nm
$CD_{0.9}^{lib}$	90% level Critical Dimension measurement stored in data library	nm
$Cost$	Cost function for library searching approach of lithography process	-
D	system observations (data) in general	-
D_1, D_2	dimension 1 and 2 for a two-dimensional space	-
$e(k)$	Gaussian white noise	-
exp	Exposure Dose	mJ/cm^2
E	error function for neural networks	-
foc	Focus Position	μm

i, j	dummy index numbers	-
$\mathbf{I_n}$	$n \times n$ identity matrix	-
k	discrete time step	-
K	total number of data sets	-
L	total number of parameter to be estimated	-
M	system model in general	-
$\mathbf{M}$	predictor variable data matrix for the least square fault detection model of the furnace system	-
$\mathbf{M}_{\alpha k}$	predictor variable data associated with α for the least square fault detection model of the furnace system at time k	-
$\mathbf{M}_{\beta k}$	predictor variable data associated with β for the least square fault detection model of the furnace system at time k	-
$\mathbf{M}_{\theta k}$	predictor variable data associated with θ for the least square fault detection model of the furnace system at time k	-
$p(\cdot)$	probability density function	-
$p(\cdot \mid \cdot)$	conditional probability density function	-
P_0	initial state covariance matrix for Kalman filtering	-
$P(t \mid t)$	Kalman filter state estimation covariance matrix	-
$P(t+1 \mid t)$	Kalman filter state prediction covariance matrix	-
$\mathbf{P}$	power vector of the furnace system	%power
$P_1 \sim P_5$	power for each zone of the furnace system	%power

Q	covariance matrix of $v(t)$	-
$\mathbf{R}$	factor matrix of ***cov*ε** by Cholesky factorization	-
$R_1 \sim R_5$	thermal resistance for each zone of the furnace system	$^oC/\%power$
$R_{12} \sim R_{45}$	thermal resistance between adjacent zones of the furnace system	$^oC/\%power$
SWA	side wall angle measurement of Critical Dimension profile	degree
SWA^{lib}	side wall angle measurements of Critical Dimension profile stored in data library	degree
t	continuous time/discrete time step	-
t_{PEB}	Post-Exposure Bake time	sec
$\mathbf{T}$	temperature vector of the furnace system	^oC
$T_1 \sim T_5$	temperature for each zone of the furnace system	^oC
T_a	ambient temperature	^oC
T_{PEB}	Post-Exposure Bake temperature	^oC
$v(t)$	process noise in the state space form of the ARIMA(1,1,1) model	-
$\mathbf{V}(\hat{\mathbf{\Phi}})$	variance-covariance matrix estimation of estimated fault parameters by least square approach for the furnace system	-
$w_1 \sim w_4$	weights of CDs and SWA for cost function calculation	-
$w(t)$	measurement noise in the state space form of the ARIMA(1,1,1) model	-
w_{ij}	fusion weight for $\hat{\phi}_{ij}$	-

W	link weights for neural networks	-
x_0	initial states for Kalman filtering	-
$x(t)$	states in the state space form of the ARIMA(1,1,1) model	-
$\hat{x}(t \mid t)$	Kalman filter state estimation	-
$\hat{x}(t+1 \mid t)$	Kalman filter state prediction	-
X	system input in general	-
$y(t)$	ARMA sequence	-
$\bar{y}(t)$	average value of $y(i)$ for $i = 1, \cdots, t$	-
Y	system output in general	-
$\mathbf{Y}$	response variable data matrix for the least square fault detection model of the furnace system	-
$z(t)$	ARIMA sequence	-