

65

# Nonlinear Model Reduction Methods for Rapid Thermal and Chemical Vapor Deposition Processes

by

Suman K. Banerjee

B.Tech., Indian Institute of Technology, Madras (1993)  
M.S.C.E.P., Massachusetts Institute of Technology (1994)

Submitted to the Department of Chemical Engineering  
in partial fulfillment of the requirements for the degree of

**Doctor of Philosophy in Chemical Engineering**

at the

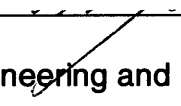
**Massachusetts Institute of Technology  
June 1998**

© 1998 Massachusetts Institute of Technology. All rights reserved

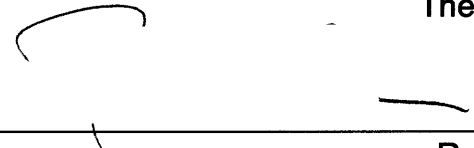
Signature of Author:

  
Department of Chemical Engineering  
May 12, 1998

Certified by:

  
Klavs F. Jensen  
Professor of Chemical Engineering and Materials Science and Engineering  
Thesis Supervisor

Accepted by:

  
Robert E. Cohen  
St. Laurent Professor of Chemical Engineering  
Chairman, Committee for Graduate Students

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

JUL 09 1998

LIBRARIES

Science

# **Nonlinear Model Reduction Methods for Rapid Thermal and Chemical Vapor Deposition Processes**

by

Suman K. Banerjee

Submitted to the Department of Chemical Engineering  
on May 12, 1998 in partial fulfillment of the requirements for the  
degree of Doctor of Philosophy in Chemical Engineering

## **Abstract**

The increasingly competitive nature of the semiconductor industry requires new process technology and equipment be brought to the manufacturing stage as quickly as possible. This requires a drastic reduction in the number of cut-and-try iterations required in designing process equipment and recipes. Modeling of the underlying physical phenomena governing these processes has helped towards this endeavor.

However, all of these modeling efforts have been concentrated towards building models mainly for equipment design and optimization. But there exists a distinct demand in the industry for models, which could be used for designing process recipes and process optimization since the equipment models in existence are too computationally intensive to address these issues in a reasonable amount of time. Many of the detailed models have hardware and software requirements that maybe outside the scope of many manufacturing organizations. Simulations using these models could also take anywhere from a few hours to days to yield a reasonable answer to process questions. Hence, there exists a distinct niche for accurate and fast models to address several processing issues. These models could be used by process engineers to design a better set of experiments to implement a new process recipe. They could also be used to answer “what-if” type of process questions on a day-to-day basis.

The work in this thesis develops and analyzes a method for generating reduced order nonlinear models from physically based, large scale finite element (FEM) models. The reduced order models are built using the same physical macroscopic conservation equations used to build the detailed models. The modeling strategy uses *a priori* knowledge of the process behavior to construct a reduced order model with relatively few unknowns. The resulting nonlinear model can then be used to extrapolate beyond the original knowledge base. Using the proper orthogonal decomposition (POD) method, the *a priori* information is extracted in the form of empirical eigenfunctions from simulation results of a detailed finite element model. Reduced order nonlinear models are then constructed using the empirical eigenfunctions as a basis set in a spectral Galerkin approximation to the physical model, i.e., the governing partial differential equations. The technique, though developed for rapid thermal processing (RTP) and chemical

vapor deposition (CVD) processes, is generic enough to be applied to any process that is described by similar macroscopic conservation equations.

Rapid thermal processes (RTP), because of their fast transient nature, make excellent test cases for the model reduction study. Simulation results with the reduced order models demonstrate good agreement with steady state and transient data generated from the finite element model, with an order of magnitude reduction in computation time. In particular, the reduced order models are shown to replicate an actual RTP processing cycle in a typical process chamber with an order of magnitude reduction in computation time compared to the finite element model. The reduced order models are capable of detecting temperature changes across silicon wafer surfaces due to the deposition of multilayer thin film stacks on the wafers. In this capacity, process engineers can use the reduced order models as a diagnostic tool. Also, a strategy for generating coupled fluid-thermal reduced order models is developed. This strategy incorporates multiple empirical eigenfunction sets in formulating spectral Galerkin approximations to the partial differential macroscopic conservation equations.

The modeling method is used to formulate reduced order models from mass conservation equations describing chemical vapor deposition (CVD) processes. Reduced order models are developed for two chemical systems – (1.) decomposition of trimethylgallium in the presence of hydrochloric acid, and (2.) in situ boron doping of silicon deposited from dichlorosilane. Simulation results from the reduced order models show good agreement against results from finite element models.

Thesis Supervisor: Klavs F. Jensen

Title: Professor of Chemical Engineering and Materials Science

# Acknowledgments

I would like to sincerely thank my advisor, Prof. Klavs F. Jensen, for his constant support and guidance throughout the duration of this work. I also thank him for providing various opportunities to me during the course of my research and patiently bearing with me during some of the frustrating phases. I express my gratitude to my thesis committee members for being helpful with their encouragement and comments over the past years.

Vernon Cole, during his stay at MIT and later at Motorola, was an invaluable help in transforming some of the ideas behind this work into practical reality.

The time spent in 66-250 and then in 66-501 was extremely memorable. All the members of the KFJ group have left their indelible mark on this work. I thank them all for the good times.

A special thanks to all the folks in Prof. Sawin's lab. Those, sometimes lengthy, interludes were a welcome distraction from the rigors of thesis research. I would like to thank Roy, Silvio, Dave, Angelo and Harsono for introducing me to the world of interesting real-time computer simulations.

The members of the Mahorowala household – Arpan, Joydeep and Venkatesh, were a joy to live with. The days spent with them were definitely some of the most memorable ones in my entire life.

I thank my parents for all their encouragement and guidance throughout my life, and more so over the course of this work.

Finally, I would like to thank Amit, Susmita and Sujatha for being there for me when I needed them most. Without their emotional support and encouragement, this thesis would never have been possible.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>4</b>
<b>TABLE OF CONTENTS.....</b>	<b>5</b>
<b>LIST OF FIGURES .....</b>	<b>8</b>
<b>LIST OF TABLES.....</b>	<b>12</b>
<b>CHAPTER 1 .....</b>	<b>13</b>
<b>INTRODUCTION .....</b>	<b>13</b>
<b>1.1 FURNACE PROCESSING.....</b>	<b>16</b>
<b>1.2 RAPID THERMAL PROCESSING .....</b>	<b>17</b>
<b>1.3 MODELING OF RTP SYSTEMS .....</b>	<b>19</b>
<b>1.4 MODELING OF CHEMICAL VAPOR DEPOSITION SYSTEMS .....</b>	<b>22</b>
<b>1.5 MATHEMATICAL BACKGROUND FOR MODEL REDUCTION STUDIES .....</b>	<b>24</b>
<b>1.6 THESIS OBJECTIVES AND OUTLINE .....</b>	<b>31</b>
<b>REFERENCES .....</b>	<b>34</b>
<b>CHAPTER 2 .....</b>	<b>41</b>
<b>MODEL REDUCTION STRATEGIES FOR LOW PRESSURE RAPID THERMAL PROCESSING SYSTEMS .....</b>	<b>41</b>
<b>2.1 INTRODUCTION .....</b>	<b>41</b>
<b>2.2 MODEL FORMULATION.....</b>	<b>43</b>
<b>2.2.1 DESCRIPTION OF THE REACTOR.....</b>	<b>43</b>
<b>2.2.3 METHOD FOR GENERATING NONLINEAR REDUCED ORDER MODELS.....</b>	<b>47</b>
<b>2.3 RESULTS AND DISCUSSION .....</b>	<b>53</b>
<b>2.3.1 STEADY STATE PERFORMANCE OF REDUCED MODELS.....</b>	<b>53</b>
<b>2.3.2 VARIATION OF RMS ERROR WITH INCREASING MODEL ORDER .....</b>	<b>56</b>
<b>2.3.3 TRANSIENT BEHAVIOR OF REDUCED MODELS.....</b>	<b>57</b>
<b>2.3.4 REDUCTION IN COMPUTATION TIME .....</b>	<b>72</b>

2.4	CONCLUSION .....	73
	REFERENCES .....	75
	CHAPTER 3 .....	78
	<b>REDUCED ORDER MODELING OF FLUID-FLOW INDUCED AND WAFER-SCALE PATTERN INDUCED TEMPERATURE VARIATIONS FOR RTP SYSTEMS .....</b>	<b>78</b>
3.1	<b>REDUCED MODELING OF COUPLED TEMPERATURE AND FLOW FIELDS IN RTP SYSTEMS</b> 79	
3.1.1	INTRODUCTION.....	79
3.1.2	MODELING EQUATIONS.....	79
3.1.3	FINITE ELEMENT FORMULATION OF MODELING EQUATIONS FOR AXISYMMETRIC GEOMETRIES.....	82
3.1.4	REDUCED MODELING STRATEGY FOR COUPLED FLUID-THERMAL EQUATIONS.....	84
3.1.5	RESULTS AND DISCUSSION.....	92
3.2	<b>REDUCED MODELING OF WAFER-SCALE PATTERN EFFECTS .....</b>	<b>96</b>
3.2.1	INTRODUCTION.....	96
3.2.2	MODEL REDUCTION STRATEGY.....	97
3.2.3	RESULTS AND DISCUSSION.....	102
3.2.4	REDUCTION IN COMPUTATION TIME .....	114
	REFERENCES .....	115
	CHAPTER 4 .....	117
	<b>REDUCED ORDER MODELING STRATEGIES FOR ATMOSPHERIC PRESSURE RAPID THERMAL PROCESSING SYSTEMS .....</b>	<b>117</b>
4.1	INTRODUCTION .....	117
4.2	MODEL FORMULATION.....	119
4.2.1	DESCRIPTION OF THE RTP REACTOR.....	119
4.2.2	FINITE ELEMENT MODEL FORMULATION .....	120
4.2.3	REDUCED MODEL FORMULATION.....	124
4.3	RESULTS AND DISCUSSION .....	131
4.3.1	COMPARISON OF STEADY STATE PERFORMANCE.....	131
4.3.2	TRANSIENT RESPONSES USING REDUCED MODELS.....	134

4.3.3	REDUCTION IN COMPUTATION TIME .....	144
4.4	CONCLUSION .....	145
	REFERENCES .....	147
	CHAPTER 5 .....	150
	REDUCED ORDER MODELING FOR CHEMICAL VAPOR DEPOSITION SYSTEMS .....	150
5.1	INTRODUCTION .....	150
5.2	MODEL FORMULATION .....	152
5.2.1	DESCRIPTION OF THE REACTOR.....	152
5.2.2	MODELING EQUATIONS AND FINITE ELEMENT FORMULATION .....	153
5.2.3	MODEL REDUCTION STRATEGY.....	158
5.3	RESULTS AND DISCUSSION .....	162
5.3.1	DECOMPOSITION OF TRIMETHYLGALLIUM IN THE PRESENCE OF HCL .....	162
5.3.2	BORON DOPING OF SILICON DEPOSITED FROM DICHLOROOSILANE.....	167
5.4	CONCLUSION .....	172
	REFERENCES .....	174
	CHAPTER 6 .....	176
	CONCLUSIONS AND RECOMMENDATIONS .....	176
6.1	CONCLUSIONS.....	176
6.2	RECOMMENDATIONS FOR FUTURE WORK.....	181
	APPENDIX A.....	184
	ALTERNATIVE METHOD FOR COMPUTING POLYNOMIAL NONLINEARITIES USING EIGENFUNCTIONS .....	184

# LIST OF FIGURES

<b>Figure</b>	<b>Caption</b>	<b>Page</b>
1.1	Schematic of a transistor as it evolves through several of the main steps of a process flow.	15
1.2	(a) Schematic of a conventional batch furnace, and (b) Typical time-temperature curve for a batch process such as oxidation.	17
1.3	(a) Schematic of an RTP system, and (b) Typical time-temperature curve for RTP.	19
1.4	Overview of the model reduction technique.	32
2.1	Schematic of a two-dimensional axisymmetric RTP reactor.	44
2.2	Schematic representation of the model reduction method.	45
2.3	Comparison of a typical temperature snapshot obtained from the FEM model with the dominant eigenfunction extracted by the POD procedure.	48
2.4	Comparison of the wafer center temperature of FEM and reduced models: (a) Lumped radiation model. (b) Explicit two band radiation model.	55
2.5	Variation of RMS error of wafer temperature with increasing number of eigenfunctions.	57
2.6	Behavior of wafer center temperature of FEM and reduced models, during transient ramp up and hold phases.	59
2.7	Temperature flood plots of FEM and reduced models during hold phase.	61
2.8	Behavior of wafer center temperature of FEM and combined reduced models, during transient ramp up and hold phases of the RTP cycle.	63
2.9	Difference in wafer temperatures between FEM and combined reduced models during ramp up and hold phases of the RTP cycle.	64



<b>Figure</b>	<b>Caption</b>	<b>Page</b>
2.10	Behavior of wafer center temperature of FEM and reduced models during ramp up, hold, and cool down phases of the RTP cycle.	66
2.11	Temperature flood plots of FEM and reduced models during cool down phase of the RTP cycle.	67
2.12	Transient ramp up, hold and cool down response of FEM and combined reduced models. (a) Combination of reduced models extracted at 1300 and 300 K with arithmetic averaging. (b) Combination of reduced models extracted at 1300, 1130, and 300 K with switch over.	69
2.13	Replication of RTP ramp cycle using FEM and reduced models.	71
3.1	Model reduction strategy for the coupled transient fluid-thermal system.	85
3.2	Relationship between different transient fields and eigenfunction sets.	86
3.3	Wafer center temperature trajectories from the FEM and reduced models.	93
3.4	Temperature and flow fields extracted at 0.1 atmospheres with H <sub>2</sub> as carrier gas.	94
3.5	Temperature and flow fields extracted at 0.01 atmospheres with N <sub>2</sub> as carrier gas.	94
3.6	Schematic of the front side of a patterned wafer.	97
3.7	Reduced modeling strategy for modeling wafer-scale pattern effects	101
3.8	Schematic of typical multilayer patterns on a MOSFET wafer.	103
3.9	Transient temperature trajectory comparisons for wafer-scale pattern effects for annealing pattern 1.	105
3.10	Transient temperature trajectory comparisons for wafer-scale pattern effects for annealing pattern 2.	107
3.11	Transient temperature trajectory comparisons for wafer-scale pattern effects for silicide step 1.	110

<b>Figure</b>	<b>Caption</b>	<b>Page</b>
3.12	Temperature profiles in the RTP chamber at the high temperature processing conditions as obtained from the FEM and reduced models.	111
3.13	Transient temperature trajectory comparisons for wafer-scale pattern effects for silicide step 2.	113
4.1	Schematic of the Applied Materials Centura™ RTP reactor.	120
4.2	Comparison of a typical temperature snapshot as obtained from the FEM model with the dominant eigenfunction extracted by the POD procedure.	127
4.3	Comparison of the wafer center temperature response from the FEM model and the 1300K reduced model.	132
4.4	Comparison of the wafer center temperature response from the FEM model and the 300K reduced model.	133
4.5	Wafer center temperature response for the RTP cycle simulation as obtained from the FEM model and the reduced models.	138
4.6	Wafer edge temperature response for the RTP cycle simulation as obtained from the FEM model and the reduced models	139
4.7	Wafer center temperature response of the combined reduced models with switch-over for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.	140
4.8	Wafer edge temperature response of the combined reduced models with switch-over for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.	141
4.9	Comparison of temperature flood plots from the FEM model and combined reduced models at different time instants into the RTP cycle.	142

<b>Figure</b>	<b>Caption</b>	<b>Page</b>
4.10	Wafer center temperature response of the combined reduced models with interpolation for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.	143
5.1	Schematic of tube-type axisymmetric single wafer reactor used for model reduction studies.	153
5.2	Model reduction strategy for chemical vapor deposition systems.	159
5.3	Comparison of concentration fields for trimethylgallium as obtained from the FEM and reduced models.	164
5.4	Comparison of concentration fields for dimethylgallium as obtained from the FEM and reduced models.	165
5.5	Comparison of concentration fields for monomethylgallium as obtained from the FEM and reduced models.	166
5.6	(a) Overall comparison of $\text{SiCl}_2\text{H}_2$ concentration fields. (b) Comparison of $\text{SiCl}_2\text{H}_2$ concentrations along axial cross-sections 1 and 2.	170
5.7	(a) Overall comparison of $\text{B}_2\text{H}_6$ concentration fields. (b) Comparison of $\text{B}_2\text{H}_6$ concentrations along axial cross-sections 1 and 2.	170
5.8	Comparison of boron incorporation rate and silicon growth rate on the susceptor surface as simulated by the reduced and FEM models.	171

# LIST OF TABLES

<b>Table</b>	<b>Caption</b>	<b>Page</b>
2.1	Comparison of model execution time under steady operating conditions.	72
2.2	Comparison of model execution time for RTP ramp up.	72
3.1	Nomenclature for multiple eigenfunction sets.	86
3.2	Emittance values in the three bands for a bare silicon wafer.	102
3.3	Emittance values in the three bands for annealing pattern 1.	103
3.4	Emittance values in the three bands for annealing pattern 2.	106
3.5	Emittance values in the three bands for the two silicidation steps.	108
3.6	Comparison of computation time in reduced order modeling of pattern effects.	114
4.1	Comparison of model execution time.	144
5.1	Chemical mechanism describing the decomposition of trimethylgallium (TMG) in the presence of HCl.	162
5.2	Chemical mechanism for the in situ boron doping of silicon deposited from dichlorosilane.	169
A.1	Comparison of model execution time between different reduced models and the FEM model for low pressure RTP systems.	185
A.2	Comparison of model generation time for different reduced models.	186

# Chapter 1

## Introduction

Processes used to manufacture semiconductor devices are becoming increasingly complex, while competition demands that these devices be brought to market more quickly and manufactured more reliably. This has caused an exponential growth in the necessary investment of time and money to develop new generations of process equipment. In the recent past, the construction of new equipment was still largely based on trial and error techniques, adjusting existing equipment to meet the increasing demands in a purely empirical way. However, in order to speed up this development process, one needs to understand the complicated physical rate processes governing each fabrication step and use this knowledge in designing future generations of equipment and processes. Such an understanding is best expressed in terms of a detailed, physically based, mathematical model. This has led to the development of a variety of detailed mathematical models based on the fundamental laws from chemistry and physics that provide a scientific basis for design, development and optimization of new process equipment. However, the solution of such a detailed model is often time consuming and requires the use of hardware and software resources beyond those available to typical manufacturing organizations because of the complex time dependent and three-dimensional nature of the process equipment. Simulations of these processes using the existing computational models can take hours to days to yield results. On the other hand, there exists a demand for fast and compact process models for

applications in process optimization, diagnostics and process control. Therefore, techniques are required for deriving low-order, physically based models for semiconductor manufacturing processes. These models could be used to study on-line process variations or to answer “what-if” type of process questions under a limited range of conditions on a day-to-day basis. The reduced complexity and smaller computational storage requirements imply that the reduced models can be simulated on desktop computers (such as PC’s), besides workstations. Hence, process engineers and operators could use these models for a better understanding of semiconductor manufacturing processes. A well-designed reduced model could help in cutting down the number of experiments required in designing a process recipe and thus reduce the transition time in bringing a process from the research to the manufacturing stage in a fabrication line. Another use of such a model would be in advanced model based control strategies.

Thermal processing is required in many steps in the manufacture of semiconductor devices in the microelectronics industry. Figure 1.1 shows a schematic of a single transistor as it evolves through several of the main processing steps[1], along with some of the typical dimensions of a transistor manufactured with the current technology. [2] A typical semiconductor device fabrication process could involve as many as ~ 25 steps of oxidation, annealing, nitridation, and chemical vapor deposition (CVD). [3] In all of these processes, the wafer is subjected to high temperatures anywhere from 400-1200 °C. The continuous demand for increases in speed and memory of semiconductor devices has led to the shrinking of device sizes further into the sub-micrometer regime. One of the main issues, therefore, in thermal processing of silicon wafers is the reduction of the “thermal budget”, defined as  $\int Tdt$ , where  $T$  is the wafer temperature and  $t$  is the processing time. For example, the gate oxide thickness and the junction depth are two critical dimensions that must be reduced in order to scale down overall device dimensions. If the wafer is held at the processing temperature for too long during gate oxidation, the gate oxide thickness becomes too large for proper device functioning. Similarly, if the wafer is held too long at the process temperature during ion implant anneal, the dopants diffuse too far into the silicon, and the junction depth becomes too large. Any thermal

processing the wafer receives after the source and drain anneal may cause further unwanted increase in junction depth. As device dimensions are scaled down, the reduction of thermal budget and overall control of thermal processes are becoming critical issues in the microelectronics industry.

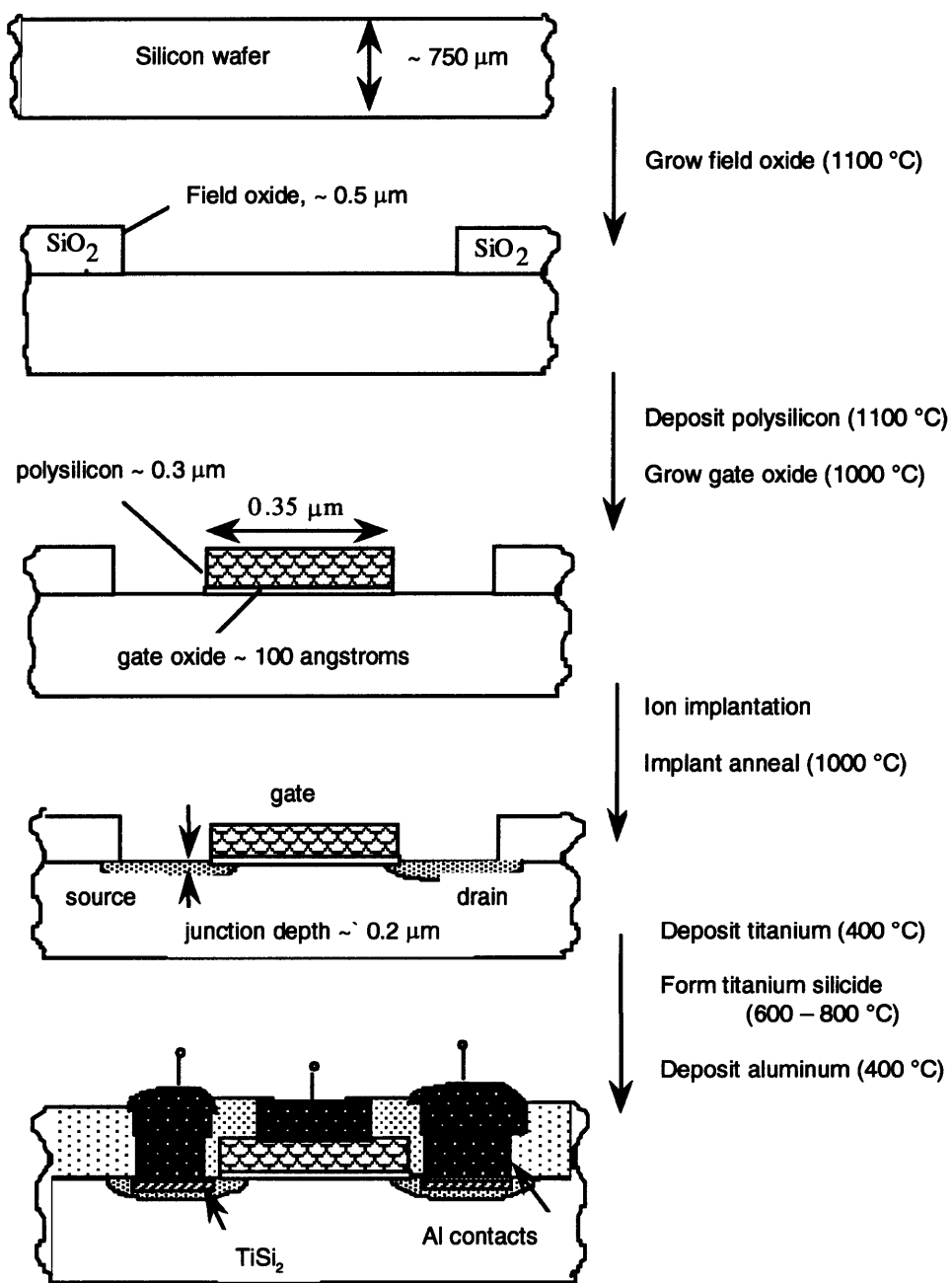


Figure 1.1 Schematic of a transistor as it evolves through several of the main steps of a process flow.

## 1.1 FURNACE PROCESSING

Conventionally, batch furnaces have been used for the thermal processing of wafers. In a batch furnace, 25 – 150 wafers are simultaneously heated to the processing temperature. The schematic of a typical batch furnace is shown in Figure 1.2a. The quartz walls of the furnace are resistively heated using block heaters, which in turn heat the wafer stack by a combination of radiative, conductive, and convective heat transfer. In a typical process[4] the batch of wafers are slowly loaded at a temperature between 700 and 800 °C in an inert ambient. The temperature of the wafer stack is then ramped to the process temperature over a period of approximately 20 minutes. A typical time-temperature curve for a batch furnace process is shown in Figure 12.b. The large thermal mass of the system prevents the temperature of the wafers from being increased too rapidly, and so the wafers spend tens of minutes at the processing temperature. Fast ramping of the wafer temperature also results in large thermal gradients and plastic deformation (slip) of the wafers. This large thermal budget associated with furnace processing is not compatible with shrinking device dimensions, as it increases the solid state diffusion of dopants introduced in previous processing steps. Future generations of devices would require device sizes less than 0.2  $\mu\text{m}$ , which will not be achievable in a batch furnace. [2] Furthermore, the potential loss of a large number of wafers due to an operational error in processing is driving the semiconductor industry towards single wafer processing tools.



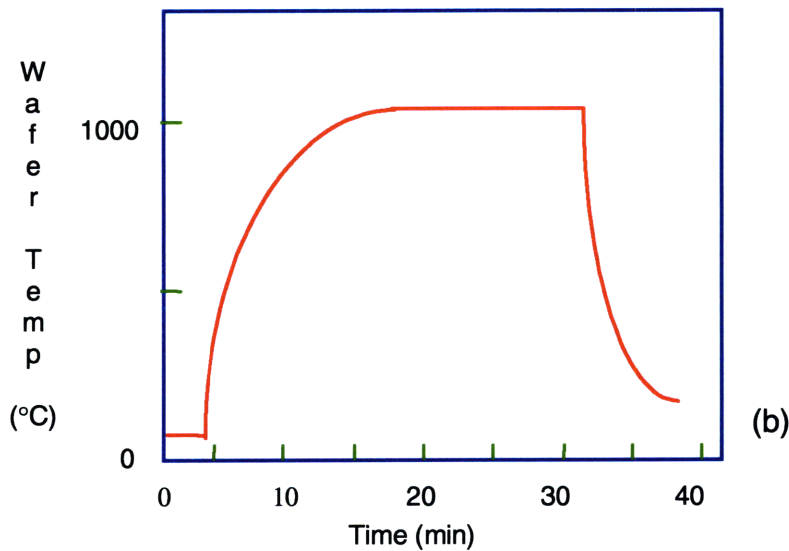
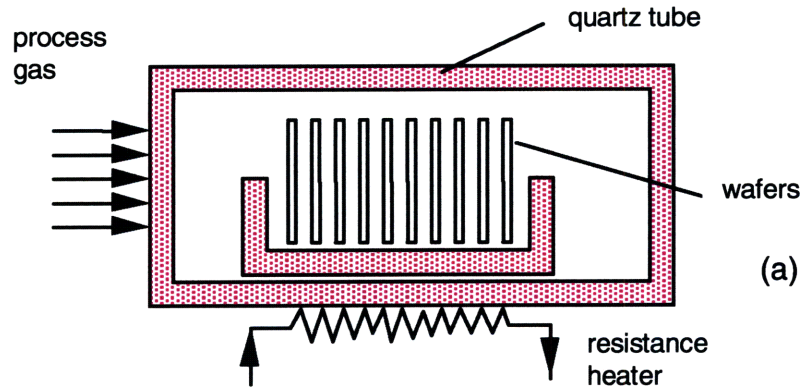


Figure 1.2 (a) Schematic of a conventional batch furnace, and (b) Typical time-temperature curve for a batch process such as oxidation.

## 1.2 RAPID THERMAL PROCESSING

The need for faster and more uniform thermal processes, coupled with the drive towards single wafer processing has led to the rise of Rapid Thermal Processing (RTP)[5] as a viable alternative to conventional furnace processing. RTP systems are, in general, single wafer reactors. [5, 6] The schematic of a typical RTP reactor is shown in Figure1.3a. In a RTP

system, the wafer is heated to the processing temperature using high intensity incoherent radiation, typically by an array of tungsten halogen lamps that have an operating temperature of  $\sim 3000$  K. The wafer is thermally isolated from its environment by supporting it on several quartz pins or on a thermal guard ring, which facilitates temperature ramp rates as high as  $100$  °C/sec. The walls of the reactor stay relatively cold during the entire process. The wafer is heated and cooled, primarily by radiative heat transfer, with the radiative heat flux on the order of  $100$  kW/m<sup>2</sup> and the convective heat flux on the order of  $1$  kW/m<sup>2</sup>. A typical time-temperature curve for a RTP processing cycle is shown in Figure 1.3b. In a typical RTP process, the wafer is introduced into the reactor at room temperature. Subsequently, the wafer temperature is ramped to the processing temperature of  $\sim 1000$  °C in  $30 - 60$  seconds. The wafer is held at the processing temperature for a time duration of  $\sim 30$  seconds, following which the temperature is ramped down and the wafer is removed from the chamber while it is still considerably hotter than ambient conditions. Therefore, the processing time in a RTP system is very short,  $\sim 2$  minutes, which minimizes solid state diffusion and preserves previously formed dopant profiles from previous steps. This short processing time and the associated low thermal budget makes RTP a viable alternative to conventional furnace processing.

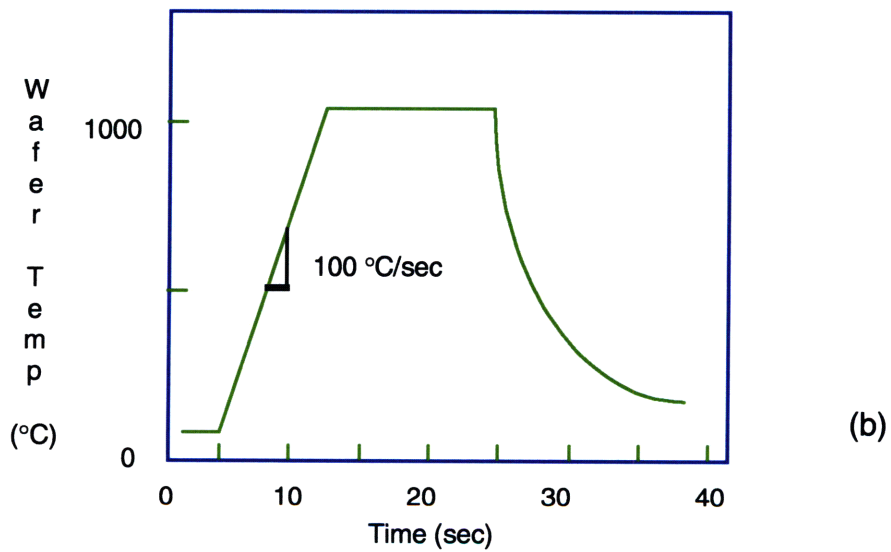
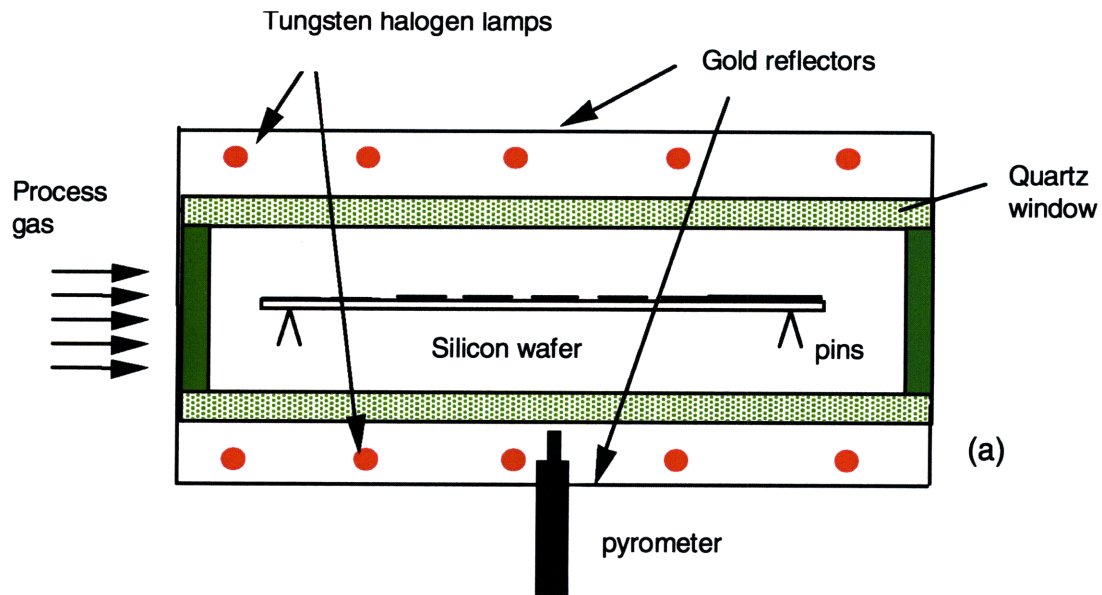


Figure 1.3 (a) Schematic of a RTP system, and (b) Typical time-temperature curve for RTP.

### 1.3 MODELING OF RTP SYSTEMS

The competitive nature of the industry demands new technology, including RTP, be brought from its inception to the manufacturing stage in the fabrication line as fast as possible.

This requires a detailed understanding of the physical processes governing these systems and also the capability to extend this understanding to future generations of equipment and technology. Such an understanding is best expressed in terms of mathematical models built on the physical rate processes governing these systems. Hence, computational modeling of the physical processes has contributed not only in the development of better equipment but also in the development of process recipes and process optimization.

Several approaches have been taken in the modeling of RTP systems. Lord[7] and later Kakoschke *et al.* [8] were among the first to simulate physical phenomena in RTP systems. In these simple models, a one-dimensional transient heat conduction equation was solved for the silicon wafer in a RTP cycle. The models provide useful insight into the dynamics of RTP processing, but have little predictive capability since both the heat generation and heat loss terms are treated empirically. Campbell *et al.* [9, 10] modeled the steady state flow and temperature fields of an incompressible gas in a RTP reactor. Kersch *et al.* [11] combined steady state simulations with a detailed radiative heat transfer model to predict steady state profiles. Knutson *et al.* [12] modeled the three dimensional steady state fluid flow and temperature fields in a commercial RTP reactor using a diffuse gray radiation model. All of these models predict steady state temperature profiles, but do not take into account the inherent transient nature of the RTP cycle, hence limiting their applicability to the steady state design of a RTP chamber.

Chatterjee *et al.* [13] solved for the transient wafer temperature profiles in a RTP reactor, assuming that the gas in the reactor was stagnant. Merchant *et al.* [14] and Jensen *et al.* [15] developed a transient fluid flow and heat transfer model coupled with an accurate radiative heat transfer model to predict temperature profiles in RTP reactors. The radiative heat transfer model computes the radiation energy transport for both specular and diffusely reflecting surfaces in the reactor. The model allows for the inclusion of wavelength, temperature, and material dependent properties.

Semiconductor device manufacturing begins with a blank silicon wafer, but with each step of wafer processing, thin layers of materials such as silicon dioxide, polysilicon, silicon

nitride and titanium silicide are grown or deposited and patterned by photolithography. [1] The presence of these layers change the radiative properties of the wafer through thin film interference effects that in turn alters the temperature profile across the wafer surface. There has been direct experimental evidence of temperature variations across the wafer due to these pattern effects. The most extensive work has been done by Vandennebeele *et al.* [16], who showed that simple silicon dioxide patterns could cause temperature non-uniformities of up to 80 °C. These patterns also caused plastic deformation of the wafer, with slip lines coinciding with the shape of the patterns. Others have also reported similar effects. [17, 18] Indirect evidence for pattern-induced plastic deformation for product wafers has also been reported in the literature. [19, 20] Feil *et al.* [19] reported that the degree of pattern misregistration after an RTP reflow step was strongly pattern dependent. Buller *et al.* [20] reported similar observations of pattern-dependent misregistration after an RTP contact anneal step. There has also been modeling work to predict the influence of wafer-scale patterns on temperature non-uniformities across the wafer surface. Vandennebeele *et al.* [21] used a simple model of an RTP reactor to examine pattern effects. But there are many important transport effects that cannot be captured by these simple models. Therefore detailed reactor scale transport models have been used to account for these effects. [14, 22] The most prominent of these effects is radiative heat exchange in the system that can only be described accurately by finite element and/or Monte Carlo models. [15] Hebb *et al.* [23] integrated radiative properties modeling with finite element based reactor transport modeling [14] to show that multilayer patterns can have extreme effects on transient and steady state temperature uniformity during RTP. The simulations also showed that reactor design can have large effects on pattern induced temperature non-uniformity that could lead to pattern misregistration and plastic deformation of the wafer.

These modeling efforts have been primarily concentrated towards the development of computational models to be used for equipment design and optimization. Additional studies have also been conducted to develop models for application in process control schemes for RTP systems. Norman [24] used a simplified one-dimensional wafer model with annular lamp rings

to control the radial temperature profile across the wafer surface by modifying the lamp powers. Breedijk *et al.* [25] improved on this simple model by taking into account radiative exchange factors from the various lamps to the wafer surface compute a more realistic power input to the wafer. The one-dimensional model was further combined with a non-linear least squares minimization approach to dynamically control the wafer temperature during the RTP transient phase.

Schaper[26] used a multizone, multivariable control scheme to control the wafer temperature during the RTP cycle. Later, Schaper *et al.* [27, 28] extended their work to eight different axisymmetric RTP reactor configurations and demonstrated that good temperature uniformity could be achieved throughout the process range. Aral *et al.* [29] constructed simple gain scheduled PI controllers using a batch least squares approach for parameter identification based on the RTP model developed by Merchant *et al.* [14]

However, there is a need for a class of models that potentially could be used by process engineers to design experiments to implement a new process recipe or answer “what-if” type of questions on a day-to day basis. Simple models, such as those used in control schemes, have no predictive capabilities and previous equipment design oriented models are too computationally intensive to be applied in this capacity. Many of these equipment design models take anywhere from hours to days to simulate reasonable answers to these process questions. Thus there exists a distinct niche for accurate and fast models to serve the demands of the process engineers. These models could also be used in a combined feedforward-feedback control strategy for RTP systems, where the feedforward transient trajectory could be simulated using these models and a simple PID controller could be used to implement corrective feedback control around such a trajectory.

## **1.4 MODELING OF CHEMICAL VAPOR DEPOSITION SYSTEMS**

Chemical vapor deposition (CVD) has been an important technology for the deposition of thin films on silicon wafers from the earliest days of the microelectronics industry. Today, thin films of silicon dioxide, doped and undoped polysilicon, epitaxial silicon, silicon nitride, aluminum, tungsten, and tungsten silicide deposited by CVD play an essential role in manufacturing silicon-based microelectronic circuits.

The tremendous increase in complexity in semiconductor processing in the last decades, leading to the present generation of ultra large scale integration (ULSI) chips with 16 million components or more on one chip, at typical feature dimensions of 0.5  $\mu\text{m}$  or less, has led to ever increasing demands on the performance of CVD processes. [30] This has caused an exponential growth in the necessary investment of time and money to develop new generations of CVD equipment. Nevertheless, the construction of new CVD reactors is still largely based on trial and error techniques, adjusting existing equipment to meet increasing demands in a purely empirical way. Computational models, based on fundamental laws of physics and chemistry, have proved to be helpful both in equipment as well as process design and optimization.

There are several different reactor types used for various CVD applications in the semiconductor industry. Kleijn [30] presents a detailed review of the various reactor configurations and relevant modeling efforts. Some of the reactor types and the modeling work done on them are presented here. Atmospheric or slightly reduced pressure horizontal duct reactors are used mainly in research and to some extent in the production of compound semiconductors. The state of the art modeling effort for this class of reactors is formed by two-dimensional elliptic fluid flow models in combination with detailed chemistry, including several dozens of gas phase and surface species and reactions, by Jensen *et al.* [31] Atmospheric pressure axisymmetric vertical reactors are widely used for epitaxial growth at near-atmospheric pressures and for low-pressure polycrystalline CVD processes. Some of the notable modeling work for this type of reactors is by Jensen *et al.* [32], Patnaik *et al.* [33] and Fotiadis *et al.* [34] The horizontal hotwall multiple-wafer-in-tube batch reactor, operated at low pressures (30 to 100 Pa), is the most widely used reactor configuration in the silicon-based industry. [30]

Amongst others, it is used to deposit doped and undoped polycrystalline silicon, silicon oxide, and silicon nitride. The landmark modeling study for this kind of systems is by Jensen and Graves. [35] With the increasing demands on film uniformities and qualities, the development of new low-pressure CVD processes such as tungsten LPCVD from  $WF_6$ , and with the increase of wafer dimensions to 200-mm diameter and larger, there has been an increasing interest in single-wafer low pressure CVD reactors. Kleijn studied gas phase and surface reactions in polysilicon LPCVD in this class of reactors. [36] There has also been some modeling work in rapid thermal CVD by Merchant. [37]

All the detailed models for CVD processes have drawbacks in terms of extensive computational resource requirements similar to the detailed RTP models. Hence, there exists a demand for suitable modeling techniques that can be used to develop models, which are compact and fast from a computational point of view.

## **1.5 MATHEMATICAL BACKGROUND FOR MODEL REDUCTION STUDIES**

Particular computational methods have assumed prominent positions in certain areas of applications, such as the finite element method in structural problems, and spectral methods for global atmospheric modeling and weather predictions. These apparently unrelated classes of algorithms are actually specializations of one general method, with the artificial division arising due to the segregation into different modeling areas.

Solution techniques in this general class are called the Method of Weighted Residuals. [38, 39] Within the method of weighted residuals, there exists a class of numerical techniques called the Galerkin method. This method was originally used by Galerkin[40] to study elastic equilibrium of rods. Widespread availability of computers led to more emphasis on Galerkin methods, as solutions of greater accuracy with minimal execution time could be achieved. A good introduction including connections to other numerical methods through appropriate choices



of expansion functions and test functions can be found in a book by Fletcher. [39] An important fact in developing a Galerkin method is that the expansion fields are deliberately chosen to satisfy the boundary conditions. Following the Galerkin method, the work in this thesis takes that observation one step further and also requires that the expansion fields be likely candidates of the actual problem solution. The use of non-analytic functions, however, does not alter the Galerkin nature of this procedure.

The key features of a Galerkin method are a differential equation (ordinary or partial)

$$Lu = 0 \tag{1}$$

in a domain  $D$  with boundary conditions

$$Su = 0 \tag{2}$$

on the boundary of the domain ( $\partial D$ ). The Galerkin method assumes that the solution  $u$  can be well approximated by a series of known functions  $\{\phi_i\}_{i=1}^N$ :

$$u = u_0(\mathbf{x}) + \sum_{i=1}^N a_i \phi_i(\mathbf{x}) \tag{3}$$

Normally the  $\phi_i$  are analytical functions, but that need not necessarily be the case, as will be shown by the empirical nature of the trial functions used in this thesis. Frequently, the trial functions,  $\phi_i$ , are deliberately chosen to satisfy the boundary conditions. Substituting the expansion for  $u$  into the linear governing equation produces a nonzero residual given by

$$R = Lu = Lu_0 + \sum_{i=1}^N a_i L(\phi_i) \tag{4}$$

In the case of the method of weighted residuals, the residual is forced to be orthogonal to some other class of functions not necessarily those used in the expansion. In the Galerkin method, the trial functions,  $\phi_i$ , are the same as the test functions,  $w_i$ , in the requirement

$$\langle R, w_i(\mathbf{x}) \rangle = 0 \quad (5)$$

where  $i = 1, \dots, N$ . This generates a matrix equation that is generally solved iteratively by some method, such as the Newton-Raphson method, [38, 41] to obtain the coefficients  $a_i$ . The leading term,  $u_0$ , is included to satisfy the boundary conditions and, in general, can be subtracted off the dependent variable,  $u$ , to formulate a problem with homogeneous boundary conditions. Handling of the boundary conditions in this manner allows for local error reduction (on  $\partial D$ ) while the whole Galerkin procedure is constructed to achieve global error reduction.

The trial functions used in a Galerkin expansion are critical in determining the success of the method in terms of rate of error reduction and calculation time required. Expansions in terms of trial functions that are not orthogonal leads to a system of equations which has to be solved simultaneously to obtain values for all the coefficients. Moreover, an increase in the number of trial functions requires another complete solution of the (now larger) set of coefficients. In the case of orthogonal functions, the set of equations is independent and can be solved one at a time leading to simpler matrix problem. Adding another orthogonal element to the basis requires only a single derivation and the solution of only one more equation since the values of the previously solved coefficients do not change. The orthogonal trial functions could be locally defined within smaller discretized elements of the entire domain. Examples of such methods are the finite element method[42], and the method using wavelets[43]. In contrast to these methods that rely on low-order, local nature of the trial functions, spectral methods[44] exploit global orthogonal trial functions, trading off ease of computation for rapid convergence. A few well-known trial functions used in spectral methods and their relative advantages are listed below[45]:

1. A Fourier transform expands an operator in terms of the trigonometric functions sine and

cosine and is well suited for problems with periodic boundary conditions. This basis has the nice property of infinite differentiability as well, and has been shown to exhibit the best convergence rate for smooth functions, but performs poorly near discontinuities (including at non-periodic boundaries).

2. A Taylor series of the monomials  $x^n$  for  $n$  increasing from zero does not see wide use in practice due to wild oscillations that occur when trying to fit functions using a large number of basis elements.
3. Legendre polynomials fix this problem to some degree and offer good wavelength resolution as well, and are non-periodic, but converge very slowly near discontinuity boundaries.
4. Chebyshev polynomials are perhaps the most “robust” available in terms of being able to fit most functions, and exhibit good convergence rates near discontinuities, however they are not strictly orthogonal which adds to the computation time. These polynomials satisfy the criterion of minimum possible maximum error as well. [38]

The list above is ordered by increasing robustness, but decreasing accuracy, the standard trade-off faced when choosing global trial functions. The use of analytical, as opposed to empirical basis functions, occurred during the inception of Galerkin methods due to the absence of significant computational resources such as workstations. But clearly one desires to have a set of basis functions tailor-made for each problem type. Such a set would have the ability to provide the best trade-off between accuracy and computation time as the minimum number of functions could be used to expand the solution with the greatest accuracy. The derivation of such an “optimal” basis set requires the widening of trial function choice to include empirical globally defined orthogonal functions. The derivation and the subsequent use of such empirical functions in spectral Galerkin expansions of partial differential equations with algebraic constraints are described in this thesis.

In this thesis, the Karhunen-Loève (KL)[46, 47] technique has been used to generate an empirical basis set. The technique, as is presented in this thesis[45], is more generalized than that presented by the original developers – Karhunen[47] and Loève[46]. According to Lumley

[48] this technique was suggested independently by several scientists, e.g. Kosambi[49], Loève[46], Karhunen[47], Pougachev[50], and Obukhov[51]. This technique is also known as Proper Orthogonal Decomposition (POD) in operator theory[52], Principal Component Analysis (PCA) in statistics[53, 54], and empirical orthogonal eigenfunctions in meteorology[55]. The technique in its various forms has been used in a variety of disciplines: Papoulis[56] – random variables; Rosenfeld and Kak[57] – image processing; Algazi and Sakrison[58] – signal analysis; Andrews *et al.* [59] – data compression; Preisendorfer[53] – oceanography; Sirovich[60-64] – fluid mechanics; and Gay and Ray[65] – process identification and control in chemical engineering.

The form of the POD technique used in this thesis, generically referred to as the proper orthogonal decomposition method of empirical eigenfunctions, yields a decomposition of data set that is, in a well defined sense, optimal. [66] Given an ensemble of patterns, the technique yields an orthogonal basis for the representation of the ensemble, as well as a measure of the relative contribution of each basis function to the total “energy” (mean square fluctuation) of the ensemble. The basis is optimal in the sense that a *truncated* series representation of the data in it has a smaller mean square error than a representation by any other basis *of the same dimension*. These properties make the basis set a natural one to consider when performing model reduction or data analysis and compression. [66] This technique was first used by Lumley[67] as a rational procedure for the extraction of *coherent structures*. [60] However, the method in its original formulation was often numerically intractable until Sirovich[60] proposed the method of snapshots to greatly reduce the amount of computation necessary for the analysis of large data sets. Following this, the technique has been widely used in the analysis and construction of low dimensional models in turbulent fluid mechanics and catalytic reaction-diffusion systems. Some of the notable applications of this technique have included: construction of low-dimensional models of turbulent boundary layer flows[52, 68], time dependent flows in complex geometries[69], data analysis of turbulent Rayleigh-Benard convection simulations[62, 69, 70], simulations of the Ginzburg-Landau equations[64], and temperature patterns in heterogeneous

catalytic reaction-diffusion experiments[71, 72]. The review by Berkooz *et al.* [48] contains a presentation of mathematical properties of the method, as well as a discussion of applications to turbulent flows. The usual implementation of the KL procedure can be summarized as follows. [66] Consider a time-dependent, spatially varying process that, after an initial transient, approaches a final, stationary – but not necessarily steady – state, an “attractor”. Data from this process, in the form of discretized images, are obtained by following the evolution in time of one or a few initial conditions, sometimes including transient data so that some information about the approach to the final state (attractor) is retained. Covariances are then computed, taking the ensemble average to be simply the time average. The eigenvectors of the covariance matrix constitute the hierarchical “optimal” basis. In this thesis, a variation of this technique is used. Instead of an eigenvalue analysis, a singular value decomposition of the temporal covariance matrix is used to compute the empirical trial functions. This technique is similar to the one used by Wyckoff[45] to extract eigenfunctions to expand partial differential equations related to atmospheric chemistry and flow problems.

The essential content of the KL procedure, as presented by Sirovich *et al.* [60-64], is as follows. Imagine an ensemble of states or snapshots on the attractor,

$$\mathbf{v}^{(n)} = \mathbf{v}(\mathbf{x}, t_n) \quad (6)$$

sampled at uniformly spaced, uncorrelated times  $t_n$ . To arrive at the KL procedure one can seek the *most likely state*,

$$\lambda = \left\langle \left( \mathbf{v}^{(n)}, \phi \right)^2 \right\rangle \quad (7)$$

is a maximum, subject to the normalization condition

$$(\phi, \phi) = \int \sum_k \phi_k(\mathbf{x}) \phi_k(\mathbf{x}) d\mathbf{x} = 1 \quad (8)$$

The solution to this problem is given by the principal eigenfunction of

$$\mathbf{K}\phi = \int \mathbf{K}(\mathbf{x}, \mathbf{x}')\phi(\mathbf{x}')d\mathbf{x}' = \lambda\phi(\mathbf{x}) \quad (9)$$

where

$$\mathbf{K}_{ij}(\mathbf{x}, \mathbf{x}') = \langle v_i(\mathbf{x})v_j(\mathbf{x}') \rangle = \frac{1}{M} \sum_{n=1}^M v_i^{(n)}(\mathbf{x})v_j^{(n)}(\mathbf{x}') \quad (10)$$

is the two-point correlation function and  $M$  is the number of snapshots that have been collected. Here,  $\mathbf{K}$  is a non-negative Hermitian operator and Equation 10 generates a complete orthonormal set  $\{\phi_n\}$  with  $\lambda_n \geq 0$ .

Finally except for a set of zero measure the *snapshots* can be expanded in the eigenfunctions  $\{\phi_n\}$

$$\mathbf{v}(\mathbf{x}, t) = \sum_{n=1} a_n(t)\phi_n(x) \quad (11)$$

where convergence is in the  $\ell_2$  sense. The coefficients,  $a_n$ , are uncorrelated in time, i.e., they are statistically orthogonal

$$\langle a_n a_m \rangle = \lambda_n \delta_{nm} \quad (12)$$

There have been some studies in using these empirical eigenfunctions in Galerkin expansions for a variety of systems. Sirovich[63, 70] has suggested the use of these eigenfunctions in the development of reduced models for analyzing fluid flow problems. The method has also been used for model reduction, in conjunction with an artificial neural network,

to construct empirical dynamical systems from experimental data in an isothermal heterogeneous catalytic reaction-diffusion system. [73] Empirical eigenfunction sets have also been used by Wyckoff[45] to develop models for studying atmospheric chemistry and flow problems.

## **1.6 THESIS OBJECTIVES AND OUTLINE**

There exists a lack of systematic methods for development of reduced order models in the semiconductor industry. This thesis develops one approach to obtaining reduced order process models from complex, physically realistic, finite element models without compromising on the underlying nonlinearities governing the system. The method is developed in such a way that the algorithm and software tools are not equipment or process specific, but could be generalized and used for a wide class of processes and equipment, that can be described by the same macroscopic conservation equations. The model reduction technique, developed in this thesis, introduces a generic step-by-step procedure for building nonlinear reduced order models from detailed finite element (FEM) models. Data obtained FEM model simulations are used to extract empirical eigenfunctions using the POD method. These eigenfunctions are then used in spectral Galerkin expansions of the same physical rate governing equations used to build the FEM models. Thus no compromises, such as lumping or truncation of terms, are made while solving these equations within the reduced model framework. RTP and CVD systems are used as test-beds for analyzing the model reduction technique. However, the generality of the algorithm and software tools does not restrict the technique to the test-beds; rather the method can be applied to any system that are determined by similar macroscopic conservation equations. A general overview of the model reduction technique is shown in Figure 1.4.

## PROPER ORTHOGONAL DECOMPOSITION METHOD

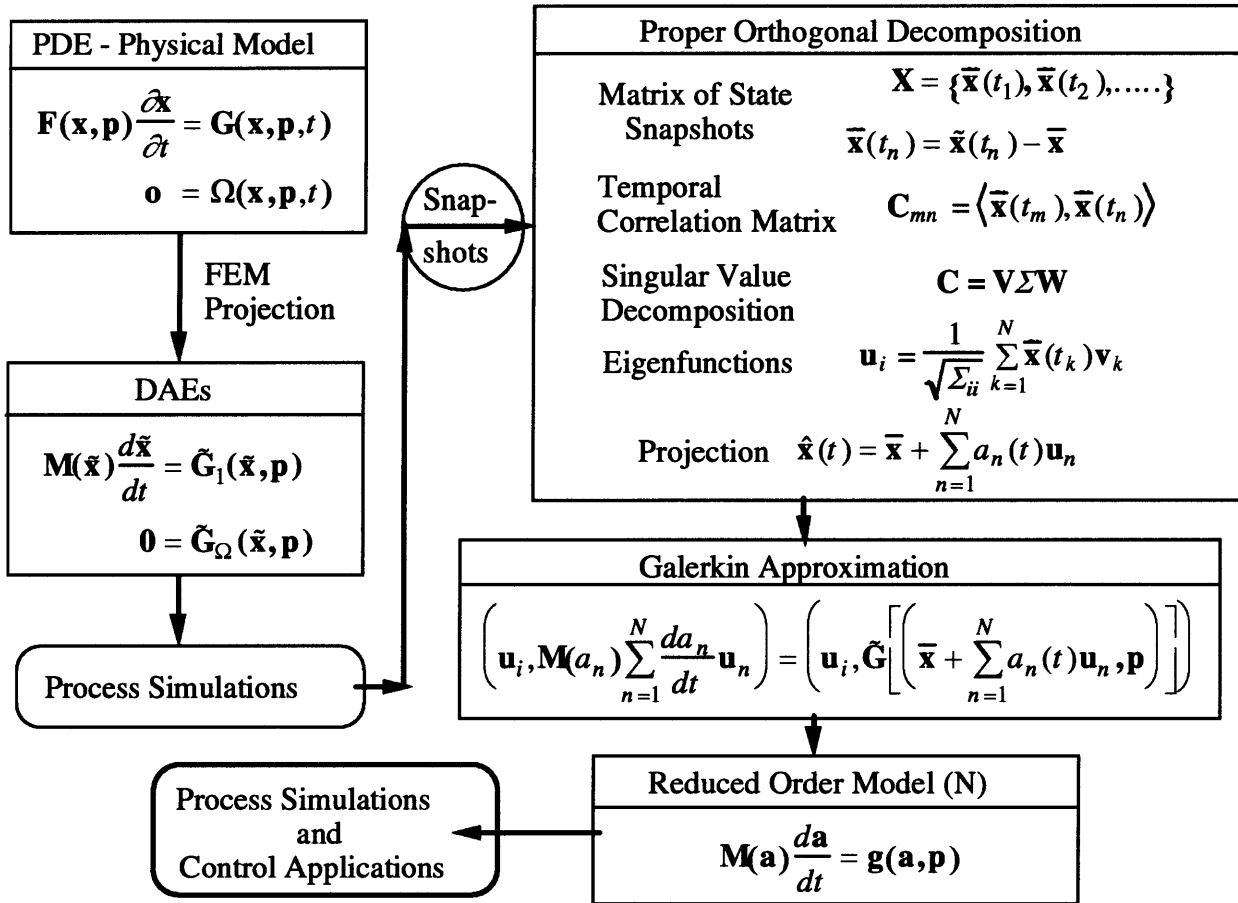


Figure 1.4 Overview of the model reduction technique.

The second chapter deals with model reduction for low pressure RTP system. RTP systems because of their dynamic nature provide unique opportunities for the study of the transient and steady behavior of the reduced models. The eigenfunction extraction technique is introduced here. Following that the model reduction procedure for the energy conservation equation is formulated from the FEM model. Both the steady and transient behaviors of the reduced model are compared against the finite element model. The tradeoff between accuracy and increasing model order is analyzed. Finally timing runs are performed to evaluate the computational savings obtained from the reduced model.

The third chapter is divided into two sections. The first section deals with the model



reduction of the coupled fluid-flow and heat transfer formulation for RTP systems. This section introduces the model reduction technique using multiple eigenfunction sets. The mathematical complexities such as orthonormalities across multiple eigenfunctions sets and cross coupling of terms are addressed. Subsequently a model reduction procedure, using multiple eigenfunction sets, is formulated for generating coupled fluid-flow and heat transfer reduced models. The reduced models are compared against the finite element model to study their accuracy. The second section introduces a model reduction technique for analyzing temperature variations across the silicon wafer in RTP systems, due to the introduction of multilayer thin-film stacks on the wafer surface. The transient and steady behaviors of the reduced models are compared against FEM results. Timing runs are performed for the reduced and FEM models to evaluate the computational savings.

The fourth chapter introduces the model reduction technique for atmospheric pressure RTP systems. Reduced models are formulated for a commercial RTP system different from the geometry used in the earlier chapters. A framework for using reduced models to replicate a commercial RTP cycle is developed. The strategy is then compared against both finite element and actual process data to study its efficacy. The computational time savings achieved by using reduced models is evaluated by comparing them against the detailed finite element model.

In the fifth chapter, the model reduction technique is extended to chemical vapor deposition (CVD) systems. A model reduction strategy is formulated for reducing mass conservation equations. Multiple eigenfunction sets extracted from chemical species conservation fields are used to extract reduced models, which are compared against finite element models. The strategy is used to generate reduced models two systems – (1.) decomposition of trimethylgallium and (2.) boron doping of silicon.

Conclusions and recommendations for further research are given in the sixth chapter.

## REFERENCES

- [1] S. Wolf and R. N. Tauber, *Silicon Processing for VLSI Era: Volume 1*. Sunset Beach, CA: Lattice Press, 1986.
- [2] "The National Technology Roadmap for Semiconductors," Semiconductor Industry Association, San Jose, CA 1994.
- [3] I. Calder, "Rapid Thermal Process Integration," in *Reduced Thermal Processing for ULSI*, R. A. Levy, Ed. New York: Plenum Press, 1989, pp. 181.
- [4] J. Nulman, "Rapid thermal processing with reactive gases," in *Reduced Thermal Processing for ULSI*, R. A. Levy, Ed. New York: Plenum Press, 1989, pp. 1-59.
- [5] F. Roozeboom, "Introduction: History and Perspectives of RTP," in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [6] P. Singer, "Rapid thermal processing: A progress report," in *Semiconductor International*, May 1993, pp. 64-69.
- [7] H. A. Lord, "Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven," *IEEE Trans. Semicon. Manuf.*, vol. 1, pp. 105, 1988.
- [8] R. Kakoschke, E. Bußmann, and H. Foll, "The appearance of spatially nonuniform temperature distributions during rapid thermal processing," *Appl. Phys. A*, vol. 52, pp. 52, 1991.
- [9] S. A. Campbell, K.-H. Ahn, K. L. Knutson, B. Y. H. Liu, and J. D. Leighton, "Steady state thermal uniformity and gas flow patterns in a rapid thermal processing chamber," *IEEE. Trans. Semicon. Manuf.*, vol. 4, pp. 14, 1991.
- [10] S. A. Campbell and K. L. Knutson, "Transient effects in rapid thermal processing," *IEEE Trans. Semicon. Manuf.*, vol. 5, pp. 302, 1992.
- [11] A. Kersch, H. Schafer, and C. Werner, "Improvement of thermal uniformity of RTP-CVD equipment by application of simulation," *IEDM Technical Digest*, pp. 883, 1991.

- [12] K. L. Knutson, S. A. Campbell, and F. Dunn, "Modelling of three dimensional effects on thermal uniformity in rapid thermal processing of 8 inch wafers," *IEEE. Trans. Semicon. Manuf.*, vol. 7, pp. 68, 1994.
- [13] S. Chatterjee, I. Trachtenberg, and T. F. Edgar, "Mathematical modelling of a single-wafer rapid thermal reactor," *J. Electrochem. Soc.*, vol. 139, pp. 3682, 1992.
- [14] T. P. Merchant, J. V. Cole, K. L. Knutson, J. P. Hebb, and K. F. Jensen, "A systematic approach to simulating Rapid Thermal Processing systems," *J. Electrochem. Soc.*, vol. 143, pp. 2035, 1996.
- [15] K. F. Jensen, T. P. Merchant, J. V. Cole, J. P. Hebb, K. L. Knutson, and T. G. Mihopoulos, "Modeling Strategies for Rapid Thermal Processing: Finite Element and Monte Carlo Methods," in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [16] P. Vandenabeele, K. Maex, and R. De Keersmaecker, "Impact of patterned layers on temperature nonuniformity during rapid thermal processing," *Mat. Res. Soc. Proc.*, vol. 146, pp. 149, 1989.
- [17] Z. Nenyeyi, A. Tillmann, and J. Gelpey, "Radiation incidence engineering in rapid thermal processing," presented at Second International Rapid Thermal Processing Conference, Monterrey, CA, 1994.
- [18] R. Kakoschke and E. Bußmann, "Simulation of temperature effects during rapid thermal processing," *Mat. Res. Proc.*, vol. 146, pp. 473, 1989.
- [19] B. Feil, B. Drew, and B. Moench, "Pattern-induced pattern misregistration after BPSG RTA reflow," *Proceedings of the 1st International Rapid Thermal Processing Conference*, pp. 114, 1993.
- [20] J. F. Buller, M. Farahani, and S. Garg, "RTA induced overlay errors in a global alignment stepper technology," *Proceedings of the 2nd International Rapid Thermal Processing Conference*, pp. 52, 1994.

- [21] P. Vandenabeele and K. Maex, "Temperature non-uniformity during rapid thermal processing of patterned wafers," *Proc. SPIE*, vol. 1189, pp. 89, 1989.
- [22] A. Kersch and T. Schaufbauer, "3D simulation and optimization of an RTO chamber with Monte Carlo heat transfer in comparison with experiments," *Proceedings of the 4th International Rapid Thermal Processing Conference*, pp. 347, 1996.
- [23] J. P. Hebb and K. F. Jensen, "The effect of multilayer patterns on temperature uniformity during rapid thermal processing," *J. Electrochem. Soc.*, vol. 143, pp. 1142, 1996.
- [24] S. A. Norman, "Optimization of transient temperature uniformity in RTP systems," *IEEE Trans. Electron Dev.*, vol. 39, pp. 205 - 207, 1992.
- [25] T. Breedijk, T. F. Edgar, and I. Trachtenberg, "A model predictive controller for multivariable temperature control in rapid thermal processing," *Proc. Amer. Control Conf.*, pp. 2980, 1993.
- [26] C. Schaper, "Real time control of rapid thermal processing semiconductor manufacturing equipment," *Proc. Amer. Control Conf.*, pp. 2985, 1993.
- [27] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Modeling, identification, and control of rapid thermal processing systems," *J. Electrochem. Soc.*, vol. 141, pp. 3200, 1994.
- [28] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Control of MMST RTP: Repeatability, uniformity, and integration of flexible manufacturing," *IEEE Trans. Semicon. Manuf.*, vol. 7, pp. 202, 1994.
- [29] G. Aral, T. P. Merchant, J. V. Cole, K. L. Knutson, and K. F. Jensen, "Concurrent engineering of a RTP reactor: Design and Control," *Proceedings of RTP-'94*, pp. 288, 1994.
- [30] C. R. Kleijn, "Chemical Vapor Deposition Processes," in *Computational Modeling in Semiconductor Processing*, M. Meyyappan, Ed. Norwood, MA: Artech House, 1994, pp. 97 - 229.
- [31] K. F. Jensen, D. I. Fotiadis, and T. J. Mountziaris, "Detailed models of the MOVPE

- process,” *J. Cryst. Growth*, vol. 107, pp. 1 - 11, 1991.
- [32] K. F. Jensen, E. O. Einset, and D. I. Fotiadis, “Flow phenomena in chemical vapor deposition of thin films,” *Ann. Rev. Fluid Mech.*, vol. 23, pp. 197 - 232, 1991.
- [33] S. Patnaik, R. A. Brown, and C. A. Wang, “Hydrodynamic dispersion in rotating-disk OMVPE reactors: Numerical simulation and experimental measurements,” *J. Cryst. Growth*, vol. 96, pp. 153 - 174, 1989.
- [34] D. I. Fotiadis, S. Kieda, and K. F. Jensen, “Transport phenomena in vertical reactors for metalorganic vapor phase epitaxy: I. effects of heat transfer characteristics, reactor geometry, and operating conditions,” *J. Cryst. Growth*, vol. 102, pp. 441 - 470, 1990.
- [35] K. F. Jensen and D. B. Graves, “Modeling and analysis of low pressure CVD reactors,” *J. Electrochem. Soc.*, vol. 130, pp. 1950 - 1957, 1983.
- [36] C. R. Kleijn, “A mathematical model of the hydrodynamics and gas-phase reactions in silicon LPCVD in a single-wafer reactor,” *J. Electrochem. Soc.*, vol. 138, pp. 2190 - 2200, 1991.
- [37] T. P. Merchant, “Modeling of Rapid Thermal Processes,” in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [38] R. W. Hamming, *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill, 1973.
- [39] C. A. J. Fletcher, *Computational Galerkin methods*. New York: Springer-Verlag, 1984.
- [40] B. G. Galerkin, *Vestnik Inzhenerov i Tekhnikov*, vol. 19, pp. 897-908, 1915.
- [41] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore and London: The Johns Hopkins University Press, 1989.
- [42] O. C. Zienkiewicz, *The Finite Element Method*. New York: McGraw Hill, 1977.
- [43] E. Bacry, S. Mallat, and G. Papanicolaou, “A wavelet based space-time adaptive numerical method for partial differential equations,” in *Courant Institute preprint*.
- [44] C. Canuto, M. Y. Hussaini, A. Quaderonic, and T. A. Zang, *Spectral Methods in Fluid Dynamics*. New York: Springer, 1988.

- [45] W. S. Wyckoff, "Numerical Solution of Differential Equations through Empirical Eigenfunctions," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [46] M. Loève, "Fonctions aleatoires de second ordre," *Comptes Rendus des Seances de l'Academie des Sciences*, vol. 220, 1945.
- [47] K. Karhunen, "Zur spektraltheorie stochastischer prozesse," *Annales Academiae Scientiarum Fennicae, Series A*, vol. 1, pp. 34, 1946.
- [48] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annu. Rev. Fluid Mech.*, vol. 25, pp. 539 - 575, 1993.
- [49] D. D. Kosambi, "Statistics in function space," *J. Indian Math. Soc.*, vol. 7, pp. 76 - 88, 1943.
- [50] V. S. Pougachev, "General theory of the correlations of random functions," *Izv. Akad. Nauk. SSSR, Ser. Mat.*, vol. 17, pp. 1401, 1953.
- [51] A. M. Obukhov, "Statistical description of continuous fields," *Tr. Geophys. Int. Akad. Nauk. SSSR*, vol. 24, pp. 3 - 42, 1954.
- [52] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, "The dynamics of coherent structures in the wall region of a turbulent boundary layer," *J. Flu. Mech.*, vol. 192, pp. 115 - 173, 1988.
- [53] R. W. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography*. Amsterdam: Elsevier, 1988.
- [54] J. E. Jackson, *A user's guide to principal components*. New York: John Wiley & Sons, Inc., 1991.
- [55] F. M. Selten, "Toward an optimal description of atmospheric flow," *Journal of the Atmospheric Sciences*, vol. 50, pp. 861 - 877, 1993.
- [56] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.
- [57] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.

- [58] V. R. Algazi and D. J. Sakrison, "On the optimality of the Karhunen-Loève expansion," *IEEE Trans. Inform. Theory*, vol. 15, pp. 319 - 321, 1969.
- [59] C. A. Andrews, J. M. Davies, and G. R. Schwartz, "Adaptive data compression," *Proc. IEEE*, vol. 55, pp. 267 - 277, 1967.
- [60] L. Sirovich, "Turbulence and the Dynamics of Coherent Structures: I, II and III," *Quarterly of Applied Mathematics*, vol. XLV, pp. 561, 1987.
- [61] L. Sirovich, "Chaotic dynamics of coherent structures," *Physica D*, vol. 37, pp. 126, 1989.
- [62] L. Sirovich and H. Park, "Turbulent thermal convection in a finite domain: Part I. Theory," *Phys. Fluids A*, vol. 2, pp. 1649, 1990.
- [63] L. Sirovich, "Empirical eigenfunctions and low dimensional systems," in *New Perspectives in Turbulence*, L. Sirovich, Ed. New York: Springer, 1991.
- [64] L. Sirovich, J. D. Rodriguez, and B. Knight, "Two boundary value problem for Ginzburg Landau equation," *Physica D*, vol. 43, pp. 63, 1990.
- [65] D. H. Gay and W. H. Ray, "Identification and control of distributed parameter systems by means of the singular value decomposition," *Chemical Engineering Science*, vol. 50, pp. 1519 - 1539, 1995.
- [66] M. D. Graham and I. G. Kevrekidis, "Alternative approaches to the Karhunen-Loève decomposition for model reduction and data analysis," *Computers Chem. Engng*, vol. 20, pp. 495 - 506, 1996.
- [67] J. L. Lumley, in *Transition and Turbulence*, R. E. Meyer, Ed. New York: Academic, 1981, pp. 215.
- [68] N. Aubry, R. Guyonnet, and R. Lima, "Spatio-temporal analysis of complex signals: theory and applications," *J. stat. Phys.*, vol. 64, pp. 683 - 739, 1991.
- [69] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag, "Low-dimensional models for complex geometric flows: application to grooved channels and circular cylinders," *Phys. Fluids A*, vol. 3, pp. 2337 - 2354, 1991.

- [70] H. Park and L. Sirovich, "Turbulent thermal convection in a finite domain: Part II. Numerical results," *Phys. Fluids A*, vol. 2, pp. 1659, 1990.
- [71] C.-C. Chen, E. E. Wolf, and H.-C. Chang, "Low-dimensional spatiotemporal thermal dynamics on nonuniform catalytic surfaces," *J. Phys. Chem.*, vol. 97, pp. 1055 - 1064, 1993.
- [72] M. D. Graham, S. L. Lane, and D. Lus, "Proper orthogonal decomposition analysis of spatiotemporal temperature patterns," *J. Phys. Chem.*, vol. 97, pp. 889 - 894, 1993.
- [73] K. Krischer, R. Rico-Martinez, I. G. Kevrekidis, H. H. Rotermund, G. Ertl, and J. L. Hudson, "Model identification of a spatiotemporally varying catalytic reaction," *A. I. Ch. E. J.*, vol. 39, pp. 89 - 98, 1993.



## **Chapter 2**

# **Model Reduction Strategies for Low Pressure Rapid Thermal Processing Systems**

### **2.1 INTRODUCTION**

Processes used to manufacture semiconductor devices are becoming increasingly complex, while competition demands that these devices be brought to market more quickly and manufactured more reliably. This calls for reduction in the large number of cut-and-try iterations in developing processes, process equipment, or process control software. In order to speed up this development process, one needs to understand the complicated physical rate processes governing each fabrication step. Such an understanding is best expressed in terms of a detailed, physically based, mathematical model. However, the solution of such a model is often time consuming and requires the use of hardware and software resources beyond those available to typical manufacturing organizations because of the complex time dependent and three-dimensional nature of the production equipment. Simulations of these processes using the existing computational models can take hours to days to yield results. Therefore, techniques are required for deriving low-order, physically based models for semiconductor manufacturing processes. These models could be used to study on-line process variations or to answer “what-if” type of questions under a limited range of conditions. The reduced complexity and smaller

computational storage requirements imply that the reduced models can be simulated on desktop computers (such as PCs), besides workstations. Hence, process engineers and operators could use these models for a better understanding of semiconductor manufacturing processes. A well-designed reduced model could help in cutting down the number of experiments required in designing a process recipe and thus reduce the transition time in bringing a process from the research to the manufacturing stage in a fabrication line. Another use for such a model would be in advanced model based control strategies.

In this chapter, a model reduction technique has been studied, using a Rapid Thermal Processing (RTP) system as a test vehicle. RTP is an emerging technology in chip manufacturing processes and has shown promise in a wide variety of applications. A typical fabrication process may consist of as many as 26 different RTP steps of oxidation, annealing, nitridation and chemical vapor deposition. [1] The demand for submicron device sizes have placed severe constraints on the thermal processing of silicon wafers. To minimize solid state diffusion of dopants, the amount of time spent by the wafer close to processing temperature needs to be considerably reduced. RTP provides a viable alternative to existing thermal processing techniques. RTP systems are, in general, single wafer reactors. [2, 3] The wafer is heated by tungsten halogen filament lamps or by water-cooled arc lamps. The primary mode of heat transfer to the wafer is by radiation from the lamps. The wafer is typically supported by quartz pins or a silicon guard ring, so that the wafer temperature may be ramped at very high rates (~100 K/s). After processing, the wafer is ramped down quickly and the process gases are purged from the reactor using inert gases. The wafer processing time in an RTP reactor is very short, which minimizes diffusion lengths and preserves already formed dopant profiles from previous steps. The fast dynamics and transient nature of a RTP system make it a good choice for exploring the capabilities of the model reduction procedure.

The emerging nature of RTP technology drives the need for models, both reduced and complex, which would lead to a better understanding of the process. A number of model based control studies of RTP systems have been developed, [4-8] but, further advances in control

design would need accurate models capable of simulating the process in real time (or faster). In this work we have developed nonlinear low order models without approximating the physical conservation equations describing the process, thereby making them more accurate compared to conventional linear models over a wider range of conditions. Such models show promise for application in the development of model based control schemes. [9]

## **2.2 MODEL FORMULATION**

### **2.2.1 DESCRIPTION OF THE REACTOR**

The geometry for the two-dimensional axisymmetric reactor used for the model reduction studies was developed by Hebb *et al.*[10] and has similar characteristics compared to commercial systems currently being used in the industry. Figure 2.1 depicts an axisymmetric drawing of the reactor. Only one-half of the chamber is depicted due to radial symmetry. The lamphouse consists of five concentric toroidal lamps with a gold reflector assembly at the top. Below the lamp house is the region through which process gases are introduced into the chamber known as the showerhead. A quartz window separates the showerhead region from the lamphouse assembly. The process gases are introduced on top of the wafer through a honeycomb shaped quartz plate. For all simulations, in this chapter, nitrogen is the process gas introduced at a flow rate of 5 slm and at an operating pressure of 0.01 atmospheres. The wafer is supported on a silicon guard ring and rotated at 20 rpm. The lamps have a radius of 1 mm and the wafer has a radius of 100mm.

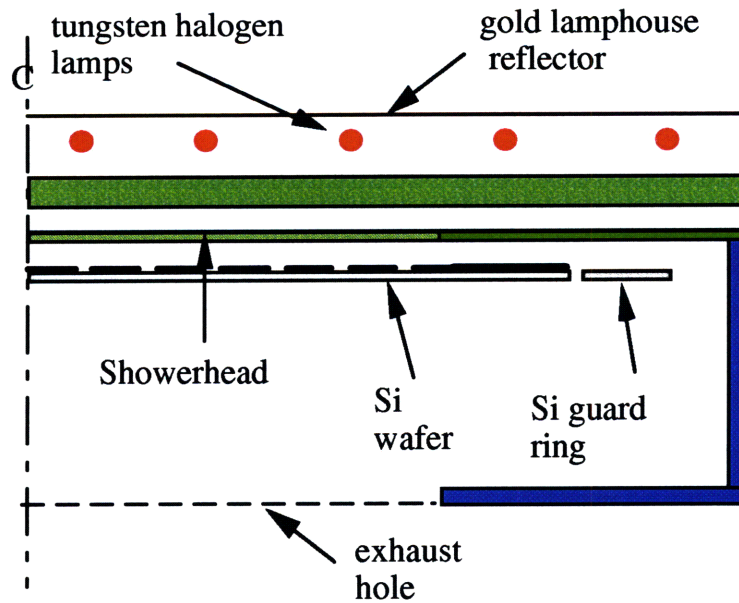


Figure 2.1 Schematic representation of axisymmetric RTP reactor.

## 2.2.2 MODEL REDUCTION APPROACH

The model reduction strategy is shown schematically in Figure 2.2. A detailed, physical finite element model of a generic two-dimensional (2D) RTP system [11, 12] with features relevant to the next generation of RTP systems serves as the base case for the model reduction study. The modeling strategy used to generate this detailed model is similar to that used in previous simulations of RTP systems. [11, 12] It is based on a finite element (FEM) solution of the general equations representing conservation of mass, momentum and energy. The boundary condition of the energy equation describes the radiation heat transfer, which separates the thermal radiation into multiple wavelength bands and includes the effect of multiple reflections. In the present case studies the velocity field is assumed to be constant through the RTP cycle, fixed at a steady state solution at the process hold conditions, a reasonable approximation at low pressures. [11, 12] At higher pressures transient flow effects must be included, but the general model reduction strategy will remain the same.

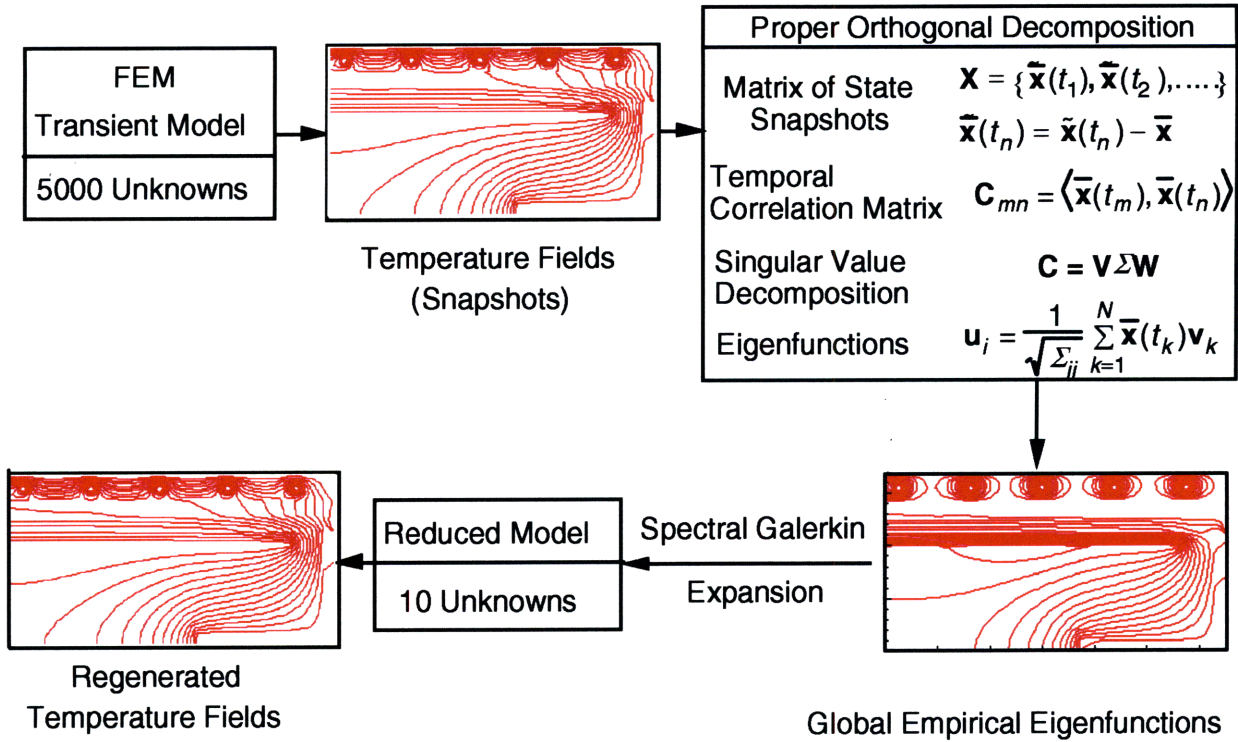


Figure 2.2 Schematic representation of the model reduction method.

The modeling equations are solved by the Galerkin finite element method. [11, 12] In this method the unknown flow and temperature fields are approximated by expansions in piecewise, low order polynomials. This approach has the advantage of being general and flexible, but the large number of coefficients required leads to large nonlinear matrix problems. The numerical solution of this problem therefore requires workstations and special computational algorithms. The number of coefficients involved in representing the temperature fields could, in principle, be reduced if the approximating functions were similar in form to the actual solution. One approach for obtaining better approximating functions is the proper orthogonal decomposition (POD) method (also known as the Karhunen-Loève procedure). [13,

14] This method was first suggested by Lumley[15] as a rational procedure for the extraction of *coherent structures*. [13] In this method, empirical eigenfunctions can be extracted from either experimental observation or detailed model predictions of temperature fields ("snapshots") for the entire reactor at discrete time intervals. The method of eigenfunction extraction starts with a matrix of transient temperature fields generated by the finite element model at discrete time intervals.

$$X = \{\tilde{\mathbf{x}}(t_1), \tilde{\mathbf{x}}(t_2), \dots\} \quad (1)$$

$$\tilde{\mathbf{x}}(t_n) = \mathbf{x}(t_n) - \bar{\mathbf{x}} \quad (2)$$

where  $\tilde{\mathbf{x}}(t_n)$  is the transient temperature field extracted at time  $t_n$  and  $\bar{\mathbf{x}}$  is the steady state temperature field. For generating these temperature fields, the transient FEM model of the RTP reactor is run with a set of lamp powers till the wafer attains a steady temperature,  $\bar{\mathbf{x}}$ . After the wafer has attained a steady state, the lamp powers are individually perturbed to generate variations in the wafer temperature. The temperature fields obtained from these lamp power perturbations are then stored in the matrix  $X$ . A temporal correlation matrix is subsequently constructed from the snapshots as follows,

$$C_{mn} = \langle \tilde{\mathbf{x}}(t_m), \tilde{\mathbf{x}}(t_n) \rangle \quad (3)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product in the  $\ell_2$  norm. The eigenfunctions  $\mathbf{u}_i$  are obtained from a singular value decomposition of the temporal correlation matrix,

$$C = V \Sigma W \quad (4)$$

$$\mathbf{u}_i = \left( \frac{1}{\sqrt{\Sigma_{ii}}} \right) \sum_{k=1}^N \tilde{\mathbf{x}}(t_k) \mathbf{v}_k \quad (5)$$

where  $V$  is a matrix whose columns are the left singular vectors of  $C$  and  $\Sigma$  is a diagonal matrix with the singular values of  $C$  on the diagonal. Therefore the eigenfunctions are admixtures of the snapshots. [16, 17] The number of eigenfunctions determined from this technique is equal to the dimension of the square temporal matrix,  $C$ . These eigenfunctions form an optimal basis set for the given series of snapshots. [18] The remaining eigenfunctions, for the series of snapshots, are not uniquely determined. The only requirement on them is that they be orthogonal to the already determined set, and hence orthogonal to the snapshots  $v_k$ . The empirical eigenfunction set generated by this technique can be used to regenerate a series of temperature fields by projecting a suitable set of temporal coefficients,  $a_i(t)$ , on the eigenfunction basis set as shown below,

$$\hat{x}(t) = \bar{x} + \sum_{i=1}^N a_i(t)u_i \quad (6)$$

In order to solve for this set of temporal coefficients,  $a_i(t)$ , we have to integrate an initial value problem for a group of ordinary differential equations (ODEs). This set of ODEs, which is the low order reduced model, is obtained by the procedure discussed in the following section.

### 2.2.3 METHOD FOR GENERATING NONLINEAR REDUCED ORDER MODELS

Figure 2.3 compares a typical temperature snapshot obtained from the transient FEM model to the dominant eigenfunction extracted from the snapshots by the POD method discussed above. The dominant eigenfunction, *i.e.* the eigenfunction corresponding to the largest singular value, has most of the qualitative information about the temperature field. This can be seen from the figure where the contours of the temperature field closely match those of the dominant eigenfunction. Hence, the empirical eigenfunctions may be viewed as ideal fitting functions to

be used in a pseudospectral[19] Galerkin procedure. [18]

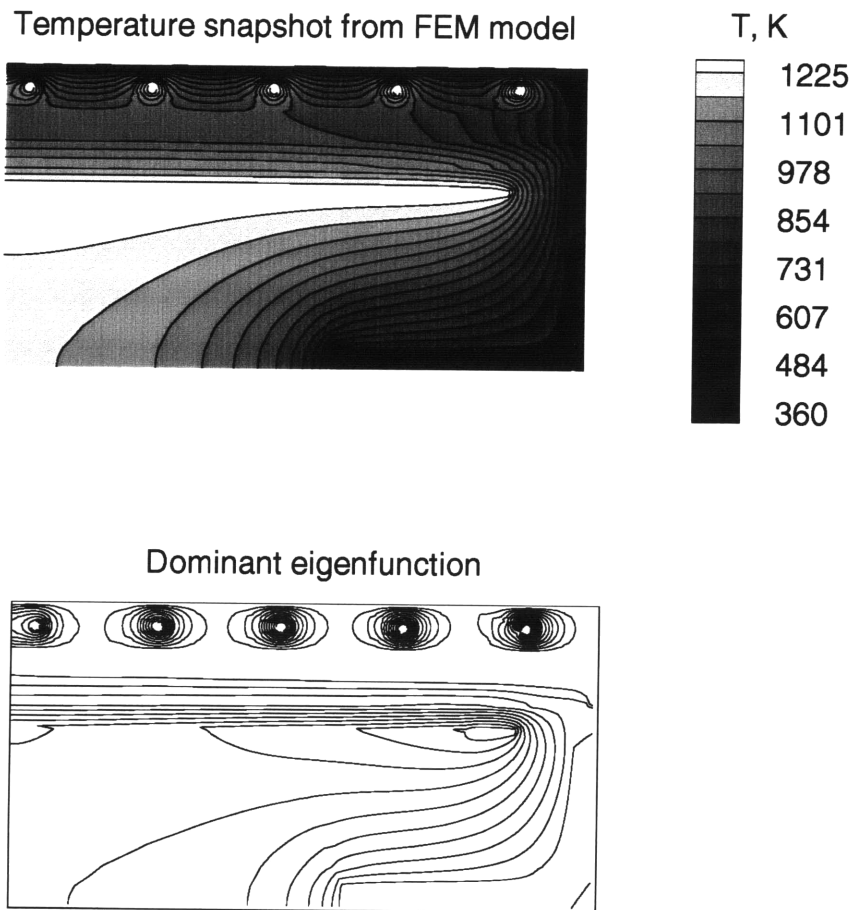


Figure 2.3 Comparison of a typical temperature field with the dominant eigenfunction.



In general, for a set of differential equations on one variable,  $\mathbf{y}$ , expressed as

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\mathbf{y}) \quad (7)$$

the pseudospectral Galerkin procedure is given by

$$\left\langle \mathbf{u}_i, \frac{d\mathbf{y}}{dt} \right\rangle = \left\langle \mathbf{u}_i, \mathbf{F}(\mathbf{y}) \right\rangle, i = 1, \dots, N \quad (8)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product,  $\mathbf{u}_i$  represents the eigenfunctions, and  $N$  is the number of eigenfunctions used in the pseudospectral Galerkin procedure. This procedure, using empirical eigenfunctions, has been applied to modeling turbulence and large-scale problems in fluid mechanics. [16, 20-23]

The general method of expressing the FEM model in a form amenable to model reduction is given below. The important idea is to separate the terms linear and nonlinear in temperature so that they can be handled separately. The conservation of energy equation in the FEM model takes the following form:

$$\rho C_p \left[ \frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right] = \nabla \cdot [k^f \nabla T] \quad (9)$$

where  $\rho$  is the density,  $C_p$  is the specific heat,  $\mathbf{v}$  is the velocity vector,  $T$  is the temperature and  $k^f$  is the fluid thermal conductivity. The density and specific heat of the gas phase are modeled as temperature dependent properties in the FEM model. The solid thermal properties, except for the thermal conductivity of the silicon wafer, are constant in the model. The boundary condition for Equation 9 takes the following form[11, 12]

$$k^s \nabla T_i \cdot \mathbf{n} = k^f \nabla T_i \cdot \mathbf{n} + \alpha_i \sum_{l=1}^{N_{lamp}} R_{il} P_l + \sigma \sum_{k=1}^{N_{bands}} \left[ \alpha_i^k \sum_{j=1}^{N_{SW}} \phi_{\lambda^k - T_j} \varepsilon_j^k R_{ij}^k T_j^4 - \phi_{\lambda^k - T_i} \varepsilon_i^k T_i^4 \right] \quad (10)$$

The left-hand side represents the conduction into the solid. This is balanced on the right hand side by conduction in the gas, energy input from the lamps, and energy transfer with other surfaces in the system. In Equation 10,  $k^s$  is the solid thermal conductivity,  $\alpha$  is the absorptance of the solid surface,  $P_l$  is the radiation intensity of lamp  $l$ ,  $\sigma$  is the Stefan-Boltzmann constant,  $\varepsilon_j^k R_{ij}^k$  is the percentage of radiation, in band  $k$ , leaving surface  $i$  which is absorbed by surface  $j$  (by direct viewing and all intervening reflections). The exchange factors,  $R_{ij}^k$ , are assumed to be temperature independent based on the high temperature opaque silicon properties. [11, 12] This has been shown to be a reasonable approximation for RTP processes. [11, 12]

The gas in the lamphouse and in the region between the showerhead and the quartz window is treated as stagnant. Therefore the gradients in temperature in these regions are determined by gas phase conduction. There are additional boundary conditions at the fluid-solid interfaces on the exterior walls of the reactor that represent heat transfer to the surrounding ambient. Using the Galerkin finite element method, Equation 9 is transformed to a set of algebraic equations as follows,

$$\int_D \rho C_p \left[ \frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right] \Phi^i dV = - \int_D k^f \nabla T \cdot \nabla \Phi^i dV + \int_{\partial D} k^f \nabla T \cdot \mathbf{n} \Phi^i dS \quad (11)$$

where  $\Phi^i$  are the piecewise continuous basis functions used in the finite element method,  $D$  represents the volume of the domain and  $\partial D$  is the boundary of the domain. [11, 12] The boundary condition shown in Equation 10 is evaluated as part of the boundary integral in Equation 11.

In order to make the conservation of energy equation (Equation 9) amenable to the model reduction technique, the terms in the finite element expansion (Equation 11) are lumped together and expressed in the following matrix form

$$M(\tilde{\mathbf{x}}) \frac{d\tilde{\mathbf{x}}}{dt} = C(\tilde{\mathbf{x}})\tilde{\mathbf{x}} + RD(\tilde{\mathbf{x}}) + \alpha_i \sum_{l=1}^{N_{lamp}} R_{il} P_l \quad (12)$$

where,  $RD(\tilde{\mathbf{x}})$ , is the nonlinear radiation heat transfer contribution to the reduced model and can be written as

$$RD(\tilde{\mathbf{x}}) = \sigma \sum_{k=1}^{N_{bands}} [ \alpha_i^k \sum_{j=1}^{N_{SW}} \phi_{\lambda^k - T_j} \varepsilon_j^k R_{ij}^k T_j^4 - \phi_{\lambda^k - T_i} \varepsilon_i^k T_i^4 ] \quad (13)$$

$M(\tilde{\mathbf{x}})$  is obtained by lumping all the dynamic contribution from the energy conservation equation, and  $C(\tilde{\mathbf{x}})$  is obtained by lumping all the convection and conduction terms from the energy conservation equation.

This separates the nearly linear conduction and convection terms in the matrix  $C(\tilde{\mathbf{x}})$  from the highly nonlinear radiation terms in the matrix  $RD(\tilde{\mathbf{x}})$ . Thus, the temperature dependence of material properties, such as the gas phase thermal conductivity,  $k^f$ , gas phase specific heat,  $C_p$  etc. can be linearized and included in the matrices  $M(\tilde{\mathbf{x}})$  and  $C(\tilde{\mathbf{x}})$ . In the actual FEM model, these material properties are expressed as power law fits which are weakly nonlinear compared to the terms in the radiation heat exchange. Since the reduced model is extracted using deviation eigenfunctions, the models extracted would be exact around the given steady state and would differ from the FEM model around other operating conditions depending on the nonlinear effects of the material properties.

The method of extracting nonlinear reduced order models is implemented in deviation variables, *i.e.* the steady state temperature field is subtracted from the transient temperature fields and the eigenfunctions are extracted from the deviation fields. This eliminates any steady state offset completely, if one generates the reduced model from small perturbations about a given steady state. The  $T^4$  nonlinearity in the radiation heat exchange term prevents a linearization of the model equations from being valid over a broad range of conditions.

Therefore, this contribution to the reduced model has to be evaluated by reconstructing the temperature fields, generating the  $T^4$  term explicitly in absolute temperatures, and then evaluating the radiation contribution to the reduced model at every time step.

The empirical eigenfunctions obtained from the POD method are used in a pseudospectral Galerkin expansion of Equation 12. The resulting low order ( $N < 10$ ) system of ordinary differential equations takes the form:

$$(\mathbf{u}_m^T \mathbf{M}(\bar{\mathbf{x}}) \sum_{i=1}^N \mathbf{u}_n \frac{da_i}{dt}) = (\mathbf{u}_m^T \mathbf{C}(\bar{\mathbf{x}}) \sum_{n=1}^N a_n(t) \mathbf{u}_n) + (\mathbf{u}_m^T \mathbf{R}) [\hat{\mathbf{x}}(t)]^4 + (\mathbf{u}_m^T [\alpha_i \sum_{l=1}^{N_{lamp}} R_{il} \tilde{P}_l + \mathbf{K}]) \quad (16)$$

where  $\mathbf{R}$  is the nonlinear radiation exchange term. The nonlinearities that arise from the temperature dependence of the emissivity, thermal conductivity, density and heat capacity are not explicitly accounted for at each time instant. Instead these properties are evaluated at the given steady state resulting in the matrices  $\mathbf{M}(\bar{\mathbf{x}})$  and  $\mathbf{C}(\bar{\mathbf{x}})$ . The matrix  $\mathbf{K}$  arises from the steady state contribution of the heat transfer to the ambient boundary conditions and the steady state part of the radiation term ( $RD(\bar{\mathbf{x}})$ ).

Equation 16 can be reformulated in matrix notation as

$$[\mathbf{U}^T \mathbf{M}(\bar{\mathbf{x}}) \mathbf{U}] \frac{d\mathbf{a}}{dt} = [\mathbf{U}^T \mathbf{C}(\bar{\mathbf{x}}) \mathbf{U}] \mathbf{a} + [\mathbf{U}^T \mathbf{R}] \{\hat{\mathbf{x}}(t)\}^4 + [\mathbf{U}^T \mathbf{G}] \tilde{\mathbf{P}} + [\mathbf{U}^T \mathbf{K}] \quad (17)$$

where  $\mathbf{U}$  is the matrix of eigenfunctions,  $\mathbf{a}$  is the temporal coefficient vector, and  $\mathbf{G}$  is the lamp power transformation matrix. This set of ordinary differential equations is then integrated using an initial value solver. In order to calculate the contribution,  $\{\hat{\mathbf{x}}(t)\}^4$ , the term,  $\hat{\mathbf{x}}(t)$ , is calculated at each time instant using Equation 6. An alternate method for computing  $\{\hat{\mathbf{x}}(t)\}^4$ , explicitly in terms of the eigenfunctions, is described in Appendix A.

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 STEADY STATE PERFORMANCE OF REDUCED MODELS

The method outlined above was used to obtain reduced models with 10 unknowns from the FEM model with 5060 unknowns. The reduced models showed excellent agreement with the FEM model at steady state operating conditions and for local perturbations around those operating conditions. The FEM model uses a two-band approximation for the partial transmission by quartz in different wavelength ranges. The quartz is treated as transparent for wavelengths shorter than 4  $\mu\text{m}$  and opaque for wavelengths longer than 4  $\mu\text{m}$ . [11, 12] The principal source of deviation between the reduced and FEM models proved to be the nonlinear function which decides the fraction of radiation in each of the two wavelength bands.

In order to arrive at this conclusion, a reduced model was extracted using lumped band radiation properties. In this reduced model, referred to elsewhere in this chapter as the ‘lumped band reduced model’, there is a single matrix ( $\mathbf{R}$  in Equation 17) which accounts for the total radiation contribution. In the other type of reduced model, the explicit two-band formulation from the FEM model is retained. In this type of reduced model, referred to as the ‘explicit two-band reduced model’, there are two matrices ( $\mathbf{R}_1$  and  $\mathbf{R}_2$ ) which separately account for the radiation contributions in the two wavelength bands. The fraction of the radiation contribution in each of the two bands is read dynamically from a look up table indexed to temperature. In the lumped band reduced model the fractions are the same as those at the steady state at which the reduced model is extracted. The wafer center temperatures predicted by the lumped band reduced model and the explicit two-band reduced model are compared to those predicted by the FEM transient model in Figure 2.4. Both the reduced models were extracted at a steady state where the wafer temperature was at 1300 K, so that the properties in the matrices  $\mathbf{M}(\bar{\mathbf{x}})$  and  $\mathbf{C}(\bar{\mathbf{x}})$  in Equation 16 were for that steady state. Both the reduced models predict the temperature perturbations at 1300K steady state operating conditions with reasonable accuracy.

As can be seen from the figure, the temperature difference between the reduced and FEM models is within 2K. The explicit two-band reduced model predicts the wafer center temperature more accurately at other steady state operating conditions, when compared to the lumped band reduced model. Hence, in the rest of the study, the explicit two-band reduced model was used and is referred to as the 'reduced model'.

The most nonlinear term in the conservation of energy equation, other than the radiation boundary condition, is the inverse of temperature appearing in the gas density. In an attempt to further improve the accuracy of the explicit two-band reduced model, this term was linearized about the steady state. However, this change gave no improvement in the agreement of the reduced and FEM models because the wafer, quartzware, and walls provide the majority of the system mass, and these solids have a constant density in both the formulations. This leads to the conclusion that the deviation of the reduced model temperature trajectory from that predicted by the FEM model for other steady state operating conditions is due to the nonlinear variation of gas phase properties such as thermal conductivity and specific heat.

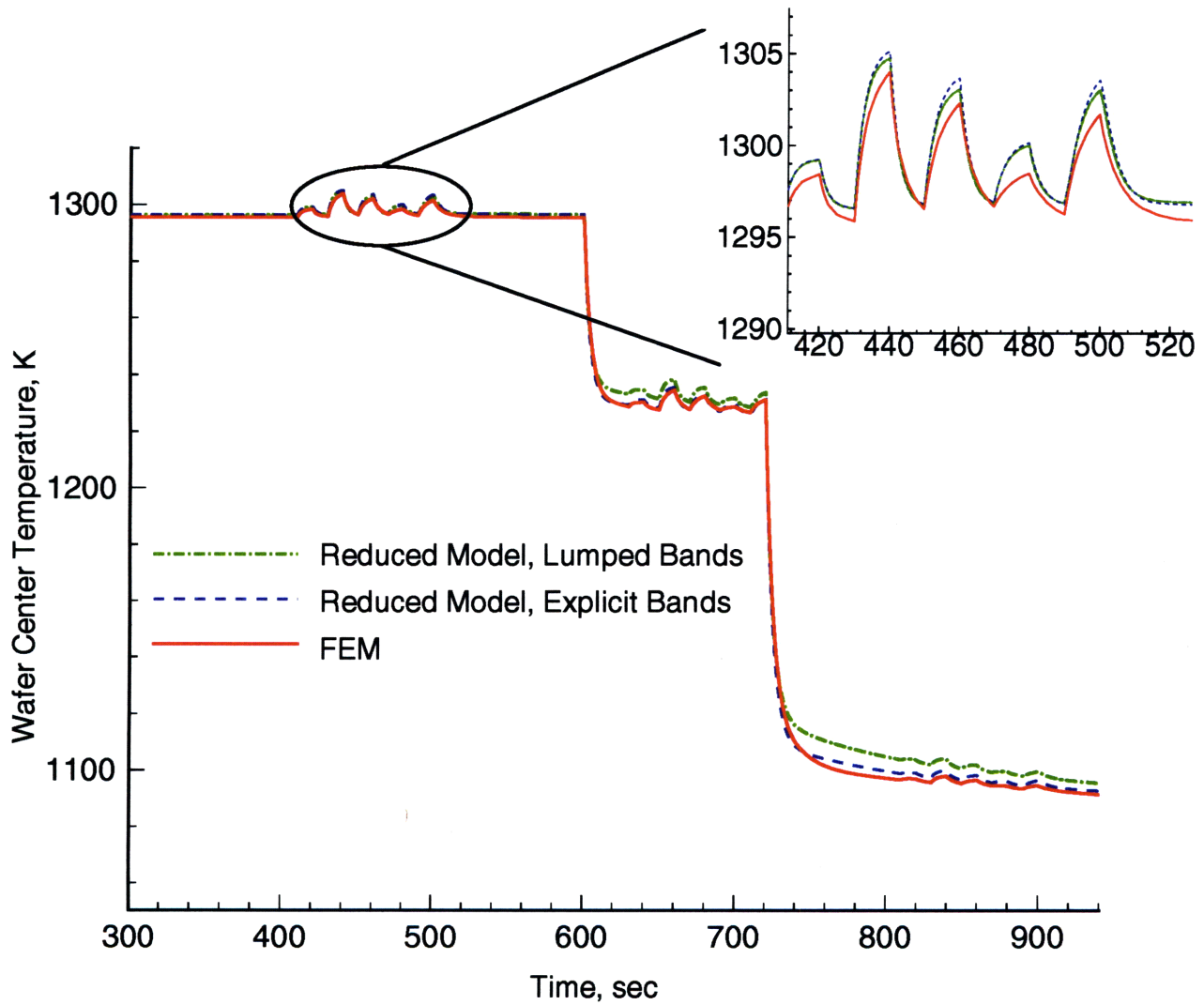


Figure 2.4 Comparison of the wafer center temperature of FEM and reduced models - a.) Lumped radiation model, and b.) Explicit two band radiation model.

### 2.3.2 VARIATION OF RMS ERROR WITH INCREASING MODEL ORDER

An important issue in extracting reduced models is deciding upon the number of eigenfunctions to be used in generating the reduced model. The fewer the number of eigenfunctions, the less accurate the reduced model is going to be when compared to the FEM model. On the other hand, a larger number of eigenfunctions would increase the complexity of the reduced model to an extent that it might be too slow to be used in real time process control or other applications. To study this problem, the lamp power was perturbed around a given steady state operating condition, giving rise to local temperature perturbations similar to those shown in Figure 2.4. The RMS error of the wafer temperature between the reduced and FEM model was calculated as follows

$$\text{Error} = \sqrt{\frac{\sum_{i=1}^N (T_{i,\text{Reduced}} - T_{i,\text{FEM}})^2}{N}} \quad (18)$$

where  $N$  denotes the number of points on the wafer surface over which the RMS error is evaluated. The RMS error was found over 15 points distributed over the wafer surface, and the results were plotted against the number of eigenfunctions as shown in Figure 2.5. The error falls steeply till the introduction of the fifth eigenfunction. Following this, there are minor variations in the RMS error till the introduction of the tenth eigenfunction. The RMS error then settles down at approximately 1.1 after the tenth eigenfunction. The results show that a fifth order reduced model would be good enough for the model reduction strategy, however a tenth order reduced model was chosen to perform a rigorous analysis of the technique.



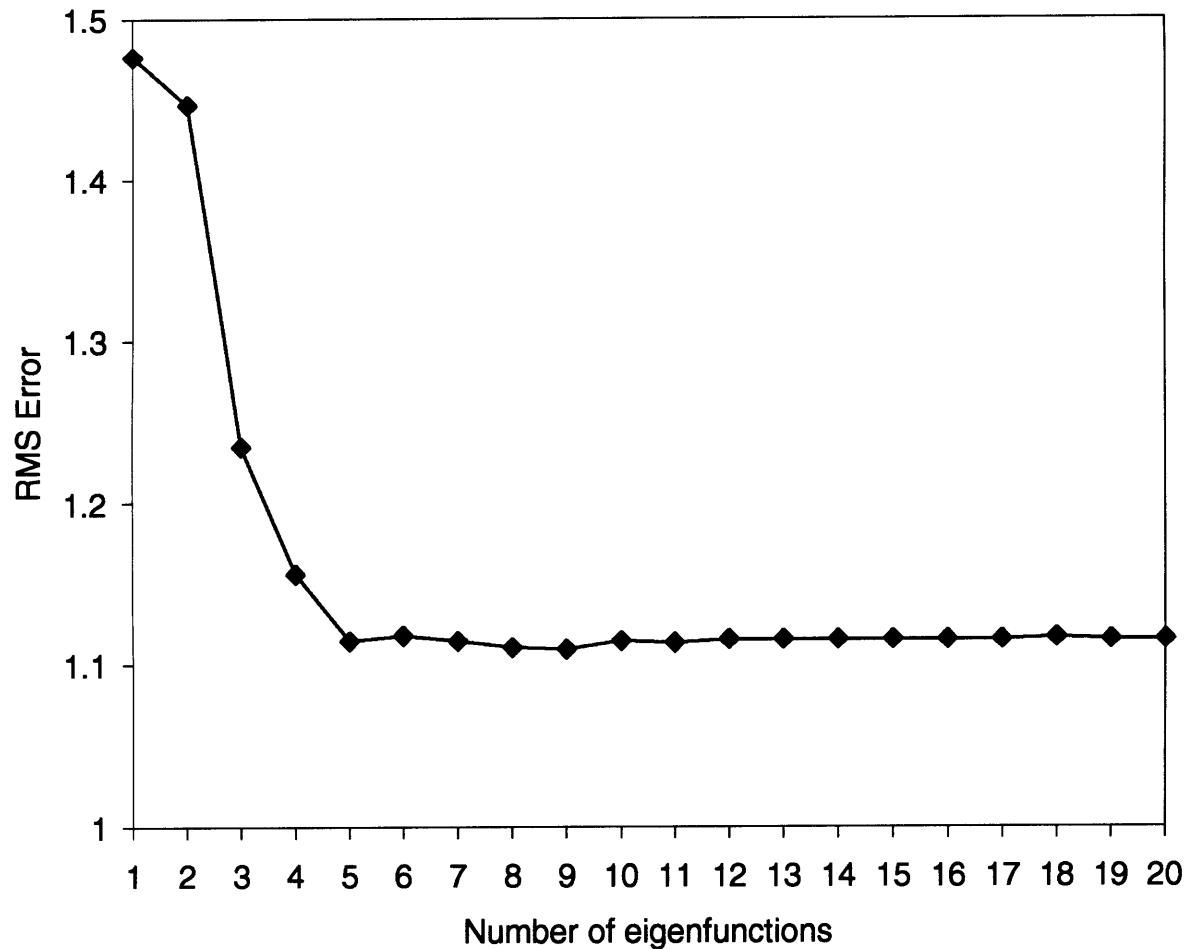


Figure 2.5 Variation of RMS error of wafer temperature with increasing number of eigenfunctions.

### 2.3.3 TRANSIENT BEHAVIOR OF REDUCED MODELS

After having studied the response of the reduced models at steady-state operating conditions, the next issue addressed was the performance of the reduced model in replicating the transient RTP cycle as generated by the FEM model. For this study, a suitable lamp power profile was designed so that the wafer temperature was ramped from 300K to 1300K at approximately 10 °C/second. After an initial stabilization of the numerical simulations for 250 seconds, the lamps are turned on and the wafer temperature is ramped from 300 K to 1300 K in

150 seconds and then held constant at 1300 K for 800 seconds. All the reduced models used to study the transient ramp response were explicit two-band reduced models. A typical RTP cycle is much shorter in duration than the present case study, but the larger cycle was chosen to explore the effect of any drifts which might be present in the reduced model, as they would be amplified over a length of time.

Figure 2.6 shows the comparison of the ramp response as shown by the FEM transient model and the reduced model extracted at a wafer steady state of 1300K. The reduced model attains a different steady state from that given by the FEM model when the lamps are kept at zero power. This is because the fluid properties in the reduced model are linear extrapolations from the 1300 K values instead of the nonlinear power law fits employed in the FEM model. As the lamp powers are ramped, the reduced model shows good agreement with the FEM model and finally attains the same steady state as the FEM model.

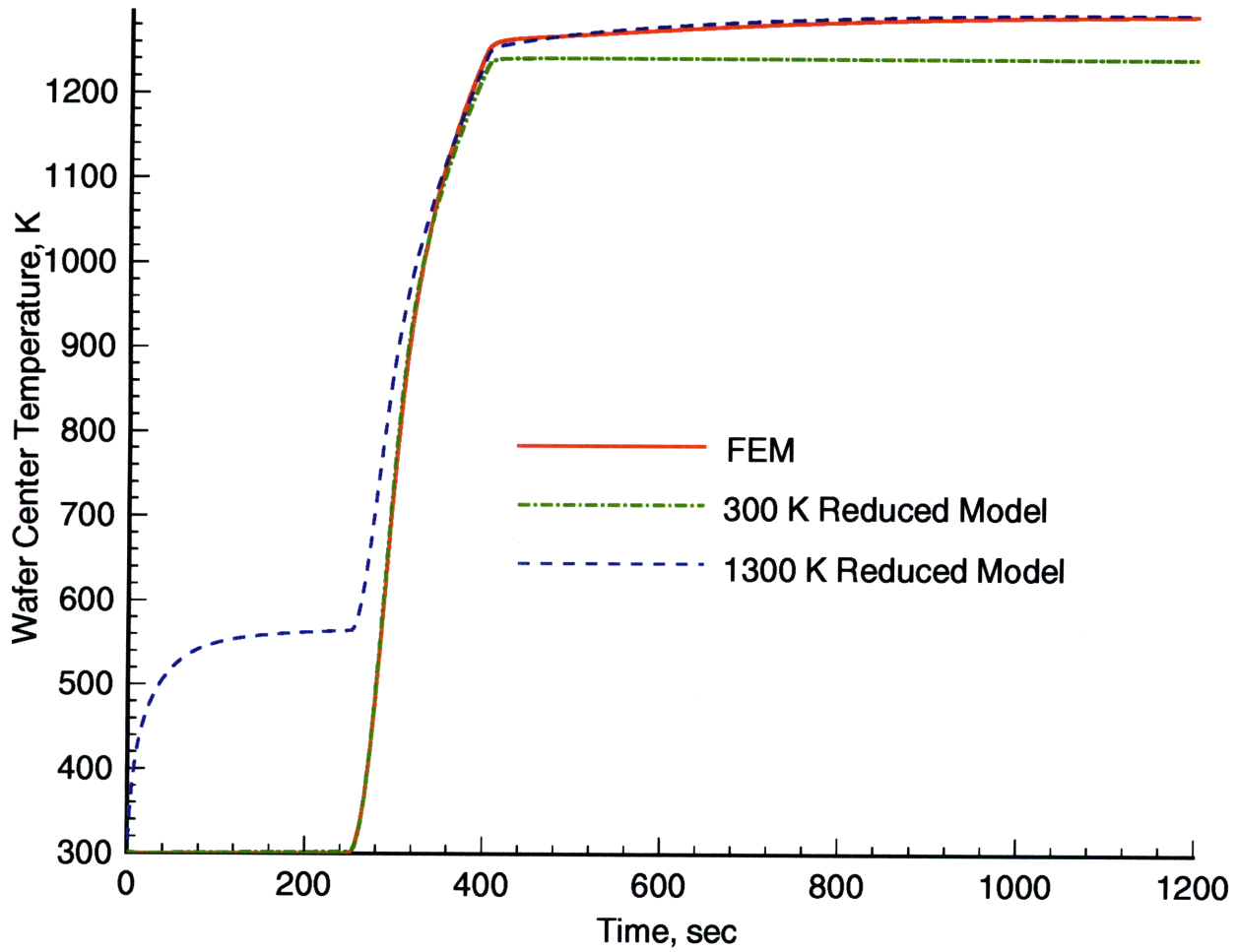


Figure 2.6 Behavior of wafer center temperature of FEM and reduced models, during transient ramp up and hold phases of the RTP cycle.

In order to further understand the performance of the reduced order modeling scheme, a model was extracted at a wafer steady state of 300K. This reduced model was then used to study the ramp response. The results are also shown in Figure 2.6. As seen from the figure, the initial steady state attained by the reduced model and the FEM model are the same. This is to be expected as the reduced model extracted at 300K has the same set of properties as the FEM model at 300K. This reduced model shows good agreement with the FEM model for the lower portion of the ramp, but deviates as the FEM model nears 1300K. Finally at the higher steady state the reduced model attains a different steady state from the one attained by the FEM model because the linearization of the fluid properties is inadequate at this temperature.

The temperature fields throughout the reactor at the higher temperature steady state, at 800 seconds into the cycle, are compared in Figure 2.7. The reduced model generated at 1300K shows excellent agreement with the full FEM model, as expected. In the 300K reduced model, the walls, showerhead, and the quartz window are cooler than both the 1300K reduced model and the FEM model. This is due to the fact that the thermal conductivity of the gas phase is over predicted by the 300K reduced model. This effect is much more predominant in the lamphouse and in the region between the quartz window and the showerhead because the fluid is treated as stagnant in these regions. As a result, the temperature gradients in these regions are determined by radiation and gas phase conduction. In other parts of the reactor this effect is not so evident due to the forced convection in the gas.

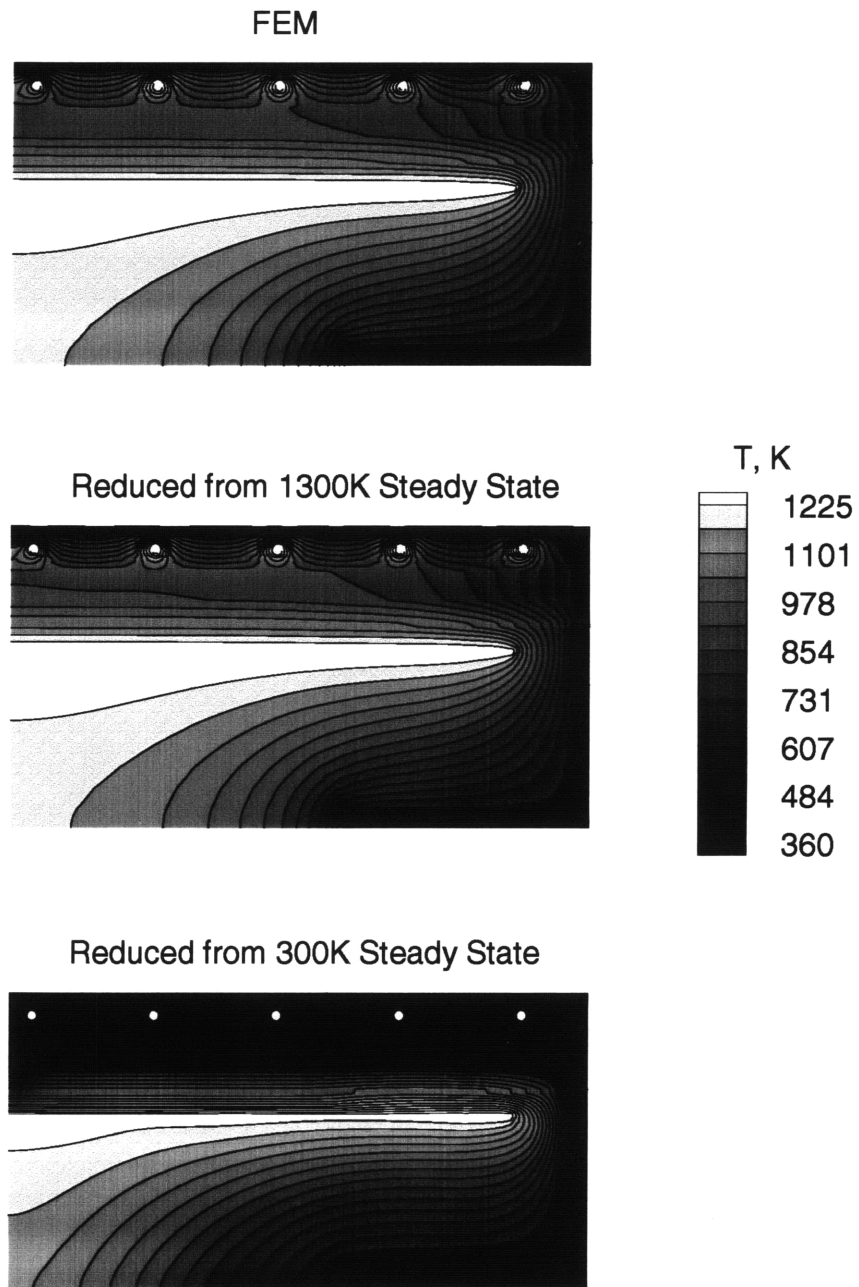


Figure 2.7 Temperature fields from FEM and reduced models during the hold phase.

These results show that at least two reduced order models need to be combined in order to replicate the FEM ramp response over the entire trajectory. The strategy devised in this regard was to start integrating with the reduced order model extracted at 300K, then switch to the reduced model extracted at 1300K when the wafer center temperature is at 1000K. The switching temperature of 1000K was chosen because this was the temperature at which the trajectories of both the reduced models intersected the FEM trajectory. Switching models forces the time integrator to restart, and initial values for the 1300K reduced model coefficients are needed at the switching time. To obtain the coefficients, the transient temperature field at 1000K was extracted and the inverse problem was solved in the lower dimensional eigenfunction space. This was done by using the QR-Transform method[24] to determine a least squares solution of Equation 2.6 for the temporal coefficients.

The results of this strategy are shown in Figure 2.8. As can be seen from the figure, the trajectories obtained from the reduced models and the FEM model coincide almost exactly. There is a deviation between the two trajectories immediately after the switch-over, because the integrator has to be reinitialized at the switch-over temperature. As a consequence, the integrator has to start with a new set of coefficients and lacks information about the time derivatives of the coefficients. Also, the least squares solution is not the exact initial state for the given temperature field, when the integrator switches between the two sets of eigenfunctions.

The difference between the FEM and the reduced model temperature trajectories are plotted in Figure 2.9. Other than at the switch-over temperature, there is good agreement between the two trajectories. If we ignore the region of the switch-over between the reduced models, the temperature difference is within  $\pm 10$  °C for the center and within  $\pm 15$  °C for the edge.

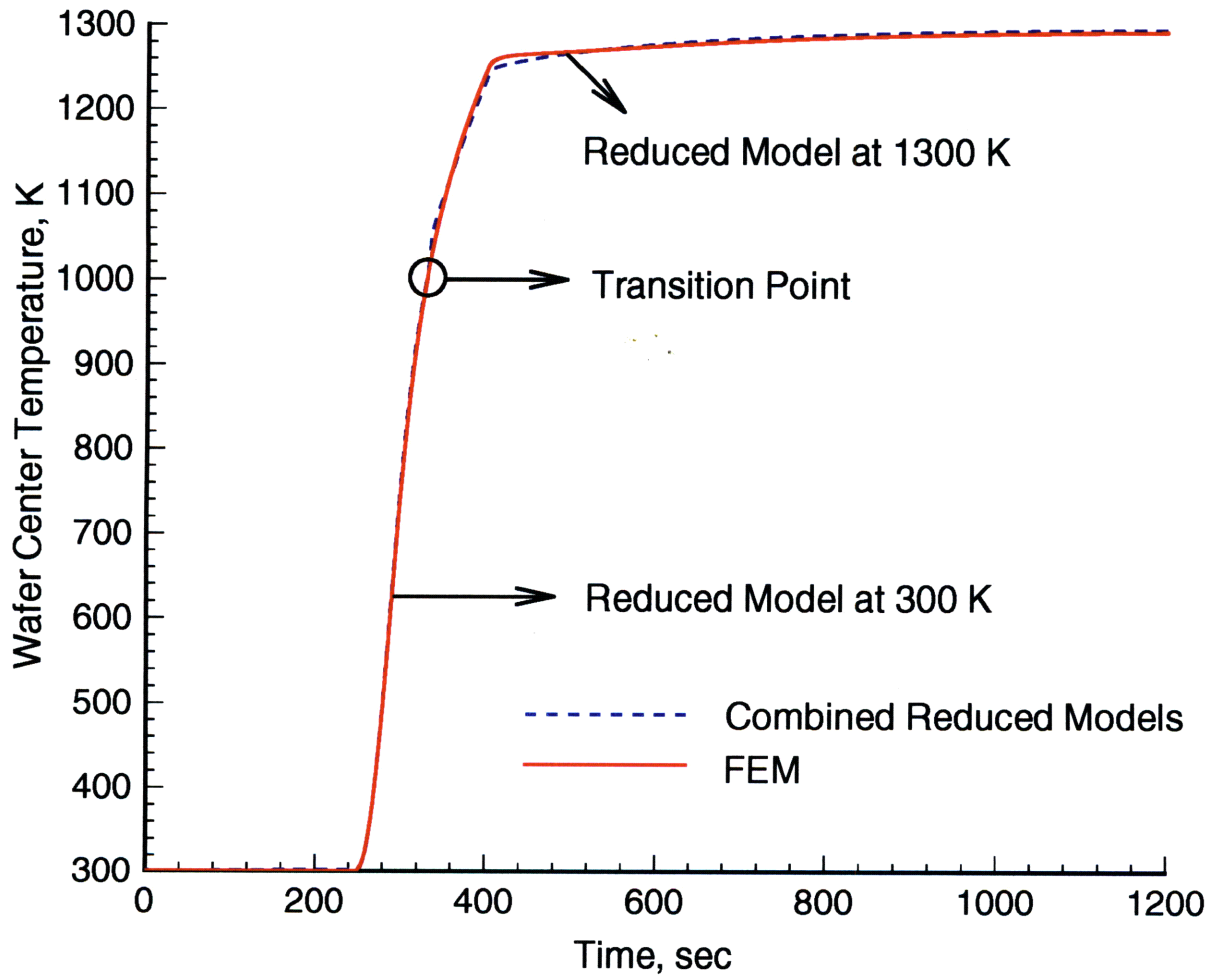


Figure 2.8 Behavior of wafer center temperature of FEM and combined reduced models, during transient ramp up and hold phases of the RTP cycle.

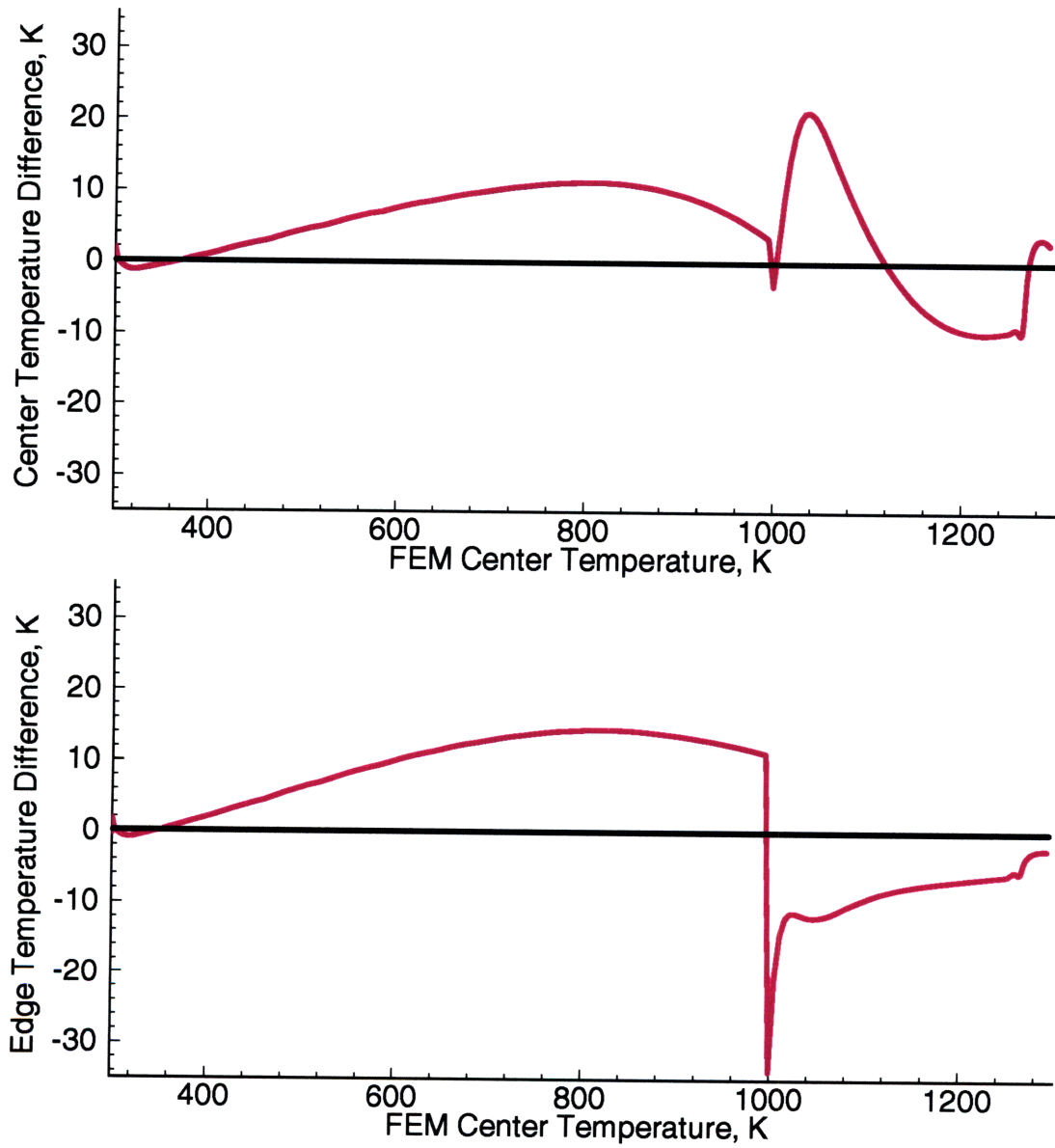


Figure 2.9 Difference in wafer temperatures between FEM and combined reduced models during ramp up and hold phases of the RTP cycle.



The reduced models extracted at wafer steady state temperatures of 1300K and 300K were then used to study the cool down phase of the RTP cycle. As seen from Figure 2.10, the 1300K reduced model reaches a higher steady state on cool down compared to the FEM transient model and the 300K reduced model reaches a lower steady state. Therefore unlike in the ramp up phase, the cool down part of the RTP cycle cannot be replicated by switching between the two models. In order to understand the cool down behavior better, temperature snapshots obtained from the reduced models at the end of the cool down phase (1600 seconds) are compared to the snapshot obtained at the same time instant from the FEM model in Figure 2.11. The 1300K model shows a much hotter reactor compared to the FEM model. Whereas the 300K model shows a reactor which has reached a nearly uniform temperature of 300K throughout the reactor. The hottest temperature zone in the 1300K model is in the lamphouse and the region between the quartz window and the showerhead. These effects are again due to the linearization of the gas phase properties in the reduced models. The 1300 K reduced model underestimates the thermal conductivity, so there is less conductive cooling from the cold walls and the wafer region is warmer than the FEM. This effect is predominant in the lamphouse and in the region between the quartz window and the showerhead because the gas is treated as stagnant in these regions, leading to the temperature gradients in this region being determined by radiation and gas phase conduction. This leads to the occurrence of the hot-zones in these regions in the 1300K reduced model. The 300 K model overpredicts the thermal conductivity at elevated temperatures. As a result, there is too much conduction coupling between the wafer and walls, leading to a cooler reactor. In both cases, the most dramatic difference between the FEM and the reduced models are in the lamphouse and in the region between the showerhead and window. In other regions of the reactor, the forced convection of the gas helps in removing some of these effects.

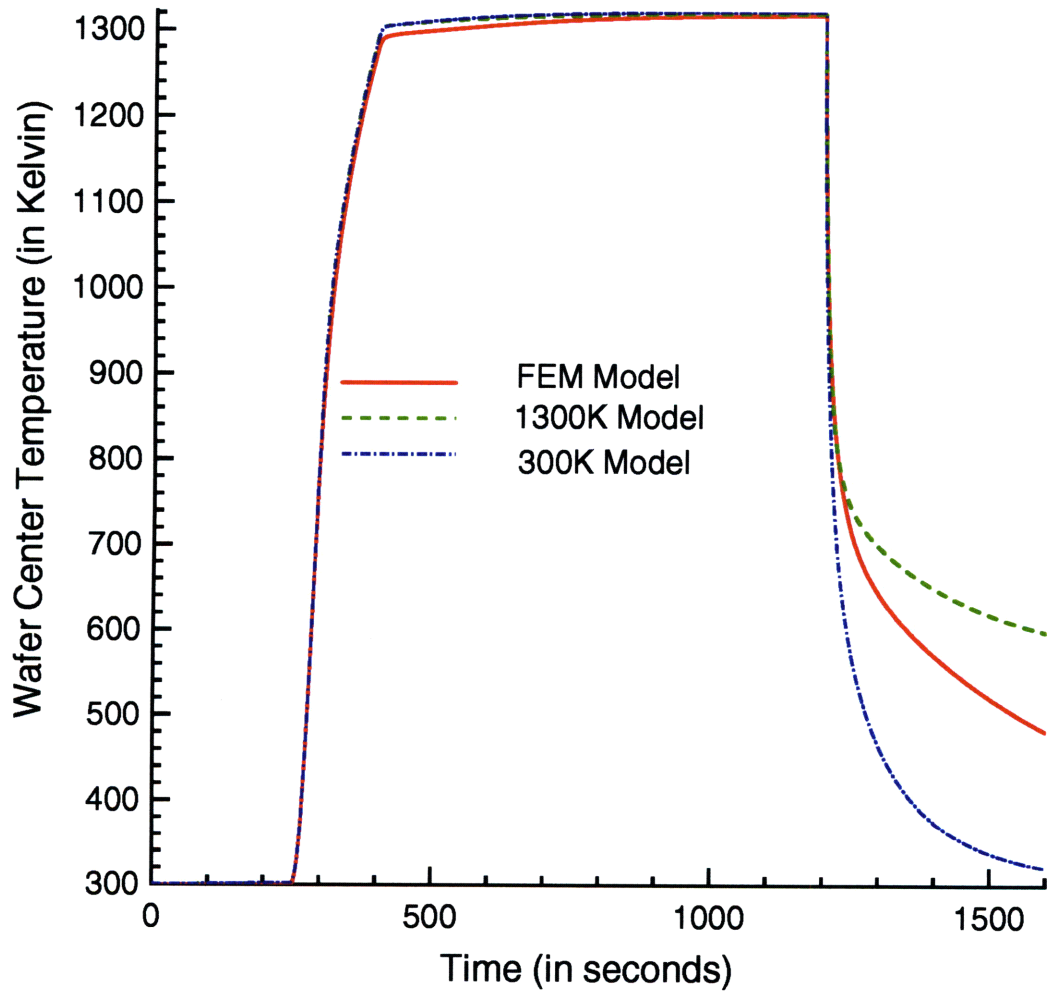


Figure 2.10 Behavior of wafer center temperature of FEM and reduced models during ramp up, hold, and cool down phases of the RTP cycle.

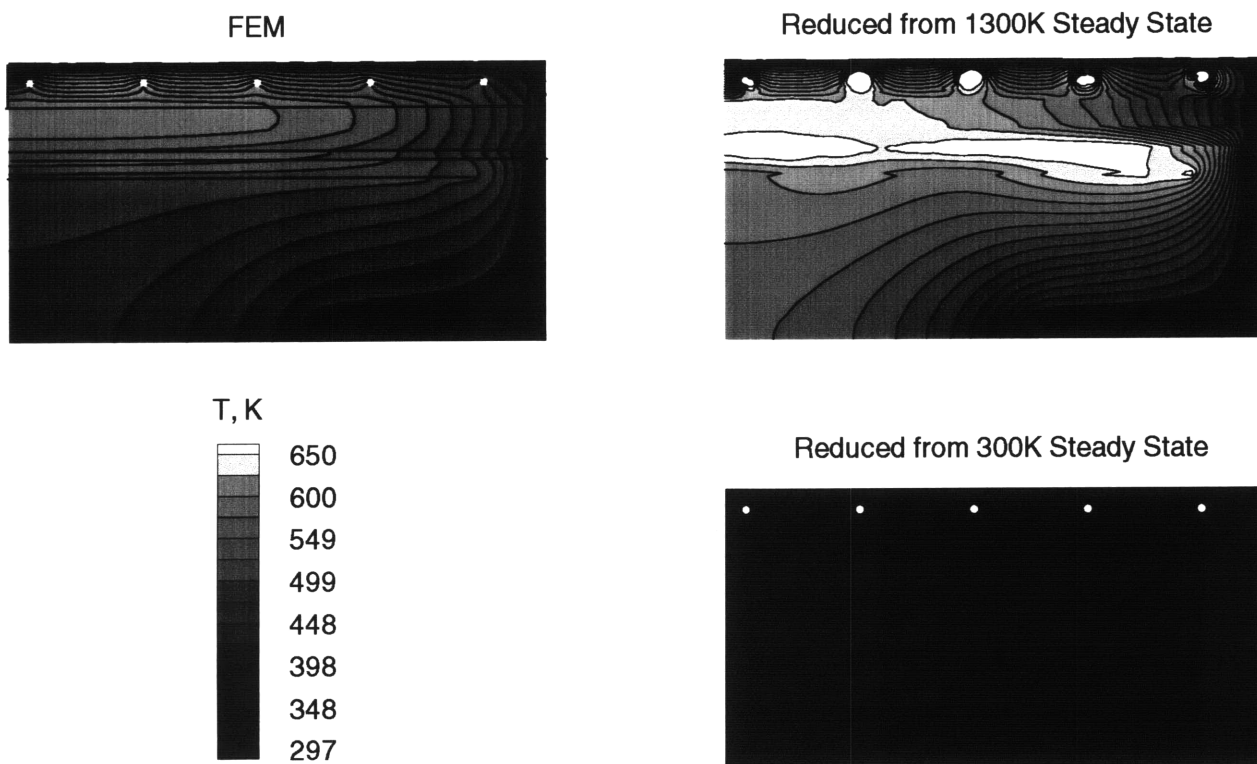


Figure 2.11 Temperature fields from FEM and reduced models during cool down phase of the RTP cycle.

There are several ways of approximating the cool down dynamics using reduced models. One of them would be to use some kind of arithmetic average of the responses of the two reduced models (300K and 1300K reduced models) to yield a cool down trajectory similar to the FEM model. Hence the strategy to replicate the cool down trajectory was to integrate both the reduced models simultaneously over the entire RTP cycle. For the ramp up phase, a linear interpolating function of wafer temperature was used to determine the contribution of each of the reduced models to the temperature trajectory. At the initial part of the ramp up phase, the temperature predicted by the 300K reduced model is taken, and at temperatures close to the hold phase, the temperature predicted by the 1300K reduced model is taken as the overall response of the combined reduced models. In between these two extremes, the interpolating function determines the contribution of the two reduced models in determining the overall temperature trajectory. In the cool down phase, an average of the predictions of the reduced models gives the overall response. The results of this strategy are shown in Figure 2.12.

Another strategy to replicate the cool down dynamics would be to extract a reduced model at an intermediate wafer steady state, viz. 1130K, and switch over to this reduced model during the cool down phase. After studying the cool down temperature trajectories predicted by reduced models extracted at different wafer steady states, the reduced model extracted around a wafer steady state of 1130K was found to have the best agreement with the cool down temperature trajectory as predicted by the FEM model. Therefore, in this strategy we start integrating using the 300K reduced model and switch to the 1300K reduced model at 1000K during the ramp up phase. The response of the 1300K model is taken as the overall response of the reduced models during the hold phase. At the end of the hold phase (1200 seconds), we switch over to the 1130K reduced model and use it to predict the response for the entire cool down phase. The results are shown in Figure 2.12. Both the strategies show reasonable agreement with the FEM model, but the latter strategy is marginally better.

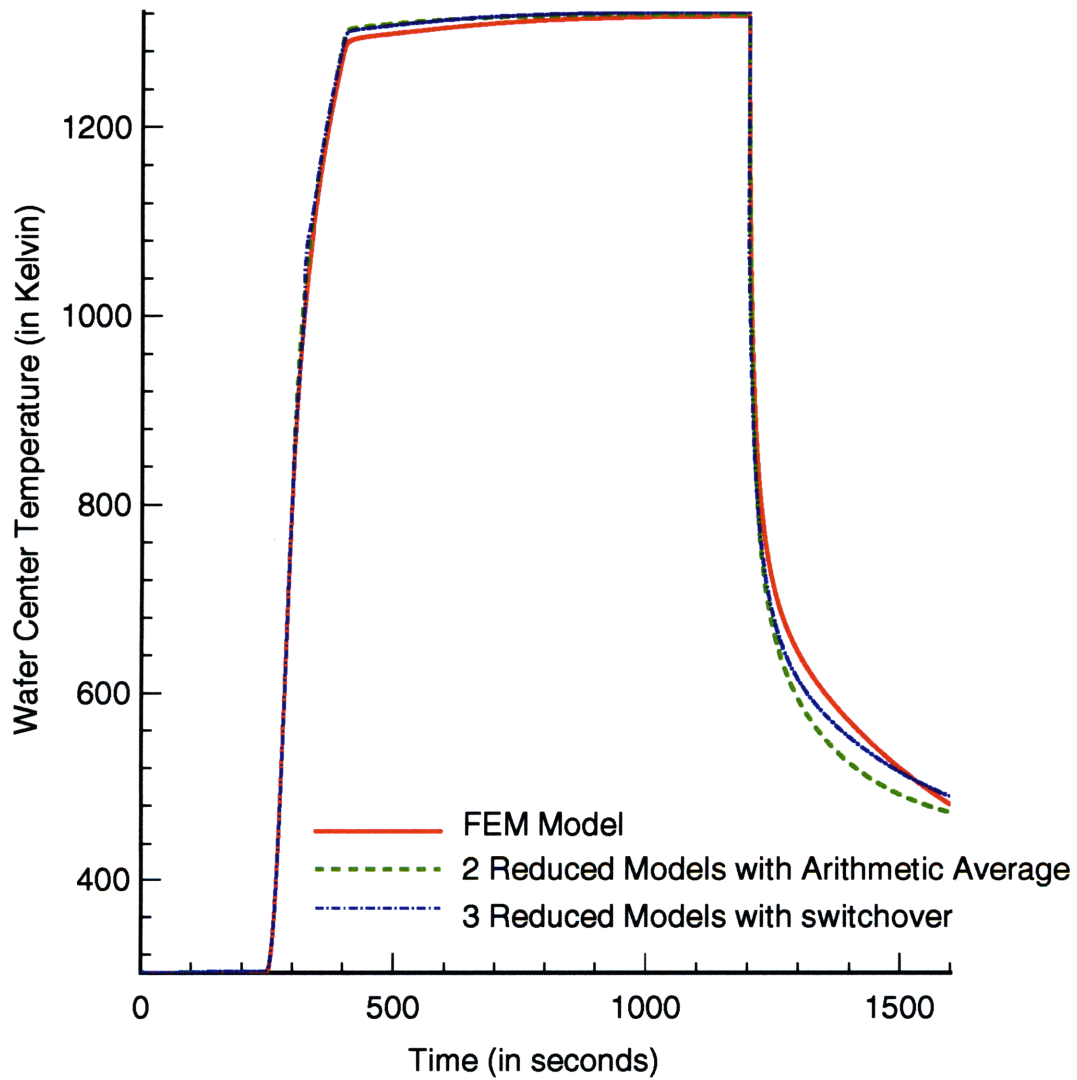


Figure 2.12 Transient ramp up, hold and cool down response of FEM and combined reduced models – a.) Combination of reduced models extracted at 1300K and 300K with arithmetic averaging, and b.) Combination of reduced models at 1300K, 1130K and 300K with switchover.

Finally, the reduced models were used to replicate an actual RTP cycle. In this cycle, the lamp powers were ramped from their initial switched-off state to the values corresponding to the wafer steady state of 1300K in 20 seconds. The lamp powers were then held at the steady state values for 30 seconds and then ramped down to the switched-off state in 20 seconds. The effect of this power protocol on the wafer center temperature for both the FEM and reduced models are shown in Figure 2.13. For the reduced model strategy, the 300K reduced model was used to replicate the initial wafer steady state. At the start of the ramp up, the integrator was switched over to the 1300K reduced model and this was used to replicate the entire trajectory from then on. The figure shows good agreement between the FEM and reduced model responses and further validates the efficacy of the reduced model strategy in replicating RTP transients.

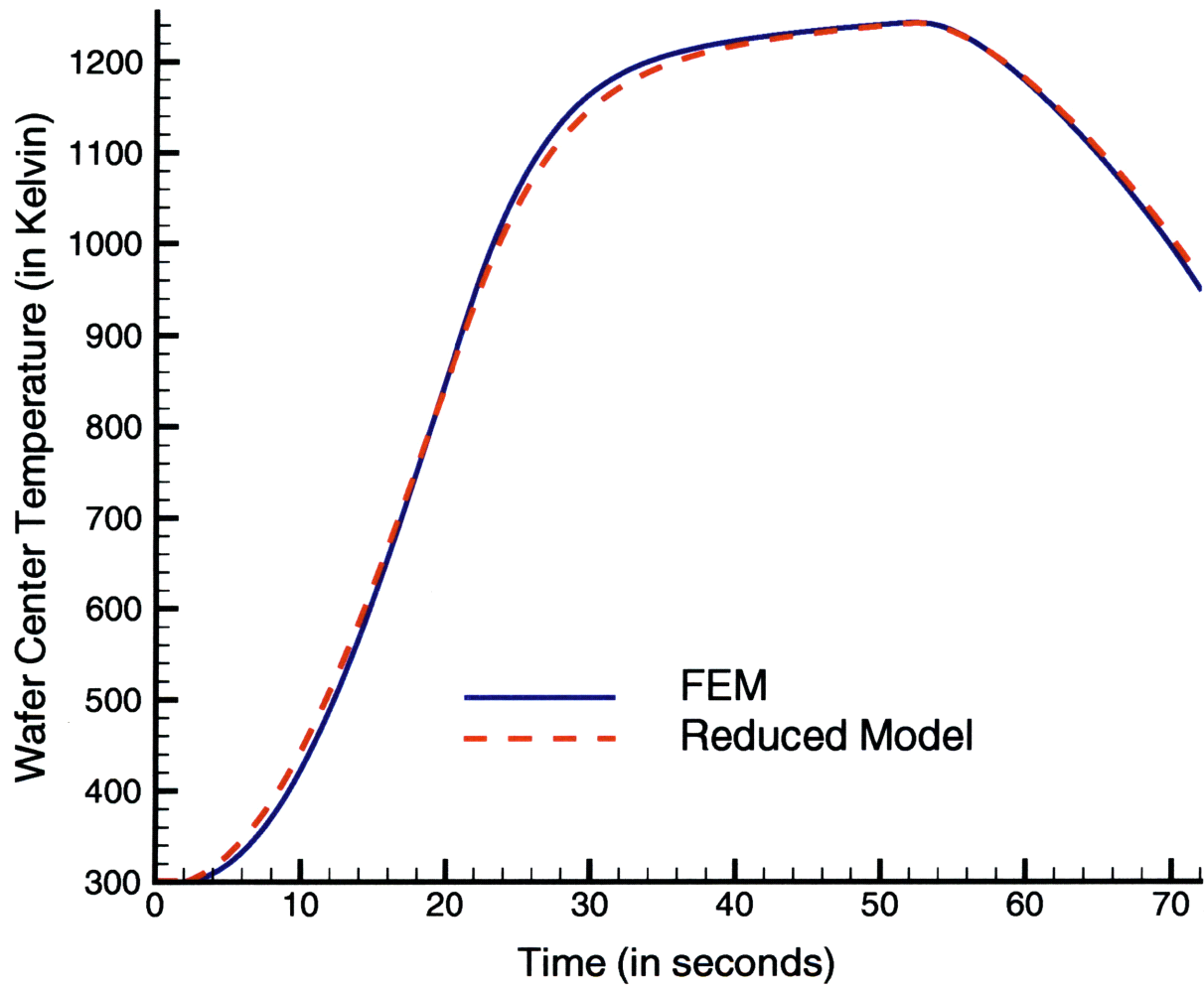


Figure 2.13 Replication of RTP ramp cycle using FEM and reduced models.

### 2.3.4 REDUCTION IN COMPUTATION TIME

The primary motivation for developing the technique of reduced order model extraction is to obtain reduced models with good predictive capabilities which have significantly less computation times compared to the FEM model. Therefore, timing runs were carried out both during steady state operating conditions and during transient ramp up to determine the reduction in computation time. For 200 seconds of real process time at the steady state operating temperature of 1300K, the following were the computation times for the FEM model and the reduced model extracted at 1300K to perform local temperature perturbations of the kind shown in Figure 2.4. The FEM model and the reduced model were simulated on a HP-735 workstation.

Models	Computation Time
FEM Transient Model	45 minutes
1300K Reduced Model	0.75 minutes

Table 2.1 Comparison of model execution time under steady operating conditions.

As shown in the table above, the time required for execution of the reduced model is nearly two orders of magnitude less than the FEM model. Timing runs were also carried out to compare the reduction in computation time between the combined reduced models used for the cool down study, the reduced model with switch over used to study the ramp up phase and the FEM transient model. For a real processing time of 150 seconds for the ramp up between 300K and 1300K the computation times on a HP-735 workstation are as follows,

Models	Computation Time
FEM Transient Model	22.17 minutes
Reduced Model with switch-over	1.8 minutes
Combined Reduced Models	3.73 minutes

Table 2.2 Comparison of model execution time for RTP ramp up.



Both the reduced models show nearly an order of magnitude decrease in computation time when compared to the FEM model. The computation time doubles in the case of the combined reduced models, as two reduced models, and hence two sets of differential equations, have to be integrated simultaneously. The computation time for this combined model can be decreased by choosing lesser number of eigenfunctions in each of the reduced models, hence leading to a smaller number of differential equations in each of the two sets. The reduced model with switch-over integrates faster than real time and shows promise of being useful in model based control.

The main overhead in terms of computation time comes in the reduced model extraction stage. A typical snapshot generation and eigenfunction extraction run can take hours. In this case, generation of 220 temperature snapshots and eigenfunction extraction from the transient FEM model took ~ 8 hours on a HP-735 workstation. Hence reduced models are good for applications in which the models have to be executed repetitively, as in model based controllers, or to study process changes under small perturbations. This would involve extracting a few reduced models at predetermined steady state operating conditions once, and then using them for the desired applications repetitively, thereby cutting down on the overhead.

## **2.4 CONCLUSION**

A strategy for extracting lower dimensional physically based reduced order models from complex finite element models has been developed. RTP was used as a test vehicle because of its dynamic nature, but the reduced model extraction procedure can be applied to any other process that can be described by similar fluid-thermal conservation equations. The reduced models (10 unknowns) showed very good agreement with the FEM model (5060 unknowns) not only around the steady state operating conditions from which they were extracted, but also at other steady operating conditions. This technique is superior to other strategies, such as lumping of nodes

within the FEM framework or assuming certain variables constant, because it does not simplify any of the physical conservation equations and the eigenfunction sets used to expand the equations carry qualitative information about the solution fields. A single reduced model can, therefore, be used for process optimization studies and answering “what if” type of process questions spanning a large window in process space ( $\pm 100$  °C). The entire RTP cycle (ramp up, hold and cool down) can be simulated using combinations of a few reduced models in real time on workstations. The reduced models have computation times that are an order of magnitude less than the FEM model. The reduced model strategy can be used in a combined feedforward and feedback control application. In such a strategy, the reduced models described in this chapter would be used to provide the feedforward trajectory and a simple PID controller would be used to implement feedback control around this predicted trajectory. Due to the linearization of the gas phase thermal properties, the temperature response of the reduced models would tend to become more and more inaccurate as the range of operation is stretched beyond the conditions around which the linearization is done. Therefore, by explicitly accounting for these nonlinearities, the response of the reduced models can be improved. However, this would introduce further complexities in the reduced model and increase their computation time. Intelligent “model switching” could provide a viable alternative to circumvent this trade-off problem.

## REFERENCES

- [1] I. Calder, "Rapid Thermal Process Integration," in *Reduced Thermal Processing for ULSI*, R. A. Levy, Ed. New York: Plenum Press, 1989, pp. 181.
- [2] P. Singer, "Rapid thermal processing: A progress report," in *Semiconductor International*, May 1993, pp. 64-69.
- [3] F. Roozeboom, "Introduction: History and Perspectives of RTP," in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [4] T. Breedijk, T. F. Edgar, and I. Trachtenberg, "A model predictive controller for multivariable temperature control in rapid thermal processing," *Proc. Amer. Control Conf.*, pp. 2980, 1993.
- [5] C. Schaper, "Real time control of rapid thermal processing semiconductor manufacturing equipment," *Proc. Amer. Control Conf.*, pp. 2985, 1993.
- [6] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Control of MMST RTP: Repeatability, uniformity, and integration of flexible manufacturing," *IEEE Trans. Semicon. Manuf.*, vol. 7, pp. 202, 1994.
- [7] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Modeling, identification, and control of rapid thermal processing systems," *J. Electrochem. Soc.*, vol. 141, pp. 3200, 1994.
- [8] G. Aral, T. P. Merchant, J. V. Cole, K. L. Knutson, and K. F. Jensen, "Concurrent engineering of a RTP reactor: Design and Control," *Proceedings of RTP-'94*, pp. 288, 1994.
- [9] H. Aling, A. Abedor, J. L. Ebert, A. Emami-Naeni, and R. L. Kosut, "Application of feedback linearization to model of Rapid Thermal Processing (RTP) reactors," *Proc. RTP - '95*, pp. 356-366, 1995.
- [10] J. P. Hebb and K. F. Jensen, "The effect of multilayer patterns on temperature uniformity

- during rapid thermal processing,” *J. Electrochem. Soc.*, vol. 143, pp. 1142, 1996.
- [11] K. F. Jensen, T. P. Merchant, J. V. Cole, J. P. Hebb, K. L. Knutson, and T. G. Mihopoulos, “Modeling Strategies for Rapid Thermal Processing: Finite Element and Monte Carlo Methods,” in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Pooreboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [12] T. P. Merchant, J. V. Cole, K. L. Knutson, J. P. Hebb, and K. F. Jensen, “A systematic approach to simulating Rapid Thermal Processing systems,” *J. Electrochem. Soc.*, vol. 143, pp. 2035, 1996.
- [13] L. Sirovich, “Turbulence and the Dynamics of Coherent Structures: I, II and III,” *Quarterly of Applied Mathematics*, vol. XLV, pp. 561, 1987.
- [14] W. S. Wyckoff, “Numerical Solution of Differential Equations through Empirical Eigenfunctions,” in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [15] J. L. Lumley, in *Transition and Turbulence*, R. E. Meyer, Ed. New York: Academic, 1981, pp. 215.
- [16] H. Park and L. Sirovich, “Turbulent thermal convection in a finite domain: Part II. Numerical results,” *Phys. Fluids A*, vol. 2, pp. 1659, 1990.
- [17] L. Sirovich and H. Park, “Turbulent thermal convection in a finite domain: Part I. Theory,” *Phys. Fluids A*, vol. 2, pp. 1649, 1990.
- [18] L. Sirovich, “Empirical eigenfunctions and low dimensional systems,” in *New Perspectives in Turbulence*, L. Sirovich, Ed. New York: Springer, 1991.
- [19] C. Canuto, M. Y. Hussaini, A. Quaderonic, and T. A. Zang, *Spectral Methods in Fluid Dynamics*. New York: Springer, 1988.
- [20] J. D. Rodriguez and L. Sirovich, “Low dimensional dynamics for the complex Ginzburg Landau equation,” *Physica D*, vol. 43, pp. 77, 1990.
- [21] L. Sirovich, “Chaotic dynamics of coherent structures,” *Physica D*, vol. 37, pp. 126,

1989.

- [22] L. Sirovich, J. D. Rodriguez, and B. Knight, "Two boundary value problem for Ginzburg Landau equation," *Physica D*, vol. 43, pp. 63, 1990.
- [23] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, "The dynamics of coherent structures in the wall region of a turbulent boundary layer," *J. Flu. Mech.*, vol. 192, pp. 115, 1988.
- [24] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore and London: The Johns Hopkins University Press, 1989.

## **Chapter 3**

# **Reduced Order Modeling of Fluid-Flow Induced and Wafer-Scale Pattern Induced Temperature Variations for RTP Systems**

This chapter is divided into two main sections. The first section deals with the reduced order modeling of coupled fluid flow and temperature fields in a RTP system. Flow effects can have significant influence on the temperature profiles in a RTP system. The reduced order modeling technique presented here is a computationally inexpensive method for modeling such effects. The second section demonstrates the reduced modeling strategy used to detect temperature non-uniformities across silicon wafer surfaces arising from multilayer thin film stacks patterned on the wafer surface. Changes in radiative properties of the multilayer thin film stacks can cause the temperature profiles across a patterned wafer to differ from that of a bare silicon wafer. This variation could lead to significant processing problems. [1] The reduced order modeling technique presented here is an effective diagnostic method for detecting the onset of such pattern effects.

## **3.1 REDUCED MODELING OF COUPLED TEMPERATURE AND FLOW FIELDS IN RTP SYSTEMS**

### **3.1.1 INTRODUCTION**

The last chapter dealt with the reduced modeling of the energy conservation equation for RTP systems by assuming a pseudo steady state flow field profile in the RTP chamber. This strategy works for RTP systems in which the gas phase heat transfer dynamics are much faster than the solid phase heat transfer dynamics, leading to a pseudo steady state approximation to be valid for the flow field. However, in many high pressure RTP systems the flow field changes with temperature leading to a discernible influence on the temperature profile and vice versa. The reduced model formulation for such systems must therefore incorporate coupled transient heat transfer and fluid flow equations. This section describes the technique used for developing such a reduced model.

### **3.1.2 MODELING EQUATIONS**

The equations of conservation of mass, momentum and energy[2] have to be solved simultaneously to predict the temperature and gas flow profiles in a RTP system. The appropriate dimensionless forms of these equations, for the gaseous phase in a RTP system, are described in the work by Merchant. [3] The dimensionless equations (from the work by Merchant[3]) relevant for the reduced model analysis are shown below,

Conservation of mass (Continuity equation):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1)$$

Conservation of momentum:

$$Re\rho\left(\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla\mathbf{v}\right) = \frac{Re}{Fr}(\rho - 1)\mathbf{e}_g - Re\nabla P + \nabla \cdot \boldsymbol{\tau} \quad (2)$$

Conservation of energy:

$$PeC_p\rho\left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T\right) = \nabla \cdot (k^f \nabla T) \quad (3)$$

where  $\rho$  is the dimensionless density of the gas phase. When the net change of mass in the system is negligible, the density of the gas is a function of temperature alone and can be written in dimensionless terms as  $\rho = 1/T$ . [3]  $\mathbf{v}$  is the dimensionless velocity vector for the gas phase velocity,  $Re$  is the Reynolds number,  $Fr$  is the Froude number,  $\mathbf{e}_g$  is the unit normal vector in the direction of the gravitational acceleration,  $P$  is the dimensionless pressure which is defined as the dynamic variation in pressure from the overall hydrostatic base pressure ( $P \equiv (\hat{P} - P_0)/\rho_0 U_0^2$ ),  $\boldsymbol{\tau}$  is the viscous stress tensor for a Newtonian fluid,  $Pe$  is the Peclet number for the gas phase,  $C_p$  is the gas phase specific heat,  $T$  is the dimensionless temperature for the gas,  $k^f$  is the thermal conductivity of the gas phase,  $t$  is the dimensionless time defined as  $t \equiv \hat{t}U_0/L_0$  and  $\nabla$  is the dimensionless gradient operator defined as  $\nabla \equiv \hat{\nabla}/L_0$ . The characteristic fluid velocity,  $U_0$ , and the characteristic length of the reactor,  $L_0$ , are used to make the various terms dimensionless.

For the solid components in the system only the energy equation has to be solved since the velocity components are identically zero. The conservation of energy equation for the solids can be written as

$$Pe^s C_p^s \rho^s \frac{\partial T}{\partial t} = \nabla \cdot (k^s \nabla T) \quad (4)$$



where  $Pe^s$  is the solid Peclet number,  $C_p^s$  is the specific heat of the solid,  $\rho^s$  is the density of the solid, and  $k^s$  is the thermal conductivity.

The Galerkin finite element procedure[4] involves the solution of an approximate integral form of the differential equations describing the conservation of mass, momentum and energy (Equations 1, 2, 3, and 4). This approximate integral form is also known as the ‘weak’ form of the differential equations. A detailed study of the FEM formulation of the conservation and continuity equations, relevant for RTP systems, can be found in the work by Merchant *et al.* [5] and Jensen *et al.*[6] In the FEM formulation, the continuity is assigned to the solution of the pressure field and the equation is expressed as

$$\frac{dT}{dt} \int_D \frac{1}{T^2} \Psi^i dV = \int_D \nabla \cdot (\rho \mathbf{v}) \Psi^i dV \quad (5)$$

where  $dV$  corresponds to the differential volume and the gas phase density has been replaced by the inverse of the temperature.  $\Psi^i(x, y, z)$ ,  $i = 1, \dots, N_l$  are bilinear piecewise-continuous interpolating basis functions defined on each element. Similarly the weak form of the momentum equation in terms of biquadratic piecewise-continuous basis functions,  $\Phi^i(x, y, z)$ ,  $i = 1, \dots, N_q$ , is given by

$$Re \int_D \rho \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) \Phi^i dV = \frac{Re}{Fr} \int_D [\rho - 1] e_s dV + Re \int_D P \nabla \Phi^i dV - \int_D \tau \cdot \nabla \Phi^i dV - \int_{\partial D} [ReP - \tau] \cdot \mathbf{n} \Phi^i dS \quad (6)$$

where Green’s theorem is applied to the pressure and viscous stress terms to compute an integral over the boundary ( $\partial D$ ) of the differential area  $dS$ . Similarly the integral form of the energy equation for the gaseous phase is expressed as

$$Pe \int_D \rho C_p \left[ \frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right] \Phi^i dV = - \int_D [k^f \nabla T] \cdot \nabla \Phi^i dV + \int_{\partial D} [k^f \nabla T] \cdot \mathbf{n} \Phi^i dS \quad (7)$$

The integral form of the energy equation for a solid is given by

$$Pe^s \int_D \rho^s C_p^s \frac{\partial T}{\partial t} \Phi^i dV = - \int_D [k^s \nabla T] \cdot \nabla \Phi^i dV + \int_{\partial D} [k^s \nabla T] \cdot \mathbf{n} \Phi^i dS \quad (8)$$

The radiative and lamp heat fluxes enter as solid boundary conditions for the energy equation and are included by replacing the temperature gradient term on the boundary with the appropriate heat flux equation.

### 3.1.3 FINITE ELEMENT FORMULATION OF MODELING EQUATIONS FOR AXISYMMETRIC GEOMETRIES

The specific extension of the weak form of the governing equations to axisymmetric geometries is described here. In a cylindrical coordinate system, the relevant coordinates are  $(r, \theta, z)$  with the respective velocities  $(v_r, v_\theta, v_z)$ . Thus the dimensionless gradient operator is defined as

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_z \frac{\partial}{\partial z} \quad (9)$$

The residual for the continuity equation (Equation 5) for node  $i$  is then written as

$$R_c^i = \frac{dT}{dt} \int_D \frac{1}{T} \Psi^i dV - \int_D \left[ \left( \frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z} \right) + \frac{v_r}{r} \right] \Psi^i dV + \int_D \left[ \frac{v_r}{T} \frac{dT}{dr} + \frac{v_z}{T} \frac{dT}{dz} \right] \Psi^i dV \quad (10)$$

The momentum conservation equation is broken up into three separate components. The  $r$  component of the momentum equations for node  $i$  can be expressed from Equation 6 as

$$\begin{aligned}
R_r^i = & Re \int_D \rho \left[ \frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} - \frac{v_\theta^2}{r} + v_z \frac{\partial v_r}{\partial z} \right] \Phi^i dV - Re \int_D P \left[ \frac{\partial \Phi^i}{\partial r} + \frac{\Phi^i}{r} \right] dV \\
& + \int_D \left[ \tau_{rr} \frac{\partial \Phi^i}{\partial r} + \tau_{\theta\theta} \frac{\Phi^i}{r} + \tau_{rz} \frac{\partial \Phi^i}{\partial z} \right] dV + \int_{\partial D} [ReP - \tau_{rr} - \tau_{rz}] \mathbf{re}_r \cdot \mathbf{n} \Phi^i dS
\end{aligned} \tag{11}$$

where the boundary integral over  $\partial D$  is non zero if the normal to the outlet has a component along the  $r$  direction. The azimuthal ( $\theta$ ) component of the momentum equation can be written in a similar manner as

$$R_\theta^i = Re \int_D \rho r \left[ \frac{\partial v_\theta}{\partial t} + v_r \frac{\partial v_\theta}{\partial r} - \frac{v_\theta v_r}{r} + v_z \frac{\partial v_\theta}{\partial z} \right] \Phi^i dV + \int_D r \left[ \tau_{r\theta} \frac{\partial \Phi^i}{\partial r} + \tau_{\theta z} \frac{\partial \Phi^i}{\partial z} \right] dV \tag{12}$$

There are no pressure gradients or boundary integral terms in Equation 12 due to azimuthal symmetry. The  $z$  component of the momentum conservation equation is formulated as

$$\begin{aligned}
R_z^i = & Re \int_D \rho \left[ \frac{\partial v_z}{\partial t} + v_r \frac{\partial v_z}{\partial r} + v_z \frac{\partial v_z}{\partial z} \right] \Phi^i dV - Re \int_D P \frac{\partial \Phi^i}{\partial z} dV + \frac{Re}{Fr} \int_D \left[ \frac{1}{T} - 1 \right] dV \\
& + \int_D \left[ \tau_{rz} \frac{\partial \Phi^i}{\partial r} + \tau_{zz} \frac{\partial \Phi^i}{\partial z} \right] dV + \int_{\partial D} [ReP - \tau_{rz} - \tau_{zz}] \mathbf{re}_z \cdot \mathbf{n} \Phi^i dS
\end{aligned} \tag{13}$$

where the gravitational force is assumed to be along the negative  $z$  direction. The various viscous tensor terms are as follows:

$$\begin{aligned}
\tau_{rr} &= \mu \left[ \frac{4}{3} \frac{\partial v_r}{\partial r} - \frac{2}{3} \left( \frac{v_r}{r} + \frac{\partial v_z}{\partial z} \right) \right] \\
\tau_{\theta\theta} &= \mu \left[ \frac{4}{3} \frac{v_r}{r} - \frac{2}{3} \left( \frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z} \right) \right] \\
\tau_{zz} &= \mu \left[ \frac{4}{3} \frac{\partial v_z}{\partial z} - \frac{2}{3} \left( \frac{v_r}{r} + \frac{\partial v_r}{\partial r} \right) \right] \\
\tau_{rz} = \tau_{zr} &= \mu \left[ \frac{\partial v_z}{\partial r} + \frac{\partial v_r}{\partial z} \right] \\
\tau_{r\theta} = \tau_{\theta r} &= \mu \left[ \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r} \right] \\
\tau_{\theta z} = \tau_{z\theta} &= \mu \left[ \frac{\partial v_\theta}{\partial z} \right]
\end{aligned} \tag{14}$$

Similarly, the residual for the conservation of energy equation for the gas phase nodes can be formulated as

$$\begin{aligned}
R_T^i &= Pe \int_D \rho C_p \left[ \frac{\partial T}{\partial t} + v_r \frac{\partial T}{\partial r} + v_z \frac{\partial T}{\partial z} \right] \Phi^i dV + \int_D k^f \left[ \frac{\partial T}{\partial r} \frac{\partial \Phi^i}{\partial r} + \frac{\partial T}{\partial z} \frac{\partial \Phi^i}{\partial z} \right] dV \\
&\quad - \int_{\partial D} k^f \left[ \frac{\partial T}{\partial r} + \frac{\partial T}{\partial z} \right] \cdot \mathbf{n} \Phi^i dS
\end{aligned} \tag{15}$$

and for the solid phase nodes can be formulated as

$$R_T^i = Pe^s \int_D \rho^s C_p^s \frac{\partial T}{\partial t} \Phi^i dV + \int_D k^s \left[ \frac{\partial T}{\partial r} \frac{\partial \Phi^i}{\partial r} + \frac{\partial T}{\partial z} \frac{\partial \Phi^i}{\partial z} \right] dV - \int_{\partial D} k^s \left[ \frac{\partial T}{\partial r} + \frac{\partial T}{\partial z} \right] \cdot \mathbf{n} \Phi^i dS \tag{16}$$

### 3.1.4 REDUCED MODELING STRATEGY FOR COUPLED FLUID-THERMAL EQUATIONS

The model reduction strategy for generating the reduced model, using the coupled continuity, momentum and energy conservation equations, is shown in Figure 3.1.

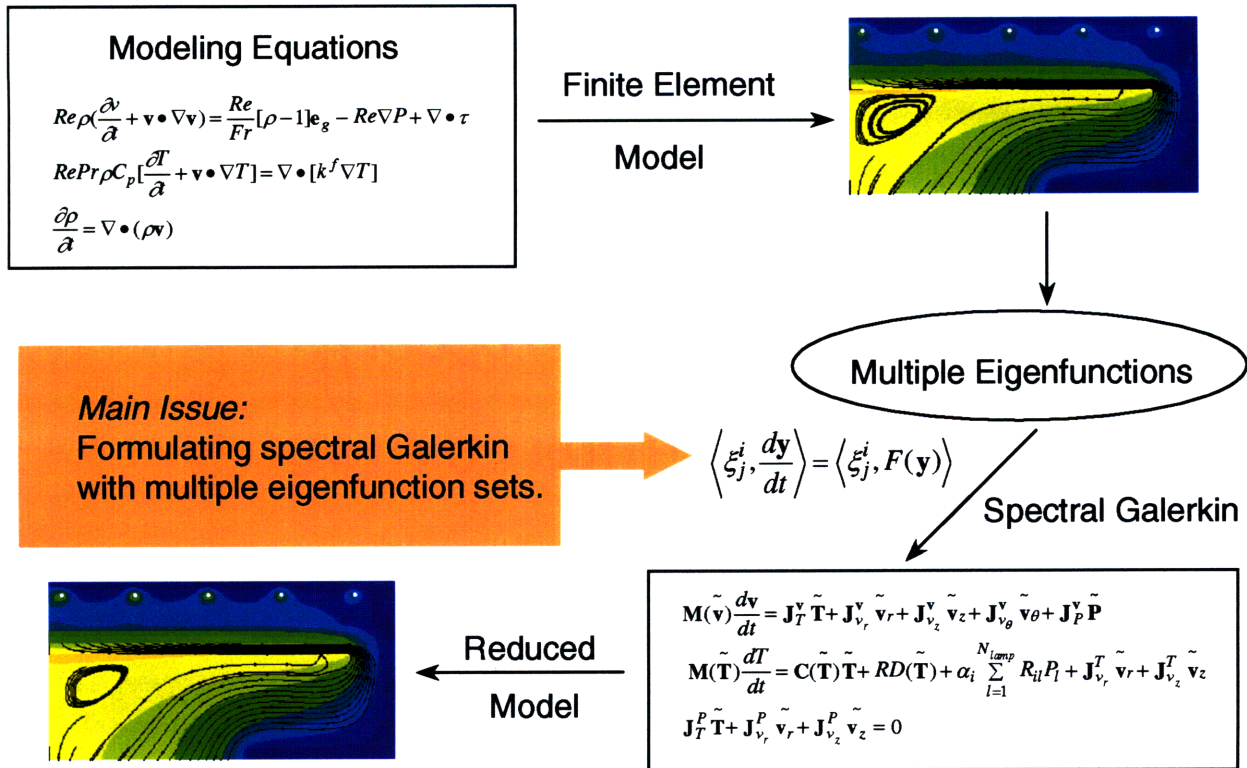


Figure 3.1. Model reduction strategy for the coupled transient fluid-thermal RTP system.

As shown in the figure, the model reduction procedure begins with the transient simulations of a finite element model of the complete set of fluid-thermal conservation equations. The transient FEM model is brought to near steady state operating conditions around a desired wafer temperature. The lamp powers are then perturbed individually to generate a set of transient temperature, pressure and velocity fields. The POD method[7, 8] is then used to extract a set of empirical eigenfunctions from each set of fields. Following this, the continuity, momentum and energy conservation equations are expanded in a pseudospectral[9] Galerkin technique[10] using the eigenfunctions as basis sets to yield a group of differential-algebraic equations (DAEs) that is referred to as the reduced model. The relationship between the different transient fields and the various eigenfunction sets are shown in Figure 3.2.

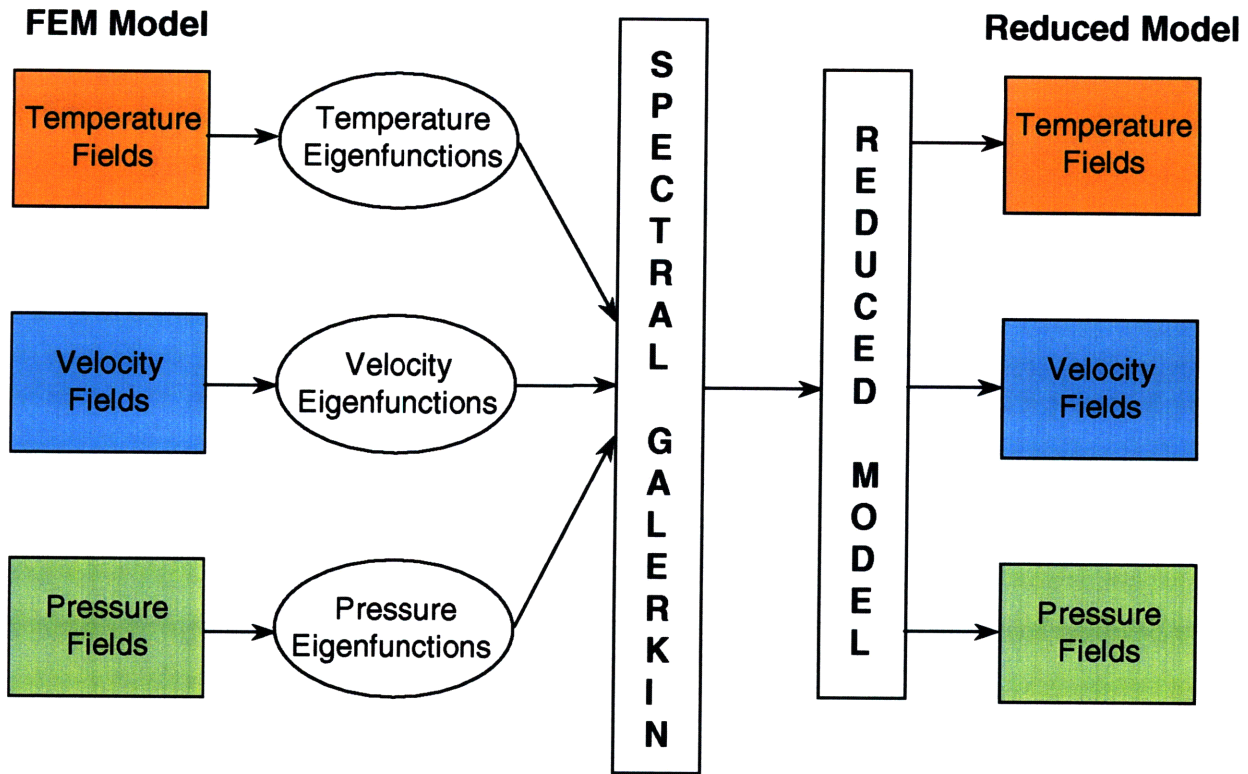


Figure 3.2. Relationship between different transient fields and eigenfunction sets.

The eigenfunctions for the different fields are tabulated below in Table 3.1.

$\xi^T$	Temperature eigenfunctions
$\xi^{v_r}$	Radial velocity eigenfunctions
$\xi^{v_\theta}$	Azimuthal velocity eigenfunctions
$\xi^{v_z}$	Axial velocity eigenfunctions
$\xi^P$	Pressure eigenfunctions

Table 3.1. Nomenclature for multiple eigenfunction sets.

The model reduction procedure is implemented in deviation variables to eliminate steady state offsets. A transient field can then be expressed in terms of deviation variables and eigenfunctions as follows

$$\mathbf{T}(t) = \bar{\mathbf{T}} + \tilde{\mathbf{T}}(t) = \bar{\mathbf{T}} + \sum_{i=1}^N a_i^T(t) \xi_i^T \quad (17)$$

where  $\bar{\mathbf{T}}$  denotes the steady state temperature field around which the eigenfunctions are extracted,  $\tilde{\mathbf{T}}(t)$  denotes the deviation of a transient temperature field from that particular steady state,  $\xi_i^T$  is the vector of temperature eigenfunctions,  $N$  is the number of eigenfunctions used to generate the reduced model, and  $a_i^T(t)$  denotes the temporal coefficients obtained by integrating the set of DAEs that comprise the reduced model. Following the same nomenclature, the velocity and pressure fields are expressed as follows,

Radial velocity:

$$\mathbf{v}_r(t) = \bar{\mathbf{v}}_r + \tilde{\mathbf{v}}_r(t) = \bar{\mathbf{v}}_r + \sum_{i=1}^N a_i^{v_r}(t) \xi_i^{v_r} \quad (18)$$

Axial velocity:

$$\mathbf{v}_z(t) = \bar{\mathbf{v}}_z + \tilde{\mathbf{v}}_z(t) = \bar{\mathbf{v}}_z + \sum_{i=1}^N a_i^{v_z}(t) \xi_i^{v_z} \quad (19)$$

Azimuthal velocity:

$$\mathbf{v}_\theta(t) = \bar{\mathbf{v}}_\theta + \tilde{\mathbf{v}}_\theta(t) = \bar{\mathbf{v}}_\theta + \sum_{i=1}^N a_i^{v_\theta}(t) \xi_i^{v_\theta} \quad (20)$$

Pressure:

$$\mathbf{P}(t) = \bar{\mathbf{P}} + \tilde{\mathbf{P}}(t) = \bar{\mathbf{P}} + \sum_{i=1}^N a_i^P(t) \xi_i^P \quad (21)$$

The empirical eigenfunctions are used to expand the governing equations (Equations 1, 2, 3, 4) by a pseudospectral Galerkin method within the FEM framework. In general, if the governing equations are expressed as

$$\frac{dy^i}{dt} = F(y^i) \quad (22)$$

then the pseudospectral Galerkin expansion for such an expression can be formulated as

$$\left\langle \xi_j^i, \frac{dy^i}{dt} \right\rangle = \left\langle \xi_j^i, F(y^i) \right\rangle \quad j = 1, \dots, N \quad (23)$$

where  $\langle \cdot \rangle$  represents the inner product and  $N$  is the number of eigenfunctions used to generate the reduced model. Each of the governing equations is expanded in terms of the eigenfunctions corresponding to the variable on which the equation is based. For example, for the conservation of momentum in the radial direction, the superscript  $i$  would represent the radial velocity,  $v_r$ , and the governing equation would be expanded by computing the inner product with the radial velocity eigenfunctions.

At every node of the FEM mesh, depending on whether the node lies in a solid or fluid element, one or more of the residual equations (Equations 10, 11, 12, 13, 15, 16) are solved. The same equations are now expressed in a form amenable to model reduction. The residuals for the conservation of energy (Equations 15, 16) are formulated as

$$\mathbf{M}(\bar{\mathbf{T}}) \frac{d\mathbf{T}}{dt} = \mathbf{C}(\bar{\mathbf{T}}) \tilde{\mathbf{T}} + \mathbf{R}D(\mathbf{T}) + \alpha_i \sum_{l=1}^{N_{lump}} R_{il} P_l + \mathbf{J}_{v_r}^T \tilde{\mathbf{v}}_r + \mathbf{J}_{v_z}^T \tilde{\mathbf{v}}_z \quad (24)$$

where  $\mathbf{M}(\bar{\mathbf{T}})$  is the dynamic contribution to the energy equation evaluated at steady state conditions corresponding to  $\bar{\mathbf{T}}$ ,  $\mathbf{C}(\bar{\mathbf{T}})$  arises from the combination of the conduction and



convection terms evaluated at steady state conditions,  $RD(\mathbf{T})$  is the nonlinear radiation contribution, and  $\alpha_i \sum_{l=1}^{N_{lamp}} R_{il} P_l$  is the radiative heat exchange from the lamps to any surface  $i$ . The

$\sim$  represents that these components are evaluated in deviation variables. The dependence of the energy conservation equation on the different velocity components are expressed in terms of components of the full system Jacobian. Therefore,  $\mathbf{J}_{v_r}^T$  is the Jacobian of the energy conservation equation with respect to the radial velocity,  $v_r$ , and  $\mathbf{J}_{v_z}^T$  is the Jacobian of the energy conservation equation with respect to the axial velocity,  $v_z$ . The residual for the conservation of momentum in the radial direction (Equation 11) is expressed as

$$\mathbf{M}(\bar{v}_r) \frac{dv_r}{dt} = \mathbf{J}_T^{v_r} \tilde{\mathbf{T}} + \mathbf{J}_{v_r}^{v_r} \tilde{v}_r + \mathbf{J}_{v_z}^{v_r} \tilde{v}_z + \mathbf{J}_{v_\theta}^{v_r} \tilde{v}_\theta + \mathbf{J}_P^{v_r} \tilde{\mathbf{P}} \quad (25)$$

In this equation,  $\mathbf{M}(\bar{v}_r)$  is the dynamic contribution to the radial momentum conservation equation,  $\mathbf{J}_T^{v_r}$  is the Jacobian of the radial momentum conservation equation with respect to the temperature. Hence, this signifies the dependence of the radial momentum conservation equation on temperature. Similarly,  $\mathbf{J}_{v_r}^{v_r}$  is the dependence on the radial velocity ( $v_r$ ),  $\mathbf{J}_{v_z}^{v_r}$  is the dependence on the axial velocity ( $v_z$ ),  $\mathbf{J}_{v_\theta}^{v_r}$  is the dependence on the azimuthal velocity ( $v_\theta$ ), and  $\mathbf{J}_P^{v_r}$  is the dependence on pressure. Retaining the same nomenclature the other residuals for the momentum conservation equation are formulated as follows,

Residual for the conservation of momentum in the axial direction (Equation 12):

$$\mathbf{M}(\bar{v}_z) \frac{dv_z}{dt} = \mathbf{J}_T^{v_z} \tilde{\mathbf{T}} + \mathbf{J}_{v_r}^{v_z} \tilde{v}_r + \mathbf{J}_{v_z}^{v_z} \tilde{v}_z + \mathbf{J}_P^{v_z} \tilde{\mathbf{P}} \quad (26)$$

Residual for the conservation of momentum in the angular (azimuthal) direction (Equation 13):

$$\mathbf{M}(\bar{v}_\theta) \frac{dv_\theta}{dt} = \mathbf{J}_T^{v_\theta} \tilde{\mathbf{T}} + \mathbf{J}_{v_r}^{v_\theta} \tilde{v}_r + \mathbf{J}_{v_z}^{v_\theta} \tilde{v}_z + \mathbf{J}_{v_\theta}^{v_\theta} \tilde{v}_\theta \quad (27)$$

The residual for the continuity equation (Equation 10) is formulated as an algebraic constraint to the set of ODEs shown above, and is expressed as

$$\mathbf{J}_T^P \tilde{\mathbf{T}} + \mathbf{J}_{v_r}^P \tilde{\mathbf{v}}_r + \mathbf{J}_{v_z}^P \tilde{\mathbf{v}}_z = 0 \quad (28)$$

Inner products with the respective eigenfunction sets are then computed over the set of DAEs (Equations 24-28) in the same manner as shown in Equations 22 and 23. This leads to the final formulation for the reduced model. The energy conservation equation (Equation 24) is expressed as

$$\begin{aligned} [\xi_i^{T^T} \mathbf{M}(\bar{\mathbf{T}}) \xi_i^T] \frac{da_i^T}{dt} &= [\xi_i^{T^T} \mathbf{C}(\bar{\mathbf{T}}) \xi_i^T] a_i^T + [\xi_i^{T^T} \mathbf{R}(\mathbf{T})] \{\mathbf{T}(t)\}^4 + \xi_i^{T^T} \alpha_i \sum_{l=1}^{N_{lamp}} R_{il} P_l + [\xi_i^{T^T} \mathbf{K}] \\ &+ [\xi_i^{T^T} \mathbf{J}_{v_r}^T \xi_i^{v_r}] a_i^{v_r} + [\xi_i^{T^T} \mathbf{J}_{v_z}^T \xi_i^{v_z}] a_i^{v_z} \end{aligned} \quad (29)$$

In Equation 29,  $\mathbf{R}(\mathbf{T})$  denotes the nonlinear radiation contribution to the reduced model,  $\mathbf{K}$  arises from the combination of the steady state contribution to the radiation heat transfer and the lamp heat fluxes. All the terms expressed within square parentheses are precomputed. These terms can, therefore, be solved for exclusively in the lower dimensional eigenfunction coefficient space. The conservation of momentum in the radial direction (Equation 25) is expressed as

$$\begin{aligned} [\xi_i^{v_r^T} \mathbf{M}(\bar{\mathbf{v}}_r) \xi_i^{v_r}] \frac{da_i^{v_r}}{dt} &= [\xi_i^{v_r^T} \mathbf{J}_T^{v_r} \xi_i^T] a_i^T + [\xi_i^{v_r^T} \mathbf{J}_{v_r}^{v_r} \xi_i^{v_r}] a_i^{v_r} + [\xi_i^{v_r^T} \mathbf{J}_{v_z}^{v_r} \xi_i^{v_z}] a_i^{v_z} \\ &+ [\xi_i^{v_r^T} \mathbf{J}_{v_\theta}^{v_r} \xi_i^{v_\theta}] a_i^{v_\theta} + [\xi_i^{v_r^T} \mathbf{J}_P^{v_r} \xi_i^P] a_i^P \end{aligned} \quad (30)$$

The conservation of momentum in the axial direction (Equation 26) is written as

$$[\xi_i^{v_z T} \mathbf{M}(\bar{\mathbf{v}}_z) \xi_i^{v_z}] \frac{da_i^{v_z}}{dt} = [\xi_i^{v_z T} \mathbf{J}_T^{v_z} \xi_i^T] a_i^T + [\xi_i^{v_z T} \mathbf{J}_{v_r}^{v_z} \xi_i^{v_r}] a_i^{v_r} + [\xi_i^{v_z T} \mathbf{J}_{v_z}^{v_z} \xi_i^{v_z}] a_i^{v_z} + [\xi_i^{v_z T} \mathbf{J}_P^{v_z} \xi_i^P] a_i^P \quad (31)$$

The conservation of momentum in the angular direction (Equation 27) is expressed as

$$[\xi_i^{v_\theta T} \mathbf{M}(\bar{\mathbf{v}}_\theta) \xi_i^{v_\theta}] \frac{da_i^{v_\theta}}{dt} = [\xi_i^{v_\theta T} \mathbf{J}_T^{v_\theta} \xi_i^T] a_i^T + [\xi_i^{v_\theta T} \mathbf{J}_{v_r}^{v_\theta} \xi_i^{v_r}] a_i^{v_r} + [\xi_i^{v_\theta T} \mathbf{J}_{v_z}^{v_\theta} \xi_i^{v_z}] a_i^{v_z} + [\xi_i^{v_\theta T} \mathbf{J}_{v_\theta}^{v_\theta} \xi_i^{v_\theta}] a_i^{v_\theta} \quad (32)$$

The continuity equation (Equation 28) is written as

$$[\xi_i^{P T} \mathbf{J}_T^P \xi_i^T] a_i^T + [\xi_i^{P T} \mathbf{J}_{v_r}^P \xi_i^{v_r}] a_i^{v_r} + [\xi_i^{P T} \mathbf{J}_{v_z}^P \xi_i^{v_z}] a_i^{v_z} = 0 \quad (33)$$

The main mathematical challenge to generating a reduced model formulation from this set of DAEs is evaluating the inner product across multiple eigenfunction sets. Empirical eigenfunctions, by virtue of the POD extraction technique, are orthonormal within a given set, but are not orthonormal across sets. While evaluating the reduced model, inner products had to be taken across different sets due to the coupling of the equations through the different variables. For example, the energy conservation equation (Equation 29) has a dependence on the radial velocity. This contribution was evaluated by taking the inner product between the temperature eigenfunctions and the radial velocity eigenfunctions (used to expand  $\tilde{\mathbf{v}}_r$ ). The lack of orthonormality across sets did not prove to be a problem while generating the reduced model as long as the appropriate set of eigenfunctions were used for the conservation and continuity equations; i.e., the residual for the energy conservation equation was made orthogonal to the temperature eigenfunctions, the residual for the conservation of radial momentum equation was made orthogonal to the radial velocity eigenfunctions, so on and so forth. This extension of the POD-pseudospectral Galerkin technique to multiple eigenfunction sets is a significant advancement upon the expansion of the energy equation alone in terms of the temperature eigenfunctions.

Another possible method for extracting eigenfunctions would be concatenate all the pressure, temperature and velocity fields and extract a single set of eigenfunctions from all the fields. But in this case there is no particular steady state value about which to regenerate the absolute fields, if one works with deviation eigenfunctions. Also this introduces a scaling problem when regenerating the fields as the temporal coefficients differ from each other by orders of magnitude. Hence, multiple eigenfunction sets provide a more systematic and mathematically tractable framework for the generation of reduced models for the fluid-thermal conservation equations.

### **3.1.5 RESULTS AND DISCUSSION**

To study the effect of the flow field on the temperature profile in the chamber, the carrier gas in the RTP system was changed from nitrogen to hydrogen and the operating pressure was increased from 0.01 atmospheres to 0.1 atmospheres. Individual lamp power perturbations of the type used to extract temperature fields, as described earlier, were used to generate variations in temperature, velocity, and pressure fields. Empirical eigenfunctions were then extracted from these transient fields. Figure 3.3 shows the effect of the lamp power perturbations on the wafer center temperature as obtained from the FEM and reduced models.

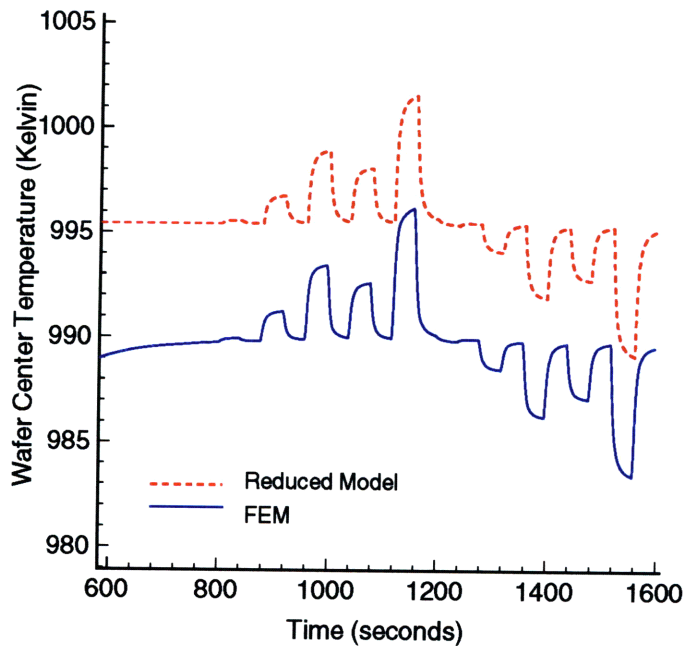


Figure 3.3. Wafer center temperature trajectories from the FEM and reduced models.

The offset between the two trajectories arises from the fact that the FEM model has to be integrated from room temperature conditions to the desired steady state conditions using a slow ramp rate. This was found to be the most effective integration technique for reaching the high temperature operating conditions using the FEM model. Hence, when the perturbations are introduced, the FEM model is still on a slow ramp up and is only near the desired steady state conditions. The FEM model could be allowed to integrate longer along the slow ramp to get exact agreement of the steady states, but this would prove to be computationally time-consuming. On the other hand, since the reduced model is generated around the desired steady state, this model replicates the steady conditions exactly. The reduced order model nevertheless replicates the magnitude of perturbations introduced into the system.

Figure 3.4 shows the temperature and flow fields as obtained from the FEM and reduced models. These fields were extracted at a time instant just prior to the onset of the lamp power perturbations. As seen from the figure, the reduced model agrees well with the FEM model in the replication of the temperature and flow fields.

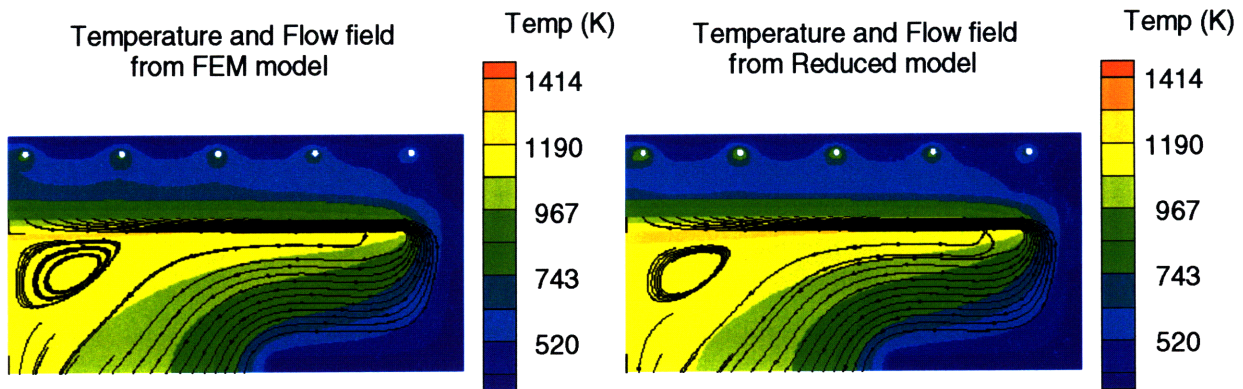


Figure 3.4. Temperature and flow fields extracted at 0.1 atmospheres with  $H_2$  as carrier gas.

Figure 3.5 compares the temperature and flow fields generated under 0.01 atmospheres operating pressure and using nitrogen as the carrier gas. These are the same low pressure conditions used to study the reduced model, as shown earlier. As seen from the figure, the reduced model agrees well with the FEM model under these conditions.

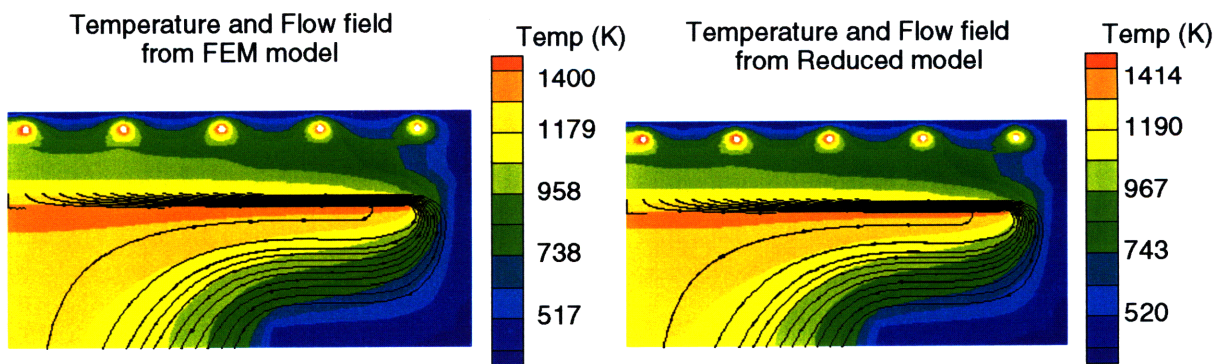


Figure 3.5. Temperature and flow fields extracted at 0.01 atmospheres with  $N_2$  as carrier gas.

The effect of the high pressure conditions and the change in carrier gas composition is reflected in the nature of the flow fields. The low pressure case (Figure 3.5) has a fairly uniform flow field all throughout the chamber. Whereas in the 0.1 atmospheric pressure case (Figure 3.4), convection roll cells develop beneath the wafer center. This lowers the wafer temperature to ~1000 K as opposed to ~1300 K for the low pressure case. The reduced model succeeds in capturing these flow effects accurately, as seen from the figures.

The FEM model takes ~24 hours to do the entire slow ramp up and perturbations. On the other hand, the reduced model takes a few minutes to replicate the same trajectory. This case study demonstrates the order of magnitude savings in computation time that can be achieved by using the model reduction scheme. Moreover, the strategy could be a computationally inexpensive tool for understanding localized flow effects in the RTP system without extensive workstation computations.

## **3.2 REDUCED MODELING OF WAFER-SCALE PATTERN EFFECTS**

### **3.2.1 INTRODUCTION**

Processing of semiconductor chips begins with a blank silicon wafer, but with each step of wafer processing, thin layers of materials such as silicon dioxide, amorphous silicon, silicon nitride and titanium silicide are grown or deposited and patterned by photolithography. The presence of these layers changes the radiative properties of the wafer through thin film interference effects. The front side of the wafer where devices are patterned is shown schematically in Figure 3.6. There are three main areas: the die area, where the devices are patterned, the between-die area that leaves room for test structures and for sawing the wafer when the processing is finished, and the wafer periphery (border) where devices are typically not patterned. These three regions consist of different multilayer thin film stacks, leading to a spatial variation of radiative properties across the wafer.



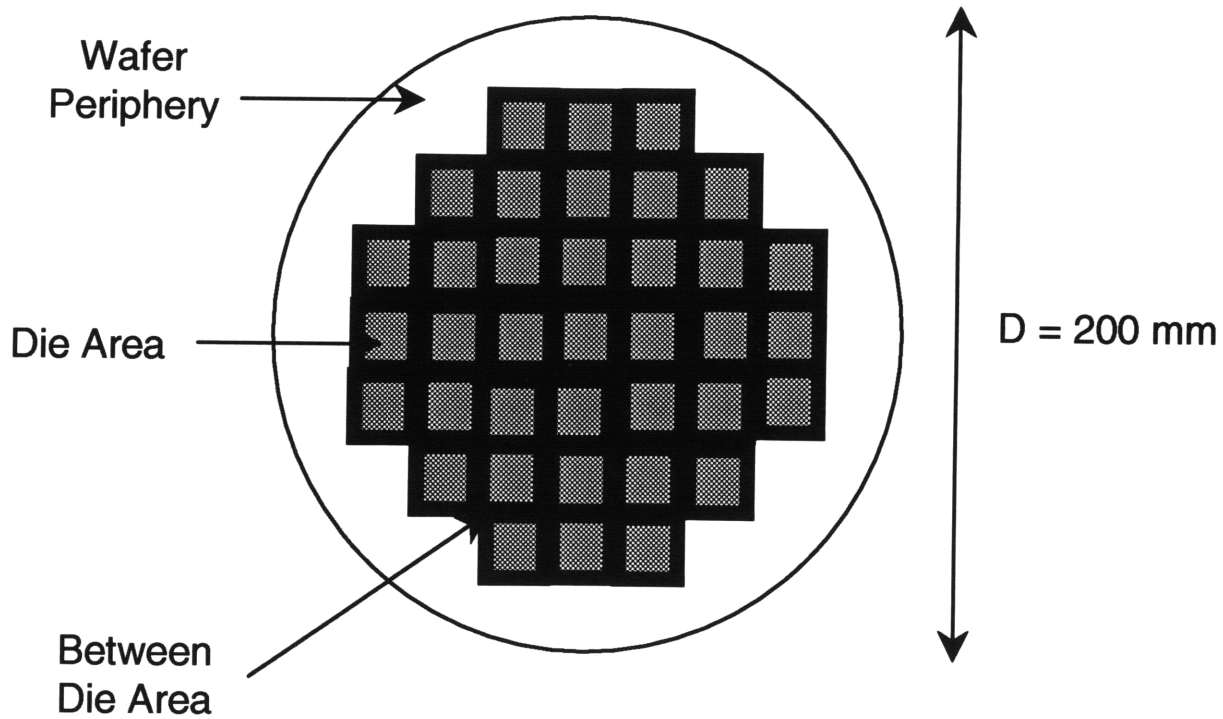


Figure 3.6. Schematic of the front side of a patterned wafer.

The effect of patterns on temperature non-uniformity across the silicon wafer has been experimentally demonstrated. [11-13] There have also been extensive modeling studies to understand the effect of wafer-scale patterns on temperature profiles across the wafer. [1, 5, 14, 15] But these models are too complex to be implemented as diagnostic tools for identifying the onset of these pattern effects. Hence, there exists a demand for using reduced models for detecting the influence of wafer-scale pattern effects on wafer temperature. This section presents a strategy for generating reduced models for detecting wafer-scale pattern effects.

### 3.2.2 MODEL REDUCTION STRATEGY

Introduction of thin film layers on a silicon wafer, influences the parameters in the radiation heat flux boundary condition. The integral finite element form of the solid phase energy conservation equation [1, 5, 6] is written as

$$Pe^s \int_D C_p^s \rho^s \frac{\partial T}{\partial t} \Phi^i dV = - \int_D (k^s \nabla T) \cdot \nabla \Phi^i dV + \int_{\partial D} (k^s \nabla T) \cdot \mathbf{n} \Phi^i dS \quad [34]$$

The heat flux boundary conditions on the solid surfaces are included in the formulation by replacing the temperature gradient term on the boundary with the appropriate radiation and convection boundary conditions. For solid surfaces adjacent to the gas phase where the velocity is being solved, the boundary condition is given as

$$k^s \nabla T_i \cdot \mathbf{n} = k^f \nabla T_i \cdot \mathbf{n} + \alpha_i^\ell \sum_{m=1}^{N_{lamp}} \sum_{j=1}^{N_{s,m}} P_m R_{ij}^\ell + \sigma \sum_{k=1}^{N_{bands}} \left[ \alpha_i^k \sum_{j=1}^{N_{sw}} \phi_{T_j}^k \varepsilon_j^k R_{ij}^k T_j^4 - \phi_{T_i}^k \varepsilon_i^k T_i^4 \right] \quad [35]$$

where  $k^s$  and  $k^f$  are the thermal conductivities of the solid and surrounding gas respectively,  $T$  is the surface temperature,  $\mathbf{n}$  is the unit normal vector to the surface,  $\alpha$  is the total absorptance,  $\varepsilon$  is the total emittance,  $R_{ij}$  is the radiative exchange factor from surface  $i$  to surface  $j$  (defined as the fraction of energy emitted from surface  $i$  that is incident on surface  $j$  through direct viewing and multiple reflections),  $P_m$  is the power flux from lamp  $m$ ,  $N_{bands}$  is the number of bands excluding the lamp band, and  $N_{sw}$  is the number of surfaces in the chamber excluding lamp surfaces. Superscript  $\ell$  indicates the lamp band and  $k$  indicates bands for emission from all other surfaces. The first term on the right hand side is the contribution of convection from the gas. The second term gives the radiative input absorbed by surface  $i$  from the lamps, with the power in each lamp defined as the fraction of total lamp power that passes through the quartz window. The last term accounts for energy that is emitted from all other surfaces in the enclosure that is absorbed by surface  $i$  and the energy that is emitted from surface  $i$ . The radiative properties of all surfaces influence radiative exchange factors in each band because of multiple reflections in the chamber. Also, the radiative properties of surface  $i$  directly impact the energy balance through absorption and emission of radiation in each band. To solve for the temperature fields, one needs to know the radiative properties of all surfaces in

each band. Therefore, whenever a different material is deposited onto the surface of the silicon wafer the absorptance and emittance of the surface, in each band, are modified. This in turn changes the radiative exchange factors, in each band, because of multiple reflections. These changes have to be incorporated both in the FEM and reduced models to accurately generate temperature fields in the chamber.

Both the FEM and the reduced model use a three-band formulation for calculating the total radiative properties for each surface in the chamber. It is assumed that there are two dominant sources of thermal radiation in the system – the tungsten halogen lamps, which are at approximately 3000 K, and the wafer, which is at approximately 1000 K. [1] The quartz window is treated as transparent for wavelengths below 4.0  $\mu\text{m}$ , and opaque for wavelengths greater than 4.0  $\mu\text{m}$ . For band 1, the Planck function is evaluated at the wafer temperature, and integrated from 0.4 to 4.0  $\mu\text{m}$ . For band 2, the Planck function is evaluated at the wafer temperature, and integrated from 4.0 to 20.0  $\mu\text{m}$ . For band 3, also known as the lamp band, the Planck function is evaluated at the lamp temperature of 3000 K, and integrated from 0.4 to 4.0  $\mu\text{m}$ . For wavelengths greater than 4  $\mu\text{m}$ , the lamp radiation is filtered away by the quartz window. The upper bounds of 0.4  $\mu\text{m}$  and 20  $\mu\text{m}$  are chosen so that 99% of the energy is captured when calculating radiative properties. Because the body temperature and source temperature are the same for the wafer in bands 1 and 2, the total absorptance is equal to the total emittance for the wafer in these bands. Because of multiple reflections in the system, it is assumed that lamp radiation will arrive at the wafer surface from all directions. Hence, the total hemispherical radiative properties are used. [1]

For each different area on the wafer, the total radiative properties in band 1, band 2, and the lamp band can be assigned in the reactor scale boundary condition (Equation 35). In this way, the connection between the macroscale temperature fields and the microscale wafer patterns is made. The radiative properties are calculated at the steady state processing temperature of the wafer, and are not varied dynamically through the transient simulations.

A schematic for the model reduction procedure for RTP systems is shown in Figure 3.7.

The figure depicts the different inputs to the reduced model extraction program from the various finite element program [1, 5, 6] modules. Wafer-scale pattern effects have an influence on all these inputs to a greater or lesser degree, but by analyzing the modules separately we can segregate the inputs into classes that are influenced the most and those that are influenced the least. By depositing different layers of oxides, nitrides, silicides etc. on the wafer surface changes the topography and the radiative properties (absorptance and emittance) of the surface. These, in turn, change the radiative exchange factors significantly. Therefore a reduced model used to model pattern effects should be constructed from the correct set of exchange factors and radiative properties. Changing the radiative properties also changes the steady state temperature profile of the wafer surface drastically. Since the reduced model is constructed in deviation variables around a given steady state operating condition, the updated steady state profile is crucial in obtaining accurate replication of temperature trajectories using reduced models. The wafer-scale pattern effects do not influence the global empirical eigenfunctions significantly unless they bring about major changes to the temperature contours in the entire reactor. Since this was not found to be the case, eigenfunctions extracted by modeling the wafer surface as bare silicon can be used to generate the reduced model for modeling wafer-scale pattern effects. This distinction among the various inputs is also depicted in Figure 3.7.

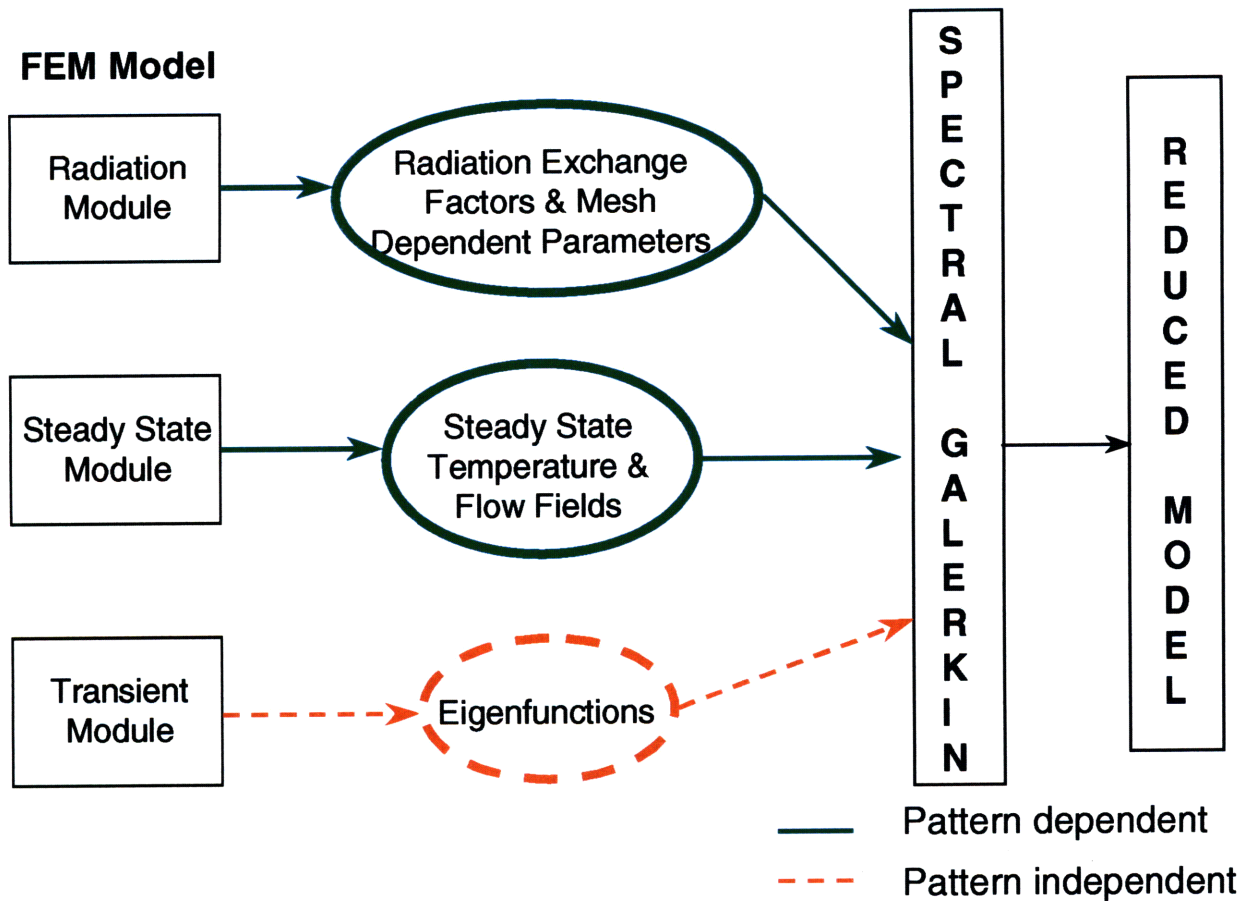


Figure 3.7. Reduced modeling strategy for modeling wafer-scale pattern effects.

Therefore the strategy for modeling wafer-scale pattern effects using reduced models is as follows:

- The transient FEM model is brought to a desired steady state and the lamp powers are perturbed to generate a set of eigenfunctions. The wafer is modeled with radiative properties of bare silicon.
- Next, the wafer is modeled as consisting of different layers of oxide, nitride and amorphous silicon. Radiative properties, such as the radiative exchange factors, for the patterned wafer, are calculated using the FEM radiation program.
- New steady state temperature and flow fields are generated from the FEM steady state program using the pattern modified radiative properties for the wafer.

- The updated set of radiative properties and temperature and flow fields are then used to generate the reduced model for a patterned wafer.

In this strategy we do not have to extract a new set of eigenfunctions every time we change the radiative properties, thereby cutting down on the simulation time for eigenfunction extraction, which is ~ 8 hours for ~ 200 temperature fields.

### 3.2.3 RESULTS AND DISCUSSION

#### 3.2.3.1 Rapid Thermal Annealing

Annealing of shallow junctions is one of the main process steps for which RTP is needed for future generations of devices. [16] The low thermal budget of RTP makes rapid thermal annealing of shallow implants an attractive alternative to high thermal budget furnace processing.

For the base case simulations used to extract the eigenfunctions, a bare silicon wafer was chosen. The emittance (or absorptance) values in the three bands for the wafer were assigned as tabulated below in Table 3.2, [1]

$\epsilon_1$	$\epsilon_2$	$\alpha_t$
0.66	0.66	0.64

Table 3.2. Emittance values in the three bands for a bare silicon wafer.

The FEM transient model was used to generate a set of temperature fields by simulating lamp power perturbations similar to those described earlier. Empirical eigenfunctions were then generated from these temperature fields using the POD[7, 8] technique.

Figure 3.8 shows the composition of the various multilayers on a MOSFET[17] wafer ready for implant anneal. The various layers of field oxide, gate oxide, polysilicon gate, and

silicon nitride are deposited or thermally grown and patterned. The FEM program used to model the radiative properties of the wafer surface assumes that the field oxide thickness is 0.5  $\mu\text{m}$ , the gate oxide thickness is 0.01  $\mu\text{m}$ , the polysilicon thickness is 0.3  $\mu\text{m}$ , and the silicon nitride is of thickness 0.2  $\mu\text{m}$ . [1] Two different annealing patterns have been used to study the model reduction technique. These are referred to as annealing pattern 1 and 2.

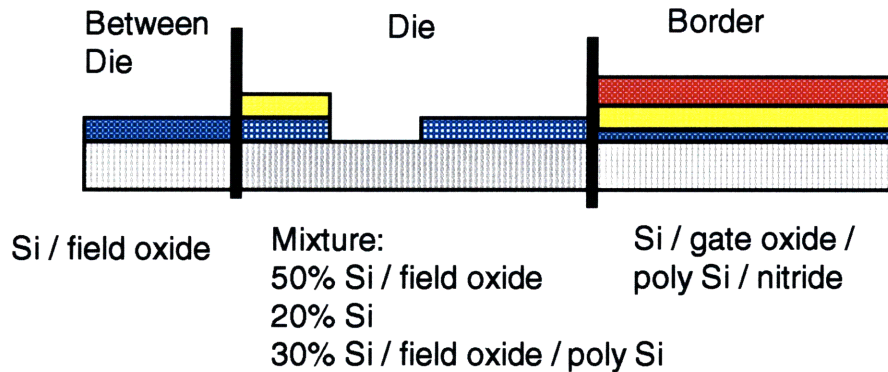


Figure 3.8. Schematic of typical multilayer patterns on a MOSFET wafer.

The emittance values for the various regions on the wafer surface in the three bands for annealing pattern 1 are as follows:

AREA	$\epsilon_1$	$\epsilon_2$	$\alpha_t$
Die	0.68	0.64	0.66
Between die	0.81	0.69	0.74
Border	0.81	0.69	0.80
Backside	0.66	0.66	0.64

Table 3.3. Emittance values in the three bands for annealing pattern 1.

The FEM radiation program was then used to compute a set of radiative exchange factors using the radiative properties given above. Subsequently, a new set of steady state temperature and flow fields was calculated using the FEM steady state program. Two reduced models were

generated using this strategy, one around a wafer steady state temperature of 300K and another around a wafer steady state temperature of 1300K.

To study the transient performance and occurrence of steady state drifts in the reduced models a lamp power protocol was setup to do the following:

1. Maintain the wafer at 300K for 600seconds.
2. Ramp up the wafer temperature to 1300K in 20 seconds.
3. Maintain the wafer at 1300K for 300 seconds.

This trajectory replicates a typical RTP ramp up and also allows the investigation of any steady state drifts in temperature under steady state conditions. In order to replicate this trajectory using reduced models the following procedure was carried out:

1. Start the time integration using the 300K reduced model till the beginning of the ramp up.
2. Switch to the 1300K reduced model just prior to temperature ramp up.
3. Carry out the time integration for the rest of the trajectory using the 1300K reduced model.

The temperature trajectories obtained from the FEM and reduced models for annealing pattern 1 are compared in Figure 3.9.



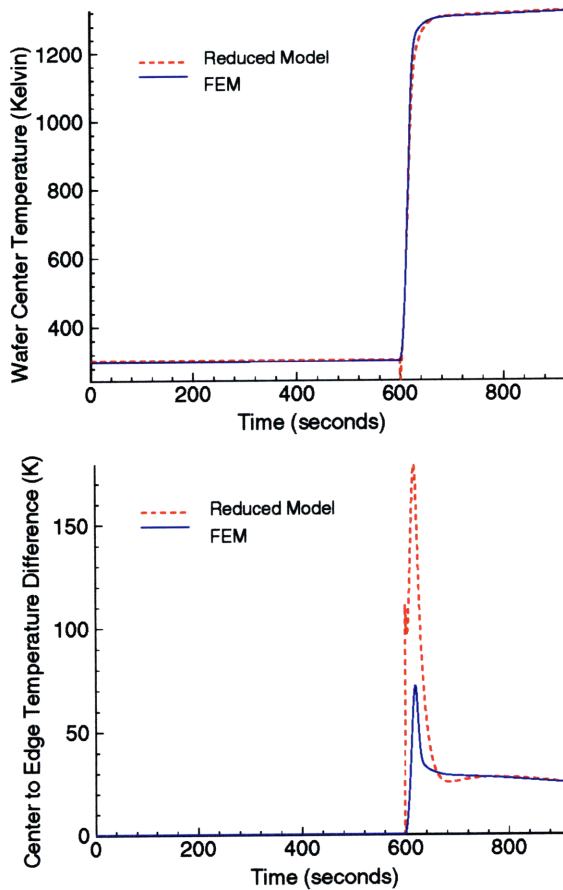


Figure 3.9. Transient temperature trajectory comparisons for wafer-scale pattern effects for annealing pattern 1.

The wafer center temperature as predicted by the reduced model agrees well with the FEM model all over the RTP ramp cycle. But on comparing the wafer to center edge difference, we find that the reduced model over predicts the pattern effect during the ramp up portion of the cycle. But the agreement is good during the two steady state portions of the cycle.

A different set of radiative properties was then chosen for the various regions on the wafer surface. [1] These values are also typical for annealing patterns. The emittance values for the various regions on the wafer surface in the three bands for annealing pattern 2 are as follows:

AREA	$\varepsilon_1$	$\varepsilon_2$	$\alpha_t$
Die	0.69	0.67	0.67
Between die	0.66	0.68	0.65
Border	0.56	0.53	0.64
Backside	0.66	0.68	0.65

Table 3.4. Emittance values in the three bands for annealing pattern 2.

The transient temperature trajectories from the FEM and reduced models are compared in Figure 3.10. Again the reduced model shows good agreement with the FEM model in tracking the overall behavior of the transient temperature trajectory, but over-predicts the pattern effects across the wafer surface.

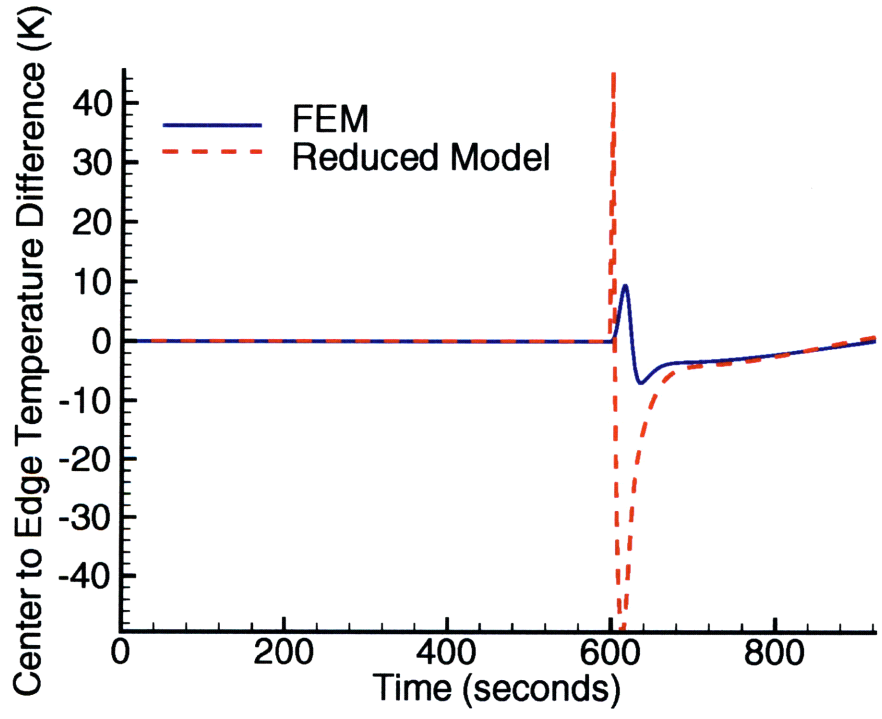
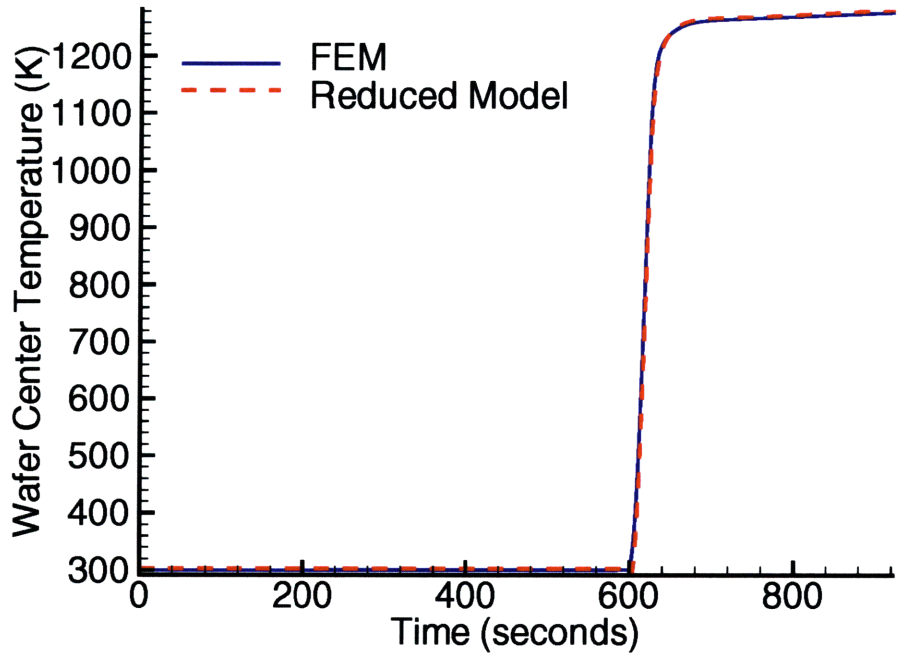


Figure 3.10. Transient temperature trajectory comparisons for wafer-scale pattern effects for annealing pattern 2.

### 3.2.3.2 Titanium Silicidation

Titanium silicidation of source, drain, and gate electrode is now a standard step in the fabrication of a typical MOSFET wafer, for lowering contact resistance and acting as a diffusion barrier. This is one of the few process steps for which RTP is routinely used in production. Titanium silicidation by RTP is typically a two step process. [18] In the first step, Ti is deposited on one side of the wafer by sputtering. Then, a low temperature anneal (target temperature ~ 950 K) is carried out to convert the Ti to  $\text{TiSi}_2$  in areas where the Ti is in contact with single crystal silicon or polysilicon. The unreacted Ti is then etched off, and a second higher temperature (target temperature ~ 1100 K) is done to change the silicide to a low resistance phase. In the FEM modeling work [1], it is assumed that these two steps take place immediately after the implant anneal, to form a silicide on the polysilicon and single crystal silicon. Therefore the thickness of the various layers are as follows: poly = 0.3  $\mu\text{m}$  polysilicon, nitride = 0.2  $\mu\text{m}$   $\text{Si}_3\text{N}_4$ , field oxide = 0.5  $\mu\text{m}$   $\text{SiO}_2$ , gate oxide = 100 angstroms  $\text{SiO}_2$ ,  $\text{TiSi}_2$  = 0.1  $\mu\text{m}$   $\text{TiSi}_2$ . The total hemispherical radiative properties for the various layers in the two steps are tabulated below.

PATTERN	AREA	LAYERS	$\epsilon_1$	$\epsilon_2$	$\alpha_t$
Silicide Step 1	Frontside	$\text{TiSi}_2$	0.08	0.01	0.34
	Backside	Bare silicon	0.67	0.67	0.67
Silicide Step 2	Die	20% $\text{TiSi}_2$ 30% $\text{TiSi}_2$ / poly / field oxide 50% field oxide	0.46	0.36	0.55
	Between die	Field oxide	0.82	0.70	0.75
	Border	Nitride / poly / gate oxide	0.80	0.70	0.80
	Backside	Bare silicon	0.67	0.67	0.65

Table 3.5. Emittance values in the three bands for the two silicidation steps.

As explained in the previous section, empirical eigenfunctions were extracted from base case simulations in which the wafer was modeled using the radiative properties of bare silicon. To study both the transient ramp up and steady state drifts, the lamp power protocol described in the earlier section was repeated for the FEM and reduced models. Figure 3.11 shows the comparison of the transient temperature trajectories as obtained from the FEM and reduced models for silicide step 1. As seen from the figure, the reduced model does not agree with the FEM model at the high temperature operating conditions. This is contrary to the results obtained for the annealing studies. Further inspection of the temperature profiles in the RTP chamber as obtained from the FEM and reduced models (Figure 3.12) under the high temperature operating conditions shows that the main difference in the temperature contours exist near the wafer top surface and in the showerhead region. This is because for silicide step 1, the top surface of the wafer has extremely low emittance in band 1 and band 2 when compared to bare silicon (Table 3.2 vs. Table 3.5). This unrealistically low emittance is indicative of experimental errors in the optical constants used and, therefore, needs further experimental studies as suggested by Hebb. [19] As a result, the wafer top surface reflects most of the incident radiation and is cooler compared to the bare silicon case. Due to the proximity of the wafer to the showerhead, a cooler top surface of the wafer also results in a cooler showerhead region. This effect is more evident towards the wafer center as convective cooling effects are not predominant here as in the case of the wafer edge. The difference in the temperature contours represents a change in the qualitative information about the temperature profile that is not captured by the empirical eigenfunctions obtained from the bare silicon case. Hence, the reduced model performs poorly in replicating the temperature profile.

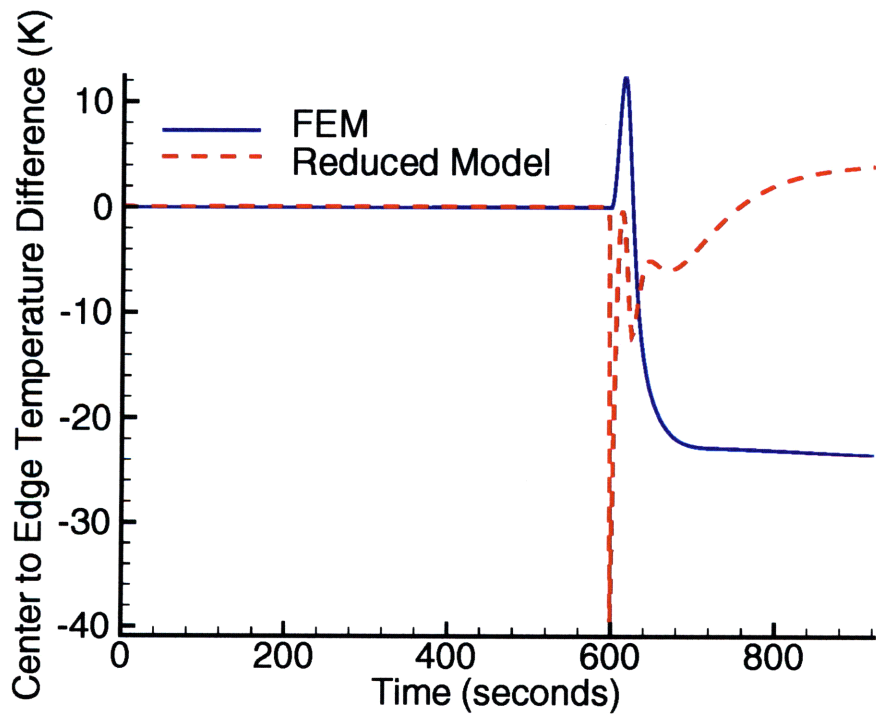
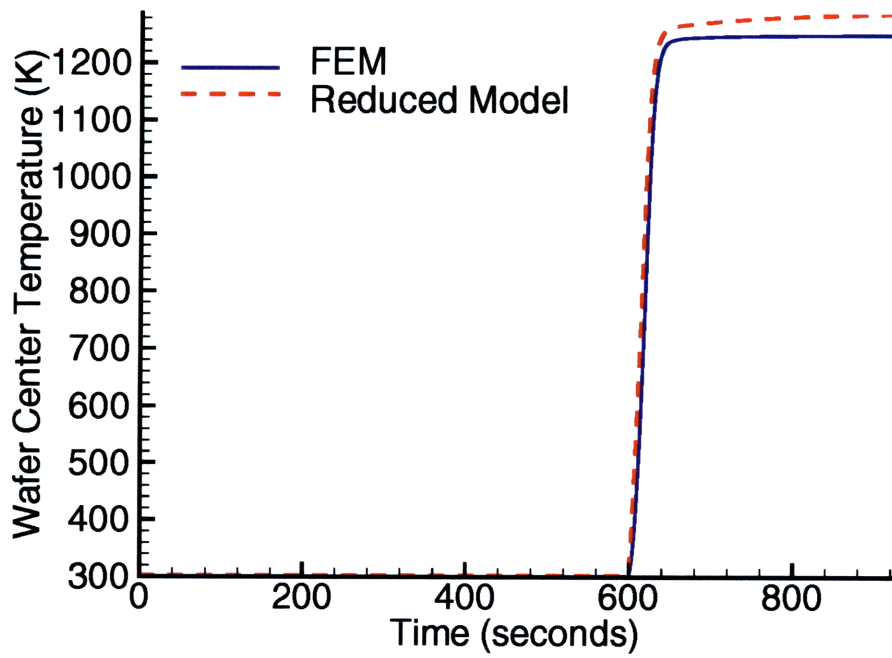
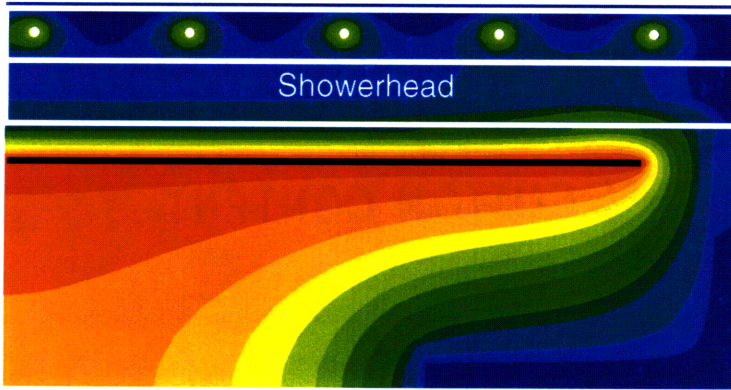


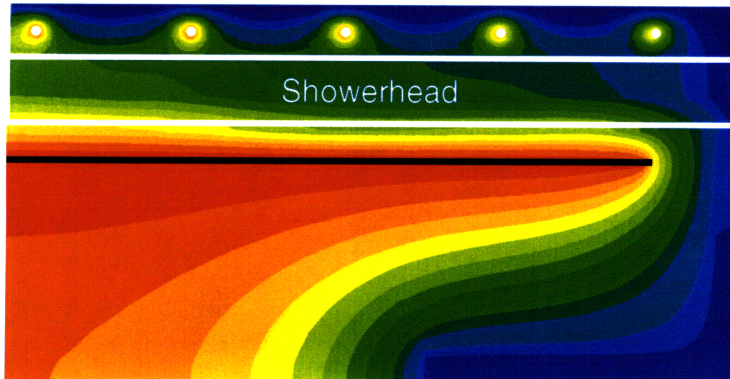
Figure 3.11. Transient temperature trajectory comparisons for wafer-scale pattern effects for silicide step 1.

# Temperature profiles at 900.0 secs for Silicide step 1

FEM Model



Reduced Model



Temp (K)

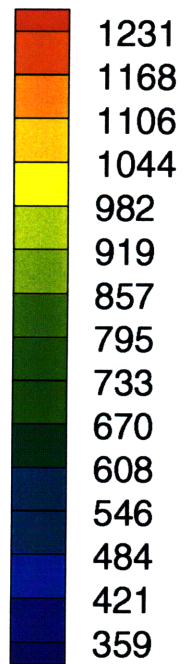


Figure 3.12. Temperature profiles in the RTP chamber at the high temperature processing conditions as obtained from the FEM and reduced models.

Figure 3.13 shows the transient temperature trajectories for silicide step 2. In this case the reduced model trajectory agrees very well with the FEM trajectory all over the cycle. But the pattern effect is over-predicted when we compare the wafer center to edge temperature difference. This result is similar to that obtained from the annealing studies. In this case, since the emittance values are closer to that for bare silicon compared to silicide step 1, the eigenfunction basis set is able to provide an accurate representation of the transient trajectory.

These studies show that the reduced modeling technique is an effective method for detecting the onset of wafer-scale pattern induced temperature changes in an RTP chamber. Hence, this is an effective diagnostic method that could be used by process engineers. Once pattern effects have been detected using reduced models, the FEM model can be used to understand them better and for studying corrective methods. The eigenfunction based reduced models are effective in detecting these pattern effects when the temperature contours are not drastically altered due to the changes in the radiative properties of the wafer. In the case where the radiative properties are drastically altered, a new set of eigenfunctions has to be extracted for the reduced model based on these radiative properties.



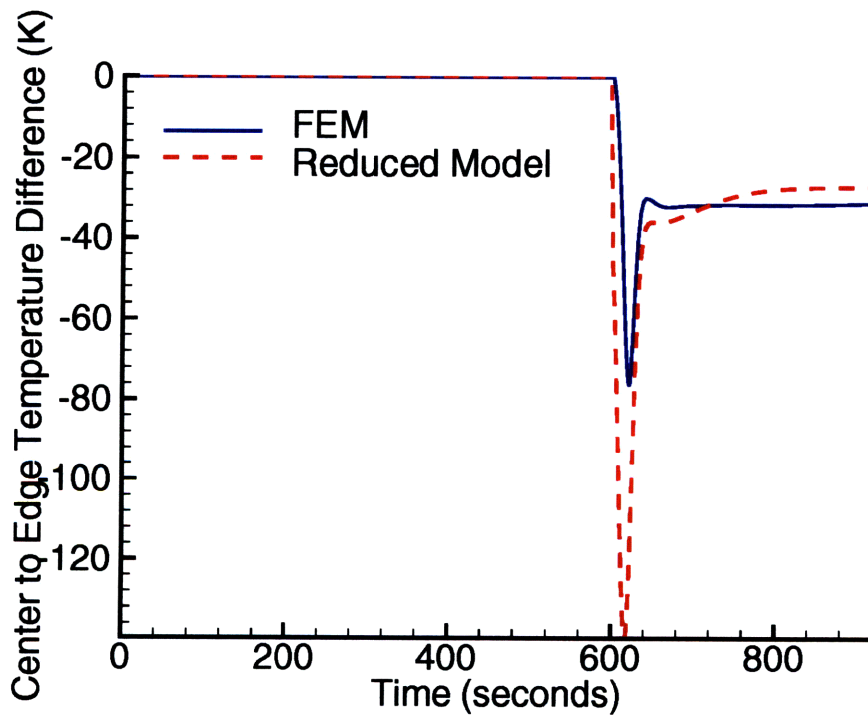
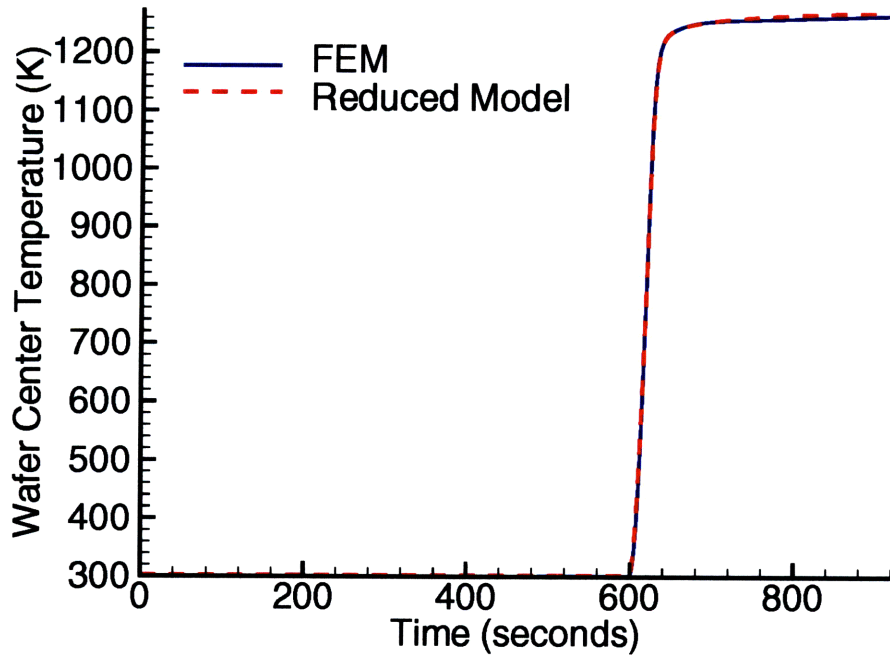


Figure 3.13. Transient temperature trajectory comparisons for wafer-scale pattern effects for silicide step 2.

### 3.2.4 REDUCTION IN COMPUTATION TIME

Timing runs were carried out to find the reduction in computation time using reduced models. For 922 seconds of real process time, which includes both the steady operating conditions and the ramp up phase, the timing runs from the FEM and reduced models on HP-735 workstation are tabulated below in Table 3.6

Processing Step	Finite Element Model	Reduced Model
Annealing	2553 sec	288 sec
Silicide Step 1	2596 sec	136 sec
Silicide Step 2	2587 sec	187 sec

Table 3.6 Comparison of computation time in reduced order modeling of pattern effects.

There is an order of magnitude reduction in computation time by using the reduced models. Also, since the eigenfunction set has to be extracted only once for all the reduced models, the computation time for the extraction (~ 8 hours) is amortized over all of them. Hence, this technique is relatively computationally inexpensive for applications such as in a diagnostic toolkit for detecting the onset of wafer-scale pattern effects in RTP systems.

## REFERENCES

- [1] J. P. Hebb and K. F. Jensen, "The effect of multilayer patterns on temperature uniformity during rapid thermal processing," *J. Electrochem. Soc.*, vol. 143, pp. 1142, 1996.
- [2] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*. New York: John Wiley and Sons, 1960.
- [3] T. P. Merchant, "Modelling of Rapid Thermal Processes," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [4] O. C. Zienkiewicz, *The Finite Element Method*. New York: McGraw Hill, 1977.
- [5] T. P. Merchant, J. V. Cole, K. L. Knutson, J. P. Hebb, and K. F. Jensen, "A systematic approach to simulating Rapid Thermal Processing systems," *J. Electrochem. Soc.*, vol. 143, pp. 2035, 1996.
- [6] K. F. Jensen, T. P. Merchant, J. V. Cole, J. P. Hebb, K. L. Knutson, and T. G. Mihopoulos, "Modeling Strategies for Rapid Thermal Processing: Finite Element and Monte Carlo Methods," in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [7] L. Sirovich, "Turbulence and the Dynamics of Coherent Structures: I, II and III," *Quarterly of Applied Mathematics*, vol. XLV, pp. 561, 1987.
- [8] W. S. Wyckoff, "Numerical Solution of Differential Equations through Empirical Eigenfunctions," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [9] C. Canuto, M. Y. Hussaini, A. Quaderonic, and T. A. Zang, *Spectral Methods in Fluid Dynamics*. New York: Springer, 1988.
- [10] L. Sirovich, "Empirical eigenfunctions and low dimensional systems," in *New Perspectives in Turbulence*, L. Sirovich, Ed. New York: Springer, 1991.

- [11] B. Feil, B. Drew, and B. Moench, "Pattern-induced pattern misregistration after BPSG RTA reflow," *Proceedings of the 1st International Rapid Thermal Processing Conference*, pp. 114, 1993.
- [12] P. Vandenabeele, K. Maex, and R. De Keersmaecker, "Impact of patterned layers on temperature nonuniformity during rapid thermal processing," *Mat. Res. Soc. Proc.*, vol. 146, pp. 149, 1989.
- [13] J. F. Buller, M. Farahani, and S. Garg, "RTA induced overlay errors in a global alignment stepper technology," *Proceedings of the 2nd International Rapid Thermal Processing Conference*, pp. 52, 1994.
- [14] A. Kersch and T. Schaufbauer, "3D simulation and optimization of an RTO chamber with Monte Carlo heat transfer in comparison with experiments," *Proceedings of the 4th International Rapid Thermal Processing Conference*, pp. 347, 1996.
- [15] P. Vandenabeele and K. Maex, "Temperature non-uniformity during rapid thermal processing of patterned wafers," *Proc. SPIE*, vol. 1189, pp. 89, 1989.
- [16] "The National Technology Roadmap for Semiconductors," Semiconductor Industry Association, San Jose, CA 1994.
- [17] S. Wolf and R. N. Tauber, *Silicon Processing for VLSI Era: Volume 1*. Sunset Beach, CA: Lattice Press, 1986.
- [18] M. Miller, "Titanium silicide formation by RTA: Device implications," presented at First International Rapid Thermal Processing Conference, Scottsdale, AZ, 1993.
- [19] J. P. Hebb, "Pattern Effects in Rapid Thermal Processing," in *Mechanical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1997.

# Chapter 4

## Reduced Order Modeling Strategies for Atmospheric Pressure Rapid Thermal Processing Systems

### 4.1 INTRODUCTION

The competitive environment and rapid developments characteristic of the semiconductor industry requires new process technology (including RTP) and equipment be brought into manufacturing as quickly as possible. It is therefore desirable to reduce the number of cut-and-try iterations required for selecting process recipes and process optimization. Modeling of the underlying physical phenomena governing these processes can play a useful role in achieving this goal.

Several approaches have been taken in the modeling of RTP reactors. Lord[1] and later Kakoschke *et al.*[2] were among the first to simulate physical phenomena in RTP systems. In these simple models, a one-dimensional transient heat conduction equation was solved for the silicon wafer in a RTP cycle. The models provide useful insight into the dynamics of RTP processing, but have little predictive capability since both the heat generation and heat loss terms are treated empirically. Campbell *et al.*[3, 4] modeled the steady state flow and temperature fields of an incompressible gas in a RTP reactor. Kersch *et al.*[5] combined steady state

simulations with a detailed radiative heat transfer model to predict steady state temperature profiles. Knutson *et al.*[6] modeled the three dimensional steady state fluid flow and temperature fields in a commercial RTP reactor. All of these models predict steady state temperature profiles, but do not take into account the inherent transient nature of the RTP cycle, hence limiting their applicability to the steady state design of a RTP chamber.

Chatterjee *et al.*[7] solved for the transient wafer temperature profiles in a RTP reactor assuming that the gas in the reactor was stagnant. Merchant *et al.*[8, 9] developed a transient fluid flow and heat transfer model coupled with an accurate radiative heat transfer model to predict temperature profiles in RTP reactors. The radiative heat transfer model computes the radiation energy transport for both specular and diffusely reflecting surfaces in the reactor. The model also allows for the inclusion of wavelength, temperature and material dependent properties.

These modeling efforts have been primarily concentrated on simulations for equipment design and optimization. Additional studies[10-14] have also been conducted to develop process control schemes for RTP systems. There is a need for a class of models that potentially could be used by process engineers to design experiments to implement a new process recipe or answer “what-if” type of questions on a day-to-day basis. Simple models have no predictive capabilities and previous equipment design oriented models are too computationally intensive to be applied in this capacity. In fact, many of the detailed RTP simulations have hardware and software requirements, which could be outside the scope of typical manufacturing organizations. Simulations using detailed, physically based models could take anywhere from a few hours to days to yield a reasonable answer to process questions. Thus, there exists a distinct niche for accurate and fast models to serve the demands of the process engineers.

This chapter describes a modeling technique that can be used to develop fast and accurate models for atmospheric pressure rapid thermal manufacturing processes. The reduced models are built from the same physical rate governing rate governing partial differential equations used to build the detailed finite element (FEM) models. [8, 9] Rapid thermal processes, because of

their fast transient nature, make excellent test pads for analyzing the technique for creating nonlinear low order models.

## **4.2 MODEL FORMULATION**

### **4.2.1 DESCRIPTION OF THE RTP REACTOR**

The Applied Materials Centura™ RTP reactor[15] used for this study consists of a cylindrical chamber separated from the lamp assembly by quartz windows and a quartz adapter plate. The schematic of this reactor is shown in Figure 4.1. The quartz adapter plate consists of a sandwich arrangement of two thin quartz windows with a water-cooled honeycomb in the middle. The area between the windows is pumped out to allow for reduced pressure operation. The wafer is supported by a guard ring, which rests on a quartz cylinder that can be rotated. The process gas is introduced and removed from opposing sides, though for near atmospheric pressure operation at typical wafer rotation rates the system can be approximated as being axisymmetric. The lamp assembly consists of many single lamps arrayed on a hexagonal pitch in individual reflective cells. Groups of individual lamps at approximately the same radial position relative to the wafer center constitute zones to allow control of the wafer radial temperature profile. Multiple sensors measure the temperature across the wafer backside, and these measurements provide the input to a multivariable model-based control system, which dynamically alters the lamp zone power outputs to achieve the desired wafer temperature.

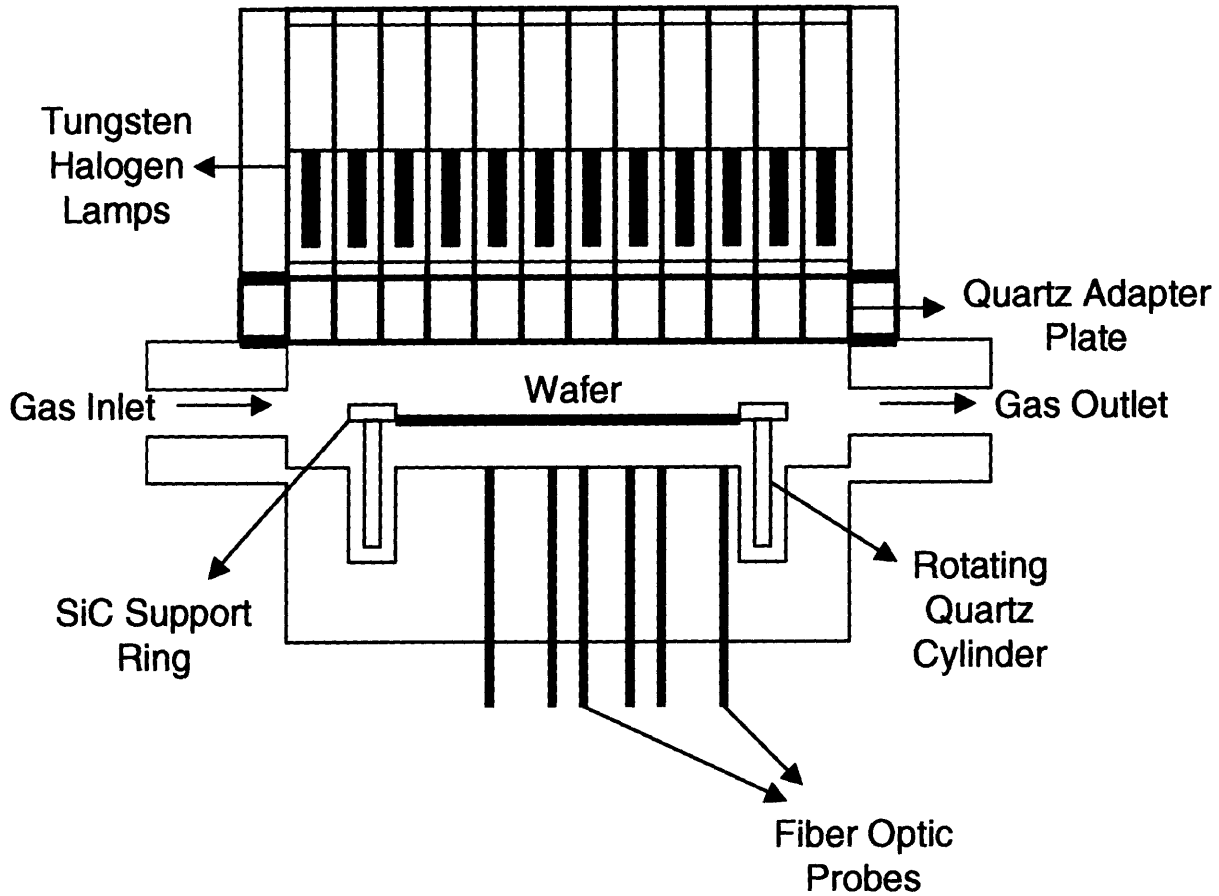


Figure 4.1 Schematic of the Applied Materials Centura™ RTP reactor.

#### 4.2.2 FINITE ELEMENT MODEL FORMULATION

The equations of conservation of mass, momentum and energy [16] have to be solved simultaneously to predict the temperature and gas flow profiles in an RTP system. The form of these equations for the gaseous phase is shown below,

Conservation of mass (Continuity equation):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1)$$



Conservation of momentum:

$$Re\rho\left(\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla\mathbf{v}\right) = \frac{Re}{Fr}(\rho - 1)\mathbf{e}_g - Re\nabla P + \nabla \cdot \boldsymbol{\tau} \quad (2)$$

Conservation of energy:

$$PeC_p\rho\left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T\right) = \nabla \cdot (k^f \nabla T) \quad (3)$$

Here  $\rho$  is the density of the gas,  $\mathbf{v}$  is the velocity vector for the gas phase velocity,  $Re$  is the Reynolds number,  $Fr$  is the Froude number,  $\mathbf{e}_g$  is the unit normal vector in the direction of the gravitational acceleration,  $P$  is the dimensionless pressure which is defined as the dynamic variation in pressure from the overall hydrostatic base pressure,  $\boldsymbol{\tau}$  is the viscous stress tensor,  $Pe$  is the Peclet number for the gas phase,  $C_p$  is the gas phase specific heat,  $T$  is the temperature and  $k^f$  is the thermal conductivity of the gas phase.

For the solid components in the system only the energy equation has to be solved. The conservation of energy equation for the solids can be written as

$$Pe^s C_p^s \rho^s \frac{\partial T}{\partial t} = \nabla \cdot (k^s \nabla T) \quad (4)$$

where  $Pe^s$  is the solid Peclet number,  $C_p^s$  is the specific heat of the solid,  $\rho^s$  is the density of the solid, and  $k^s$  is the thermal conductivity.

A description of the strategy for solving these equations for RTP systems using the Galerkin Finite Element Method (FEM) [17] is given by Jensen *et al.* [8] and Merchant *et al.* [9]. Here we limit the discussion to a brief overview of key features of the FEM technique, which will be relevant to the model reduction approach. The reactor is tessellated into small domains, and the Galerkin form of the governing equations is solved over each of these domains locally.

by expanding the solution in terms of piecewise-continuous functions, biquadratic for temperature and velocity, and bilinear for pressure. The solution for the entire reactor is then obtained by piecing together the solutions from these local domains. For steady state simulations, the time derivative is set to zero and the resulting set of nonlinear algebraic equations is solved to obtain the velocity, temperature, and pressure fields in the RTP reactor. The transient formulation of the problem, consists of the energy conservation equation (Equation 3) with the appropriate boundary conditions. The flow field is approximated to be at pseudo steady state, since the flow field is dominated by wafer rotation driven forced convection throughout the RTP cycle. This assumption has been shown to be valid for low-pressure RTP systems in which the character of the flow solution changes little with wafer temperature. [8, 9] The weak form of the energy equation (Equation 3) for the gas phase is given as,

$$Pe \int_D C_p \rho \left( \frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right) \Phi^i dV = - \int_D (k^f \nabla T) \cdot \nabla \Phi^i dV + \int_{\partial D} (k^f \nabla T) \cdot \mathbf{n} \Phi^i dS \quad (5)$$

where Green's theorem is applied to compute an integral over the boundary ( $\partial D$ ) of differential area  $dS$ , and  $dV$  corresponds to the differential volume of an element. The weak form of the energy equation for solid elements is as follows,

$$Pe^s \int_D C_p^s \rho^s \frac{\partial T}{\partial t} \Phi^i dV = - \int_D (k^s \nabla T) \cdot \nabla \Phi^i dV + \int_{\partial D} (k^s \nabla T) \cdot \mathbf{n} \Phi^i dS \quad (6)$$

The heat flux boundary conditions on the solid surfaces are included in the formulation by replacing the temperature gradient term on the boundary with the appropriate radiation and convection boundary conditions. For solid surfaces adjacent to the gas phase where the velocity is being solved, the boundary condition is given as

$$k^s \nabla T \cdot \mathbf{n} = k^f \nabla T \cdot \mathbf{n} + \left( \frac{L_0 T_0^3 \sigma}{k_0^f} \right) q^r \quad (7)$$

where  $L_0$  is the characteristic length scale,  $T_0$  is the characteristic temperature,  $k_0^f$  is the gas phase thermal conductivity at the characteristic temperature,  $\sigma$  is the Stefan-Boltzmann constant,  $\mathbf{n}$  is the unit normal vector to the solid surface, and  $q^r$  the net heat radiation to the surface. The net heat radiation to a surface  $i$  ( $q^r$ ) is given by

$$q_i^r = \sum_{k=1}^{N_{bands}} [\alpha_i^k \sum_{j=1}^{N_{sw}} \phi_{\lambda^k-T_j} \varepsilon_j^k R_{ij}^k T_j^4 - \phi_{\lambda^k-T_i} \varepsilon_i^k T_i^4] \quad (8)$$

where  $\alpha_i^k$  is the absorptance of solid surface  $i$  in band  $k$ ,  $\phi_{\lambda^k-T_j}$  is the blackbody fraction in band  $k$  at the temperature  $T_j$ ,  $\varepsilon_j^k R_{ij}^k$  is the percentage of radiation in band  $k$  leaving surface  $i$  that is absorbed by surface  $j$  (by direct viewing and all intervening reflections). The exchange factors,  $R_{ij}^k$ , are assumed to be temperature independent based on the opaque silicon properties at high temperatures. This has been shown to be a reasonable approximation for RTP processes. [8, 9]

For surfaces on the exterior of the reactor, which are water cooled, the heat transfer boundary condition is given by

$$k^s \nabla T \cdot \mathbf{n} = \left[ \frac{h_{amb}^f L_0}{k_0^f} \right] (T - T_{amb}) + \left[ \frac{L_0 T_0^3}{k_0^f} \right] \sigma \varepsilon (T^4 - T_{amb}^4) \quad (9)$$

where  $T_{amb}$  is the ambient temperature and  $h_{amb}^f$  is the ambient heat transfer coefficient.

The FEM model of the Centura™ RTP reactor is an axisymmetric representation of the portion of the reactor below the quartz adapter plate. In order to model the lamp heating system, separate Monte Carlo based simulations [8] of an individual lamp and reflector assembly were carried out to determine the spatial and angular distribution of radiation entering the reactor from any one lamp cell. This information was then used to determine the incident flux distribution at the wafer, as a function of radial position, from the lamps at each distinct radial position. Given

a power setting for each lamp zone and the above distributions of the lamp fluxes as input, the incident radiative flux to each radial element on the wafer from the lamps is calculated. This flux is then, either absorbed or redistributed throughout the chamber, based on the wafer optical properties and the radiative exchange factors.

### 4.2.3 REDUCED MODEL FORMULATION

There are over 7000 temperature unknowns in the finite element model of the RTP reactor. This leads to a large nonlinear matrix problem, the solution of which requires significant computational resources. The problem size could, in theory, be reduced drastically if the temperature fields were expanded in terms of basis functions which closely resembled the temperature fields, instead of the locally defined piecewise continuous basis functions used in the finite element method. The Proper Orthogonal Decomposition (POD) method has been used in this work to extract such global basis functions. [18, 19] This technique, also known as the Karhunen-Loève procedure, was first used by Lumley [20] as a rational procedure for the extraction of *coherent structures*. [18] By this method, empirical eigenfunctions can be obtained for a given discrete data set. The data set can be obtained from experiments or, as in this case, from detailed model predictions. In order to obtain a set of empirical eigenfunctions for a group of temperature fields, the transient finite element model is simulated with a set of lamp powers till the temperature everywhere in the reactor has stabilized to near steady state conditions. Then each of the twelve lamp zones is perturbed individually to generate a group of transient temperature fields. These transient fields, obtained by perturbing the finite element model around a particular steady state, contain information about the behavior of the system around that steady state. The steady state temperature field is then subtracted from the individual transient temperature fields and the deviation fields (snapshots) are stored in matrix  $\mathbf{X}$  as shown below,

$$\mathbf{X} = \{\tilde{\tilde{T}}(t_1), \tilde{\tilde{T}}(t_2), \dots, \tilde{\tilde{T}}(t_n)\} \quad (10)$$

$$\tilde{\tilde{T}}(t_n) = \tilde{T}(t_n) - \bar{T} \quad (11)$$

Here  $\bar{T}$  is the vector which contains the steady state temperature field for the entire reactor,  $\tilde{T}(t_n)$  is the temperature field obtained at the  $n$ th time instant from the lamp zone perturbations, and  $\tilde{\tilde{T}}(t_n)$  is the deviation temperature field at the  $n$ th time instant. A temporal correlation matrix,  $\mathbf{C}$ , is then constructed from the matrix  $\mathbf{X}$  as follows,

$$\mathbf{C}_{mn} = \langle \tilde{\tilde{T}}(t_m), \tilde{\tilde{T}}(t_n) \rangle \quad (12)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product in the  $\ell_2$  norm. The eigenfunctions,  $\mathbf{u}_i$ , are obtained by projecting the left singular vectors of the temporal correlation matrix onto the snapshots,

$$\mathbf{C} = \mathbf{V}\Sigma\mathbf{W} \quad (13)$$

$$\mathbf{u}_i = \left( \frac{1}{\sqrt{\Sigma_{ii}}} \right) \sum_{k=1}^N \tilde{\tilde{T}}(t_k) \mathbf{v}_k \quad (14)$$

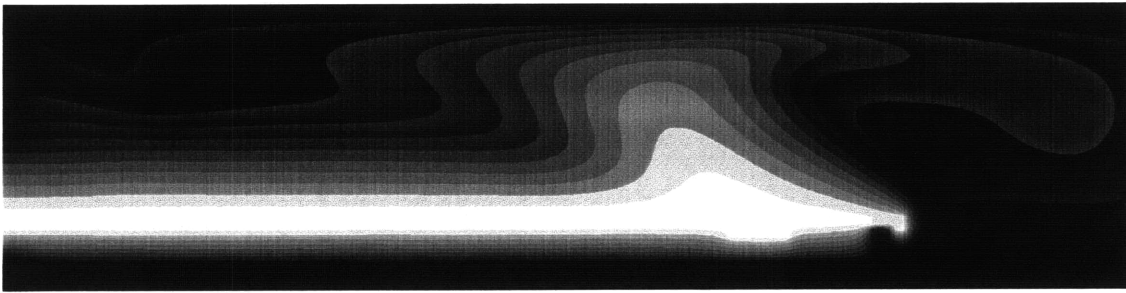
$\mathbf{V}$  is a matrix whose columns are the left singular vectors of  $\mathbf{C}$ ,  $\Sigma$  is a diagonal matrix with the singular values of  $\mathbf{C}$  on the diagonal, and  $N$  is the number of snapshots. The eigenfunctions obtained from this technique are admixtures of the snapshots. [21, 22] Figure 4.2 compares a typical temperature field to the dominant eigenfunction (eigenfunction associated with the largest singular value) extracted by the POD method. The dominant eigenfunction, extracted by this method, has most of the qualitative information about the temperature fields. This is demonstrated in the figure, where the contours of the temperature field and the eigenfunction match closely. Hence, these empirical eigenfunctions can be viewed as ideal basis functions for use in a pseudospectral [23] Galerkin expansion [24] of the conservation of energy equation (Equation 3). The method of using empirical eigenfunctions as a basis set has been used for

modeling turbulence and large-scale flow problems in fluid mechanics. [18, 21, 22, 24-27] In our case, the objective is to build a reduced model which provides an initial value problem for the transient form of the energy conservation equation. This model, when simulated, would then yield a set of temporal coefficients,  $a_i(t)$ , which would be used to generate temperature fields,  $\mathbf{x}(t)$ , using the following equation,

$$\mathbf{x}(t) = \bar{T} + \sum_{i=1}^N a_i(t) \mathbf{u}_i \quad (15)$$

where  $N$  is the order of the reduced model, *i.e.* the number of eigenfunctions used in constructing the reduced model.

### Temperature Field



Temp. (Kelvin) 398 492 586 681 775 869 963 1057 1151 1245

### Dominant Eigenfunction

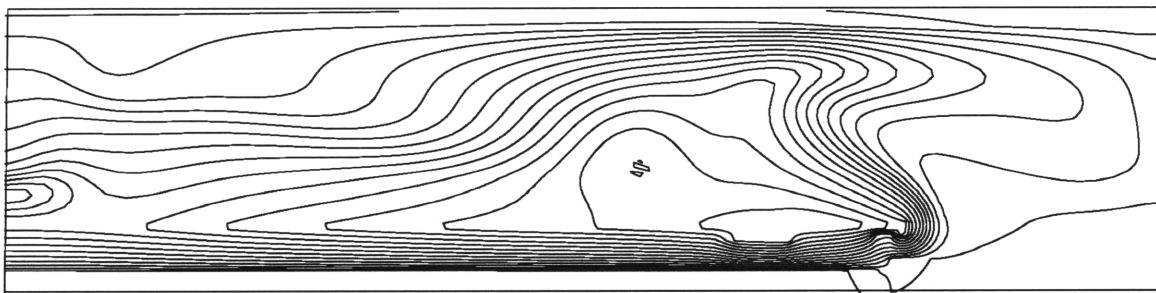


Figure 4.2 Comparison of a typical temperature snapshot obtained from the FEM model with the dominant eigenfunction extracted by the POD procedure.

The next step in the model reduction procedure is to express the energy conservation equation as formulated within the finite element framework (Equations 5 and 6) in a form amenable to the model reduction procedure. In RTP systems, the main mode of heat transfer is by radiative heat exchange, which is by far the major nonlinear term in the finite element model. In order to express the nonlinear characteristics of the system accurately, the radiation heat flux terms are separated from the conduction and convection terms. In this manner, the weakly nonlinear terms are separated from the strongly nonlinear ones. The conduction and convection terms are then linearized about a given steady state operating condition. Following this, Equations 5 and 6, along with the appropriate boundary conditions, can be written as

$$\mathbf{M}(\bar{\mathbf{x}}) \frac{d\mathbf{x}}{dt} = \mathbf{C}(\bar{\mathbf{x}})\mathbf{x} + \mathbf{R}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{l} \quad (16)$$

where  $\mathbf{x}$  is the vector comprising of the states of the system. In this case, the temperatures at every node in the finite element mesh are the states of the system.  $\mathbf{M}(\bar{\mathbf{x}})$  represents the dynamic contribution to the energy conservation equation evaluated at the steady state conditions given by  $\bar{\mathbf{x}}$ ,  $\mathbf{C}(\bar{\mathbf{x}})$  contains all the convection and conduction terms evaluated at steady state conditions,  $\mathbf{R}(\mathbf{x})$  is the radiation contribution,  $\mathbf{G}(\mathbf{x})$  is the transformation matrix which transforms the lamp powers,  $\mathbf{l}$ , to the appropriate heat flux contribution for the energy conservation equation. The material properties are linearized and included in the matrices  $\mathbf{M}(\bar{\mathbf{x}})$  and  $\mathbf{C}(\bar{\mathbf{x}})$ . Properties such as the density of the gas,  $\rho$ , and the specific heat,  $C_p$ , are evaluated at the given steady state conditions and included in the matrix  $\mathbf{M}(\bar{\mathbf{x}})$ . Other properties such as the thermal conductivity,  $k^f$  and  $k^s$ , and the heat transfer coefficient to the ambient fluid,  $h_{amb}^f$ , are similarly included in  $\mathbf{C}(\bar{\mathbf{x}})$ .

The reduced model generation method is implemented in deviation variables, *i.e.* the steady state temperature field is subtracted from the data set, and the empirical eigenfunctions are extracted from the corresponding deviation fields. The reduced model developed in this manner is exact at the given steady state, but would start to deviate from the actual system when



the range of operation is stretched beyond that particular steady state. The agreement between the reduced model and the actual system can be systematically improved by expressing the nonlinear terms explicitly, starting with the highly nonlinear terms. This is the motivating factor for separating the weakly nonlinear terms from the highly nonlinear ones. The following discussion further shows how the major nonlinear terms are handled explicitly.

In the energy conservation equation used to model RTP systems, the  $T^4$  term in the radiation boundary condition (Equation 9) is the main nonlinearity that prevents a full linearization of the modeling equations to hold over a wide range of conditions. Hence this term has to be evaluated explicitly at each time step by reconstructing the absolute temperature fields from the temporal coefficients,  $a_i(t)$ , using Equation 15. Also, the FEM model uses a two-band approximation for the partial transmission by quartz in different wavelength ranges. The quartz is treated as transparent for wavelengths shorter than 4  $\mu\text{m}$  and opaque for wavelengths longer than 4  $\mu\text{m}$ . [8, 9] A nonlinear function is used to decide the fraction of radiation in each of the two wavelength bands. This two-band formulation is also retained in the reduced models. The radiation contribution is, therefore, separated into two matrices for each of the two bands. The radiation contribution to the reduced model from each band is dynamically calculated by multiplying the matrices by a fraction indexed to the temperature. After expressing these nonlinearities explicitly, Equation 16 is reformulated in deviation variables as

$$\mathbf{M}(\bar{\mathbf{x}}) \frac{d\tilde{\mathbf{x}}}{dt} = \mathbf{C}(\bar{\mathbf{x}})\tilde{\mathbf{x}} + \xi_1 \mathbf{R}_1(\bar{\mathbf{x}})[\mathbf{x}(t)]^4 + \xi_2 \mathbf{R}_2(\bar{\mathbf{x}})[\mathbf{x}(t)]^4 + \mathbf{G}(\bar{\mathbf{x}})\mathbf{I} + \mathbf{K} \quad (17)$$

where  $\tilde{\mathbf{x}}$  is the deviation temperature field,  $\mathbf{R}_1(\bar{\mathbf{x}})$  and  $\mathbf{R}_2(\bar{\mathbf{x}})$  are the radiation contributions in the two bands evaluated at steady state conditions,  $\xi_1$  and  $\xi_2$  are the fractions of radiation contribution in each of the two bands, and the matrix  $\mathbf{K}$  arises from combining the steady state contribution of the heat transfer to the ambient and the steady state radiation, including the radiation heat exchange from the lamps. The temperature dependence of the optical properties of the wafer can cause nonlinear variation in not only the absorption of energy from the lamps

by the wafer but also the radiative exchange factors between the various surfaces. These effects are weakly nonlinear compared to  $T^4$  dependence for emitted radiation and the energy distribution in the two-band approximation. Hence, the wafer properties and exchange factors are evaluated at the steady state conditions and incorporated in the matrices  $\mathbf{R}_1(\bar{\mathbf{x}})$  and  $\mathbf{R}_2(\bar{\mathbf{x}})$ .

The empirical eigenfunctions,  $\mathbf{U}$ , obtained from the POD method are then used in a pseudospectral Galerkin expansion of Equation 17 to yield

$$[\mathbf{U}^T \mathbf{M}(\bar{\mathbf{x}}) \mathbf{U}] \frac{d\mathbf{a}}{dt} = [\mathbf{U}^T \mathbf{C}(\bar{\mathbf{x}}) \mathbf{U}] \mathbf{a} + \{\xi_1 [\mathbf{U}^T \mathbf{R}_1(\bar{\mathbf{x}})] + \xi_2 [\mathbf{U}^T \mathbf{R}_2(\bar{\mathbf{x}})]\} \{\mathbf{x}(t)\}^4 + [\mathbf{U}^T \mathbf{G}(\bar{\mathbf{x}})] + [\mathbf{U}^T \mathbf{K}] \quad (18)$$

where  $\mathbf{a}$  is the vector of temporal coefficients. All the terms expressed within square parentheses are evaluated prior to solving the initial value time integration problem. The term,  $\{\mathbf{x}(t)\}^4$ , is evaluated by calculating the absolute temperature fields,  $\mathbf{x}(t)$ , at each time step using Equation 15. The computation of the temperature fields at each time step involves reverting back to the higher dimensional temperature field space from the lower dimensional temporal coefficient space. This is computationally expensive, and, therefore, inclusion of nonlinearities involves a trade-off between computation time and accuracy of the model. Therefore, the nonlinear variation of gas density, thermal conductivity, and specific heat with temperature is approximated by a linearization about the local values of these properties at the steady state. The reduced models generated in this manner will be demonstrated to be adequate in replicating the RTP reactor performance as simulated by the FEM model, over a fairly wide range of conditions. However, effects of linearization were evident when the performance of the reduced models were stretched over a large range of conditions. Further improvements in the reduced model performance could be realized by expressing the weak nonlinearities explicitly, but at the expense of increased computation efforts. Examples of reduction in computational efforts through the use of the reduced order modeling technique are also demonstrated.

## **4.3 RESULTS AND DISCUSSION**

### **4.3.1 COMPARISON OF STEADY STATE PERFORMANCE**

The first step in the validation of the performance of the reduced models was to compare them against the perturbations, which formed the basis for extraction of the eigenfunctions used to create the reduced model. Figure 4.3 shows the results of such a comparison for a reduced model generated around a wafer steady state of 1300K. The figure shows the impact of the individual lamp perturbations for each of the twelve zones on the wafer center temperature. The lamp power perturbations caused variations in the temperature fields throughout the reactor when simulated using the finite element model. These fields were then used to extract a set of eigenfunctions, which we call the 1300K eigenfunctions. The eigenfunctions were then used to generate the 1300K reduced model using the reduced modeling strategy described in the earlier section. In order to test the reduced model the same set of lamp power perturbations were simulated using the reduced model and the results compared against the FEM response. As seen from Figure 4.3, the reduced model and FEM responses overlap throughout the range of perturbations.

A similar reduced model was created at low temperature conditions, *viz.* around a wafer steady state of 300K. This model was also compared against the perturbation responses and the results are plotted in Figure 4.4. The minor differences in the two models arise due to the linearization of the gas phase properties in the reduced model. This comparison validates the efficacy of the reduced models in replicating FEM model performance for both high and low temperature operating conditions.

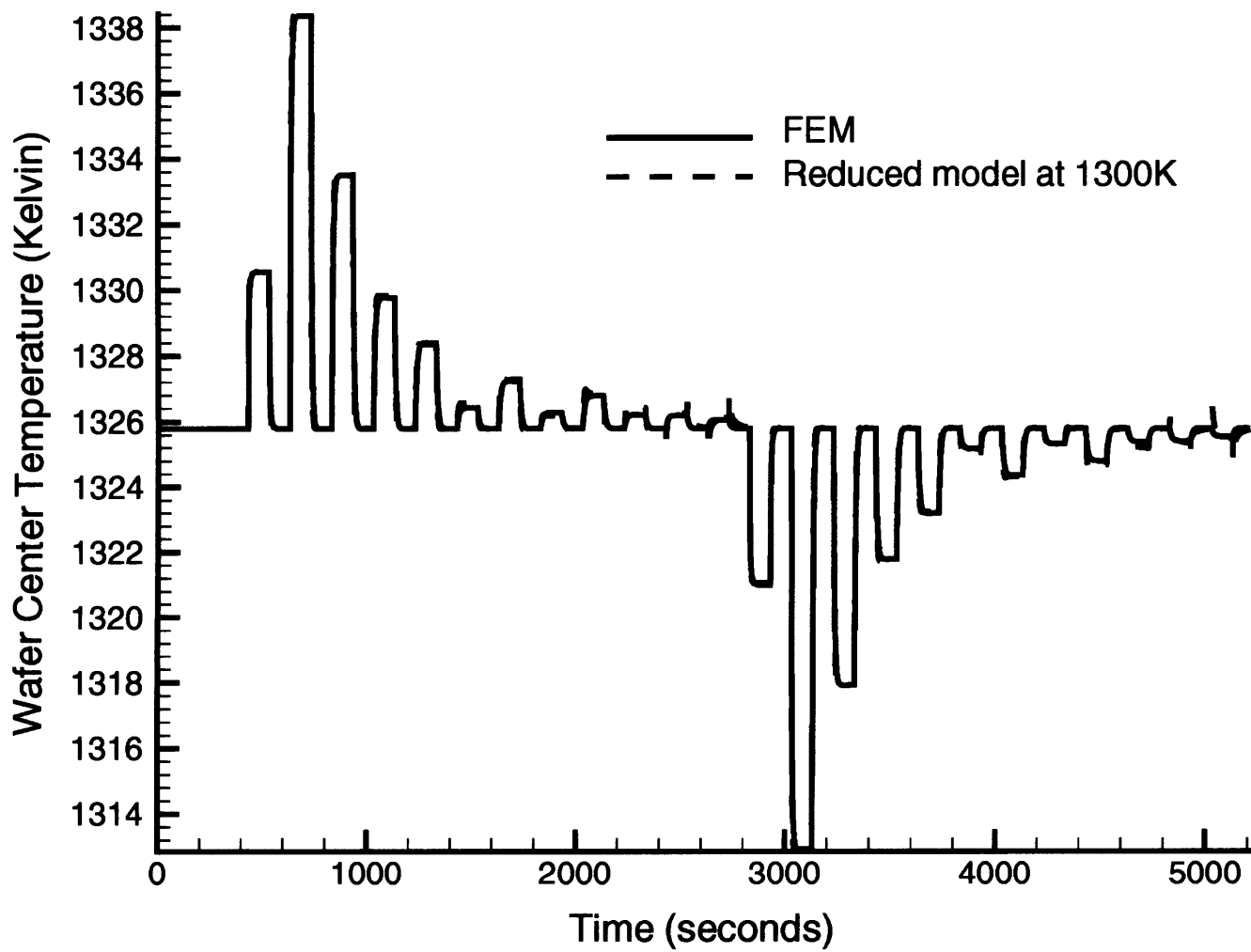


Figure 4.3 Comparison of the wafer center temperature response from the FEM model and the 1300K reduced model.

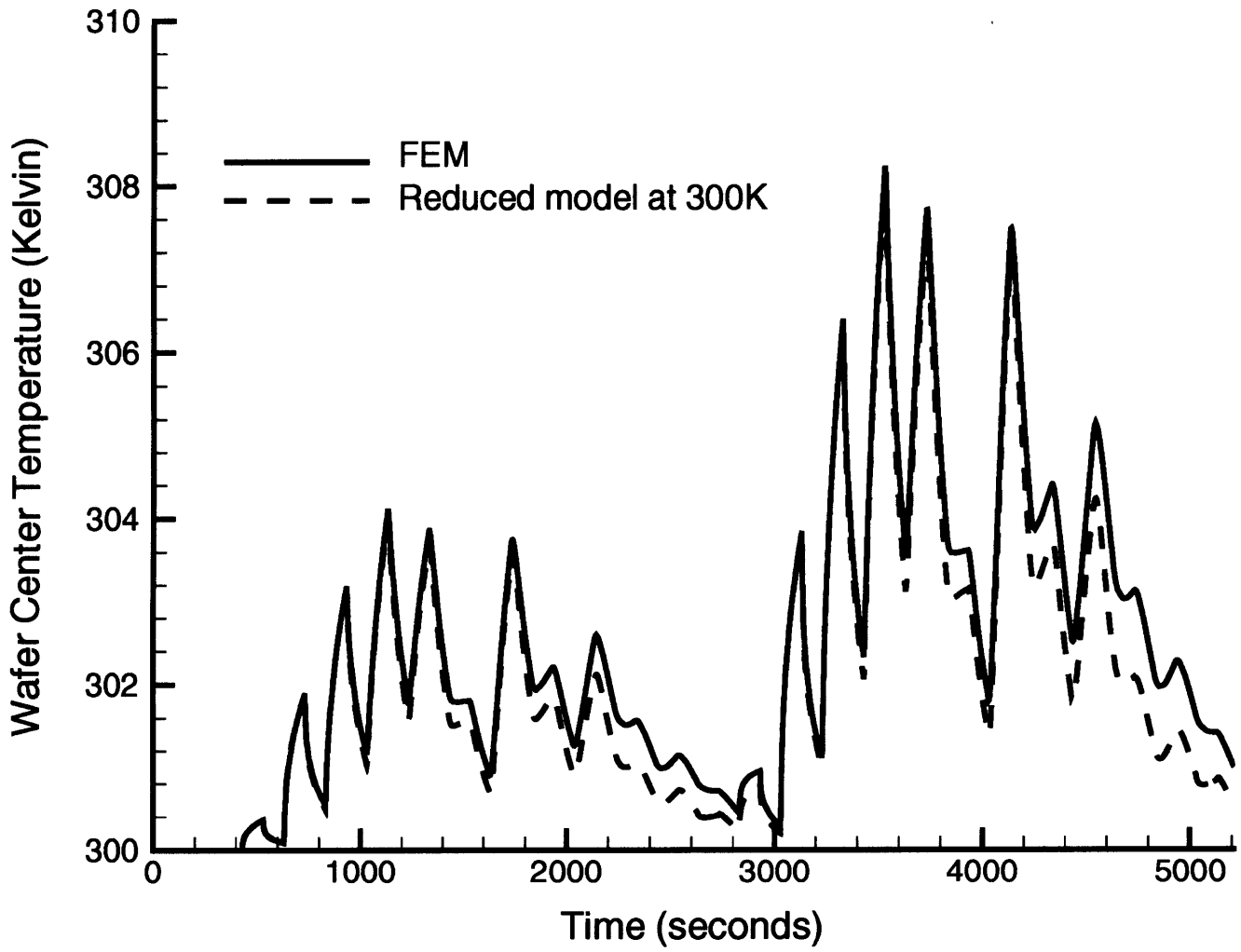


Figure 4.4 Comparison of the wafer center temperature response from the FEM model and the 300K reduced model.

### 4.3.2 TRANSIENT RESPONSES USING REDUCED MODELS

In order to replicate an actual RTP cycle using reduced models, one needs to study the ramp up, stabilization and high temperature hold phases in the cycle. The ramp part of the cycle can be replicated using a reduced model, or multiple reduced models, which gives the same slope for the trajectory resulting from a set of lamp powers. To approximate the stabilization phase, one needs to find a reduced model, which has the same behavior. From both the reduced model generation procedure and from the steady state behavior comparisons, we know that a reduced model would behave exactly like the FEM model around a given steady state. However, since the process does not reach any steady state condition throughout the entire RTP cycle a reduced model had to be found by trial and error.

Figure 4.5 shows the wafer center temperatures through an RTP cycle for three different reduced models and the FEM model. The three reduced models were chosen such that they span the entire temperature range evenly. As seen from the figure, the reduced model generated around the wafer steady state of 300K replicates the RTP cycle exactly at the beginning of the cycle. This model also predicts the initial gradient of the ramp phase correctly, but it begins to deviate from the FEM response around 500K (250 °C). This behavior is expected, as the linearization of the gas phase properties does not hold beyond a certain range. Also, the FEM model uses a temperature dependent absorptance for the wafer surface over the transient simulation. The absorptance for the wafer surface in the reduced models is a linear extrapolation from the corresponding steady state value around which the reduced model is extracted. This linearization does not hold beyond a certain range and hence the wafer temperature trajectory predicted by the reduced model differs from the FEM trajectory. The two other reduced models do not predict the initial steady state properly, again due to the linearization of gas phase properties and the wafer surface absorptance. The values for the gas phase properties and the wafer surface absorptance in these reduced models are linear extrapolations from those at the higher steady state, and as a result they revert to different steady state conditions with the initial

set of lamp powers as inputs. Figure 4.6 shows the comparison of the wafer edge temperatures for the different reduced models with the FEM results. The trajectories show the same trend as the wafer center temperatures due to the same reasons.

In order to replicate the entire RTP cycle, we need to combine the reduced models in a manner such that at different portions of the cycle the response represents the actual response of the reactor. One possible strategy would be to start integrating with the 300K reduced model at the beginning of the cycle and then switch to other reduced models as we proceed along the ramp. Figure 4.7 compares the transient response from such a strategy to the FEM transient response and the behavior of the Applied Materials Centura™ RTP chamber. The same set of lamp power inputs was used to obtain the temperature data from the RTP system and the temperature response from the FEM and combined reduced models. The reduced model trajectory, shown in Figure 4.7, was created by starting the time integration with the 300K reduced model. When the wafer center temperature reached 600K, the 800K reduced model was switched in and used to integrate over the rest of the trajectory. As seen from the figure, the reduced model trajectory deviates from the one predicted by the FEM model immediately after switchover. This behavior stems from switching between different reduced models forcing the time integrator to restart with a new set of initial values for the 800K reduced model. To obtain the new set of coefficients, the transient temperature field at the wafer center temperature of 600K was extracted and the inverse problem was solved in the lower dimensional eigenfunction space. This was done using the QR-Transform method [28] to determine the best least squares solution of Equation 15 for the temporal coefficients. The deviation between the two trajectories immediately after switchover is due to the reinitialization of the time integrator. Since the time integrator only has the initial set of coefficients at this time, and no information about the time derivatives of the coefficients, the slope predicted at this point is incorrect. However, after proceeding along the trajectory the time integrator accumulates the derivative information and returns to the right slope for the trajectory.

The reduced model trajectory agrees very well with the FEM trajectory and the RTP data

around the desired high temperature processing conditions, and also immediately prior to reaching the process temperature. The agreement is also good during the cool-down phase of the cycle. The reduced model response deviates from that of the FEM model at the intermediate temperature stabilization portion of the process. Switching to a different reduced model generated at the stabilization temperature can, in principle, eliminate the discrepancy between the two trajectories. Since the result of the process is dominated by the time at the elevated temperatures, it is sufficient to make the reduced model trajectory agree accurately with the FEM trajectory and the process data at the end of the ramp up and through the high temperature hold. Figure 4.8 shows the reduced model trajectory for the wafer edge temperature. The two figures demonstrate good agreement between the trajectories over the entire wafer surface. To further validate the agreement with the FEM model, temperature fields were extracted at different points throughout the entire transient cycle. Figure 4.9 shows the comparison of the temperature fields at different points into the RTP cycle. The first comparison is at 24 seconds into the cycle. This is around the transition from the 300K reduced model point to the 800K reduced model. The next comparison shows the temperature fields extracted at 30 seconds into the cycle. This is during the intermediate temperature stabilization phase. In both the comparisons, the RTP system is slightly hotter as simulated by the reduced model compared to the FEM model. Most of the discrepancy is in the gas phase near the wafer edge. The last two comparisons in Figure 4.9 show the temperature fields at 55 seconds and 69 seconds, respectively, into the trajectory. The temperature field at 55 seconds is at the processing temperature and the field at 69 seconds is during the cool-down phase of the RTP cycle. The flood plots show good agreement between the temperature fields, which further validates the accuracy of the reduced models at the processing conditions.

Another strategy for replicating the RTP cycle using reduced models would be to integrate both the 300K and 800K reduced models simultaneously and then pick the temperature response from one of the models, or interpolate between the responses, depending on the position in the trajectory. In this strategy, one would choose the response from the 300K



reduced model as the response of the combined reduced models at the beginning of the cycle. As one moves closer to the high temperature processing conditions, the response of the 800K reduced model would become the total response of the reduced model. An interpolation scheme would give the temperature response at intermediate points. The results of such an interpolation strategy is shown in Figure 4.10. The advantage of this strategy is the elimination of the sharp deviation immediately after switchover shown by the reduced model scheme discussed earlier. The disadvantage of this strategy is having to integrate two sets of differential equations instead of one as in the earlier case. This leads to a marginal increase in the computation time for the reduced model strategy.

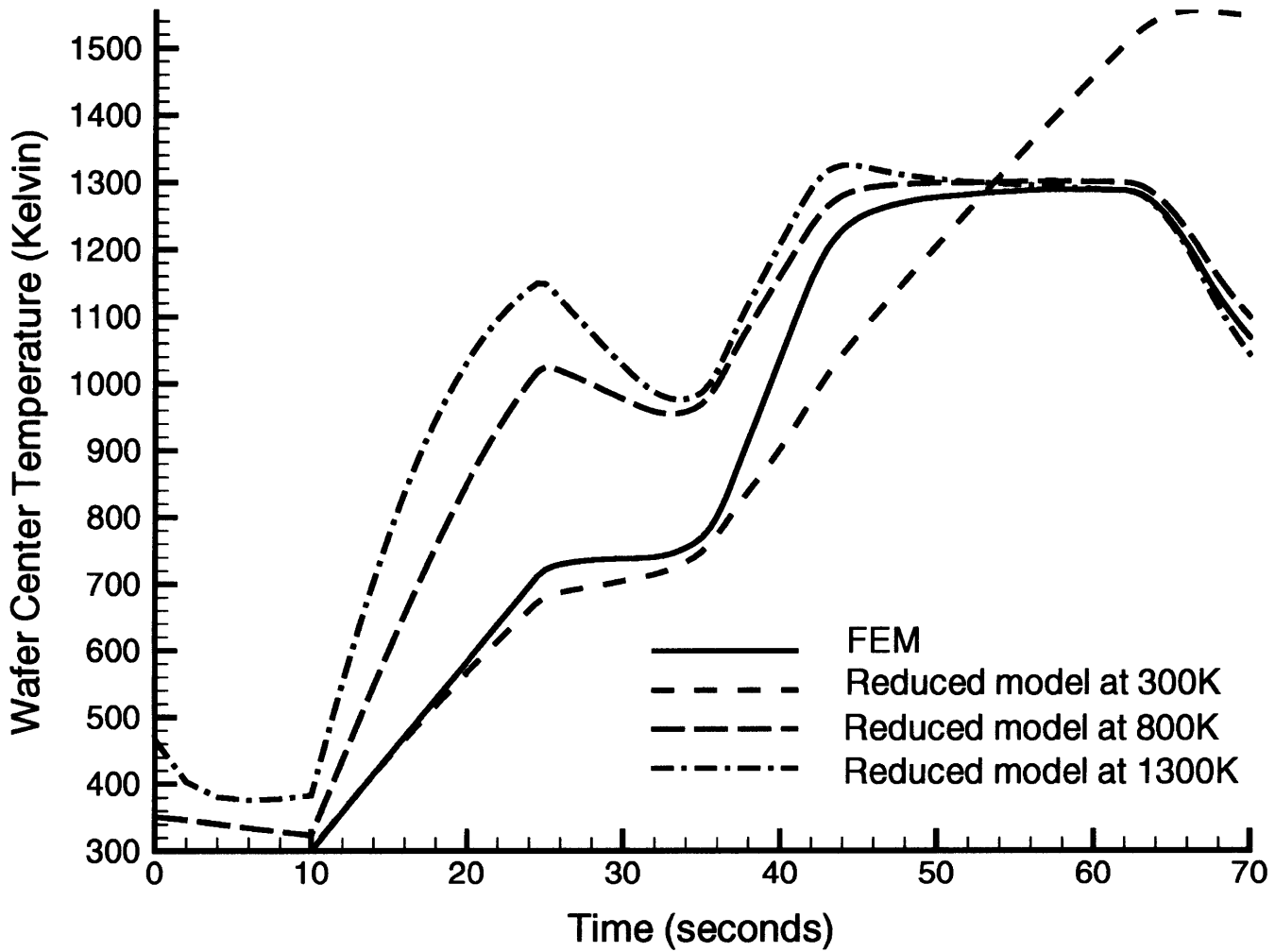


Figure 4.5 Wafer center temperature response for the RTP cycle simulation as obtained from the FEM model and the reduced models.

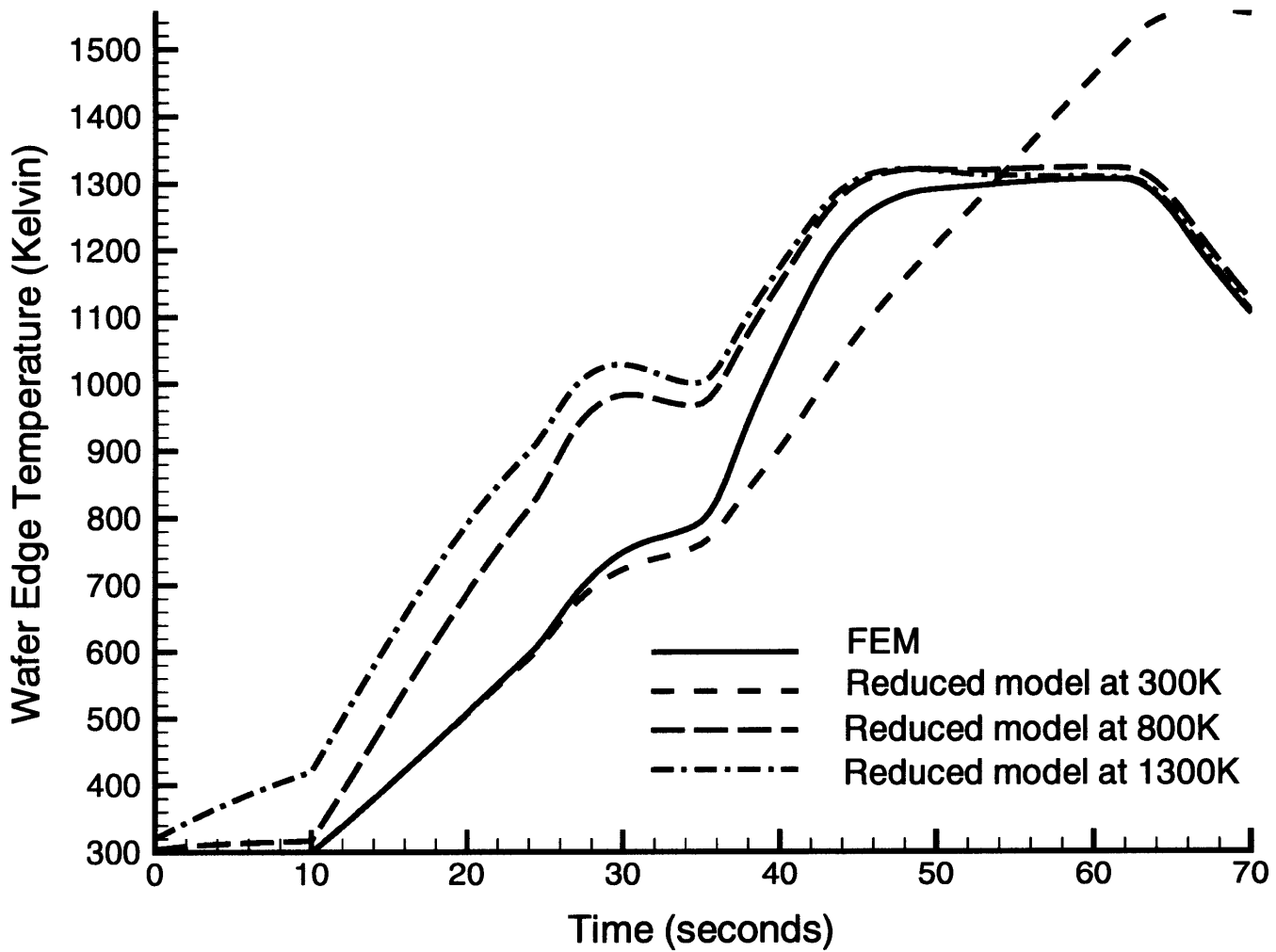


Figure 4.6 Wafer edge temperature response for the RTP cycle simulation as obtained from the FEM model and the reduced models.

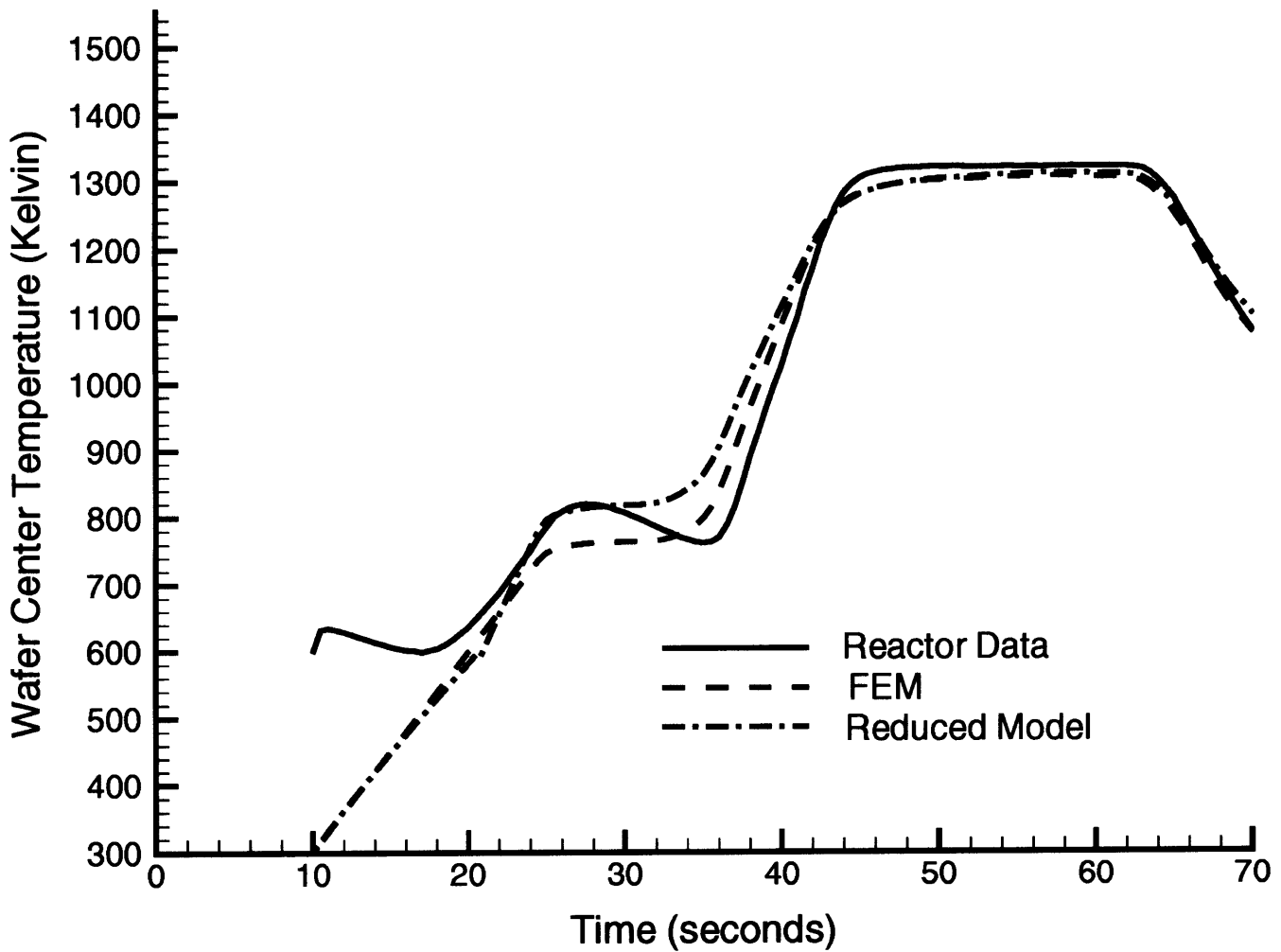


Figure 4.7 Wafer center temperature response of the combined reduced models with switch-over for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.

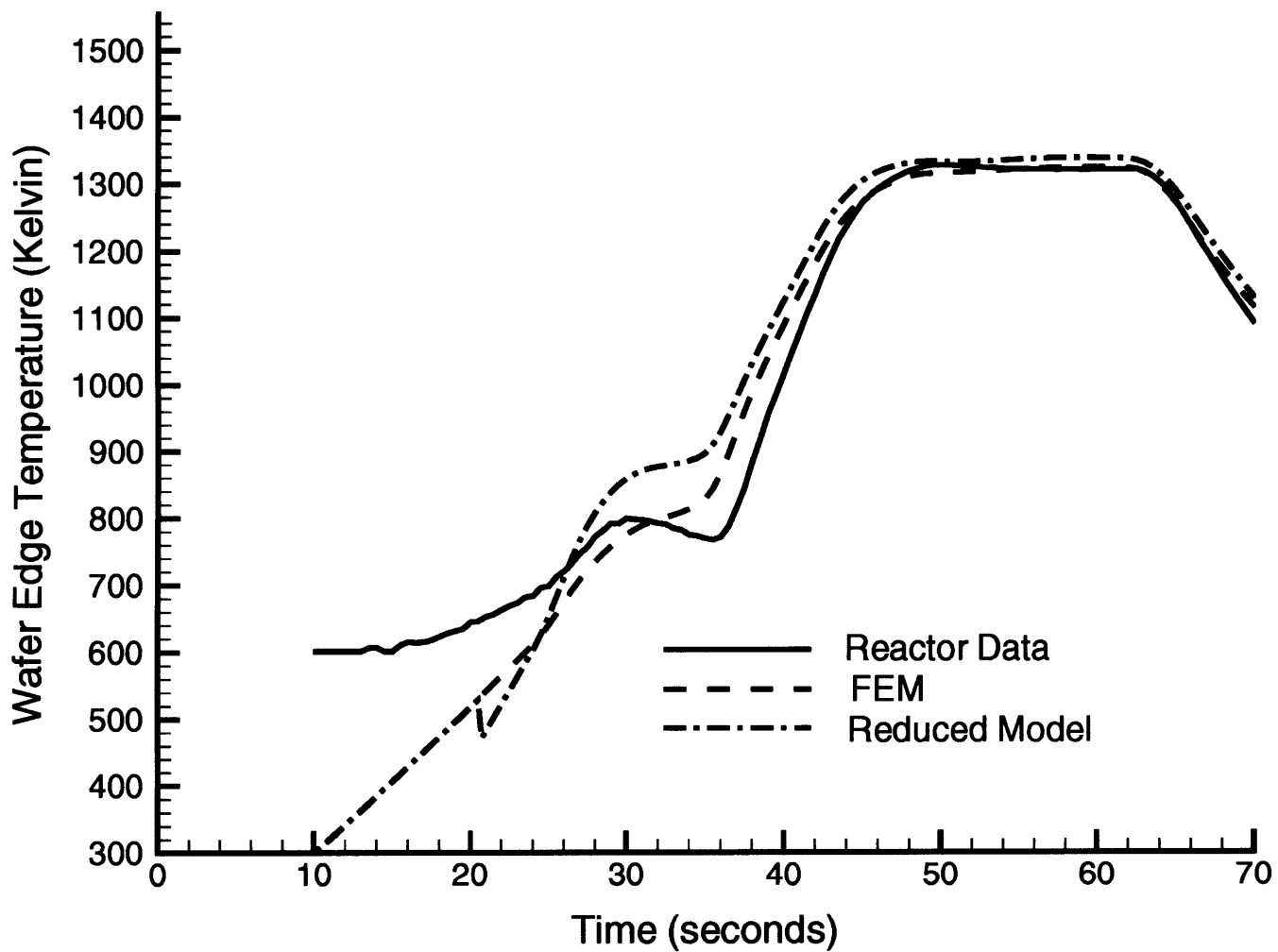
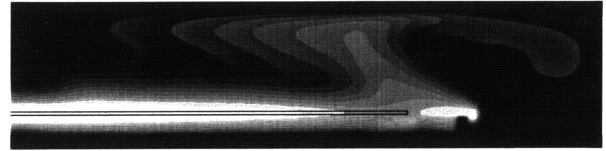
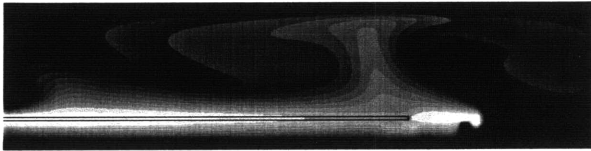


Figure 4.8 Wafer edge temperature response of the combined reduced models with switch-over for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.

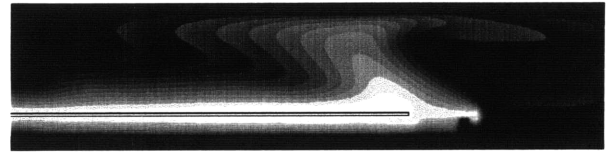
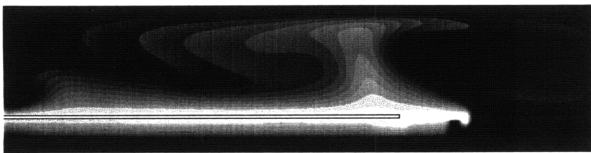
Finite Element Model

Reduced Model

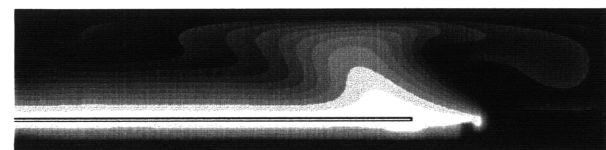
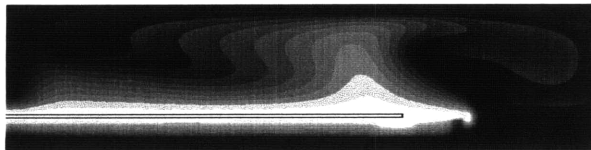
t = 24 seconds



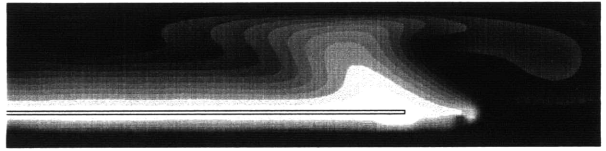
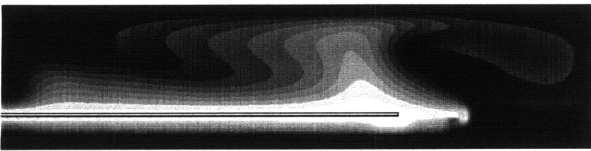
t = 30 seconds



t = 55 seconds



t = 69 seconds



Temperature (Kelvin) 380 475 569 663 757 851 945 1040 1134 1228

Figure 4.9 Comparison of temperature flood plots from the FEM model and combined reduced models at different time instants into the RTP cycle.

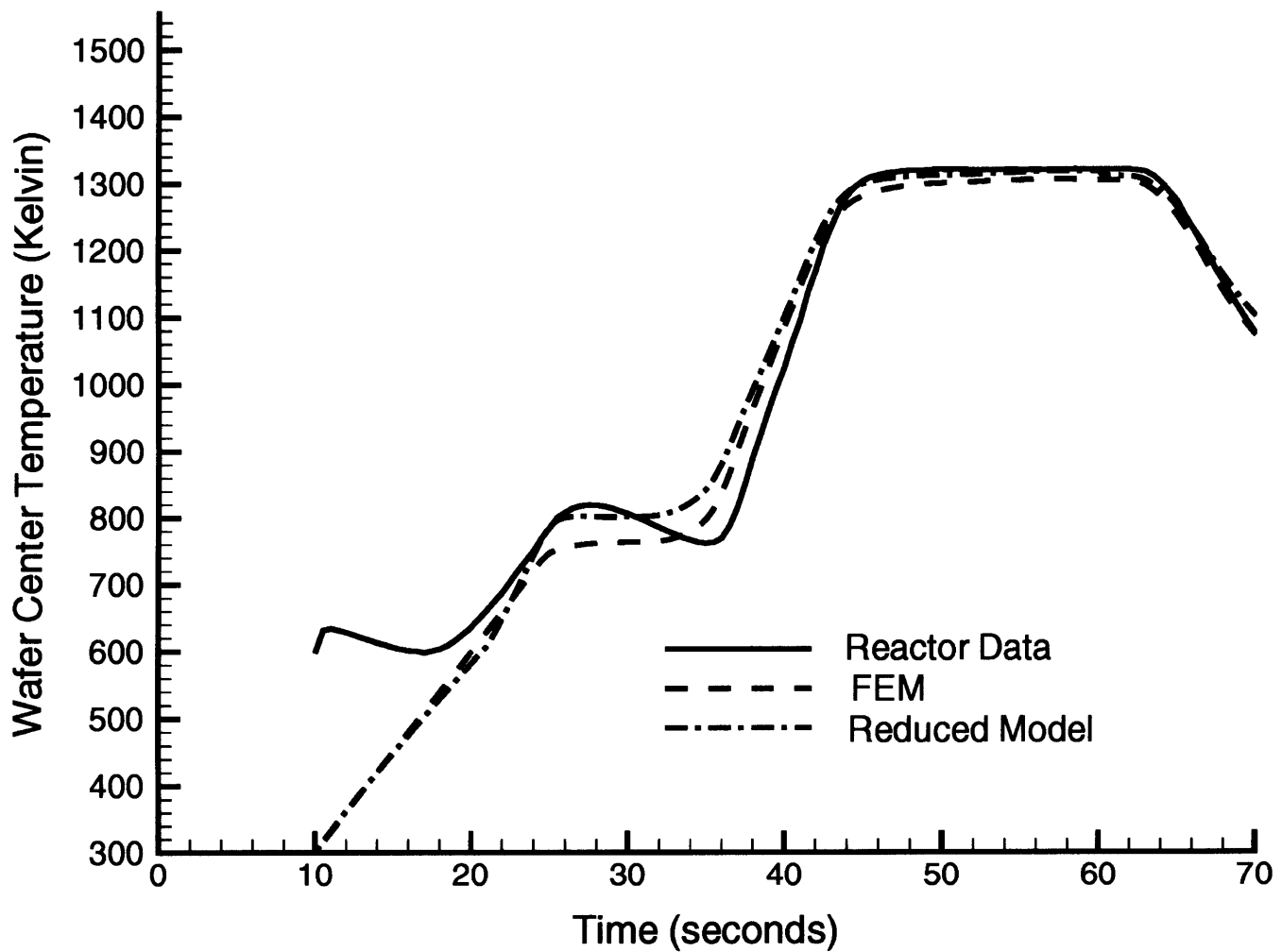


Figure 4.10 Wafer center temperature response of the combined reduced models with interpolation for the RTP cycle compared against the transient response from the FEM model and process temperature data from the Applied Materials Centura™ RTP reactor.

### 4.3.3 REDUCTION IN COMPUTATION TIME

The main motivation for the development of the reduced modeling procedure is the reduction in computation time. In order to compare the computation time, timing runs for the various models are tabulated in Table 4.1. These timing runs were performed on a HP-735 workstation for the 70 seconds RTP cycle.

Model	Execution Time
Finite Element Model	250 minutes
Reduced model with switch-over	4 minutes
Reduced model with interpolation	8 minutes

Table 4.1 Comparison of model execution time.

Both the reduced models show more than an order of magnitude reduction in computation time compared to the FEM model for the RTP cycle. The reduced model with the interpolation scheme takes twice the time as the reduced model with switch-over, since we are integrating two sets of differential equations instead of one.

The main overhead in the POD technique comes in extracting the empirical eigenfunctions from the FEM transient simulation. A typical run to collect ~200 temperature fields, followed by the eigenfunction extraction from these fields takes ~8 hours. Therefore this strategy is suitable for applications in which a few reduced models are extracted and used repetitively, such as in process optimization and control system design. This strategy would not be useful for applications in which the reduced models have to be extracted over and over again, such as in equipment design.



## 4.4 CONCLUSION

The modeling technique described in this chapter provides a systematic framework for the extraction of nonlinear physically based reduced models from complex finite element models. The reduced models are generated from empirical eigenfunctions, which can be extracted from either transient finite element simulations or experimental data. The eigenfunctions extracted by the POD technique contain qualitative information about the data set, and therefore comprise an optimal basis set for the partial differential equations used to describe the system. This empirical eigenfunction basis set is then used to generate reduced models by expanding the rate governing partial differential conservation equations using the pseudospectral Galerkin method. In this expansion, the highly nonlinear terms are identified, separated from the weakly nonlinear ones, and then accounted for explicitly. This strategy provides a systematic procedure for generating reduced models with varying degrees of nonlinearity. The technique, though developed for RTP, can be used to formulate reduced models for systems governed by similar fluid-thermal conservation equations.

The linearization of gas phase properties and absorptance of the wafer surface limits the range of accuracy of the reduced models when compared against the finite element models. Additional nonlinearities can be introduced in the reduced models at the expense of increase in computation time. Model switching, as depicted in the chapter, is one approach to circumvent the complication.

The reduced models (10 unknowns) have very few states when compared against finite element models (7000 unknowns) used to describe the same system. The reduced models extracted by this procedure show good agreement with the finite element models around steady operating conditions and can also be used to predict the behavior of the system beyond the operating conditions around which they were generated. A single reduced model can, therefore, be used for process optimization studies and answering “what-if” type of processing questions spanning a large window in process space ( $\pm 100$  °C). Two reduced models were used to

generate the entire RTP ramp cycle, and the transient trajectory generated by the combined reduced model showed good agreement when compared against both transient finite element simulations and process data. The computation time involved in simulating the reduced model was an order of magnitude less than that involved in similar finite element simulations.

The reduced modeling technique is suitable for designing models that have to be used repetitively, such as in process optimization or formulating process recipes. The strategy also has promise for application in a combined feedforward and feedback control strategy, where these models could be used in a model based feedforward loop. The large computational overhead involved in extracting the eigenfunctions prevents this strategy from being useful during the early stages of equipment design.

## REFERENCES

- [1] H. A. Lord, "Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven," *IEEE Trans. Semicon. Manuf.*, vol. 1, pp. 105-114, 1988.
- [2] R. Kakoschke, E. Bubmann, and H. Foll, "The appearance of spatially nonuniform temperature distributions during rapid thermal processing," *Appl. Phys. A*, vol. 52, pp. 52-59, 1991.
- [3] S. A. Campbell, K.-H. Ahn, K. L. Knutson, B. Y. H. Liu, and J. D. Leighton, "Steady state thermal uniformity and gas flow patterns in a rapid thermal processing chamber," *IEEE Trans. Semicon. Manuf.*, vol. 4, pp. 14-19, 1991.
- [4] S. A. Campbell and K. L. Knutson, "Transient effects in rapid thermal processing," *IEEE Trans. Semicon. Manuf.*, vol. 5, pp. 302-307, 1992.
- [5] A. Kersch, H. Schafer, and C. Werner, "Improvement of thermal uniformity of RTP-CVD equipment by application of simulation," *IEDM Technical Digest*, pp. 883-886, 1991.
- [6] K. L. Knutson, S. A. Campbell, and F. Dunn, "Modelling of three dimensional effects on thermal uniformity in rapid thermal processing of 8 inch wafers," *IEEE Trans. Semicon. Manuf.*, vol. 7, pp. 68-72, 1994.
- [7] S. Chatterjee, I. Trachtenberg, and T. F. Edgar, "Mathematical modelling of a single-wafer rapid thermal reactor," *J. Electrochem. Soc.*, vol. 139, pp. 3682-3689, 1992.
- [8] K. F. Jensen, T. P. Merchant, J. V. Cole, J. P. Hebb, K. L. Knutson, and T. G. Mihopoulos, "Modeling Strategies for Rapid Thermal Processing: Finite Element and Monte Carlo Methods," in *Proceedings of NATO Advanced Study Institute, Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishing, 1996.
- [9] T. P. Merchant, J. V. Cole, K. L. Knutson, J. P. Hebb, and K. F. Jensen, "A systematic approach to simulating Rapid Thermal Processing systems," *J. Electrochem. Soc.*, vol. 143, pp. 2035-2043, 1996.

- [10] G. Aral, T. P. Merchant, J. V. Cole, K. L. Knutson, and K. F. Jensen, "Concurrent engineering of a RTP reactor: Design and Control," *Proceedings of RTP-'94*, pp. 288-295, 1994.
- [11] T. Breedijk, T. F. Edgar, and I. Trachtenberg, "A model predictive controller for multivariable temperature control in rapid thermal processing," *Proc. Amer. Control Conf.*, pp. 2980 - 2984, 1993.
- [12] C. Schaper, "Real time control of rapid thermal processing semiconductor manufacturing equipment," *Proc. Amer. Control Conf.*, pp. 2985-2989, 1993.
- [13] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Control of MMST RTP: Repeatability, uniformity, and integration of flexible manufacturing," *IEEE Trans. Semicon. Manuf.*, vol. 7, pp. 202-219, 1994.
- [14] C. Schaper, M. Moslehi, K. Saraswat, and T. Kailath, "Modeling, identification, and control of rapid thermal processing systems," *J. Electrochem. Soc.*, vol. 141, pp. 3200-3209, 1994.
- [15] B. Peuse, M. Yam, S. Bahl, and C. Elia, "Advances in Temperature Measurement and Control for RTP," *Proceedings of the 5th International Conference on Advanced Thermal Processing of Semiconductors - RTP'97*, pp. 358-365, 1997.
- [16] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*. New York: John Wiley and Sons, 1960.
- [17] O. C. Zienkiewicz, *The Finite Element Method*. New York: McGraw Hill, 1977.
- [18] L. Sirovich, "Turbulence and the Dynamics of Coherent Structures: I, II and III," *Quarterly of Applied Mathematics*, vol. XLV, pp. 561, 1987.
- [19] W. S. Wyckoff, "Numerical Solution of Differential Equations through Empirical Eigenfunctions," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [20] J. L. Lumley, in *Transition and Turbulence*, R. E. Meyer, Ed. New York: Academic, 1981, pp. 215-242.

- [21] L. Sirovich and H. Park, "Turbulent thermal convection in a finite domain: Part I. Theory," *Phys. Fluids A*, vol. 2, pp. 1649-1658, 1990.
- [22] H. Park and L. Sirovich, "Turbulent thermal convection in a finite domain: Part II. Numerical results," *Phys. Fluids A*, vol. 2, pp. 1659-1668, 1990.
- [23] C. Canuto, M. Y. Hussaini, A. Quaderonic, and T. A. Zang, *Spectral Methods in Fluid Dynamics*. New York: Springer, 1988.
- [24] L. Sirovich, "Empirical eigenfunctions and low dimensional systems," in *New Perspectives in Turbulence*, L. Sirovich, Ed. New York: Springer, 1991.
- [25] L. Sirovich, "Chaotic dynamics of coherent structures," *Physica D*, vol. 37, pp. 126-145, 1989.
- [26] L. Sirovich, J. D. Rodriguez, and B. Knight, "Two boundary value problem for Ginzburg Landau equation," *Physica D*, vol. 43, pp. 63, 1990.
- [27] J. D. Rodriguez and L. Sirovich, "Low dimensional dynamics for the complex Ginzburg Landau equation," *Physica D*, vol. 43, pp. 77, 1990.
- [28] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore and London: The Johns Hopkins University Press, 1989.

# Chapter 5

## Reduced Order Modeling for Chemical Vapor Deposition Systems

### 5.1 INTRODUCTION

Chemical vapor deposition (CVD) has been an important technology for the deposition of thin films on silicon wafers from the earliest days of the microelectronics industry. Compared to other deposition techniques, such as sputtering, sublimation, and evaporation, CVD is very versatile and offers good control of film structure and composition, excellent uniformity, and sufficiently high growth rates. [1] Perhaps the most important advantage of CVD over other deposition techniques is its capability of conformal deposition (that is, the capability of depositing films of uniform thickness on highly irregular surfaces). Today, thin films of silicon dioxide, doped and undoped polysilicon, epitaxial silicon, silicon nitride, aluminum, tungsten, and tungsten silicide deposited by CVD play an essential role in manufacturing silicon-based microelectronic circuits.

The tremendous increase in complexity in semiconductor processing in the last decades, leading to the present generation of ultra large scale integration (ULSI) chips with 16 million components or more on one chip, at typical feature dimensions of 0.5  $\mu\text{m}$  or less, has led to ever

increasing demands on the performance of CVD processes. [1] This has caused an exponential growth in the necessary investment of time and money to develop new generations of CVD equipment. Nevertheless, the construction of new CVD reactors is still largely based on trial and error techniques, adjusting existing equipment to meet increasing demands in a purely empirical way. Computational models, based on fundamental laws of physics and chemistry, have proved to be helpful both in equipment as well as process design and optimization.

There are several different reactor types used for various CVD applications in the semiconductor industry. Kleijn[1] presents a detailed review of the various reactor configurations and relevant modeling efforts. Some of the reactor types and the modeling work done on them are presented here. Atmospheric or slightly reduced pressure horizontal duct reactors are used mainly in research and to some extent in the production of compound semiconductors. The state of the art modeling effort for this class of reactors is formed by two-dimensional elliptic fluid flow models in combination with detailed chemistry, including several dozens of gas phase and surface species and reactions, by Jensen *et al.* [2] Atmospheric pressure axisymmetric vertical reactors are widely used for epitaxial growth at near-atmospheric pressures and for low-pressure polycrystalline CVD processes. Some of the notable modeling work for this type of reactors is by Jensen *et al.* [3], Patnaik *et al.* [4] and Fotiadis *et al.* [5] The horizontal hotwall multiple-wafer-in-tube batch reactor, operated at low pressures (30 to 100 Pa), is the most widely used reactor configuration in the silicon-based industry. [1] Amongst others, it is used to deposit doped and undoped polycrystalline silicon, silicon oxide, and silicon nitride. The landmark modeling study for this kind of systems is by Jensen and Graves. [6] With the increasing demands on film uniformities and qualities, the development of new low-pressure CVD processes such as tungsten LPCVD from  $WF_6$ , and with the increase of wafer dimensions to 200-mm diameter and larger, there has been an increasing interest in single-wafer low pressure CVD reactors. Kleijn studied gas phase and surface reactions in poly-silicon LPCVD in this class of reactors. [7] There has also been some modeling work in rapid thermal CVD by Merchant. [8]

All the detailed models for CVD processes have drawbacks in terms of extensive computational resource requirements similar to the detailed RTP models. Hence, there exists a demand for suitable modeling techniques that can be used to develop models, which are compact and fast from a computational point of view.

In this chapter, the reduced modeling technique developed for RTP systems is extended to CVD transport phenomena modeling. The mass conservation equations for chemical species, including gas and surface phase species, are formulated within the reduced model framework using multiple empirical eigenfunctions extracted from species concentration fields as basis functions in a spectral Galerkin expansion.

## **5.2 MODEL FORMULATION**

### **5.2.1 DESCRIPTION OF THE REACTOR**

A schematic of the tube-type axisymmetric single wafer reactor used for the model reduction studies is shown in Figure 5.1. The figure also shows the streamlines for the flow profile in the reactor that is laminar for a wide range of flow rates. A typical temperature field is also depicted. The process gases are introduced at the top of the reactor and exit at the bottom after impinging on the wafer (susceptor). The reactor wall is made up of opaque quartz and is air cooled on the outside. The height of the reactor is 100 mm and the radius is 20 mm.



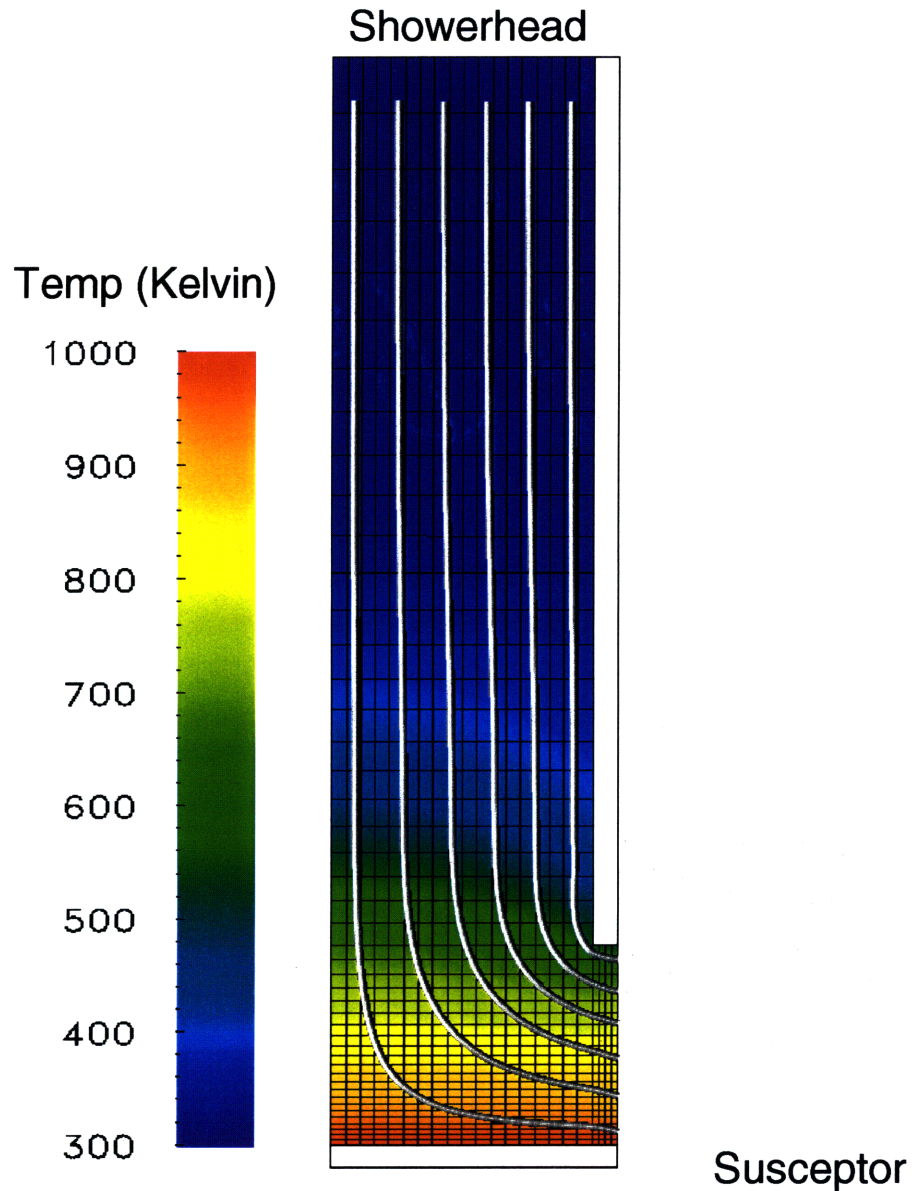


Figure 5.1 Schematic of tube-type axisymmetric single wafer reactor used for model reduction studies.

## 5.2.2 MODELING EQUATIONS AND FINITE ELEMENT FORMULATION

Simulation of CVD processes begins with the fluid flow and thermal simulations as in the case of RTP systems. The reactants are introduced in dilute concentrations in an inert carrier

gas, and unlike reacting flows associated with combustion processes the reactions have only a negligible impact on the flow and temperature distribution. [8] Therefore in all the model development work discussed in this chapter, the pseudo steady state approximation is used for both the temperature and flow fields in the CVD chamber, and they are assumed to be fixed at steady state conditions. The equations representing the mass balances over the individual chemical species are partial differential equations in three spatial dimensions. Because of the chemical reaction terms, the equations are nonlinear and have to be solved numerically. As in the case of fluid flow and heat transfer simulations, finite element methods are the preferred solution strategy because of the natural implementation of flux boundary conditions and handling of complex geometries. The tessellation of the computational domain, which includes not only the fluid flow region, but also the chamber walls and substrate surfaces, is done by the same mesh generation strategy as in the fluid flow and heat transfer simulations in the RTP case.

The nonlinear nature of the species mass balances and the strong coupling between the gas phase and surface reactions complicates the computation of the concentration fields. Special iterative strategies are typically needed for reaction mechanisms involving more than 5 chemical species. Continuation from a known, or an easily solved initial solution, must be used to simulate the concentration fields in such cases. In addition it may be necessary to decompose the problem according to gas phase and surface reactions and solve them separately in the initial stages of the iterative procedure.

Gas diffusion in a CVD reactor may result from concentration gradients (ordinary diffusion), but also from temperature gradients (thermal diffusion, Soret effect). [1] There are many different ways of expressing species concentrations and diffusion velocities. [9] The use of mass fractions and diffusive mass fluxes relative to the mass-averaged velocity of the gas mixture has several important advantages over the use of mole fractions, mole averaged velocities and so forth, the most important of which is the fact that the mass-averaged velocity is directly obtained from the Navier-Stokes equation. The mass averaged velocity  $\hat{v}$  in an  $N$  component gas mixture is [1]

$$\hat{\mathbf{v}} = \sum_{i=1}^N w_i \hat{\mathbf{v}}_i \quad (1)$$

and the diffusive mass flux vector  $\hat{\mathbf{J}}_k$  of the  $k$ th species is defined as

$$\hat{\mathbf{J}}_k = \hat{\rho} w_k (\hat{\mathbf{v}}_k - \hat{\mathbf{v}}) \quad (2)$$

where  $w_k$  is the mass fraction of the  $k$ th species and  $\hat{\rho}$  is the density of the gas. To predict the species transport profiles, the equations of conservation of mass for all individual species have to be solved simultaneously. The equation of conservation of species  $k$  in a gas containing  $N$  species in terms of the mass fraction  $w_k$  and molecular weight  $M_k$  of that species is given by[9]

$$\frac{\partial \hat{\rho} w_k}{\partial \hat{t}} + \hat{\mathbf{v}} \cdot (\hat{\rho} w_k) = \sum_{j=1}^{N_g} \nu_{jk} M_k \hat{\mathcal{R}}_{jk}^g - \hat{\nabla} \cdot \hat{\mathbf{J}}_k \quad (3)$$

The first term on the right hand side represents the net rate of generation of species  $k$  due to  $N_g$  gas phase reactions. The second term on the right hand side accounts for diffusive species transport. The mass fractions of all species must sum to one for the conservation of total mass, hence only  $N-1$  of the species balances shown in Equation 3 are independent. Upon normalization with the appropriate characteristic scales, Equation 3 can be reformulated as follows[8]

$$Re Sc_k \left[ \frac{\partial \rho w_k}{\partial t} + \mathbf{v} \cdot (\rho w_k) \right] = \sum_{j=1}^{N_g} Da_{jk}^g M_k \mathcal{R}_{jk}^g - \nabla \cdot \mathbf{J}_k \quad (4)$$

In this equation,  $Re$  is the Reynolds number,  $Sc_k$  is the Schmidt number and is the ratio of the kinematic viscosity ( $\mu/\rho$ ) to the diffusivity. Thus, the Schmidt number corresponds to the ratio

of the momentum and mass diffusivities. The product of the Reynolds and the Schmidt numbers is also known as the Peclet number ( $Pe_k^m = ReSc_k$ ) and gives the ratio of the momentum flux to the diffusive flux. The Damkohler number ( $Da_{jk}^s = \Re_0 L_0^2 M_0 / D_0 \rho_0$ ) corresponds to the ratio of the reaction rate for species  $k$  reacting in gas phase reaction  $j$  to the diffusion rate of species  $k$  in the medium.

The diffusive molar flux,  $\mathbf{J}_k$ , is composed of the flux due to molecular diffusion generated by the concentration gradients, and the thermal diffusion flux due to temperature gradients, i.e.,  $\mathbf{J}_k = \mathbf{J}_k^C + \mathbf{J}_k^T$ . When the reactants are introduced in dilute amounts in the carrier gas, multicomponent interactions can be neglected altogether. In that case,  $\mathbf{J}_k$  can be expressed as

$$\mathbf{J}_k = -cD_k [\nabla w_k + \alpha_k^T w_k \nabla \ln(T)] \quad (5)$$

where  $c \equiv P/RT$  is the total concentration,  $D_k$  is the binary diffusion coefficient of species  $k$  with the carrier gas,  $\alpha_k^T$  is the thermal diffusion factor of species  $k$  in the carrier gas.

For species  $k$  residing only on the surface the mass balance equation is

$$\frac{d\theta_{kl}}{dt} = \sum_{j=1}^{N_s} M_k Da_{jk}^s \Re_{jk}^s \quad (6)$$

The index  $k$  is given by  $k = 1, 2, \dots, M_l^s$ , where  $M_l^s$  is the total number of surface species on site  $l$  and  $Da_{jk}^s$  is the corresponding surface Damkohler number.  $N_s$  is the number of surface reactions.  $\theta_{kl}$  is the site fraction of species  $k$  on site  $l$  and assuming that the total number of sites are conserved on the surface, the following equality must be satisfied.

$$\sum_{l=1}^{M_l^s} \theta_{kl} = 1 \quad (7)$$

The index  $l$  ranges from  $l=1,2,\dots,L$ , where  $L$  is the total number of sites present on the surface. The site fraction,  $\theta_{kl}$ , and the surface density,  $X_{kl}$ , are related by the expression

$$X_{kl} = \frac{M_k \theta_{kl} \Gamma_l}{\Xi_{kl}} \quad (8)$$

where  $\Gamma_l$  is the molar site density in *moles/m<sup>2</sup>*.  $\Xi_{kl}$  is the number of sites occupied by species  $k$  on site  $l$ .

The Galerkin finite element method[10] is used to solve the species conservation equations. The integral form of the species conservation equation (Equation 3) is

$$\begin{aligned} P e_k^m \int_D \left[ \frac{\partial \rho w_k}{\partial t} + \mathbf{v} \cdot (\rho w_k) \right] \Phi^i dV = \int_D \sum_{j=1}^{N_s} D a_{jk}^s M_k \mathfrak{R}_{jk}^s \Phi^i dV - \\ \int_D c D_k \left[ \nabla w_k + \alpha_k^T w_k \nabla \ln(T) \right] \cdot \nabla \Phi^i dV \\ \int_{\partial D} \mathbf{J}_k \cdot \mathbf{n} \Phi^i dS \end{aligned} \quad (9)$$

where Equation 5 is used for the diffusive flux in the second term on the right hand side. For steady state simulations, as considered in this chapter, the time derivative term is zeroed out from the formulation. For each surface site the following weak form applies

$$\int_{\partial D} \frac{d\theta_{kl}}{dt} \Phi^i dS = \int_{\partial D} \left( \sum_{j=1}^{N_s} M_k D a_{jk}^s \mathfrak{R}_{jk}^s \right) \Phi^i dS \quad (10)$$

The discretized version of the integral form results in a set of nonlinear algebraic equations that have to be solved simultaneously to predict the gas phase and surface species concentrations. The nonlinearity in the governing equations arises solely from the reaction rates for the dilute approximation. Only  $N-1$  gas phase species unknowns are solved directly, with all the dependence on the carrier gas being implicitly introduced by forcing the total mass balance

conservation as

$$w_N = 1.0 - \sum_{k=1}^{N-1} w_k \quad (11)$$

### 5.2.3 MODEL REDUCTION STRATEGY

An overview of the model reduction strategy for chemical vapor deposition systems is shown in Figure 5.2. The model reduction for CVD systems starts with the extraction of empirical eigenfunctions by the proper orthogonal decomposition (POD) method. [11, 12] There is a small variation introduced into the overall eigenfunction extraction method in this case. Instead of using transient fields, as in the RTP case, steady state chemical species concentration fields are used for the eigenfunction extraction. In order to obtain a set of empirical eigenfunctions for a group of species concentration fields, the finite element steady state mass transfer model is simulated for different inlet concentrations of one of the chemical species. The species, whose inlet concentration is perturbed, is chosen in such a manner that the change in the concentration of this species would have a significant impact on the concentration fields of all other species. This particular species is termed as the dominant species. For different values of the inlet concentration of the dominant species, steady state concentration fields for all the chemical species in the system are computed and stored.

## Reduced Chemistry Model Generation Strategy

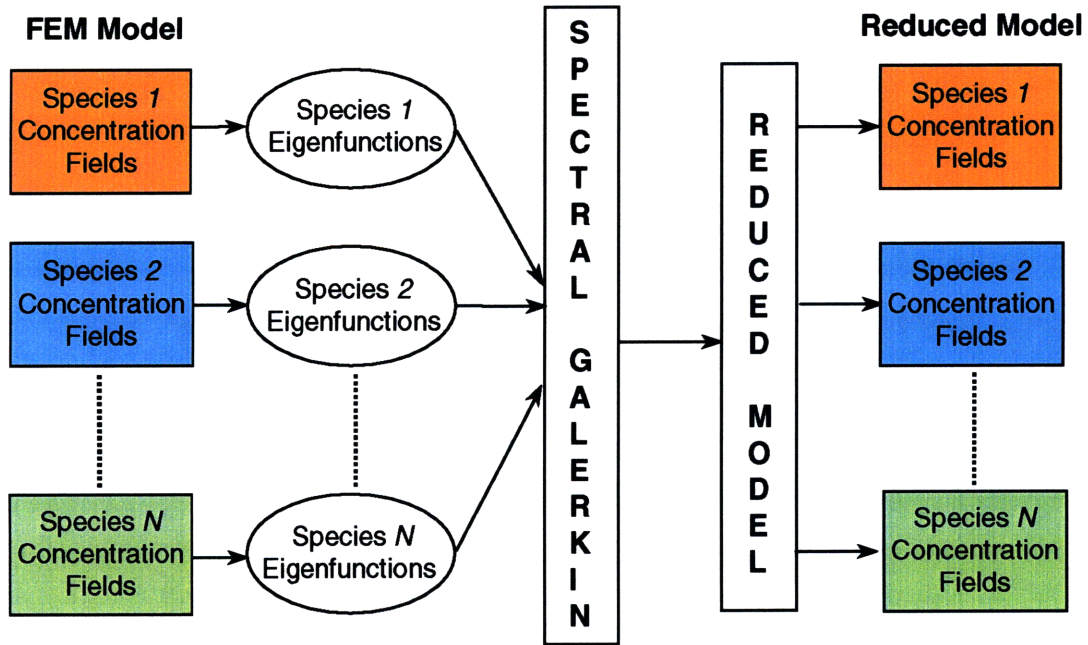


Figure 5.2 Model reduction strategy for chemical vapor deposition systems.

In order to generate eigenfunctions in a manner similar to the RTP case, a particular set of steady state species concentration fields (obtained for a particular inlet concentration of the dominant species) is subtracted from all the other species concentration fields. Subsequently, the deviation fields obtained for the  $k$ th species are stored in matrix  $\mathbf{X}_k$  as follows

$$\mathbf{X}_k = \{\tilde{w}_k^1, \tilde{w}_k^2, \dots, \tilde{w}_k^n\} \quad (12)$$

$$\tilde{w}_k^n = w_k^n - w_{k,ss} \quad (13)$$

Here  $w_k^n$  is the concentration field for species  $k$  obtained for the  $n$ th change in the inlet concentration of the dominant species,  $w_{k,ss}$  is the vector that contains the steady state concentration field for species  $k$  that is subtracted from all the other fields of species  $k$

generated at various inlet concentrations.  $\tilde{w}_k^n$  is the deviation concentration field for species  $k$ . A correlation matrix is then constructed from the matrix  $\mathbf{X}_k$  as follows,

$$\mathbf{C}_{mn} = \langle \tilde{w}_k^m, \tilde{w}_k^n \rangle \quad (14)$$

where  $\langle \dots \rangle$  is the inner product in the  $\ell_2$  norm. The eigenfunctions,  $\mathbf{u}_i$ , are obtained by projecting the left singular vectors of the temporal correlation matrix onto the snapshots,

$$\mathbf{C} = \mathbf{V}\Sigma\mathbf{W} \quad (15)$$

$$\mathbf{u}_i = \left( \frac{1}{\sqrt{\Sigma_{ii}}} \right) \sum_{j=1}^{N_{snap}} \tilde{w}_k^j \mathbf{v}_j \quad (16)$$

$\mathbf{V}$  is a matrix whose columns are the left singular vectors of  $\mathbf{C}$ ,  $\Sigma$  is a diagonal matrix with the singular values of  $\mathbf{C}$  on the diagonal, and  $N_{snap}$  is the number of snapshots.

In order to express the mass transport equation (Equation 4) in a form amenable to model reduction, the highly nonlinear terms must be identified and segregated from the weakly nonlinear ones. This is similar to the RTP case wherein the highly nonlinear radiation terms were identified and separated from the conduction and convection terms. In the case of species transport, the nonlinearities arising from the chemical reaction terms are more important compared to the nonlinearities due to diffusion and convection. After segregating the terms along these lines, the steady state finite element formulation of the species transport equation (Equation 9) can be rewritten in the following matrix notation.

$$\mathbf{D}_k \tilde{w}_k + (\Omega_k - \Omega_{k,ss}) = 0 \quad (17)$$

where  $\mathbf{D}_k$  is the matrix that contains all the diffusion and convection contribution for gas phase species  $k$  in Equation 9.  $\Omega_k$  is the net reactive flux contribution to the species transport



equation for species  $k$  due to gas phase and surface reactions.  $\tilde{w}_k$  is the deviation concentration field for the  $k$ th gas phase species that has to be computed by the reduced model. Since the model reduction procedure is implemented in deviation variables,  $\tilde{w}_k$  represents the variation of the concentration of species  $k$  around the steady state field of the same species about which the eigenfunctions were extracted and the reduced model is formulated. Hence, the actual concentration field for species  $k$  can be regenerated from the deviation field by using Equation 13. Maintaining the same notation,  $\Omega_{k,ss}$  is the net reactive flux contribution for the  $k$ th species, evaluated at the steady state conditions around which the reduced model is formulated. The net reactive flux contribution to the reduced model is computed in a similar manner as expressed in the FEM formulation. This comprises of gas phase and surface reactions and is obtained by combining the reactive flux contributions from Equations 9 and 10. Therefore, the accuracy with which the reduced model is able to generate the deviation concentration fields,  $\tilde{w}_k$ , would largely depend upon the accuracy of the reactive flux contribution to Equation 17. In all the reduced model work discussed in this chapter, this reactive contribution was evaluated at the exact steady state conditions as is done in the FEM model.

By computing the inner products with the empirical eigenfunctions for each species, Equation 17 can be expressed as

$$[\mathbf{U}_k^T \mathbf{D}_k \mathbf{U}_k] \mathbf{a}_k + \mathbf{U}_k^T (\Omega_k - \Omega_{k,ss}) = 0 \quad (18)$$

where  $\mathbf{U}_k$  is the matrix of empirical eigenfunctions for species  $k$  and  $\mathbf{a}_k$  is the vector of reduced order coefficients that have to be computed by solving Equation 18. The terms expressed within square parentheses are precomputed and stored, where as the second term on the left hand side of Equation 18 has to be computed at different steady state conditions when the reduced model is simulated. The actual species concentration fields are regenerated by projecting the reduced order coefficients,  $\mathbf{a}_k$ , onto the eigenfunctions as shown below

$$w_k = w_{k,ss} + \sum_{i=1}^{N_{red}} a_{ik} \mathbf{u}_{ik} \quad (19)$$

where  $N_{red}$  is the dimension of the reduced order model. Because of the mass balance constraint (Equation 11), only  $N - 1$  gas phase species concentrations are obtained from the reduced model in this manner. The carrier gas concentration is evaluated by Equation 11.

## 5.3 RESULTS AND DISCUSSION

### 5.3.1 DECOMPOSITION OF TRIMETHYLGALLIUM IN THE PRESENCE OF HCl

Decomposition of trimethylgallium in the presence of HCl is an important process step in the manufacture of GaAs based semiconductor devices. The chemical system describing the decomposition of trimethylgallium (TMG) in the presence of HCl was used as a test bed for the model reduction strategy. The chemical mechanism for the decomposition is shown in Table 5.1.

No.	Reaction	Pre-exp.	T exp.	E <sub>a</sub>	Rate Law
1	$\text{Ga}(\text{CH}_3)_3 = \text{Ga}(\text{CH}_3)_2 + \text{CH}_3$	$.30 \times 10^{16}$	0	59	$[\text{Ga}(\text{CH}_3)_3]$
2	$\text{Ga}(\text{CH}_3)_2 = \text{GaCH}_3 + \text{CH}_3$	$.30 \times 10^{16}$	0	35	$[\text{Ga}(\text{CH}_3)_2]$
3	$\text{Ga}(\text{CH}_3)_3 + \text{HCl} = \text{CH}_4 + \text{GaCl}(\text{CH}_3)_2$	$.70 \times 10^{11}$	0	10	$[\text{Ga}(\text{CH}_3)_3][\text{HCl}]$
4	$\text{GaCl}(\text{CH}_3)_2 = \text{CH}_3 + \text{GaClCH}_3$	$.35 \times 10^{16}$	0	67	$[\text{GaCl}(\text{CH}_3)_2]$
5	$\text{GaCl}(\text{CH}_3)_2 = \text{CH}_3 + \text{GaCl}$	$.30 \times 10^{16}$	0	35	$[\text{GaCl}(\text{CH}_3)_2]$
6	$\text{GaCH}_3 + \text{s} = \text{sGaCH}_3$	0.1	0	0	$[\text{GaCH}_3][\text{s}]$
7	$\text{sGaCH}_3 = \text{GaCH}_3 + \text{s}$	$.10 \times 10^{13}$	0	50	$[\text{sGaCH}_3]$

Table 5.1 Chemical mechanism describing the decomposition of trimethylgallium (TMG) in the presence of HCl.

The mechanism comprises of 5 gas phase and 2 surface reactions involving 12 chemical species. Trimethylgallium was chosen as the dominant species in the system and the initial inlet concentration of TMG was varied from a mole fraction of  $1.5 \times 10^{-4}$  to  $2.5 \times 10^{-4}$  at intervals of  $1.0 \times 10^{-6}$ . The steady state concentration fields were computed and stored at each interval and empirical eigenfunctions were calculated by the POD method as explained in the last section. Following this, a reduced model using two eigenfunctions was formulated for the TMG decomposition system. The reduced model was then used to simulate the system for an initial inlet concentration of  $4.0 \times 10^{-4}$ . This inlet concentration is outside the range within which the eigenfunction set was extracted. Hence, simulation results from the reduced model for this inlet concentration, tests the capability for extrapolation using the reduced model. The concentration fields for different species as obtained from the reduced model are compared against those obtained from the FEM model in Figures 5.3, 5.4, and 5.5.

Initial  $\text{Ga}(\text{CH}_3)_3$  Concentration =  $4.00\text{E-}04$

FEM Model

Reduced Model

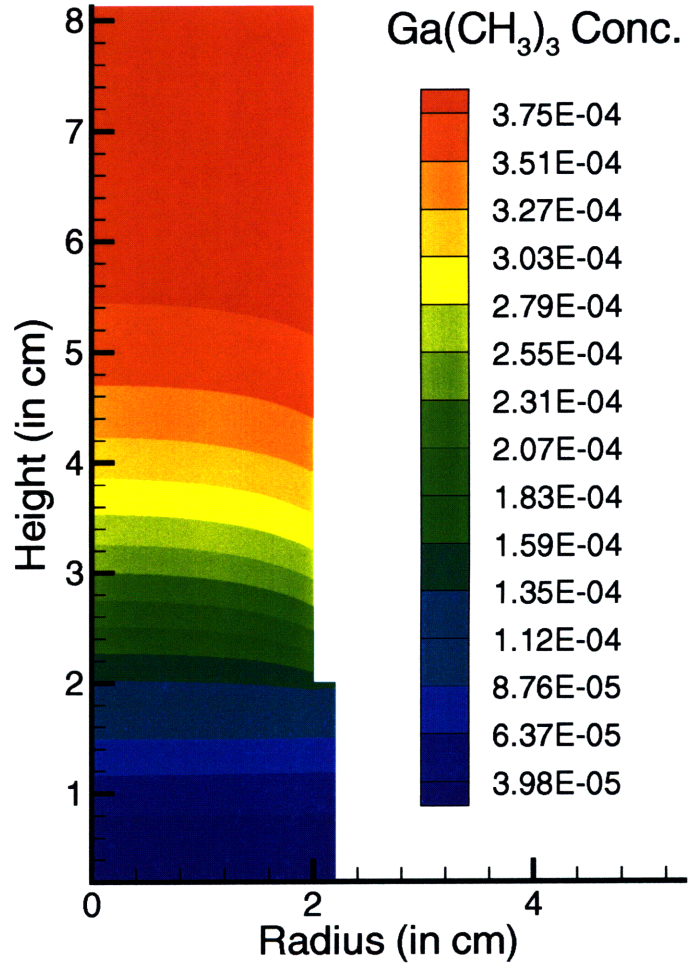
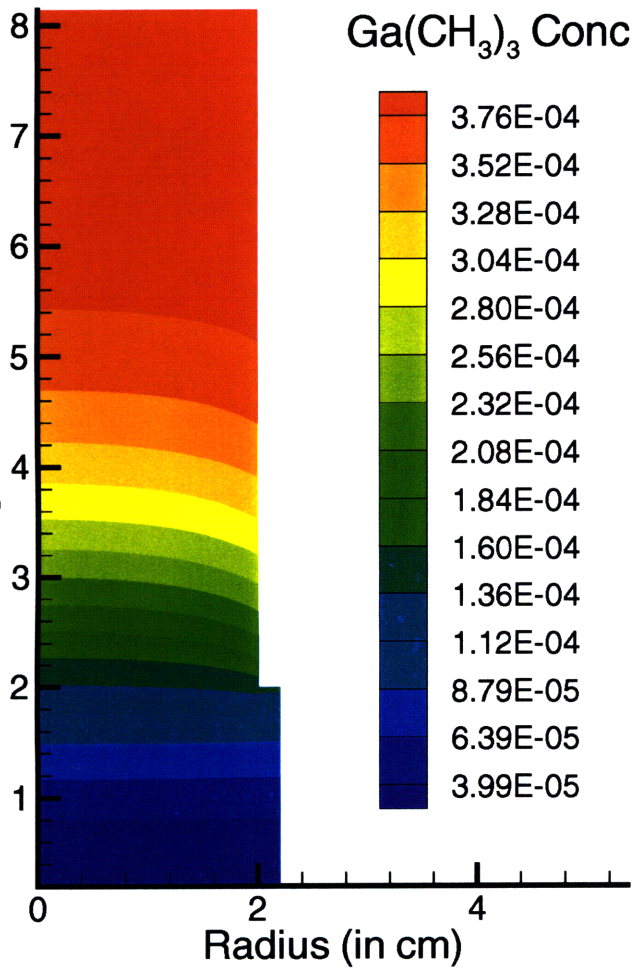
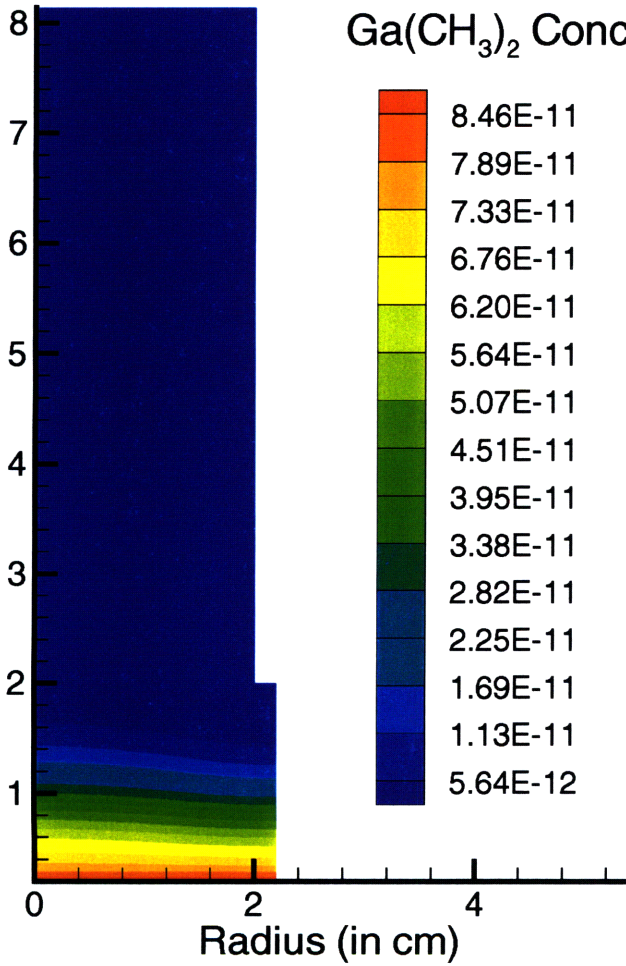


Figure 5.3 Comparison of concentration fields for trimethylgallium as obtained from the FEM and reduced models.

Initial  $\text{Ga}(\text{CH}_3)_3$  Concentration =  $4.00\text{E-}04$

FEM Model



Reduced Model

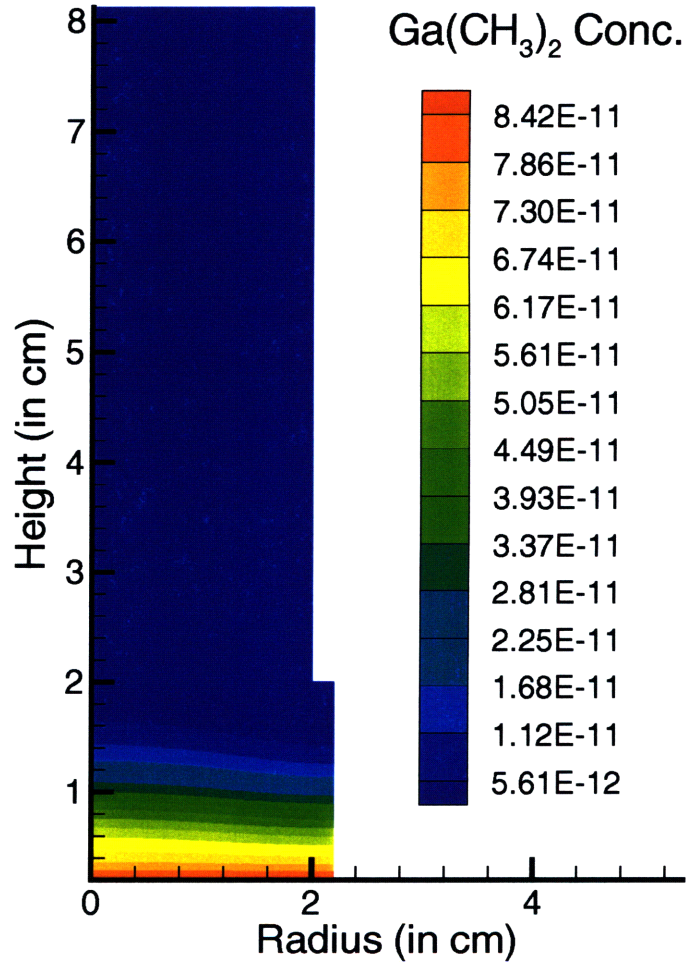


Figure 5.4 Comparison of concentration fields for dimethylgallium as obtained from the FEM and reduced models.

Initial  $\text{Ga}(\text{CH}_3)_3$  Concentration =  $4.00\text{E-}04$

FEM Model

Reduced Model

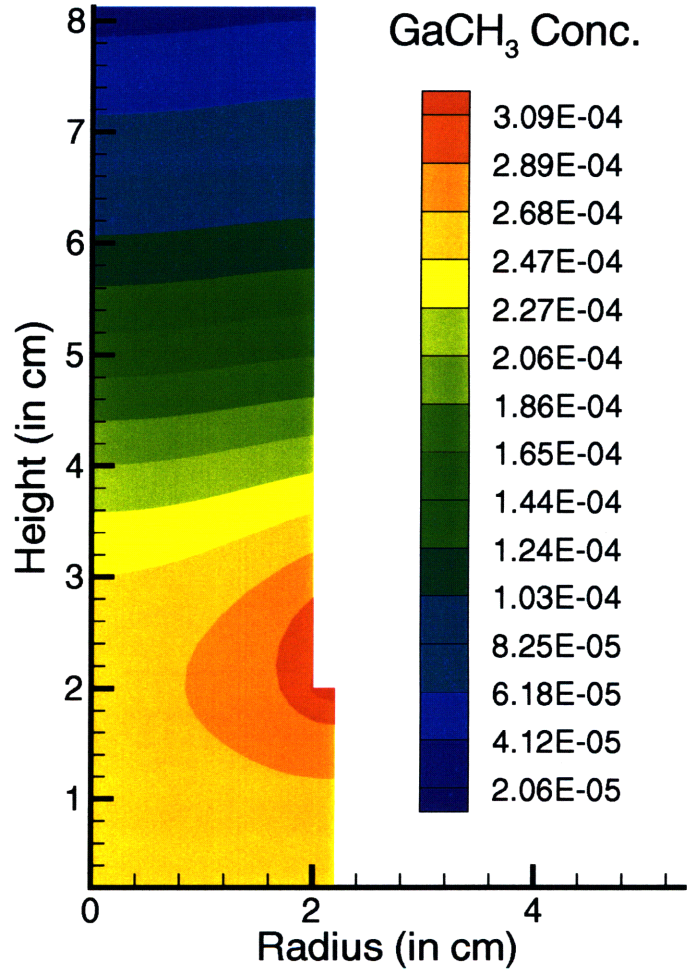
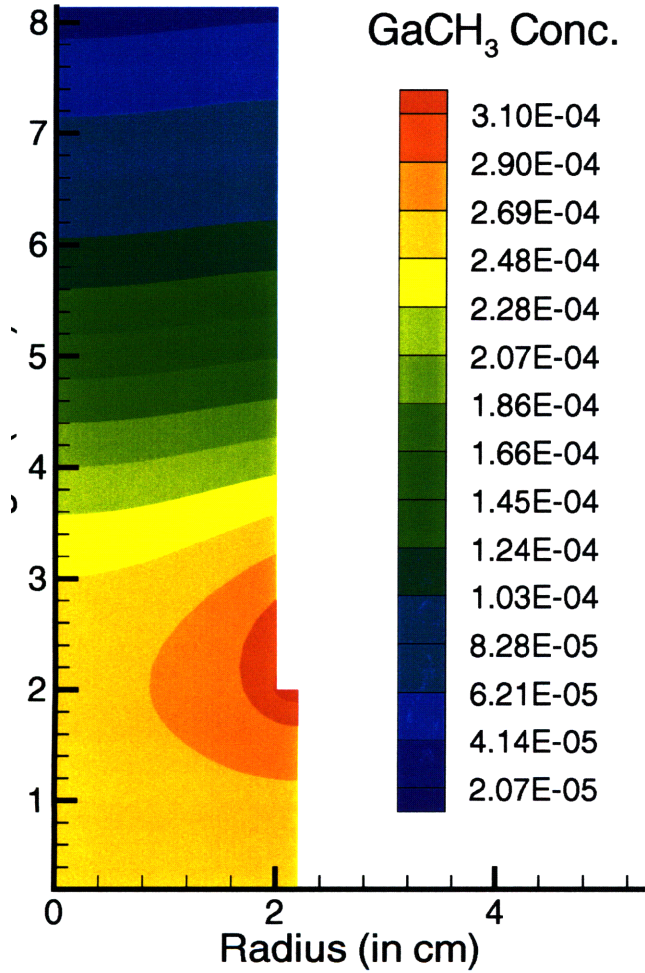


Figure 5.5 Comparison of concentration fields for monomethylgallium as obtained from the FEM and reduced models.

### 5.3.2 BORON DOPING OF SILICON DEPOSITED FROM DICHLOROOSILANE

In situ boron doping of silicon from dichlorosilane is an important process step in the manufacture of silicon based semiconductor devices. Table 5.2[13] lists the various independent reactions involved in the chemical mechanism for the boron doping of silicon from dichlorosilane. There are a total of 41 reactions and 19 chemical species involved in a combination of gas phase and surface reactions.

The inlet mole fraction of  $B_2H_6$ , the dominant species, was varied from  $2.5 \times 10^{-4}$  to  $3.5 \times 10^{-4}$  at intervals of  $1.0 \times 10^{-6}$ . The steady state concentration fields at each interval were calculated and stored. Empirical eigenfunctions, for each of the chemical species, were then extracted from these steady state concentration fields by the POD technique. A reduced model, with 3 eigenfunctions for each of the chemical species, was generated to simulate the steady state species concentration profiles in the reactor. The species concentration fields at an inlet  $B_2H_6$  mole fraction of  $4.0 \times 10^{-5}$  were then simulated using the reduced models. This inlet mole fraction is an order of magnitude lesser than the range within which the eigenfunctions were extracted and, therefore, tests the capability of the reduced model in predicting concentration fields within a large window in process space. Figure 5.6 shows the comparison between the  $SiCl_2H_2$  concentration fields as simulated by the FEM and reduced models. Both the overall and cross-sectional comparisons show good agreement with the FEM model. Similar comparisons for  $B_2H_6$  are shown in Figure 5.7. In this case, the cross-sectional comparison shows a difference of  $1.0 \times 10^{-5}$  in the mole fractions towards the top of the reactor. The reduced model predictions are not fully accurate under these conditions due to the inlet mole fraction being far outside the eigenfunction extraction range. This inaccuracy is more evident in the case of  $B_2H_6$  as the magnitudes of the concentrations are much smaller than those of  $SiCl_2H_2$ . However, the species concentrations at the susceptor surface are predicted accurately as shown by the growth rate plots in Figure 5.8. Since the growth rates at the susceptor surface are the most important

parameters from a processing standpoint, the reduced model can be used to study process behavior over a large concentration range. This is because the most nonlinear term (the reaction flux contribution) is calculated at the FEM conditions. Hence, the reduced model gets the correct input for the highly nonlinear part and the linearization of the gas phase diffusion and convection related terms holds over this range of concentrations.



No.	Reaction	Pre-exp.	T exp.	E <sub>a</sub>
1	$B_2H_6 = 2 BH_3$	$5.90 \times 10^{49}$	-10.8	50.9
2	$2 BH_3 = B_2H_6$	$5.90 \times 10^{44}$	-9.8	10.7
3	$HCl + BH_3 = BClH_2 + H_2$	$1.10 \times 10^{12}$	.0	8.2
4	$BClH_2 + H_2 = HCl + BH_3$	$2.00 \times 10^{12}$	.0	29.1
5	$SiCl_3H + BClH_2 = SiCl_2H + BH_3$	$4.00 \times 10^{10}$	.0	23.0
6	$SiCl_2H + BH_3 = SiCl_3H + BClH_2$	$2.40 \times 10^{11}$	.0	18.5
7	$SiCl_3H = SiCl_2 + H_2$	$1.40 \times 10^{34}$	-6.3	85.4
8	$SiCl_2 + H_2 = SiCl_3H$	$2.00 \times 10^{29}$	-5.3	50.6
9	$SiCl_3H = HSiCl + HCl$	$8.60 \times 10^{33}$	-5.9	83.0
10	$HSiCl + HCl = SiCl_2H_2$	$1.70 \times 10^{28}$	-4.9	14.6
11	$SiCl_2H_2 + d = HCl + CldSiH$	0.008	.0	.0
12	$HCl + d = HdCl$	.50	.0	.0
13	$HSiCl + d = CldSiH$	.10	.0	.0
14	$SiCl_2 + d = CldSiCl$	.10	.0	.0
15	$HdH = H_2 + d$	$2.00 \times 10^{15}$	.0	57.0
16	$HdCl = HCl + d$	$8.50 \times 10^{14}$	.0	65.0
17	$2 HdCl = HdH + Cl dCl$	$1.80 \times 10^{24}$	.0	32.1
18	$HdH + Cl dCl = 2 HdCl$	$1.00 \times 10^{25}$	.0	30.0
19	$Cl dSiCl = SiCl_2 + d$	$1.10 \times 10^{15}$	.0	67.0
20	$Cl dCl + Si = Cl dSiCl$	$1.00 \times 10^{19}$	.0	65.0
21	$2 Cl dSiH = 2 HdCl + 2 Si$	$1.00 \times 10^{25}$	.0	30.0
22	$Cl dSiH + Cl dSiCl = HdCl + Cl dCl + 2 Si$	$1.00 \times 10^{25}$	.0	32.0
23	$2 Cl dSiCl = 2 Cl dCl + 2 Si$	$1.00 \times 10^{25}$	.0	34.0
24	$BH_3 + d = HdBH_2$	0.1	.0	.0
25	$BH_3 + Cl dCl = Cl dClBH_2$	0.01	.0	.0
26	$BH_3 + HdCl = HdClBH_2$	0.01	.0	.0
27	$Cl dClBH_2 = BClH_2 + HdCl$	$1.00 \times 10^{13}$	.0	17.0
28	$HdClBH_2 = BClH_2 + HdH$	$1.00 \times 10^{13}$	.0	17.0
29	$Cl dClBH_2 = HCl + Cl dBH_2$	$1.00 \times 10^{13}$	.0	25.0
30	$HdClBH_2 = HCl + HdBH_2$	$1.00 \times 10^{13}$	.0	25.0
31	$Cl dBH_2 = HCl + dBH$	$1.00 \times 10^{12}$	.0	40.0
32	$Cl dBH_2 = BClH_2 + d$	$1.00 \times 10^{12}$	.0	50.0
33	$HdBH_2 = H_2 + dBH$	$1.00 \times 10^{12}$	.0	35.0
34	$HdBH_2 = BH_3 + d$	$1.00 \times 10^{12}$	.0	60.0
35	$HdCl + dBH = Cl dBH_2 + d$	$1.00 \times 10^{25}$	.0	30.0
36	$HdH + dBH = HdBH_2 + d$	$1.00 \times 10^{25}$	.0	30.0
37	$2 dBH = HdH + dB_2$	$1.00 \times 10^{25}$	.0	30.0
38	$dB_2 = d + 2 B$	$1.00 \times 10^{13}$	.0	30.0
39*	$2 Cl dSiH + (dB_2) = 2 HdCl + d + 2 B + 2 Si$	$1.00 \times 10^{25}$	.0	30.0
40*	$Cl dSiH + Cl dSiCl + (dB_2) = HdCl + d + Cl dCl + 2 B + 2 Si$	$1.00 \times 10^{25}$	.0	32.0
41*	$2 Cl dSiCl + (dB_2) = d + 2 Cl dCl + 2 B + 2 Si$	$1.00 \times 10^{25}$	.0	34.0

\* trapping reactions, the rates of reactions 39-41 were calculated by multiplying the rates of reactions 21-23 by the surface coverage of  $dB_2$

Table 5.2 Chemical mechanism for the in situ boron doping of silicon deposited from dichlorosilane. [13]

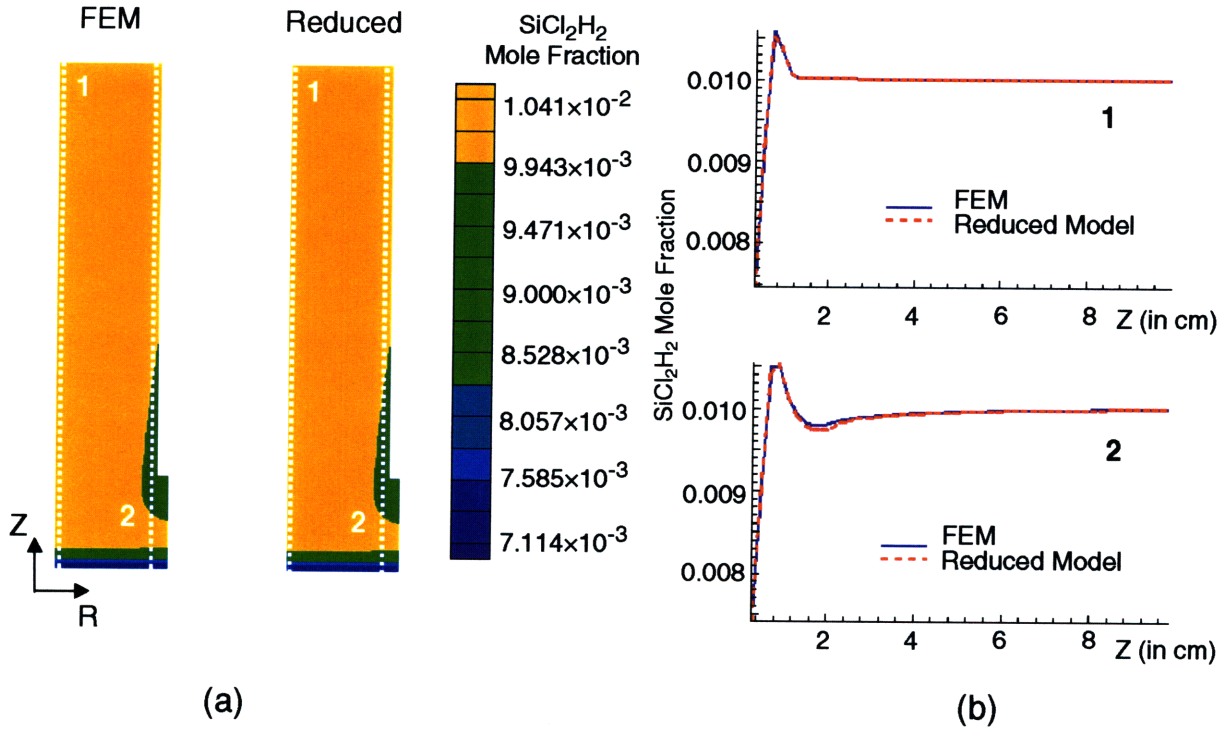


Figure 5.6 (a) Overall comparison of  $\text{SiCl}_2\text{H}_2$  concentration fields. (b) Comparison of  $\text{SiCl}_2\text{H}_2$  concentrations along axial cross-sections 1 and 2.

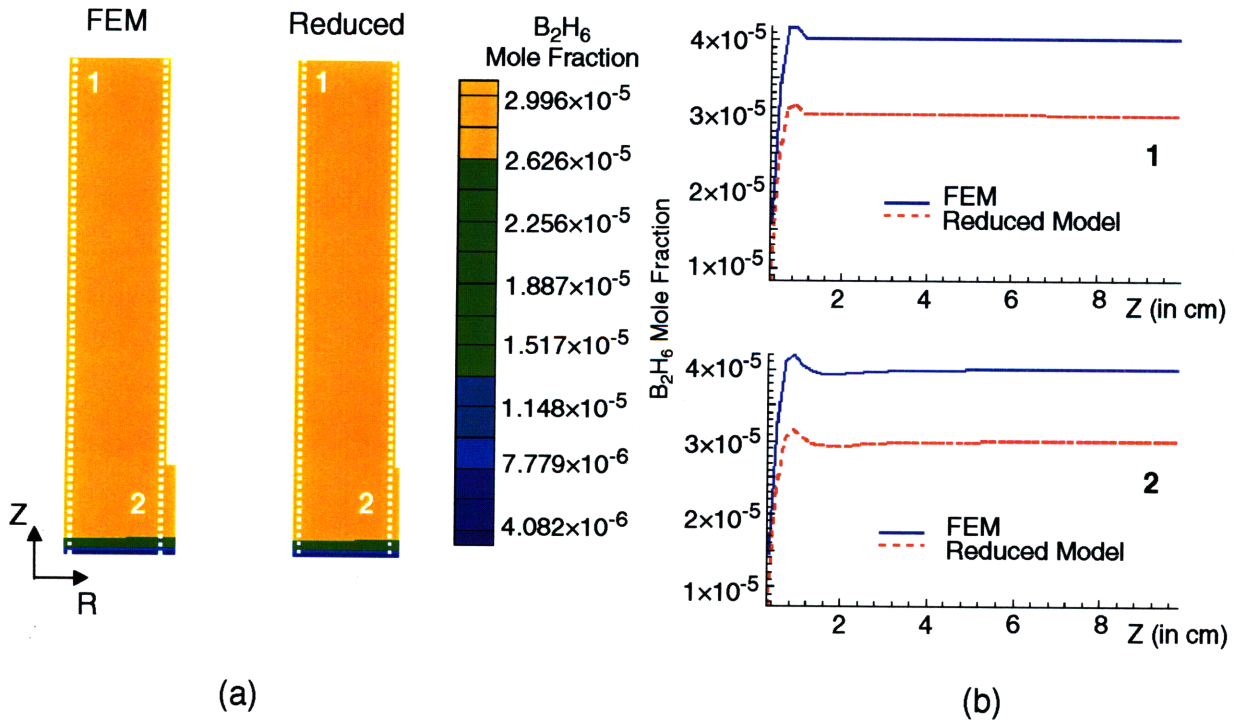


Figure 5.7 (a) Overall comparison of  $\text{B}_2\text{H}_6$  concentration fields. (b) Comparison of  $\text{B}_2\text{H}_6$  concentrations along axial cross-sections 1 and 2.

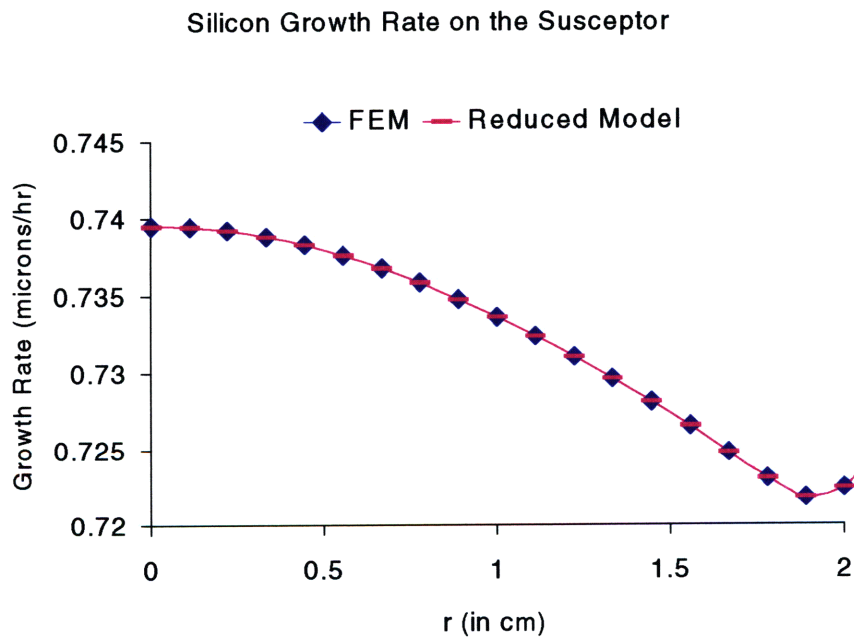
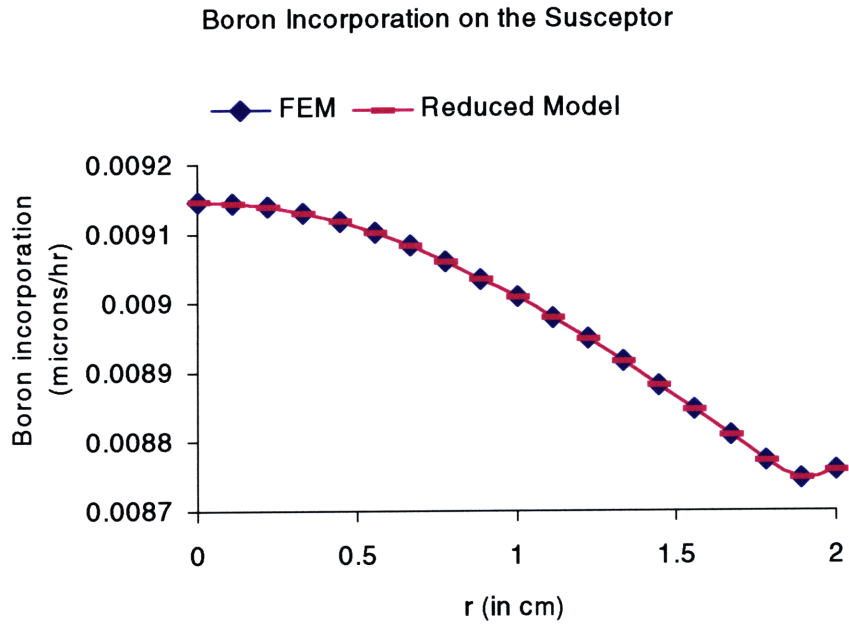


Figure 5.8 Comparison of boron incorporation rate and silicon growth rate on the susceptor surface as simulated by the reduced and FEM models.

## 5.4 CONCLUSION

In this chapter, a model reduction technique that is applicable for generating reduced models, for chemical vapor deposition systems, has been described. The technique uses steady state chemical species concentration fields for extracting empirical eigenfunctions using the POD technique. The mass conservation equation used to describe chemical systems was expanded in spectral Galerkin method using the empirical eigenfunctions as basis sets. In this expansion similar issues regarding multiple eigenfunction sets that arose during the generation of the fluid-thermal RTP reduced model were faced and addressed in the same manner as for the RTP case. Therefore, the mass conservation equations for each species was solved using the eigenfunction set obtained for that particular species as a basis set. The model reduction procedure demonstrated here linearizes the convection and diffusion contribution to the mass conservation equation. On the other hand, the main nonlinearity arising from the reaction flux term is handled in a manner similar to the FEM model. This ensures good agreement of the reduced models with their FEM counterparts as long as the convection and diffusion contribution terms do not change significantly while extending the range of execution for the reduced models. The technique, was then demonstrated for two cases - 1.) Decomposition of TMG, and 2.) In situ boron doping of silicon. Reduced models with 3 eigenfunctions or less were used to accurately regenerate the species concentration profiles in the entire reactor. In the case for the in situ boron doping of silicon, the reduced model gave accurate results even when the inlet concentration of  $B_2H_6$  was reduced by an order of magnitude. The growth rate of silicon and the boron incorporation rate were predicted accurately by the reduced model. Therefore, the reduced model is an accurate modeling tool for understanding process behavior within a large, but limited, range of process conditions. This technique provides a systematic framework for generating such reduced models from FEM simulations. The technique is generic enough, so that it can be used for any dilute chemical reaction system that is governed by the same mass conservation equation. Also, just as the FEM model can be used to study process

behavior for different types of process reactors, the reduced model can also be extended to different reactor geometries under various process conditions. This would entail the extraction of a new eigenfunction set every time the geometry is changed. But once the reduced model has been generated one could carry out repeated simulations using the reduced model to understand process behavior and optimize the process.

## REFERENCES

- [1] C. R. Kleijn, "Chemical Vapor Deposition Processes," in *Computational Modeling in Semiconductor Processing*, M. Meyyappan, Ed. Norwood, MA: Artech House, 1994, pp. 97 - 229.
- [2] K. F. Jensen, D. I. Fotiadis, and T. J. Mountziaris, "Detailed models of the MOVPE process," *J. Cryst. Growth*, vol. 107, pp. 1 - 11, 1991.
- [3] K. F. Jensen, E. O. Einset, and D. I. Fotiadis, "Flow phenomena in chemical vapor deposition of thin films," *Ann. Rev. Fluid Mech.*, vol. 23, pp. 197 - 232, 1991.
- [4] S. Patnaik, R. A. Brown, and C. A. Wang, "Hydrodynamic dispersion in rotating-disk OMVPE reactors: Numerical simulation and experimental measurements," *J. Cryst. Growth*, vol. 96, pp. 153 - 174, 1989.
- [5] D. I. Fotiadis, S. Kieda, and K. F. Jensen, "Transport phenomena in vertical reactors for metalorganic vapor phase epitaxy: I. effects of heat transfer characteristics, reactor geometry, and operating conditions," *J. Cryst. Growth*, vol. 102, pp. 441 - 470, 1990.
- [6] K. F. Jensen and D. B. Graves, "Modeling and analysis of low pressure CVD reactors," *J. Electrochem. Soc.*, vol. 130, pp. 1950 - 1957, 1983.
- [7] C. R. Kleijn, "A mathematical model of the hydrodynamics and gas-phase reactions in silicon LPCVD in a single-wafer reactor," *J. Electrochem. Soc.*, vol. 138, pp. 2190 - 2200, 1991.
- [8] T. P. Merchant, "Modelling of Rapid Thermal Processes," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [9] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*. New York: John Wiley and Sons, 1960.
- [10] O. C. Zienkiewicz, *The Finite Element Method*. New York: McGraw Hill, 1977.
- [11] L. Sirovich, "Turbulence and the Dynamics of Coherent Structures: I, II and III," *Quarterly of Applied Mathematics*, vol. XLV, pp. 561, 1987.

- [12] W. S. Wyckoff, "Numerical Solution of Differential Equations through Empirical Eigenfunctions," in *Chemical Engineering*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
- [13] I. Lengyel and K. F. Jensen, "Progress report on chemical mechanism development of in situ boron doping of silicon deposited from dichlorosilane," Massachusetts Institute of Technology, Cambridge, MA 1997.

# Chapter 6

## Conclusions and Recommendations

### 6.1 CONCLUSIONS

The increasingly competitive environment characteristic of the semiconductor industry requires that emerging technology be brought to the manufacturing stage as quickly as possible. Currently process engineers have to carry out a large number of cut-and-try iterations to generate and optimize processes, before a new equipment or technology can make the transition from the research lab to the fabrication line. Computational modeling has helped in saving time and effort required in equipment design and optimization. However, most of the detailed computational models are far too complex in their execution procedures and computer hardware requirements to be of direct use to process engineers. Simulations on many of the detailed models also take hours to days to yield accurate results. Hence, there exists a demand for the development of compact and accurate models that can be used by process engineers to develop process recipes and optimize them. The work in this thesis has demonstrated a reduced order modeling technique that can be used to systematically generate nonlinear physically based mathematical models from complex finite element models. The strategy, though developed for rapid thermal and chemical vapor deposition processes, is generic and can be extended to other



systems governed by similar physical rate governing equations.

The reduced order modeling technique was introduced by implementing it on a low-pressure RTP case. The reduced order modeling started with the extraction of empirical eigenfunctions from transient simulations of the FEM model describing the RTP system. The method of using lamp power perturbations to generate a set of transient temperature fields around a given steady operating condition was introduced. The steady state temperature field obtained at the steady operating condition was subtracted from the transient temperature fields to generate a set of deviation temperature fields. Empirical eigenfunctions were extracted from the set of deviation fields using the POD technique. This particular method extracting deviation eigenfunctions was chosen so that the model generated from the eigenfunctions will be exact at the given steady operating condition. Care was taken to obtain eigenfunctions that were orthonormal in the  $\ell_2$  norm since the RTP system was modeled as a two-dimensional axisymmetric reactor.

The empirical eigenfunctions obtained from the POD extraction procedure were then used to formulate spectral Galerkin expansions over the same fluid-thermal conservation equations that were used to generate the detailed finite element model. In this procedure, first the radiative heat transfer terms were identified to be the most nonlinear ones in the fluid-thermal conservation equations for the RTP system. These terms were then separated from the weakly nonlinear conduction and convection terms. In this manner, a systematic procedure for segregating highly nonlinear terms from the weakly nonlinear ones within the FEM framework was developed. After segregating the various terms in a matrix form, inner products were computed with the set of empirical eigenfunctions to generate the reduced model. The  $T^4$  nonlinearity and the black body fraction in the radiative heat exchange term were evaluated explicitly at each time step in the nonlinear reduced model thus generated. An alternate mathematical technique for evaluating  $T^4$  nonlinearity within the lower dimensional eigenfunction space was presented and compared to the explicit evaluation case.

The nonlinear reduced models were first used to simulate the steady operating conditions

around which they were extracted. Under these conditions the reduced models agreed exactly with the FEM model. This is a distinct advantage of generating the reduced model using deviation eigenfunctions. The reduced models were then used to extrapolate to other steady operating conditions to test their range of accuracy. It was found that the reduced models agreed very well with the FEM model when they were stretched beyond the steady operating condition around which they were generated. Therefore, the reduced models can be used for process optimization studies within a large window in process space ( $\pm 100$  °C). Subsequently, models with varying number of eigenfunctions were extracted and a RMS error criterion over the wafer surface temperature was used to determine that in the low pressure RTP case reduced models generated from ten eigenfunctions had the greatest accuracy. No further accuracy could be obtained by using more eigenfunctions.

After testing under steady operating conditions, the reduced models were used to simulate transient temperature trajectories. Under these dynamic conditions it was found that the linearization of gas phase properties introduced errors in temperature trajectories, when the reduced models were used to predict far outside their range of accuracy. However, reduced models obtained under different steady operating conditions were shown to accurately represent the dynamic behavior at the respective conditions. Two schemes – 1.) switching, and 2.) interpolation between different reduced models, were developed to accurately simulate the entire transient trajectory. In this manner, the different portions of the RTP cycle, ramp-up, steady-hold, and cool-down were simulated using a few reduced models. Subsequently, an actual RTP cycle was simulated by a combination of reduced models. Timing runs were performed on the reduced and finite element models to demonstrate the orders of magnitude savings in computational time that can be obtained by using reduced models.

Subsequently, the reduced order modeling technique was extended to modeling of coupled heat transfer and fluid-flow problems. Multiple eigenfunction sets were extracted from transient velocity, temperature, and pressure fields for a RTP system. The equations representing the conservation of mass, momentum and energy were then formulated in a form

amenable to the model reduction technique. The technique for expressing the reduced models across eigenfunction sets that were not orthonormal across each other was developed. In this development, the question regarding which set of eigenfunctions to use for which equation was answered. This mathematical development, showed that as long as the appropriate set of eigenfunctions was used for computing inner products with an equation, the lack of orthonormality across eigenfunction sets did not hinder the reduced model formulation. This represented a significant advancement from the reduced modeling technique involving only single eigenfunction sets. The nonlinear reduced model was then used to accurately model fluid flow induced temperature variations in the RTP reactor under different pressure conditions.

The deposition of multilayer thin film stacks on the wafer surface leads to changes in the radiative properties of the wafer surface, which cause significant variations in the temperature profiles across the wafer surface. This issue has been of prime concern to process engineers while developing process recipes for patterned wafers. A reduced order modeling technique was developed to model these pattern effects. The modeling technique provided further insight into the efficacy of the reduced modeling using POD eigenfunctions. It was shown that the correct set of radiative properties and steady state temperature fields are required to accurately simulate pattern effects around a given steady operating condition. However, an eigenfunction set obtained by simulating the wafer with the properties of bare silicon would suffice for all cases in which the radiative properties of the patterned did not vary significantly from the properties of bare silicon. This is a significant finding, because a single set of eigenfunctions could now be used to generate reduced models for a variety of patterned wafers. The reduced models generated in this fashion were used to model transient RTP temperature trajectories for annealing and silicidation steps. It was found that the reduced models show good agreement with the FEM models under steady operating conditions, but over predict the pattern induced temperature changes during the RTP ramp-up phase. Timing runs from the reduced and FEM model showed order of magnitude savings for the reduced model. Hence, the nonlinear reduced modeling technique can be used as an effective diagnostic tool for detecting the onset of pattern

effects on wafers.

Atmospheric pressure RTP represents a major portion of all rapid thermal processes in the semiconductor industry. The nonlinear reduced modeling technique was extended to the development of a model for a commercial RTP chamber. The reduced models were found to agree very well with the FEM model of the RTP reactor around steady operating conditions. An actual RTP cycle was then simulated using reduced models. It was found that the difference between the transient temperature trajectories as predicted by the reduced models and that predicted by the FEM model was due to the linearization of wafer surface absorptance and gas phase thermal properties. Combinations of multiple reduced models were used to simulate the entire RTP cycle, and the temperature trajectories obtained compared favorably against process data and FEM simulations. Timing runs again demonstrated the savings in computational time obtained by using reduced models.

The model reduction technique was further extended to reduce mass conservation equations for chemical vapor deposition (CVD) systems. The technique in the CVD case uses steady state chemical species concentration fields for extracting empirical eigenfunctions by the POD method. The mass conservation equation used to describe chemical systems was expanded in spectral Galerkin method using the empirical eigenfunctions as basis sets. In this expansion similar issues regarding multiple eigenfunction sets that arose during the generation of the fluid-thermal RTP reduced model were faced and addressed in the same manner as for the RTP case. The model reduction procedure, demonstrated for the CVD case, linearizes the convection and diffusion contribution to the mass conservation equation. On the other hand, the main nonlinearity arising from the reaction flux term is handled in a manner similar to the FEM model. This ensures good agreement of the reduced models with their FEM counterparts as long as the convection and diffusion contribution terms do not change significantly while extending the range of execution for the reduced models. The technique, though demonstrated for two cases – (1.) decomposition of trimethylgallium, and (2.) in situ boron doping of silicon, is generic and can be applied to a variety of chemical systems.

In summary, a strategy for extracting lower dimensional physically based reduced order models from complex finite models was developed. The technique is superior to other strategies, such as lumping of nodes within the FEM framework or assuming certain variables constant, because it does not simplify any of the physical conservation equations and the eigenfunction sets used to expand the equations are likely candidates for the solution fields. Nonlinear terms can be expressed within this framework, either by evaluating them explicitly at each time step or by evaluating them within the reduced model framework. Linearization of some of the physical effects leads to inaccuracies when the range of operation of the reduced models were extended far beyond the range in which the linearization was done. These inaccuracies can be removed by expressing these nonlinear physical effects explicitly. However, this would lead to a complicated model and would defeat the original motivation of developing fast and compact reduced models. Many of the reduced models have computation times that are faster than real time and hence these models show promise of being applicable model based controller design.

## **6.2 RECOMMENDATIONS FOR FUTURE WORK**

There are several areas in which this work could be extended. One of the areas to explore further would be the use of functions besides the POD empirical eigenfunctions described here in performing the spectral Galerkin expansions. Even within the area of empirical eigenfunction extraction by the POD technique, there are several possible methods for perturbing the fields from which the eigenfunctions are extracted. In this thesis lamp power inputs are used to perturb all the fields. But in a RTP process, there are several other inputs such as gas flow rate and composition that are varied during the process. Fields obtained from a combination of perturbations of different inputs could be used to extract eigenfunctions. In this study, it would be worth exploring the relative advantages and disadvantages of such

perturbation techniques. Also, RTP is a dynamic and transient process. Therefore, one could also use fields extracted over the entire RTP trajectory to extract eigenfunctions. In this case, if deviation eigenfunctions as described in this thesis are used, then the question regarding around which steady state the spectral Galerkin needs to be formulated has to be answered.

In this thesis, eigenfunctions extracted from deviation fields are used to expand the macroscopic conservation equations. This is done to remove any steady state offsets in the reduced model. However, one can also study the use of eigenfunctions extracted from absolute fields in spectral Galerkin expansions. This method would eliminate the question around which steady state the Galerkin needs to be formulated, as the eigenfunction expansions in this case will not need a steady state value to regenerate the fields.

Two methods for handling polynomial nonlinearities have been explained in this thesis. But, in the energy conservation equation, there exists several exponential nonlinearities (such as gas phase thermal conductivities and specific heats) that have to be handled differently if one needs to account for them in a nonlinear fashion in the reduced model. Hence, techniques have to be derived to formulate such nonlinearities within the reduced model framework.

The use of multiple eigenfunction sets for expanding the fluid-thermal conservation of equations is one method of formulating a reduced model. Instead of extracting individual eigenfunction sets for the individual classes of fields, one could concatenate the fields and extract eigenfunctions from the entire set. In this expansion, the issue regarding which steady state field to pick for extracting deviation eigenfunctions, needs to be answered. Also, there are scaling issues that have to be addressed while extracting the eigenfunctions, as the values for temperature and velocities are different by orders of magnitude.

In the case of exploring pattern effects with reduced models, the issue regarding what can be done to remove the overprediction of pattern effects by the reduced models needs to be explored. In this case, if the emittance values can be accounted for explicitly in the reduced models while performing the time integration for the transient ramp up, an improvement in model prediction could be achieved. However, one needs to decouple the emittance terms from

the radiation contribution term in the reduced model, which could lead to increasingly nonlinear model that requires significantly higher computation time.

There are several areas to explore in using empirical eigenfunction for model reduction of CVD systems. Some of the issues here are similar to the issues in reduced order modeling of fluid-thermal equations. But there are additional avenues of further research in expressing the chemical reaction flux contribution in terms of empirical eigenfunctions. The method described in this thesis allows a generic formulation for any chemical mechanism, but treats the total reaction flux contribution in a lumped manner. Deconvoluting the reaction flux contribution in terms of species concentration fields and expressing them individually in terms of the respective eigenfunction sets is a challenging problem. Extension of the reduced order modeling technique to a transient mass conservation formulation is another area worth exploring.

# APPENDIX A

## ALTERNATIVE METHOD FOR COMPUTING POLYNOMIAL NONLINEARITIES USING EIGENFUNCTIONS

In Chapter 2 it has been shown how the empirical eigenfunctions obtained from the POD method are used in a pseudospectral Galerkin expansion to generate a reduced model of the form:

$$(\mathbf{u}_m^T \mathbf{M}(\bar{\mathbf{x}}) \sum_{i=1}^N \mathbf{u}_n \frac{da_i}{dt}) = (\mathbf{u}_m^T \mathbf{C}(\bar{\mathbf{x}}) \sum_{n=1}^N a_n(t) \mathbf{u}_n) + (\mathbf{u}_m^T \mathbf{R}) [\hat{\mathbf{x}}(t)]^4 + (\mathbf{u}_m^T [\alpha_i \sum_{l=1}^{N_{lamp}} \tilde{\mathbf{R}}_{il} \tilde{\mathbf{P}}_l + \mathbf{K}]) \quad (1)$$

which can be reformulated in matrix notation as

$$[\mathbf{U}^T \mathbf{M}(\bar{\mathbf{x}}) \mathbf{U}] \frac{d\mathbf{a}}{dt} = [\mathbf{U}^T \mathbf{C}(\bar{\mathbf{x}}) \mathbf{U}] \mathbf{a} + [\mathbf{U}^T \mathbf{R}] \{\hat{\mathbf{x}}(t)\}^4 + [\mathbf{U}^T \mathbf{G}] \tilde{\mathbf{P}} + [\mathbf{U}^T \mathbf{K}] \quad (2)$$

One strategy for expanding the  $\{\hat{\mathbf{x}}(t)\}^4$  in terms of the eigenfunctions would be to compute the inner product explicitly as shown below in Equation 3.



$$\langle T^4, \mathbf{u}_i \rangle = \left\langle \left\{ \sum_{j_1} a_{j_1}(t) \mathbf{u}_{j_1} \right\} * \left\{ \sum_{j_2} a_{j_2}(t) \mathbf{u}_{j_2} \right\} * \left\{ \sum_{j_3} a_{j_3}(t) \mathbf{u}_{j_3} \right\} * \left\{ \sum_{j_4} a_{j_4}(t) \mathbf{u}_{j_4} \right\}, \mathbf{u}_i \right\rangle \quad (3)$$

On accumulating terms, Equation 3 can be rewritten as

$$\langle T^4, \mathbf{u}_i \rangle = \sum_{j_1, j_2, j_3, j_4} a_{j_1} a_{j_2} a_{j_3} a_{j_4} \langle \mathbf{u}_{j_1} * \mathbf{u}_{j_2} * \mathbf{u}_{j_3} * \mathbf{u}_{j_4}, \mathbf{u}_i \rangle \quad (4)$$

where  $\langle \dots \rangle$  denotes the inner product and  $*$  denotes the pointwise multiplication product between two vectors. The inner product shown in Equation 4 is precomputed and stored for various values of the indices. Subsequently, in the transient integration using the reduced model, the corresponding products of the temporal coefficients multiply the precomputed inner products and the summation shown in Equation 4 is evaluated. This method does not require the absolute temperatures to be regenerated at each time step, and hence, the transient reduced model formulation is solved in the lower dimensional eigenfunction space instead of reverting to the higher dimensional temperature space. But the computation time involved in calculating the inner products is large. The operation count scales as  $N^5$  where  $N$  is the number of eigenfunctions. Also a large amount of storage space is required for storing the precomputed inner products, since  $N^5$  have to be stored and read by the reduced model program.

Timing runs, on a HP-735 workstation, were performed on the different reduced models two investigate the savings in the different methods. To replicate the temperature trajectory for a set of lamp power perturbations, the following results were obtained.

Model	Computation time
FEM model	2684 seconds
Reduced model with temperature reconstruction	45 seconds
Reduced with precomputed inner products	27 seconds

Table A.1 Comparison of model execution time between different reduced models and the FEM model for low pressure RTP systems.

The reduced model with the precomputed inner products is faster than the reduced model with the temperature reconstruction as the transient problem is entirely solved in the lower dimensional eigenfunction space. However, the time required to generate the reduced model with the precomputed inner products is significantly higher due to the large number of operation counts. The time required to generate the two types of reduced models with 10 eigenfunctions on a HP-735 workstation are tabulated below.

Model	Model Generation Time
Reduced model with temperature reconstruction	~ 20 minutes
Reduced with precomputed inner products	~ 1 day

Table A.2 Comparison of model generation time for different reduced models.