Chemical Engineering Research and Design 202 (2024) 1–11

Contents lists available at ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

Sparse identification modeling and predictive control of wafer temperature in an atomic layer etching reactor

Feiyang Ou [a], Fahim Abdullah [a], Henrik Wang [a], Matthew Tom [a], Gerassimos Orkoulas [c], Panagiotis D. Christofides [a,b,*]

[a] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA
[b] Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA
[c] Department of Chemical Engineering, Widener University, Chester, PA 19013, USA

ARTICLE INFO

Keywords:
Atomic layer etching
Radiative heating lamps
Sparse identification modeling
Model predictive control
Computer aided engineering

ABSTRACT

To address the escalating demand and supply constraints for semiconductor devices, manufacturing techniques such as thermal atomic layer etching (ALE) have been utilized, which require robust temperature control systems. Radiative heating with lamps is a promising method for achieving fast and precise temperature control. However, there is limited research in analyzing the dynamic control of radiative lamp heating systems. In this study, a transient, two-dimensional (2-D) radiative heating model in a cross-flow thermal ALE reactor is constructed through Ansys Fluent. A sparse identification (SINDy) modeling approach is applied to build a reduced-order dynamic model for the prediction of system states in the control system. A model predictive controller (MPC) is then developed to bring the wafer surface temperature to the target temperature while preserving the uniformity of the temperature on the wafer surface. Control is implemented by adjusting the powers of three groups of heating lamps independently. Notably, a novel feedback-based time-varying steady-state penalty approach is applied with the MPC in this study, which enables the system to reach the target temperature range within 10 s while maintaining temperature uniformity for 1000 s.

1. Introduction

Over the last decade, there has been a rising demand for high-performance semiconductor chips such as 5-nm fin-field effect transistors (FinFETs) that are fabricated through advanced processes like extreme ultraviolet (EUV) lithography (Wang et al., 2020). Due to their wide usage, there is a potential for a global supply shortage of advanced semiconductor chips, which are difficult to fabricate from silicon-based materials and consist of over 500 processing steps (Espadinha-Cruz et al., 2021). A considerable portion of these processing steps requires thermal processing under high temperatures, such as thermal atomic layer etching (ALE), to ensure conformal fabrication of these downscaled devices. Thermal ALE is a top-down fabrication procedure that improves the surface uniformity and alignment of transistors by employing a system of reactions that produce volatile byproducts to enable surface etching. However, to ensure that complete byproduct removal is observed, the specific optimal operating temperature depends on the material of the substrate. For instance, aluminum oxide, Al_2O_3 requires high operating temperatures of 573 K, which has been proposed experimentally (Lee et al., 2016) and studied in prior *in silico* modeling work (Yun et al., 2022a). In addition to reaching the chosen operating temperatures, it is important that the temperature of the substrate surface is uniform. This ensures that the self-limiting surface reactions on the wafer will finish concurrently and minimize the processing time.

Prior work has extensively focused on the *in silico* modeling of thermal ALE processes through various reactor configurations (Yun et al., 2022b,c), during which a constant and uniform temperature, both on the substrate surface and operating environment, are assumed. This assumption overlooks the possibility of temperature fluctuations both on the wafer surface and in the overall reactor, which is widely reported in both laboratory experiments and numerical simulations (Gyurcsik and Sorrell, 1991). Furthermore, these studies typically neglect the transient and dynamic nature of energy transfer among the wafer, reactor, and external environment. However, realistic thermal ALE processes often experience disturbances that disrupt the standard operating conditions of the process and introduce surface defects that result in product nonconformance, performance degradation, and lower productivity, depending on the magnitude of the disturbance. For example, a temperature fluctuation of 5 K will substantially change the half-cycle time, as observed in Yun et al. (2022a). Therefore, precise temperature

* Corresponding author at: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA.
E-mail address: pdc@seas.ucla.edu (P.D. Christofides).

https://doi.org/10.1016/j.cherd.2023.12.024
Received 30 November 2023; Received in revised form 12 December 2023; Accepted 14 December 2023
Available online 16 December 2023

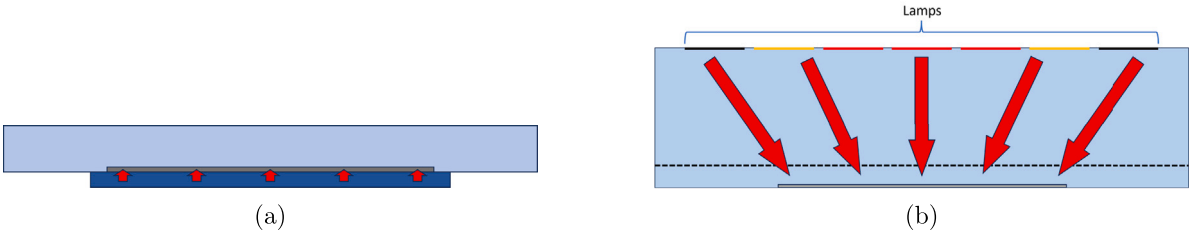

Fig. 1. Schematic figure that compares the (a) conventional heating plate structure and (b) novel radiative heating structure, where red arrows represent the direction of energy transfer.

control is necessary to maintain the desired operating temperatures uniformly across the substrate surface for thermal ALE processes. In order to achieve a comprehensive understanding of the ALE process under practical industrial settings, it is critical to develop a transient model that characterizes the physical heating process of the wafer.

Using heating lamps as the thermal source for a reactor is a popular and widely employed method in industrial practices (Roozeboom and Parakh, 1990). Unlike a traditional heating plate where the primary heating mechanism is conduction, the primary heating mechanism for a heating lamp is thermal radiation, as shown in Fig. 1.

An advantageous feature of radiative heating is the high instantaneous changing rate in temperature that is beneficial for achieving fast controller response to mitigate disturbances. For instance, Gyurcsik and Sorrell (1991) determined that radiative heating can supply a maximum temperature rate of 200 K/s, which has been validated in *in silico* modeling work by Theodoropoulou et al. (1998). Radiative heating also achieves better temperature uniformity on the wafer surface that is maintained by individually adjusting lamp powers, where in most industrial and experimental contexts, multiple groups of lamps are involved, and each group is independently controlled. Additionally, the temperature uniformity resulting from radiative heating is beneficial for the modern downscaling of semiconductors (Dassau et al., 2006). By combining the quick heating capabilities and improved temperature uniformity, radiative heating reduces the process time of the ALE processes by decreasing the time required to preheat the system and by providing more flexible control options (Timans, 1998).

To manifest the superior performance of radiative heating, the geometrical design of the reactor is crucial. Various geometric configurations of lamp groups have been explored in prior works: a parallel configuration of individual lamps positioned directly above the wafer (Gyurcsik and Sorrell, 1991), a ring structure with varying radii above the wafer (Oh et al., 2009), and a structure where there are lamps both above the wafer and on the side of the wafer (Theodoropoulou et al., 1998). Baker and Christofides (1999) also proposed a complete configuration with circular lamps positioned above, on the side, and beneath the wafer in a quartz process chamber. With the aforementioned reactor configurations, the temperature uniformity is significantly improved with individually controlled lamp power.

Nevertheless, prior studies have been primarily concentrated on optimizing the power supplied to the lamps to achieve a uniform temperature profile in a steady-state condition after thousands of seconds, which is similar to an open-loop control strategy (Gyurcsik and Sorrell, 1991; Cho et al., 1994; Theodoropoulou et al., 1998; Oh et al., 2009). There is limited work focused on transient temperature analysis and real-time control. Due to the dynamic nature of industrial production and the prolonged time to reach steady-state, there is a greater need for dynamic temperature and uniformity control, especially for systems that fabricate wafers with diameters exceeding 200 mm, that operate at temperatures below 600 K, and that are sensitive to temperature changes (Roozeboom and Parakh, 1990). Prior works such as Schaper et al. (1994) and Baker and Christofides (2000) discussed the transient model and output feedback control of radiative heating processes, but with simpler reactor configurations and analytical transport equations. The analytical equations only work for ideal models that lack complex geometries and configurations, and for a modern industrially applicable system, a numerical computation model is required. To model the complex reactor accurately, in this study, a transient computational fluid dynamics (CFD) model incorporating physical heating devices is introduced. With the proposed model, the dynamic behavior of the temperature on the wafer surface and the ensuing energy transfer between the wafer, reactor, and surrounding environment can be analyzed. The development of a dynamic model facilitates the application of a model predictive controller (MPC) for real-time feedback control over the wafer surface temperature to optimize the dynamic performance of the reactor. The objective of the controller is to conduct a rapid elevation of the wafer surface temperature from room temperature to the desired value, and then maintain the wafer surface temperature within a user-specified range around the target. Additionally, the controller is expected to preserve optimal surface temperature uniformity. Such a controller conforms with evolving demands of precise thermal control in *state-of-the-art* semiconductor technology.

The organization of this work is as follows: the computational fluid dynamics model setup is discussed in detail in Section 2, the dynamic modeling by the sparse identification method is examined in Section 3, the formulation of the model predictive controller is elucidated in Section 4, and the *in silico* experimental results are provided in Section 5.

2. Transient modeling of radiative heating in ALE

This section analyzes the development of the radiative heating model conducted by heating lamps. In this work, the 2-D model for the radiative heating ALE reactor is first created. Following the construction of the 2-D reactor, a meshing procedure is performed to discretize the geometry into an appropriate mesh with sufficient quality to ensure numerical accuracy and efficiency. Lastly, the macroscopic computational fluid dynamics (CFD) simulation that is integrated with the radiation model is conducted to generate spatiotemporal temperature data on the wafer surface.

2.1. 2-D reactor model and mesh generation

In this work, the macroscopic modeling framework begins by constructing the thermal ALE reactor geometry in computer aided design (CAD) software Ansys SpaceClaim. In order to conserve computational resources, a 2-D symmetrical modeling approach is adopted, which is illustrated in Fig. 2 . The wafer, with a diameter of 300 mm, is positioned at the base of the reactor, which has a cross-flow configuration (Elers et al., 2006). Three distinct groups of lamps, the center, edge, and side lamp groups, are arranged above the reactor, and supplies heat to the surrounding environment in the form of thermal radiation. The center lamp group consists of three lamps, while the edge and side groups consist of two lamps each. The output power of each lamp group is independently adjustable, which provides a flexible control scheme with the potential to precisely modulate the wafer surface temperature. A quartz boundary with a small gap distance is positioned between the wafer and lamps, which serves as a physical barrier for internal gas flow

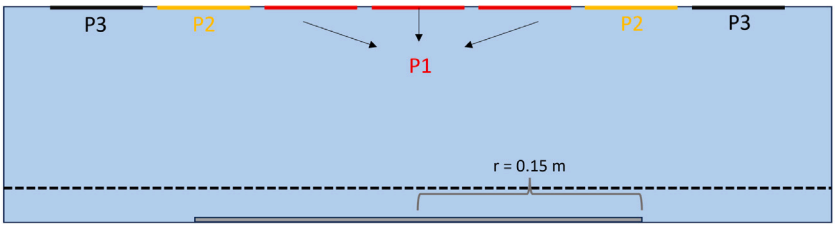

Fig. 2. Schematic diagram of the 2-D thermal ALE, cross-flow reactor: The red lines are the Center lamp group; the yellow lines are the Edge lamp group; the solid black lines are the Side lamp group; the dashed black line is the quartz window; the gray rectangle part is the wafer. A comprehensive list of geometry parameters are displayed in Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The mesh quality acceptability criteria range and mesh parameters calculated from Ansys Fluent for the ALE reactor. For the orthogonality, the minimum value is presented. For the aspect ratio, the maximum value is presented.

Quality indicator	Orthogonality	Skewness	Aspect ratio	Number of cells
Criteria	0.001 ~ 1[a]	0[a] ~ 0.95	1[a] ~ 8	N/A
ALE reactor	0.998	0.002	1.46	28 180

[a] Desired value for ideal mesh quality.

while allowing light and energy transfer across the boundary. Additionally, external radiation from the room temperature environment is also defined as boundary conditions into the model to reflect the natural cooling process in a practical scenario (Theodoropoulou et al., 1998).

The procedure to divide the CAD model into discrete cells for numerical calculation is called meshing, and the mesh quality plays a significant role in the accuracy, convergence, and stability of the numerical CFD simulation. In this work, Ansys Workbench Meshing is applied to create a discretized 2-D grid that is composed of rectangular cells with maximum cell distances of 0.5 mm. The mesh quality parameters, orthogonality, skewness, and aspect ratio, for rectangular cells in this work is within the acceptable range by the criteria recommended by ANSYS (2022) in all phases and boundaries, and it is summarized in Table 1. Additionally, the reactor was discretized into 28,000 cells to balance the computational efficiency and accuracy of the CFD simulation.

2.2. Macroscopic energy model

The computational fluid dynamics (CFD) simulation is conducted in the multiphysics software, Ansys Fluent. Since the energy transfer and surface temperature profile are major concerns in this study, only the heat transfer equation is calculated to save computational resources. The conservation of energy equation is described as follows:

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = -\nabla(\Sigma h_j J_j) + S_h \quad (1)$$

where ρ is the density of the fluid, E is defined as the internal energy of the system, $\vec{v}$ is the fluid velocity, p is the system pressure, h_j is the sensible enthalpy of the gas-phase species j, J_j is the diffusion flux of gaseous species j, and S_h is the heat transfer source flux rate.

The convective heat transfer originated from gas flow is also considered as a boundary condition on wafer. As demonstrated in a prior work (Yun et al., 2022a), the gas flow in the thermal ALE process is laminar and has small surface flow velocities for small reactor volumes (Ponraj et al., 2013). As a result, the convective heat transfer coefficient on the wafer is approximated as a fixed value where forced convection is negligible. A complete list of the parameters and constants used in the CFD model is outlined in Table 2. As the quartz window is thin, the amount of light absorbed by the quartz glass is neglected for simplicity, which implies that the quartz window is assumed to be fully transparent.

Table 2
Parameters and constants of CFD model.

Parameter	Value	Units
Convective heat transfer coefficient	2.5	W/m^2/s
Convective flow temperature	520	K
Wafer diameter	300	mm
ALE reactor height	10	mm
ALE reactor diameter	460	mm
Gap distance	83	mm
Single lamp group width	60	mm
Interval between lamp groups	6	mm
Maximum lamp power	5000	W/m^2
Minimum lamp power	0	W/m^2
Quartz window thickness	1	mm
Quartz window heat capacity	964	J/kg/K
Quartz window absorption coefficient	0	m^{-1}
Quartz window thermal conductivity	1.67	W/m/K
Initial temperature	298	K

2.3. Discrete ordinate radiation model

Various radiation models are provided in Ansys to calculate radiation intensity. The discrete ordinate (DO) model is one of the most reliable models to simulate radiative heat transfer with optically-thin and transparent media like air and thin glass in this study (ANSYS, 2022). The DO model solves the radiative transfer equation (RTE) described in Eq. (2) to obtain a spatiotemporal solution of the radiation intensity I within any media that absorbs, emits, and scatters light. The DO model considers $\vec{s}$ as a field function, and the slightly modified RTE equation is defined by Eq. (3) below:

$$\frac{dI(\vec{r},\vec{s})}{ds} + (a+\sigma_s) I(\vec{r},\vec{s}) = an^2\frac{\sigma T^4}{\pi} + \frac{\sigma_s}{4\pi}\int_0^{4\pi} I(\vec{r},\vec{s}')\,\Phi(\vec{s}\cdot\vec{s}')\,d\Omega' \quad (2)$$

$$\nabla\cdot(I(\vec{r},\vec{s})\vec{s}) + (a+\sigma_s) I(\vec{r},\vec{s}) = an^2\frac{\sigma T^4}{\pi} + \frac{\sigma_s}{4\pi}\int_0^{4\pi} I(\vec{r},\vec{s}')\,\Phi(\vec{s}\cdot\vec{s}')\,d\Omega' \quad (3)$$

where $\vec{r}$ is the position vector, $\vec{s}$ is the direction vector, $\vec{s}'$ is the scattering direction vector, s is the path length, a is the absorption coefficient, n is the refractive index, σ_s is the scattering coefficient, σ is the Stefan–Boltzmann constant, I is the radiation intensity, T is the local temperature, Φ is the phase function, and Ω' is the solid angle. The items on the left-hand side of the conservation equation represent the sum of the spatial derivative of radiation intensity and absorption of light, while the items on the right-hand side of the equation represent the sum of the emission and scattering of light. A detailed description and guide to the calculation theories are further examined in ANSYS (2022).

The radiation-related parameters for the wafer, walls, and heating lamps are listed in Table 3. Additionally, radiation permeation is disabled for all wall boundaries except for the quartz window. For the DO method, it is assumed that the light beams transmit in a finite number of directions to make the numerical calculation possible. In this 2-D system, the discretization of the angular directions is implemented by

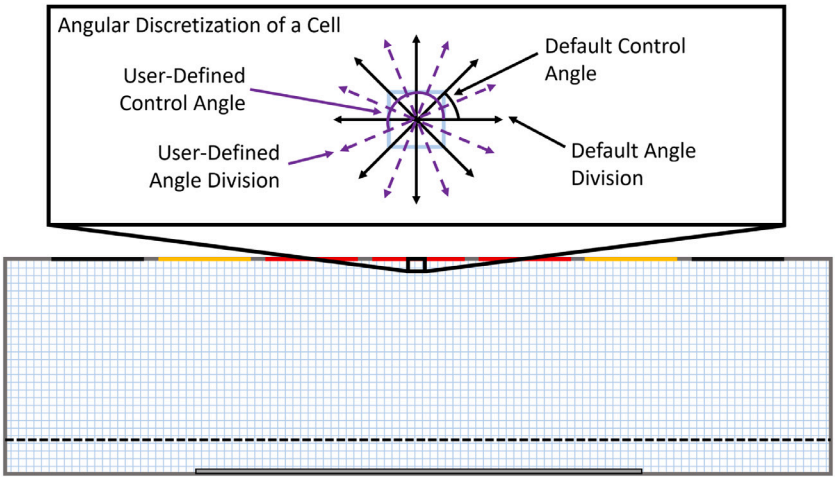

Fig. 3. Angular discretization of an adjacent, 2-D rectangular cell on a heating lamp, where control angles are generated to reflect the light emissions produced from the heating lamps.

Table 3
Parameters and constants of the radiation model.

Parameter	Value	Units
Wafer emissivity	0.7	N/A
Wafer diffuse fraction	0.3	N/A
Side wall emissivity	0.3	N/A
Side wall diffuse fraction	0.3	N/A
External radiation temperature	300	K
External radiation emissivity	1.0	N/A

separating a space of 360° into partitions, which are further separated into angular divisions that are called control angles. The default settings in Ansys Fluent divide the angular space into eight partitions with 45° intervals, and additional control angles can be specified to improve the accuracy of the numerical methods, as demonstrated in Fig. 3. The number of control angles defined by the user plays a substantial role in both the accuracy and computational complexity of the simulation. To find the optimal balance between numerical accuracy and computational efficiency, a grid search method is implemented through extensive test runs with the goal of finding the smallest number of control angles that will result in a minimal difference in simulation results when compared to test runs with more control angles. This work determined an optimized value of 8 for the number of user-defined control angles in each default partition zone, for a total of 64 control angles. However, the discretization may encounter issues in areas with irregular geometry and cause control angles to become overhanging angles, which results in unwanted reflection and refraction of light beams (ANSYS, 2022). To reduce the risk of overhanging angles appearing, rectangular cells are used in the meshing component of this work.

2.4. Implementation and monitoring

Given the transient and dynamic nature of the model, the controller frequently adjusts the lamp power to adapt to perturbations in the temperature on the wafer surface. Any changes in the lamp power are formulated as a linear interpolation function between the two consecutive power levels determined by the controller to avoid potential numerical error produced from abrupt changes to the lamp power and reflect the continuously changing nature of the lamps. The calculation of the lamp power at time t is expressed by the following equation:

$$P_i(t) = P_i(t_0) + \frac{t - t_0}{t_s}(P_i(t_1) - P_i(t_0)), \quad i = 1, 2, 3 \tag{4}$$

where $i = 1, 2, 3$, corresponds to the center, edge, and side lamp groups, respectively, $P_i(t)$ is the power of lamp group i at time t, $P_i(t_0)$ is the power of lamp group i at the start of the current sampling time period, $P_i(t_1)$ is the power of lamp group i determined by controller with feedback of temperature measured at t_0, and t_s is the length of one sampling time period.

To accurately represent the surface temperature profile, four temperature inspection positions are defined on the wafer surface. These locations include the center point ($r = 0$ m), the first trisection point ($r = 0.05$ m), the second trisection point ($r = 0.10$ m), and the edge point ($r = 0.135$ m), where r is the distance between the inspection points and the center of the wafer. The outermost 10% of the wafer is omitted from the region of interest due to inherent cooling effects on the periphery (Gyurcsik and Sorrell, 1991). The controller described in the following sections will use the temperatures monitored at these designated inspection points as feedback values to maintain the temperature within an acceptable range around the target temperature. For the purpose to regulate the process time of ALE within operable range determined by Yun et al. (2022a), a target temperature over 570 K is required on wafer surface, but not exceeding too much for saving energy cost. On the other hand, the acceptable range is determined by the criteria that boundary temperatures have less than 20% half-cycle time difference with target temperature. By microscopic simulation result conducted by Yun et al. (2022a), the acceptable range within ± 3 K fulfill the above criteria within temperature range of $[570, 580]$ K with half-cycle time around 0.8 s and 1.0 s. Consequently, 573 ± 3 K is determined as the goal for temperature control.

3. Sparse identification modeling

One key aspect in accurately predicting system behavior is a dynamic model that fully captures the time derivatives of the state variables. For simple process systems with basic kinetics, such as a continuous stirred tank reactor (CSTR) or a plug-flow reactor (PFR), first-principle models can be constructed using mass and energy balances (e.g., Fogler, 2006). However, these standard first-principle modeling procedures are unavailable for most complex chemical processes, such as the ALE reactor examined in this work. As a result, data-driven approaches have emerged as effective tools for modeling nonlinear processes; for example, there is the numerical method for subspace state-space system identification (Van Overschee and De Moor, 1994), the polynomial nonlinear state-space method (Alanqar et al., 2015), and machine learning methods like recurrent neural networks (Wu et al., 2019; Ren et al., 2022). Recently, a novel method for building dynamic models from open-loop data called sparse identification of

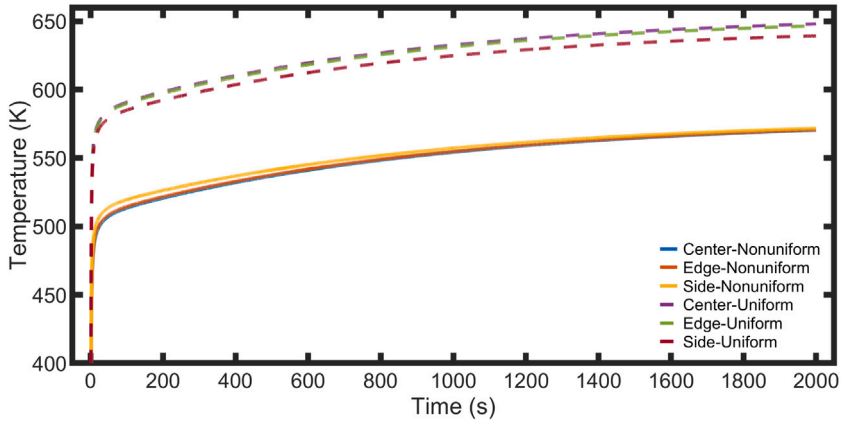

Fig. 4. Temperature profiles of sample open-loop datasets used for model training to illustrate the surface temperature uniformity for various lamp power configurations. Dashed lines are for the uniform 2000 W/m^2 case and solid lines are for the nonuniform case where $P_1 = P_2 = 450$ W/m^2, $P_3 = 4000$ W/m^2. The side surface temperature is over 10 K lower than center temperature at $t = 2000$ s in the case of uniform lamp power; in contrast, nonuniform lamp power results in better surface temperature uniformity.

nonlinear dynamics (SINDy) has gained attention for its effectiveness in several complex engineering systems (e.g., Proctor et al., 2014), such as modeling a nonisothermal CSTR reactor (Abdullah and Christofides, 2023) and fluid vortex shedding behind a cylinder (Brunton et al., 2016). Of which, the latter case is a nonlinear system whose underlying dynamics were identified after nearly 30 years by field experts. SINDy has also been used directly with input/output data (known as generalized SINDy or GSINDy) and the Kalman filter to construct time-variant digital twin models (Wang et al., 2022). The Kalman-GSINDy approach was subsequently used along with the proper orthogonal decomposition (POD) to find the lifting functions to find reduced-order Koopman linear models for incorporation into MPC. SINDy is a data-driven method that utilizes discrete measurements from a physical system to identify a first-order ordinary differential equation (ODE) of the following form:

$$\dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) \tag{5}$$

where $\tilde{x}(t)$ represents the state vector of the SINDy model, and $f(\tilde{x}(t), u(t))$ is a function of state and input variables that captures the dynamics of the underlying physical laws that govern the system. The SINDy model is applied with the goal of system identification; of note, this does not necessarily mean accurately representing the true underlying physics of the process. The key assumption of the SINDy algorithm is that $f(\tilde{x}(t), u(t))$ is sparse and only has a few nonlinear terms. Although there is a large pool of candidate nonlinear functions, only a few terms are expected to be active and contribute to system dynamic identification. The sparsity of the SINDy model facilitates calculating the nonzero coefficients, while also preventing overfitting.

3.1. Open-loop data generation

The training, or fitting, of the SINDy model requires measurements of three quantities: the state variables, the manipulated inputs, and the time-derivative of the state variables. The state variables in this study are the temperature inspection points on the wafer: T_1 at $r = 0.00$ m, T_2 at $r = 0.05$ m, T_3 at $r = 0.10$ m, and T_4 at $r = 0.135$ m, where r is the distance from the center of the wafer. The control variables are the power settings for the three lamp groups, P_1, P_2, and P_3, as depicted in Fig. 2.

The training data is generated via open-loop simulations that heat the reactor and wafer from an ambient temperature of 298 K using fixed lamp powers. Both uniform and nonuniform lamp power combinations, as detailed in Table 4, are used to generate the spatiotemporal temperature data that is then used to create the dynamic model. The temperatures on the four inspection points and corresponding lamp powers are recorded in each simulation from $t = 0$ s to $t = 2000$ s, which is when the system is indistinguishable from its final steady-state. Variable sampling times were also used to account for the fast dynamics at the start of each run. Examples of transient profiles of open-loop data are presented in Fig. 4.

Table 4
List of lamp powers used to generate open-loop data (W/m^2).

P_1	P_2	P_3
1000	1000	1000
2000	2000	2000
3000	3000	3000
4000	4000	4000
5000	5000	5000
2000	2000	4000
1500	2500	3500
1000	5000	3000
1000	1000	5000
450	450	4000

To achieve desired model performance, the range of operating conditions in training data is larger than the operating range of interest. The target temperature in the paper is 573 ± 3 K, and the generated open-loop data set contains final steady-state temperatures from less than 540 K to over 700 K. Moreover, the coupling effects of control variables are also necessary to be taken into consideration, as 5 groups of nonuniform lamp power data with various power ratios were involved in model training. Additionally, process safety is an inevitable issue when obtaining data from real world experiments under the wide range of operating conditions; however, computational simulation methods such as CFD avoid the potential safety risk.

The time derivative of the temperature generated from the open-loop data is calculated using a first-order numerical method as per the equation,

$$\dot{\overline{T}}_t = \frac{\overline{T}_{t+1} - \overline{T}_{t-1}}{2\Delta t} \tag{6}$$

where $\overline{T}_t$ is the measurement temperature vector containing all 4 temperature measurement points at time t, $\overline{T}_{t+1}$ is the temperature vector at the subsequent time step, $\overline{T}_{t-1}$ is the temperature vector at the previous time step, and Δt is the length of the time step. In order to reduce the computational cost while preserving simulation accuracy, a time-varying step size is implemented during the open-loop simulation. A much smaller integration time step is used for the first 5 s of the simulation duration when the temperature increases rapidly, while increasingly larger integration time steps are employed for the remainder of the simulation as the temperature increases slowly therein. The detailed distribution of the integration time step is described in Table 5.

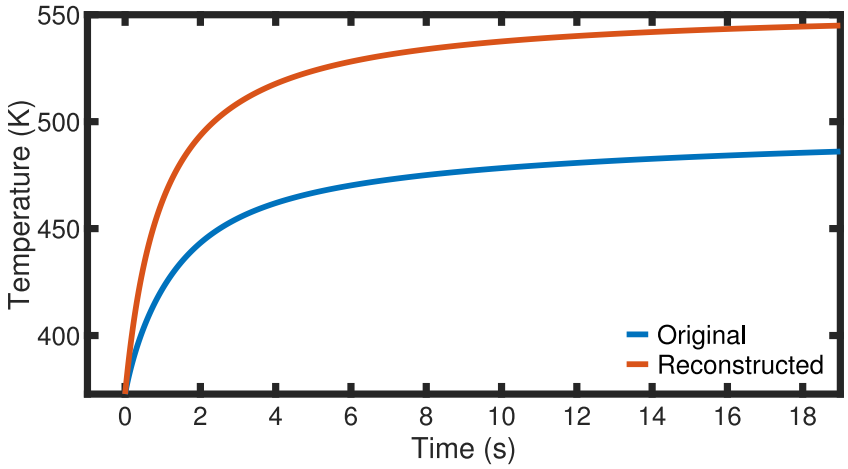

Fig. 5. Comparison of the temperature profiles of the original and reconstructed data. The reconstructed temperature approaches the final steady-state value within 20 s, which enables the dynamic model trained using the reconstructed data to capture the correct steady state while retaining the trend of the original data.

Analysis of Fig. 4 shows a distinct trend: Temperatures at all four points initially rise rapidly but then rise slowly for the remainder of the simulation, which leads to larger time derivatives in the first few seconds compared to the majority of the simulation. This phenomenon leads to a data imbalance problem, i.e., if the data used for modeling spans the entire time frame, then the model will accurately capture neither the fast transient nor the subsequent slow rise. Instead, selectively using only the early simulation data with the fast transient changes and discarding the remainder of the simulation is often proposed as one intuitive solution to maximize the ability of the model to capture the dynamic behavior. However, the temperature at the end of the initial fast transient step remains significantly lower than the final, steady-state temperature, which deteriorates the ability of the model to capture the final steady state when using such an approach. Therefore, an innovative data reconstructing methodology is proposed, where each trajectory is reconstructed according to the following function in time:

$$T_{\text{new}}(t) = \frac{(T_{\text{final}} - T_{\text{init}}) \cdot t}{1 + t} + T_{\text{init}} \tag{7}$$

where T_{new} is an arbitrary, reconstructed temperature at any time t, T_{final} and T_{init} are the true final and initial temperatures under the designated open-loop simulation, respectively, and t is the time. This formulation guarantees that the reconstructed data will have the same initial and final temperatures as the original data since $T_{\text{new}} \to T_{\text{init}}$ as $t \to 0$, while $T_{\text{new}} \to T_{\text{final}}$ as $t \to \infty$. Importantly, the dynamic behavior of the reconstructed function is altered from the original, such that the temperatures reach their steady states earlier. Additionally, since deriving an analytical derivative function is possible for the reconstructed data, the temperature derivative is calculated as follows:

$$\dot{T}_{\text{new}}(t) = \frac{T_{\text{final}} - T_{\text{init}}}{(t+1)^2} \tag{8}$$

An example of the comparison between the original and reconstructed plot is illustrated in Fig. 5. To overcome the data imbalance problem, since the fast transient behavior still occurs within the first 5 s in the reconstructed data set, only the first 20 s of the reconstructed trajectory is used because the temperature of the reconstructed trajectory at $t = 20$ s is sufficiently close to the final, steady-state temperature, unlike the original temperature trajectory, which enables the model to capture the steady-state dynamics while retaining the shape and trend of the original data.

3.2. Dynamic model development

To train the SINDy model, the open-loop data obtained from the data generation process is rearranged into two matrices: the state matrix X and the control input matrix U, both of which are defined as follows:

Table 5
List of Δt applied during different time regions.

Time region	Δt
0 s → 5 s	0.01 s
5 s → 20 s	0.02 s
20 s → 75 s	0.05 s
75 s → 2000 s	0.1 s

$$X = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \dots & x_n(t_m) \end{bmatrix} \tag{9}$$

$$U = \begin{bmatrix} u_1(t_1) & u_2(t_1) & \dots & u_p(t_1) \\ u_1(t_2) & u_2(t_2) & \dots & u_p(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(t_m) & u_2(t_m) & \dots & u_p(t_m) \end{bmatrix} \tag{10}$$

where $x_i(t_k)$ is the ith state variable at the kth sampling time, $u_j(t_k)$ is the jth control variable at the kth sampling time, m is the number of time-series data points, n is the number of state variables, and p is the number of control input variables. With the state matrix and control input matrix, a function library $\Theta(X, U)$ is developed that consists of candidate linear and nonlinear features of both the state and control input variables. An example of a function library is

$$\Theta(X,U) = \begin{bmatrix} | & | & | & | & | & | & | \\ \mathbf{1} & X & U & X^2 & U^2 & \sin X & \cdots \\ | & | & | & | & | & | & | \end{bmatrix} \tag{11}$$

Each candidate function in $\Theta(X, U)$ is then assigned a pre-multiplying coefficient to fit the derivative of the data, as shown in Eq. (12) below:

$$\dot{X} = \Theta(X,U)\Xi \tag{12}$$

where Ξ is the coefficient matrix that stores the coefficients associated with each basis function for each dependent variable. To find Ξ, Eq. (12) is solved by Lasso regression, an optimization-based linear regression with first-order norm regularization. The Lasso optimization problem to solve for Ξ can be formulated as follows:

$$\Xi = \arg\min_{\Xi' \in \mathbb{R}} \|\Theta(X,U)\Xi' - \dot{X}\|_2 + \lambda|\Xi'|_1 \tag{13}$$

where Ξ' is a dummy variable that replaces Ξ for the optimization expression and λ is the regularization coefficient. The second term that

includes λ is the first-order regularization term that encourages sparsity in the Ξ matrix by zeroing all coefficients with an absolute value less than λ at each iteration. The optimization is solved using the Lasso regression function in the Python package Scikit Learn (in which the parameter α corresponds to λ as per the formulation of Eq. (13)).

The state variables in this work are the four temperature measurement points on the wafer, $T_i, i \in \{1,2,3,4\}$, and the control input variables are the power levels for the three lamp groups, $P_j, j \in \{1,2,3\}$. The four temperatures are decoupled from each other and modeled using four separate SINDy models, leading to the following sparse identification problem formulation:

$$\dot{T}_i = \Theta_i(T_i, P)\Xi_i, \qquad i \in \{1,2,3,4\} \tag{14}$$

where $T_i \in \mathbb{R}^m$ is the ith temperature state variable vector with time-derivative $\dot{T}_i$, $P = [P_1 \quad P_2 \quad P_3] \in \mathbb{R}^{m\times 3}$ is the lamp power control input matrix, and Ξ_i is the coefficient vector for each basis function for the ODE corresponding to the ith temperature state. The variables T_i and P correspond to X and U in Eq. (12), respectively. The rationale behind employing a distinct SINDy model for each state variable is to introduce sparsity during the candidate function selection process. The derivative of each state variable is presumed to be solely dependent on its own value, without the influence of other state variables. The candidate functions for each Θ_i implemented in this work are a bias term and a combination of linear and quadratic functions of the states and inputs, as follows:

$$\Theta_i(T_i, P) = \begin{bmatrix} | & | & | & | & | & | \\ \mathbf{1} & T_i & P & T_i \odot T_i & P \odot P & T_i \odot P \\ | & | & | & | & | & | \end{bmatrix} \tag{15}$$

where $\mathbf{1} \in \mathbb{R}^m$ is a column vector of ones that serves as the bias term in linear regression, and $\odot$ denotes element-wise multiplication except in the term $T_i \odot P$, where it denotes element-wise multiplication between the ith temperature state variable vector and each lamp group's power control vector, i.e., $T_i \odot P = [T_i \odot P_1 \quad T_i \odot P_2 \quad T_i \odot P_3]$. As a result, $\Theta_i(T_i, P)$ is the $\mathbb{R}^{m\times 12}$ feature map matrix of $\dot{T}_i$. To improve the conditioning of the sparse identification problem, each candidate function/column in $\Theta(T, P)$ is multiplied by an additional coefficient to standardize all columns to a comparable scale. For this work, the regularization parameter λ is set as 0.01.

The choice of basis functions is highly dependent on the system under consideration. The natural occurrence of polynomials and trigonometric functions in engineering systems makes them an appropriate first choice for the basis functions when using SINDy (Brunton et al., 2016). However, the basis set can and should be adapted based on both performance of the method and, more importantly, any prior knowledge of the system dynamics. For example, for chemical reactors, the temperature-dependence of the reaction kinetics is based on the Arrhenius law, which suggests using exponential terms in the reactor temperature dynamic equation in the SINDy library (Abdullah and Christofides, 2023). Conversely, trigonometric functions form an appropriate basis for power systems (Stanković et al., 2020). In the ALE process studied, since quadratic polynomials were sufficient to accurately capture the system dynamics when used in conjunction with the proposed data reconstruction scheme, no other basis functions were considered.

The accuracy of the trained SINDy model can be represented by derivative fitting plots and state prediction plots. The derivative fitting plot for when the lamp power is set at a uniform 2000 W/m^2 is shown in Fig. 6(a), which demonstrates that the trained SINDy model fits well to the derivative data. The state prediction plot is illustrated in Fig. 6(b), and it compares the true and predicted temperatures. The predicted trajectory is obtained by integrating the trained SINDy model using the explicit Euler method from the same initial temperature measurement as the true data. By examining the state prediction plot, the predicted temperature from the model is 8.81 K lower than the true temperature at the end of the trajectory at $t = 20$ s. As the total change in temperature for the true trajectory is 190 K, this represents an error of less than 5% error compared to the true trajectory. Furthermore, while the entire 20 s of the prediction in Fig. 6(b) is initialized once from the temperature at t_0, in practice, when the SINDy model is incorporated into a model-based controller, it will typically predict over much smaller time periods than 20 s with a single initialization, which will yield much smaller errors. Overall, the results suggest that the dynamic model developed fulfills the requirement of the controller.

4. Model predictive control

Model predictive control is an optimization-based advanced control strategy that utilizes a dynamic model to predict the future values of the state variables and then optimizes a cost function to find the optimal control action. The general formulation of model predictive controller is shown in the following set of equations:

$$\begin{aligned} \min_{u(t)\in S(t_s)} \quad & \int_{t_k}^{t_{k+N}} L(\tilde{x}(\tau), u(\tau))\, d\tau \\ \text{s.t.} \quad & \dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) \\ & \tilde{x}(t_k) = x(t_k) \\ & u(t) \in U,\ \forall\, t \in [t_k, t_{k+N}) \\ & g_{\text{MPC},1}(\tilde{x}(t), u(t)) = 0 \\ & g_{\text{MPC},2}(\tilde{x}(t), u(t)) \le 0 \end{aligned}$$

where $L(\tilde{x}, u)$ is the cost function to be optimized, $u(\tau)$ is the set of control actions and inputs that the controller receives, $f(\tilde{x}, u)$ is the dynamic model used for state prediction, $\tilde{x}(t)$ is the predicted state variable, $x(t_k)$ is the measured state variable from the physical system at time t_k where $[t_k, t_{k+N})$ represents the prediction horizon in model predictive control, and $g_{\text{MPC},1}$ and $g_{\text{MPC},2}$ denote equality and inequality constraints, respectively.

Model predictive control has received recognition for commercial applications because of its advantages over traditional controllers (Qin and Badgwell, 1997). Unlike traditional controllers that implement control based on the feedback of offsets, such as the proportional–integral (PI) controller, an MPC foresees the future state values to make optimized control actions. PI controllers may encounter difficulties when trying to achieve a fast response with minimal overshoot due to their reliance on the continuously updated integral term to eliminate bias. On the other hand, with the predictions made by the dynamic process model, the MPC is able to optimize the control actions to quickly reach the setpoint while eliminating severe overshoots. Moreover, the MPC is capable of controlling highly nonlinear processes with constraints in the control actions and states, which is challenging for traditional linear controllers that cannot explicitly account for the presence of constraints (Garcia et al., 1989; Mayne, 2014). Model predictive control also offers flexibility in the control objectives, as the cost function can be customized to optimize for a variety of objectives, such as to maximize a profit or yield function in economic MPC.

4.1. Implementation of model predictive controller

In order to implement model predictive control, it is necessary to define the sampling time and prediction horizon and construct the cost function and constraints for the state and control variables. The sampling time t_s is set to 0.2 s in this work, signifying that the temperature is measured every 0.2 s, and the control action is applied and held constant for 0.2 s until the next control action is applied. The prediction horizon is set to 3 in this work, denoting that the controller forecasts three sampling periods (0.6 s) to calculate the cost function for optimization. The cost function and constraints used in the MPC are described below:

$$\text{Cost Function: } L(T, P) = \sum_{i=1}^{4}(T_i(t) - 573)^2 + \beta \sum_{j=1}^{3}(P_j(t) - P_j^{ss})^2 \tag{16}$$

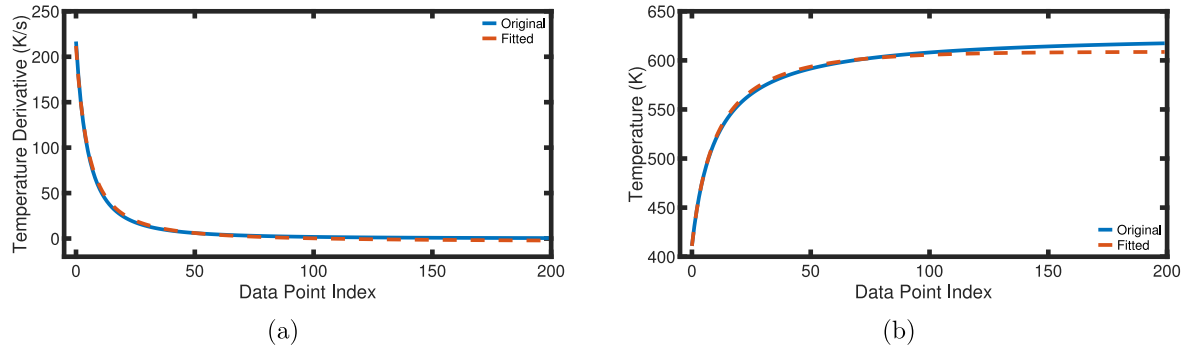

Fig. 6. (a) Derivative fitting and (b) state prediction plots for checking model performance. The fitted line in (a) closely aligns with the original data, suggesting that the SINDy model successfully captures the system dynamics. The fitted line in (b) shows strong correspondence between the predicted temperature and original data, where the maximum absolute error is 8 K when predicting 20 s of temperature.

$$\text{Input Magnitude Constraint: } P_j \in [0, 5000] \text{ W/m}^2 \tag{17}$$

$$\text{Input Rate of Change Constraint: } \left|P_j(t+t_s) - P_j(t)\right| \leq 250 \text{ W/m}^2 \tag{18}$$

where $T_i(t)$ is the predicted value of the measured temperature on inspection point i at time t; $P_j(t)$ denotes the power of lamp group j at time t, as the power is adjusted at the start of each sampling period. The cost function consists of two components. The first term is the sum of squared differences between the predicted temperatures and the target temperature of 573 K. The predicted temperatures are calculated by integrating the SINDy model using the explicit Euler method starting at the state measurement at $t = t_k$. The second component is the sum of the squared differences between the optimized lamp power and the assigned steady-state values. The second term serves as a steady-state regularization on the input lamp power, where β is a tunable coefficient. A high β value drives the optimized lamp power closer to the assigned steady-state value, while a small β value encourages a lamp power that minimizes the deviation between the predicted and target temperatures. Based on extensive tests, $\beta = 1 \cdot 10^{-4}$ is used for involving moderate influence of the steady-state power in the optimization process to reach the target quickly while keeping stability at the same time. The upper bound of lamp power and maximum adjusting rate are set to 5000 W/m^2/s and 250 W/m^2/s, respectively, to account for physical constraints in practical lamps. The Python package CyIPOPT is used for solving the constrained nonlinear optimization problem to obtain the optimized lamp power.

4.2. Feedback-based time-varying steady state in MPC

It is important to note that both the reactor and wafer are at room temperature at $t = 0$. The components of the reactor are simultaneously heated with the wafer by the lamps via radiative heat transfer. The walls of the reactor and the lamp surfaces also contribute to additive amounts of radiation through reflection, emission, and transfer of external radiation. As a result, the temperature of the environment around the wafer changes continuously as ambient surface temperatures rise with time. This changing nature of the ambient temperature implies that the required "steady-state" power is not constant during the process, i.e., the power required to maintain the wafer at the target temperature is unique at any given time. From a modeling perspective, the same input vectors of surface temperatures and lamp powers can produce completely different temperature derivatives as the incoming energy from the reactor and environment is continuously changing. Moreover, since reconstructed data was used during the training of the process model, deviations exist between the reconstructed derivative data and the original derivative data, particularly in the initial region, which also must be accounted for in the controller. Consequently, a novel approach of a feedback-based time-varying steady-state P_{ss} (VSS) is proposed where the steady-state lamp power is dynamically adjusted to accommodate the continuously changing temperature of the wafer environment and compensate for the derivative differences. The VSS algorithm is described in Algorithm 1.

Algorithm 1: Feedback-Based Time-Varying Steady-State Update

Data: t; T_t; S_{old}; P^{ss}_{old}
Result: S_{new}; P^{ss}_{new}
Parameters: P^{ss}_{final}; P^{ss}_0; ΔP

/* Read t, T_t from simulation of last sampling time */
/* Read S_{old}, P^{ss}_{old} from saved text file */
if $t \leq 50$ s **then**
  $S_{\text{new}} = 0$;
else if $\min(T_t) \leq 571.0$ K **and** $S_{old} = 0$ **then**
  $S_{\text{new}} = 1$;
else if $\max(T_t) \geq 574.5$ K **and** $S_{old} = 1$ **then**
  $S_{\text{new}} = 0$;
else
  $S_{\text{new}} = S_{old}$;
/* Switch variable S_{new} is updated */
if $P^{ss}_{old} \leq P^{ss}_{final}$ **then**
  $P^{ss}_{\text{new}} = P^{ss}_{final}$;
else if $t \leq 2.5$ s **then**
  $P^{ss}_{\text{new}} = P^{ss}_0$;
else if $S_{new} = 0$ **then**
  $P^{ss}_{\text{new}} = P^{ss}_{old} - \Delta P \cdot dt$;
else if $S_{new} = 1$ **then**
  $P^{ss}_{\text{new}} = P^{ss}_{old}$;
/* New steady-state power P^{ss}_{new} is updated */
/* Send P^{ss}_{new} to CFD simulation */
/* Write S_{new}, P^{ss}_{new} to saved text file */

In the VSS algorithm, the steady-state power is initialized with lamp powers of $P^{ss}_{0,1} = P^{ss}_{0,2} = 2400$ W/m^2, $P^{ss}_{0,3} = 5000$ W/m^2 and kept constant when $t \leq 2.5$ s to a achieve a quick response at the start of the time when the reactor is cool. Afterward, a switch variable S is introduced to serve as an indicator of the system status. For the first 50 s, S is set to 0 to allow the steady-state power to continuously decrease at a rate of $\Delta P_1 = \Delta P_2 = 8.1$ W/m^2/s, $\Delta P_3 = 4.6$ W/m^2/s to avoid a potential overshoot and account for the heated reactor. The temperature is constantly monitored, and if the minimum temperature is lower than 571.0 K, the system indicator variable S changes from 0 to 1, and

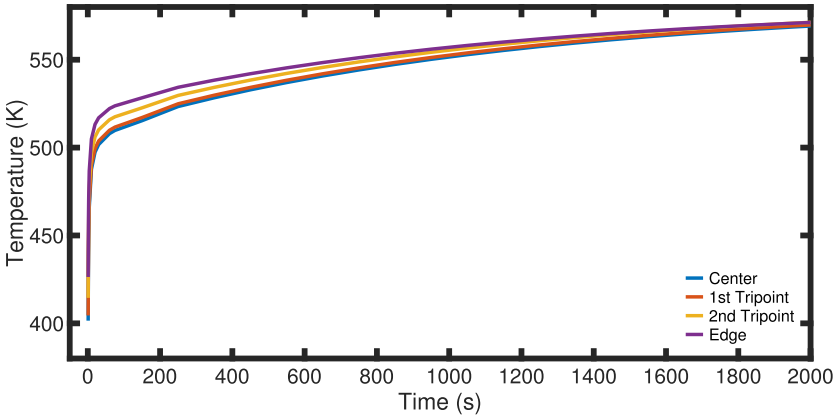

Fig. 7. Temperature profiles with open-loop control strategy, where the lamp power is fixed at P^{ss}_{final}. Due to the low power input to the system at all times, the system needs over 2000 s to reach the target temperature, which implies that a controller is needed.

P^{ss} stops decreasing to allow the temperature to gradually increase back towards the target temperature until the maximum temperature measured on the inspection points is higher than 574.5 K, which is when P^{ss} resumes decreasing. These two critical temperatures are within the bounds of the acceptable range of $[570, 576]$ K and are expected to enable the VSS strategy to control the temperature within the range of $T \in [571.0, 574.5]$ K at all times. The power is expected to reach $P^{ss}_{final} = [P^{ss}_{final,1}\ P^{ss}_{final,2}\ P^{ss}_{final,3}] = [400\ 400\ 3866]$ W/m^2 when the system reaches the final steady-state condition after a sufficiently long time. P^{ss}_{final} has been verified to produce a uniform temperature profile around the target temperature through an independent steady-state simulation.

5. Results and discussion

5.1. Open-loop control performance

The open-loop control strategy fixes the lamp power at P^{ss}_{final}, where the transient temperature profile is shown in Fig. 7. With this setup, the wafer temperature only reaches the target temperature after 2000 s of process time. P^{ss}_{final} is defined as the power used in this final steady-state scenario where the reactor is thoroughly heated up. Of note is that, in this situation, the power required to maintain a steady-state is much lower than during the initial stage of the reactor. As a result, the temperature increases rapidly during the initial stage and very slowly afterward due to the colder reactor temperature at $t = 0$. The overall response is unacceptably slow and prevents practical implementation of the open-loop control strategy. In addition, the uniformity of the temperature on the wafer surface is not within an acceptable range of the temperature setpoint (±3 K) for most of the time. Consequently, a feedback controller must be implemented to improve the performance.

5.2. Closed-loop performance with fixed steady-state power

In closed-loop control, the measured wafer surface temperature is fed to the controller, which is then used by the MPC to compute the optimal control actions to transmit to the lamp. In the current process, the controller development is challenging due to the changing environment in the reactor and the use of reconstructed training data. Specifically, the cost function of Eq. (16) constitutes of a state penalty, which is the first term, and an input penalty, which is the second term. We consider two cases for the steady-state power in the input penalty term, P^{ss}_j.

Figs. 8 and 9 depict the closed-loop wafer temperature and lamp power profiles, respectively, under the MPC when using a fixed steady-state power of P^{ss}_{final} in the MPC cost function. From Fig. 8, it can be observed that the wafer surface temperature is not uniform for most of the process duration as the difference between maximum and minimum temperatures measured on inspection points is over 6 K. This phenomenon is attributed to the low power of the center and edge lamps compared to the side lamps, as the ratio of $P^{ss}_{final,1}, P^{ss}_{final,2}$ to $P^{ss}_{final,3}$ is about 1 to 10, as shown in Fig. 9. However, it is worth mentioning that the uniformity improves continuously with time as the reactor gradually heats up. The slow response is attributed to the fact that the final steady-state power, P^{ss}_{final}, calculated using a steady-state simulation, corresponds to the case where the whole reactor is thoroughly heated up to reach their steady-state temperature in conjunction with the wafer, which is far from the case for most of the process duration. This is because both the reactor walls and lamps start at room temperature (298 K) at $t = 0$ s, alongside the wafer. The input profile demonstrated in Fig. 9 shows that the powers of the center and edge lamps, P_1 and P_2, are significantly lower than the power of the side lamp, P_3. The low center and edge lamp power results in the phenomenon in Fig. 8, where the center temperature is much lower than the edge temperature, resulting in poor uniformity. Nevertheless, the closed-loop controller performance surpasses that of the open-loop control case, as the temperature of the closed-loop controller is much closer to the target temperature for any given time. Under the MPC, the temperature at two inspection points on the wafer reached a temperature of over 570 K at $t = 1000$ s, while, in open-loop, all inspection points on the wafer are under 550 K at $t = 1000$ s. This case study demonstrates the effectiveness of an MPC over the open-loop control strategy, even if the convergence to the target temperature of 573 K occurs more slowly than desired. Additionally, an observable feature of Figs. 8 and 9 the oscillatory-like behavior in particular times is originated from the numerical CFD simulation that produces converged solutions when residual tolerances are met. The resulting error generates chattering in the temperature profiles that simultaneously affects the MPC action. However, to prevent the drastic attenuation in the temperature rise after the first 20 seconds, next, we consider a second case of using a higher steady-state power in the input penalty term, P^{ss}_0.

Figs. 10 and 11 demonstrate the closed-loop state and input profiles, respectively, under MPC when using a fixed steady state power of P^{ss}_0 in the cost function of the MPC. As expected, the controller with the high fixed steady-state power penalty cannot capture the target temperature and final steady-state. However, it significantly reduces the time it takes the wafer surface temperature to reach the target temperature. In addition, with the steady-state power set to P^{ss}_0, the ratios P^{ss}_1/P^{ss}_3 and P^{ss}_2/P^{ss}_3 are reduced to between 1 and 2, which causes the spatial uniformity in temperature to worsen as the system evolves, and the

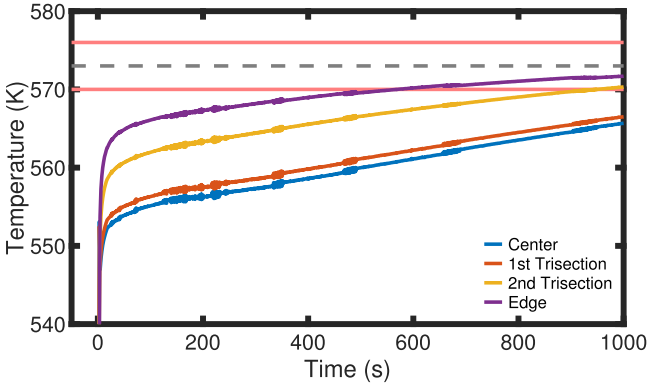

Fig. 8. Temperature profile under closed-loop MPC with steady-state power in input penalty term fixed at P^{ss}_{final}. This performance surpasses open-loop control since the temperature is closer to target. The gray middle dashed line is the target temperature; the two red solid lines on the bottom and top are boundaries of control range around the target. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

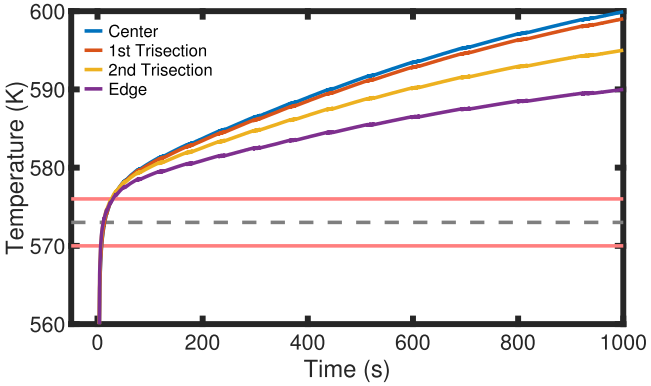

Fig. 10. Temperature profile under closed-loop MPC with steady-state power in input penalty term fixed at P^{ss}_{0}. The plotting criterion is identical to that used in Fig. 8. The system response is quicker, but severe overshoot occurs due to a higher total lamp power.

.

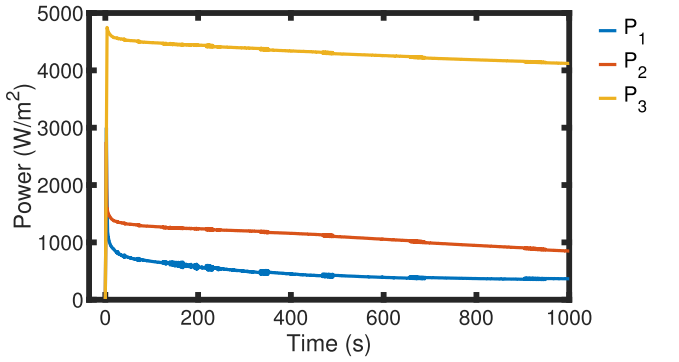

Fig. 9. Lamp power profiles under closed-loop MPC with steady-state power in input penalty term fixed at P^{ss}_{final}. The low power for the center (P_1) and edge (P_2) lamps compared to the power of the side lamp (P_3) corresponds to the phenomenon that the center temperature is significantly lower than the edge temperature in Fig. 8, which is attributed to a low $P^{ss}_{final,1}$ and $P^{ss}_{final,2}$.

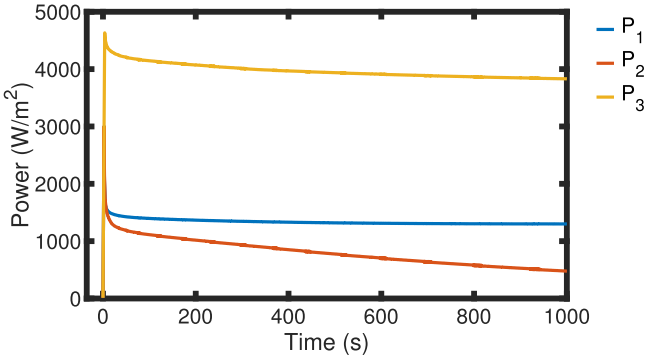

Fig. 11. Lamp power profiles under closed-loop MPC with steady-state power in input penalty term fixed at P^{ss}_{0}. The center lamp power (P_1) is significantly higher than that observed in Fig. 9 because $P^{ss}_{0,1}$ and $P^{ss}_{0,2}$ have a larger magnitude than $P^{ss}_{final,1}$ and $P^{ss}_{final,2}$, which results in a higher center temperature than the edge temperature in Fig. 10.

temperatures to vary by more than 6 K between inspection points soon after the initial rise. This phenomenon is reflected in the control actions in Fig. 11, where the center lamp power is much higher than in the low fixed steady-state case. However, as the system continues to evolve, the high center-to-edge power ratio results in a much higher center temperature than the edge and also severe overshoot because of the high total lamp power.

By combining the observations from the low fixed steady-steady power and the high fixed steady-state power cases, it can be proposed that a time-varying steady-state power is necessary to fulfill the requirements of a fast response, small overshoot, and spatial temperature uniformity at all times.

5.3. Closed-loop performance with feedback-based time-varying steady-state in MPC

The time-varying steady-steady power controller defined in Algorithm 1 has a temperature output as shown in Fig. 12. Under the VSS approach, the temperature successfully reaches the target value of 573 K within 10 s, which is as fast as the high fixed steady-state power penalty case. After reaching the target temperature and overshooting a minor amount, which is below the upper threshold limit, the continuously decreasing steady-state power gradually drives the wafer surface temperature back towards the temperature setpoint. After the lowest measured temperature on the wafer reaches the lower temperature threshold of 571.0 K, the VSS controller stops decreasing the steady-state power and successfully moves the temperature back to target. The VSS controller then resumes decreasing the steady-state power when the highest wafer surface temperature reaches the upper temperature threshold of 574.5 K, and successfully brings the temperature back to the target. The control logic of the VSS controller is observed in the control action in Fig. 13, where the inflection points are observed on lamp power curve when plotted against time, which indicates that the VSS algorithm does exhibit start-stop behavior when decreasing the steady-state lamp power as described above. Throughout the whole process, and especially after the initial stage, the temperature is well-controlled within the accepted region, and uniformity is also well-maintained. From Fig. 12, it is seen that the VSS controller greatly surpasses the fixed steady-state controller in terms of controller performance.

6. Conclusion

In this work, we developed a 2-D transient energy model for a thermal atomic layer etching (ALE) reactor with a radiative lamp heating system that consists of three groups of independently controlled heating lamps. A data-driven dynamic model that predicts the system behavior is generated through sparse identification (SINDy) with an open-loop dataset consisting of 10 trajectories. A model predictive controller (MPC) that relies on the generated dynamic model is then applied to both the reactor and wafer, with the goal of driving the system from room temperature to the target temperature of 573 K. The

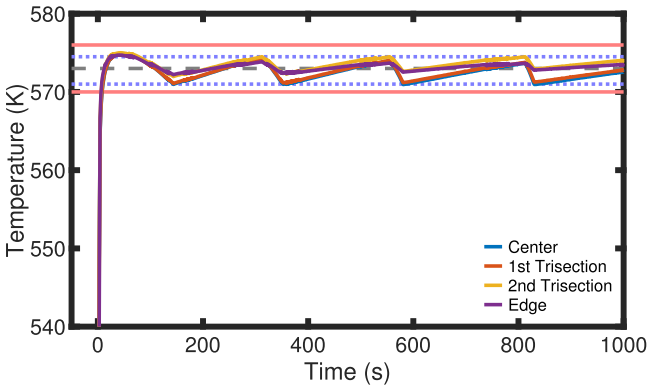

Fig. 12. Temperature profile under closed-loop MPC with the feedback-based time-varying steady-state (VSS) power approach. The temperatures reach the target within 10 s and are kept within the acceptable range as desired. The blue dotted lines are the lower and upper temperature bounds (571.0 K, 574.5 K) for the VSS to stop or resume decreasing P^{ss} in the cost function of the MPC formulation.

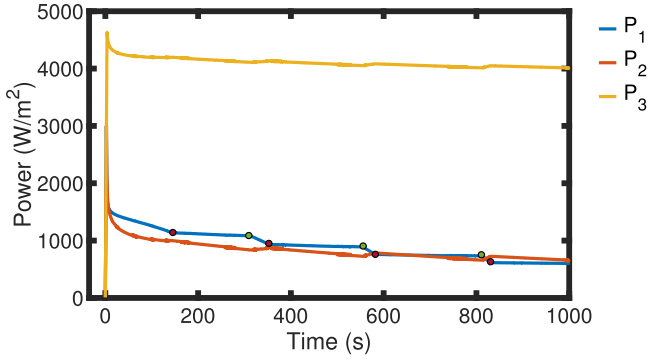

Fig. 13. Lamp power profiles under closed-loop MPC with the feedback-based time-varying steady-state (VSS) power approach. The red points indicate that VSS stops decreasing P^{ss} in MPC cost function, while green points indicate resume decreasing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

open-loop controller strategy showed a slow response with poor temperature uniformity. An MPC was implemented with two settings for the input penalty term in the objective function, one with a low fixed steady-state power and one with a high fixed steady-state power. While the former MPC resulted in a faster response and better uniformity than that of the open-loop, the latter produced a rapid response but overshot heavily. Therefore, a feedback-based time-varying steady-state (VSS) power penalty approach was proposed, where the MPC gradually decreases the steady-state power from an initial high value based on the wafer surface temperature and process conditions. The VSS approach successfully achieved a fast response and kept the temperature within an acceptable range for the entire duration of the process.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Financial support from the National Science Foundation, United States is gratefully acknowledged. This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group.

References

Abdullah, F., Christofides, P.D., 2023. Data-based modeling and control of nonlinear process systems using sparse identification: An overview of recent results. Comput. Chem. Eng. 174, 108247.
Alanqar, A., Durand, H., Christofides, P.D., 2015. On identification of well-conditioned nonlinear systems: Application to economic model predictive control of nonlinear processes. AIChE J. 61, 3353–3373.
ANSYS, 2022. Ansys Fluent Theory Guide. ANSYS Inc., Canonsburg, PA.
Baker, J., Christofides, P.D., 1999. Output feedback control of parabolic PDE systems with nonlinear spatial differential operators. Ind. Eng. Chem. Res. 38, 4372–4380.
Baker, J., Christofides, P.D., 2000. Finite-dimensional approximation and control of non-linear parabolic PDE systems. Internat. J. Control 73, 439–456.
Brunton, S.L., Proctor, J.L., Kutz, N.J., 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Appl. Math. 113, 3932–3937.
Cho, Y.M., Paulraj, A., Kailath, T., Xu, G., 1994. A contribution to optimal lamp design in rapid thermal processing. IEEE Trans. Semicond. Manuf. 7, 34–41.
Dassau, E., Grosman, B., Lewin, D.R., 2006. Modeling and temperature control of rapid thermal processing. Comput. Chem. Eng. 30, 686–697.
Elers, K.-E., Blomberg, T., Peussa, M., Aitchison, B., Haukka, S., Marcus, S., 2006. Film uniformity in atomic layer deposition. Chem. Vapor Depos. 12, 13–24.
Espadinha-Cruz, P., Godina, R., Rodrigues, E.M.G., 2021. A review of data mining applications in semiconductor manufacturing. Processes 9, 305.
Fogler, H.S., 2006. Elements of Chemical Reaction Engineering, fourth ed. Prentice Hall PTR, Upper Saddle River, NJ.
Garcia, C.E., Prett, D.M., Morari, M., 1989. Model predictive control: Theory and practice—A survey. Automatica 25, 335–348.
Gyurcsik, T.J., Sorrell, F.Y., 1991. A model for rapid thermal processing: Achieving uniformity through lamp control. IEEE Trans. Semicond. Manuf. 4, 9–13.
Lee, Y., DuMont, J.W., George, S.M., 2016. Trimethylaluminum as the metal precursor for the atomic layer etching of Al_2O_3 using sequential, self-limiting thermal reactions. Chem. Mater. 28, 2994–3003.
Mayne, D.Q., 2014. Model predictive control: Recent developments and future promise. Automatica 50, 2967–2986.
Oh, H.-K., Kang, S.B., Choi, Y.K., Lee, J.S., 2009. Optimization of rapid thermal processing for uniform temperature distribution on wafer surface. J. Mech. Sci. Technol. 23, 1544–1552.
Ponraj, J.S., Attolini, G., Bosi, M., 2013. Review on atomic layer deposition and applications of oxide thin films. Crit. Rev. Solid State Mater. Sci. 38, 203–233.
Proctor, J., Brunton, S., Brunton, B., 2014. Exploiting sparsity and equation-free architectures in complex systems. Eur. Phys. J. Spec. Top. 223, 2665–2684.
Qin, J.S., Badgwell, T.A., 1997. An overview of industrial model predictive control technology. In: AICHE Symposium Series, vol. 93, p. 316.
Ren, Y.M., Alhajeri, M.S., Luo, J., Chen, S., Abdullah, F., Wu, Z., Christofides, P.D., 2022. A tutorial review of neural network modeling approaches for model predictive control. Comput. Chem. Eng. 165, 107956.
Roozeboom, F., Parakh, N., 1990. Rapid thermal processing systems: A review with emphasis on temperature control. J. Vacuum Sci. Technol. B 8, 1249–1259.
Schaper, C.D., Moslehi, M.M., Saraswat, K.C., Kailath, T., 1994. Modeling, identification, and control of rapid thermal processing systems. J. Electrochem. Soc. 141, 3200.
Stanković, A.M., Sarić, A.A., Sarić, A.T., Transtrum, M.K., 2020. Data-driven symbolic regression for identification of nonlinear dynamics in power systems. In: 2020 IEEE Power & Energy Society General Meeting. PESGM, IEEE, pp. 1–5.
Theodoropoulou, A., Adomaitis, A.R., Zafiriou, E., 1998. Model reduction for optimization of rapid thermal chemical vapor deposition systems. IEEE Trans. Semicond. Manuf. 1, 85–98.
Timans, P.J., 1998. Rapid thermal processing technology for the 21st century. Mater. Sci. Semicond. Process. 1, 3–4.
Van Overschee, P., De Moor, B., 1994. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica 30, 75–93.
Wang, J., Moreira, J., Cao, Y., Gopaluni, B., 2022. Time-variant digital twin modeling through the Kalman-generalized sparse identification of nonlinear dynamics. In: American Control Conference. Atlanta, Georgia, pp. 5217–5222.
Wang, C.-P., Tsai, Y.-P., Lin, B.J., Liang, Z.-Y., Chiu, P.-W., Shih, J.-R., Lin, C.J., King, Y.-C., 2020. On-wafer FinFET-based EUV/eBeam detector arrays for advanced lithography processes. IEEE Trans. Electron Devices 67, 2406–2413.
Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019. Machine learning-based predictive control of nonlinear processes. part I: theory. AIChE J. 65, e16729.
Yun, S., Tom, M., Luo, J., Orkoulas, G., Christofides, P.D., 2022a. Microscopic and data-driven modeling and operation of thermal atomic layer etching of aluminum oxide thin films. Chem. Eng. Res. Des. 177, 96–107.
Yun, S., Tom, M., Orkoulas, G., Christofides, P.D., 2022b. Multiscale computational fluid dynamics modeling of spatial thermal atomic layer etching. Comput. Chem. Eng. 163, 107861.
Yun, S., Tom, M., Ou, F., Orkoulas, G., Christofides, P.D., 2022c. Multiscale computational fluid dynamics modeling of thermal atomic layer etching: Application to chamber configuration design. Comput. Chem. Eng. 161, 107757.