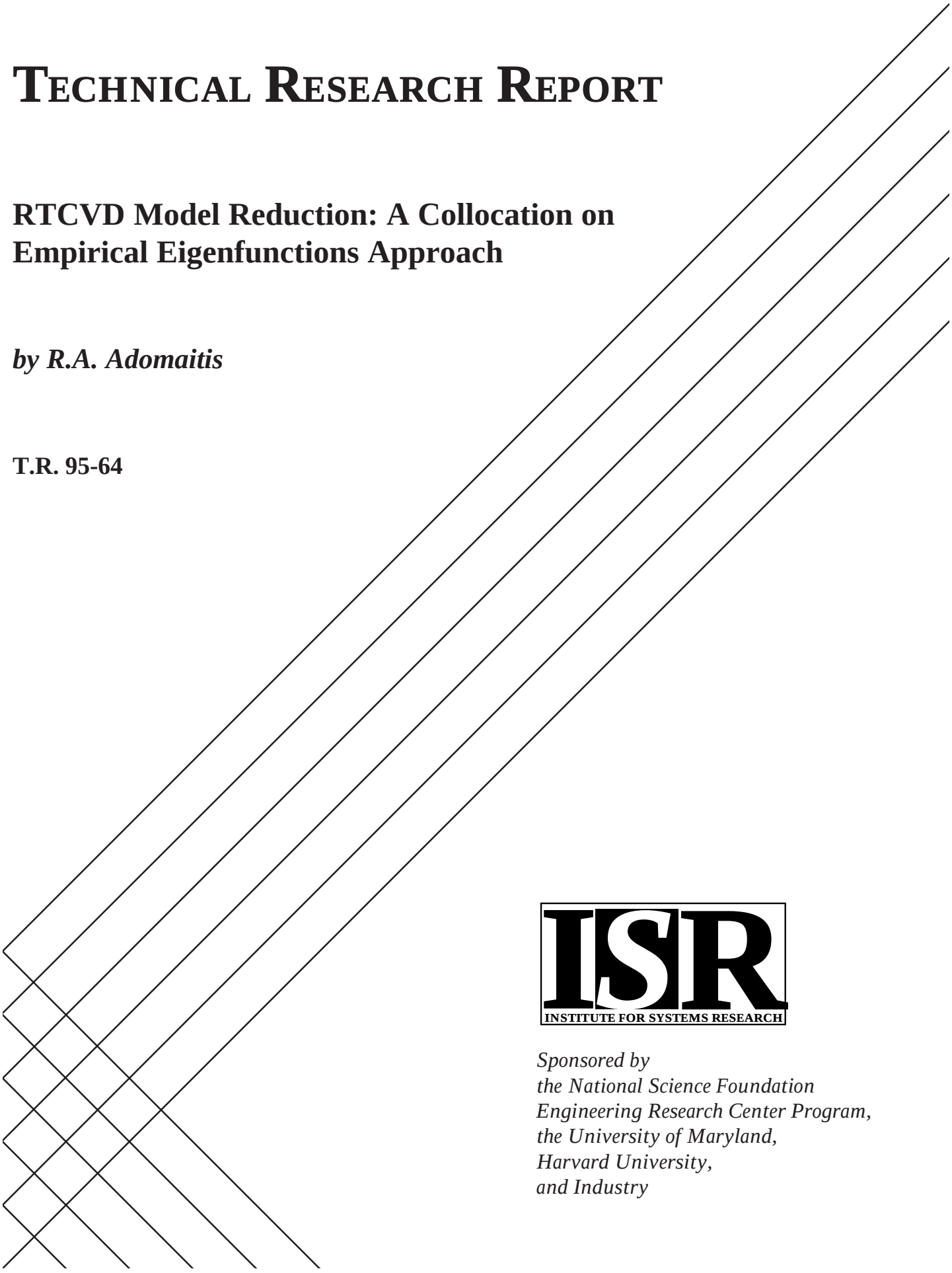
TECHNICAL RESEARCH REPORT

RTCVD Model Reduction: A Collocation on Empirical Eigenfunctions Approach

by R.A. Adomaitis

T.R. 95-64



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

To be presented at the AIChE Annual Meeting, Miami, FL Nov. 12-17, 1995.

Preliminary draft

RTCVD Model Reduction: A Collocation on Empirical Eigenfunctions Approach

Raymond A. Adomaitis
Department of Chemical Engineering and Institute for Systems Research
University of Maryland
College Park, MD 20742

July 20, 1995

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 and gas-phase reactant depletion on polysilicon deposition uniformity. A sequence of simulated runs is performed, varying the lamp power profiles so that different temperature modes are excited. The dominant spatial wafer thermal modes are extracted via proper orthogonal decomposition. A collocation formulation of Galerkin's method is used to discretize the original modeling equations giving a low-order model which looses little of the original's fidelity. We make use of the excellent predictive capabilities of the reduced model in a real-time RTP system simulator.

1 Introduction

Most rapid thermal chemical vapor deposition (RTCVD) reactor modeling research has been motivated by the reproducibility and deposition nonuniformity problems experienced in these systems. Studies of the factors which lead to spatially-nonuniform thin film deposition rates include detailed, ray-trace models of the radiant energy distribution of RTP furnaces [1, 2], the thermal dynamics of the wafer [4, 5], and deposition nonuniformity contributions from gas-phase depletion and thermal energy transport effects [6, 7].

Detailed, first principles-type simulations which include some or all of these phenomena are computationally expensive. Because RTP is dynamic process, simulations of the gas phase temperature and reacting species concentration distributions, and wafer temperature and deposition thickness profiles must be performed for the entire processing cycle. CFD calculations in complex domains and the Monte Carlo techniques used in ray-trace radiative heat transfer models performed for each update of the numerical integrator contribute to a simulator demanding large computational resources. It is no wonder that much of this type of research is performed at or in conjunction with the large National Laboratories, making use of their extensive computing facilities.

Simulation studies have developed hand-in-hand with research focusing on optimizing operating conditions and processing recipes, improving furnace designs, and developing controllers (including statistical process control, run-to-run, and real-time feedback control) to minimize deposition nonuniformity. High-fidelity simulators, however, are not practical for use in controller

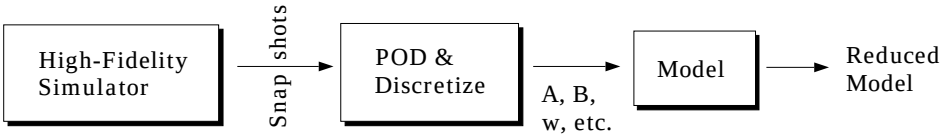


Figure 1: Model reduction methodology.

development, implementation of model-based controllers, efficient recipe optimization, or for interpretation of sensor and metrology data. This has led to the use of empirically-determimined models for control, such as input-output linear dynamic models [8] or response-surface models. "Virtual" process simulators – simulations which can be run in real-time – have also relied on relatively simple descriptions of the processing chamber dynamics. In all of these cases, the models lack true predictive capabilities and cannot resolve mechanisms responsible for spatial nonuniformities.

When viewed from the proper perspective, most dynamical behavior appears to be low-dimensional. Solution to the general problem of generating computationally-reasonable models of distributed parameter systems hinges on this fact and is based on extracting the few spatial modes responsible for most of the spatial structure. Determining the dynamic, nonlinear coupling between these *coherent structures* [13] using Galerkin's method [10] to project the transport equations onto these modes gives low-order, predictive models. This model reduction works by shifting most of the complexity to the empirically-determined eigenfunctions, leaving ordinary differential equations to describe the time-dependence of the mode amplitudes. Explicit representation of system parameters and control inputs is retained. Recognizing that the dominant dynamical behavior of a distributed parameter system (DPS) is low dimensional and relying on numerical techniques analogous to those discussed in [13], Gay and Ray [9] developed a method for identifying linearized DPS models in the form of integral equations from input/output data. In a related study, a weakly nonlinear control theory which depended on identifying the dominant (relatively slow) eigenmodes and adjoint eigenfunctions was developed by Chen and Chang [11].

This study presents a model reduction methodology for generating low-order RTP models, suitable for optimization, model-based control, real-time simulators, and model-based sensing, from detailed RTP dynamic simulators. We consider this research as a first step in developing a "tool box" which fits in between a detailed simulation package and low-order model (see Fig. 1). In this Figure, we make distinctions between the "High-fidelity Simulator" (consisting of first-principles modeling equations, empirical correlations, etc., combined with a numerical solution procedure) and the "Model" (which consists of modeling equations which may be simpler or similar to those used in the simulator). The goal of our research is not to develop a new, detailed RTP dynamic simulator, but to understand the interaction between the final uses of the reduced model, the model reduction, and the detailed simulation.

In addition to the precision, consistency, predictive capabilities, and computational efficiency requirements imposed on the reduced order simulators, these reduced models must also retain their "intuitive feel." In other words, models based on describing the dynamics of mode amplitudes associated with the empirically-determined eigenfunctions are not as useful as an explicit description of the state of wafer or gas phase at discrete points on the wafer or reaction chamber. With this in mind, we have taken a collocation formulation approach to the discretization procedure. This results in a low-order model in terms of the wafer temperature and deposition thickness represented at discrete points on the wafer surface, and a similar representation of the the tool state. This numerical technique also facilitates model validation, since there is a direct connection between sensor measurements and predictions of the reduced simulator. The relationship between collocation point selection and optimal sensor placement will be discussed.

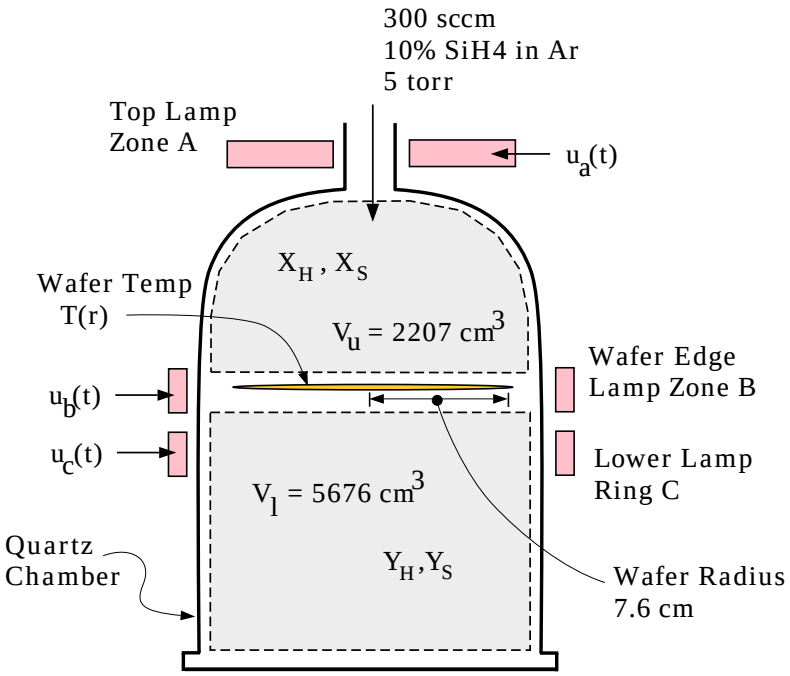


Figure 2: NC State RTP system geometry.

This paper appears to be a novel application of signal processing, weighted-residual, and collocation discretization techniques to generating simple and accurate RTP dynamic models. The combination of methods is particularly well-suited to RTP and other batch processes used in electronic and structural materials manufacturing processes – dynamic processes so complicated that reduced-order models must be used for process optimization or other situations where numerous simulations must be performed.

2 The RTP System Model

As a first step in developing an RTP system model reduction methodology, we focus on developing a "detailed" simulator of an RTCVD reaction system for depositing thin films (0.5 μm) of polycrystalline silicon from silane onto 6 inch wafers. The RTP system model, parameters, and operating conditions were chosen to match those of the three-zone RTP system [2] located at the North Carolina State University Advanced Electronics Materials Engineering Research Center (see Fig. 2).

2.1 Wafer Thermal Dynamics

Heat transfer to the gas phase from the wafer in this system is negligible because of the relatively low operating pressure (5 torr). The temperature rise generated by the heat of the polysilicon deposition reaction was also found to be small relative to wafer conduction and radiation. Previous simulation studies [4] indicate the temperature difference between the top and bottom wafer surfaces is small. Because the wafer is slowly rotated during processing, temperature variations in the azimuthal direction are minimized leaving only temperature variations in the radial direction. Under these observations and assumptions, we can write the wafer thermal

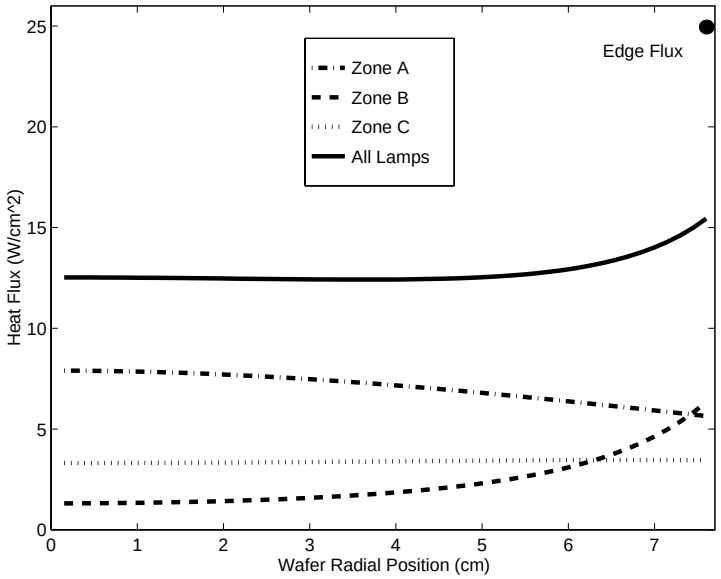


Figure 3: Incident radiation flux profiles for $u_{a,b} = 1$ and $u_c = 0.4$.

dynamics equation over the wafer domain $\mathcal{D}$ as

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T + \epsilon(1 - T^4) + \mathbf{q}^T \cdot \mathbf{u} \tag{1}$$

subject to the boundary condition

$$\frac{\partial T}{\partial r} = \epsilon_{\mathrm{ed}}(1 - T^4) + q_{\mathrm{ed}} u_b \quad \text{on } \partial\mathcal{D}. \tag{2}$$

The quantity $\mathbf{q}^T \cdot \mathbf{u}$ represents the wafer incident radiation flux from the three-zone furnace:

$$\mathbf{q} = \omega \begin{bmatrix} q_a(r) \\ q_b(r) \\ q_c(r) \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} u_a(t) \\ u_b(t) \\ u_c(t) \end{bmatrix}.$$

Parameters, their values, and definition of the dimensionless variables are discussed in more detail in the Appendix.

The lamp bank flux profiles ($q_a(r)$, $q_b(r)$, $q_c(r)$, $q_{b_{\mathrm{edge}}}$) are computed from a simple viewfactor model. This relates the power output of the 10 2kW bulbs of bank A and 32 1kW bulbs of each bank B and C to the incident flux at the wafer surface as a function of the distance to each point on the wafer and the angle of incidence of the ray. Similar calculations are performed for computing the wafer edge radiation flux. No reflections were included in our calculations. Computed flux profiles which would correspond to 100% lamp power to banks A and B ($u_{a,b} = 1$) and bank C at 40% power ($u_c = 0.4$) are shown in Fig. 3. It is interesting to compare the similarities between these results with the predicted flux profiles shown in Figure 2 of [2], corresponding to the same power inputs.

2.2 Deposition Dynamics

The gas phase silane and hydrogen balances are described in terms of hydrogen and silane mole fractions for the upper chamber region (X_H, X_S) and lower region (Y_H, Y_S). The species mass balances depend on the total rate of deposition R_c across the wafer and so are given in terms

of the integral equations

$$\begin{aligned}
\frac{dX_H}{dt} &= X_{H_{\text{feed}}} - X_H + 2\beta \int_{\mathcal{D}} R_c(X_H, X_S, T)\, dA \quad \text{with} \quad X_{H_{\text{feed}}} = 0 \\
\frac{dX_S}{dt} &= X_{S_{\text{feed}}} - X_S - \beta \int_{\mathcal{D}} R_c(X_H, X_S, T)\, dA \quad \text{with} \quad X_{S_{\text{feed}}} = 0.1 \\
\frac{dY_H}{dt} &= \frac{\tau}{\tau_l}(X_H - Y_H) + 2\beta \frac{\tau}{\tau_l} \int_{\mathcal{D}} R_c(Y_H, Y_S, T)\, dA \\
\frac{dY_S}{dt} &= \frac{\tau}{\tau_l}(X_S - Y_S) - \beta \frac{\tau}{\tau_l} \int_{\mathcal{D}} R_c(Y_H, Y_S, T)\, dA.
\end{aligned} \tag{3}$$

The dimensionless rate expression for polysilicon deposition from silane is based on the assumption of silane and hydrogen adsorption/desorption equilibrium with the decomposition of silane on the surface as the rate-controlling step:

$$R_c(X_H, X_S, T) = \frac{k_0 \exp(-\gamma/T) X_S}{1 + bX_S + c\sqrt{X_H}}. \tag{4}$$

Because the wafer is heated completely through, polysilicon deposition occurs on both wafer surfaces. The dimensionless rate of polysilicon film growth on the upper (S_u) and lower (S_l) wafer surfaces are simply

$$\frac{dS_u}{dt} = R_c(X_H, X_S, T) \qquad \text{and} \qquad \frac{dS_l}{dt} = R_c(Y_H, Y_S, T), \tag{5}$$

respectively. This concludes the description of the detailed model development.

3 Simulations

The wafer thermal dynamics PDE (1) was discretized with second-order accurate finite differences in r. For the simulations reported in this paper, the number of discretization points n was fixed at 100. It is important to note that in discretized form, the two polysilicon deposition thickness ODEs (5) become $2n$ ODEs due to their dependence on wafer temperature T. Discretization of the integral equations (3) gives four additional ODEs, giving a total of $3n+4$ ODEs in time after discretization. The nonlinear wafer edge boundary condition (2) was discretized and solved by Newton-Raphson iterations during each time step update computation.

Runs performed with the detailed simulator at this point were meant to excite different wafer thermal and deposition thickness modes in response to inputs from the three lamp banks. While identification techniques for linear time-invariant systems are better developed, designing a systematic method for exciting and identifying the spatial modes of nonlinear DPS presents more of a challenge. The latter statement is supported by the variety of techniques employed: assumptions of ergodicity in turbulent flows [14]; the "extended" proper orthogonal decomposition approach discussed in [10]; observing the system response to random distributed noise [11]; and the independent step-test identification method [9].

For this study, three simulated RTP runs of 60 seconds were performed, each corresponding to having only one of the lamp banks activated. Initial conditions consisted of the wafer at ambient temperature and the reaction chamber filled with argon. At the start of the run, the flow of 10% silane was initiated and the lamp bank power was ramped up to 100% over five seconds. Full power was held for 40 seconds, and then ramped linearly down to zero over five seconds. The wafer was removed from the reaction chamber after a total of 60 seconds. Data were collected from each run as a sequence of "snapshots" of the wafer state (temperature and

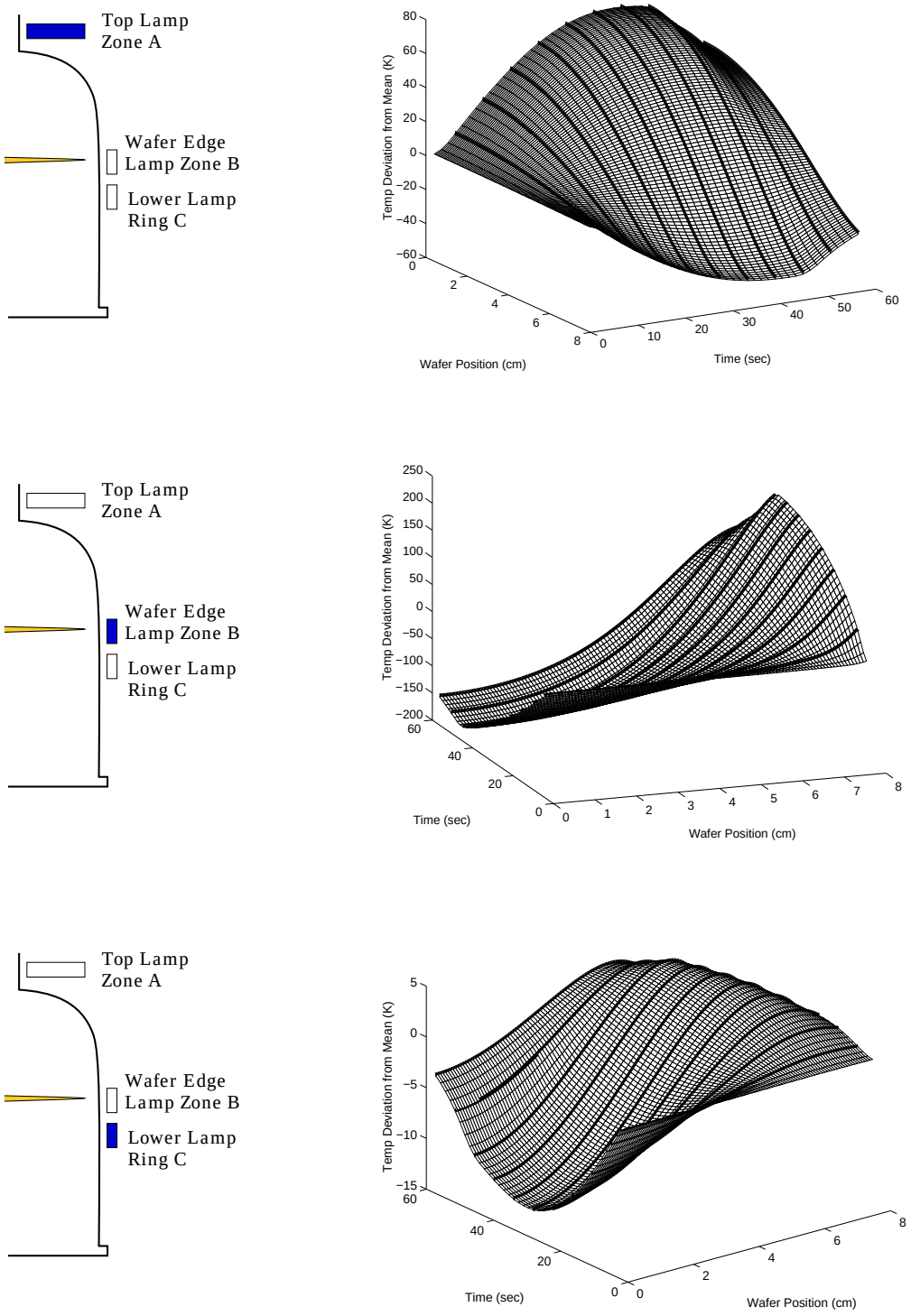


Figure 4: Time-dependent temperature profiles resulting from heating by each of the individual lamp banks. Each curve on the surfaces indicates a temperature profile collected as one of the snapshots.

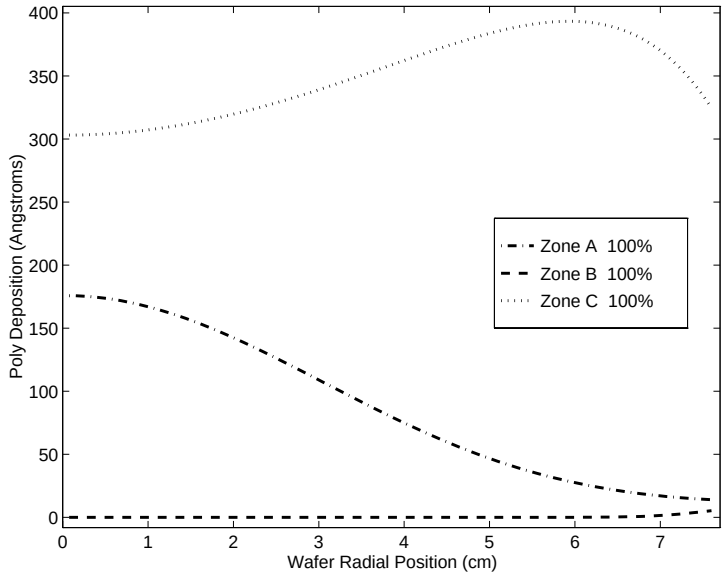


Figure 5: Final polysilicon deposition thickness profiles corresponding to each of the three snapshot collection runs.

deposition thickness). See Fig. 4 for the temperature snapshots collected and Fig. 5 for the polysilicon deposited in each of the runs.

4 Model Reduction

In the simulations of the previous section, we probed the phase space in which solutions to the wafer thermal dynamics differential equation

$$\frac{\partial T}{\partial t} = f(T, T_{rr}, \ldots, \mathbf{u}) \quad \text{in } \mathcal{D}, \quad g(T, T_r, \ldots, u_b) = 0 \quad \text{on } \partial\mathcal{D} \tag{6}$$

live. We would like to find a series of orthonormal eigenfunctions ψ_i spanning this space such that the temperature field can always be described by a linear combination of these trial functions

$$T(r,t) = T_{\partial\mathcal{D}} + (1-r^2)\sum_{i=0}^{\infty} a_i(t)\psi_i(r) \tag{7}$$

where, because of the cylindrical symmetry of the wafer, we define the orthogonality by

$$\begin{aligned} \frac{1}{\pi}\int_{\mathcal{D}} \omega(r)\,\psi_i\psi_j dA = 2\int_0^1 (1-r^2)\,\psi_i\psi_j\, r dr &= 0 \quad i \neq j; \\ &= 1 \quad i = j. \end{aligned}$$

The dynamics and nonlinear interactions of the modes can be determined by substituting a truncated version ($T^{(m)}$) of (7) into (6) to define the residuals

$$\mathcal{R}(r,t) = \frac{\partial}{\partial t}(T^{(m)}) - f(T^{(m)}, T^{(m)}_{rr}, \ldots, \mathbf{u}) \quad \text{with} \quad \mathcal{R}_{\partial\mathcal{D}} = g(T^{(m)}, T^{(m)}_r, \ldots, u_b) = 0. \tag{8}$$

Assuming the residual $\mathcal{R}$ lives in the space spanned by ψ_i, our goal is to choose a set of $a_i(t)$ and $T_{\partial\mathcal{D}}(t)$ so that

$$\mathcal{R} = \mathcal{R}_{\partial\mathcal{D}} + (1 - r^2) \sum_{i=0}^{\infty} b_i(t)\psi_i(r) \tag{9}$$

is minimized. Projecting the residual $\mathcal{R}$ onto the first m ψ_i gives

$$b_j = F_j \left(\mathbf{a}, \frac{da_j}{dt}, T_{\partial\mathcal{D}}, \mathbf{u} \right) \qquad \text{for} \qquad j = 0, \ldots, m. \tag{10}$$

Thus, trajectories defined by the ordinary differential equations in time of the R.H.S. of (10), with $T_{\partial\mathcal{D}}$ uniquely determined by $\mathcal{R}_{\partial\mathcal{D}} = 0$ in (8), will result in $b_0, b_1, \ldots, b_m = 0$. This forces the residual to lie in the space spanned by $\psi_{m+1}, \psi_{m+2}, \ldots$. We define the probability of finding mode i in the data collected during the detailed simulation by the ensemble average of the projection of a large set of M temperature field snapshots onto mode i:

$$\lambda_i = \frac{1}{M} \sum_{j=0}^{j=M} (T_j, \psi_i)^2 \qquad \text{arranged so that} \qquad \lambda_1 > \lambda_2 > \ldots > \lambda_n > \ldots. \tag{11}$$

Since

$$\sum_{j=m+1}^{\infty} \lambda_j \to 0$$

this procedure (Galerkin's method) converges as m increases, giving an accurate discretization for sufficiently large m.

4.1 Proper Orthogonal Decomposition

The number of trial functions m required for Galerkin's procedure to converge to an acceptably small residual depends heavily on the choice of the orthonormal basis ψ_i. The trial functions can be chosen from a set of "theoretical" trial functions, such as the Jacobi polynomials used by Villadsen and Stewart [15] or any of those discussed in Gottlieb and Orszag [16]. In some cases, a more efficient set of trial functions can be found as the set of eigenfunctions corresponding to the linearization of the transport equations [17, 18].

Alternatively, we can use the high-fidelity simulations themselves to generate the trial functions. We first define the projection operator by

$$\mathcal{P}_m T(r,t) = \sum_{i=0}^{m} a_i(t)\psi_i(r) \quad \text{with} \quad \psi_0 = 1, \quad a_i = (T, \psi_i). \tag{12}$$

The inner product is defined as

$$(\psi_i, \psi_j) = \frac{1}{\pi} \int_{\mathcal{D}} \psi_i \psi_j \, dA = 2 \int_0^1 \psi_i \psi_j \, r dr.$$

Now, consider the snapshots as vectors in the space where temperature profiles of (1) exist. While an orthogonal basis for this space can be computed with the Gram-Schmidt orthogonalization procedure, the trial functions generated in this way will not, in general, satisfy condition (11). An optimal set of trial functions is efficiently computed with the Proper Orthogonal Decomposition method based on snapshots [13]. What makes this numerical technique work is recognition that inner products of M snapshot vectors can define the coefficients of a quadratic form in a

length M vector of snapshot amplitudes [19]. The principal axes of the quadric surface are the characteristic vectors of the array $\mathbf{C}$ whose row elements $C_{i,j}$ consist of the projection of the set of M snapshots onto the ith snapshot

$$\mathbf{C}\mathbf{e}_i = \begin{bmatrix} (\theta_1,\theta_1) & (\theta_2,\theta_1) & \cdots & (\theta_M,\theta_1) \\ & \ddots & & \\ (\theta_1,\theta_M) & (\theta_2,\theta_M) & \cdots & (\theta_M,\theta_M) \end{bmatrix} \mathbf{e}_i = \lambda_i \mathbf{e}_i \tag{13}$$

with

$$\theta_i = \alpha_T \left[T_i - \frac{1}{\pi} \int_{\mathcal{D}} T_i \, dA \right] + \alpha_S \left[S_{u\,i} - \frac{1}{\pi} \int_{\mathcal{D}} S_{u\,i} \, dA \right]$$

where T_i is the temperature field snapshot taken at time t_i; the same notation applies to the top surface poly deposition thickness S_u. The weight constants α_T and α_S are chosen to balance the contributions the two terms.

The eigenvalue λ_i associated with each eigenvector e_i is equivalent to the probability defined in (11), and the eigenmodes are reconstructed from the snapshots as

$$\psi_i(r) = \sum_{j=1}^{M} e_{j,i}\, \theta_i(r) \qquad i = 1, \ldots, M. \tag{14}$$

The first three most energetic modes are shown in Fig. 6; a comparison to a set of theoretical basis functions based on the Jacobi polynomials is made in this figure. We can measure the relative "effectiveness" of the different trial functions by re-defining the projection operator as

$$\mathcal{P}T_i = a_0 + a_1\psi_1 + \ldots + a_m\psi_m + a^c\psi^c \quad \text{with} \quad a_j = (T_i, \psi_j);$$

$$T_i^c = (1 - \mathcal{P}_m)T_i, \qquad \psi^c = \frac{T_i^c}{(T_i^c, T_i^c)}.$$

The eigenvalues corresponding to the probability of finding mode j in the set of snapshots are defined as before (11), but are how normalized so that the sum of λ_1 through λ^c (the spatial modes) is unity.

4.2 Collocation Formulation

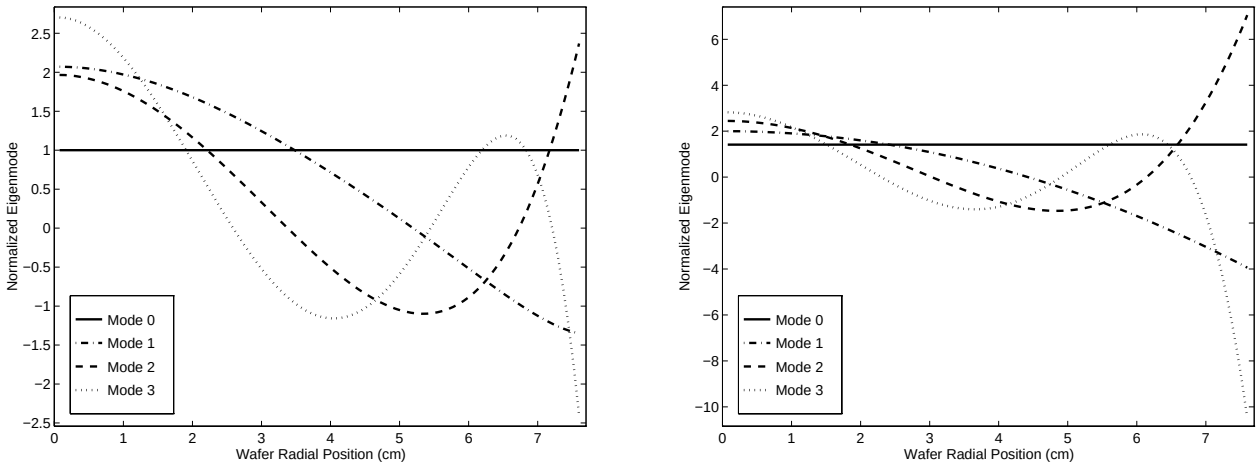
Evaluating the integrals required to find (10) might be impossible due to the exponential and other nonlinearities. Even if it were, the resulting ODEs will probably lack "intuitive feel" for the relationship between sensor measurements, the state of the system, and the control inputs, even if it is explicit. These facts motivated reformulating the model reduction procedure so that most of the workload is reduced to using a set of Matlab routines. The collocation approach allows straightforward use of commercial and other, already developed detailed simulators.

If we select a set of n points in the wafer (in $\mathcal{D}$) plus one on the wafer edge ($\partial\mathcal{D}$) and denote $T_k = T(r_k) \in \mathbf{T}$ for $k = 1, \ldots, n+1$ and (analogous to finite difference discretizations) assume we can write derivatives as linear functions of the T_k:

$$\begin{aligned} \frac{dT_k}{dr} &= A_{k,1}T_1 + \cdots + A_{k,n+1}T_{n+1} \qquad \text{for} \qquad k = 1, \ldots, n+1 \\ \frac{d}{dr}\mathbf{T} &= \mathbf{A}\mathbf{T}. \end{aligned} \tag{15}$$

Evaluating the truncated projection operator (12) at the discretization points r_k:

$$\mathcal{P}_n T(r_k, t) = T_k = \sum_{i=0}^{n} a_i(t)\psi_i(r_k) \tag{16}$$



Spatial Mode	Energy (Empirical)	Energy (Jacobi)
λ_1	0.9395	0.4718
λ_2	0.0555	0.1562
λ_3	0.0027	0.0218
λ^c	0.0023	0.3502

Figure 6: Empirically determined eigenmodes (left) versus modes based on the Jacobi polynomials (right).

or

$$\mathbf{T} = \mathbf{Q}\mathbf{a} \qquad \text{with} \qquad \mathbf{Q} = \begin{bmatrix} 1 & \psi_1(r_1) & \cdots & \psi_n(r_1) \\ 1 & \psi_1(r_2) & \cdots & \psi_n(r_2) \\ & & \vdots & \\ 1 & \psi_1(r_{n+1}) & \cdots & \psi_n(r_{n+1}) \end{bmatrix} \tag{17}$$

so the elements of $\mathbf{Q}$ are functions of the time-invariant trial functions and the collocation point positions. This allows us to write

$$\begin{aligned} \frac{d}{dr}\mathbf{T} &= \left(\frac{d}{dr}\mathbf{Q}\right)\mathbf{Q}^{-1}\mathbf{T} \\ &= \mathbf{A}\mathbf{T} \end{aligned} \tag{18}$$

for $n+1 \leq m$. A similar procedure to (15-18) for the Laplacian operator gives the discretization matrix $\mathbf{B}$.

4.3 Recovering Spatial Mode Amplitudes

An important aspect of the model reduction that is not lost in collocation formulation is that the spatial modes are easily recovered using only the $\mathbf{Q}^{-1}$ array, the vector of temperatures at the collocation points, and definition (17). With $\mathbf{a} = \mathbf{Q}^{-1}\mathbf{T}$, we denote $\mathbf{w}$ as the first row of $\mathbf{Q}^{-1}$ and immediately obtain the result $a_0 = \mathbf{w} \cdot \mathbf{V}$. Because the snapshots used in (13) were defined to have zero mean, any linear combination of them will also have zero mean, thus

$$\int_{\mathcal{D}} \psi_i \, dA = 0 \qquad \text{for} \qquad i \geq 1.$$

This means integrals over the wafer domain, such as those used in (3), can be quickly computed as

$$\int_{\mathcal{D}} R_c(X_H, X_S, T)\, dA = \sum_{k=1}^{n+1} w_k R_c(X_H, X_S, T_k)$$

In addition to speeding up the computation of integral quantities, straightforward recovery of the mode amplitude magnitudes is also important in defining optimization objective functions and makes for a "fool-proof" uniformity control criterion by separating the desired mean wafer temperature ($a_0 = T_{\text{set}}$) and temperature uniformity across the wafer ($a_1, a_2, \ldots = 0$).

4.4 Collocation Point Selection

In the collocation formulation, the definition of the residual is similar to (8):

$$\mathcal{R}_i(\mathbf{T}, t) = \frac{\partial}{\partial t}(T_i) - f_i(\mathbf{T}, \mathbf{B}, \ldots, \mathbf{u}) \quad \text{with} \quad \mathcal{R}_{\partial\mathcal{D}} = g(\mathbf{T}, \mathbf{A}, \ldots, u_b) = 0 \tag{19}$$

and $i = 1, \ldots, n$. This means we can choose T_{n+1} such that $\mathcal{R}_{\partial\mathcal{D}} = 0$ and determine a set of n ODEs in time which will satisfy $\mathcal{R}_i(\mathbf{T}, t) = 0$.

Once again, assuming the residual can be expressed as a linear combination of the trial functions ψ_i, allows us to write

$$\mathcal{R} = \sum_{i=0}^{n} b_i(t)\psi_i(r) \quad \text{or} \quad \mathbf{b} = \mathbf{Q}^{-1}\mathbf{R}. \tag{20}$$

In discretized form, and by choosing the collocation points to be the zeros of the nth eigenfunction ψ_n, (20) can be written as

$$\begin{bmatrix} 1 & \psi_1(r_1) & \cdots & \psi_{n-1}(r_1) & 0 \\ 1 & \psi_1(r_2) & \cdots & \psi_{n-1}(r_2) & 0 \\ & & \vdots & & \\ 1 & \psi_1(r_{n+1}) & \cdots & \psi_{n-1}(r_{n+1}) & \psi_n(r_{n+1}) \end{bmatrix} \mathbf{b} = \begin{bmatrix} \hat{\mathbf{Q}} & \mathbf{0} \\ \hat{\mathbf{q}}^T & \psi_n(r_{n+1}) \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \vdots \\ \mathcal{R}_{n+1} \end{bmatrix}. \tag{21}$$

Because $\mathcal{R}_i = 0$ for $i = 1, \ldots, n$ by (19) and since

$$\mathcal{R}_{n+1} \neq \mathcal{R}_{\partial\mathcal{D}},$$

solution of $\mathbf{b} = \mathbf{Q}^{-1}\mathbf{R}$ implies $b_i = 0$ for $i = 0, \ldots, n-1$ (since $\hat{\mathbf{Q}}$ is full rank), b_n can be nonzero, and so the residual over $\mathcal{D}$ is simply $\mathcal{R} = b_n\psi_n$. Because the collocation points were chosen as the zeros of ψ_n, the residual evaluated at the collocation points will be zero, but can be nonzero in between the collocation points.

5 The Low-Order Model

Having defined the discretization matrices, it is simple to discretize the spatial derivatives of the original modeling partial differential equation (1) and its boundary condition (2):

$$\begin{aligned} \frac{dT_k}{dt} &= \kappa \sum_{j=1}^{n+1} B_{k,j} T_j + \epsilon(1 - T_k^4) + \mathbf{q}(r_k)^T \cdot \mathbf{u} \qquad k = 1, \ldots, n \\ 0 &= \sum_{j=1}^{n+1} A_{n+1,j} T_j - \epsilon_{\text{ed}}(1 - T_{n+1}^4) - q_{\text{ed}} u_b \end{aligned}$$

using a *total* of $n+1$ collocation points. The two polysilicon deposition thickness ordinary differential equations (5) are now expanded to the $2(n+1)$:

$$\frac{dS_{u_k}}{dt} = R_c(X_H, X_S, T_k) \quad \text{and} \quad \frac{dS_{l_k}}{dt} = R_c(Y_H, Y_S, T_k) \qquad \text{for} \quad k = 1, \ldots, n+1.$$

The gas phase silane and hydrogen balances become

$$\begin{aligned}
\frac{dX_H}{dt} &= X_{H_{\text{feed}}} - X_H + 2\beta \sum_{i=1}^{n+1} w_i R_c(X_H, X_S, T_i) \quad \text{with} \quad X_{H_{\text{feed}}} = 0 \\
\frac{dX_S}{dt} &= X_{S_{\text{feed}}} - X_S - \beta \sum_{i=1}^{n+1} w_i R_c(X_H, X_S, T_i) \quad \text{with} \quad X_{S_{\text{feed}}} = 0.1 \\
\frac{dY_H}{dt} &= \frac{\tau}{\tau_l}(X_H - Y_H) + 2\beta \frac{\tau}{\tau_l} \sum_{i=1}^{n+1} w_i R_c(Y_H, Y_S, T_i) \\
\frac{dY_S}{dt} &= \frac{\tau}{\tau_l}(X_S - Y_S) - \beta \frac{\tau}{\tau_l} \sum_{i=1}^{n+1} w_i R_c(Y_H, Y_S, T_i)
\end{aligned} \tag{22}$$

giving a total of $3(n+1)+4$ ordinary differential equations in time. The discretization arrays, quadrature weights, and collocation point locations based on the empirically-determined eigenfunctions were found to be

$$\mathbf{A} = \begin{bmatrix} -3.4858 & 6.1358 & -7.2326 & 4.5826 \\ -0.9135 & -2.8543 & 8.5450 & -4.7771 \\ 0.2650 & -2.2301 & -11.5675 & 13.5325 \\ -0.3777 & 2.8255 & -29.8354 & 27.3875 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} -12.3111 & 15.7093 & -6.1059 & 2.7077 \\ 9.7005 & -34.5634 & 39.1722 & -14.3093 \\ -6.4731 & 66.2134 & -298.7683 & 239.0280 \\ -17.8266 & 127.7846 & -449.5482 & 339.5902 \end{bmatrix},$$

$$\mathbf{w} = [0.2807 \;\; 0.4560 \;\; 0.2600 \;\; 0.0033],$$

$$r_k = \{0.34,\ 0.72,\ 0.95,\ 1.00\}.$$

These values should be compared with those obtained in [15] using Jacobi polynomials.

5.1 Reduced-Order Simulator Results

We demonstrate the predictive capabilities of the reduced order model by comparing its predictions with those produced by the detailed simulator, both runs corresponding to operating conditions different from those used to generate the original snapshots. The processing recipe begins with the wafer at ambient temperature and the reaction chamber filled with argon. At the start of the run, the flow of 10% silane was initiated and lamp bank A was ramped up to 90%, bank B to 50%, and bank C to 80% over five seconds. These power settings were held for 40 seconds, and then ramped linearly down to zero over five seconds. The wafer was removed from the reaction chamber after a total of 60 seconds.

The wafer temperature dynamics generated by the detailed simulator and measured at the collocation point locations is compared to the temperature traces predicted directly from the reduced order model in Fig. 7. The gas phase reacting species concentration time profiles are

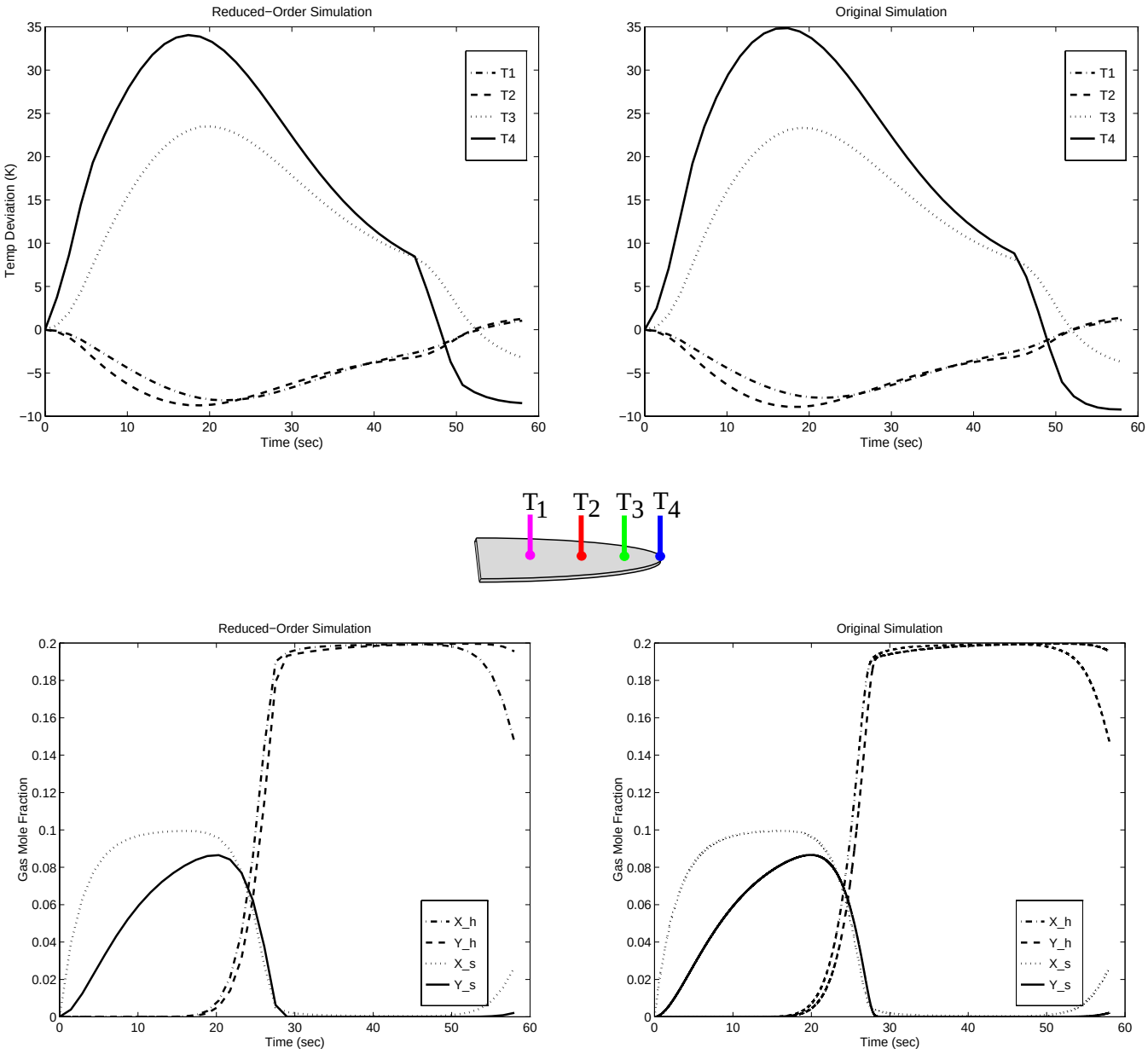


Figure 7: Reduced versus high-fidelity simulator predictions. Top figures compare temperature at collocation points; bottom figures compares gas phase hydrogen and silane mole fractions.

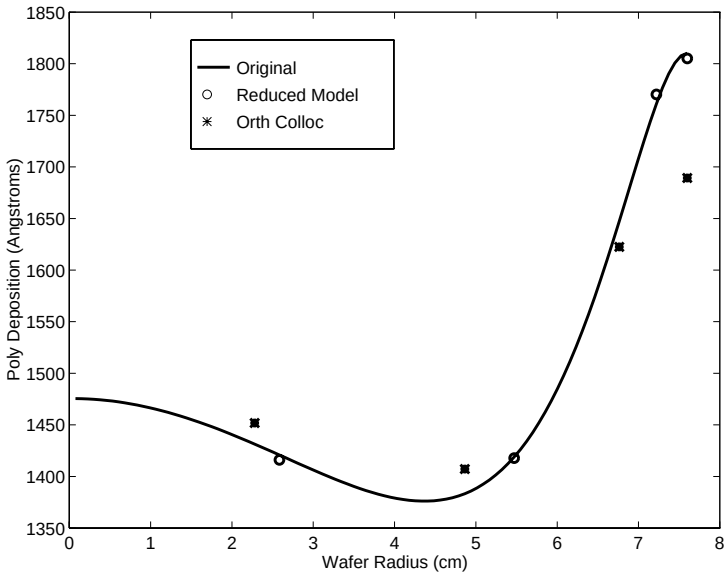


Figure 8: Comparison of polysilicon deposition prediced by reduced-order models based on theoretically versus empirically determined trial functions.

also shown, illustrating that both spatially-dependent and lumped-parameter-type states are accurately predicted with the reduced model. There is virtually no discernible error in the reduced simulations – in this case, the reduced order model used approximately 5% of the computing resources required by the high-fidelity simulator.

A comparison between the polysilicon deposition thickness profile generated by the high-fidelity simulator with predictions made by two reduced-order models is shown in Fig. 8. We see that the deposition thickness predictions of the reduced model based on empirical eigenfunctions is far more accurate than predictions produced when theoretical trial functions (the Jacobi polynomials) are used to discretize the modeling PDE. The accuracy of the empirical eigenfunction approach is almost surprising, given the differences in deposition thickness magnitude between the runs used to generate snapshots (Fig. 5) and this run.

6 Discussion

The focus of this paper centered on developing numerical techniques for computing discretization arrays from spatially-resolved data taken from detailed RTP dynamic simulations. In exploring the interdependence of detailed simulation and the end use of the reduced order models developed from the simulations, a number of issues surfaced and are currently under investigation.

The first is to explore the obvious connections 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 ψ_i as a "smarter" way of interpolating the state of the process – in other words, to recover the spatial temperature or reactant concentration profiles from the minimum number of measurements, or computing accurate values for mean quantities.

The present method for computing collocation point locations is based on driving a residual to zero at the n interior collocation points, forcing the residual to span the $n+1$th and higher eigenmodes. The demand for more accurate interpolative capabilities leads naturally to research

on improving the accuracy of predictions for regions between the collocation points. Theodoropoulou [20] found that adding a collocation point at the wafer center reduces the residual over the wafer domain. Concepts borrowed from computing approximate inertial manifolds [21], where modes associated with the nonzero portion of the residual are slaved to the dominant modes, might increase the accuracy of our collocation procedure.

In this study, the transport equations used to generate the reduced model were the same as those found in the high-fidelity simulator. Current work focuses on quantifying the error incurred when simplified descriptions of the equations of energy and mass transport (the box labeled "Model" in Fig. 1) are used to determine the nonlinear dynamics of the modes extracted from the detailed simulations. We see this as a necessary step in the development of the model reduction methodology, since this would facilitate using extremely complex simulators and experimental data as input to the model reduction routines.

Finally, computing the adjoint eigenfunctions [11] necessary for designing controllers fits neatly into the model reduction framework presented in this paper. Development of the numerical techniques is underway.

7 Acknowledgments

This research was supported by the NSF Engineering Research Centers Program under Grant NSF EEC 94-02384. Development of the model reduction methodology was partially supported by AFOSR grant F49620-93-1-0186.

References

[1] Cho, B.-J., P. Vandenabeele, and K. Maex 1994, Development of a hexagonal-shaped rapid thermal processor using a vertical tube. *IEEE Trans. Semicond. Manuf.* **7** 345-353.

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

[3] Gyurcsik, R. S., T. J. Riley, and F. Y. Sorrell 1991, A model for rapid thermal processing: Achieving uniformity through lamp control. *IEEE Trans. Semicond. Manuf.* **4** 9-13.

[4] Sorrell, F. Y., M. J. Fordham, M. C. Ozturk, and J. J. Wortman 1992, Temperature uniformity in RTP furnaces. *IEEE Trans. Electron Dev.* **39** 75-79.

[5] Breedijik, T., T. F. Edgar, and I. Trachtenberg 1993, A model predictive controller for multivariable temperature control in rapid thermal processing. *Proc 1993 ACC* 2980-2984.

[6] Lie, K.-H., T. P. Merchant, and K. F. Jensen 1993, Simulation of rapid thermal processing equipment and processes. *Mat. Res. Soc. Symp. Proc.* **303** 197-209.

[7] Chatterjee, S., I. Trachtenberg, and T. F. Edgar 1992, Mathematical modeling of a single-wafer rapid thermal reactor. J. Electrochem. Soc. **139** 3682-3689.

[8] Gyugyi, P. J., Y. M. Cho, G. Franklin, and T. Kailath 1993, Convex optimization of wafer temperature trajectories for rapid thermal processing. *Preprint.*

[9] Gay, D. H. and W. H. Ray 1995, Identification and control of distributed parameter systems by means of the singular value decomposition. *Chem. Engng Sci.*, **50**, 1519-1539.

[10] Deane, A. E., I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag 1991, Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. *Phys. Fluids A* **10** 2237-2354.

[11] Chen, C.-C. and H.-C. Chang 1992, Accelerated disturbance damping of an unknown distributed system by nonlinear feedback. *AIChE J.* **38** 1461-1476.

[12] Roenigk, K. F. and K. F. Jensen 1985, Analysis of multicomponent LPCVD processes. *J. Electrochem. Soc.* **132**, Vol 2, 448-454.

[13] Sirovich, L. 1987, Turbulence and the dynamics of coherent structures, Pt. I-III. Quart. Appl. Math. **XLV**, No. 3, 561-590.

[14] Sirovich, L. 1991, Empirical eigenfunctions and low dimensional systems. In *New Perspectives in Turbulence*, L. Sirovich, Ed., Springer-Verlag, 139-163.

[15] Villadsen, J. V. and W. E. Stewart 1967, Solution of boundary-value problems by orthogonal collocation. *Chem. Engng Sci.* **22**, 1483-1501.

[16] Gottlieb, D. and S. A. Orszag 1977, *Numerical Analysis of Spectral Methods: Theory and Applications*, SIAM CBMS-NSF #26.

[17] Sorensen, J. P., and W. E. Stewart 1982, *Collocation Analysis of Multidiffusion and Reactions in Porous Catalysis* U. of Wisc. MRC Report #2341.

[18] Constantin, P., C. Foias, O. P. Manley, and R. Temam 1985, Determining modes and fractal dimension of turbulent flows. *J. Fluid Mech.* **150**, 427-440.

[19] Hildebrand, F. B. 1965, *Methods of Applied Mathematics* Prentice-Hall, New Jersey. pp. 36,122-123.

[20] Theodoropoulou, A.-G. 1995 *Model Reduction for Control of Rapid Thermal Processing*, Ph.D. thesis proposal, University of Maryland.

[21] Sirovich, L., B. W. Knight, and J. D. Rodriguez 1990, Optimal low-dimensional dynamical approximations. *Quart. Appl. Math.* **XLVIII**, No. 3, 535-548.

Appendix

The thermo-physical constants for silicon, parameters associated with the wafer radiative heat transfer, and reaction rate constants (taken from Roenigk and Jensen [12]) are collected in Table 9. Constants corresponding to the 3-zone RTP system geometry are also included in this table. The residence time constants τ and τ_l are based on a total feed flow rate of 300 sccm. The dimensionless states and coordinates (r, t) are defined in Table 10. Since the dimensionless wafer radius is 1, the surface area of one wafer side $\int_{\mathcal{D}} dA = \pi$.

The constants of the wafer thermal dynamics PDE (1) and its boundary condition (2) are defined as

$$\kappa = \frac{\tau k}{\rho C_p W_{\text{ref}}^2} = 0.0021 \qquad \epsilon = \frac{\tau \sigma E_w T_{\text{amb}}^3}{\rho C_p \delta z} = 0.0012$$

and

$$\epsilon_{\text{ed}} = \frac{\sigma E_w W_{\text{ref}} T_{\text{amb}}^3}{k} = 0.0037 \qquad q_{\text{ed}} = \frac{q_{b_{\text{edge}}} E_w W_{\text{ref}}}{k T_{\text{amb}}} = 2.011$$

$$\omega = \frac{\tau E_w}{\rho C_p \delta z T_{\text{amb}}} = 0.0256 \text{W/cm}^2.$$

The numerical values are computed with the constants given in Table 9. The radiation flux values at the interior collocation points from each lamp bank (used to compute elements of $\mathbf{q}(r_k)$ in W/cm^2) are

$$\begin{aligned} q_a(r_k) &= [7.58 \quad 6.60 \quad 5.82] \\ q_b(r_k) &= [1.51 \quad 2.62 \quad 5.16] \\ q_c(r_k) &= [8.38 \quad 8.61 \quad 8.64] \end{aligned}$$

Parameter	*Value*	*Description*
k	0.22 W/(cm K)	Si thermal conductivity
ρ	2.3 gm/cm^3	Si density
C_p	2.3 J/(gm K)	Si heat capacity
M_w	28.1 gm/gmol	Si molecular weight
E_w	0.7	Wafer emissivity
σ	5.677×10^{-12} J/(s cm^2 K^4)	Boltzmann constant
$T_{\rm amb}$	300 K	Reference temperature
δz	0.05 cm	Wafer thickness
$W_{\rm ref}$	7.6 cm	Wafer radius
K_0	1.6×10^9 mol/(s atm m^2)	Preexponential constant
K_S	0.7×10^5 atm^{-1}	SiH_4 adsorption constant
K_H	0.6×10^2 atm$^{-1/2}$	H_2 desorption constant
E_A	18500 K	Activation energy
P_T	5 torr	Total pressure
τ	2.9 s	Residence time, upper chamber
τ_l	7.47 s	Residence time, lower chamber
$S_{\rm ref}$	1×10^{-6} cm	Reference deposition thickness
$W_{\rm ref}$	7.6 cm	Wafer radius

Figure 9: Table of constants.

The reaction kinetics parameters are

$$\begin{aligned} k_0 &= \frac{K_0 P_T M_w \tau}{\rho S_{ref}} \left(\frac{1 \text{ atm m}^2}{7.6\times10^6 \text{ cm}^2 \text{ torr}} \right) = 7.455\times10^8 \\ b &= K_S P_T \left(\frac{1 \text{ atm}}{760 \text{ torr}} \right) = 460.5 \\ c &= K_H \sqrt{P_T} \left(\frac{1 \text{ atm}}{760 \text{ torr}} \right)^{1/2} = 4.867 \\ \gamma &= E_A / T_{\rm amb} = 61.67 \end{aligned}$$

The conversion factor in the gas phase equations (3) relating polysilicon deposition rate R_c to the rate of hydrogen production and silane consumption is the ratio of molar density of silicon $\times$ a wafer-scale volume to the gas phase molar density $\times$ the upper chamber volume:

$$\beta = \frac{\pi \rho W_{\rm ref}^2 S_{\rm ref}/M_w}{(P_T/RT_{\rm amb})V_u} = 1.260$$

where R is the gas constant and V_u is the upper chamber volume.

State	*Description*	*Units*	*Dimensionless Definition*
$P^u_{\mathrm{H}_2}$	H_2 partial press, upper chamber	torr	$X_H = P^u_{\mathrm{H}_2}/P_T$
$P^l_{\mathrm{H}_2}$	H_2 partial press, lower chamber	torr	$Y_H = P^l_{\mathrm{H}_2}/P_T$
$P^u_{\mathrm{SIH}_4}$	SiH_4 partial press, upper chamber	torr	$X_S = P^u_{\mathrm{SiH}_4}/P_T$
$P^l_{\mathrm{SIH}_4}$	SiH_4 partial press, lower chamber	torr	$Y_S = P^l_{\mathrm{SiH}_4}/P_T$
$T'(r', t')$	Wafer temperature	K	$T = T'/T_{\mathrm{amb}}$
$S'_u(r', t')$	Poly film thickness, upper surface	cm	$S_u = S'_u/S_{\mathrm{ref}}$
$S'_l(r', t')$	Poly film thickness, lower surface	cm	$S_l = S'_l/S_{\mathrm{ref}}$
t'	Time	sec	$t = t'/\tau$
r'	Radial position	cm	$r = r'/W_{\mathrm{ref}}$

Figure 10: Table of dimensionless states.