International Journal of Control, Automation and Systems 16(1) (2018) 207-216
http://dx.doi.org/10.1007/s12555-016-0788-x
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Linear Parameter-varying Approach for Modeling and Control of Rapid Thermal Processes

Mark Trudgen and Javad Mohammadpour Velni*

Abstract: In the present paper, a new approach is presented to model and control single wafer *rapid thermal processing* (RTP) systems. In 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 the nonlinearities within the system. The model is then discretized into a high-order LPV model; 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. Using the reduced order model, we then design a gain-scheduled controller to satisfy an induced $\mathcal{L}_2$ gain performance for tracking of a temperature profile and show improvement over other controller design methods suggested in the literature.

Keywords: Gain-scheduled control, linear parameter-varying systems, partial differential equations, rapid thermal processing.

1. INTRODUCTION

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 minimizing defects [1]. 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 50-250°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 measure temperatures [2].

Several approaches to modeling and control of single wafer RTP systems have been suggested in the literature. From a modeling standpoint, [3] and [4] studied the feasibility of modeling RTP chambers. Review of a *Steag Inc* RTP system with first principles modeling and *nonlinear model predictive control* (NMPC) was proposed in [5]. The authors in [6] studied the thermal behavior of large silicon wafers. The authors in [7] and [8] used *proper orthogonal decomposition* (POD) to develop a low-order control-oriented model of an RTP system. A linear quadratic Gaussian (LQG) approach to control was taken in [9] and [10]. Furthermore, a run-to-run approach was taken in [11], while [12] used *internal model control* (IMC). Finally, multi-variable and multi-zone control was presented in [13–15]. A survey of RTP processes was pre-

Manuscript received December 15, 2016; revised April 26, 2017; accepted June 16, 2017. Recommended by Associate Editor Yang Tang under the direction of Editor Myo Taeg Lim.

Mark Trudgen and Javad Mohammadpour Velni are with School of Electrical & Computer Engineering, The University of Georgia, Athens, GA 30602, USA (e-mails: mtrudgen@uga.edu, javadm@uga.edu).
* Corresponding author.

sented in [16].

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, which is the extended version of our recent work [17], we propose a* linear parameter-varying *(LPV) modeling approach that directly and systematically copes with the complex nonlinearities seen in the RTP system.* LPV techniques have gained popularity as they have developed into effective tools to control multi-input multi-output (MIMO) nonlinear systems [18]. Furthermore, the application of these methods has not yet been explored for thermal processes including RTP systems, for which the well known nonlinear material properties can be exploited in the LPV framework.

2. RTP SYSTEM DESCRIPTION AND MODELING

In this section, we describe a typical RTP setup and provide a first-principles model of the process developed previously in the literature. However, we note that this derivation can be made general and applied to other RTP chambers.

2.1. The typical RTP setup

In our modeling of RTP systems, we choose to use a single wafer setup as seen in [1,5,9]. 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 Fig. 1.

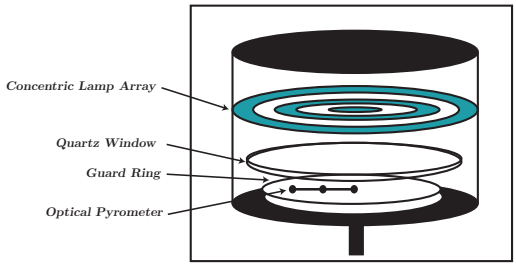

Fig. 1. Representative single wafer RTP setup.

2.2. 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 gives [5]: $\rho C(T)\frac{\partial T}{\partial t} = q_k + q_c + q_r$, 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:

$$\rho C(T)\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\tilde{k}(T)\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial z}\left(\tilde{k}(T)\frac{\partial T}{\partial z}\right). \quad (1)$$

Furthermore, the initial and boundary conditions are given as

$$T(x,z,0) = T_{initial}, \quad (2)$$

$$\tilde{k}(T)\frac{\partial T}{\partial x} = 0 \quad \text{at} \quad x = 0, \quad (3)$$

$$\tilde{k}(T)\frac{\partial T}{\partial x} = -h_e(T - T_{wall}) \quad \text{at} \quad x = R, \quad (4)$$

$$\tilde{k}(T)\frac{\partial T}{\partial z} = F_1\varepsilon_1(T)\sigma(T^4 - T_{cool}^4) + h_w(T - T_{cool}) \text{ at } z = 0, \quad (5)$$

$$h_w(x) = h_i + (h_o - h_i)\left(\frac{x}{R}\right)^4, \quad (6)$$

$$\tilde{k}(T)\frac{\partial T}{\partial z} = \varepsilon_2(T)Q(x,t) - F_2\varepsilon_2(T)\sigma(T^4 - T_a^4) \text{ at } z = Z, \quad (7)$$

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 [19], T_{cool} is the coolant temperature, 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 defined

by $Q(x,t) = \frac{q(x,t)}{A(x)}$. The heat power $q(x,t)$ is defined later and $A(x)$ is the effective wafer area at the chord position.

The initial condition in (2) assumes 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 (5) represents the conduction heat losses made with the reactor walls by convection. We use the overall heat transfer coefficient approach as in [19] in order to account for spatial variations. Lastly, the boundary condition at $z = Z$ described in (7) 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 [5]. Since the wafer is primarily composed of silicon, we can generalize the wafer's material properties to that of silicon [20]. The thermal conductivity and heat capacity of silicon are given in [5] and used in this work. Furthermore, the material properties of the emissivity is given in [21].

For further computational simplicity, we notice that the wafer density can be considered a constant, $\rho = 2330 kg/m^3$, since this density does not strongly depend on the temperature. Additionally, this weak temperature dependence allows for a homogeneous energy balance assumption such that (1) can be simplified to

$$\rho C(T)\frac{\partial T}{\partial t} = \tilde{k}(T)\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right). \tag{8}$$

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 the 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. We note that 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 [5] that describes the geometric relationship between two areas given as

$$F_{1-2} = \frac{1}{A_1}\int_{A_1}\int_{A_2}\frac{\cos(\theta_1)\cos(\theta_2)}{\pi S^2} dA_2 dA_1, \tag{9}$$

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 [5], (9) is integrated on a differential annular heating ring. We then recast into a generalized form for the multiple zones as $q(x,t) = \alpha \cdot \sum_{j=1}^{n} F_{j-x}(x, r_{in}, r_{out}) \cdot q(j)$, 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.

3. DEVELOPING A LARGE-ORDER LPV MODEL OF RTP SYSTEMS

The two-dimensional heat equation (8) 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 (8) is then represented by $T_{i,j}^k = T : \hat{\mathbb{S}} \times \hat{\mathbb{R}} \to \mathbb{T}$, 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.

3.1. Discretization of the RTP model

The partial differential equation (PDE) in (8) is discretized using a *forward time-center space* (FTCS) discretization method, which gives

$$\rho C(T_{i,j}^k)\frac{T_{i,j}^{k+1} - T_{i,j}^k}{\Delta t} = \tilde{k}(T_{i,j}^k)\left[\frac{T_{i-1,j}^k - 2T_{i,j}^k + T_{i+1,j}^k}{(\Delta x)^2} + \frac{T_{i,j-1}^k - 2T_{i,j}^k + T_{i,j+1}^k}{(\Delta z)^2}\right], \tag{10}$$

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 discretize the nonlinear boundary conditions (4)-(7). The time step Δt is chosen such that it obeys the limit of the FTCS discretization stability restrictions given by $\Delta t \leq min\left\{\frac{1}{2}\frac{\rho C(T_{i,j}^k)\Delta x^2 \Delta z^2}{k(T_{i,j}^k)(\Delta x^2 + \Delta z^2)}\right\}$.

3.2. Linear parameter-varying model derivation

The system model (1)-(7) is nonlinear and remains so after discretization. There exists several approaches to transform a nonlinear system represented by

$$x(k+1) = F(x(k), u(k)), \tag{11}$$

into a linear model. A well known approach is the Jacobian linearization of (11) around an equilibrium trajectory. The resulting linear system will then only describe the local behavior around that trajectory. Another approach is based on rewriting (11) 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 [18]. A discrete-time LPV model can be represented in state-space form as

$$\begin{bmatrix} x(k+1) \\ y(k) \end{bmatrix} = \begin{bmatrix} A(\theta(k)) & B(\theta(k)) \\ C(\theta(k)) & D(\theta(k)) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}, \tag{12}$$

where $y(k)$ represents the model output. We rewrite the nonlinear model into the LPV form (12), 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 $x(k) = [x_1(k),\ldots,x_{mm\times nn}(k)]^\top$. 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:

$$x(k) = [T^k_{1,1},\ldots,T^k_{mm,1},\ldots,T^k_{1,nn},\ldots,T^k_{mm,nn}]^\top. \tag{13}$$

Next, we define the scheduling variable vector in a similar fashion where $\theta_1(T^k_{i,j}),\ldots,\theta_4(T^k_{i,j})$ are derived so that (12) is affine in the scheduling variables,

$$\theta(k) = [\theta_1(T^k_{1,1}),\ldots,\theta_4(T^k_{1,1}),\ldots, \theta_1(T^k_{mm,nn}),\ldots,\theta_4(T^k_{mm,nn})]^\top. \tag{14}$$

Each scheduling variable in (14) is unique as the scheduling variables are functions of the local temperature at each unique spatial location. We make the definitions θ_1 to θ_4 noting that at each unique spatial location (x_i,z_j), these scheduling variables are unique as they are functions of the local temperature $T^k_{i,j}$.

In formulating (12), 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 (1)-(7). However, we see an exponential relationship between the number of scheduling variables and LMIs to solve in controller design. Our objective now becomes to reduce the model while preserving the original dynamics.

4. DEVELOPMENT OF A CONTROL-ORIENTED LPV MODEL USING PCA

The first step in developing a control-oriented LPV model of the RTP system is to reduce the number of scheduling variables through the use of principal component analysis (PCA) [22]. To apply PCA to the LPV scheduling variables data, one first needs to generate and collect data by means of measurements or simulations [23], such that the data covers all regions within the operating range. Given the LPV model (12) and assuming that the measurable signals have been sampled at time instants $k \in \{1,2,\ldots,K\}$, scheduling variables $\theta(k) \in \mathbb{R}^{\tilde{m}}$ are computed and collected in the following $\tilde{m}\times K$ 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 $\tilde{m}$ represents the actual number of scheduling variables and K denotes the number of data samples, with $K \geq \tilde{m}$. PCA is then applied by solving an eigenvalue problem for the covariance matrix. The covariance matrix is given by $\bar{C} = \frac{1}{K}\Theta_c\Theta_c^\top$, where $\Theta_c = \mathcal{C}(\Theta) = \Theta - \theta_{\text{mean}}$ 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 $\tilde{m}\times\tilde{l}$ matrix whose columns contain the $\tilde{l}$ eigenvectors associated with the first $\tilde{l}$ significant eigenvalues. The approximation of the actual variable $\hat{\theta}(k)$, corresponding to $\rho(k)$, can be easily computed as $\hat{\theta}(k) = \mathcal{C}^{-1}(V_m\rho(k))$, where $\mathcal{C}^{-1}(V_m\rho(k)) = V_m\rho(k) + \theta_{\text{mean}}$. Henceforth, we also drop the time index k for better readability and denote $\rho(k)$ and $\theta(k)$ simply as ρ and θ, respectively. The PCA-based reduced LPV model can be represented as

$$\begin{aligned} x(k+1) &= \hat{A}(\rho)x(k) + \hat{B}(\rho)u(k), \\ y(k) &= \hat{C}(\rho)x(k) + \hat{D}(\rho)u(k). \end{aligned} \tag{15}$$

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 [24]

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

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 note that [24]

$$\hat{Q}(\rho) = Q(\hat{\theta}) = Q_0 + \sum_{i=1}^{\tilde{m}} Q_i\hat{\theta}^i = \hat{Q}_0 + \sum_{j=1}^{\tilde{l}} \hat{Q}_j\rho^j, \tag{16}$$

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. Equation (16) is a reduced model which is also affine in the reduced scheduling variables ρ.

5. PROPER ORTHOGONAL DECOMPOSITION FOR LPV MODEL ORDER REDUCTION

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 [25]. Essentially, POD extracts a set of *orthonormal basis functions* (OBFs) [26], usually with a few modes [25]. To approximate the function of interest over a domain, we write the ensemble into coefficients to be determined as $T(x,z,t) \approx \widehat{T}(x,z,t) = \sum_{j=1}^{M} \alpha_j(x,z)\varphi_j(t)$, where α_j's define the set of OBFs, and φ_j's denote the time-dependent coefficients. We employ the method of snapshots [27], which solves an eigenvalue problem and only requires an ensemble of appropriately organized data points [26]. Here we define $\tilde{D} = mm \times nn$. The data needed is captured as $T_{snap} \in \mathbb{R}^{\tilde{D}\times K}$, where $\tilde{D}$ corresponds to the number of discretization points and K corresponds to the number of snapshots. In the finite-dimensional case, POD reduces to a singular value decomposition (SVD) problem as

$$\begin{aligned} T_{snap} &= \Phi \Sigma V^{\top} \\ &= \begin{bmatrix} \Phi_r & \Phi_s \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 & 0 \\ 0 & \Sigma_s & 0 \end{bmatrix} \begin{bmatrix} V_r^{\top} \\ V_s^{\top} \end{bmatrix}. \end{aligned} \tag{17}$$

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 the latter equation, $\Phi \in \mathbb{R}^{\tilde{D}\times\tilde{D}}$ 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. These basis functions, called POD modes, are used to obtain accurate low-order dynamic models via Galerkin projection [26].

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 (18). A high percentage of energy preserved is 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]

$$PEP = 100 \times \frac{\sum_{i=1}^{M} \sigma_i^2}{\sum_{i=1}^{\tilde{D}} \sigma_i^2}, \tag{18}$$

where we note that M is the user's choice, and $\tilde{D}$ is the original order of the state-space system. To obtain the reduced-order state-space LPV model, (15) is multiplied from both sides by the truncated orthonormal matrix $\Phi_r \in \mathbb{R}^{\tilde{D}\times M}$ as

$$\Phi_r^{\top} x(k+1) = \Phi_r^{\top} \hat{A}(\rho) x(k) + \Phi_r^{\top} \hat{B}(\rho) u(k). \tag{19}$$

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

$$x_r(k) = \Phi_r^{\top} x(k). \tag{20}$$

Since each element of $x_r(k)$ is a linear combination of the elements of $x(k)$, substituting (20) into (19) yields

$$\begin{aligned} x_r(k+1) &= A_r(\rho)x_r(k) + B_r(\rho)u(k) \\ y_r(k) &= C_r(\rho)x_r(k) + D_r(\rho)u(k), \end{aligned} \tag{21}$$

with $A_r(\rho) = \Phi_r^{\top}\hat{A}(\rho)\Phi_r$, $B_r(\rho) = \Phi_r^{\top}\hat{B}(\rho)$, $C_r(\rho) = \hat{C}(\rho)\Phi_r$, $D_r(\rho) = \hat{D}(\rho)$.

6. LINEAR PARAMETER-VARYING CONTROLLER DESIGN

In this section, gain-scheduled controller synthesis using the reduced LPV model is presented.

6.1. Polytopic LPV model

The low-order, low-scheduling variable model in (21) can be converted to a polytopic LPV model where the parameter dependent matrices of (21) are to be determined and $\rho(k)$ is the scheduling parameter vector. The LPV model can be represented as a linear input-output map

$$P(\rho) = \begin{bmatrix} A_r(\rho) & B_r(\rho) \\ C_r(\rho) & D_r(\rho) \end{bmatrix}. \tag{22}$$

Introducing the compact set $\mathcal{P}_\rho \subset \mathcal{R}^{\tilde{l}}$: $\rho(k) \in \mathcal{P}_\rho, \forall k > 0$, this set is then the polytope defined by the convex hull $\mathcal{P}_\rho := Co\{\rho_{v_1}, \rho_{v_2}, ..., \rho_{v_{n_v}}\}$, where $n_v = 2^{\tilde{l}}$ is the number of vertices. Next, we note that the system is parameter-affine since the state-space matrices depend affinely on the scheduling parameters as

$$P(\rho) = P_0 + \sum_{i=1}^{\tilde{l}} \rho_i P_i = P_0 + \rho_1 P_1 + \cdots + \rho_{\tilde{l}} P_{\tilde{l}}. \tag{23}$$

Since any $\rho(k)$ can be expressed as a convex combination of n_v vertices, this is called a polytopic LPV system where $\mathcal{P}(\rho) \in Co\{\rho_{v_1}, \rho_{v_2}, ..., \rho_{v_{n_v}}\} = \sum_{j=1}^{n_v} \xi_j P_j$, where $\sum_{j=1}^{n_v} \xi_j = 1$, $\xi_j \geq 0$ are the convex coordinates, and P_j's are calculated at corresponding vertex [26]. We note that since (21) is the reduced-order model, the low number of scheduling variables allows for tractability in using standard solvers in order to design controllers.

6.2. LPV controller synthesis

The controller design configuration for the RTP system is shown in Fig. 2. The polytopic gain-scheduled controller, $K(\rho)$, is designed based on the low-order, low-scheduling variables LPV model $P(\rho)$. We design $K(\rho)$ such that the closed-loop system meets an induced $\mathcal{L}_2$ gain performance with the gain-scheduled controller described as

$$K(\rho): \zeta(k+1) = A_K(\rho)\zeta(k) + B_K(\rho)e(k),$$

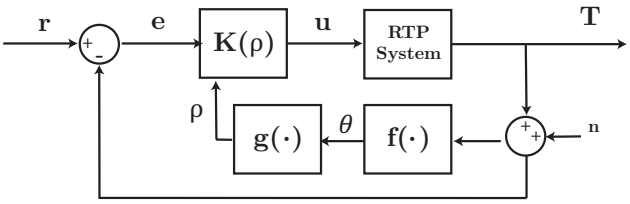

Fig. 2. Control of the RTP system using an LPV controller $K(\rho)$ designed based on the reduced LPV model.

$$u(k) = C_K(\rho)\zeta(k) + D_K(\rho)e(k). \tag{24}$$

In Fig. 2, r denotes the reference input, u the controller output, T the temperature output of the RTP system, and n the noise introduced into the temperature measurement. We see from Fig. 2 that the temperature measurement is used to calculate the original scheduling variables θ. We then calculate ρ to send to the controller.

This control methodology is specific to LPV models with an affine dependence on the scheduling variables ρ that vary within a fixed polytope. Next, we convert Fig. 2 into the standard lower-fractional transformation (LFT) where w is the external disturbance, z denotes the controlled outputs, u represents the controller outputs, and y is the measurements. From [28], we define the induced $\mathcal{L}_2$ gain performance as

$$\|T_{wz}\|_{i,2} = \sup_{\rho} \sup_{w \neq 0} \frac{\|z\|_{\mathcal{L}_2}}{\|w\|_{\mathcal{L}_2}}, \tag{25}$$

where i denotes the selected induced norm. The gain indicates the worst-case output energy $\|z\|_{\mathcal{L}_2}$ over all bounded energy disturbances $\|w\|_{\mathcal{L}_2}$ for all admissible values of the scheduling variables ρ. The closed-loop LPV system of the LFT system has an induced $\mathcal{L}_2$ gain performance less than γ if there exists a symmetric positive-definite matrix X such that a linear matrix inequality (LMI) problem has a feasible solution for all admissible trajectories of ρ [28]. From the LMI problem, it is easy to see why reducing the number of scheduling variables is critical in LPV applications as the matrix X is found by solving a finite number of LMIs. We see from the proof in [28] that for polytopic LPV representations with affine dependence on ρ, the underlying LMI condition holds for all trajectories of ρ within the polytope, if it holds true at the vertices.

7. SIMULATION RESULTS AND DISCUSSION

In this section, we present results from the first principles-based modeling and use MATLAB/Simulink to demonstrate the closed-loop simulation results.

7.1. Validation of the affine LPV Model

First, we followed our recent work [17] and chose a discretization of 10 steps in the x direction and 3 steps in the z direction. We then derived a high-order, high-scheduling variable affine LPV model where the states correspond to the finite discretization nodes. The discretization step size was considered to be $\Delta_t = 9 \times 10^{-5}$. The original model consisted of 30 states and 73 scheduling variables. The coefficients F_1, F_2, and α were tuned to match the open-loop response of the STEAG RTP system in [5].

7.2. Scheduling variables reduction

Next, we proceed with the analysis by generating a low-frequency multi-layered sinusoidal input to excite the dynamics of the system model. Using this input, we proceeded to follow the PCA analysis given in Section 4 We chose to map to three scheduling variables which retain 95.6% of accuracy. However, this is at the tradeoff of model insight, as the reduced-order scheduling variables no longer have physical meaning.

Remark 1: The LPV framework is a natural framework for RTP modeling since temperature is a readily measurable scheduling variable and the nonlinearities seen in the model are smooth.

Remark 2: All of the scheduling variables share the common thread in that they are all functionals of temperature.

7.3. Model order reduction

Next, we used the POD method as described in Section 5 to reduce the order of the LPV model. Using (18)-(21), we reduced the high-order model to a second order system while preserving 98% of the energy.

Remark 3: Since RTP wafer recipes are known *a priori*, we can expect to preserve a large amount of energy in the reduced-order model to create a low-order system computationally inexpensive enough to be run in real time.

Fig. 3 shows the comparison between the low-order, low-scheduling variable model with the original nonlinear model. The low-order system was chosen to have two POD modes, and three scheduling variables. Good agreement is observed around the operating temperatures.

7.4. LPV controller design for RTP system

The RTP control design objective is to track the reference trajectory r. For RTP systems, this reference temperature is a ramp function with a large soak time. A controller must provide a good steady state tracking, with preferably zero steady-state error. Overshoot beyond a few degrees is unacceptable, actuator saturation is plausible, and fast ramp rates are needed.

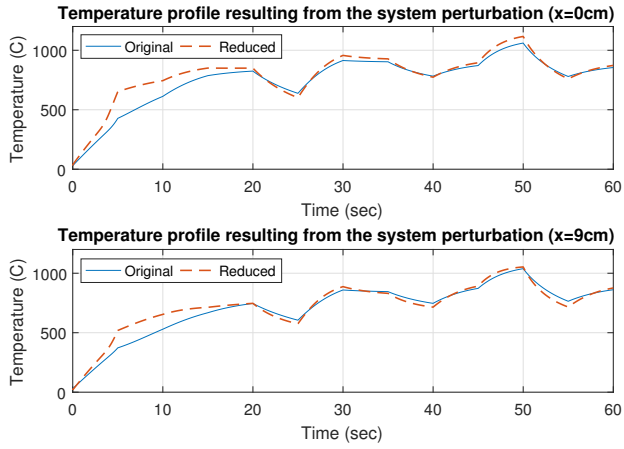

Fig. 3. Comparison between the low-order, low-scheduling variable model with the original nonlinear system. The top plot shows a node at the middle of the wafer, while the bottom plot shows a node at the edge of the wafer.

To accomplish this, we add loop-shaping filters as shown in Fig. 4. We add a first-order low-pass filter W_e to impose tracking requirements, and W_u as a first-order high pass filter to penalize the control effort on each zone output. The filters are selected and tuned by trial and error, seeking the minimization of the induced $\mathcal{L}_2$ gain from the external input vector $w = [r\ n]^\top$ to the controlled outputs $z = [z_e\ z_u]^\top$ in order to enforce the performance requirement.

For controller synthesis that is based on polytopic LPV models, the plant input and output matrices, B and C, need to be independent of the scheduling variables [28]. This is not the case for (21). Therefore, to put (21) in a suitable form for controller synthesis we filter the input with a low-pass filter of suitable bandwidth. We make note that C and D are not functions of ρ although (21) has been formulated in a general sense.

7.5. Closed-loop simulation results

To simulate a realistic RTP system, we assume that we only have a single temperature measurement at the middle of the wafer which is measured in real time. We make the assumption that all scheduling variables can be calculated from the single temperature measurement since the wafer is desired to be at a uniform temperature. Next, we design $K(\rho)$ to track the temperature measurement. The closed-loop simulation results are given in Fig. 5. For simulation purposes, we add noise to the temperature measurement to impose a 15 dB signal-to-noise ratio (SNR). As an input reference trajectory we impose a 250°C ramp with a steady-state of 700°C. The controller tracks the temperature measurement with only 12°C of overshoot.

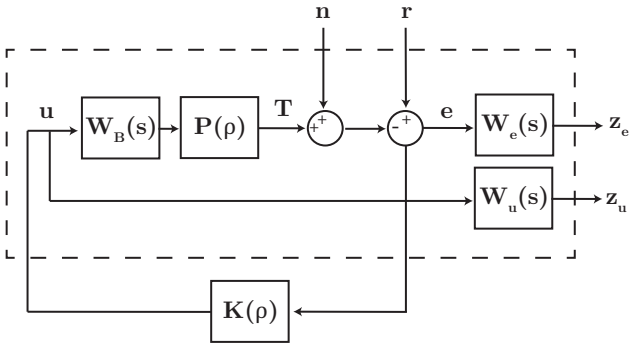

Fig. 4. Generalized configuration of the closed-loop system composed of the reduced LPV model, LPV controller, and loop-shaping filters.

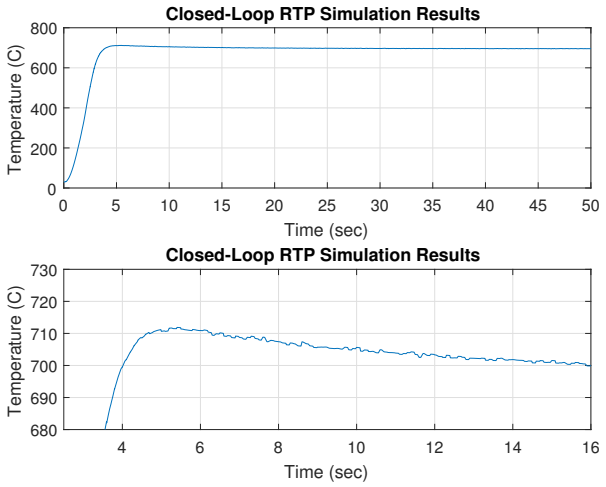

Fig. 5. Closed-loop simulation results for tracking a reference temperature (top: entire region; bottom: close-up on overshoot).

In the closed-loop simulation, saturation blocks are used to enforce that the lamps input ranges between 0-100%. Fig. 6 shows the controller outputs. Given the saturation of zones 1 and 2, we also design a suitable anti-windup scheme to prevent integrator windup.

Using LPV modeling and gain-scheduled (GS) control presents a systematic first principles-based procedure for modeling and design. Should the wafer material properties or RTP chamber geometry change, the dynamics of the model can be updated as well. This is in comparison to purely data driven techniques, e.g., [5], where data is generated from a single wafer setup.

7.6. Comparison with existing design methods

As a comparison between the performance of the designed LPV controller with that of existing design meth-

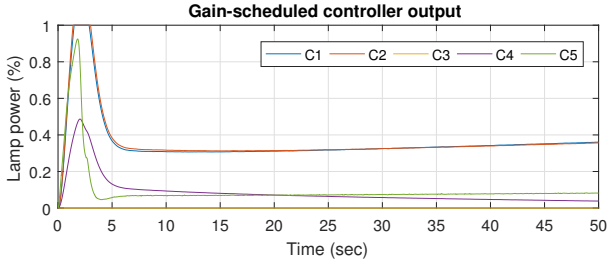

Fig. 6. Controller outputs of the gain-scheduled controller $K(\rho)$ where C_1 through C_5 denote controller outputs corresponding to five zones.

ods so far examined in the literature for RTP system, we first design a PID controller to control the reduced-order LPV model. The input to the PID controller is the temperature error and the output is a global lamp power percentage. Since the lamps in the RTP system are divided into different zones, as a solution to better wafer uniformity for single-input, single-output (SISO) PID controllers, the literature has suggested different zone power percentages. Ultimately, a lamp's final power percentage is defined to be the product of the global PID controller's output with the local zone power percentage. We use the final zone power percentages presented in [5] and tune the PID controller gains using MATLAB/Simulink's PID toolbox. We also implement an anti-windup scheme given that there exists controller saturation.

In addition to PID controllers, we also used the results given in [5] to compare with the performance of a nonlinear model predictive control (NMPC) approach. A numerical comparison between all methods is shown in Table 1 and illustrates an overall superior performance that the LPV controller offers compared to the other two controllers. We recall here again our goal that is to minimize the thermal budget of each wafer. Table 1 is hence hierarchically arranged from most important to least of closed-loop performance criteria. The overall improvement translates into a lower thermal budget, which then leads to lower heating costs and higher product quality.

To further demonstrate that the gain-scheduled LPV controller provides a reduced thermal budget and improved response, Fig. 8 shows the error responses of the corresponding closed-loop systems for LPV controller, as well as the two PID controllers.

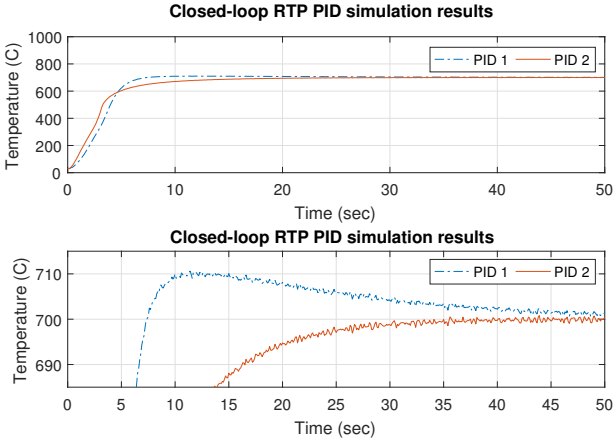

Fig. 7. Closed-loop response of the RTP system with two PID controllers: PID 1 allows overshoot while PID 2 does not (top: entire region; bottom: close-up).

Table 1. Closed-loop performance comparison between the gain-scheduled controller designed using the LPV model, the nonlinear MPC controller designed using the data-driven model of [5], and the well-tuned PID controllers, where PID 1 allows overshoot and PID 2 does not allow.

	GS	NMPC	PID 1	PID 2
Rise Time (sec)	4	∼7	7.35	37.5
Overshoot (°C)	12	∼30	10	0
Time to ± 1°C (sec)	15.2	∼18	49.3	30.5

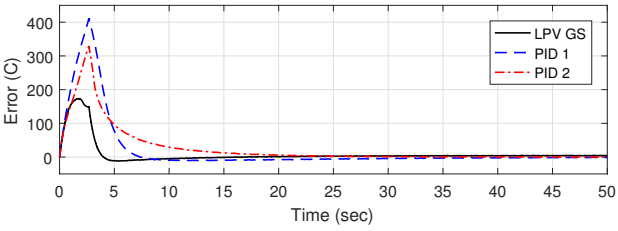

Fig. 8. Comparison between the error signals in the closed-loop response of the RTP systems using three controllers each tracking the same ramp input. The LPV GS controller clearly shows a better performance in terms of reaching the target much faster with less than half overshoot compared to the two PID controllers.

8. CONCLUDING REMARKS

In this paper, we used a well-established first principles-based modeling approach to develop an affine LPV model for rapid thermal processes. We further employed PCA and POD to reduce the dimensionality 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 control synthesis. Finally, we designed a gain-scheduled controller to track a temperature reference profile. Closed-

loop simulation results showed overall improvement in tracking performance (overshoot and settling time) compared to the PID and MPC control design approaches.

REFERENCES

[1] 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*, pp. 3910-3921, 2004.

[2] Y. J. Lee, B. Khuri-Yakub, and K. C. Saraswat, "Temperature measurement in rapid thermal processing using acoustic techniques," *Review of scientific instruments*, vol. 65, no. 4, pp. 974-976, 1994.

[3] J. Liu and Y.-S. Chen, "Simulation of rapid thermal processing in a distributed computing environment," *Numerical Heat Transfer: Part A: Applications*, vol. 38, no. 2, pp. 129-152, 2000. [click]

[4] F. Fasching, S. Halama, and S. Selberherr, *Technology CAD systems*. Springer Science & Business Media, 2012.

[5] 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. [click]

[6] 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.

[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] A. Theodoropoulou, R. A. Adomaitis, and E. Zafiriou, "Model reduction for optimization of rapid thermal chemical vapor deposition systems," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 85-98, 1998.

[9] 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. [click]

[10] D. De Roover, A. Emami-Naeini, J. L. Ebert, and R. L. Kosut, "Trade-offs in temperature control of fast-ramp RTO and RTA systems," in *7th Intl Conf. on Advanced Thermal Processing of Semiconds, RTP*, vol. 99. Citeseer, 1999.

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

[12] 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.

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

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

[15] 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. [click]

[16] 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. [click]

[17] M. Trudgen, S. Z. Rizvi, and J. Mohammadpour, "Linear parameter-varying approach for modeling rapid thermal processes," *Proc. of American Control Conference (ACC)*, IEEE, pp. 3243-3248, 2016. [click]

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

[19] 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.

[20] V. Borisenko and P. J. Hesketh, *Rapid Thermal Processing of Semiconductors*, Springer Science & Business Media, 2013.

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

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

[23] A. Kwiatkowski and H. Werner, "PCA-based parameter set mappings for LPV models with fewer parameters and less overbounding," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 4, pp. 781-788, 2008. [click]

[24] S. Rizvi, J. Mohammadpour, R. Tòth, , and N. Meskin, "A kernel-based PCA approach to model reduction of linear parameter-varying systems," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 5, pp. 1883-1891, 2016. [click]

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

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

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

[28] P. Apkarian, P. Gahinet, and G. Becker, "Self-scheduled H_∞ control of linear parameter-varying systems: a design example," *Automatica*, vol. 31, no. 9, pp. 1251-1261, 1995.

Mark Trudgen received his B.S. and M.S. degrees in mechanical engineering and his Ph.D. degree in engineering. He joined the University of Georgia as a research assistant in the Complex Systems Control Laboratory under the direction of Dr. Velni in August 2013. In January 2017 he joined the University of Georgia as a lecturer in the school of electrical and computer engineering. His current research is in low-order LPV model development and robust control design for automotive and manufacturing applications.

Javad Mohammadpour Velni received his B.S. and M.S. degrees in electrical engineering and his Ph.D. degree in mechanical engineering. He joined University of Georgia as an assistant professor of electrical engineering in Aug. 2012. He has published over 100 articles in international journals and conferences, served in the editorial boards of ASME and IEEE conferences on control systems and edited two books on large-scale systems (2010) and LPV systems modeling and control (2012). His current research is on secure control of cyber physical systems, coverage control of multi-agent systems, and data-driven model learning and control of complex distributed systems.