

TECHNICAL RESEARCH REPORT

Model Reduction via the Karhunen-Loeve Expansion Part II: Some Elementary Examples

by Andrew J. Newman

T.R. 96-33



ISR develops, applies and teaches advanced methodologies of design and analysis to solve complex, hierarchical, heterogeneous and dynamic problems of engineering technology and systems for industry and government.

ISR is a permanent institute of the University of Maryland, within the Glenn L. Martin Institute of Technology/A. James Clark School of Engineering. It is a National Science Foundation Engineering Research Center.

Web site <http://www.isr.umd.edu>

Model Reduction via the Karhunen-Loeve Expansion

Part II: Some Elementary Examples

Andrew J. Newman *

Institute for Systems Research and
Electrical Engineering Department
University of Maryland
College Park, MD 20742
newman@isr.umd.edu

April 2, 1996

Abstract

In Part II of this paper we present some elementary examples of distributed parameter system model reduction using the techniques described in Part I. In particular, we focus on using the Karhunen-Loeve expansion and Galerkin's method to formulate reduced order models for a heat diffusion system and temperature field dynamics in a Rapid Thermal Chemical Vapor Deposition reactor. Simulation results are presented.

1 Introduction

In Part I of this paper [11], we provided a detailed exposition of some techniques that are useful in finding models of reduced complexity for dynamical systems involving flows. Here, in Part II, we put these techniques to work and present some examples of their use. We focus on distributed parameter systems related to semiconductor manufacturing, an area of much recent interest. For purposes of clarity and completeness, Part II contains the main ideas of the background material necessary to proceed to the examples without the reader having to review Part I.

The manufacturing of semiconductor devices requires the use of ever more complex processes and systems. Quite often, these systems involve the interaction of fluid flows, thermodynamic effects, chemical reactions, and various other physical phenomena. It is apparent that mathematical models describing such processes will be highly complicated, consequently causing difficulties in analysis, simulation, and design. For these systems, models of reduced complexity which retain qualitative correctness are very attractive, especially when such a model provides additional insight which helps us to understand, improve, and control the process.

This paper describes some important ideas and methods in simulation and modeling of distributed parameter dynamical systems, i.e. systems whose dynamics are governed by partial differential equation (PDE) models, which are common in semiconductor manufacturing. The purpose, however, is not to provide an introduction to the variety of such systems, models, and methods. Instead, we wish to simply outline a general methodology for implementation of one or two techniques, and apply these methods to a simple, linear system. In this manner, the main ideas are not obscured by the technical details of a highly complex system, while still providing enough information so that one can see the extension to more complicated

*This research was supported in part by a grant from the National Science Foundation's Engineering Research Centers Program: NSFD CDR 8803012 and by the NSF Grant EEC-9527576.

problems. Also, we can easily test the sensitivity of the methods to various parameters. Then we show how these ideas might be used in the modeling and simulation of rapid thermal chemical vapor deposition (RTCVD), a process of much recent interest in the area of semiconductor manufacturing.

Simulation of Distributed Parameter Systems

We have stated the desire to “simulate” systems governed by PDEs, in particular those involved with the manufacturing of semiconductor devices. The term “simulate” is somewhat ambiguous, but it usually means that one is trying to approximate some actual condition or phenomenon without conducting an experiment on the actual system. For example, NASA might simulate zero gravity conditions (a physical chamber capable of producing a good approximation to the desired conditions), or a demographer might simulate population growth (a mathematical model capable of providing a good approximation to the actual phenomenon). In the context of this paper, by simulation we mean the numerical integration of the governing equations to provide a good approximation to the actual solution, where the actual solution is what we would see if we performed an experiment on the given system.

There is a variety of available methods for simulation of the distributed parameter systems in which we are interested. For example, to simulate fluid flows and heat transfer in an RTCVD chamber, one might use a sophisticated computational fluid dynamics (CFD) software package with a built-in PDE solver to solve the equations of interest while allowing the user to input various problem geometries, physical constants, boundary conditions, and other parameters. In Section 4, we do just this, employing Fluent CFD software to simulate heat transfer in a particular RTCVD reactor. On the other hand, one might hard-code the numerical integrations, either using a specialized mathematics programming language such as MATLAB or Mathematica, or a lower level programming language such as C. The choice of methodology depends on the complexity of the system we wish to simulate. For the above RTCVD heat transfer problem, the equations and geometry of the problem are sufficiently complex to justify the additional overhead of using the powerful software package. By additional overhead, we not only refer to the monetary expense of purchasing such a package, but also in training, and the required preprocessing and postprocessing of data. By postprocessing of data, we are referring to the fact that the simulation results are output in a format which may not be amenable to further analysis and computations, a situation which must be rectified by creation of additional programs.

In Section 2, we describe simulations of heat diffusion on simple geometries, a system described by a linear, homogeneous PDE with simple BCs and ICs. Thus, the numerical integrations can be easily hard-coded using MATLAB, with output data immediately ready for further analysis and computations.

Model Formulation and Reduction

The model formulation and reduction techniques that we focus on fall in the general category of spectral methods. Roughly speaking, spectral methods involve seeking or representing the solution to a problem as a truncated series of known, smooth functions of the independent variables. In [6], Gottlieb and Orszag provide a survey of such methods and their applications. In [3], Canuto and et.al. discuss computational issues and focus on applications in fluid dynamics. In this paper, we employ the Galerkin method of PDE discretization, where by discretization we mean the transformation of the original infinite dimensional problem into a finite dimensional problem. More details on the theory and applications of this method can be found in [7, 9]. The basis of the finite dimensional function space of interest is computed using the proper orthogonal decomposition (POD) (also known as principal components analysis), which is a technique that uses empirical data to extract a basis for a modal decomposition of the system dynamics. This decomposition has several attractive features, and provides a basis for models of reduced complexity. In this paper, the empirical data comes from simulations, although experimental data could also be used if available. Detailed descriptions of the POD and related results can be found in [2, 13].

Organization of this Paper

In Section 2, we describe simulations of heat diffusion on a thin rod (one-dimensional spatial domain) and on a square plate (two-dimensional spatial domain). In Section 3, we use the empirical data provided by the simulations to find models of reduced complexity via computation of the POD and Galerkin discretization. In addition, we present results quantifying the correctness and sensitivity to parameters of the models developed

using these methods, and make comparisons with alternative procedures. Section 4 discusses applications of these ideas to a particular problem in RTCVD and related issues in sensing and control.

2 Heat Equation Simulation

As stated earlier, the proper orthogonal decomposition requires empirical data for determination of basis functions. We provide empirical data via simulation of the heat diffusion equation. In this section, we discuss the methodology used for the simulations and present the simulation results. The following IBVP was selected for the one-dimensional (1D) simulations:

$$\begin{array}{lll} \text{PDE} & u_t = u_{xx} & x \in (0, 1); \quad t > 0 \\ \text{BCs} & u(0, t) = 0 = u(1, t) & t > 0 \\ \text{IC} & u(x, 0) = 1 & x \in (0, 1) \end{array} \quad (1)$$

where $u(x, t)$ represents the temperature field on a thin rod. Similarly, the following IBVP was selected for the two-dimensional (2D) simulations:

$$\begin{array}{lll} \text{PDE} & u_t = u_{xx} + u_{yy} & x \in (0, 1); \quad y \in (0, 1); \quad t > 0 \\ \text{BCs} & u(0, y, t) = 0 = u(1, y, t) & t > 0 \\ & u(x, 0, t) = 0 = u(x, 1, t) & t > 0 \\ \text{IC} & u(x, y, 0) = 1 & x \in (0, 1); \quad y \in (0, 1) \end{array} \quad (2)$$

where $u(x, y, t)$ represents the temperature field on a flat plate.

2.1 Methodology

In order to generate simulation data for the heat diffusion systems, we numerically integrated the IBVPs (1) and (2). An easy and convenient method of doing this is to evaluate the infinite series solutions to the respective IBVPs at a predetermined set of spatial points and temporal values. These computations can easily be hard-coded using MATLAB or another mathematics software package.

The infinite series solution to IBVP (1) is given by

$$u(x, t) = \sum_{n=1}^{\infty} A_n e^{-n^2 \pi^2 t} \sin(n\pi x) \quad (3)$$

where $A_n = \frac{2}{n\pi}(1 - \cos(n\pi))$.

To simulate IBVP (1), the spatial domain was discretized into segments of length 0.01. Then, solution (3) was evaluated at each x value in the discretization for a predetermined set of values for t . The solution evaluation required carrying out the summation to a sufficient number of terms.

The infinite series solution to IBVP 2 is given by

$$u(x, y, t) = \sum_{m,n=1}^{\infty} A_{mn} e^{-(m^2+n^2)\pi^2 t} \sin(m\pi x) \sin(n\pi y) \quad (4)$$

where $A_{mn} = \frac{4}{mn\pi^2}(1 - \cos(m\pi))(1 - \cos(n\pi))$.

To simulate IBVP (2), the spatial domain was discretized into a 100 X 100 grid. Then, solution (4) was evaluated at each (x, y) value in the discretization for a predetermined set of values for t . Again, the solution evaluation required carrying out the summation to a sufficient number of terms.

In addition to the infinite series solution, we also used an explicit finite differences scheme to numerically integrate the above PDEs for purposes of comparison. Other possible methods exist, such as a numerical evaluation of the convolution of the IC with the Gaussian kernel. We note here that the finite differences simulation produced results identical to those of the infinite series solution. These alternate methods will be useful for future work on more complicated systems.

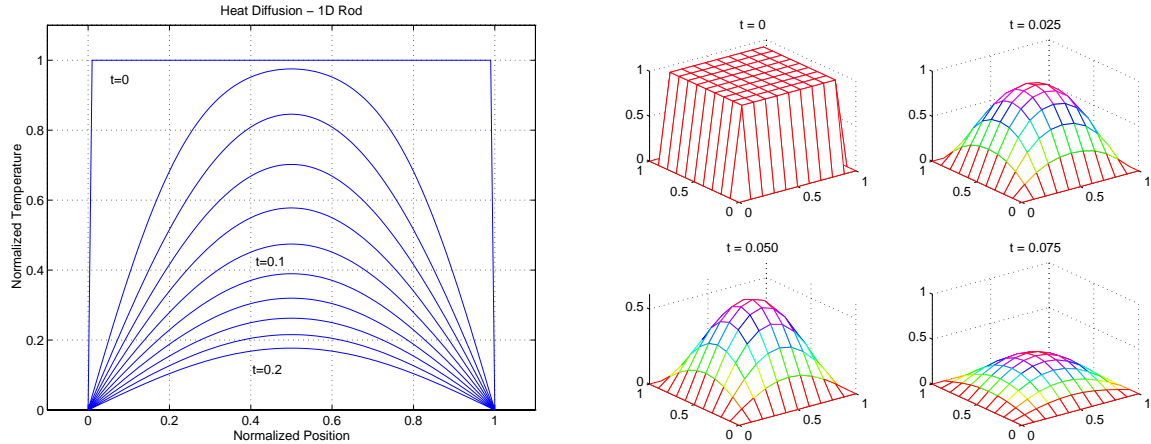


Figure 1: Left: Simulation of IBVP (1): time dependent heat diffusion on 1D rod with constant initial condition and zero boundary conditions. Right: Simulation of IBVP (2): time dependent heat diffusion on 2D plate with constant initial condition and zero boundary conditions.

2.2 Results

The solution for IBVP (1) was evaluated at each value of time in the set $\{0.00, 0.001, 0.002, \dots, 0.200\}$. The solution for IBVP (2) was evaluated at each value of time in the set $\{0.00, 0.05, 0.10, \dots, 0.45, 0.50\}$. Several of the resulting temperature distributions are shown in Figure 1. The data was stored for use as empirical data for the POD.

3 Model Reduction

This section contains a brief overview of the Galerkin method of PDE discretization, the proper orthogonal decomposition, and their use in model reduction. We present the methodology used to implement and apply these techniques and the results of their application to the heat diffusion IBVPs. Details and further applications of these ideas can be found in [1, 2, 3, 9, 13].

3.1 The Galerkin Method

The Galerkin method is a discretization scheme for PDEs which is generically categorized as one of the spectral methods or methods of weighted residuals. These methods are based on the separation of variables approach and attempt to find an approximate solution in the form of a truncated series expansion given by

$$\hat{u}(x, t) = \sum_{n=1}^N a_n(t) \phi_n(x) \quad (5)$$

where the $\phi_n(x)$ are known as trial functions. Thus, the original infinite-dimensional system is approximated by an N -dimensional system where the order of the reduced model is determined by the truncation point. The trial functions form an orthonormal basis for the appropriate function space and are chosen to satisfy desired properties for the given problem. For example, they might be eigenfunctions of the spatial operator associated with the given PDEs or basis functions generated by the POD. The $a_n(t)$ are time-dependent coefficients which are chosen to ensure that the original PDE is satisfied as closely as possible by (5). This is achieved by minimizing the *residual*, i.e. the error produced by using (5) instead of the exact solution, with respect to a suitable norm. Equivalently, the residual must be orthogonal to each of the given trial functions.

To state this mathematically, suppose we have a system governed by the PDEs (in symbolic form)

$$\frac{\partial u}{\partial t} = D(u); \quad u : \mathcal{D} \times (0, \infty) \rightarrow \mathbb{R} \quad (6)$$

with appropriate boundary conditions and initial conditions, where $D(\cdot)$ is a spatial operator, e.g. the Laplacian in the case of heat diffusion. Define the residual as

$$r(x, t) = \frac{\partial \hat{u}}{\partial t} - D(\hat{u}). \quad (7)$$

We force the residual to be orthogonal to a suitable number of trial functions, i.e.

$$\langle r(x, t), \phi_n(x) \rangle = 0 \quad n = 1, \dots, N. \quad (8)$$

It can be shown that the projection of the residual onto the subspace spanned by each trial function as in (8) represents the energy of the residual in the direction in function space associated with each trial function, respectively. Moreover, if the trial functions are determined using the POD, the optimality property of the Karhunen–Loeve expansion applies, and the energy in the residual vanishes in those directions for which the energy in the flow is extremized.

Substituting (5) into (7) yields,

$$r(x, t) = \sum_{n=1}^N \dot{a}_n(t) \phi_n(x) - D \left(\sum_{m=1}^N a_m(t) \phi_m(x) \right). \quad (9)$$

Applying the orthogonality condition (8) and using the orthonormality property of the set of trial functions results in

$$\dot{a}_i(t) = \int_{\mathcal{D}} D \left(\sum_{m=1}^N a_m(t) \phi_m(x) \right) \phi_i(x) dx \quad i = 1, \dots, N. \quad (10)$$

Thus, insisting that the residual be orthogonal to the first N trial functions yields a system of N ordinary differential equations in t . The reduced order model is a system of N ordinary differential equations (an N^{th} -order system)

$$\dot{a} = F(a) \quad (11)$$

where $a = (a_1, \dots, a_N)$ and $F : \mathbb{R}^N \rightarrow \mathbb{R}^N$. By simulating or solving the reduced order model (11) and substituting back into (5) we get an approximate solution for $u(x, t)$ without solving or simulating the infinite dimensional system given by the original PDEs. We can choose N , the order of the approximation, appropriately for our purposes.

The initial conditions for the resulting system of ODEs are determined by a second application of the Galerkin approach. We force the residual $I(x) = u(x, 0) - \hat{u}(x, 0)$ of the initial conditions to also be orthogonal to the first N basis functions. We obtain a system of N linear equations

$$a_i(0) = \int_{\mathcal{D}} u(x, 0) \phi_i(x) dx \quad i = 1 \dots N. \quad (12)$$

Observe that solution of the system of ODEs (11) and (12) requires only the selection of the set of trial functions $\{\phi_n : n = 1, \dots, N\}$ and the initial condition on the original system $u(x, 0)$. Any complete set of trial functions will suffice. However, we focus on those generated by the POD.

3.2 The Proper Orthogonal Decomposition

The proper orthogonal decomposition is a procedure for extracting a basis for a modal decomposition from an ensemble of signals. The applications of this procedure are extensive, e.g. modeling of turbulence [13, 14, 15], rapid thermal processing [1], and image processing [12].

The mathematical theory behind the POD is the spectral theory of compact, self-adjoint operators. The main theorem employed is the Karhunen–Loeve decomposition theorem (see [17]). The theorem roughly states that for a stochastic process, the eigenfunctions of the integral operator whose kernel is the process correlation function form an orthogonal basis for the function space in which the process resides. Hence, under some technical assumptions, we can model a dynamical distributed parameter system as a stochastic process, compute the process correlation function, and use the eigenfunctions of the integral operator as a basis for the function space of interest.

There are several attractions of this procedure. It only requires linear operations, although it makes no assumptions regarding the linearity of the problem to which it is applied. The basis functions, also referred to as *coherent structures*, can be described as organized spatial features which repeatedly appear and undergo a temporal life cycle. Hence, the POD reveals, in some sense, the underlying structure of the system dynamics. Most importantly, from several points of view, the POD is optimal for modeling or reconstructing a signal. For instance, in [2] Berkooz, Holmes, and Lumley present one optimality result which implies that among all linear decompositions, the POD is most efficient from a data compression point of view. I.e., for a given number of modes, the projection on the subspace used for modeling will contain the most “energy” possible in an average sense. This energy in a given mode can be measured in a relative sense by the magnitude of the eigenvalue corresponding to that mode.

We shall employ the POD as one way of computing the trial functions for the expansion (5) and attempt to demonstrate its optimality properties.

3.3 Method of Snapshots

From a practical point of view, difficulties appear in the determination of the eigenfunctions using the POD. For computational purposes, we must discretize the spatial domain, usually leading to a very large spatial correlation matrix. One approach to solving this problem is to sample the empirical data at equally spaced time intervals and use a suitable number of these samples in an appropriate manner to perform the required computations. Sirovich refers to this approach as the *method of snapshots* [13].

Essentially, this method provides a good approximation of the basis functions found using the Karhunen–Loeve expansion theorem. The problem is reduced to finding eigenvectors and eigenvalues of the $M \times M$ matrix C whose entries are given by

$$(C)_{ij} = \frac{1}{M} \int_{\mathcal{D}} v^{(i)}(x)v^{(j)}(x) dx \quad i, j = 1, \dots, M \quad (13)$$

where M is the number of data samples or “snapshots”, \mathcal{D} is the spatial domain (e.g. $[0, 1]$ or $[0, 1] \times [0, 1]$), and $v^{(i)}(x)$ is the i^{th} snapshot. The eigenvectors $A^{(n)}$ of C and the corresponding eigenvalues λ_n satisfy

$$CA^{(n)} = \lambda_n A^{(n)} \quad n = 1, \dots, M \quad (14)$$

which under easily satisfied assumptions can be solved for the corresponding system of M eigenvalues and M eigenvectors. The empirically determined eigenfunctions are then computed as linear combinations of the data snapshots using

$$\phi_n(x) = \sum_{k=1}^M A_k^{(n)} v^{(k)}(x) \quad n = 1, \dots, M. \quad (15)$$

See [11, 13] for details of the underlying theory.

3.4 Implementation Methodology

In this section we describe the methodology used to apply these methods to the heat diffusion system. We find the empirically determined eigenfunctions and use the Galerkin method to find a reduced order model. We then test the sensitivity of the low order model accuracy to the number of data snapshots (parameter

M) used in (13)-(15) for computing the basis functions, and to the number of terms used in the series expansion (5) (parameter N). MATLAB scripts and functions were programmed to perform the necessary computations for the POD and Galerkin approximations.

Computing the Empirically Determined Eigenfunctions

As stated earlier, snapshots of the heat equation solution were recorded at 200 equally spaced sample times between $t = 0.001$ and $t = 0.200$ (the IC was also used as the first snapshot). These snapshots are used as the empirical data for computing a set of basis functions via the POD.

We adjust the empirical data samples so that the mean of the ensemble of snapshots is zero. This is accomplished by computing the “average snapshot” and then subtracting this profile from each member of the ensemble. This is done mainly for reasons of scale, i.e. the deviations from the mean contain the information of interest, but may be small compared with the original signal.

If we denote the set of original snapshots as $\{u^{(k)} : k = 1, 2, \dots, M\}$ then the average snapshot is computed as

$$\bar{u} = \frac{1}{M} \sum_{k=1}^M u^{(k)} \quad (16)$$

and the mean adjusted snapshots are given by

$$v^{(k)} = u^{(k)} - \bar{u}. \quad (17)$$

This adjustment leaves us with a new ensemble of data samples $\{v^{(k)} : k = 1, \dots, M\}$. These snapshots are then used in (13) to compute the $M \times M$ empirical correlation matrix C . The numerical integration is hard-coded using a simple approximation technique. Existing MATLAB functions are used to find the eigenvalues and eigenvectors of C , which are then used in (15) to compute the empirically determined eigenfunctions. The parameter M is varied to test the sensitivity of the resulting approximation to the number of data snapshots used. Observe that using M snapshots results in the computation of M trial functions, although not all of these need be used in the Galerkin approximation (later we examine the effect of this by varying N).

Approximating Heat Diffusion Using Galerkin’s Method

We now transform the original infinite dimensional problem involving the heat diffusion PDE to a finite dimensional problem merely requiring solution of a simple linear N^{th} -order ODE. The Galerkin approximation (5) is performed using two different sets of spatial functions as the basis functions $\{\phi_n(x) : n = 1, \dots, N\}$. We use the empirically determined eigenfunctions from the POD, and for comparison, a set of sinusoidal functions of different frequencies. The sinusoids are chosen because they are the eigenfunctions of the Laplacian operator, i.e. the spatial operator associated with the heat diffusion PDE. In particular, we choose the normalized set

$$\Phi_{sin} = \{\sqrt{2} \sin(n\pi x), \quad n = 1, \dots, N\}. \quad (18)$$

We shall denote the basis functions $\phi_n(x)$ regardless of whether they are the sinusoids or empirically determined eigenfunctions. At this point, the system of ODEs (11) and (12) for the coefficients $a_n(t)$ must be formulated and computed using (10). The heat diffusion system dynamics are described by

$$\frac{\partial(v + \bar{u})}{\partial t} = D(v) = \nabla^2(v + \bar{u}) \quad (19)$$

Applying (10) yields the system of linear ODEs

$$\dot{a}_i(t) = \sum_{j=1}^N a_j(t) \int_{\mathcal{D}} \phi_i(x) \nabla^2 \phi_j(x) dx + \int_{\mathcal{D}} \nabla^2 \bar{u}(x) \phi_i(x) dx \quad i = 1, \dots, N \quad (20)$$

with initial conditions

$$a_i(0) = \int_{\mathcal{D}} \phi_i(x) v(0, x) dx \quad i = 1, \dots, N \quad (21)$$

where $\mathcal{D} = [0, 1]$ for the rod and $\mathcal{D} = [0, 1] \times [0, 1]$ for the plate. This results in the linear system of ODEs

$$\dot{a}(t) = \Gamma a(t) + b \quad (22)$$

where $a(t)$ is an N -vector, Γ is the $N \times N$ matrix with entries

$$(\Gamma)_{ij} = \int_{\mathcal{D}} \phi_i(x) \nabla^2 \phi_j(x) dx \quad (23)$$

and b is an N -vector with elements

$$b_i = \int_{\mathcal{D}} \nabla^2 \bar{u}(x) \phi_i(x) dx. \quad (24)$$

The solution to (22) is given by the variation of constants formula

$$a(t) = e^{\Gamma t} a(0) + \int_0^t e^{\Gamma(t-\tau)} b d\tau \quad (25)$$

where the IC $a(0)$ is an N -vector with entries given by (21). However, rather than hard-code the solution (25) we can numerically integrate (22) using one of the Runge-Kutta methods available in MATLAB.

Observe that since the sinusoids (18) are eigenfunctions of the Laplacian, the matrix Γ in (22) is diagonal when the sinusoids are used as the trial functions, i.e. the system of ODEs is decoupled. This property does not hold in general for other sets of basis functions including the empirically determined eigenfunctions from the POD.

Once the ODE (22) is solved and evaluated at the desired values of t , the Galerkin approximation is given by the truncated series expansion

$$\hat{v}(x, t) = \sum_{n=1}^N a_n(t) \phi_n(x). \quad (26)$$

The average snapshot \bar{u} is then added

$$\hat{u}(x, t) = \hat{v}(x, t) + \bar{u}(x) \quad (27)$$

to reconstruct the original signal. The approximation order N can be varied to achieve the desired degree of accuracy.

3.5 Results

Now, we present the results of the above computations and simulations. A set of trial functions was determined using the POD for the heat diffusion system on 1D and 2D domains for $M = 201$. Figure 2 shows the resulting empirically determined eigenfunctions for the 1D and 2D heat diffusion systems along with the corresponding eigenvalues. As stated earlier, the eigenvalues measure the relative energy of the system dynamics contained in that particular mode (they have been normalized to correspond to a percentage). The figures contain only the first four eigenfunctions, i.e. the four modes that contain virtually all of the energy. Observe that the shape of the spatial structures which characterize the system dynamics are not obvious given only the original data samples.

Now we investigate the effect of changing the size of the data snapshot ensemble used in performing the POD. We have already seen that using $M = 201$ snapshots the first mode contains more than 96 percent of the energy. Figure 3 shows that even for a small ensemble the POD is successful in capturing almost 96 percent of the energy in the first mode.

It should be noted that the behavior of the heat diffusion system is not very rich, i.e. the spatial structure of the snapshots are not complicated. One might expect that for systems with more complicated dynamics the POD will not be as successful in reducing the infinite dimensional system to one of very low

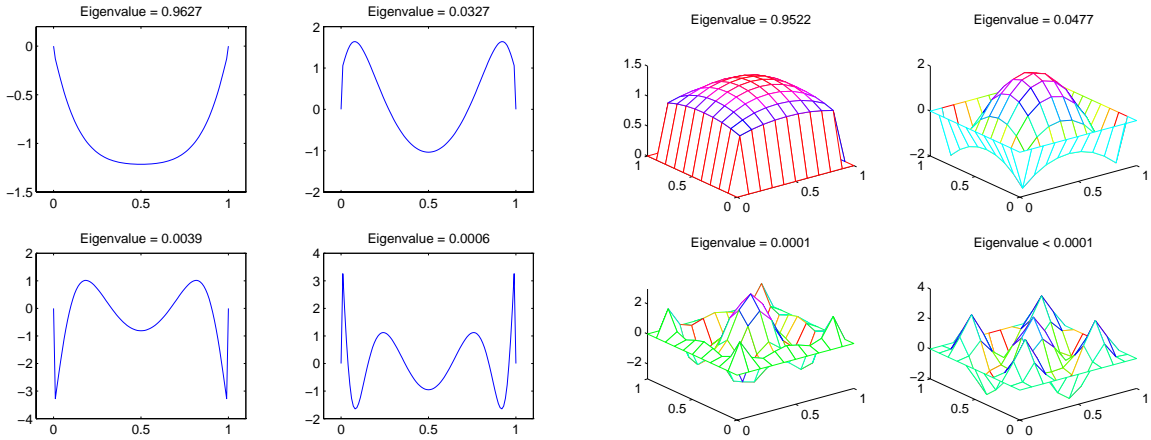


Figure 2: Left: Empirically determined eigenfunctions for IBVP (1) with correspondig eigenvalues (normalized). $M = 201$ snapshots used. Right: Empirically determined eigenfunctions for IBVP (2) with corresponding eigenvalues (normalized). $M = 21$ snapshots used.

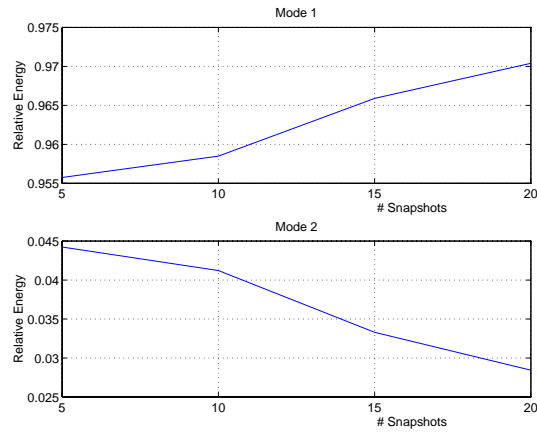


Figure 3: Eigenvalues (relative energies) associated with the first two empirically determined modes versus ensemble size M for approximation of IBVP (1).

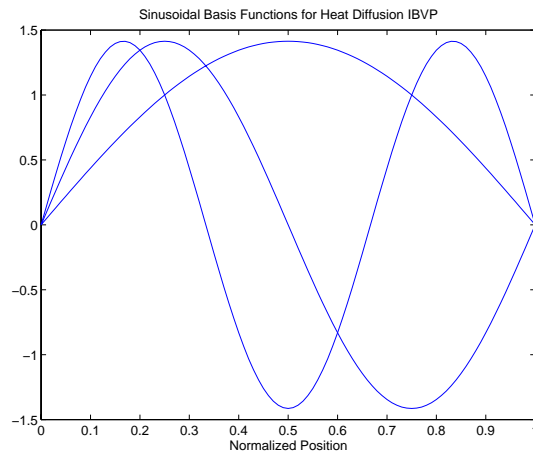


Figure 4: Sinusoidal basis functions for approximation of IBVP (1).

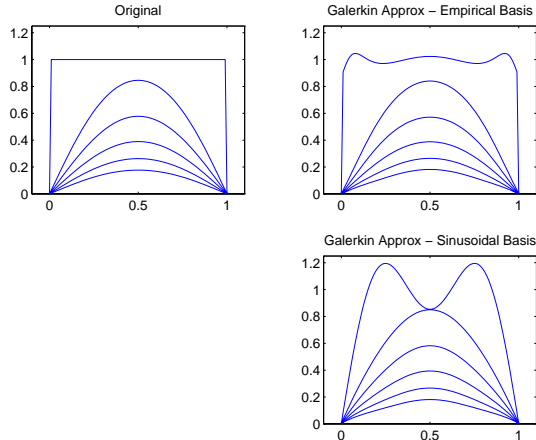


Figure 5: Original heat diffusion data (top, left) and Galerkin approximations of IBVP (1) using empirically determined eigenfunctions (top, right) and sinusoidal basis functions (bottom, right) for $t = 0, 0.04, 0.08, 0.12, 0.16, 0.20$. $N = 3$, $M = 201$.

order. This is not necessarily the case, as we will see later when we apply these methods to an RTCVD system.

As stated earlier, we selected a set of sinusoids as an alternative orthonormal basis for purposes of comparison. The first three members of this set are shown in Figure 4.

The reduced order models (Galerkin approximations) were simulated via numerical integration of the ODE (22). Both empirical eigenfunctions and sinusoids were used as the set of trial functions in (26). Figure 5 shows the original 1D data samples along with the reconstructed data samples from the Galerkin approximations with the approximation order set to $N = 3$, i.e. the series is truncated at 3 terms. We observe that the approximations are accurate, although there is difficulty in reproducing the initial condition. This phenomenon is due to the fact that the solutions progress from a discontinuous (hence nonsmooth) initial condition to smooth profiles at larger values of time. Since our basis functions are smooth the reconstructions of the smoother profiles require fewer terms to get equivalent accuracy. In the language of Fourier theory, the smoother profiles have a smaller spectral content or bandwidth. This can be seen more clearly in Figure 6 which compares the original data samples with the reconstructed data samples for small values of time. We observe that the approximation accuracy increases rapidly with time. We also see here the superiority of using the empirically determined basis which provides a more accurate approximation than the sinusoids. Figure 7 presents the reconstructed snapshots for the 2D domain using empirically determined eigenfunctions. Results are similar to those from the 1D case.

The accuracy of the approximation can be quantified by examining the percent error between the actual and approximate solutions given by $\|u - \hat{u}\|/\|u\|$. We measure this error by computing the maximum absolute pointwise difference between the actual and approximate solution over all times that the solution was computed. With some abuse of terminology we call this a “max”-norm or infinity-norm.

To examine the effect of approximation order on the accuracy, N was varied from 1 to 5, i.e. (26) was evaluated using between 1 and 5 terms. Figure 8 shows the results. Clearly the empirically determined basis is superior to the sinusoidal basis. Using a fifth order approximation the heat diffusion dynamics can be reconstructed with an error that is negligible. Figure 9 shows similar results for the 2D case.

It is reasonable to think that the ensemble of data snapshots used to compute the empirically determined basis will have some effect on the accuracy of the approximations. Since we have seen that the system dynamics for small values of time are more difficult to reconstruct, we now restrict the ensemble to 10 snapshots from $t = 0$ to $t = 0.01$ instead of using all 201 snapshots from $t = 0$ to $t = 0.2$. Empirical eigenfunctions are once again computed using the POD. Figure 10 shows that this restricted set does a slightly better job of capturing the system dynamics and hence providing more accurate reconstructions.

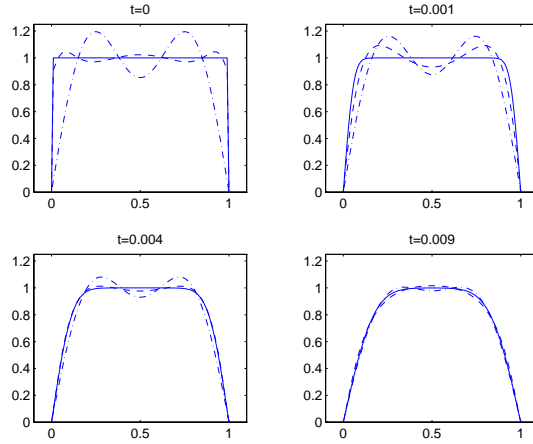


Figure 6: Heat diffusion profiles for $t = 0, 0.001, 0.004, 0.009$. Original data (solid), Galerkin approximation using empirical eigenfunctions (dashed), Galerkin approximation using sinusoidal basis functions (dot-dashed). $N = 3, M = 201$.

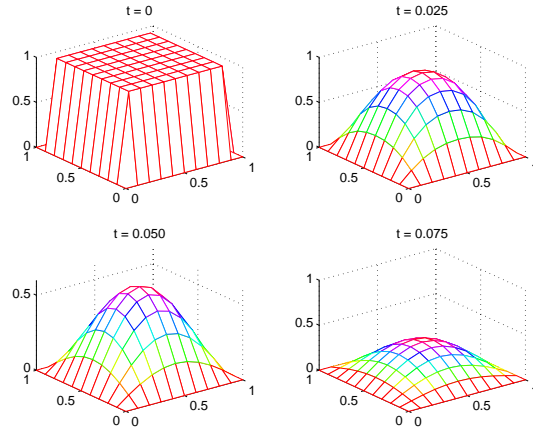


Figure 7: Galerkin approximations of IBVP (2) using empirically determined eigenfunctions for $t = 0, 0.025, 0.050, 0.075$. $N = 3, M = 21$.

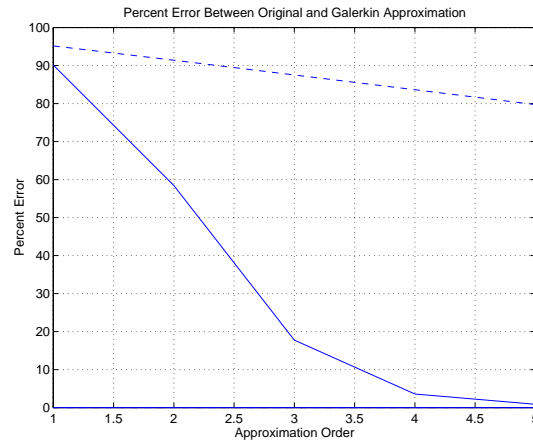


Figure 8: Percent error between original and reconstructed data using infinity-norm versus approximation order N for 1D domain. Solid: empirical basis. Dashed: sinusoidal basis. $M = 201$.

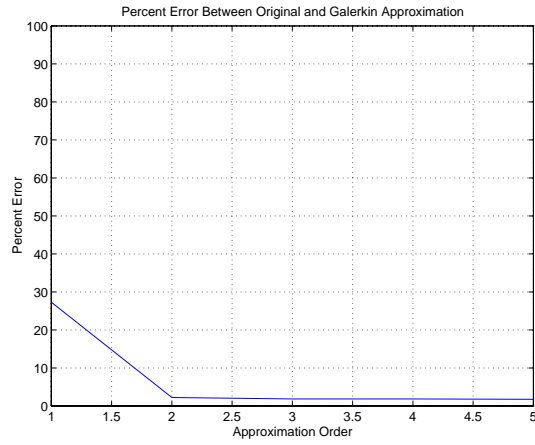


Figure 9: Percent error between original and reconstructed data using infinity–norm versus approximation order N for 2D domain. $M = 21$.

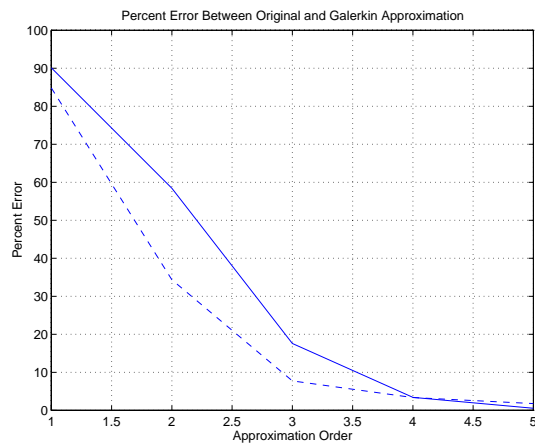


Figure 10: Percent error versus approximation order for 1D domain. Solid: approximation via empirical eigenfunctions computed using $M = 201$ snapshots. Dashed: approximation via empirical eigenfunctions computed using $M = 10$ (first 10) snapshots.

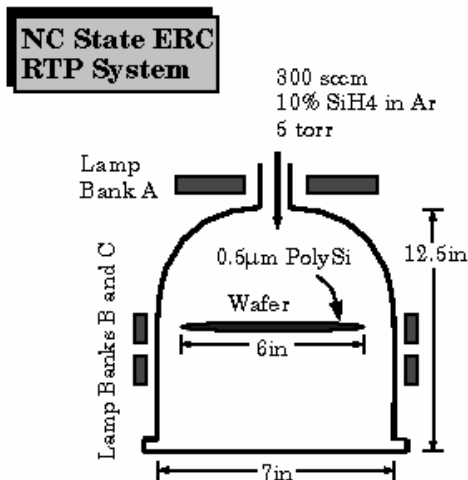


Figure 11: The North Carolina State University RTCVD reactor with typical operating conditions and parameters.

4 Applications

Now that we have demonstrated the use of spectral methods on some simple systems, we can further explore the power of these ideas by applying them to some practical and more complicated systems.

4.1 Model Reduction in Rapid Thermal Processing

Rapid Thermal Processing (RTP) is a versatile, cold-wall, single-wafer approach to semiconductor wafer processing suitable for several applications including annealing (RTA), cleaning (RTC), chemical vapor deposition (CVD), oxidation (RTO), and nitridation (RTN). RTP systems have several advantages including improved wafer-to-wafer uniformity and reduced thermal budgets. For more details on RTP models and applications, see [1, 4, 8, 10, 16, 18].

We have simulated the fluid flows and thermal dynamics of a RTCVD reactor under development at North Carolina State University [8] (Figure 11) using the Fluent computational fluid dynamics software package [5]. Gases flow into the chamber from top (inlet) to bottom (outlet). The wafer is heated via radiation from three lamp banks situated in strategic locations beside and around the reactor. Material is deposited on the wafer which rests in the middle of the chamber. Deposition rate is a strong function of the temperature distribution on the wafer surface. Currently there are many open problems regarding the sensing and control of the temperature distribution. Essentially, the Fluent software numerically integrates the governing equations (e.g. Navier-Stokes, conservation of energy) using a finite-differences/finite-elements approach on the user specified geometry and grid associated with the problem.

Here we present the results of one simulation of the temperature field dynamics for the RTCVD reactor. In this simulation, instead of including the radiative heat transfer dynamics from lamps to wafer we make the simplifying assumption that the wafer has been instantaneously heated to a uniform, constant 1000 K. Operating conditions are set as shown in Figure 11.

Figure 12 shows snapshots of the temperature field for equally spaced values of time during the transient. These snapshots were used in a POD to determine the empirical eigenfunctions of the system. These can be seen in Figure 13. Note that the corresponding normalized eigenvalues are 0.9750, 0.0157, and 0.0060.

The important point here is that even though the temperature field dynamics are governed by complicated PDEs that are coupled with the PDEs governing the conservation of momentum and conservation of

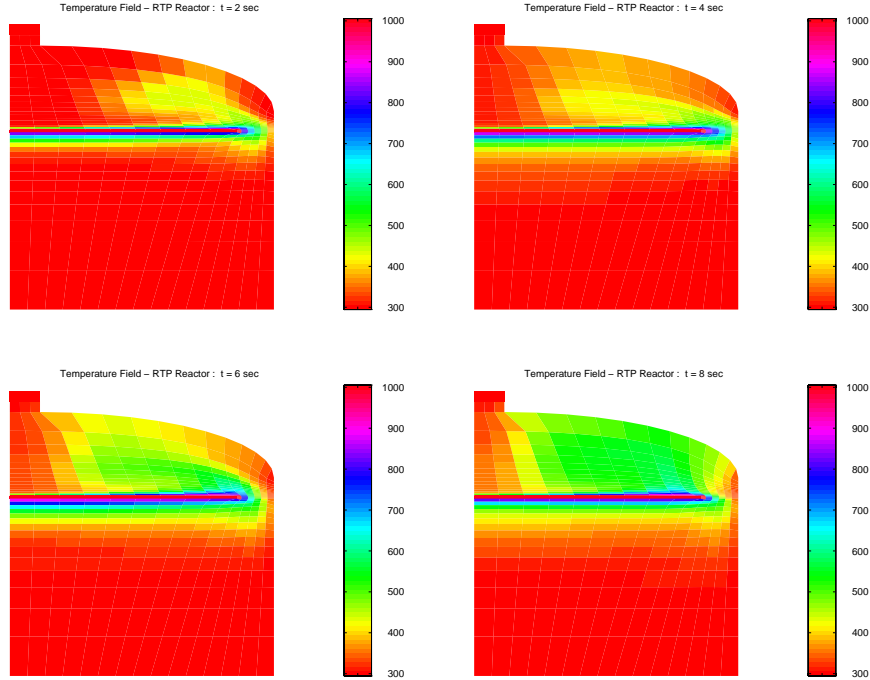


Figure 12: Temperature field snapshots of RTP reactor: $t = 2, 4, 6, 8$ seconds.

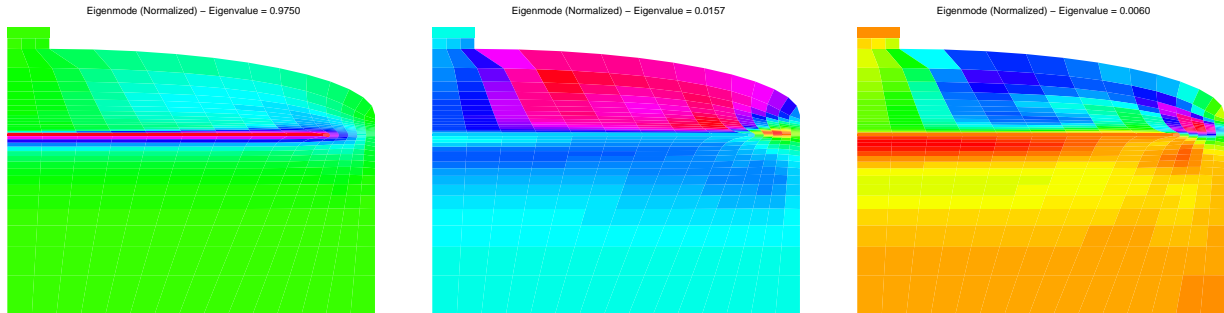


Figure 13: First three eigenmodes for temperature field dynamics of NCSU RTP reactor: $\lambda_1 = 0.9750$, $\lambda_2 = 0.0060$, $\lambda_3 = 0.0001$.

mass, almost all of the energy in the system is associated with just one spatial structure. We see that 97.5 percent of the energy in the system is contained in the first mode. This result is encouraging because the temperature field dynamics can be simulated to a high degree of accuracy using only two or three modes thus yielding a model of significantly reduced complexity. It also provides insight into the way the system behaves as we see that the complicated dynamics of the system are essentially low-dimensional.

4.2 Sensing and Control

One goal of the research associated with this paper is to investigate implications of these ideas for sensing and control of distributed parameter systems such as those found in the area of RTCVD. We wish to examine the use of other sets of basis functions such as wavelets, and to study how to use reduced order models in formulating control strategies.

5 Conclusions

We have illustrated the main ideas of certain model reduction techniques by presenting some elementary examples of their use. We have shown how the correctness of the reduced order models can be affected by various parameters of interest. Simulation results from a heat diffusion system and an RTCVD reactor indicate that the methods described in this paper can be applied effectively.

References

- [1] R. A. Adomaitis. RTCVD model reduction: A collocation on empirical eigenfunctions approach. Technical Report T.R. 95-64, Institute for Systems Research, 1995.
- [2] Gal Berkooz, Philip Holmes, and John L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Reviews of Fluid Mechanics*, 25:539–575, 1993.
- [3] C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang. *Spectral Methods in Fluid Dynamics*. Springer-Verlag, 1988.
- [4] Richard B. Fair, editor. *Rapid Thermal Processing Science and Technology*. Academic Press, Inc., 1250 6th Ave. San Diego, CA 92101-4311, 1993.
- [5] Fluent, Inc., Lebanon, NH. *Fluent User's Guide*, 4.3 edition, 1995.
- [6] D. Gottlieb and S. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications*. SIAM, 1977.
- [7] Thomas J. Hughes. *The Finite Element Method*. Prentice-Hall, 1987.
- [8] W. J. Kiether, M. J. Fordham, Seungil Yu, A. J Silva Neto, K.A. Conrad, J. R. Hauser, F.Y. Sorrell, and J.J. Wortman. Three-zone rapid thermal processor system. In *Proceedings of the 2nd International Rapid Thermal Processing Conference, RTP94, Monterey, CA*, pages 96–101, 1994.
- [9] C.R. MacCluer. *Boundary Value Problems and Orthogonal Expansions*. IEEE Press, 1994.
- [10] M. Meyyappan. *Computational Modeling in Semiconductor Processing*. Artech House, 1995.
- [11] Andrew J. Newman. Model reduction via the Karhunen-Loeve expansion Part I: An exposition. Technical Report T.R. 96-32, Inst. Systems Research, April 1996.
- [12] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524, March 1987.
- [13] Lawrence Sirovich. Turbulence and the dynamics of coherent structures; Part I: Coherent structures. *Quarterly Appl. Math.*, 45(3):561–571, October 1987.
- [14] Lawrence Sirovich. Turbulence and the dynamics of coherent structures; Part II: Symmetries and transformations. *Quarterly of Applied Mathematics*, 45(3):573–582, October 1987.
- [15] Lawrence Sirovich. Turbulence and the dynamics of coherent structures; Part III: Dynamics and scaling. *Quarterly of Applied Mathematics*, 45(3):583–590, October 1987.
- [16] F.Y. Sorrell, J.A. Harris, and R.S. Gyurcsik. A global model for rapid thermal processors. *IEEE Transactions on Semiconductor Manufacturing*, 3(4):183–188, November 1990.
- [17] Eugene Wong. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, 1971.
- [18] X. L. Xu, J. J. Wortman, and M.C. Ozturk. A review of rapid thermal processing: System design and applications. volume 1595, pages 18–34, 1991.