

Article

# Integrating Feedback Control and Run-to-Run Control in Multi-Wafer Thermal Atomic Layer Deposition of Thin Films

Yichi Zhang <sup>1</sup>, Yangyao Ding <sup>1</sup> and Panagiotis D. Christofides <sup>1,2,\*</sup>

<sup>1</sup> Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA; stevenzyc9558@ucla.edu (Y.Z.); dingyangyao1995@ucla.edu (Y.D.)

<sup>2</sup> Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA

\* Correspondence: pdc@ucla.edu

Received: 19 November 2019; Accepted: 18 December 2019; Published: 21 December 2019



**Abstract:** There is currently a lack of understanding of the deposition profile in a batch atomic layer deposition (ALD) process. Also, no on-line control scheme has been proposed to resolve the prevalent disturbances. Motivated by this, we develop a computational fluid dynamics (CFD) model and an integrated online run-to-run and feedback control scheme. Specifically, we analyze a furnace reactor for a SiO<sub>2</sub> thin-film ALD with BTBAS and ozone as precursors. Initially, a high-fidelity 2D axisymmetric multiscale CFD model is developed using ANSYS Fluent for the gas-phase characterization and the surface thin-film deposition, based on a kinetic Monte-Carlo (kMC) model database. To deal with the disturbance during reactor operation, a proportional integral (PI) control scheme is adopted, which manipulates the inlet precursor concentration to drive the precursor partial pressure to the set-point, ensuring the complete substrate coverage. Additionally, the CFD model is utilized to investigate a wide range of operating conditions, and a regression model is developed to describe the relationship between the half-cycle time and the feed flow rate. A run-to-run (R2R) control scheme using an exponentially weighted moving average (EWMA) strategy is developed to regulate the half-cycle time for the furnace ALD process between batches.

**Keywords:** batch processes; data-driven modeling; first principles modeling; control; quality control; monitoring; optimization; design of batch processes

## 1. Introduction

Deposition is a vital part of thin-film manufacturing and vacuum deposition methods, and especially chemical vapor deposition (CVD) methods are crucial to the semiconductor manufacturing industry due to their superior capabilities depositing high quality and high-performance oxide and solid materials. The growing complexities and demands of electronic devices have been continuously urging the refinement of deposition techniques to develop better precision and controllability including the application of run-to-run control methods (e.g., [1]). For example, the transistors in recent memory devices like 3D not-and (NAND) memory and dynamic random-access memory (DRAM) are based on non-planar 3D design and require ultra-thin gate structures (<10 nm) with minimal defect formation [2,3]. In order to meet these design challenges, atomic layer deposition (ALD), which can be viewed as a variation of CVD, has been developed for nanoscale thin-film manufacturing [4]. A general ALD process includes sequential injections of precursor gases and purge gas into the reactor chamber, where each injection phase is called a half-cycle. Under the ideal operating conditions and the optimal precursor selection, the wafer

substrate undergoes self-limiting reaction with each precursor, which leads to a defect-free and conformal thin-film even for high aspect-ratio feature surfaces [5].

It has been demonstrated that ALD is capable of providing nano-scale control for various materials including metals, organic polymers, and metal oxides [5]. Thus, in industrial production, ALD has been adopted for a variety of thin-film materials manufacturing. Nevertheless, on the mass-production scale, ALD has encountered difficulties including throughput and operating costs. A traditional single-wafer ALD reactor can only process one wafer at a time, which leads to a high precursor and energy cost, as well as low productivity. To overcome the above hindrances, different types of batch ALD systems have been researched and utilized in industry. For instance, spatial ALD systems separate the precursor gases and half-cycles using multiple reactor chambers, where each reactor chamber is specifically responsible for a specific half-cycle. The wafer is switched between the chambers in a streamlined manner, thereby allowing multiple wafers to be processed concurrently [6]. On the other hand, the furnace ALD system utilizes a big hot-wall reactor chamber to process up to 125 parallel placed wafers together. The two most predominant furnace ALD system designs are vertical and horizontal furnace ALD chambers, which differ mostly by the position of the precursor inlet. Leading manufacturers in the semiconductor industry have developed a variety of reactor furnaces, including TFS 500 from Beneq, A412 PLUS from ASM, P-1000 from Picosun, etc., all providing high throughput and cost-efficient solutions to ALD wafer mass-production [7].

Despite the advantages of the furnace ALD system, there are still various challenges associated with the realistic industrial manufacturing environment, because of its large wafer quantity and complex reactor geometry. The lack of understanding of the influence of the gas-phase profile on the wafer surface deposition is one of the most important problems [8]. In an experimental work and industrial pilot plant, the operating cost associated with the detailed tests in exploring the optimal operating conditions for the batch ALD systems, especially furnaces, is also very high due to the large physical scale and the long processing time [9]. Thus, experimental works have only been able to investigate a limited range of precursor selections and operating conditions. As a result, many attempts to simulate and characterize the ALD have been made. Specifically, Ref. [10] developed an overview of the application of kinetic Monte-Carlo (kMC) modeling to thin-film growth. Refs. [11,12] explored the surface reaction kinetics and structural details with kMC for  $\text{HfO}_2$  and  $\text{SiO}_2$  thin-film ALD, respectively. Ref. [13] formulated a computational fluid dynamics (CFD) simulation of the ALD cross-flow reactor gas-phase transport phenomena. Recently, Ref. [14] have constructed a multiscale CFD simulation that integrates surface deposition modeling and gas-phase CFD modeling for the simulation of a single-wafer showerhead ALD system. Based on the CFD simulation, Ref. [15] proposed a machine learning based data-driven model that largely cut down the computation time and preserved model fidelity. However, furnace ALD simulations were mainly still based on semi-empirical Langmuir isotherm, empirical relationships, and simplified CFD gas-phase simulation [8,9].

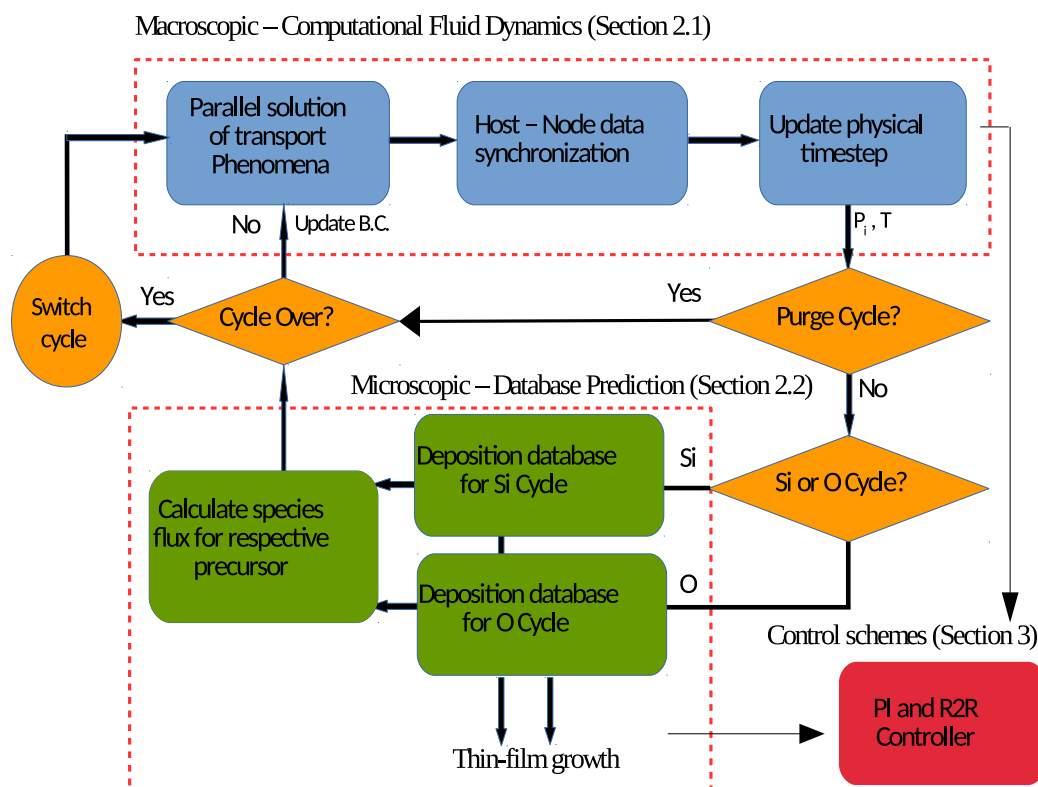
Furthermore, because of the large physical dimension of the batch ALD systems, they are much more susceptible to batch-to-batch variation and process disturbances compared to single wafer systems [16,17]. Therefore, it is beneficial if on-line controllers, integrated with run-to-run control schemes, can be formulated and implemented for the batch ALD processes [18]. Several works have investigated control schemes for single-wafer and small-scale CVD and ALD systems [19–23]. For example, Ref. [24] utilized an ion beam sculpting feedback control to fabricate non-porous thin silicon nitride membranes using ALD coated pores. Ref. [25] looked into a predictive control scheme to regulate the heating in a house-made ALD reactor. These proposed control schemes demonstrated the challenges and opportunities of integrating control designs in complicated modern applications [26,27]. Nevertheless, it has been demonstrated that there is a limited amount of on-line control schemes as well as run-to-run control schemes proposed for furnace ALD systems [28].

In this work, a multiscale CFD model and an integrated on-line control run-to-run scheme are constructed for a furnace ALD process of  $\text{SiO}_2$  thin-film production with BTBAS and ozone as precursors, as shown in Figure 1. A 2D axisymmetric multiscale CFD model is first formulated,

combining the surface deposition domain and the gas-phase transport domain, and is validated with respect to industrial and experimental results. Next, a proportional-integral (PI) controller is developed, which regulates the precursor partial pressure set-point, which allows full-coverage on the wafer surface, by manipulating the inlet precursor concentration. Then, a database for multiple operating conditions and a regression method are developed to characterize the input-output relationship between half-cycle time and the inlet feed flow rate. This relationship can be used in an exponentially weighted moving average (EWMA) algorithm of a run-to-run (R2R) control scheme to adjust the operating conditions for half-cycle time tracking. Finally, the R2R and PI controllers are integrated with the multiscale CFD model via a message-passing interface (MPI), and the closed-loop performance under a vacuum pump disturbance is evaluated.

## 2. Multiscale CFD Modeling for Furnace ALD Process

In this section, the construction of a 2D axisymmetry multiscale CFD model for the furnace thermal ALD reactor is discussed. First, the cylindrical ALD furnace geometry, the 2D axisymmetric approximation, and the meshing process are discussed. Then, the CFD model settings and the solution of the gas-phase momentum, energy, and species transport equations are presented and investigated. Next, the surface deposition calculation adopting the previously developed kMC database is introduced. In addition, the automatic workflow under the message passing interface (MPI) and the computation parallelization is discussed. Specifically, the ALD process of  $\text{SiO}_2$  involves two half-cycles, O-cycle and Si-cycle, for precursor depositions, and also a purge step in between half-cycles for purging. In the Si-cycle, the BTBAS precursor is introduced into the batch ALD reactor, which reacts with the substrate surface to form a Si layer, whereas in the O-cycle, the O layer is produced by ozone. Due to the intriguing complexity and the substantial operating cost associated, this study focuses on the discussion of Si-cycle.



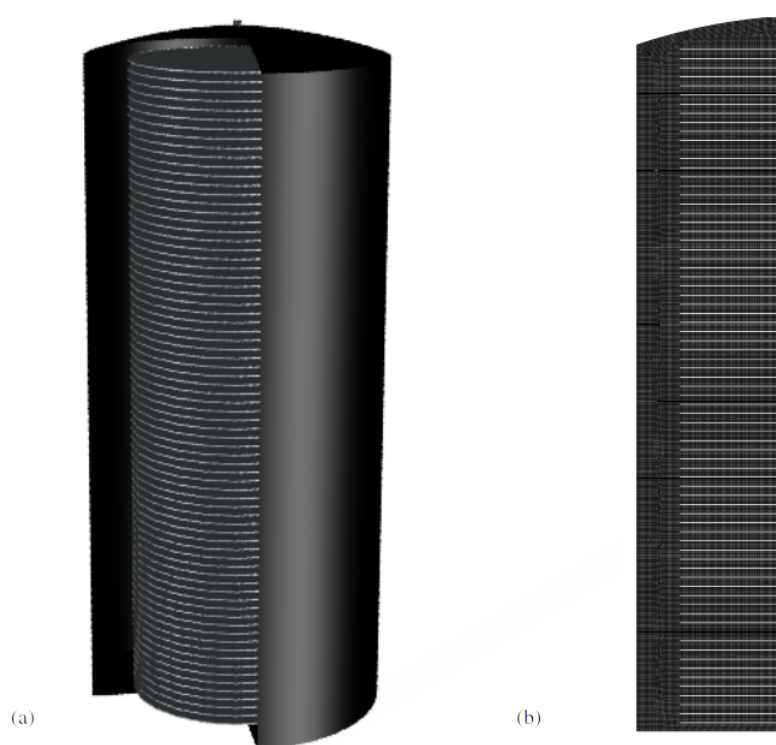
**Figure 1.** Multiscale CFD simulation workflow, parallel processing, information exchange illustration, and control schemes.

## 2.1. Gas-Phase CFD Model

The gas-phase transport phenomena greatly affect the overall gas-phase profile, the surface deposition profile and the resulting half-cycle time. Due to the large and complicated geometry of the multi-wafer furnace reactor, it is hard to develop a simple analytical solution without sacrificing the important details in the simulation. Nevertheless, CFD provides a comprehensive but computationally feasible solution. By discretizing the domain spatially and temporally, a highly accurate gas-phase profile can be computed from the governing equations of momentum, species and energy transfer. In addition, for the ALD surface deposition, CFD provides insights to the gas-phase environment that the wafer surfaces are exposed to. In this work, we utilize the ANSYS Fluent 2019R2 software package to construct the reactor geometry, mesh the domain, and set up the CFD solution [29].

### 2.1.1. Furnace ALD Reactor Design and Meshing

The first aspect of the CFD model is to choose a precise and accurate furnace model geometry, which provides the basis for a meaningful simulation of the actual industrial manufacturing process. Many companies have developed their batch ALD systems for thin-film mass production. In our model, we adopted a general industrial vertical furnace system, shown in Figure 2a, which is demonstrated to have better performance than a parallel furnace platform or a spatial platform [18]. In this reactor design, the precursor is introduced into the reactor through the top inlet and exits through the bottom outlet. A total of 75 plates are placed in parallel across the reactor chamber, of which 60 are 300 mm substrate wafers and 15 are heating susceptors, with four wafers wrapped between two heating susceptors. Specifically, the inter-wafer distance is 10 mm, the height of the wafer holders and susceptors is 2 mm, the height of the inlet region is 50 mm, resulting in a total reactor height of 950 mm, and the side vertical tunnel has a width of 60 mm. A 2D axisymmetric model could be built because of the azimuthal symmetry, which saves a significant amount of computational power while maintaining geometric detail and model validity. This dimensional reduction, shown in Figure 2b, can be achieved using AUTOCAD, a computer-aided design software.



**Figure 2.** (a) The full 3D cylindrical furnace atomic layer deposition (ALD) reactor geometry; (b) the 2D-axisymmetric representation of the 3D reactor.

In advance of the CFD simulation, the reactor geometry needs to be dissected into geometric primitives. The meshing quality is crucial to the computation accuracy and efficiency [30]. For this model, a software-aided structured mesh is generated to precisely control the mesh size and the connection to enhance the mesh quality. With a mesh independence test, a resulting mesh with 50388 cells is proven to achieve a good cell quality and a size-independent solution. The acceptable range and the desired value of skewness, overall quality, and aspect-ratio, which characterize the mesh quality, are shown in Table 1. For each mesh property, the max/min qualities lie within the acceptable range, and the average quality lies towards the optimal side, which is labeled with the asterisk symbol [29,31].

**Table 1.** Mesh quality metrics. The asterisk symbol (\*) represents the optimal end of the range.

	Acceptable Range	Mesh Bound	Mesh Average
quality	0–1 *	0.50	0.99
skewness	0 *–0.95	0.51	0.01
aspect-ratio	1 *–100	1.70	1.02

### 2.1.2. Gas-Phase CFD Transport Model

After the furnace ALD reactor geometry has been meshed, it is used by ANSYS Fluent for CFD calculation. The boundary and operating conditions, material properties, and solution methods are defined by the user-defined function (UDF) functionality in Fluent. The right side of the Figure 2b is formulated as the rotating axis for a 2D axis-symmetric model. In addition, a coupled pressure-based solver is utilized, which is appropriate for a low-speed ideal gas flowing through the furnace ALD reactor [32].

In order to solve the dynamic reactor profile and estimate the half-cycle time, a transient simulation is performed in Fluent. The time step of the transient model is critical to the efficiency and accuracy of the CFD model and it is governed by the Courant–Friedrichs–Lewy (CFL) condition, which is a popular criterion for PDE finite-difference approximation [33]:

$$C = \frac{u\Delta t}{\Delta x} \quad (1)$$

where  $C$  is the Courant number, a quantitative description of the number of computational cells that fluid moves through in one time step,  $u$  is the average flow velocity, and  $\Delta x$  is the minimum cell length. It is demonstrated that a commercial CFD software allows convergence with high Courant number around 50 [34]. The transport of species, energy, and momentum predicts the wafer surface profile to calculate for the ALD deposition. ANSYS Fluent obtains an accurate reactor dynamic profile by solving the coupled transient Navier–Stoke equations, shown in their tensor forms [29,35]:

$$\vec{J}_i = -\rho D_{m,i} \nabla Y_i - D_{T,i} \frac{\nabla T}{T} \quad (2)$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla \cdot (\rho \vec{v} Y_i) = -\nabla \cdot \vec{J}_i + R_i + S_i \quad (3)$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla \cdot (\vec{v}(\rho E + p)) = \nabla \cdot (k \nabla T - \Sigma h \vec{J} + (\bar{\tau} \vec{v})) + S_h \quad (4)$$

$$\frac{\partial(\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla P + \nabla \cdot (\bar{\tau}) + \rho \vec{g} + \vec{F} \quad (5)$$

$$\bar{\tau} = \mu \left[ (\nabla \vec{v} + \nabla \vec{v}^T) - \frac{2}{3} \nabla \cdot \vec{v} I \right] \quad (6)$$

where  $\vec{J}_i$  is the diffusive flux of species  $i$ ,  $\rho$  is the density of the gas-phase species mixture,  $T$  is the fluid temperature,  $D_{m,i}$  and  $D_{T,i}$  are the mass and heat diffusivities of species  $i$ ,  $Y_i$  is the mass fraction of species  $i$ ,  $\vec{v}$  is the velocity of gas-phase species mixture,  $R_i$  and  $S_i$  are the reaction and mass transfer source terms of species  $i$ , respectively,  $E$  is the total energy,  $\bar{\tau}$  is the stress tensor,  $S_h$  is the energy flux at

the boundary,  $P$  is the static pressure,  $\vec{g}$  is the component of gravity on the direction of the fluid flow,  $F$  is the mass flux at the boundary,  $\mu$  is the fluid viscosity and  $I$  is the unit tensor.

Despite the capability of CFD to solve an all-inclusive solution, it is still helpful to make reasonable physical assumptions to cut down the computation requirements. Due to the typical environment (high temperature, low pressure) in the ALD reactor, we assume the gas species to be ideal gas. Five gas species are incorporated in the Fluent simulation: BTBAS and ozone are the Si-cycle and O-cycle precursors, respectively; tert-Butylamine and oxygen are the Si-cycle and O-cycle by-products, respectively; argon is the purging gas between the two half-cycles. The thermodynamic and kinetics properties of ozone, oxygen, and argon are provided by the Fluent database, while the properties of BTBAS and tert-Butylamine (e.g., viscosity, specific heat ( $C_p$ ), standard enthalpy, and entropy) are defined by values computed from density functional theory (DFT) calculation and literature [14]. The mixture follows the ideal gas mixing law.

Moreover, it is important to examine the different forms of the species transport in the furnace ALD reactor as the wafer–wafer inter-distance is relatively small compared to the length scales of the rest of the reactor. From a preliminary velocity profile test, it is demonstrated that the velocity between wafers is nearly zero, and thus the species transport is limited by diffusion. Since there are more than two gas species with significant quantity co-existing and diffusing in the region, a full-component diffusion model needs to be utilized to account for the multi-component diffusion. To determine the effective diffusivity between species, the Chapman–Enskog formula, derived from the fundamental kinetic theory of gases, can be utilized to calculate the diffusion coefficient of each pair of species [36]:

$$D_{ij} = 0.00188 \frac{(T^3 (\frac{1}{M_{w,i}} + \frac{1}{M_{w,j}}))^{\frac{1}{2}}}{P_{abs} \sigma_{ij}^2 \Omega_D}$$

where  $P_{abs}$  is the local absolute pressure,  $\sigma_{ij}$  is the arithmetic average of the individual rigid elastic sphere diameter,  $\Omega_D$  is the diffusion collision integral, which is a function of the Lennard–Jones parameter  $\epsilon$  that represents the intermolecular interaction.

## 2.2. Surface Deposition Rate Calculation

In addition to the macroscopic model which determines the gas-phase transport phenomena, it is important to simulate the substrate surface dynamics including the details on film surface coverage and the transient surface deposition rate. Simple Langmuir adsorption models, often utilized to characterize surface deposition processes, do not provide a comprehensive deposition profile because of the competing reaction pathways, the complex reaction mechanisms, and the influences of the absorbed species and structural details. In our previous works, we constructed an on-lattice kinetic Monte-Carlo (kMC) model, which described the lattice structures and surface reactions in detail [12]. Based on the kMC simulation results, a deposition rate database and a Bayesian regularized artificial neural network (BRANN) model were developed to characterize the transient deposition under various precursor partial pressures, surface temperatures, and surface coverages [15]. Thus, in this work, we adopt the developed deposition rate database to simulate the ALD surface reaction, instead of running the computationally demanding kMC models.

## 2.3. Automated Workflow and Parallel Computation

The resulting multiscale CFD model involves more than 5000 cells and a long computational processing time. Therefore it would be extremely challenging to perform the computation task serially on a single core and would be impossible to perform any data mining. To improve the computational efficiency and allow the possibility for a comprehensive database generation, parallel computation needs to be employed for the multiscale CFD model. On the UCLA Hoffman2 Cluster, the parallel computation of the multiscale CFD model is implemented with the Message Passing Interface (MPI) [29]. Eight computation cores with 8G of RAM are chosen according to the batch array process availability, computational duration and the parallel communication overhead between host and nodes.

In addition, an automated workflow is integrated with the MPI and guides the connection and interaction between the microscopic and macroscopic domains. As shown in Figure 1, the blue region describes the parallel computation of the macroscopic gas-phase CFD model, the green region describes the surface deposition model, and the yellow components and arrows describe the logical connection and the direction of workflow. At the beginning of every time step,  $\Delta t$ , the CFD model solves for the transient mass and energy governing equation at that moment. After the convergence of the solution, the model provides surface boundary temperature, partial pressure and time step at five sub-regions on each wafer. This information is transferred to the surface deposition database, which returns the corresponding transient deposition rate to the CFD model. Using the transient deposition rate, the species fluxes are calculated for each sub-region on the wafer and are reported back as updated CFD boundary conditions for the next time step. Using the implicit Euler scheme, the system physical time is updated to  $t + \Delta t$  [29].

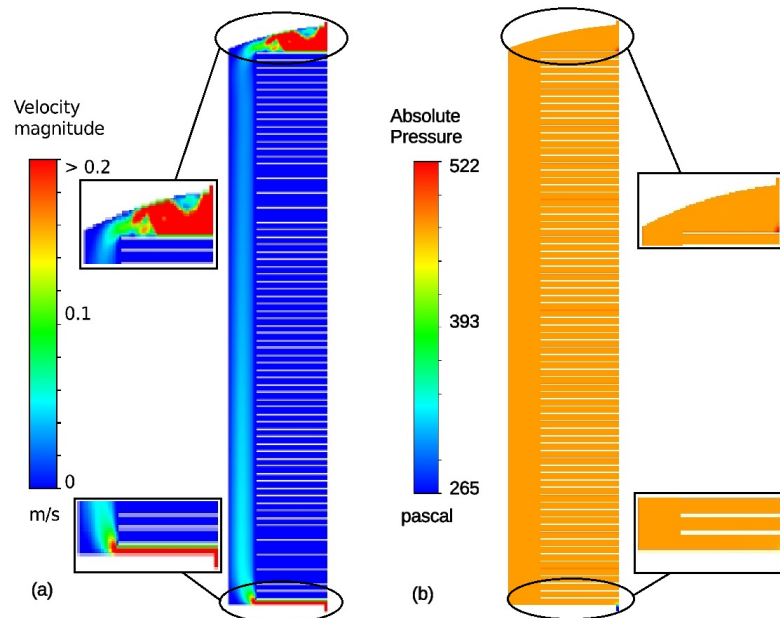
#### 2.4. Multiscale Furnace CFD Model Validation

In this subsection, the CFD model result for the furnace ALD system is validated with experimental results and existing simulation models. The current models reported in [8] and in [9] utilized empirical knowledge and a rudimentary gas-phase CFD computation to analyze the transport time scale. Specifically, the time scale results were compared to the 2D asymmetric multiscale CFD model in this paper under similar operating conditions: a precursor inlet flow rate of 1.3 slm, a precursor mole fraction of 0.3, and a vacuum pump setting of 100 Pa and 266 Pa. In addition, for the substrate surface heating from top to bottom, a temperature gradient from 475 K to 615 K with an interval of 10 K for each susceptor was implemented to enhance the surface deposition reaction for the bottom wafers.

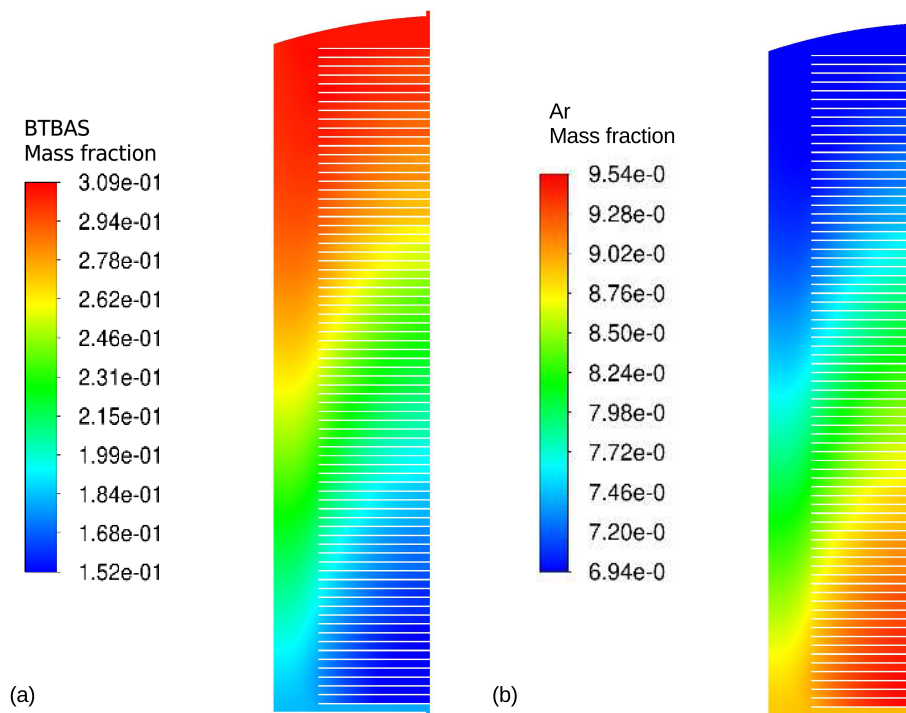
Two types of species transport exist in the furnace ALD reactor as demonstrated by the velocity profile in Figure 3a, which was chosen at a random process time for illustration purposes. The downward transport in the vertical region besides the wafers was a combination of diffusion and convection, as indicated by an asymmetric velocity magnitude of 0.05 m/s, which is equivalent to a vertical velocity magnitude of 0.008 m/s. In the inter-wafer region, the species transport was diffusion-only in the radial direction because the velocity was approaching zero. In addition, under a rotary vacuum pump setting of 12 M<sup>3</sup>/h, the corresponding operating pressure, which is conventionally defined as the pressure towards the reactor outlet should be 266 Pa. Nevertheless, because of the complex parallel wafer geometry and the diffusion stagnation in between the wafers, the absolute pressure was higher in the reactor because of the static pressure, reaching around 500 Pa, as shown in Figure 3b. The entirely orange region is a result of the diffusion dominant mass transfer. In the majority part of the reactor, the velocity magnitude was less than 0.1 m/s, which is a nominal value for gas-phase transport. This small magnitude validates the lack of apparent pressure gradient in the reactor main body. Meanwhile, there were sharp pressure gradients at the inlet and outlet, as a result of the converged gas streams.

Because of the combined convection and diffusion flow regime and the reactor layout, the precursor gas first reached the edges of upper wafers under the diffusion across the side of the furnace, and finally reached the center of the bottom wafers under diffusion into the wafer, as shown in Figure 4a. In the meantime, as the precursor came into the reactor and in between the wafers, Ar was purged out concurrently, as shown in Figure 4b. The corresponding dynamic gas-phase BTBAS partial pressure development profiles at the top, middle, and bottom wafer surfaces are shown in Figure 5a. Additionally, the transient film coverage developments at the middle and the peripheral wafer regions are shown in Figure 5b. The gas-phase dynamics and the surface deposition at different reactor locations were demonstrated to be similar to the reported profiles in [8]. In addition, the half-cycle time, at which all the wafers reach full coverage, ranged from 41 to 100 s at a vacuum level of 100 to 266 Pa, respectively, which are comparable to the experimental results in a furnace ALD reactor [18]. The discrepancies between the multiscale CFD model results and the reported values

are due to the differences in precursor reaction mechanism, the number of loaded wafers, and the need for a more comprehensive modeling of the interaction between diffusion and surface deposition. Moreover, with the validated model, the half-cycle time, the precursor partial pressure at the gas sensor position, and the wafer coverage of the bottom substrate were taken as measured observables as discussed in the next section for the formulated control schemes.

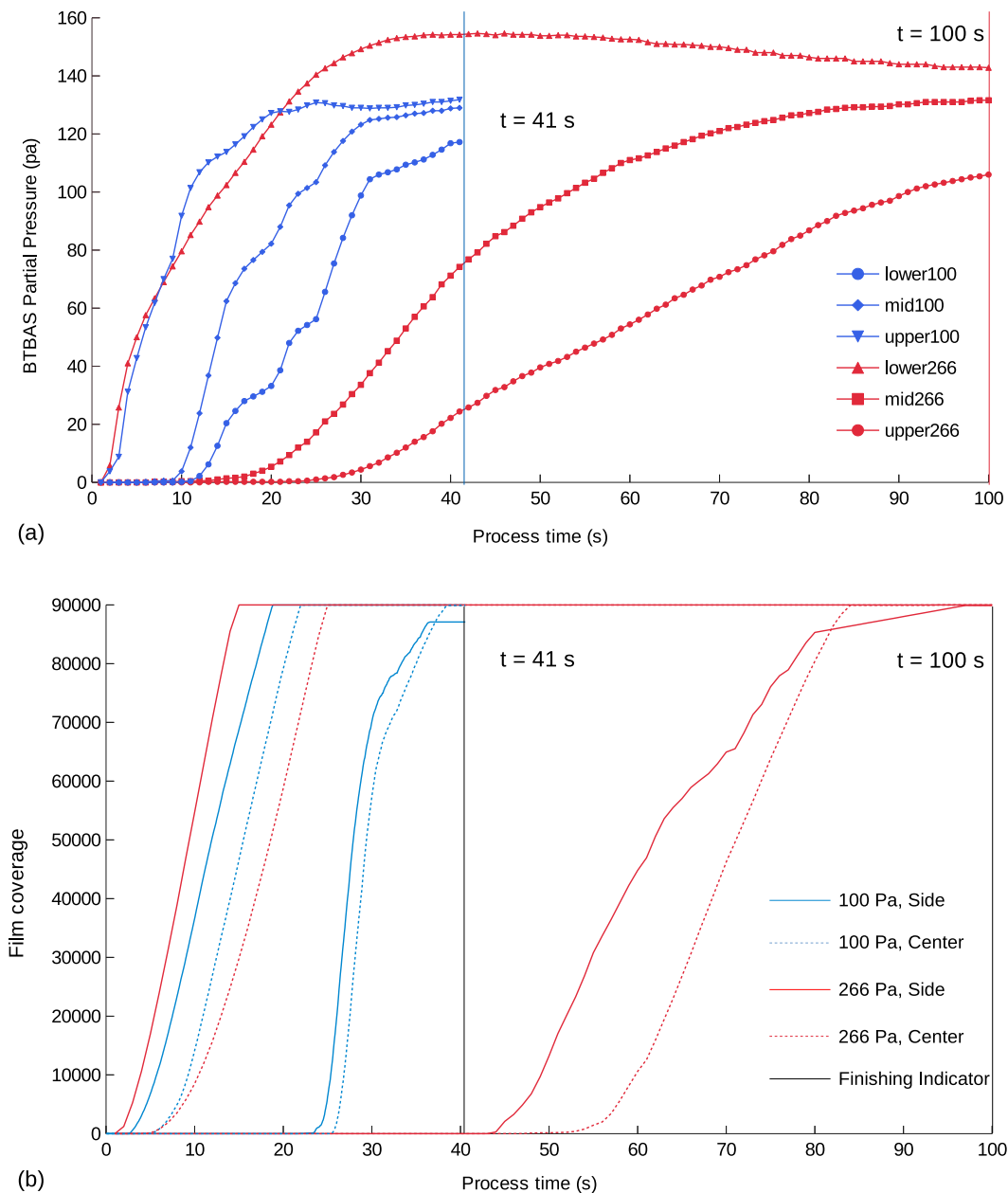


**Figure 3.** Computational fluid dynamics (CFD) species transport profiles at a randomly chosen time for simulation illustration ( $t = 60$  s) at an operating pressure of 266 Pa. (a) BTBAS mass fraction profile; (b) Ar mass fraction profile.



**Figure 4.** CFD species transport profiles at a randomly chosen transient time step for simulation illustration ( $t = 60$  s). (a) BTBAS mass fraction profile. (b) Ar mass fraction profile.





**Figure 5.** (a) BTBAS partial pressure evolution at various reactor levels. (b) Film coverage profile on the top and bottom substrates, which are located on the left side and the right side of the figure, respectively.

### 3. Control Schemes

As introduced in Section 1, many types of disturbances are prevalent during the operation of the bulky furnace ALD reactor, both between batches and in real-time. For example, reactor variations might happen between consecutive batches because of human manipulation or due to undesired deposition onto the reactor wall. In addition, disturbances (e.g., on vacuum pump) may occur in each batch half-cycle, which influences the gas-phase development and the resulting surface deposition. Thus, although a prior knowledge of an optimal operating condition can be postulated by the multiscale CFD model, this knowledge does not guarantee the optimal reactor operation under disturbances. Since these disturbances will lead to a substantial variation from full coverage at the target half-cycle time, on-line control strategies can significantly aid the performance of batch ALD processes. Thus, in this section, we formulate two control schemes that are integrated together to minimize reactor

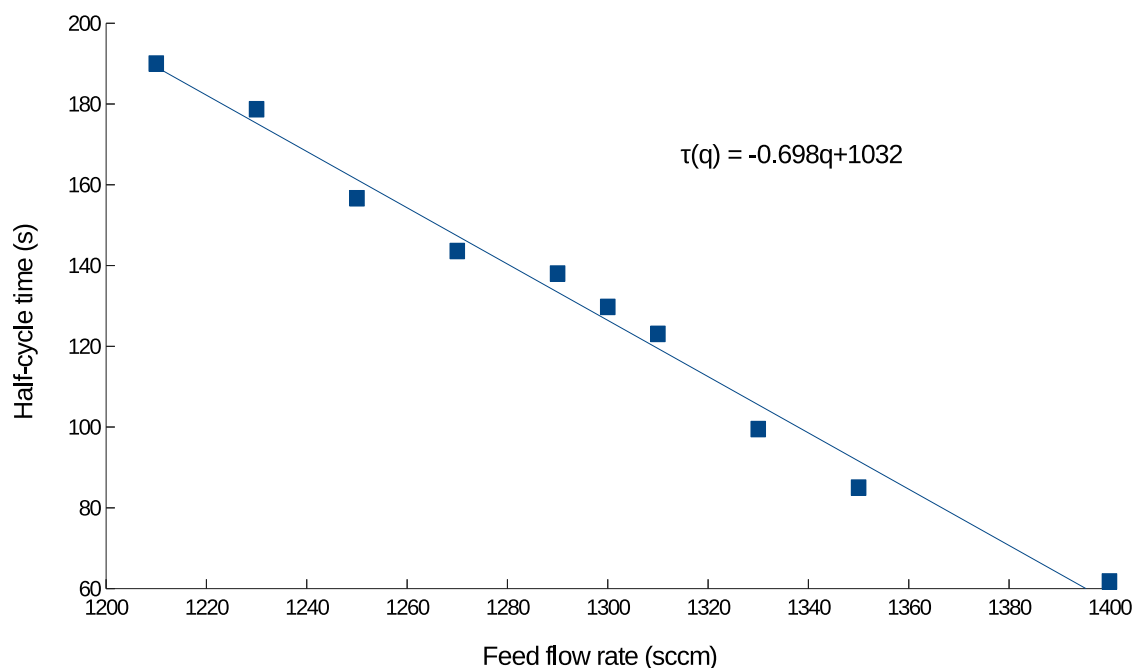
disturbances. First, an R2R controller is constructed, which adjusts the inlet feed flow rate according to the altered half-cycle time under reactor variation between batches. Next, a real-time PI control scheme is adopted, which varies the inlet feed composition to track the partial pressure set-point, thus balancing the vacuum pump disturbances. After each controller is tuned, they are integrated and their combined closed-loop performance is evaluated.

### 3.1. R2R Control of Half-Cycle Time

To gradually handle the operational variations between batches, an R2R control scheme is proposed. For a plasma-enhanced CVD system, which has a comparable operating environment and length scale to ALD systems, Ref. [1] recently proposed an R2R control scheme and demonstrated its ability to enhance the deposited film uniformity. As a result, a similar control scheme is used in the furnace ALD process. Triggered by the measured reactor variation, the operating conditions are fine-tuned based on an exponentially weighted moving average (EWMA) algorithm to minimize the effect of disturbances in the gas-phase transport, and thus, maintaining full film coverage.

#### 3.1.1. Input–Output Relationship

To implement the R2R control, the key input–output relationship between the inlet feed flow rate and the resulting half-cycle time needs to be determined from the multiscale CFD model. Due to the large physical scale and the complex geometry in the furnace ALD system, this input–output relationship is non-linear and not monotonic in the complete operating region. However, in this scenario, we only investigate a focused operating domain that is around the industry standard operating condition. The industry standard half-cycle time, 100 s, is chosen as the set-point,  $\tau_{sp}$ , and the corresponding set-point feed flow rate,  $q_{sp}$ , is equal to 1330 sccm. Under the ideal operating pressure of 266 Pa, it is ensured that full coverage can be achieved in all deposited thin-films. In this operating region, the mapping between parameters is linear with an R-squared value of 0.98, as shown in Figure 6.



**Figure 6.** Relationship between the half-cycle time and the feed flow rate. The blue dots represent the half-cycle time from the simulation and the blue line represents the regression fitting in the operating region.

### 3.1.2. EWMA Formulation

Using the post-batch measurement as the feedback to the R2R control scheme, the operating condition update can be calculated with the exponentially weighted moving average (EWMA) algorithm. EWMA statistically monitors the process measurement and updates the decision in a way that historical information is considered but weighted less with respect to later times. Specifically, in this paper, the algorithm adds a chosen learning factor to exponentially decrease the weight of the previous process output, which is the half-cycle time from the previous batch and corrects the model input, which is the inlet feed flow rate. The EWMA algorithm is formulated as follows:

$$\begin{aligned}\epsilon_{k+1} &= (1 - \lambda)\epsilon_k + \lambda(q_{sp} - q_{cfd}) \\ q_{k+1} &= q_k + \epsilon_{k+1}\end{aligned}\quad (7)$$

where  $\epsilon_k$  and  $\epsilon_{k+1}$  are the correction factors for the operating condition for the  $k$  batch and  $k + 1$  batch, respectively,  $\lambda \in [0, 1]$  is the learning factor,  $q_{cfd}$  is the equivalent feed flow rate under the ideal operating condition, calculated with the post-batch half-cycle measurement from the CFD simulation, and  $q_k$  and  $q_{k+1}$  are the manipulated feed flow rates of the  $k$  batch and  $k + 1$  batch, respectively.

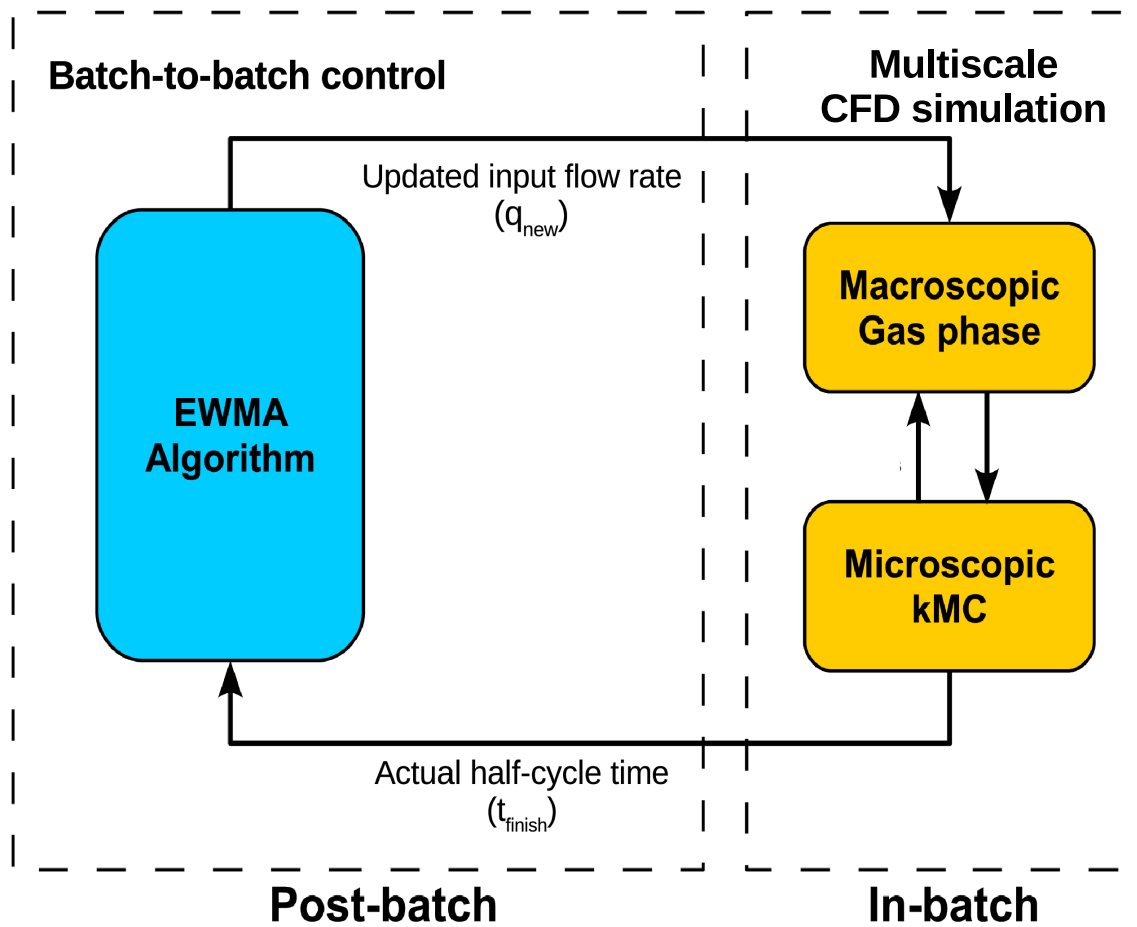
### 3.1.3. Closed-Loop Operation Results under R2R Control

Based on the EWMA algorithm, the R2R controller is integrated with the multiscale CFD model for a closed-loop simulation. As shown in Figure 7, first the multiscale CFD model simulates the half-cycle time needed with an inlet flow rate of  $q_k$  under a typical disturbance, which is an outlet vacuum pump malfunction that drives the equivalent operating pressure to 500 Pa. In the actual industrial manufacturing, a quartz crystal microbalance (QCM) device placed on the bottom wafers can be used to acquire the full coverage time. The simulated half-cycle time from the CFD model is used to calculate the effective inlet flow rate  $q_{cfd}$  with the input-output relationship, and with the EWMA algorithm, the inlet flow rate is updated to  $q_{k+1}$  for the next deposition (next CFD simulation). This process is reiterated until the simulated half-cycle time reaches the 100 s set-point. Two learning factors are tested for fast and slow learning rates, respectively, with  $\lambda_1 = 0.25$  and  $\lambda_2 = 0.75$ .

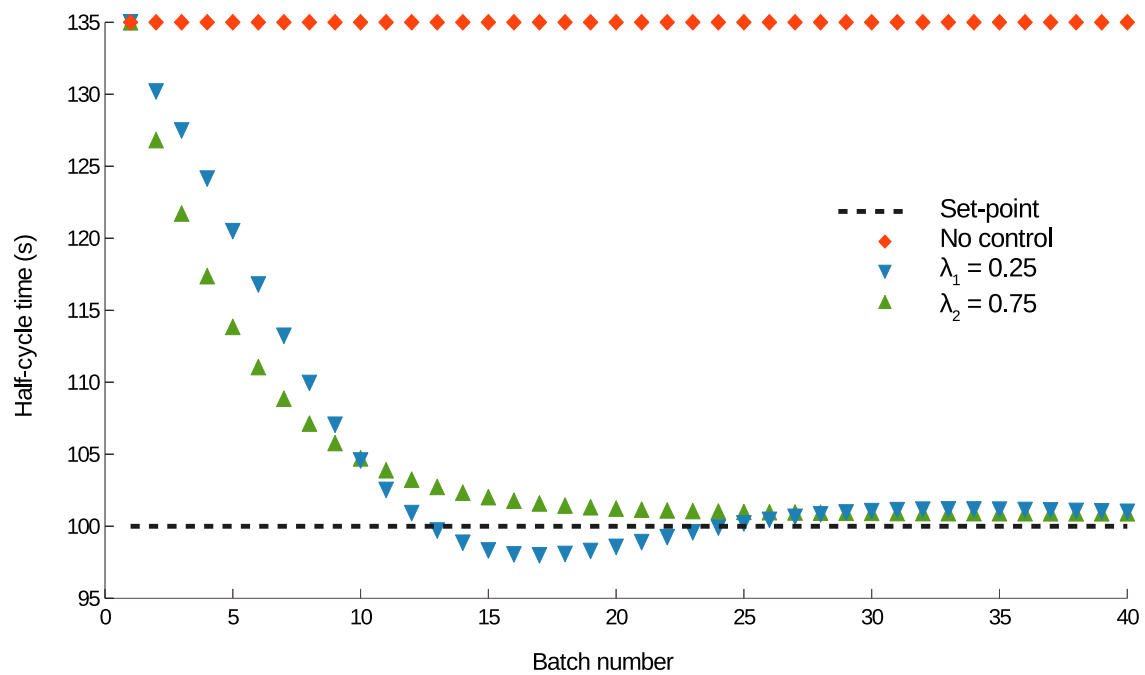
As shown in Figure 8, upward (green) and downward (blue) triangular data points represent the R2R closed-loop system with learning factors  $\lambda$  of 0.75 and 0.25, respectively, and the diamond (red) data points represent the reactor response under the existing pressure disturbance. It is demonstrated that, under the disturbances with no control scheme, the batch ALD reactor is negatively affected. At 100 s, 40% of the wafers cannot reach full coverage, and an alternative half-cycle time of 135 s is required instead for all wafers to reach full coverage. With the R2R control scheme, the input feed flow rate increases gradually, leading to a reduced half-cycle time that eventually achieve the set-point. Also, it is demonstrated that the number of batches required to reach set-point is affected by the learning factor magnitude. For  $\lambda_2 = 0.75$  and  $\lambda_1 = 0.25$ , the half-cycle time first reaches the set-point at about 25 and 12 batches, respectively.

## 3.2. PI Control of Precursor Partial Pressure

In addition to the R2R control scheme, it is crucial to reduce the effect of disturbances in real-time. As introduced in Section 1, feedback control schemes, for instance proportional-integral (PI) control and model predictive control (MPC), have been applied to single wafer ALD systems in previous works. It is also important to implement real-time control schemes in multi-wafer ALD reactors, which will lead to a quite substantial economic benefit. Thus, in this work, a PI control scheme is formulated for the furnace ALD system to counteract process disturbances.



**Figure 7.** Schematic illustration of the closed-loop system under the post-batch run-to-run (R2R) controller integrated with the multiscale CFD furnace ALD reactor model.



**Figure 8.** Closed-loop half-cycle time correction via an R2R controller under vacuum pump variation.

### 3.2.1. PI Controller Design

Similar to the R2R controller, the PI controller is investigated for the standard operating condition. Although the thin-film growth is a batch process, the flow field development in the reactor resembles that of a plug flow reactor at its start-up phase. A gas sensor positioned at the top of the reactor besides the top wafer is chosen for BTBAS partial pressure measurement. Using direct analysis in real time (DART) or a quadruple mass spectrometer (QMS), where an ion source can be used to produce electronically or vibronically excited-state inert species for dopant or ionization, the BTBAS partial pressure can be measured in real-time [37,38]. This gas sensing technique has been proven to be compatible for the ALD process [39]. Also, the chosen precursor BTBAS is a perfectly flammable precursor, suitable for the aforementioned measurement devices [40].

After around 15 s in each batch, the BTBAS partial pressure at the measurement position reaches the steady-state value of 134.82 Pa, which is the desired partial pressure set-point for the PI controller. Since the rest of the reactor will eventually reach the required partial pressure as long as the sensor position reaches the partial pressure threshold, only the pressure at the sensor position is controlled. Moreover, this threshold partial pressure allows full-coverage for all wafers, despite the vacuum pump variation, which mainly influences the gas-phase transport. To maintain the normal reactor operation and the desired deposition rate, a PI control scheme is used to manipulate the inlet BTBAS mole fraction to drive the partial pressure to the set-point value as shown below:

$$u(t) = K_C \left( e(t) + \frac{1}{\tau_I} \int_0^t e(t') dt' \right) \quad (8)$$

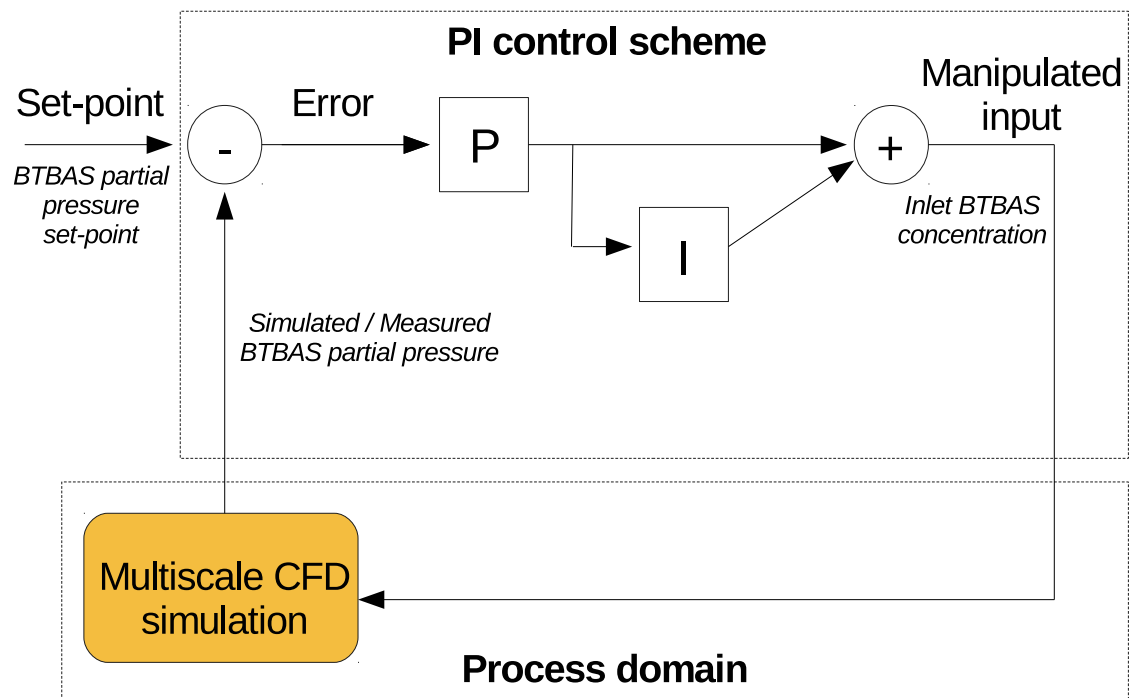
where  $t$  is the process time, and  $K_C$  and  $\tau_I$  are both non-negative controller tuning parameters, denoting the magnitude for the proportional (P) and integral (I) terms, respectively.

### 3.2.2. PI Controller Tuning

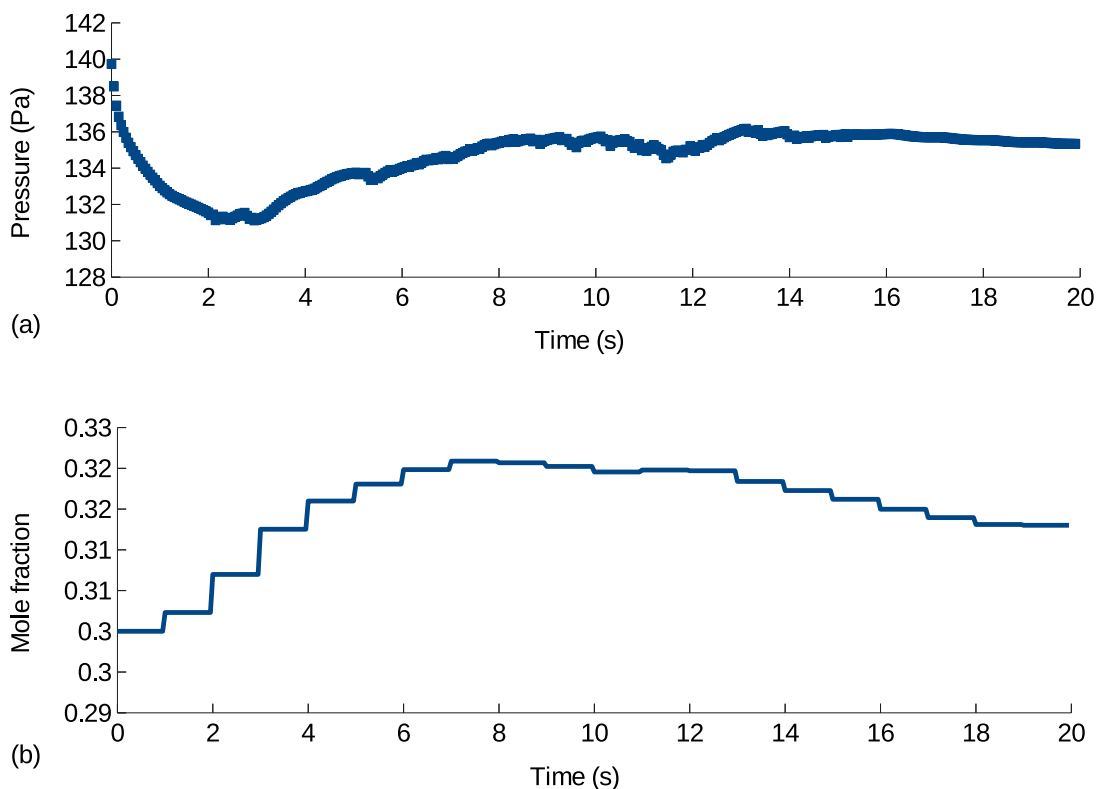
In order to determine the appropriate PI controller parameters  $K_C$  and  $\tau_I$ , Cohen–Coon tuning rules, the standard PI tuning rule appropriate for non-interactive controller, is utilized to provide an initial estimate for the tuning parameters. First, a step response test is implemented to analyze three important response characteristics: time constant, process gain and dead time, to provide the initial estimate of PI parameters. After the partial pressure at the gas sensor position has reached steady-state, a series of step changes from 13% to 100% of the original inlet precursor mole fraction are tested, while preserving the same inlet feed flow rate. Under associate step responses, the precursor partial pressure at the gas sensor position increases correspondingly, providing the process response characteristics. Then, combining the results from multiple tests and tunings with the CFD simulation, we arrive at an optimal set of tuning parameters:  $K_C = 0.712$  and  $\tau_i = 28.5$ .

### 3.2.3. Closed-Loop Operation Results under PI Control

With the calculated set of PI controller parameters, the closed-loop control system is constructed as shown in Figure 9. The operating pressure considered here is again a vacuum pump disturbance, which drives the equivalent operating pressure to be 100 Pa. Triggered by the gas sensor measurable variation, which is selected to be a difference of 2% in partial pressure per time step from the steady-state value, the PI controller applies real-time feedback control. As shown in Figure 10a, it allows the BTBAS partial pressure to reach the set-point value in around 18 s with the increased feed BTBAS mole fraction shown in Figure 10b. This BTBAS partial pressure valve allows full coverage to be achieved for all wafers but leads to a variation of the half-cycle time.



**Figure 9.** Schematic illustration of the closed-loop system under the PI controller and the multiscale CFD model.

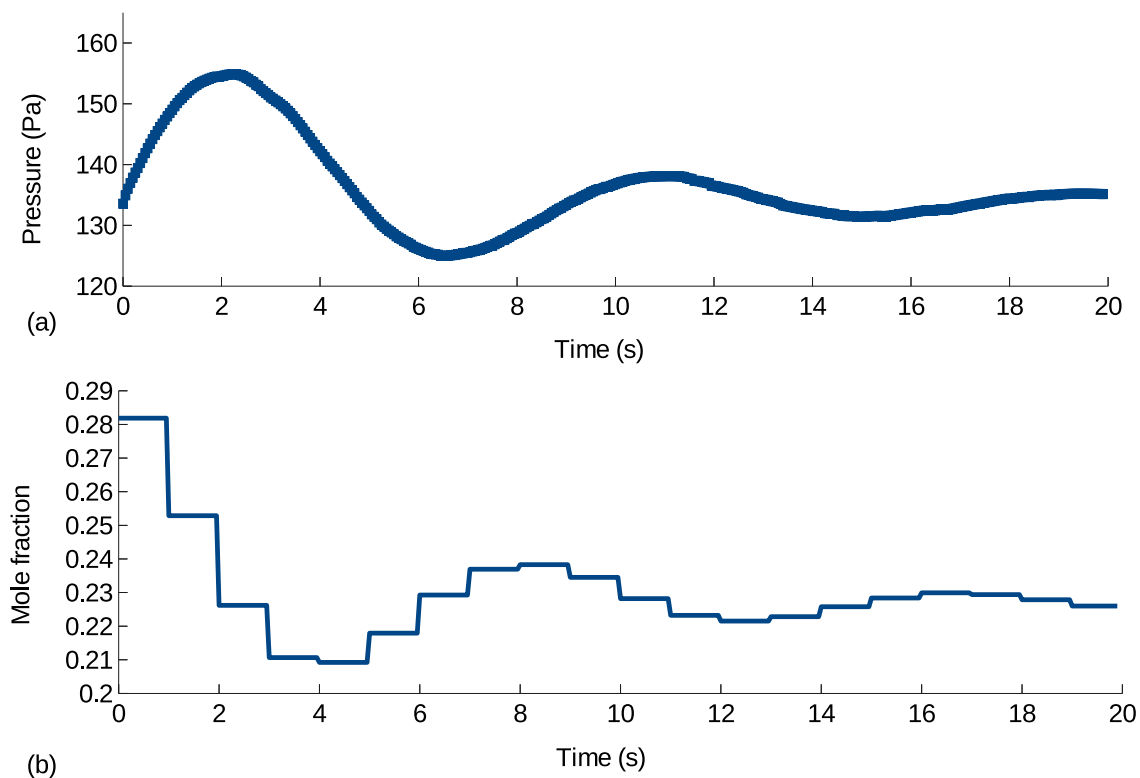


**Figure 10.** (a) Closed-loop BTBAS precursor partial pressure set-point tracking under vacuum pump disturbance, driving the operating pressure to be 100 Pa. The  $x$ -axis is the process time starting when the PI controller is initially triggered. The  $y$ -axis is the BTBAS partial pressure at the gas sensor position. (b) Manipulated input profile. The  $y$ -axis is the manipulated inlet BTBAS mole fraction.

### 3.3. Integrated R2R/PI Control Implementation

After the PI and R2R control schemes were separately tuned and validated, the two controllers were integrated to ensure full coverage on all wafers under the assigned half-cycle time. During the first batch of ALD, after the vacuum pump disturbance occurs, the PI controller was implemented with the same activation criterion. At the end of the first batch, the R2R controller was triggered, utilizing the EWMA algorithm with input–output fitting under the updated inlet precursor mole fraction.

To validate the performance of the integrated R2R/PI control scheme, the vacuum pump disturbance was again introduced in the CFD simulation, which drove the equivalent operating pressure of the reactor to be at 500 Pa. For the real-time feedback control, it was demonstrated that a reduction of inlet BTBAS mole fraction was required to reduce the amount of precursor needed at this pressure. As shown in Figure 11, a series of control actions with a sampling time of 1 s were carried out according to the PI control scheme, which reduced the inlet feed mole fraction to 0.226 and drove the precursor partial pressure to the set-point and stabilized the process output after 18 s. Also, the controller successfully reduced 23% of the precursor usage in the first batch.



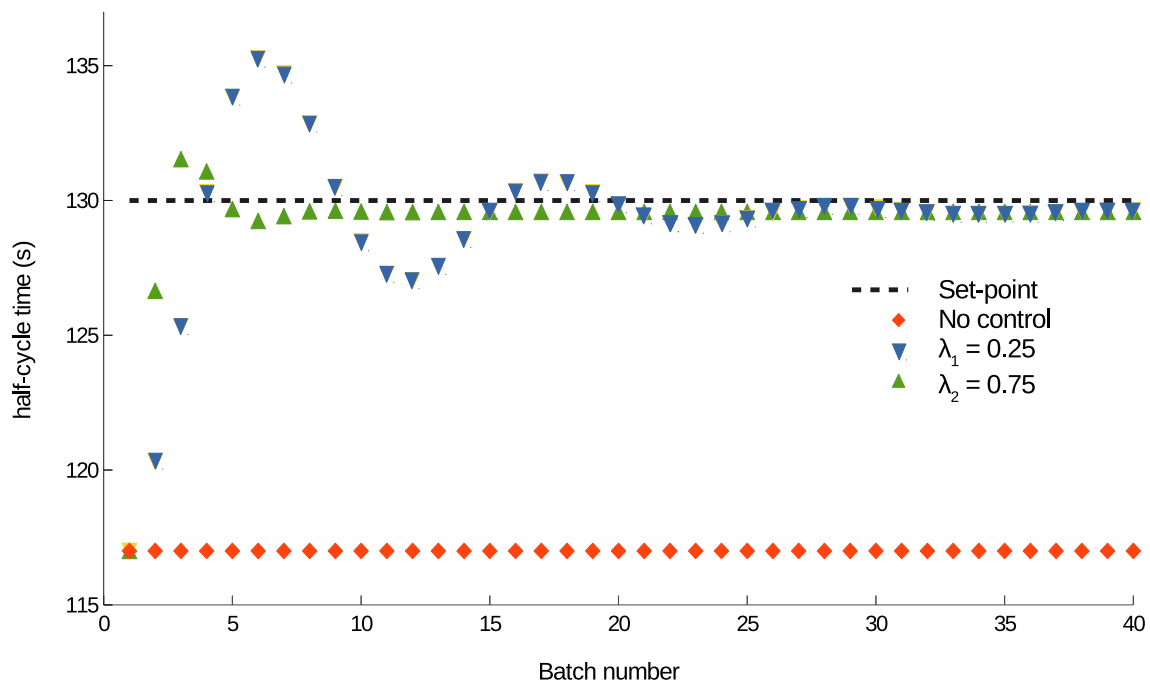
**Figure 11.** (a) Closed-loop BTBAS precursor partial pressure set-point tracking under a disturbance in the vacuum pump, driving the operating pressure to be 500 Pa. The  $x$ -axis is the process time starting when the PI controller is initially triggered. The  $y$ -axis is the BTBAS partial pressure at the gas sensor position. (b) Manipulated input profile. The  $y$ -axis is the manipulated inlet BTBAS mole fraction.

With the updated feed mole fraction determined from the PI controller, the relationship between the inlet flow rate and the half-cycle time,  $\tau$ , under the standard inlet flow rate,  $q$ , range was again determined to be:

$$\tau(q) = -0.11q + 322.16 \quad (9)$$

According to the reactor response under the updated feed mole fraction, a half-cycle time of 130 s was chosen to be the set-point for the R2R control, and a corresponding  $q_{sp}$  is 1700 sccm. As shown in Figure 12, the diamond (red) data points represent the simulated half-cycle time under the existing

pressure disturbance, and the downward (blue) and upward (green) triangular data points represent those under the R2R control schemes with a learning factor  $\lambda$  of 0.25 and 0.75, respectively. It was demonstrated that under the proposed R2R control scheme, the input feed flow rate gradually decreased, leading to an increased half-cycle time, which eventually reached the set-point. Also, as shown in Figure 12, both learning rates allowed convergence towards the half-cycle set-point, but a learning factor  $\lambda = 0.75$  allowed faster convergences and less overshoot. In addition, without the R2R control scheme, the performance of the batch ALD reactor was negatively affected by the disturbance, resulting in a 5.7% waste of precursors.



**Figure 12.** Closed-loop half-cycle time correction via an R2R controller under the updated feed precursor mole fraction and the vacuum pump variation.

#### 4. Conclusions

In this paper, a 2D-axisymmetric multiscale CFD simulation model of a multi-wafer furnace ALD system and the related feedback and run-to-run control schemes were constructed. Specifically, a vertical industrial-scale furnace ALD reactor was investigated for a  $\text{SiO}_2$  thin-film ALD using BTBAS and ozone as precursors. Firstly, the multiscale CFD model was built and validated, which characterized the dynamic gas-phase transport and the substrate deposition profiles at various wafer locations and generated a valid estimation of the half-cycle time. Next, an R2R controller was built, which successfully updated the inlet flow rate according to the half-cycle time feedback information based on the EWMA algorithm. In addition, a PI controller was formulated, which successfully adjusted the feed precursor mole fraction to maintain the precursor partial pressure at its set-point. The two controllers were combined and validated with a vacuum pump disturbance, which raised the operating pressure to 500 Pa. The resulting integrated control system regulated the precursor partial pressure and the half-cycle time to their respective set-points. Furthermore, full film coverage was demonstrated to be achieved for all substrates, while the total precursor cost was reduced by 31.1%.

**Author Contributions:** Y.Z. developed the idea of incorporating feedback control and run-to-run control in multi-wafer thermal ALD of thin films; Y.Z. and Y.D. developed the simulation models, performed the simulation study, and prepared the initial draft of the paper; P.D.C. oversaw all aspects of the research and revised this manuscript. All authors have read and agreed to the published version of the manuscript.



**Acknowledgments:** Financial support from the National Science Foundation and the Department of Energy is gratefully acknowledged.

**Conflicts of Interest:** The authors declare that they have no conflict of interest regarding the publication of this research article.

## References

1. Crose, M.; Kwon, J.S.I.; Tran, A.; Christofides, P.D. Multiscale modeling and run-to-run control of PECVD of thin film solar cells. *Renew. Energy* **2017**, *100*, 129–140. [[CrossRef](#)]
2. George, S.M.; Ott, A.W.; Klaus, J.W. Surface Chemistry for Atomic Layer Growth. *J. Phys. Chem.* **1996**, *100*, 13121–13131. [[CrossRef](#)]
3. Schuegraf, K.; Abraham, M.C.; Brand, A.; Naik, M.; Thakur, R. Semiconductor Logic Technology Innovation to Achieve Sub-10 nm Manufacturing. *IEEE J. Electron Devices Soc.* **2013**, *1*, 66–75. [[CrossRef](#)]
4. Kääriäinen, T.; Cameron, D.; Kääriäinen, M.L.; Sherman, A. *Atomic Layer Deposition: Principles, Characteristics, and Nanotechnology Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
5. George, S.M. Atomic Layer Deposition: An Overview. *Chem. Rev.* **2009**, *110*, 111–131. [[CrossRef](#)] [[PubMed](#)]
6. Poodt, P.; Cameron, D.C.; Dickey, E.; George, S.M.; Kuznetsov, V.; Parsons, G.N.; Roozeboom, F.; Sundaram, G.; Vermeer, A. Spatial atomic layer deposition: A route towards further industrialization of atomic layer deposition. *J. Vacuum Sci. Technol. A Vacuum Surf. Films* **2012**, *30*, 010802. [[CrossRef](#)]
7. Cook, R.; Stevens, R.; Schwartz, P.; Tejamo, C.; Nilsen, V.; Ormonde, G.; Paranjpe, A.; Nag, S.; Patten, M. Substrate Carrier for Parallel Wafer Processing Reactor. U.S. Patent 10/966,245, 1 September 2005.
8. Lankhorst, A.; Paarhuis, B.; Terhorst, H.; Simons, P.; Kleijn, C. Transient ALD simulations for a multi-wafer reactor with trenched wafers. *Surf. Coat. Technol.* **2007**, *201*, 8842–8848. [[CrossRef](#)]
9. Dingemans, G.; Jongbloed, B.; Knaepen, W.; Pierreux, D.; Jdira, L.; Terhorst, H. Merits of Batch ALD. *ECS Trans.* **2014**, *64*, 35–49. [[CrossRef](#)]
10. Christofides, P.D.; Armaou, A.; Lou, Y.; Varshney, A. *Control and Optimization of Multiscale Process Systems*; Birkhäuser: Boston, MA, USA, 2008.
11. Shirazi, M.; Elliott, S.D. Atomistic kinetic Monte Carlo study of atomic layer deposition derived from density functional theory. *J. Comput. Chem.* **2014**, *35*, 244–259. [[CrossRef](#)]
12. Ding, Y.; Zhang, Y.; Kim, K.; Tran, A.; Wu, Z.; Christofides, P.D. Microscopic modeling and optimal operation of thermal atomic layer deposition. *Chem. Eng. Res. Des.* **2019**, *145*, 159–172. [[CrossRef](#)]
13. Deng, Z.; He, W.; Duan, C.; Shan, B.; Chen, R. Atomic layer deposition process optimization by computational fluid dynamics. *Vacuum* **2016**, *123*, 103–110. [[CrossRef](#)]
14. Zhang, Y.; Ding, Y.; Christofides, P.D. Multiscale computational fluid dynamics modeling of thermal atomic layer deposition with application to chamber design. *Chem. Eng. Res. Des.* **2019**, *147*, 529–544. [[CrossRef](#)]
15. Ding, Y.; Zhang, Y.; Ren, Y.M.; Orkoulas, G.; Christofides, P.D. Machine learning-based modeling and operation for ALD of SiO<sub>2</sub> thin-films using data from a multiscale CFD simulation. *Chem. Eng. Res. Des.* **2019**, *151*, 131–145. [[CrossRef](#)]
16. Lu, S.; Chang, Y.; Sun, D.; Dang, V.; Yang, M.X.; Chang, A.; Nguyen, A.N.; Xi, M. Valve Control System for Atomic Layer Deposition Chamber. U.S. Patent 6,734,020, 11 May 2004.
17. Shinde, S.; Sonar, A.; Sun, Y. Advanced process control for furnace systems in semiconductor manufacturing. In Proceedings of the ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference, Saratoga Springs, NY, USA, 14–16 May 2013; pp. 275–279.
18. Granneman, E.; Fischer, P.; Pierreux, D.; Terhorst, H.; Zagwijn, P. Batch ALD: Characteristics, comparison with single wafer ALD, and examples. *Surf. Coat. Technol.* **2007**, *201*, 8899–8907. [[CrossRef](#)]
19. Armaou, A.; Christofides, P.D. Plasma enhanced chemical vapor deposition: Modeling and control. *Chem. Eng. Sci.* **1999**, *54*, 3305–3314. [[CrossRef](#)]
20. Rasoulilian, S.; Ricardez-Sandoval, L.A. Robust multivariable estimation and control in an epitaxial thin film growth process under uncertainty. *J. Process Control* **2015**, *34*, 70–81. [[CrossRef](#)]
21. Rasoulilian, S.; Ricardez-Sandoval, L.A. A robust nonlinear model predictive controller for a multiscale thin film deposition process. *Chem. Eng. Sci.* **2015**, *136*, 38–49. [[CrossRef](#)]
22. Rasoulilian, S.; Ricardez-Sandoval, L.A. Stochastic nonlinear model predictive control applied to a thin film deposition process under uncertainty. *Chem. Eng. Sci.* **2016**, *140*, 90–103. [[CrossRef](#)]

23. Chaffart, D.; Ricardez-Sandoval, L.A. Optimization and control of a thin film growth process: A hybrid first principles/artificial neural network based multiscale modelling approach. *Comput. Chem. Eng.* **2018**, *119*, 465–479. [[CrossRef](#)]
24. Chen, P.; Mitsui, T.; Farmer, D.B.; Golovchenko, J.; Gordon, R.G.; Branton, D. Atomic layer deposition to fine-tune the surface properties and diameters of fabricated nanopores. *Nano Lett.* **2004**, *4*, 1333–1337. [[CrossRef](#)]
25. He, W.J.; Zhang, H.T.; Chen, Z.; Chu, B.; Cao, K.; Shan, B.; Chen, R. Generalized predictive control of temperature on an atomic layer deposition reactor. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 2408–2415. [[CrossRef](#)]
26. Burnak, B.; Diangelakis, N.A.; Pistikopoulos, E.N. Towards the Grand Unification of Process Design, Scheduling, and Control Utopia or Reality? *Processes* **2019**, *7*, 461. [[CrossRef](#)]
27. Rafiei, M.; Ricardez-Sandoval, L.A. New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability. *Comput. Chem. Eng.* **2020**, *132*, 106610. [[CrossRef](#)]
28. Elam, J.; Groner, M.; George, S. Viscous flow reactor with quartz crystal microbalance for thin film growth by atomic layer deposition. *Rev. Sci. Instrum.* **2002**, *73*, 2981–2987. [[CrossRef](#)]
29. Fluent, A. *ANSYS Fluent Theory Guide 15.0*; ANSYS: Canonsburg, PA, USA, 2013.
30. Eymard, R.; Gallouët, T.; Herbin, R. Finite volume methods. *Handb. Numer. Anal.* **2000**, *7*, 713–1018.
31. Tran, A.; Aguirre, A.; Durand, H.; Crose, M.; Christofides, P.D. CFD modeling of a industrial-scale steam methane reforming furnace. *Chem. Eng. Sci.* **2017**, *171*, 576–598. [[CrossRef](#)]
32. Chorin, A.J. Numerical solution of the Navier-Stokes equations. *Math. Comput.* **1968**, *22*, 745–762. [[CrossRef](#)]
33. Courant, R.; Friedrichs, K.; Lewy, H. On the partial difference equations of mathematical physics. *IBM J. Res. Dev.* **1967**, *11*, 215–234. [[CrossRef](#)]
34. Moura, C.; Kubrusly, C.S. *The Courant-Friedrichs-Lewy (CFL) Condition: 80 Years After Its Discovery*; Birkhäuser: New York, NY, USA, 2012.
35. Acton, Q.A. *Chemical Processes-Advances in Research and Application: 2012 Edition: ScholarlyBrief*; ScholarlyEditions: Atlanta, GA, USA, 2012.
36. Grad, H. Principles of the kinetic theory of gases. In *Thermodynamik der Gase/Thermodynamics of Gases*; Springer: Berlin, Germany, 1958; pp. 205–294.
37. Jousten, K.; Putzke, S.; Buthig, J. Partial pressure measurement standard for characterizing partial pressure analyzers and measuring outgassing rates. *J. Vacuum Sci. Technol. A Vacuum Surf. Films* **2015**, *33*, 061603. [[CrossRef](#)]
38. Fedele, A.; Pedone, M.; Moretti, R.; Wiersberg, T.; Somma, R.; Troise, C.; De Natale, G. Real-time quadrupole mass spectrometry of hydrothermal gases from the unstable Pisciarelli fumaroles (Campi Flegrei): Trends, challenges and processes. *Int. J. Mass Spectrom.* **2017**, *415*, 44–54. [[CrossRef](#)]
39. Leskelä, M.; Ritala, M. Atomic layer deposition (ALD): From precursors to thin film structures. *Thin Solid Films* **2002**, *409*, 138–146. [[CrossRef](#)]
40. Kher, S.; Nguyen, S.; Narwankar, P.; Tandon, S.; Jumper, S.; Sermona, V. Precursors and Hardware for CVD and ALD. U.S. Patent 11/847,158, 13 March 2008.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).