# TECHNICAL RESEARCH REPORT

## Model Reduction for Optimization of Rapid Thermal Chemical Vapor Deposition Systems

*by A. Theodoropoulou, R.A. Adomaitis, and E. Zafiriou*

**T.R. 96-64r1**




*Sponsored by*
*the National Science Foundation*
*Engineering Research Center Program,*
*the University of Maryland,*
*Harvard University,*
*and Industry*

# Model Reduction for Optimization of Rapid Thermal Chemical Vapor Deposition Systems

Artemis Theodoropoulou, Raymond A. Adomaitis, and Evanghelos Zafiriou

*Department of Chemical Engineering and Institute for Systems Research*

*University of Maryland*

*College Park, MD 20742*

November 11, 1996

Artemis Theodoropoulou, Raymond A. Adomaitis[1], and Evanghelos Zafiriou

*Department of Chemical Engineering and Institute for Systems Research*

*University of Maryland*

*College Park, MD 20742*

## Abstract

A model of a three-zone Rapid Thermal Chemical Vapor Deposition (RTCVD) system is developed to study the effects of spatial wafer temperature patterns on polysilicon deposition uniformity. A sequence of simulated runs is performed, varying the lamp power profiles so that different wafer temperature modes are excited. The dominant spatial wafer thermal modes are extracted via Proper Orthogonal Decomposition and subsequently used as a set of trial functions to represent both the wafer temperature and deposition thickness. A collocation formulation of Galerkin's method is developed to discretize the original modeling equations, giving a low-order model which looses little of the original, high-order model's fidelity. We make use of the excellent predictive capabilities of the reduced model to optimize power inputs to the lamp banks to achieve a desired polysilicon deposition thickness at the end of a run with minimal deposition spatial nonuniformity.

[1]Author to whom correspondence should be addressed; Phone no. (301) 405-2969; E-mail adomaiti@isr.umd.edu.

# 1 Introduction

One of the main obstacles to commercialization of single wafer Rapid Thermal Chemical Vapor Deposition (RTCVD) systems are the problems they have experienced with nonuniform deposition. Nonuniform radiant energy distribution from the heating lamps, reactant depletion, and reactant gas flow instabilities are some of the main factors which can contribute to nonuniform deposition of thin films in these systems [1]. Corrective efforts consisting of designing the RTCVD system furnace to optimize heating lamp bank locations [2], [3], or modeling and then controlling the factors which give rise to deposition nonuniformity [5], [6], have been reported in the literature. Nevertheless, it is anticipated that these problems will only become more severe with larger wafer sizes and the demands of more environmentally conscientious manufacturing techniques.

Model-based control of wafer temperature and polysilicon deposition thickness spatial nonuniformity requires the development of appropriate models. Numerous models have been developed for RTP systems [4], [7], [8], [9] with some featuring detailed, "first-principles" type descriptions of all significant physical and chemical processes. However, advanced RTP system dynamic simulators based on first-principle models are not always suitable or convenient for model-based control and optimization, mostly because of their demand for computational resources. This necessitates model reduction techniques that will give reduced order models that retain predictive capabilities consistent with the detailed simulations from which they are generated. These reduced order models would then be suitable for RTCVD process optimization, control, and model based sensing.

In most systems, dynamical behavior appears to be low-dimensional when viewed from the proper perspective. Thus, one path to creating computationally-reasonable models of distributed parameter systems begins with extracting the few important spatial modes responsible for most of the spatial distribution of the distributed parameter, i.e., the temperature or the deposition thickness. Determining the dynamic, nonlinear coupling between these *coherent structures* [10] using Galerkin's method [12], [13], [14] to project the modeling equations onto these modes gives low-order, predictive models. Thus, one can think of the model reduction as a technique for shifting most of the complexity to the empirically-determined trial functions, leaving ordinary differential equations to describe the time-dependence of the mode amplitudes.

This study presents the numerical methods behind a model reduction methodology for generating low-order models in the context of a specific RTCVD system - the techniques, however, are suitable for other applications.

The efficiency of the method is tested on a representative RTCVD process described in [3], for which a high-order model of the wafer and tool states is first developed. The model reduction method consists of a two-step procedure: the first consists of running the high-order simulator under a sufficiently-diverse set of operating conditions to collect a range of wafer state "snapshots." An optimal set of globally valid trial functions is determined directly from the statistical information of the system by the Karhunen-Loeve expansion (also known as Proper Orthogonal Decomposition, Principal Component Analysis, and other names); by representing the wafer state in terms of a time-dependent linear combination of these trial functions, the original modeling partial differential equations can then be discretized with a collocation technique to obtain the reduced-order model in the second step. The collocation method is well-suited to deal with the nonlinearities of both the model and the wafer edge boundary condition and gives much more intuitive feel to the connection between the reduced-order model and measurements which can be made of the wafer state. The low-order model obtained is used in the optimization problem of attaining a uniform desired temperature (chosen to produce a silicon deposition thickness of 0.5 $\mu$m) at the end of the process cycle using the individual lamp bank powers as the optimization parameters.

## 2 Process Description

The system we consider is based on the three-zone RTCVD reactor located at the North Carolina State University Center for Advanced Electronic Materials Processing [3]. In this paper, we will consider deposition a $0.5\mu m$ film of polycrystalline silicon on a 6-inch wafer over the course of a run lasting approximately one minute. A diagram of the RTP system is shown in Fig. 1. The furnace is designed so that the central lamp bank A heats the total area of the wafer, and bank B mainly heats the wafer edge to compensate for the extensive heat loss that occurs from the edge. Lamp bank C heats the wafer in a nearly uniform manner (see Fig. 2) and so can be used as a coarse adjustment to the average wafer temperature. The wafer is rotated while heated for azimuthal temperature uniformity. The primary bell jar component of the chamber is made of quartz glass. A stainless steel ring directs air from a blower vertically upward along the cylinder from the quartz base to reduce chamber heating.

# 3 Model Formulation

The model presented here assumes temperature uniformity in the azimuthal direction because of the rotation of the wafer and the symmetric design of the reactor. Temperature variations in the axial direction are neglected because the wafer thickness is very small [2]. The radiative energy exchange between the wafer and the chamber walls has been modeled based on the assumption that this energy transfer occurs in an enclosure composed of diffuse-gray surfaces [15].

An ordinary differential equation is used to account for the lumped quartz chamber wall thermal dynamics, and a steady state energy balance on the cooling gas is included to evaluate the coolant temperature. Because this is a low pressure process, heat transfer from the wafer to the reactant gas is assumed negligible, so the modeling from this point on focuses on the radiant heat transfer processes.

## 3.1 Wafer Temperature

An enthalpy balance on a wedge-shaped, three-dimensional wafer differential element gives the following wafer temperature partial differential equation (PDE):

$$\rho_w \frac{\partial}{\partial t'} \left[ C_{p_w} T' \right] = \frac{1}{r'} \frac{\partial}{\partial r'} \kappa r' \frac{\partial T'}{\partial r'} + \frac{q_{rad}}{\delta z}. \tag{1}$$

We note that the thin wafer assumption ($\delta z << R_w$) means the radiative heat transfer boundary conditions at the wafer top and bottom $q_{rad}$ can be incorporated directly into this PDE. The wafer thermal conductivity and specific heat ($\kappa$ and $C_{p_w}$) are functions of wafer temperature $T'$.

Defining the dimensionless wafer temperature $T$, radial position $r$, and time $t$ (along with dimensionless chamber $T_c$ and coolant gas $T_g$ temperatures)

$$T = \frac{T'}{T_{amb}} \qquad r = \frac{r'}{R_w} \qquad t = \frac{t'}{1 \text{ sec}} \qquad T_c = \frac{T_c'}{T_{amb}} \qquad T_g = \frac{T_g'}{T_{amb}}$$

where $T_{amb}$ is the ambient temperature, (1) becomes:

$$\rho_w T_{amb} \frac{\partial}{\partial t}\left[C_{p_w} T\right] = \frac{T_{amb}}{R_w^2} \frac{1}{r} \frac{\partial}{\partial r} \kappa r \frac{\partial T}{\partial r} + \frac{q_{rad}}{\delta z} \tag{2}$$

with dimensionless boundary conditions

$$\left.\frac{\partial T}{\partial r}\right|_{r=0} = 0, \tag{3}$$

$$\left.\frac{\kappa T_{amb}}{R_w} \frac{\partial T}{\partial r}\right|_{r=1} = -\sigma \epsilon_w T_{amb}^4 (T^4 - T_c^4) + q_{edge} u_B. \tag{4}$$

The radiative energy transfer term $q_{rad}$ consists of two parts: the radiant energy absorbed from the heating lamps and the radiant energy exchanged between the wafer and the reactor walls. For the heating lamp energy flux at the wafer surface we use data presented in [3], which were calculated by a raytrace algorithm (see, e.g., [5]). The algorithm calculates the radiant energy flux distributions directly from the lamps to the wafer as well as the contribution of reflected rays. The radiant energy flux distribution for each lamp bank as a function of wafer radial position is shown in Fig 2. The energy flux at the edge of the wafer $q_{edge}$ was calculated by the authors directly as 49 $Joule/cm^2$; this value was calculated using the total power consumed by lamp bank B, the distance from the lamp bank to the wafer edge, and taking into account the lamp bank reflector.

The radiation exchange between the wafer and the walls is computed using the *net-radiation method* [15]. We consider two grey body enclosures: the one that is formed by the top surface of the wafer and the hemispherical part of the quartz chamber, and the second formed by the bottom surface of the wafer, the cylindrical part of the quartz chamber, the cylindrical steel loading chamber, and the bottom of the loading chamber (see Fig. 3). We assume that the radiation exchange between the top and the bottom chamber sections is negligible. The loading chamber walls and the bottom surface are assumed to be at ambient temperature since their distance from the wafer and the lamps is large relative to the cooling fan. A complicated exchange of radiative energy occurs inside each enclosure as radiation emitted from a surface travels to the other surfaces, is partially reflected, and is then re-reflected many times within the enclosures with partial absorption at each contact with a surface. It would have been very complicated to follow the beams of radiation as they undergo this process; the advantage of the net-radiation method is that it does not require these types of calculations [15].

The total radiative energy term is

$$q_{rad} = \boldsymbol{Q}_{lamps,w} \cdot \boldsymbol{u} + q_{dw,b} + q_{dw,t} \tag{5}$$

where the input $\boldsymbol{u} = [u_A,\ u_B,\ u_C]$ is the percentage of the lamp power that is used. The net radiative energy transfered to the wafer top and bottom surfaces from sources other than the lamps are $q_{dw,t}$ and $q_{dw,b}$, respectively. From the top enclosure net radiation energy balance we obtain

$$q_{dw,t} = \frac{q_n}{q_d} \tag{6}$$

where

$$\begin{aligned} q_n &= -\epsilon_w\,\sigma\,T_{amb}^4\left(T^4 F_{h-h} - T^4 F_{h-h}\,\epsilon_c - T^4 - T^4 F_{dw-h}\,F_{h-h} + T^4 F_{dw-h}\,F_{h-h}\,\epsilon_c + \right. \\ &\qquad \left. T^4 F_{dw-h} - T^4 F_{dw-h}\,\epsilon_c + F_{dw-h}\,\epsilon_c\,{T_c}^4\right) \\ q_d &= -F_{h-h} + F_{h-h}\,\epsilon_c + 1 + F_{dw-h}\,F_{h-h} - F_{w-h}\,F_{h-h}\,\epsilon_w - F_{dw-h}\,F_{h-h}\,\epsilon_c + F_{dw-h}\,F_{h-h}\,\epsilon_c\,\epsilon_w \\ &\quad -F_{dw-h} + F_{dw-h}\,\epsilon_w + F_{dw-h}\,\epsilon_c - F_{dw-h}\,\epsilon_c\,\epsilon_w \end{aligned}$$

The bottom enclosure net radiation energy balance gives a similar but much more complicated formula for $q_{dw,b}$.

The coefficients $F$ are the geometric configuration factors. They are shown schematically in Fig. 4. $F_{dw-c,q}$, $F_{dw-c,s}$, and $F_{dw-h}$ are viewfactors between an infinitesimal wafer area and a finite chamber area and have been calculated by the definition of the configuration factors [16].

As a result of these calculations, we find for the geometry shown in Fig. 3,

$$F_{dw-c,q} = \frac{1}{2}\left(\frac{{R_w}^2 r^2 + {L_c}^2 - {R_w}^2}{\sqrt{{L_c}^4 + 2{R_w}^2{L_c}^2 + {R_w}^4 + 2{R_w}^2 r^2 Lc^2 + {R_w}^4 r^4 - 2{R_w}^4 r^2}} - 1\right) \tag{7}$$

$$\begin{aligned} F_{dw-c,s} = \ & \left(0.5{R_w}^2 r^2 + 0.5{L_c}^2 + L_c\,L_s - 0.5{R_w}^2 + 0.5{L_s}^2\right) / \left(2{R_w}^2 r^2 {L_s}^2 + {L_c}^4 + 6{L_c}^2{L_s}^2\right. \\ & \left. +4{R_w}^2 r^2 L_c\,L_s + 4{L_c}^3 L_s + 2{R_w}^2{L_s}^2 + 4{R_w}^2 L_c\,L_s + {R_w}^4 + {R_w}^4 r^4 + 2{R_w}^2 r^2 {L_c}^2\right) \end{aligned}$$

$$-2{R_w}^4 r^2 + 2{R_w}^2 {L_c}^2 + {L_s}^4 + 4 L_c {L_s}^3\Big)^{\frac{1}{2}} - 0.5\left({R_w}^2 r^2 + {L_c}^2 - {R_w}^2\right) / \left({L_c}^4\right.$$
$$\left. + 2{R_w}^2 {L_c}^2 + {R_w}^4 + 2{R_w}^2 r^2 {L_c}^2 + {R_w}^4 r^4 - 2{R_w}^4 r^2\right)^{\frac{1}{2}} \tag{8}$$

$$F_{dw-b} = 1 - F_{dw-c,q} - F_{dw-c,s} \tag{9}$$

and for the upper hemispherical enclosure

$$F_{dw-h} = 1, \qquad F_{h-h} = \frac{1}{2}, \qquad F_{h-w} = \frac{1}{2} \tag{10}$$

A plot of $F_{dw-c,q}$ and $F_{dw-c,s}$ versus $r$ is shown in Fig. 5. The remaining configuration factors were either found tabulated in [15],[16], or derived using configuration-factor reciprocity relations [15],[16].

## 3.2 Quartz Chamber and Coolant Gas Temperature

Assuming the quartz chamber is at uniform temperature $T_c'$, an enthalpy balance in terms of dimensionless variables can be written as

$$T_{amb} M_c \frac{dT_c}{dt} = \epsilon_c \boldsymbol{Q}_{lamps} \cdot \boldsymbol{u} - A_{cyl} q_c - A_{hem} q_h - Q_{convect} - \sigma \epsilon_c A_c T_{amb}^4 \left(T_c^4 - 1\right). \tag{11}$$

with

$$q_h = -\epsilon_c \sigma \bar{T}_{amb}^4 \left(-\epsilon_w \bar{T}^4 F_{h-h} + \epsilon_w \bar{T}^4 - {T_c}^4 + {T_c}^4 F_{h-h} + F_{w-h} {T_c}^4 - {T_c}^4 F_{w-h} \epsilon_w\right.$$
$$\left. - {T_c}^4 F_{w-h} F_{h-h} + {T_c}^4 F_{w-h} F_{h-h} \epsilon_w\right) / \left(-F_{h-h} + F_{h-h} \epsilon_c + 1 + F_{w-h} F_{h-h} - F_{w-h} F_{h-h} \epsilon_w\right.$$
$$\left. - F_{w-h} F_{h-h} \epsilon_c + F_{w-h} F_{h-h} \epsilon_c \epsilon_w - F_{w-h} + F_{w-h} \epsilon_w + F_{w-h} \epsilon_c - F_{w-h} \epsilon_c \epsilon_w\right) \tag{12}$$

and $q_c$ given by a similar but more complicated expression. In (11) $\boldsymbol{Q}_{lamps}$ is a vector of the total energy emitted from the three lamp banks and so the first term represents energy absorbed by the chamber directly from the heating lamps. Moreover, in (12) $\bar{T}$ is the weighted mean value of the wafer temperature.

The net energy radiated from the cylindrical ($q_c$) and the hemispherical ($q_h$) portions of the quartz chamber are

also calculated using the net-radiation method in the same manner as in Section 3.1. The total area of the wafer $A_w$ is used instead of $dA_w$, however. The configuration factors are configuration factors between finite areas found in [15] (appendix C), [16]. Their numerical values are shown in Table 1.

| | | |
|---|---|---|
| $F_{c,q-w}$ | $\frac{1}{4R_cL_c}\left(\sqrt{L_c^4+2L_c^2\left(R_c^2+R^2\right)+\left(R_c^2-R_w^2\right)^2}-\left(R_c^2-R_w^2\right)-L_c^2\right)$ | 0.2161 |
| $F_{c,s-w}$ | $\frac{1}{4R_cL_s}\left(\sqrt{\left(L_c+L_s\right)^4+2\left(L_c+L_s\right)^2\left(R_c^2+R_w^2\right)+\left(R_c^2-R_w^2\right)^2}-\right.$ | |
| | $\left.\sqrt{L_c^4+2L_c^2\left(R_c^2+R_w^2\right)+\left(R_c^2-R_w^2\right)^2}+L_c^2-\left(L_c+L_s\right)^2\right)$ | 0.0411 |
| $F_{c,q-b}$ | $\frac{1}{4R_wL_c}\left(\sqrt{\left(L_c+L_s\right)^4+4\left(L_c+L_s\right)^2R_c^2}-\sqrt{L_s^4+4L_s^2R_c^2}+L_s^2-\left(L_c+L_s\right)^2\right)$ | 0.0607 |
| $F_{c,s-b}$ | $\frac{1}{4R_cL_s}\left(\sqrt{L_s^4+4L_s^2R_c^2}-L_s^2\right)$ | 0.2807 |
| $F_{b-c,s}$ | $2\pi R_cL_sF_{c,s-b}/\pi R_c^2$ | 0.0685 |
| $F_{b-c,q}$ | $2\pi R_cL_cF_{c,q-b}/\pi R_c^2$ | 0.2284 |
| $F_{c,q-c,q}$ | $1+\frac{L_c}{2R_c}-\sqrt{1+\left(\frac{L_c}{2R_c}\right)^2}$ | 0.5678 |
| $F_{c,s-c,s}$ | $1+\frac{L_s}{2R_c}-\sqrt{1+\left(\frac{L_s}{2R_c}\right)^2}$ | 0.4385 |
| $F_{c,s-c,q}$ | $\frac{1}{4R_cL_s}\left(2L_cL_s+L_s\sqrt{L_s^2+4R_c^2}-\left(L_c+L_s\right)\sqrt{\left(L_c+L_s\right)^2+4R_c^2}+L_c\sqrt{L_c^2+4R_c^2}\right)$ | 0.2397 |
| $F_{c,q-c,s}$ | $L_sF_{c,s-c,q}/L_c$ | 0.1554 |
| $F_{b-w}$ | $\frac{1}{2}\left(1+\frac{\left(L_c+L_s\right)^2+R_c^2}{R_w^2}-\sqrt{\left(1+\frac{\left(L_c+L_s\right)^2+R_c^2}{R_w^2}\right)^2+4\left(\frac{R_c}{R_w}\right)^2}\right)$ | 0.0870 |

Table 1: The numerical values of the configuration factors.

The energy transferred from the quartz chamber to the cooling gas $Q_{convect}$ by forced convective cooling of the quartz chamber is computed from a cooling gas energy balance. If the cooling gas enters the system at ambient temperature and leaves the system at $T_{amb}T_g$,

$$T_g = T_c + \frac{\boldsymbol{Q}_g \cdot \boldsymbol{u}}{hT_{amb}A_c}\left[1-\exp\left(-\frac{A_ch}{\dot{m}_gC_{p_g}}\right)\right]+\exp\left(-\frac{A_ch}{\dot{m}_gC_{p_g}}\right) \tag{13}$$

$$\boldsymbol{Q}_g = (1-\epsilon_c)\boldsymbol{Q}_{lamps} - \int_{A_w}\boldsymbol{Q}_{lamps,w}\,dA_w. \tag{14}$$

This means that $\boldsymbol{Q}_g$, which represents the radiative energy that is eventually removed by the coolant gas, is the energy from the lamps that is not absorbed by the chamber reduced by the portion that is absorbed by the wafer

$\int_{A_w} \boldsymbol{Q}_{lamps,w}\, dA_w$. Thus, the energy removed from the chamber wall by convective cooling is calculated by

$$Q_{convect} = -\boldsymbol{Q}_g \cdot \boldsymbol{u} + \dot{m}_g C_{p_g}(T_g - 1)T_{amb}. \tag{15}$$

The last term in (11) expresses the energy that is radiated from the outer surface of the quartz chamber to the ambient.

### 3.3 Deposition Kinetics

The rate of polysilicon deposition as a function of temperature and gas composition is computed using the kinetics developed by [17], and so the rate of change of polysilicon deposition thickness $S$ is simply

$$\frac{dS}{dt} = \frac{MW_{Si}}{\rho_{Si}} \mathcal{R}_S(T, X_{SiH_4}, X_{H_2}) \times 10^4 \tag{16}$$

where

$$\mathcal{R}_S(T, X_{SiH_4}, X_{H_2}) = \frac{k_0 \exp\left(\frac{-\gamma T_{amb}}{RT}\right) X_{SiH_4} P_{tot}}{1 + bX_{SiH_4}P_{tot} + \frac{\sqrt{X_{H_2}P_{tot}}}{c}}$$

The gas phase reactant species concentrations are given in terms of mole fractions by

$$\frac{dX_{SiH_4}}{dt} = -\alpha \int_{A_w} \mathcal{R}_S(T, X_{SiH_4}, X_{H_2})\, dA_w + \frac{1}{\tau}(X_{SiH_4^{in}} - X_{SiH_4}) \tag{17}$$

$$\frac{dX_{H_2}}{dt} = 2\alpha \int_{A_w} \mathcal{R}_S(T, X_{SiH_4}, X_{H_2})\, dA_w - \frac{1}{\tau} X_{H_2}. \tag{18}$$

Equations (17) and (18) are the mass balances for silane and hydrogen and are based on the standard CSTR assumptions. The residence time $\tau$ is calculated using only the volume of the top part of the reactor chamber and a feed flow rate of 300 sccm. Numerical values and definitions of the parameters of the system are given in Table 2.

### 3.4 High-Order Simulations

The wafer thermal dynamics partial differential equation (2) was discretized with second-order accurate finite differences in the spatial dimension with $N = 76$ points. The simulations were performed using Matlab's toolbox

| | | |
|---|---|---|
| $A_w = 182.41\ cm^2$ | wafer area | |
| $A_c = 1217.31\ cm^2$ | chamber outside area | |
| $A_{cyl} = 794.83\ cm^2$ | chamber cylindrical area | |
| $A_{hem} = 422.48\ cm^2$ | chamber hemispherical area | |
| $L_c = 15.43\ cm$ | height of the cylindrical chamber part | |
| $L_s = 10\ cm$ | height of the loading chamber | |
| $R_c = 8.2\ cm$ | chamber radius | |
| $R_w = 7.62\ cm$ | wafer radius | |
| $\delta z = 0.05\ cm$ | wafer thickness | |
| $M_c = 1422.60\ J/K$ | chamber thermal mass | |
| $q_{edge} = 49\ J/cm^2$ | wafer edge incident radiation flux | |
| $b = 78.95\ torr^{-1}$ | silane adsorption rate constant | [17] |
| $c = 0.38\ torr^{1/2}$ | hydrogen desorption rate constant | [17] |
| $\epsilon_w = 0.7$ | emissivity of the wafer | [6] |
| $\epsilon_c = 0.37$ | emissivity of the quartz chamber | [19] |
| $\epsilon_s = 0.074$ | emissivity of steel | [19] |
| $h = 260 \times 10^{-4}\ W\ cm^{-2}\ K^{-1}$ | forced convection heat transfer coefficient | |
| $\kappa = 50.5 \ln T'^2 - 734.0 \ln T' + 2.69 \times 10^3\ W/m\ K$ | thermal conductivity of the wafer | [6] |
| $\rho_c = 2.6433\ g\ cm^{-3}$ | quartz crystal density | [20] |
| $C_{p_c} = 0.992\ J\ g^{-1}\ K^{-1}$ | quartz crystal specific heat | [19] |
| $V_c = 638.25\ cm^3$ | quartz chamber wall volume | |
| $\rho_w = 2.3\ g\ cm^{-3}$ | wafer density | [20] |
| $C_{p_w} = 1.06 \times 10^3 - 1.04 \times 10^5/T'\ J\ Kg^{-1}\ K^{-1}$ | wafer specific heat | [19] |
| $\dot{m}_g = 1000\ g/s$ | cooling gas flow rate | |
| $C_{p_g} = 1.0460\ J\ g^{-1}\ K^{-1}$ | air specific heat | [20] |
| $T_{g,in} = 300\ K$ | temperature of incoming air stream | |
| $X_{SiH_4^{in}} = 0.1$ | feed gas silane mole fraction | [18] |
| $P_{tot} = 5\ torr$ | total chamber pressure | |
| $MW_{Si} = 28.086\ g/gmole$ | molar weight of Si | [20] |
| $\rho_{Si} = 2.3\ g\ cm^{-3}$ | density of Si | [20] |
| $R = 8.314\ J\ gmol^{-1}\ K^{-1}$ | gas constant | |
| $T_{amb} = 300K$ | ambient temperature | |
| $\alpha = 12.961 \times 10^3\ gmole^{-1}$ | mole to mole fraction conversion factor | |
| $k_0 = 263.158\ gmole/cm^2\ s\ torr$ | preexponential kinetic constant | [17] |
| $\gamma = 153.809 \times 10^3\ J/gmol$ | activation energy for deposition | [17] |
| $\sigma = 5.674 \times 10^{-12}\ W^{-1}\ cm^{-2}\ K^{-4}$ | Boltzmann constant | |
| $\tau = 0.380\ sec$ | reactor volume / gas flow rate | |

Table 2: Nomenclature and values for the model variables and constants.

Simulink. For time integration, the Adams-Gear method proved to be the fastest and gave better results compared to the other methods included in Simulink. The nonlinear boundary condition (4) was solved simultaneously at each time step using a Gauss-Newton method. Three simulated RTP runs of 60 seconds were performed, each corresponding to having only one of the lamp banks fully activated. The time evolution of the temperature profiles are shown in Fig. 6, with initial conditions consisting of the wafer ($T'$) and the chamber ($T_c'$) both at $T_{amb}$, $S = 0$, $X_{SiH_4} = 0.1$, and $X_{H_2} = 0$. "Snapshots" of these temperature profiles will now be used as data for determining the dominant spatial temperature modes.

# 4 Model Reduction

Simulations performed with the detailed RTP system model reveal definite patterns in the response of the wafer temperature profiles to the lamp power inputs. We should expect that if designed properly, the three-zone RTP furnace, having one control input to each of the three lamp banks, should respond with at least three "important" independent spatial temperature modes, with additional modes attributable to the thermal transport through the wafer and its boundaries. By identifying these spatial modes $\psi_i(r)$ from the original simulations, we can write the wafer temperature as a time-varying linear combination of these modes

$$T = \sum_{i=0}^{N} a_i(t)\psi_i(r). \tag{19}$$

If done in an optimal manner, this procedure has the effect of removing the correlation from the original, spatially localized basis functions (resulting from the finite-difference discretization), combining them to produce *globally* valid trial functions.

One method of generating trial functions valid over the entire spatial domain is to perform a Gram-Schmidt orthogonalization of the snapshot data. However, projecting the temperature profiles collected during the high-order simulations onto trial functions produced in this manner will reveal mode amplitude coefficients which will, in general, have some degree of correlation. This redundant information can be eliminated if the eigenvectors of the mode amplitude covariance array form the linear combinations of the original trial functions which minimize the error of approximating the solutions using fewer modes of the new set of trial functions. This is the discrete version

of the Karhunen-Loéve expansion [11], also known as the proper orthogonal decomposition, method of principal components, and the method of empirical eigenfunctions. An equivalent formulation which is computationally more efficient is the *method of snapshots* developed by Sirovich [10]. In this procedure, "snapshots" in time of the temperature field are taken and used to construct the covariance array – the eigenvectors of this array are then used to construct the optimal trial fuctions directly from the snapshots.

The temperature modes computed by this procedure are then used as trial functions for discretizing the original modeling equations, giving a set of ordinary differential equations in time which describe the nonlinear interaction of the spatial modes. Discretization with trial functions empirically derived from statistical data of the specific system will require fewer modes than discretization with other series of orthogonal functions, such as Legendre, Tschebysheff, or Jacobi [22] polynomials which have a more generalized nature. The numerical and computation procedure for computing the empirically-determined trial functions is discussed in the following section.

## 4.1 Empirically-Determined Trial Functions

Construction of the empirically-determined trial functions proceeds by the computationally-efficient *method of snapshots or strobes* developed by Sirovich [10]. This numerical technique requires spatial profiles of $T(r)$ collected at discrete moments in time $t = t_i$ to give a set of $M$ snapshots $T^i(r)$, $i = 1, \ldots, M$. We begin by defining the inner product

$$(u, v) = \int_0^1 uv \, dr$$

which *does not* include a weight $w(r) = r$ which would account for the cylindrical geometry of the wafer. We chose this inner product so that measurements of temperature deviations across the wafer would be consistent with requirements set by the 1994 SIA National Technology Roadmap for Semiconductors [23], which specify absolute temperature nonuniformity limits, not weighted limits. We now define the covariance array $\boldsymbol{C}$ whose row elements $C_i$ consist of the projection of the set of $M$ snapshots onto the $i$th snapshot

$$\mathbf{C} \cdot \mathbf{p}_i = \frac{1}{M} \begin{bmatrix} (\theta_1, \theta_1) & (\theta_2, \theta_1) & \cdots & (\theta_M, \theta_1) \\ & \ddots & & \\ (\theta_1, \theta_M) & (\theta_2, \theta_M) & \cdots & (\theta_M, \theta_M) \end{bmatrix} \mathbf{p}_i = \lambda_i \mathbf{p}_i \tag{20}$$

with

$$\theta_i = T^i - \frac{\int_0^1 rT^i\,dr}{\int_0^1 r\,dr}. \tag{21}$$

Solutions $\boldsymbol{p}_i$ to the eigenvalue problem (20) are used to construct the trial functions $\psi_i$ from the set of snapshots by

$$\psi_i(r) = \sum_{j=1}^{M} p_{j,i}\,\theta_j \qquad i = 1, \ldots, M. \tag{22}$$

As derived, the eigenvalues $\lambda_n$, represent the contribution of each optimal trial function $\psi_n$ to the squared norm of the deviation of the temperature profile from its expected value. This is very useful for the model reduction scheme we intend to implement because the eigenvalues can be considered as a measure of the importance of their corresponding eigenmodes. This way we can keep only the eigenmodes with significant mean energy and truncate the remainder. The eigenvalues indicating the contribution of each spatial eigenmode to the set of temperature profile snapshots are shown in Table 3.

| Mode i | Eigenvalue $\lambda_i$ |
|---|---|
| 1 | 9.9758e-01 |
| 2 | 2.3198e-03 |
| 3 | 7.6708e-05 |
| 4 | 2.2308e-05 |
| 5 | 1.0194e-06 |
| | ⋮ |

Table 3: The eigenvalues of matrix $\boldsymbol{C}$ indicating the relative importance of the spatially varying eigenmodes.

We see in (21) that the optimal set of trial functions $\psi_i$ is computed from the correlation of wafer temperature profile deviations. This means we must define one "flat" mode $\psi_0$ to account for the mean wafer temperature in addition to the set of trial functions $\psi_i$, $i = 1, \ldots, M$ which represent the spatial temperature variations across the wafer.

After truncating the eigenfunctions with relatively small eigenvalues (smaller than $10^{-5}$) we were left with five eigenmodes including the "flat" one. These are shown in Fig. 7, normalized so that $(\psi^i, \psi^i) = 1$ for $i = 0, \ldots, 4$

## 4.2 Collocation Formulation of Galerkin's Method

If we rewrite the wafer partial differential equation (2) in the form

$$\frac{\partial T}{\partial t} = F(T) \quad \text{with} \quad T = \sum_{i=0}^{4} a_i(t)\psi_i(r)$$

we can define the *residual* as

$$\mathcal{R} = \sum_{i=0}^{4} \frac{da_i}{dt}\psi_i(r) - F\left(\sum_{i=0}^{4} a_i\psi_i\right). \tag{23}$$

In Galerkin's method, the residual (23) is projected onto each of the $n+1$ trial functions to give $n+1$ ordinary differential equations (ODEs) in time. These ODEs describe the dynamic, nonlinear coupling between the spatial modes $\psi_i$, provided that the solutions $T = \sum_{i=0}^{n} a_i(t)\psi_i(r)$ satisfy the initial and boundary conditions.

In RTCVD, the nonlinear temperature boundary conditions at the wafer edge (4) and Arrhenius-type nonlinearities in the polysilicon deposition rate expression make direct implementation of Galerkin's method difficult. Furthermore, since temperatures are normally measured at discrete locations on the wafer, it is desirable to formulate the reduced-order models in terms of the transient behavior of the temperature at a small number of points, rather than in terms of the amplitudes of the spatial modes. Both of these difficulties are eliminated when a collocation approach is used to determine the conditions – which will now consist of a set of $n+2$ ODEs, 2 algebraic equations, and computing the optimal location of the collocation points – which force the residual (23) to vanish at the interior collocation points.

### 4.2.1 Collocation Point Selection

If we assume the residual can be written in terms of a linear combination of the trial functions $\psi_i$

$$\mathcal{R} = \sum_{i=0}^{4} b_i(t)\psi_i(r) \tag{24}$$

and compute the value of the residual (24) at the $n+2$ collocation points ($n$ is the number of interior collocation points), we can write, in matrix form,

$$\boldsymbol{\mathcal{R}} = \boldsymbol{\Psi} \cdot \boldsymbol{b}\,. \tag{25}$$

Interior collocation points are normally chosen as the zeros of the highest order trial function, forcing the residual to be orthogonal to all lower-order trial functions ([12], [22], [24] ). Following this method, the four interior collocation points are taken as the roots of the fifth trial function $\psi_4$ (see Table 4 and Fig. 7). We also place one collocation point at the boundary $r_{n+2} = 1$ to satisfy the wafer edge boundary condition, and place an additional collocation point at $r_1 = 0$. This additional collocation point forces the residual to be orthogonal to all five trial functions $\psi_1$. We can see this by writing out (25) evaluated at the collocation points:

$$\begin{bmatrix} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \vdots \\ \mathcal{R}_{n+1} \\ \mathcal{R}_{n+2} \end{bmatrix}^{(n+2)\times 1} = \begin{bmatrix} 1 & \psi_1(r_1) & \cdots & \psi_{n-1}(r_1) & \psi_n(r_1) \\ 1 & \psi_1(r_2) & \cdots & \psi_{n-1}(r_2) & 0 \\ & & \vdots & & \\ 1 & \psi_1(r_{n+1}) & \cdots & \psi_{n-1}(r_{n+1}) & 0 \\ 1 & \psi_1(r_{n+2}) & \cdots & \psi_{n-1}(r_{n+2}) & \psi_n(r_{n+2}) \end{bmatrix}^{(n+2)\times(n+1)} \cdot \mathbf{b}^{(n+1)\times 1}$$

$$= \begin{bmatrix} \hat{\mathbf{p}}^T & \psi_n(r_1) \\ \hat{\mathbf{Q}} & \mathbf{0} \\ \hat{\mathbf{q}}^T & \psi_n(r_{n+2}) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ b_{n+1} \end{bmatrix}$$

The residual evaluated at interior collocation points $r_2, \ldots, r_{n+1}$ is equal to zero, since the discretized differential equation is satisfied exactly at those points. This means

$$\hat{\mathbf{Q}}\hat{\mathbf{b}} = \mathbf{0}$$

and since we find $\hat{\mathbf{Q}}$ to be invertible, $\hat{\mathbf{b}} = \mathbf{0}$. This leaves the problem

$$\begin{bmatrix} \psi_n(r_1) \\ \psi_n(r_{n+1}) \end{bmatrix} b_{n+1} = \begin{bmatrix} \mathcal{R}_1 \\ \mathcal{R}_{n+2} \end{bmatrix}$$

Since the trial functions already satisfy the symmetry condition (3) at $r = 0$, $\mathcal{R}_1$ is made to be zero by adding an ODE at that point, and since we find $\psi_n \neq 0$, $b_{n+1} = 0$ is the only possible solution. This forces the projection of

the residual onto the first $n+1$ trial functions to be zero over the entire wafer.

| Point | r |
|---|---|
| 1 | 0 |
| 2 | 0.17783 |
| 3 | 0.56813 |
| 4 | 0.83141 |
| 5 | 0.96998 |
| 6 | 1 |

Table 4: Collocation points.

### 4.2.2 Discretization Arrays

Athough collocation methods have been extensively studied in the literature, [12], [22], [24], [25], their formulas cannot be applied directly for computing discretization arrays in our problem. One difference that occurs in our formulation is that the matrix $\mathbf{\Psi}$, which contains the value of the trial functions at the collocation points, is not square. As a consequence of this, the following computational procedure is developed.

Using (19) we can write

$$\begin{aligned} T_1 &= a_0\psi_0(r_1) + a_1\psi_1(r_1) + \ldots + a_4\psi_4(r_1) \\ T_2 &= a_0\psi_0(r_2) + a_1\psi_1(r_2) + \ldots + a_4\psi_4(r_2) \\ &\vdots \\ T_6 &= a_0\psi_0(r_6) + a_1\psi_1(r_6) + \ldots + a_4\psi_4(r_6) \end{aligned} \tag{26}$$

or in matrix form

$$\boldsymbol{T}^{6\times 1} = \mathbf{\Psi}^{6\times 5} \cdot \boldsymbol{a}^{5\times 1}. \tag{27}$$

Assuming the derivatives of the wafer temperature with respect to $r$ can be expressed as linear functions of $\mathbf{T}$, we can write

$$\frac{d\boldsymbol{T}^{6\times 1}}{dr} = \boldsymbol{A}^{6\times 6} \cdot \boldsymbol{T}^{6\times 1} \tag{28}$$

where $dT_i/dr$ means $dT/dr$ evaluated at collocation point position $r_i$. Differentiating (27) with respect to $r$ gives

$$\frac{d\boldsymbol{T}}{dr} = \frac{d\boldsymbol{\Psi}}{dr} \cdot \boldsymbol{a}. \tag{29}$$

From (27) we obtain

$$\boldsymbol{\Psi}^\top \cdot \boldsymbol{T} = \boldsymbol{\Psi}^\top \cdot \boldsymbol{\Psi} \cdot \boldsymbol{a}$$

and so solving for the mode amplitude coefficients gives

$$\boldsymbol{a} = (\boldsymbol{\Psi}^\top \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^\top \cdot \boldsymbol{T}. \tag{30}$$

Substituting (30) into (29) we obtain

$$\frac{d\boldsymbol{T}}{dr} = \frac{d\boldsymbol{\Psi}}{dr} \cdot (\boldsymbol{\Psi}^\top \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^\top \cdot \boldsymbol{T}, \tag{31}$$

and so comparing (28) and (31), we can express $\boldsymbol{A}$ in terms of $\boldsymbol{\Psi}$:

$$\boldsymbol{A} = \frac{d\boldsymbol{\Psi}}{dr} \cdot (\boldsymbol{\Psi}^\top \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^\top. \tag{32}$$

An analogous procedure is used to compute the discretization arrays for higher order derivatives, or for computing the Laplacian of the temperature field

$$\nabla^2 \boldsymbol{T} = \boldsymbol{B} \cdot \boldsymbol{T} \tag{33}$$

with

$$\boldsymbol{B} = \nabla^2 \boldsymbol{\Psi} \cdot (\boldsymbol{\Psi}^\top \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^\top. \tag{34}$$

The discretization matrices generated from the five basis functions shown in Fig. 7 at the collocation points listed in Table 4 are given in the Appendix.

### 4.2.3 Quadrature Arrays

Just as differential operators can be defined in terms of the temperature measured at discrete points, integrals of temperature functions over the wafer domain can also be defined in terms of the functions evaluated at the six collocation points. Thus, we wish to write

$$\int_{A_w} f(T)\, dA = \sum_{j=1}^{6} w_j f(T_j).$$

Assuming we can write $f(T)$ in terms of the trial functions $\psi_i$,

$$f(T) = \sum_{i=0}^{4} c_i \psi_i(r), \tag{35}$$

the function $f$ can be evaluated at the collocation points to find

$$f_j = f(T_j) = \sum_{i=0}^{4} c_i \psi_i(r_j).$$

Written in matrix form

$$\boldsymbol{f}^{6\times 1} = \boldsymbol{\Psi}^{6\times 5} \cdot \boldsymbol{c}^{5\times 1} \quad \text{thus} \quad \boldsymbol{c} = (\boldsymbol{\Psi}^{\top} \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^{\top} \cdot \boldsymbol{f}.$$

Integrating (35) over the wafer gives

$$\begin{aligned}
\int_{A_w} f(T)\, dA &= \sum_{i=0}^{4} c_i \int_0^1 2\pi \psi_i(r)\, r dr \\
&= \int_0^1 2\pi \boldsymbol{\psi}\, r dr \cdot \boldsymbol{c} \\
&= \left[ \int_0^1 2\pi \boldsymbol{\psi}\, r dr \cdot (\boldsymbol{\Psi}^{\top} \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^{\top} \right] \cdot \boldsymbol{f} \\
\text{thus,} \quad \boldsymbol{w}^{1\times 6} &= 2\pi \int_0^1 \boldsymbol{\psi}\, r dr \cdot (\boldsymbol{\Psi}^{\top} \cdot \boldsymbol{\Psi})^{-1} \cdot \boldsymbol{\Psi}^{\top}.
\end{aligned} \tag{36}$$

Numerical values for the quadrature weights $w_i$ are also given in the Appendix.

As a numerical note, we mention that derivatives and integrals over the wafer of the trial functions (eqns 32,

34, and 36) are computed in the same manner as in the original, high-order simulations. This means *all* of the discretization information (number and location of discretization points, and discretization and quadrature arrays) is computed *entirely* from the snapshots.

# 5 Reduced Order Model

The computed matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are used to discretize the spatial derivative of (2) to give a reduced order model consisting of 5 ODEs in time and one algebraic equation corresponding to wafer edge temperature boundary condition (4). To illustrate the application of this method, we write the reduced-order wafer temperature model below:

For $i = 1, \ldots, 5$

$$\rho_w \left[ \left. \frac{dC_{p_w}}{dT} \right|_{T=T_i} T_i + C_{p_w}(T_i) \right] \frac{dT_i}{dt} = \frac{1}{R_w^2} \left[ \left. \frac{d\kappa}{dT} \right|_{T=T_i} (\boldsymbol{A} \cdot \boldsymbol{T})_i^2 + \kappa (\boldsymbol{B} \cdot \boldsymbol{T})_i \right] + \frac{1}{T_{amb} \delta z} [\boldsymbol{Q}_{lamps,w}(r_i) \cdot \boldsymbol{u} + q_{dw,b}(r_i) + q_{dw,t}(r_i)] \tag{37}$$

with boundary condition:

$$\frac{\kappa(T_6) T_{amb}}{R_w} (\boldsymbol{A} \cdot \boldsymbol{T})_6 = -\sigma \epsilon_w T_{amb}^4 (T_6^4 - T_c^4) + q_{edge} u_B \tag{38}$$

where $(\boldsymbol{A} \cdot \boldsymbol{T})_i$ and $(\boldsymbol{B} \cdot \boldsymbol{T})_i$ are the $i$-th elements of the vectors $(\boldsymbol{A} \cdot \boldsymbol{T})$ and $(\boldsymbol{B} \cdot \boldsymbol{T})$, respectively. Integral equations, such as (17) and (18), are discretized using the quadrature weights, e.g.,

$$\frac{dX_{SiH_4}}{dt} = -\alpha \sum_{j=1}^{6} w_j \mathcal{R}_S(T_j, X_{SiH_4}, X_{H_2}) + \frac{1}{\tau}(X_{SiH_4^{in}} - X_{SiH_4}). \tag{39}$$

The equations used in reduced order simulations of the wafer thermal dynamics are (37), (38), and (11). A comparison of the wafer temperature spatial variation predicted by the reduced order model versus the high-order simulation results is shown in Fig. 8. The curve segments between the collocation points are computed by first determining the spatial mode amplitudes (30) and then reconstructing the temperature profiles using the trial function expansion for $\boldsymbol{T}$ (19). We see that the reduced-order model successfully reconstructs a temperature profile

originally used as a snapshot for extracting eigenmodes. Comparisons in the next section will show similarly accurate predictions for conditions not used as part of the input data. It is worth mentioning that the reduced order model requires 9 times less computing time for a run than the full order model. Although this may not sound very significant for single simulation runs, it proves to be very substantial when the models are used for optimization or control.

# 6 Optimization

To test the dynamic behavior of the reduced order model when it is used in an optimization scheme, we consider the problem of finding an optimal non time-varying set of power inputs for the lamp banks, such that at final time $t' = 60$ sec the wafer temperature is uniform with a mean temperature of $1100K$. The choice of the final temperature value was made using the polysilicon deposition rate expression (16), (17), (18) in order to estimate the temperature "recipe" that would result in the desired deposition thickness of $0.5\mu m$ after $60sec$ of processing time. Since the deposition thickness is directly related to temperature in our model, temperature uniformity should lead to deposition uniformity.

By construction, the trial function $\psi_0$ will represent the mean wafer temperature, and all others are components of the spatial nonuniformity; thus, to accomplish our goal of temperature uniformity we need

$$a_i = 0 \qquad \text{for} \quad i = 1, \cdots, n \tag{40}$$

and $a_0$ to track a setpoint $a_{des}$ corresponding to the desired temperature value. Thus, a very efficient objective function J can be formed in the following way:

$$J = (a_0 - a_{des})^2 + a_1^2 + \cdots + a_4^2. \tag{41}$$

Since the trial functions are normalized, there is no need to weight the individual amplitude coefficients in (41).

Using Matlab's Sequential Quadratic Programming algorithm we obtain the optimal power set shown in Fig. 9. We should mention that a large number of iterations was necessary for convergence. The convergence did not seem

to depend, however, on the initial guess. Nevertheless, this does not guarantee that the solution obtained by the algorithm is a global minimum. In Fig. 9, we can see the results of applying the optimal set of powers to both the full and the reduced order models. The results shown are temperature profile snapshots taken at the final time $t = 60sec$. The objective function for the optimal set of lamp bank power inputs has the value $J = 8.9166 \cdot 10^{-7}$ while the $J = 1.4512$ at the initial guess. More information illustrating the dynamics of the RTCVD system for optimal power input recipe is shown in Fig. 10. These plots were generated using the optimal power input in the high-order simulator.

In Fig. 11, we see the polycrystalline silicon deposition thickness profile when the optimal set of lamp powers is used throughout the run. Two curves are shown: the solid curve represents the deposition thickness profile created by the high-order simulator, the dashed one corresponds to the reduced order model. The circles denote position of the collocation points, and once again the dashed curve was reconstructed from the trial functions and was not generated by curve fitting. We can notice that the deposition thickness is $0.4\mu$m instead of $0.5\mu$m. We can also see that although the matching of the two curves is not exact, both reflect that we achieved our goal of uniform deposition thickness (within the specified tolerance). The maximum nonuniformity observed is 3.1%, which is acceptable since the 1994 SIA National Technology Roadmap for Semiconductors [23] calls for thin film deposition control of $\pm 5\%$. The $0.5\mu$m deposition thickness goal can be achieved either by increasing the length of the process cycle or by modifying the objective function (and optimization procedure) so that the optimal dynamic temperature profile would deposit a uniform polysilicon layer of $0.5\mu$m. The later defines a different optimal control problem and is outside the scope of this paper, though further research is currently being done to address this problem.

# 7 Concluding Remarks

A numerical technique for generating low-order, nonlinear, dynamic models from spatially-resolved data taken from detailed RTP dynamic simulations was described in this paper. The reduced order model obtained as part of this study shows very good predictive characteristics and can be used for efficient process optimization.

By structuring the model reduction technique in terms of a collocation discretization method, we find that all of the numerical descretization techniques can be packaged in a "toolbox" format, where the output depends only on the data taken from the detailed, perhaps commercially-developed, simulators. Furthermore, the collocation

formulation developed in this paper seems to reveal a connection between collocation point selection and sensor placement. We envision using high-fidelity simulators to determine the number and location of sensors and to use the empirically determined trial functions $\psi_i$ as a "smarter" way of interpolating the state of the process, that is, to recover the spatial temperature or reactant concentration profiles from the minimum number of measurements, or computing accurate values for mean quantities. This way, predictions of the reduced-order model can be compared directly with the process results.

In addition to efficient optimization, the model reduction technique presented facilitates the application of several control algorithms, such as run-to-run control [18], since applications of these methods are limited to systems described by relatively small sets of ordinary differential equations. Using spatially-resolved, low order models in run-to-run control can allow us to address deposition uniformity issues, without much computational cost over lumped-parameter models. In the same manner, any control technique used for ODE models can be applied in the reduced order model that represents the distributed parameter system.

## Acknowledgments

We acknowledge the support of National Science Foundation through grant NSF EEC 94-02384. The authors also wish to thank W.J. Kiether and Professors G.W. Rubloff and F.Y. Sorrell for their data and other information regarding the NCSU RTCVD system.

# Nomenclature

| | |
|---|---|
| $a_i$ | : spatial mode amplitude coefficients |
| $\boldsymbol{A}$ | : discretization array (first derivative) |
| $\boldsymbol{B}$ | : discretization array (laplatian operator) |
| $F_{i-j}$ | : configuration factor from surface i to surface j |
| $q_{dw,b}$ | : radiative energy to the wafer bottom surface |
| $q_{dw,t}$ | : radiative energy to the wafer top surface |
| $q_{rad}$ | : radiative energy |
| $\boldsymbol{Q}_{lamps}$ | : total radiant energy generated by lamp banks |
| $\boldsymbol{Q}_{lamps,w}$ | : radiant energy flux from the lamp banks to the wafer |
| $\boldsymbol{Q}_g$ | : radiative energy removed by the coolant gas |
| $r$, $(r')$ | : dimensionless (dimensional) radial coordinate |
| $\mathcal{R}_S$ | : growth rate for the deposition of $Si$ |
| $S$ | : polysilicon deposition thickness |
| $t$, $(t')$ | : dimensionless (dimensional) time |
| $T$, $(T')$ | : dimensionless (dimensional) wafer temperature |
| $\bar{T}$ | : dimensionless weighted mean wafer temperature |
| $T_c$, $(T_c')$ | : dimensionless (dimensional) chamber temperature |
| $T_g$, $(T_g')$ | : dimensionless (dimensional) coolant gas temperature |
| $u_A$, $u_B$, $u_C$ | : lamp bank power percentage |
| $X_{H_2}$ | : hydrogen mole fraction |
| $X_{SiH_4}$ | : silane mole fraction |
| $\boldsymbol{w}$ | : quadrature weights array |
| $\psi_i$ | : eigenfunctions |
| $\boldsymbol{\Psi}$ | : discretization array (eigenfunctions) |

# References


[1] F. Roozeboom, "Rapid Thermal Processing: Status, Problems and Options after the First 25 Years" *Material Research Society Symposium Proceedings*, vol. 303, p. 149, 1993.

[2] F. Y. Sorrell, M. J. Fordham, M. C. Ozturk, and J. J.Wortman, "Temperature uniformity in RTP furnaces" *IEEE Trans. Electron Dev.*, vol. 39, pp. 75–79, Jan. 1992.

[3] W. J. Kiether, M. J. Fordham, S. Yu, A. J. S. Neto, K. A. Conrad, J. R. Hauser, F. Y. Sorrell, and J. J. Wortman, "Three-Zone rapid thermal processor system" *Proc. 2nd Int. RTP Conf.*, pp. 96–101, 1994.

[4] S. Chatterjee, I. Trachtenberg, and T. F. Edgar, "Mathematical modeling of a single wafer rapid thermal reactor" *J. Electrochem. Soc.*, vol. 139, p. 3682, 1993.

[5] R. S. Gyurcsik, T. J. Riley, and F. Y. Sorrell, "A model for rapid thermal processing: Achieving uniformity through lamp control" *IEEE Trans. Semicond. Manuf.*, vol. 4, no. 1, pp. 9–13, 1991.

[6] T. Breedijk, T. F. Edgar, and I. Trachtenberg, "A model predictive controller for multivariable temperature control in rapid thermal processing" in *Proc. of ACC*, June 1993.

[7] H. A. Lord, "Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven" *IEEE Trans. Semicond. Manufact.*, vol. 1, p. 105, 1988.

[8] R. Kakoscke, E. Bußmann, and H. Föll, "Modeling of wafer heating during rapid thermal processing" *Appl. Phys. A*, vol. 50, p. 141, 1990.

[9] J. V. Cole, K. L. Knutson, T. P. Merchant, and K. F. Jensen, "Simulation of three-dimensional flow and heat transfer in rapid thermal processing equipment" *Proc. 1994 AIChE Annual Meeting*, pp. 187-192.

[10] L. Sirovich, "Turbulence and the dynamics of coherent structures. Part I" *Quarterly of Applied Mathematics*, vol. XLV, pp. 561–571, Oct. 1987.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Boston, Academic Press, 1990.

[12] B. A. Finlayson, *The method of weighted residuals and variational principals in Chemical Engineering.* New York, Academic Press, 1972.

[13] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag, "Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders" *Phys. Fluids*, vol. 3, no. 10, p. 2337, 1991.

[14] H. Aling, J. L. Ebert, A. Emami-Naeini, R. L. Kosut, "Application of a nonlinear model reduction method to rapid thermal processing (RTP) reactors," in *Proc. IFAC Conf.*, June 1996.

[15] R. Siegel and J. Howell, *Thermal radiation heat transfer.* McGraw-Hill, 1981.

[16] J. A. J. Buschman and C. M. Pittman, "Configuration factors for exchange of radiant energy between axisymmetrical sections of cylinders, cones, and hemispheres and their bases" Tech. Rep. TN D-944, NASA, 1961.

[17] K. F. Roenigk and K. F. Jensen, "Analysis of multicomponent LPCVD processes" *J. Electrochem. Soc.*, vol. 2, no. 132, pp. 448–454, 1985.

[18] E. Zafiriou, H.-W. Chiou, and R. A. Adomaitis, "Nonlinear model based run-to-run control for rapid thermal processing with unmeasured variable estimation" in *Proc. of Symp. of Control, Diagnostics and Modeling in Semiconductor Manufacturing*, May 1995. 187th ECS meeting.

[19] Y. S. Touloukian and E. H. Buyco, *Thermophysical properties of matter.* IFI/Plenum Data Co., Pergamon Press ed., 1970.

[20] R. H. Perry and D. Green, *Perry's chemical engineers' handbook*, McGraw-Hill ed.

[21] M. D. Graham, and I. G. Kevrekidis, "Alternative approaches to the Karhunen-Loeve decomposition for model reduction and data analysis" *Computers Chem. Engng* vol. 20, no. 5, p. 495, 1996.

[22] J. V. Villadsen and W. E. Stewart, "Solution of boundary-value problems by orthogonal collocation" *Chemical Engineering Science*, vol. 22, pp. 1483–1501, 1967.

[23] *The national technology roadmap for semiconductors*, Published by the Semiconductor Industry Association, 1994.

[24] B. A. Finlayson, "Orthogonal collocation in chemical reaction engineering" *Cat.Rev. - Sci. Eng.*, vol. 10, pp. 69–138, 1974.

[25] J. P. Sorensen and W. E. Stewart, "Collocation analysis of multidiffusion and reactions in porous catalysis" Tech. Rep. n.2341, Univ. of Wisconsin, 1982.

# Appendix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -5.43029 & 4.25922 & 1.90666 & -1.41784 & 1.60122 & -0.918957 \\ 5.28153 & -7.68260 & 0.155609 & 3.57790 & -2.92903 & 1.59659 \\ -4.92808 & 6.82694 & -5.43912 & -0.345012 & 7.25802 & -3.37275 \\ 4.59220 & -7.34376 & 6.62013 & -10.8699 & -9.47648 & 16.4778 \\ 8.13803 & -15.8404 & 17.7261 & -18.2874 & -21.4236 & 29.6873 \end{bmatrix}$$

$$B = \begin{bmatrix} -198.107 & 188.748 & 37.1893 & -61.9758 & 83.2199 & -49.0750 \\ -4.50731 & -10.0670 & 14.4358 & 3.38630 & -8.46453 & 5.21669 \\ 22.8461 & -20.1783 & -17.9271 & 15.5323 & 2.21173 & -2.48482 \\ 5.76760 & 11.6056 & -21.3762 & -21.5177 & 10.8047 & 14.7161 \\ 188.635 & -331.592 & 305.836 & -186.715 & -382.294 & 406.130 \\ 71.9170 & -259.565 & 434.165 & -316.530 & -429.760 & 499.772 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 1.0 & 1.36681 & -0.807549 & 0.949399 & -2.10790 \\ 1.0 & 1.30994 & -0.683945 & 0.745961 & 0 \\ 1.0 & 0.615426 & 0.516656 & -1.22438 & 0 \\ 1.0 & -0.508823 & 1.06256 & 1.17310 & 0 \\ 1.0 & -1.44140 & -2.53203 & 0.0710101 & 0 \\ 1.0 & -1.69066 & -4.95123 & -3.08839 & -1.23573 \end{bmatrix}$$

$$w = \begin{bmatrix} -0.0405698 & 0.488088 & 1.15188 & 1.08060 & 0.434285 & 0.0273156 \end{bmatrix}$$

# Figures

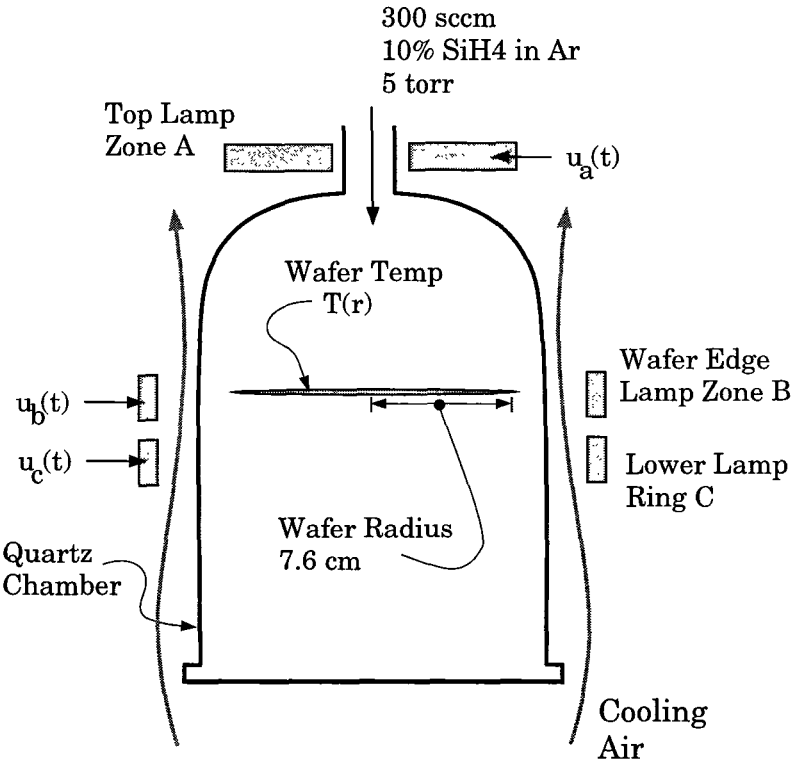


Figure 1: NC State University RTP system geometry.

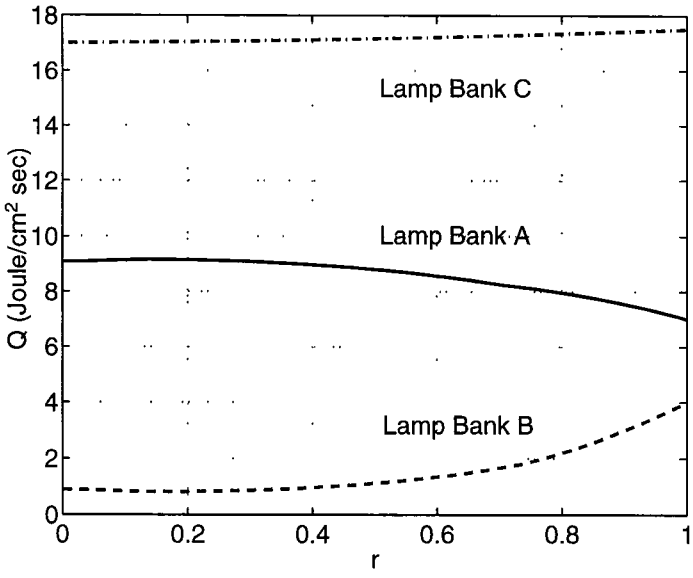


Figure 2: Radiant energy flux distributions from the individual lamp banks ($Q_{lamps,w_A}$, $Q_{lamps,w_B}$, $Q_{lamps,w_C}$)

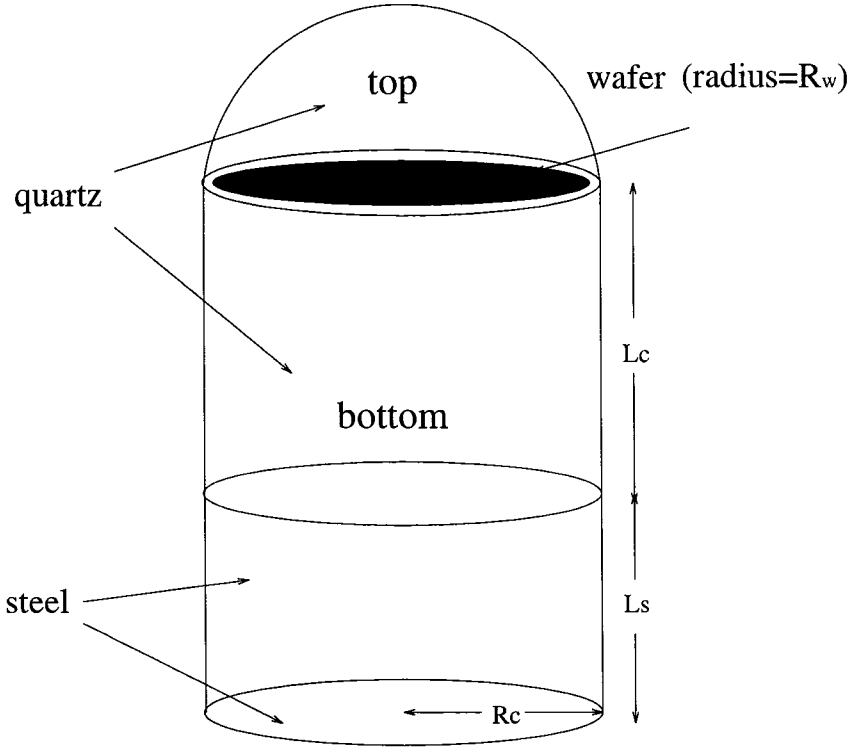


Figure 3: Diagram of the CVD chamber

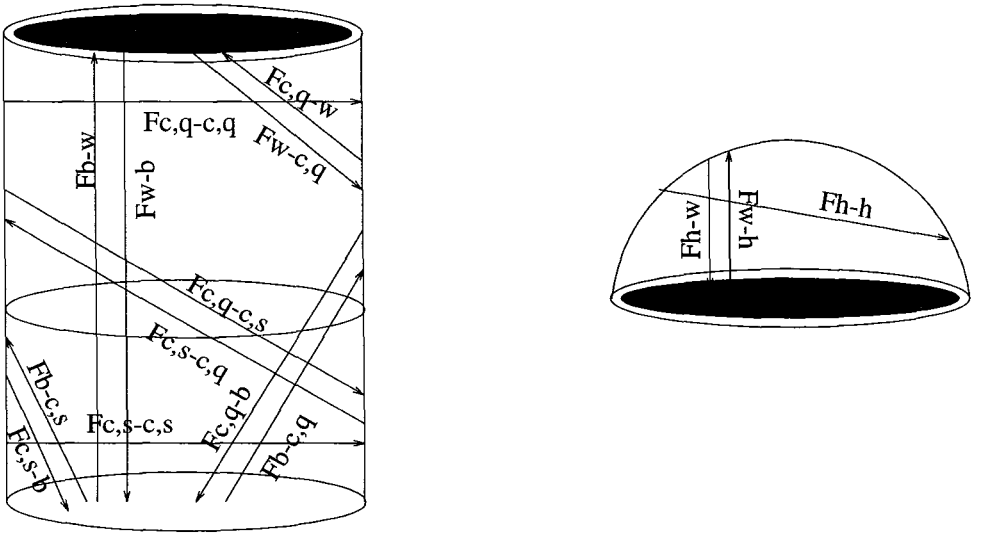


Figure 4: Diagram of the radiation exchange in the two enclosures.

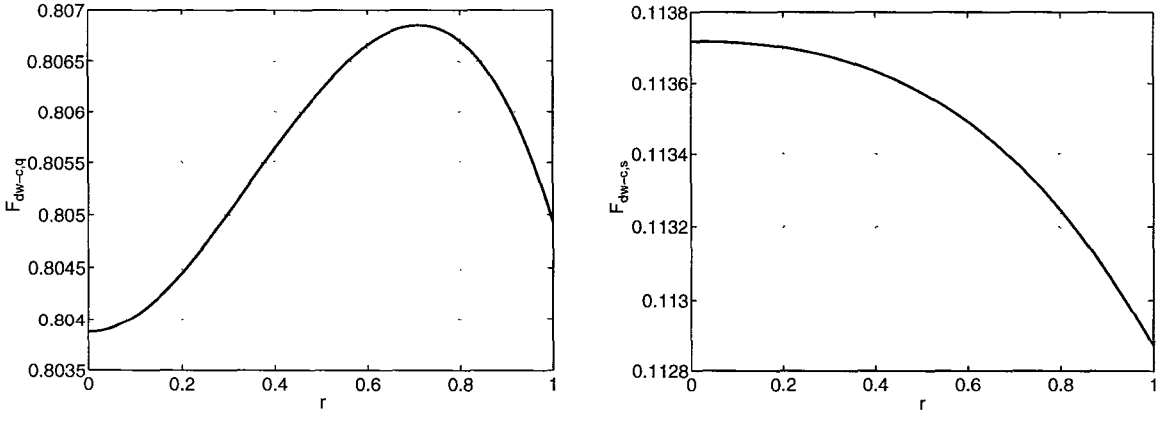


Figure 5: The configuration factors $F_{dw-c,q}$ and $F_{dw-c,s}$ as a function of radial position ($r$).

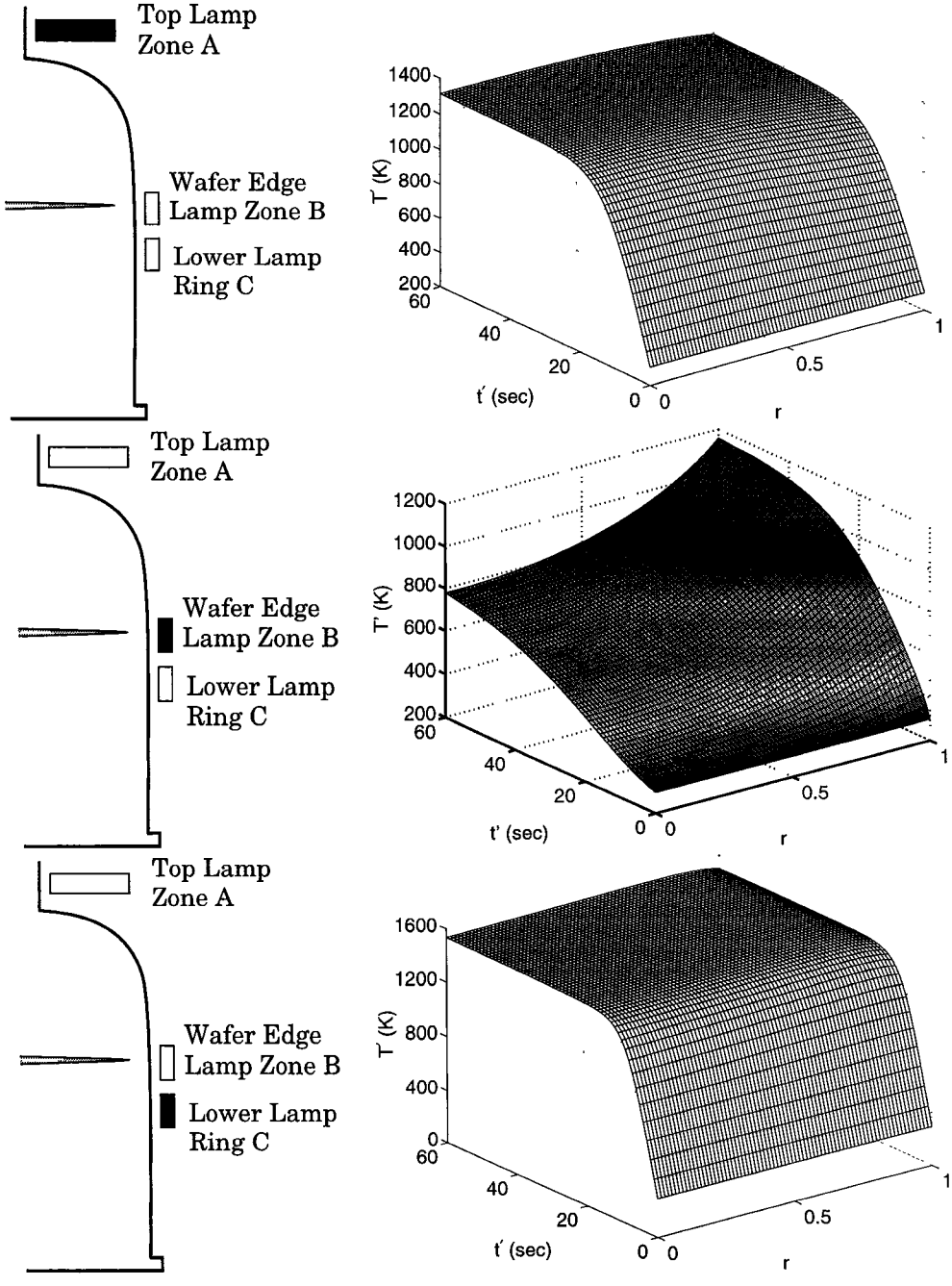


Figure 6: Time dependent temperature profiles resulting from heating by each of the individual lamp banks, used as data for determining the dominant spatial temperature modes.

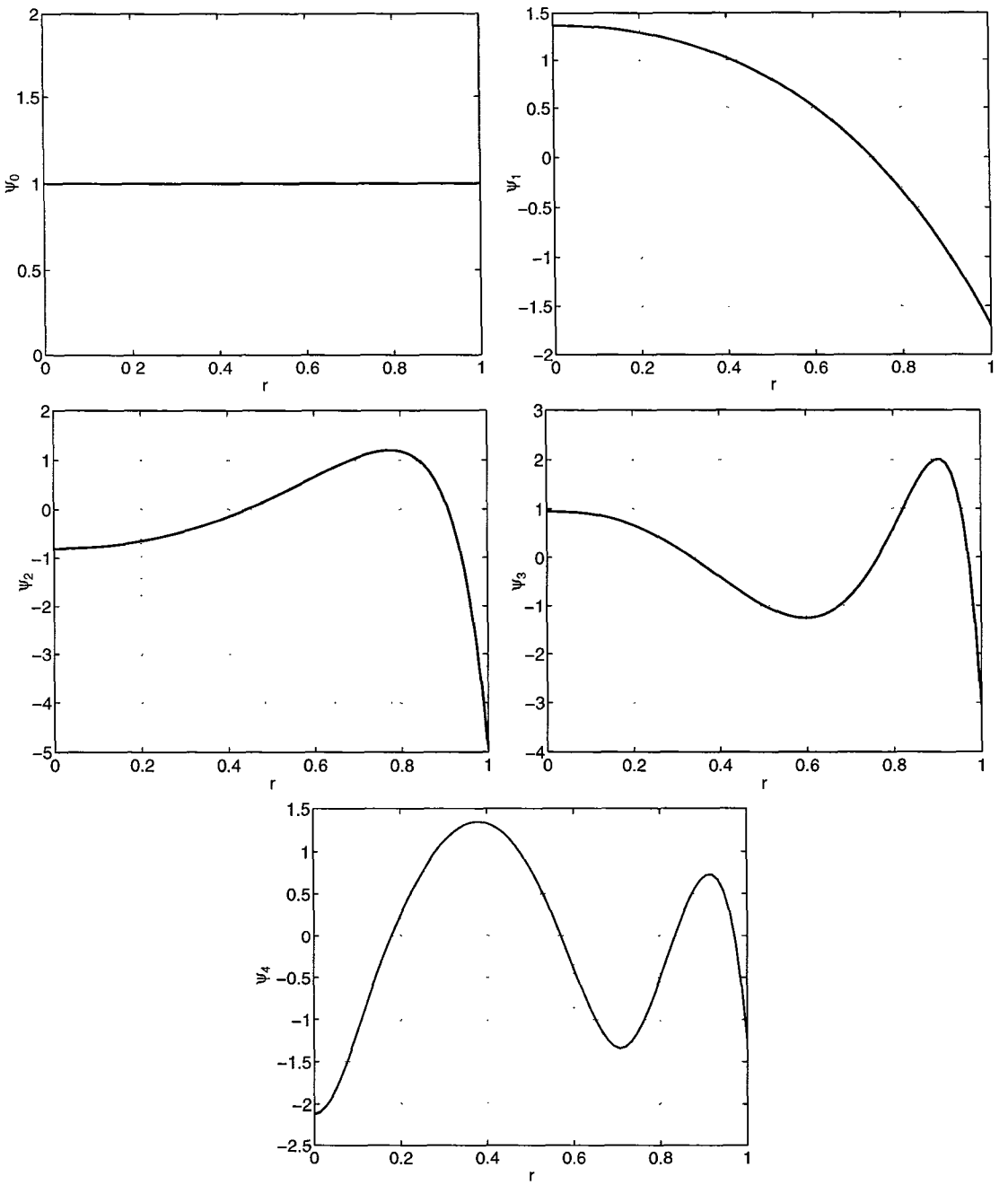


Figure 7: The dominant wafer thermal modes.

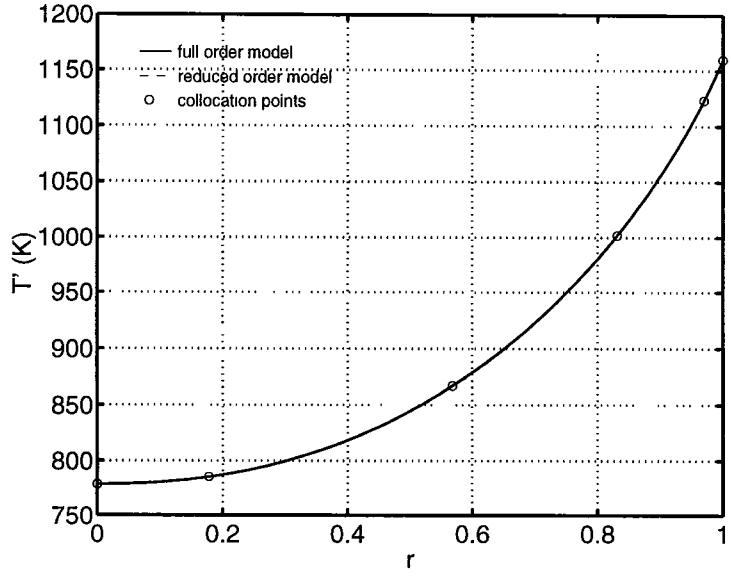


Figure 8: Comparison of the full model and the reduced order model at time=60 seconds for $u_A = 0$, $u_B = 1.0$, and $u_C = 0.0$. Both curves are plotted but are virtually indistinguishable.

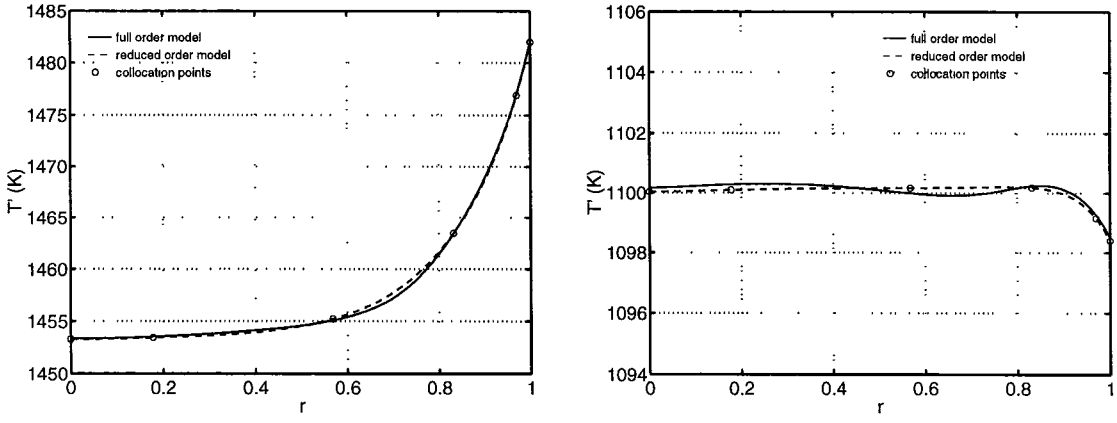


Figure 9: Comparisons of the initial guess of lamp bank power inputs (left figure: $u_A = 0.5$, $u_B = 0.5$, $u_C = 0.5$) and optimal power input settings (right figure: $u_A = 0.15213$, $u_B = 0.09512$, $u_C = 0.17703$) at time = 60 seconds.

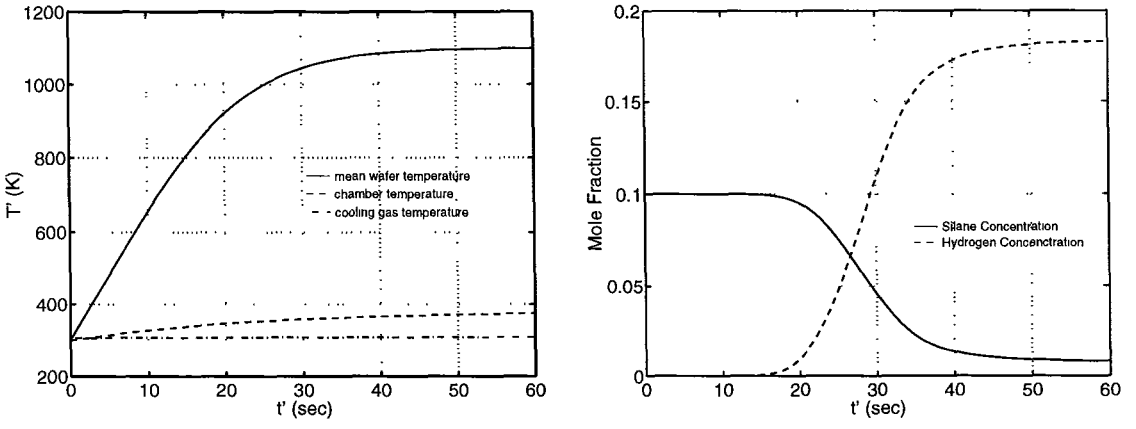


Figure 10: RTP system states for optimal input settings.

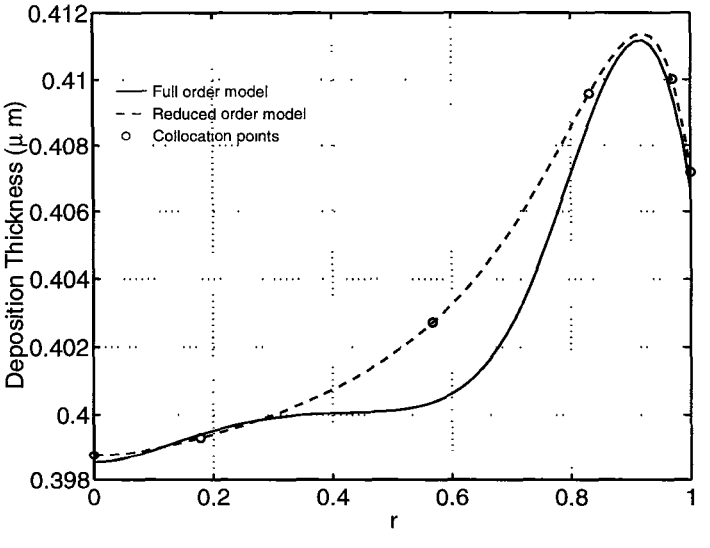


Figure 11: Optimal deposition thickness as a function of r.