2016 American Control Conference (ACC)
Boston Marriott Copley Place
July 6-8, 2016. Boston, MA, USA

Linear Parameter-varying Approach for Modeling Rapid Thermal Processes

Mark Trudgen[1] Syed Z. Rizvi[1] Javad Mohammadpour[1]

Abstract— **In the present paper, a new approach is presented to model *rapid thermal processing* (RTP) systems. Within the past decade, RTP has achieved acceptance as the mainstream technology for semiconductor manufacturing. Thermal processing is one of the most efficient ways to control the phase-structure properties; moreover, the time duration of RTP systems reduces the so-called *thermal budget* significantly compared to the traditional methods. RTP implementation is based on the use of light from heating lamps to provide a heat flux. This process is highly nonlinear due to the radiative heat transfer and material properties. By invoking the first principles-based models, we develop in this paper a *linear parameter-varying* (LPV) model to directly account for all the nonlinearities within the system. The model is then discretized into a high-order affine LPV system; thereafter, *principal component analysis* (PCA) method is utilized to reduce the number of LPV model's scheduling variables, followed by the use of *proper orthogonal decomposition* (POD) for model order reduction. Finally, simulations demonstrate that the low-order LPV model, which is in a form suitable for controller design purposes, retains the properties of the original full-order model.**

I. Introduction

Embedded deep in the heart of all electrical applications are *integrated circuits* (IC), primarily composed of semiconductor devices made from a sequence of batch processes. With the continual developments in IC technology, we see an increase in the demand for performance improvements in terms of both quality variables and output yield resulting from the use of larger diameter silicon wafers [1]. To achieve these increased yields, precise uniform temperature control of a wafer is of paramount consideration. As such, the semiconductor industry has relied on advancements in control and modeling for these purposes [2], [3].

Thermal processes are very important in the fabrication of semiconductor devices. The longer a wafer is kept at elevated temperatures, the higher probability it has of defects. As such, minimizing a metric called the *thermal budget* is very important not only for heating cost purposes, but also for purity and defect reasons [3]. The thermal budget is calculated as the integral of the product of the diffusivity and the temperature over time. As wafer dimensions have shrunk down into the micron range, there has been an increase in demand on uniform thermal processing. The push to reduce the thermal budget, combined with the tight quality requirements, has given rise to a new technology called *single wafer processing* (SWP). Traditionally, batch processes were used where wafer holders called "boats" loaded many wafers onto a quartz substrate to be placed inside a furnace. Although furnace construction included insulated walls to improve the isothermal nature of the environment inside, wafer uniformity remains an issue. This issue has led to the development of *rapid thermal processing* (RTP) technologies. Single wafer units are better alternatives to meet temperature uniformity and a lower thermal budget; however, they must be able to heat up and cool down quickly in order to compete with the volume output of batch processing. A typical RTP system undergoes three phases: (1) rapid heating on the order of 50-200°C/s, (2) a processing phase of constant temperature, and (3) a rapid cooling phase. Heating is made possible via high powered lamps. The heating lamps are split into zones, and this allows for control flexibility. Finally, optical pyrometers are used to feedback temperature measurements.

Several alternative approaches to modeling and control of single wafer RTP systems have been suggested in the literature. Review of a *Steag Inc* RTP system with first principles modeling and genetic *nonlinear model predictive control* (NMPC) was proposed in [2]. An adaptive control model was presented in [4]. The authors in [5] studied the thermal behavior of large silicon wafers. Decentralized control approach in the design of PI controllers was used in [6]. The authors in [7] used *proper orthogonal decomposition* (POD) to reduce the order of an RTP system. A linear quadratic gaussian (LQG) approach to control was taken in [8]. Furthermore, a run to run approach was taken in [9], while [10] used *internal model control* (IMC). Finally, multivariable and multizone control was presented in [11]–[13]. A survey of RTP processes was presented in [1].

As observed from the aforementioned literature, first principles-based modeling of the RTP system is best represented by a partial differential equation (PDE) with varying coefficients and nonlinear boundary conditions. However, direct control of such nonlinear system is not seen in literature, nor are there modeling frameworks that present the plant in a control-oriented form. In this paper, we propose a *linear parameter-varying* (LPV) modeling approach that directly and systematically copes with the complex nonlinearities seen in the RTP processes. LPV techniques have gained popularity as they have developed into effective tools to

[1] The authors are with the Complex Systems Control Laboratory (CSCL), College of Engineering, The University of Georgia, Athens, GA 30602, USA. Corresponding author's email: javadm@uga.edu.

control multi-input multi-output (MIMO) nonlinear systems [14]. Furthermore, the application of these methods has not been explored for thermal processes including RTP systems, for which the well known nonlinear material properties can be exploited in the LPV framework of scheduling variables.

This paper is organized as follows: Section II describes the process and the first principles-based model of the generic RTP systems. Section III reintroduces the system as a high-order discretized state-space model. This model is then converted into an LPV model. In Section IV, the number of scheduling variables in the high-order LPV model is first reduced using PCA, and then order of the model is reduced using POD in Section V. Section VI shows simulation results and Section VII draws conclusions.

II. RTP Process Description and Modeling

A. The Typical RTP Setup

In our modeling of RTP systems, we choose to use a single wafer setup as seen in [2], [3], [4], [6], [8]. For a typical RTP system, a concentric lamp array, usually of halogen lamps, is located above a quartz window. The lamp array is divided into zones, and the zone power percentage can be adjusted independently in each zone in order to aid the uniform processing of large wafers. The heating lamps and chamber are cooled by a cooling flow. The wafer is kept rotating in order to ensure uniformity. Finally, an optical pyrometer located underneath the wafer provides temperature measurement. The setup is illustrated in Figure 1.

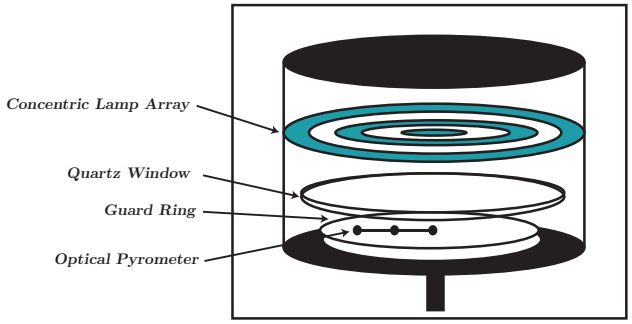

Fig. 1. Representative Single Wafer RTP Setup.

B. The First Principles-based Model

The first step in achieving a control-oriented model is to utilize a first principles-based model of the RTP chamber. Energy balance on the wafer in the RTP chamber is given as [2]

$$\rho C(T)\frac{\partial T}{\partial t} = q_k + q_c + q_r, \tag{1}$$

where ρ, C, and T are the wafer density, specific heat, and temperature, respectively. Variable t denotes continuous time. The heat transfer rates by conduction, convection, and radiation are denoted by q_k, q_c, and q_r, respectively.

In order to decrease the computational complexity of the model, we first make geometric simplifications. We note that using cylindrical coordinates, the system has rotational symmetry, and hence the full three-dimensional model (r,θ,z) can be reduced to a two-dimensional problem in (r,z). Next, we observe that in order to increase uniformity, the wafer is rotated during the operation, and this allows us to return the problem to Cartesian coordinates by representing the wafer as a radial chord. We use the simplifications and write the energy balance in terms of (x,z) as a *partial differential equation* (PDE) as follows

$$\rho C(T)\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\tilde{k}(T)\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial z}\left(\tilde{k}(T)\frac{\partial T}{\partial z}\right). \tag{2}$$

Furthermore, the initial and boundary conditions are given as

$$T(x,z,0) = T_{initial}, \tag{3}$$

$$\tilde{k}(T)\frac{\partial T}{\partial x} = 0 \quad \text{at} \quad x = 0, \tag{4}$$

$$\tilde{k}(T)\frac{\partial T}{\partial x} = -h_e(T - T_{wall}) \quad \text{at} \quad x = R, \tag{5}$$

$$\tilde{k}(T)\frac{\partial T}{\partial z} = F_1\varepsilon_1(T)\sigma(T^4 - T_{cool}^4) + h_w(T - T_{cool}) \text{ at } z = 0, \tag{6}$$

$$h_w(x) = h_i + (h_o - h_i)\left(\frac{x}{R}\right)^4, \tag{7}$$

$$\tilde{k}(T)\frac{\partial T}{\partial z} = \varepsilon_2(T)Q(x,t) - F_2\varepsilon_2(T)\sigma(T^4 - T_a^4) \text{ at } z = Z, \tag{8}$$

where T is the wafer temperature; $T_{initial}$ is the initial wafer temperature; h_w is the overall convective heat transfer coefficient; h_i, h_o, and h_e are the heat transfer coefficients at the center, edge, and wafer edge, respectively [15]; T_{cool} is the temperature of the coolant; T_a is the temperature of the quartz window; T_{wall} is the temperature of side walls; $C(T)$ is the heat capacity; $\tilde{k}(T)$ is the thermal conductivity; σ is the Stefan-Boltzmann constant; ε_1 and ε_2 are the emissivities of the lower and upper wafer surfaces; F_1 and F_2 are the tunable reflective coefficients; x and z are the Cartesian coordinates corresponding to the radial thickness Z, and the radial chord length X; and $Q(x,t)$ is the heat flux as described by $\frac{q(x,t)}{A(x)}$. The heat power $q(x,t)$ is described later in (14) and $A(x)$ is the effective wafer area at the chord position.

The initial condition in (3) makes the reasonable assumption that the entire wafer starts at a uniform temperature. Next, we assume that the quartz window, the side walls, and cooling temperatures are held constant and equal ($T_a = T_{cool} = T_{wall}$). The boundary condition (6) represents the conduction heat losses made with the reactor walls by convection. We use the overall heat transfer coefficient approach as in [15] in order to account for spatial variations. Lastly, the boundary condition at $z = Z$ as described in (8) relates the heat transfer in the wafer to the heat generation of the heating lamps and also the heating losses in the quartz window.

Next, we must account for the operation range of the RTP systems. Typical RTP systems range in temperature from 25

to 1200°C [2]. The material properties of silicon wafers are given in [16] and the thermal conductivity and heat capacity are given as

$$\tilde{k}(T) = 802.99T^{-1.12}\left[\frac{W}{cmK}\right] \in [300, 1683]K, \tag{9}$$

$$C(T) = 0.641 + 2.473\ x\ 10^4\ T\ \left[\frac{J}{gK}\right] > 300K. \tag{10}$$

Furthermore, the material properties of the emissivity is given by [17]

$$\varepsilon(T) = 0.2662 + 1.8591\ T^{-0.1996} exp\left[-\frac{1.0359\ x\ 10^{25}}{T^{8.8328}}\right]. \tag{11}$$

For further computational simplicity, we notice that the wafer density can be taken as a constant, $\rho = 2330kg/m^3$, since this density does not strongly depend on temperature. Additionally, this weak temperature dependence allows for a homogeneous energy balance assumption such that (2) can be simplified to

$$\rho C(T)\frac{\partial T}{\partial t} = \tilde{k}(T)\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right). \tag{12}$$

C. Modeling Heating Lamp Input Flux

Radiation heat transfer is the main mode heat transfer mechanism that raises the wafer temperature. The lamp array is located directly above the wafer and typically arranged into concentric rings of heating zones. Radiation heat transfer is a complicated heat transfer mode as energy transfer is based on both wavelength and geometry. Therefore, a theoretical model must also account for both diffusive and reflective radiation heat transfer. However, in order to put the model in a form suitable for controller design purposes, we first make the partial simplifying assumption of a diffusive grey body. As seen in (11), the emissivity is still a function of temperature, but we relax the condition that it also must be a function of wavelength.

Next, to calculate the heat flux transferred to the wafer, we follow the view factor formula given in [2] that describes the geometric relationship between two areas given as

$$F_{1-2} = \frac{1}{A_1}\int_{A_1}\int_{A_2}\frac{cos(\theta_1)cos(\theta_2)}{\pi S^2}dA_2 dA_1, \tag{13}$$

where F_{1-2} is the radiation fraction transmitted from surface 1 to surface 2 and θ_1 and θ_2 are the normal angles at the surfaces while S is the distance between the surfaces, and A_1 and A_2 are the corresponding surface areas. Following [2], (13) is integrated on a differential annular heating ring. We then recast into a generalized form for the multiple zones as

$$q(x,t) = \alpha \cdot \sum_{j=1}^{n} F_{j-x}(x, r_{in}, r_{out}) \cdot q(j), \tag{14}$$

where α is a tunable parameter, j represents the ring number, n is the maximum number of zones, r_{in} and r_{out} are the respective radial measurements of the local ring number, and $q(x,t)$ represents the heating ring power.

III. Nonlinear Modeling of RTP Systems

The two-dimensional heat equation (12) is given on the physical domain $\mathbb{S} = \{x | x \in [0,\chi]\} \cup \{z | z \in [0,\zeta]\}$ and the temporal domain $\mathbb{R} = \{t | t \in [0,\tau]\}$. Now T: $\mathbb{S} \times \mathbb{R} \to \mathbb{T}$ is the space and time dependent temperature. An approximate discrete solution of (12) is then represented by

$$T_{i,j}^k = T : \hat{\mathbb{S}} \times \hat{\mathbb{R}} \to \mathbb{T}, \tag{15}$$

with the finite sets $\hat{\mathbb{S}} = \{s_1, \ldots, s_{mm \times nn}\}$, $\hat{\mathbb{R}} = \{t_1, \ldots, t_K\}$, where $mm \times nn$ is the number of grid points, and K is the number of time samples.

A. Discretization of the RTP Model

The partial differential equation (PDE) in (12) is discretized using a *forward time-center space* (FTCS) discretization method, which gives

$$\rho C(T_{i,j}^k)\frac{T_{i,j}^{k+1} - T_{i,j}^k}{\Delta t} = \tilde{k}(T_{i,j}^k)\left[\frac{T_{i-1,j}^k - 2T_{i,j}^k + T_{i+1,j}^k}{(\Delta x)^2} + \frac{T_{i,j-1}^k - 2T_{i,j}^k + T_{i,j+1}^k}{(\Delta z)^2}\right], \tag{16}$$

where Δx and Δz represent the discretization step size in spatial directions, and Δt is the time step; i and j represent the two spatial indices in the x and z dimensions, and k represents the time index. We also discretized the nonlinear boundary conditions subject to (5)-(8). A simulation result of the discretized system is shown in Figure 2 at an arbitrary time instant, where Δx = 1/20, Δz = 1/4, and $T_{initial}$ = 303 K. The time step Δt is chosen such that it obeys the limits of the FTCS discretization stability restrictions. These conditions are chosen to examine the open-loop response to an input signal with typical wafer dimensions [2].

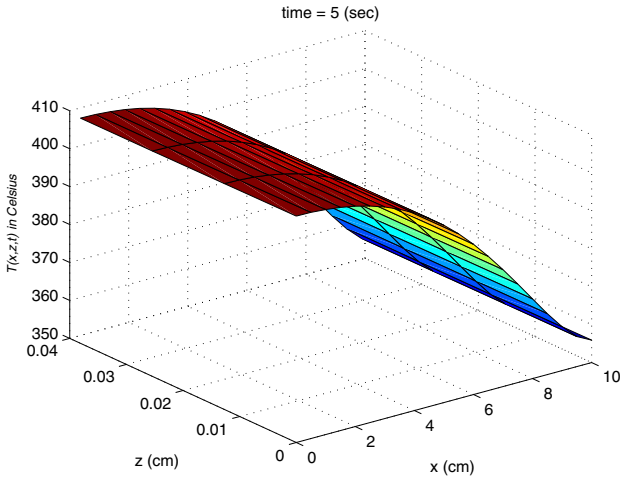

Fig. 2. Simulation of the open-loop single wafer RTP setup.

B. Linear Parameter-varying Model Derivation

It is observed that the system (2)-(8) is nonlinear, and remains so after discretization. There exists several approaches to transform a nonlinear system represented by

$$x(k+1) = F(x(k), u(k)), \tag{17}$$

into a linear model. A well known approach is the Jacobian linearization of (17) around an equilibrium trajectory. The resulting linear system will then only describe the local behavior around that trajectory. Another approach is based on rewriting (17) into an equivalent form, where the nonlinearities can be hidden inside newly defined variables, the so-called *scheduling variables*. Such a model is called a *linear parameter-varying* (LPV) model [14]. A discrete-time LPV model can be represented in state space as

$$\begin{bmatrix} x(k+1) \\ y(k) \end{bmatrix} = \begin{bmatrix} A(\theta(k)) & B(\theta(k)) \\ C(\theta(k)) & D(\theta(k)) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}, \tag{18}$$

where $y(k)$ represents the control output. We rewrite the nonlinear model into an LPV form (18), since this form is suitable for LPV controller synthesis. The state vector $x(k)$ consists of the temperature of the wafer at the discretized locations, with *mm* being the total number of steps in the x direction, and *nn* being the number of steps in the z direction; the state vector is given by

$$x(k) = [x_1(k), \ldots, x_{mm\times nn}(k)]^\top. \tag{19}$$

The state vector $x(k)$ is arranged with respect to the spatial coordinates, and thus the structure of the elements of the state vector is as follows

$$x(k) = [T^k_{1,1}, \ldots, T^k_{mm,1}, \ldots, T^k_{1,nn}, \ldots, T^k_{mm,nn}]^\top. \tag{20}$$

Next, we define the scheduling variable vector in a similar fashion where $\theta_1(T^k_{i,j})$-$\theta_4(T^k_{i,j})$ are derived so that (18) is affine in the scheduling variables,

$$\theta(k) = [\theta_1(T^k_{1,1}), \ldots, \theta_4(T^k_{1,1}), \ldots, \theta_1(T^k_{mm,nn}), \ldots, \theta_4(T^k_{mm,nn})]^\top. \tag{21}$$

Remark 1: Each scheduling variable is unique as the scheduling variables are functions of the local temperature at each unique spatial location.

In formulating (18), the higher the order and the larger the number of scheduling variables in the model, the more accurately the model will represent the original system in (2)-(8). Hence, a trade-off must be made between model complexity, and the tractability of control design and computational cost. Our objective now becomes to use order reduction techniques in order to achieve a balance between accuracy of the model and the number of scheduling variables.

IV. LPV Model Reduction Using PCA

First, we reduce the number of scheduling variables through the use of principal component analysis (PCA) [18]. To apply PCA to the LPV scheduling variables data, one first needs to generate and collect data by means of measurements or simulations [19], such that the data covers all regions of operation within the operating range. Given the LPV model (18) and assuming that the measurable signals have been sampled at time instants $k = \{1, 2, \ldots, K\}$, scheduling variables $\theta(k) \in \mathbb{R}^{\tilde{l}}$ with $\tilde{l} = 4 \times mm \times nn$ are computed and collected in the following $\tilde{l} \times K$ matrix

$$\Theta = \begin{bmatrix} \theta(1) & \cdots & \theta(K) \end{bmatrix} = \begin{bmatrix} f(T^1_{i,j}) & \cdots & f(T^K_{i,j}) \end{bmatrix},$$

where $\tilde{l}$ represents the actual number of scheduling variables and K denotes the number of data samples, with $K \geq \tilde{l}$. PCA is then applied by solving an eigenvalue problem for the covariance matrix $\Theta\Theta^\top$. The covariance matrix is given by

$$\bar{C} = \frac{1}{K}\Theta_c\Theta_c^\top,$$

where $\Theta_c = \mathscr{C}(\Theta) = \Theta - \theta_{\text{mean}}$ is the data matrix Θ normalized such that each row of Θ has zero mean. We then solve an eigenvalue problem for the covariance matrix $\bar{C}$, such that $\bar{C}v_i = \tilde{\lambda}_i v_i$, where $\tilde{\lambda}_i$ and v_i are the i^{th} eigenvalue and eigenvector, respectively. The eigenvectors are then sorted in descending order of their corresponding non-zero eigenvalues, and the m principal components for any test point $\theta(k)$, at a given time sample k, are extracted using

$$\rho(k) = g(T^k_{i,j}) = V_m^\top f(T^k_{i,j}) = V_m^\top \theta(k),$$

where V_m denotes an $\tilde{l} \times m$ matrix whose columns contain the m eigenvectors associated with the first m significant eigenvalues. The approximation of the actual variable $\hat{\theta}(k)$, corresponding to $\rho(k)$, can be easily computed as

$$\hat{\theta}(k) = \mathscr{C}^{-1}(V_m\rho(k)), \tag{22}$$

where $\mathscr{C}^{-1}(V_m\rho(k)) = V_m\rho(k) + \theta_{\text{mean}}$. Henceforth, we also drop the time index k for better readability and denote $\rho(k)$ and $\theta(k)$ simply as ρ and θ. The PCA-based reduced model can be represented as

$$\begin{aligned} x(k+1) &= \hat{A}(\rho)x(k) + \hat{B}(\rho)u(k), \\ y(k) &= \hat{C}(\rho)x(k) + \hat{D}(\rho)u(k). \end{aligned} \tag{23}$$

If m equals the number of non-zero eigenvalues, the mapping matrices $\hat{A}(\cdot)$, $\hat{B}(\cdot)$, $\hat{C}(\cdot)$, and $\hat{D}(\cdot)$ are related to the reconstructed scheduling variable $\hat{\theta}$ by [19]

$$\hat{Q}(\rho) = \begin{bmatrix} \hat{A}(\rho) & \hat{B}(\rho) \\ \hat{C}(\rho) & \hat{D}(\rho) \end{bmatrix} = \begin{bmatrix} A(\hat{\theta}) & B(\hat{\theta}) \\ C(\hat{\theta}) & D(\hat{\theta}) \end{bmatrix} = Q(\hat{\theta}). \tag{24}$$

We take m to be the number of significant eigenvalues, in which case, $\hat{Q}(\rho)$ will be an approximation of $Q(\hat{\theta})$; what constitutes significance is a user's choice. We write

the following:

$$\begin{aligned}\hat{Q}(\rho) = Q(\hat{\theta}) &= Q_0 + \sum_{i=1}^{\tilde{l}} Q_i \hat{\theta}^i \\ &= Q_0 + \sum_{i=1}^{\tilde{l}} Q_i \left(V_m \rho + \theta_{\text{mean}}\right)^i \\ &= Q_0 + \sum_{i=1}^{\tilde{l}} Q_i \theta_{\text{mean}}^i + \sum_{i=1}^{l} Q_i \left(V_m \rho\right)^i \\ &= \underbrace{Q_0 + \sum_{i=1}^{\tilde{l}} Q_i \theta_{\text{mean}}^i}_{\hat{Q}_0} + \sum_{j=1}^{m} \underbrace{\sum_{i=1}^{\tilde{l}} Q_i [V_m]_{i,j}}_{\hat{Q}_j} \rho^j \\ &= \hat{Q}_0 + \sum_{j=1}^{m} \hat{Q}_j \rho^j, \end{aligned} \tag{25}$$

where θ^i denotes the i^{th} element of the vector θ, and $[V_m]_{i,j}$ denotes the $\{i,j\}$ entry of the matrix V_m. (25) is a reduced model also affine in the reduced scheduling variables ρ.

V. Proper Orthogonal Decomposition

The next step is to reduce the order of the derived LPV model using the *proper orthogonal decomposition* (POD) method. POD delivers a basis for model decomposition in order to extract dominant trends and features [20]. Essentially, POD extracts a set of *orthonormal basis functions* (OBF) [21], usually with a few modes [20]. To approximate the function of interest over a domain, we write the ensemble into coefficients to be determined,

$$T(x,z,t) \approx \widehat{T}(x,z,t) = \sum_{j=1}^{M} \alpha_j(x,z)\varphi_j(t), \tag{26}$$

where α_j's define the set of OBFs, and φ_j's denote the time-dependent coefficients. We employ the method of snapshots [22], which solves an eigenvalue problem and only requires an ensemble of appropriately organized data points [21]. Here we define $\tilde{D}$ to be the dimension corresponding to $mm \times nn$. The data needed is captured as $T_{snap} \in \mathbb{R}^{\tilde{D}\times K}$, where $\tilde{D}$ corresponds to the number of discretization steps and K corresponds to the number of snapshots. In the finite-dimensional case, POD reduces to an SVD problem. This is done by making use of SVD as

$$T_{snap} = \Phi \Sigma V^\top = \begin{bmatrix}\Phi_r & \Phi_s\end{bmatrix} \begin{bmatrix}\Sigma_r & 0 & 0 \\ 0 & \Sigma_s & 0\end{bmatrix} \begin{bmatrix}V_r^\top \\ V_s^\top\end{bmatrix}. \tag{27}$$

The columns of Φ from the SVD form the set of basis functions $\{\alpha_1, \cdots, \alpha_{\tilde{D}}\}$. This type of projection captures the most *energy* for the reduced model. In (27), $\Phi \in \mathbb{R}^{\tilde{D}\times\tilde{D}}$ and $V \in \mathbb{R}^{K\times K}$ and the sizes of Φ_r, Σ_r, and V_r each correspond to the M dominant singular values chosen. These basis functions, called POD modes, are used to obtain accurate low-order dynamic models via Galerkin projection [21].

Next, we examine the singular values to produce a reduced-order model. A representation of the *energy* that is captured by the reduced-order model is given by the differences in the sum of the squared singular values (28). A high percentage of energy preserved is always desired, meaning a larger M, which indicates that the model retains more of the information contained in the original snapshots. The preserved energy percent (PEP) is defined as [23]

$$PEP = 100 \times \frac{\sum_{i=1}^{M} \sigma_i^2}{\sum_{i=1}^{N} \sigma_i^2}, \tag{28}$$

where we note that M is the user's choice, and N is the original order of the state-space system. To obtain the reduced order LPV state-space model, (18) is multiplied from both sides by the truncated orthonormal matrix $\Phi_r \in \mathbb{R}^{\tilde{D}\times M}$ as

$$\Phi_r^\top x(k+1) = \Phi_r^\top \hat{A}(\rho)x(k) + \Phi_r^\top \hat{B}(\rho)u(k). \tag{29}$$

Recalling that $x(k)$ is the state vector of the original high-order approximation, the reduced-order state vector becomes

$$x_r(k) = \Phi_r^\top x(k). \tag{30}$$

Since each element of $x_r(k)$ is a linear combination of the elements of $x(k)$, substituting (30) into (29) yields

$$\begin{aligned} x_r(k+1) &= A_r(\rho)x_r(k) + B_r(\rho)u(k) \\ y_r(k) &= C_r(\rho)x_r(k) + \hat{D}(\rho)u(k), \end{aligned} \tag{31}$$

with

$$A_r(\rho) = \Phi_r^\top \hat{A}(\rho)\Phi_r, \quad B_r(\rho) = \Phi_r^\top \hat{B}(\rho), \quad C_r(\rho) = \hat{C}(\rho)\Phi_r. \tag{32}$$

VI. Simulation Results and Discussion

We present simulation results comparing the nonlinear model with the reduced-order LPV model with a low number of scheduling variables.

Remark 2: The LPV framework is a natural framework for RTP modeling since temperature is a readily measureable scheduling variable and the nonlinearities seen are smooth.

Remark 3: All of the scheduling variables share the common thread in that they are all functionals of temperature.

A. Simulation Results

In our nonlinear simulation we used $\Delta x = \frac{1}{20}$ and $\Delta z = \frac{1}{3}$ and the same geometry setup in [2]. Due to the very high number of initial scheduling variables we averaged the temperature across three zones of the wafer, thus beginning with 12 scheduling variables. Using the PCA analysis described in Section IV we reduced the number of scheduling variables to 3 while retaining 97% of the energy. Figure 3 shows a sample projection from the reduced scheduling variables back onto the original high-order space. Next we used the POD method as described in Section V to reduce the order of the system. Using (28)-(32) we reduced the high-order system to a 3^{rd} order system while preserving 99% of the energy.

Remark 4: Since RTP wafer recipes are known *a priori* we can expect to preserve a large amount of energy in the

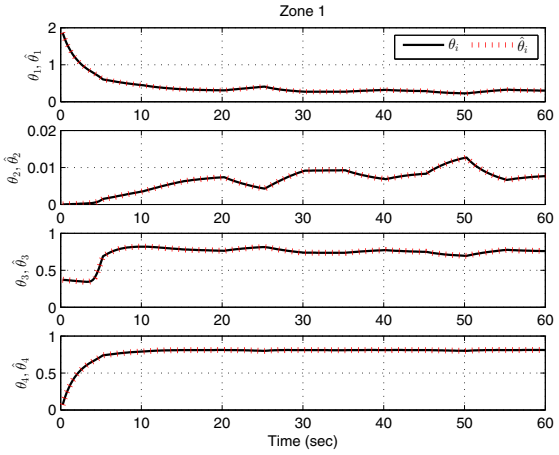

Fig. 3. Projection of the reduced scheduling variables back onto the original high-order space. The solid line represents the original scheduling data θ, and dotted line represents the projected scheduling data $\hat{\theta}$.

reduced-order system to create a low-order system computationally inexpensive enough to be run in a real time environment.

Next, we compare the nonlinear model with the LPV model that has been reduced in order and scheduling variables. Using a random zone power percentage input signal to both models, we see in Figure 4 from two sample discretization points that the LPV model is a good representation.

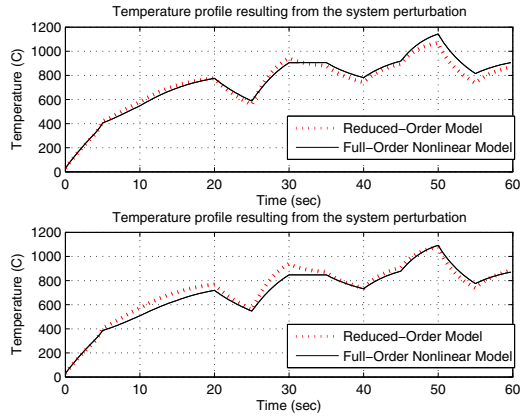

Fig. 4. Comparison of full-order nonlinear model vs. the low-order LPV model at *two representative node locations*.

VII. Concluding Remarks

In this paper, we used a first principles-based modeling approach to develop an affine LPV model for rapid thermal processes. PCA and POD were used to reduce the complexity of the LPV model into a form tractable for controller design purposes. Reducing the number of scheduling variables is desirable since it affects exponentially the number of *linear matrix inequality* (LMI) constraints required to be solved for LPV controller synthesis. Finally, using open-loop simulation results, we observed an agreement between the high-order nonlinear model and the reduced-order LPV model.

References

[1] T. F. Edgar, S. W. Butler, W. J. Campbell, C. Pfeiffer, C. Bode, S. B. Hwang, K. Balakrishnan, and J. Hahn, "Automatic control in microelectronics manufacturing: Practices, challenges, and possibilities," *Automatica*, vol. 36, no. 11, pp. 1567–1603, 2000.

[2] E. Dassau, B. Grosman, and D. R. Lewin, "Modeling and temperature control of rapid thermal processing," *Computers & chemical engineering*, vol. 30, no. 4, pp. 686–697, 2006.

[3] J. Ebert, D. De Roover, L. Porter, V. Lisiewicz, S. Ghosal, R. Kosut, and A. Emami-Naeini, "Model-based control of rapid thermal processing for semiconductor wafers," in *Proc. of American Control Conference*, 2004, pp. 3910–3921.

[4] J. Y. Choi, H. M. Do, and H. S. Choi, "Adaptive control approach of rapid thermal processing," *IEEE Trans. Semiconductor Manufacturing*, vol. 16, no. 4, pp. 621–632, 2003.

[5] W. S. Yoo, T. Fukada, I. Yokoyama, K. Kang, and N. Takahashi, "Thermal behavior of large-diameter silicon wafers during high-temperature rapid thermal processing in single wafer furnace," *Japanese journal of applied physics*, vol. 41, no. 7R, p. 4442, 2002.

[6] C.-A. Lin and Y.-K. Jan, "Control system design for a rapid thermal processing system," *IEEE Trans. Control Systems Technology*, vol. 9, no. 1, pp. 122–129, 2001.

[7] H. Aling, S. Banerjee, A. K. Bangia, V. Cole, J. Ebert, A. Emami-Naeini, K. F. Jensen, I. G. Kevrekidis, and S. Shvartsman, "Nonlinear model reduction for simulation and control of rapid thermal processing," in *Proc. of American Control Conference*, 1997, pp. 2233–2238.

[8] Y. M. Cho and P. Gyugyi, "Control of rapid thermal processing: A system theoretic approach," *IEEE Trans. Control Systems Technology*, vol. 5, no. 6, pp. 644–653, 1997.

[9] E. Zafiriou, R. Adomaitis, G. Gattu *et al.*, "An approach to run-to-run control for rapid thermal processing," in *Proc. of American Control Conference*, 1995, pp. 1286–1288.

[10] C. D. Schaper, M. M. Moslehi, K. C. Saraswat, and T. Kailath, "Modeling, identification, and control of rapid thermal processing systems," *Journal of the Electrochemical Society*, vol. 141, no. 11, pp. 3200–3209, 1994.

[11] C. D. Schaper, "Real-time control of rapid thermal processing semiconductor manufacturing equipment," in *Proc. of American Control Conference*, 1993, pp. 2985–2990.

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

[13] P. P. Apte and K. C. Saraswat, "Rapid thermal processing uniformity using multivariable control of a circularly symmetric 3 zone lamp," *IEEE Trans. Semiconductor Manuf.*, vol. 5, no. 3, pp. 180–188, 1992.

[14] J. Mohammadpour and C. W. Scherer, Eds., *Control of linear parameter varying systems with applications*. Springer Science & Business Media, 2012.

[15] H. Lord, "Thermal and stress analysis of semiconductor wafers in a rapid thermal processing oven," *IEEE Trans. Semiconductor Manufacturing*, vol. 1, no. 3, pp. 105–114, 1988.

[16] V. Borisenko and P. J. Hesketh, *Rapid thermal processing of semiconductors*. Springer Science & Business Media, 2013.

[17] A. Virzi, "Computer modelling of heat transfer in czochralski silicon crystal growth," *Journal of crystal growth*, vol. 112, no. 4, pp. 699–722, 1991.

[18] I. Jolliffe, *Principal Component Analysis*, 2nd ed. NY: Springer.

[19] R. T. S.Z. Rizvi, J. Mohammadpour and N. Meskin, "A kernel-based PCA approach to model reduction of linear parameter-varying systems," *IEEE Transactions on Control Systems Technology*, p. In print, 2016.

[20] P. Holmes, *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge University Press, 2012.

[21] S. M. Hashemi and H. Werner, "LPV modelling and control of Burgers equation," in *Proc. 18th IFAC World Congress*, 2011.

[22] L. Sirovich, "Turbulence and the dynamics of coherent structures," *Quarterly of applied mathematics*, vol. 45, pp. 561–571, 1987.

[23] M. Trudgen and J. Mohammadpour, "Lumped-parameter model development and robust control of systems governed by 2-d parabolic convection-diffusion equation," in *Proc. of American Control Conference*, 2015, pp. 607–612.