Title

Multiscale Computational Fluid Dynamics Modeling of Thermal and Plasma Atomic Layer Deposition: Application to Chamber Design and Process Control

Permalink

Author

Zhang, Yichi

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Multiscale Computational Fluid Dynamics Modeling of Thermal and Plasma Atomic Layer

Deposition: Application to Chamber Design and Process Control

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemical Engineering

by

Yichi Zhang

2021

ABSTRACT OF THE DISSERTATION

Multiscale Computational Fluid Dynamics Modeling of Thermal and Plasma Atomic Layer

Deposition: Application to Chamber Design and Process Control

by

Yichi Zhang

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2021

Professor Panagiotis D. Christofides, Chair

Facilitated by the increasing importance and demand of semiconductors for the smartphone and even the automobile industry, thermal atomic layer deposition (ALD) has gained tremendous industrial interest as it offers a way to efficiently deposit thin-films with ultra-high conformity. It is chosen largely due to its superior ability to deliver ultra-conformal dielectric thin-films with high aspect-ratio surface structures, which are encountered more and more often in the novel design of metal-oxide-semiconductor field-effect transistors (MOSFETs) in the NAND (Not-And)-type flash memory devices. Based on the traditional thermal ALD method, the plasma enhanced atomic layer deposition (PEALD) allows for lower operating temperature and speeds up the deposition process with the involvement of plasma species. Despite the popularity of these two methods, the development of their operation policies remains a complicated and expensive task, which motivates the construction of an accurate and comprehensive simulation model.

A series of studies have been carried out to elucidate the mechanisms and the concept of the PEALD process. In particular, process characterization focuses on the development of a first-principles-based three-dimensional, multiscale computational fluid dynamics (CFD) model, together with reactor geometry optimizations, of SiO2 thinfilm thermal atomic layer

deposition (ALD) using bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. Also, a comprehensive multiscale computational fluid dynamics (CFD) model incorporating the plasma generation chamber is used in the deposition of HfO2 thin-films utilizing tetrakis(dimethylamido) hafnium (TDMAHf) and O2 plasma as precursors.

Despite the great deal of research effort, ALD and PEALD processes have not been fully characterized from the view point of process control. This study aims to use previously developed multiscale CFD simulation model to design and evaluate an optimized control scheme to deal with industrially-relevant disturbances. Specifically, an integrated control scheme using a proportional-integral (PI) controller and a run-to-run (R2R) controller is proposed and evaluated to ensure the deposition of high-quality conformal thin-films. The ALD and PEALD processes under typical disturbances are simulated using the multiscale CFD model, and the integrated controllers are applied in the process domain. Using the controller parameters determined from the open-loop results, the developed integrated PI-R2R controller successfully mitigates the disturbances in the reactor with the combined effort of both controllers.

The dissertation of Yichi Zhang is approved.

Dante A Simonetti

Tsu-Chin Tsao

Carlos Gilberto Morales Guio

Panagiotis D. Christofides, Committee Chair

University of California, Los Angeles

2021

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Chapter 2 contains versions of: Zhang, Y., Y. Ding and P. D. Christofides, "Multiscale Computational Fluid Dynamics Modeling of Thermal Atomic Layer Deposition with Application to Chamber Design," *Chem. Eng. Res. & Des.*, **147**, 529-544, 2019.

Chapter 3 contains versions of: Zhang, Y., Ding, Y. and P. D. Christofides, "Integrating Feedback Control and Run-to-Run Control in Multi-Wafer Thermal Atomic Layer Deposition of Thin Films," *Processes*, **8 (1)**, 18, 18 pages, 2020.

Chapter 4 contains versions of: Zhang, Y., Y. Ding, Y. and P. D. Christofides, "Multiscale

Three-Dimensional CFD Modeling and Reactor Design for Plasma Enhanced Atomic Layer Deposition," *Comp. & Chem. Eng.*, **142**, 107066, 2020.

Chapter 5 contains versions of: Yun, S., Y. Ding, Y. Zhang and P. D. Christofides, "Integration of Feedback Control and Run-to-Run Control for Plasma Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films," *Comp. & Chem. Eng.*, **148**, 107267, 2021.

Curriculum Vitae

Education

University of California, Los Angeles	**Sep. 2013 - July 2017**
B.S., Chemical & Biomolecular Engineering	**Los Angeles, USA**

University of California, Los Angeles	**Sep. 2017 - July 2019**
M.S., Chemical & Biomolecular Engineering	**Los Angeles, USA**

Journal Publications

1. Yun, S., Y. Ding, **Y. Zhang** and P. D. Christofides, "Integration of Feedback Control and Run-to-Run Control for Plasma Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films," *Comp. & Chem. Eng.*, **148**, 107267, 2021.

2. **Zhang, Y.**, Y. Ding, Y. and P. D. Christofides, "Multiscale Three-Dimensional CFD Modeling and Reactor Design for Plasma Enhanced Atomic Layer Deposition," *Comp. & Chem. Eng.*, **142**, 107066, 2020.

3. Ren, Y., **Y. Zhang**, Y. Ding, Y. Wang and P. D. Christofides, "Computational Fluid Dynamics-Based In-Situ Sensor Analytics of Direct Metal Laser Solidification Process Using Machine Learning," *Comp. & Chem. Eng.*, **143**, 107069, 2020.

4. Wang, Y., **Y. Zhang**, Z. Wu, H. Li and P. D. Christofides, "Operational Trend Prediction and Classification for Chemical Processes: A Novel Convolutional Neural Network Method Based on Symbolic Hierarchical Clustering," Chem. Eng. Sci., **225**, 115796, 2020.

5. **Zhang, Y.**, Ding, Y. and P. D. Christofides, "Integrating Feedback Control and Run-to-Run Control in Multi-Wafer Thermal Atomic Layer Deposition of Thin Films," *Processes*, **8 (1)**, 18, 18 pages, 2020.

6. Ding, Y., **Y. Zhang**, M. Ren, G. Orkoulas and P. D. Christofides, "Machine Learning-Based Modeling and Operation for ALD of SiO2 Thin Films Using Data from a Multiscale CFD Simulation," *Chem. Eng. Res. & Des.*, **151**, 131-145, 2019.

7. **Zhang, Y.**, Y. Ding and P. D. Christofides, "Multiscale Computational Fluid Dynamics Modeling of Thermal Atomic Layer Deposition with Application to Chamber Design," *Chem. Eng. Res. & Des.*, **147**, 529-544, 2019.

8. Ding, Y., **Y. Zhang**, K. Kim, A. Tran, Z. Wu and P. D. Christofides, "Microscopic Modeling and Optimal Operation of Thermal Atomic Layer Deposition," *Chem. Eng. Res. & Des.*, **145**, 159-172, 2019.

Chapter 1

Introduction

1.1 Motivation

Thin-film deposition is one of the most important building blocks in the semiconductor industry. Various deposition techniques, such as epitaxy, chemical vapor deposition (CVD), and physical vapor deposition (PVD), have been developed to deposit high quality thin-films of various materials, e.g., Al_2O_3, Hf_2O_3, RuO_2, SiO_2, etc. [1]. However, the requirements in the production of advanced memory devices have become more and more demanding. For example, the dimensions of new high-k gate dielectrics are under transition to sub-10-nm scale and the associated film thickness is required to be under 30 Å [2, 3]. Also, new transistor designs often involve complex three-dimensional structures rather than two-dimensional planar surfaces, along with the demand of conformal films with a stringent criterion on uniformity and defects. Thus, the atomic layer deposition (ALD) process has been widely adopted by industry to meet the requirements of major design breakthroughs [4]. ALD is a thin-film deposition method originally derived from CVD. In an ALD process, a substrate surface is exposed to alternating gas-phase precursor streams such that only one type of reactant is in contact with the substrate surface at each half-cycle. Once in contact, the precursor undergoes self-limiting surface reactions that allow a nearly complete and conformal surface coverage given sufficient exposure time and appropriate

reactor conditions. In between the alternating precursor cycles, the reactor is purged with an inert gas, ensuring all previously-entered precursors are removed from the chamber prior to the exposure of the film to the next precursor, avoiding undesirable reactions and a decrease in film purity [5]. The ALD method enables a layer-by-layer film growth with film uniformity at atomic level, which is more precise and controllable than the traditional CVD approach [6–9]. Therefore, in the field of microelectronics 3D integration, where ultra-thin and highly-conformal films are needed, ALD has gained significant popularity. Nevertheless, the manufacturing of those high-κ thin-films using thermal ALD has faced challenges of low-throughput and high operating cost [10]. The primary reason for these problems is the extremely slow reaction kinetics at low temperatures because of the high activation energy barrier of the traditional precursors, thus often requiring high temperature for energy input. To overcome this problem, plasma is introduced as an enhancement to the originally neutral precursors, in which free radicals and ions are created so that the deposition reactions can go through lower energy barrier pathways [11]. Plasma enhanced atomic layer deposition (PEALD) is designed as a variation of ALD, which utilizes high-energy plasma species to replace the extremely high operating temperature.

Currently, there is a wealth of ALD and PEALD research on both laboratory and industrial scales [4]. This significant research activity on deposition has led to the discovery of novel precursors and mechanisms which make high throughput film processing possible while allowing various substrate lay-outs. However, experimental and industrial works on ALD remain expensive and time-consuming due to the cost of precursors and ALD-specific equipment, as well as due to the limited throughput [8]. Additionally, the real-time in-situ monitoring of film growth is not possible because molecular structure can only be understood through methods like scanning electron microscopy (SEM) and scanning tunneling microscope (STM), which are accurate but destructive to the deposited film [12]. Thus, a model for (PE)ALD that provides insights on the details of real-time film profile and the overall growth rate can be beneficial to both industrial and research work.

1.2 Background

PEALD is a close variation of ALD. In a PEALD process, alternative pulses of neutral precursor and plasma-enhanced precursor are introduced to the reactor, with inert gas purge pulses in between. Each precursor pulse is defined as a half-cycle, in which the associated precursor flows through the reactor, arrives at the substrate surface, and deposits the corresponding elements. Given appropriate operating conditions and sufficient precursor dosage, all the available reactive sites are occupied and reactions are completed in a self-limiting fashion. In the subsequent purge cycle, an inert gas is injected into the reactor, which removes the by-product and the left-over precursor gases to avoid undesirable reactions and multi-layer deposition [5]. Following this approach, uniform high-κ dielectric thin-films can be rapidly deposited layer-by-layer in a well-controlled manner, and high operating temperatures and high energy consumption can be avoided [9]. In the last twenty years, there has been a large number of studies on new types of high-κ dielectric thin-film materials, novel plasma precursors, associated reaction mechanisms, and the design of plasma generators and novel reactor chambers. New oxide materials like TiO_2 [13], HfO_2 [14], and ZrO_2 [15] are looked into because of their extraordinary gate capacitance, high thermal stability, good charge mobility, and the band alignment to the bottom silicon gates. Taking HfO_2 thin-film as an example, which has a dielectric constant 4-6 times higher than SiO_2, a variety of precursor combinations have been investigated. For the Hf half-cycle, Hf based metal-organic precursors are extensively studied, including tetrakis(dimethylamido) hafnium (TDMAHf) [16], tetrakis(ethylmethylamido) hafnium (TEMAHf) [13], tetrakis(diethylamido) hafnium (TDEAHf) [17], etc. The H–N hydrogen bonds associated with these precursors lead to easier physisorption and dissociative chemisorption pathways and faster reaction rates. On the other hand, common plasma-enhanced oxide precursors include O_3 plasma [18], H_2O plasma [19], and O_2 [18] plasma. In addition, the reactor geometry is also critical to PEALD throughput and effectiveness. The nominal direct plasma reactor geometry is inherited from the thermal ALD chamber, where plasma is directly generated in the main reactor chamber right above the substrate. This simple reactor geometry allows for a small volume and fast flow-type operation but may result in damage of the

3

substrate due to the high ionic energies in the plasma sheath [20]. To deal with this shortcoming, remote plasma and radical enhanced reactor configurations have been designed, in which plasma reaction happens at an appropriate remote distance, and ions are minimized while enough radicals are available to be transported to the substrate surface [21]. Thus, in this paper, we choose to investigate the deposition of HfO_2 thin-film using TDMAHf and oxygen plasma in a remote plasma reactor.

Various deposition models have been developed over the past two decades in different fields of microscopic simulation, for example, crystal growth, CVD, and plasma-ehanced CVD (PECVD) [22–24]. In particular, Molecular Dynamics (MD), and more recently, kinetic Monte-Carlo (kMC) are among the most popular simulation methods [24–26]. An ab initio MD model keeps track of all the particle movements and requires an overwhelming amount of computational resources, making it impossible to perform a simulation on an industrial scale process [27]. However, the kMC method has a crucial advantage in computational efficiency as it tracks a single event at a time in a predefined lattice space. Despite this simplification, the kMC method has been used in deposition models to successfully reproduce realistic profiles [24, 28, 29]. Recently, [24] proposed a novel multiscale computational fluid dynamics (CFD) simulation that used a surface microscopic n-fold hybrid kMC model and demonstrated its validity with a PECVD system. Moreover, many groups have showed the validity of using kMC in ALD simulation. For instance, [30] used raw probabilities of reaction and recombination to construct a kMC model for general plasma-enhanced ALD. [8] modeled a small scale Al_2O_3 ALD deposition using kMC based on first-principles analysis. For the plasma reaction scale, [31] looked into a small set of representative reactions of O_2/Ar plasma. [32] simulated a combination of the remote inductively coupled plasma (ICP) oxygen plasma source and the species transport in a simplified reactor geometry. [33] conducted a comparison between experiment and simulation of an O_2 ICP. For the gas-phase transport scale, commercial computational fluid dynamics (CFD) software have demonstrated their superior capabilities of accurately solving the gas-phase transport phenomena even for complex geometries [34, 35]. [36] simulated a cross-flow reactor for ALD of Al_2O_3. [24] demonstrated a successful

4

multiscale modeling of the plasma-enhanced chamber vapor deposition (PECVD) of amorphous silicon thin-films. [37] investigated a complex showerhead reactor for the thermal ALD of SiO_2.

However, previous works have not investigated the simulation of thermal ALD and PEALD of oxide film, which is an important material for gate oxides in MOSFET and MEMS devices, sacrificial layers, and conformal dielectric films in the front-end-of-line (FEOL) semiconductor wafer processing [38]. These ultra thin-films contain only around ten atomic layers (around 20 Å) and are only viable to prepare with ALD method [3]. Current research has yet to propose a model that is scalable to the industrial size, which have proven to be useful in traditional thin-film deposition systems [24]. Moreover, computationally efficient data-driven approaches, such as system identification, have not yet been applied to characterize the proposed ALD system [39–45]. Such data-driven models may make on-line optimization, run to run (R2R) control and real-time control possible for ALD processes [46–53]. The stringent quality requirement of PEALD demands the operating conditions to be consistent throughout the entire deposition process. Therefore, disturbances to the reactor operations must be effectively evaluated in various deposition processes to ensure the final film quality. Traditionally, feedback control systems have been used to eliminate the effect of disturbances in real-time. The performance of the feedback control has been extensively proved in various processes, including the application of plasma enhanced chemical vapor deposition (PECVD) [54], which is very similar to PEALD. In the feedback control system the measurement of the control variable is constantly fed back to the controller. The controller then compares the deviation of the measured value with the set point value and sets it as the process error. To offset the error, the controller changes the manipulated variable through an actuator, such as pressure regulating valves or flow control valves.

1.3 Dissertation Objectives and Structure

This dissertation considers industrially relevant simulation and control issues related to the atomic layer deposition and plasma enhanced atomic layer deposition. Specifically, the objectives of this
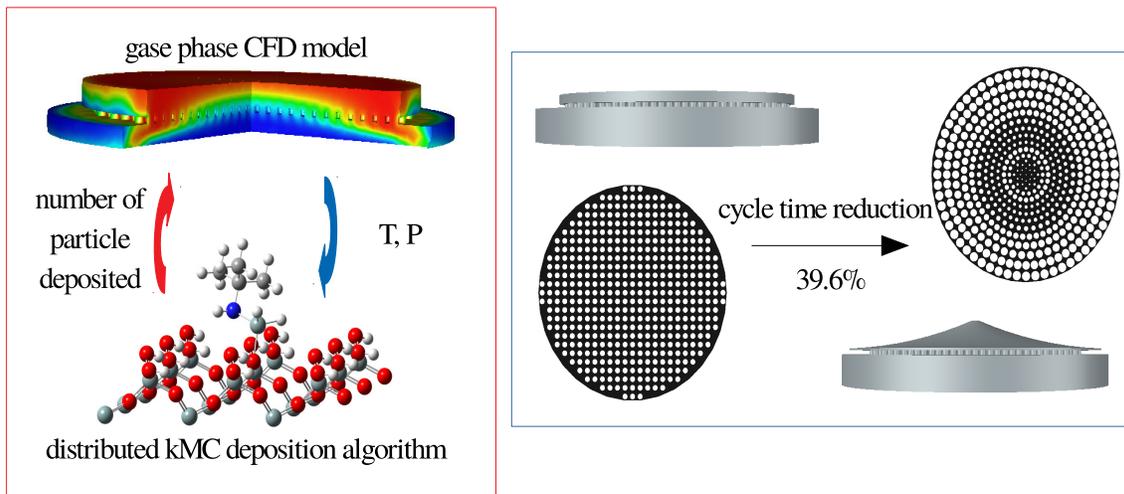
dissertation are summarized as follows:

1. To develop multiscale models for ALD and PEALD systems which capture both the gas-phase transport and reaction phenomena and microscopic thin film growth

2. To adapt run-to-run control strategies to the operation of ALD and PEALD using exponentially-weighted moving average (EWMA) algorithms and proportional integral (PI) controller

3. To address the common issue of non-uniform and slow deposition of thin films

4. To develop computational fluid dynamic (CFD) models which capture the complex gasphase phenomena within ALD and PEALD systems

5. To develop parallel programming techniques which allow for efficient computation of transient batch simulations

The dissertation is organized as follows.

In Chapter 2, a first-principles-based three-dimensional, multiscale computational fluid dynamics (CFD) model, together with reactor geometry optimizations, of SiO_2 thin-film thermal atomic layer deposition (ALD) using bis(tertiary-butylamino)silane (BTBAS) and ozone as

6

precursors. Specifically, an accurate macroscopic CFD model of the ALD reactor chamber gas-phase development is integrated with a detailed microscopic kinetic Monte Carlo (kMC) model that was developed in [55], accounting for the microscopic lattice structure, atomic interactions and detailed surface chemical reactions. The multiscale information exchange, the transient simulation of the microscopic distributed kMC algorithms and the macroscopic CFD model are achieved through a parallel processing message passing interface (MPI) structure. Additionally, density functional theory (DFT)-based calculations are adopted to compute the key thermodynamic and kinetic parameters for the microscopic thin-film growth process. Recognizing the transient non-uniformity and the possibility to reduce the current ALD cycle time, the optimal configuration of reactor geometry is designed and evaluated including a showerhead panel adjustment and geometry modifications on reactor inlet and upstream.

In Chapter 3, a multiscale CFD model and an integrated on-line control run-to-run scheme were constructed for a furnace ALD process of SiO_2 thin-film production with Bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors, as shown in Figure 2.1. 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.

In Chapter 4, a comprehensive multiscale computational fluid dynamics (CFD) model is developed for a remote PEALD reactor used in the deposition of HfO_2 thin-films. First, a

7

previously developed kinetic Monte-Carlo (kMC) model is adapted for the multiscale simulation to describe the surface reactions. Then, two macroscopic models, specifically tailored for the remote plasma reactor, are formulated to describe the dynamic behaviors of the plasma generation and bulk species transport domains, respectively. Additionally, an integrated message passing interface (MPI) scheme is built to couple and resolve the communication between different scales in the model. The model results are then validated with experimental data and an automated workflow is established for model calculations without human intervention. Finally, a set of baseline operating conditions derived from the model simulation results is proposed to guarantee the production of high-quality thin-films.

In Chapter 5, this study aims to use previously developed multiscale computational fluid dynamics (CFD) simulation model to design and evaluate an optimized control scheme to deal with industrially-relevant disturbances. Specifically, an integrated control scheme using a proportional-integral (PI) controller and a run-to-run (R2R) controller is proposed and evaluated to ensure the deposition of high-quality conformal thin-films. The PEALD process under typical disturbances is simulated using the multiscale CFD model, and the integrated controller is applied in the process domain. Using the controller parameters determined from the open-loop results, the developed integrated PI-R2R controller successfully mitigates the disturbances in the operating pressure and the gas bubbler temperature with the combined effort of both controllers.

Chapter 6 summarizes the main results of the dissertation.

Chapter 2

Multiscale Computational Fluid Dynamics Modeling of Thermal Atomic Layer Deposition with Application to Chamber Design

Thin-film deposition has always been an important upstream component of the semiconductor industry. To produce good-quality thin-films, many deposition techniques, such as physical vapor deposition (PVD), coating, and chemical vapor deposition (CVD) have been developed over the past few decades for materials like amorphous silicon, Al_2O_3, ZrO_2, HfO_2 and other [1]. Nevertheless, more precise and controllable deposition technology is needed as the requirements in semiconductor industry have become increasingly stringent. For example, the new memory designs like the NAND-type flash memory require conformal films with a more stringent criterion on uniformity and defects, along with complex non-planar three-dimensional structures. In addition, the dimensions of the new gate dielectrics are scaling down to sub-10-nm scale, with the associated film thickness under 20 Å [2, 3]. To meet the demands of such major design breakthroughs, atomic layer deposition (ALD) process has been developed and widely utilized

9

in the semiconductor industry [4]. ALD is a thin-film deposition method originally developed as a subclass variation of CVD. In ALD processes, the substrate surface is sequentially exposed to gas-phase precursor pulses instead of a single continuous flow of gas stream. In each of these pulses (half-cycle), ideally only precursor is reacting with the surface in a self-limiting way, which allows a nearly conformal surface coverage under sufficient cycle time, precursor dosage and appropriate operating conditions. Between the alternating half-cycles, an inert gas is used to purge the reactor, removing the previously-entered precursor from the chamber prior to the next precursor pulse, in order to avoid undesirable reactions and to ensure high film quality [5]. Thus, the ALD method is capable of uniformly growing materials layer-by-layer on complex substrate geometries in a more precise and more controllable manner [6–9]. As a result, ALD has demonstrated its great potential in the field of microelectronics manufacturing, where 3D ultra-thin and highly-conformal films are becoming increasingly crucial.

In recent years, there has been a growing number of ALD works on both laboratory and industrial scales for developing novel reaction mechanisms and precursors, which allow higher aspect ratio substrate lay-outs and faster film processing [4]. Nevertheless, industrial manufacturing and experimental studies on ALD still struggle with limited throughput and high cost of precursors and ALD equipment [8]. In addition, the exact influence of the gas-phase development in the reactor on the microscopic deposition processes remains unclear, and the real-time in-situ monitoring of film growth uniformity and local structure is not yet industrially available. Detailed molecular structure can be investigated using scanning electron microscope (SEM) and scanning tunneling microscope (STM), which are destructive to the deposited film [12]. Real-time overall thin-film growth rate may be examined with quartz crystal microbalance (QCM) but the measurement of local deposition rates is beyond QCM's capability [56]. In-situ spectroscopic ellipsometry is a feasible method for local behavior inspection but currently it is not prevalent in the industry because of its complex optical operations [57,58]. Thus, a comprehensive simulation model for ALD reactor chambers that helps understanding the details of the real-time reactor gas-phase profile together with the thin-film deposition profile is of value to both industrial

10

and research work.

In semiconductor manufacturing, SiO_2 is one of the most crucial materials for the sacrificial layers in micro-electro-mechanical systems (MEMS) and dielectric films for metal-oxide-semiconductor field-effect transistor (MOSFET) [38, 59]. The film thickness of those ultra-thin gate oxides has decreased to around 20 Å (10 atomic layers) and it is not possible to produce such thin films using traditional deposition methods [3]. Recently, [55] has investigated the detailed mechanism of the microscopic SiO_2 thin-film deposition using thermal ALD. However, that work has not incorporated the microscopic deposition profile with the gas-phase development in the thermal ALD reactor. In order to fully understand the entire SiO_2 ALD process, it is essential to adopt the multiscale approach developed by [24]. The resulting transient multiscale CFD model will provide a full-picture of the deposition process and enable a better understanding of the input-output relationship between operation conditions and cycle time. Having an accurate and comprehensive model also makes on-line optimization and real-time control possible for ALD processes [46–53].

In this section, we propose a density functional theory (DFT)-based multiscale CFD model for SiO_2 thin-film ALD with bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. Specifically, the multiscale CFD model combines a 3D macroscopic CFD domain and a 3D microscopic surface deposition kMC domain. The DFT-calculated thermodynamic and kinetic parameters such as the intermediate complex activation energies and the pre-exponential coefficients crucially affect the event selection of the microscopic kMC model to reproduce realistic thin-film growth rates and structure. Additionally, entropy, enthalpy, and viscosity are critical to the gas-phase CFD transport phenomena. With regard to the microscopic domain (i.e., the surface of the SiO_2 wafer), a detailed 3D kMC algorithm developed previously by [55] is adopted to capture the structure of the SiO_2 thin-film, the reaction mechanism, and the growth profile. For the macroscopic domain (i.e., the gas-phase), a transient 3D CFD model is adopted to account for the coupled mass and energy transport in the industrial scale ALD reactor under experimental and industrial standard operating conditions. The macroscopic and microscopic

11

domains are linked and solved simultaneously via a parallel processing message passing interface (MPI) structure. A nominal reactor design under the proposed model successfully reproduces the experimental and industrial growth per cycle (GPC) and cycle time, and the result reveals a non-uniform transient development of the precursor concentration in the reactor. According to this finding, two chamber geometry designs are developed and evaluated, including an adaptive adjustment of the size and position of showerhead holes, and a modification of the upstream region [60]. The resulting optimal chamber design adjusts and integrates the two features and is demonstrated to reduce the half-cycle time needed by 39.6%. Furthermore, the integrated modeling framework developed in this section can serve as a general modeling guideline for ALD and can be extended to other thin-film materials using different precursors, reactors, and operating conditions.

2.1 ALD Process Description and Multiscale CFD Modeling

This section focuses on the development of the 3D multiscale CFD model that describes the thermal ALD deposition of SiO_2 thin-film, which is demonstrated in Figure 2.1. The precursor selection is first discussed based on the candidates' reported chemical and physical properties. Then, the CFD model for the gas-phase macroscopic domain is discussed in detail: a 3D geometry of a thermal ALD reactor is constructed and is discretized spatially to a high-quality mesh; the gas-phase transport mechanisms and the detailed tuning are elaborated. Next, the kMC algorithm for the microscopic deposition model is discussed in detail: the approximation of the 3D SiO_2 lattice is introduced and validated; the reaction mechanism and associated kinetics are formulated; the kMC algorithm for event selection and time evolution is explained. Subsequently, the parallel processing and the interaction between the two models through the MPI is demonstrated along with the automated workflow. Finally, DFT calculations using Gaussian09/Gaussview software package for the determination of the important thermodynamic parameters are explained [61].
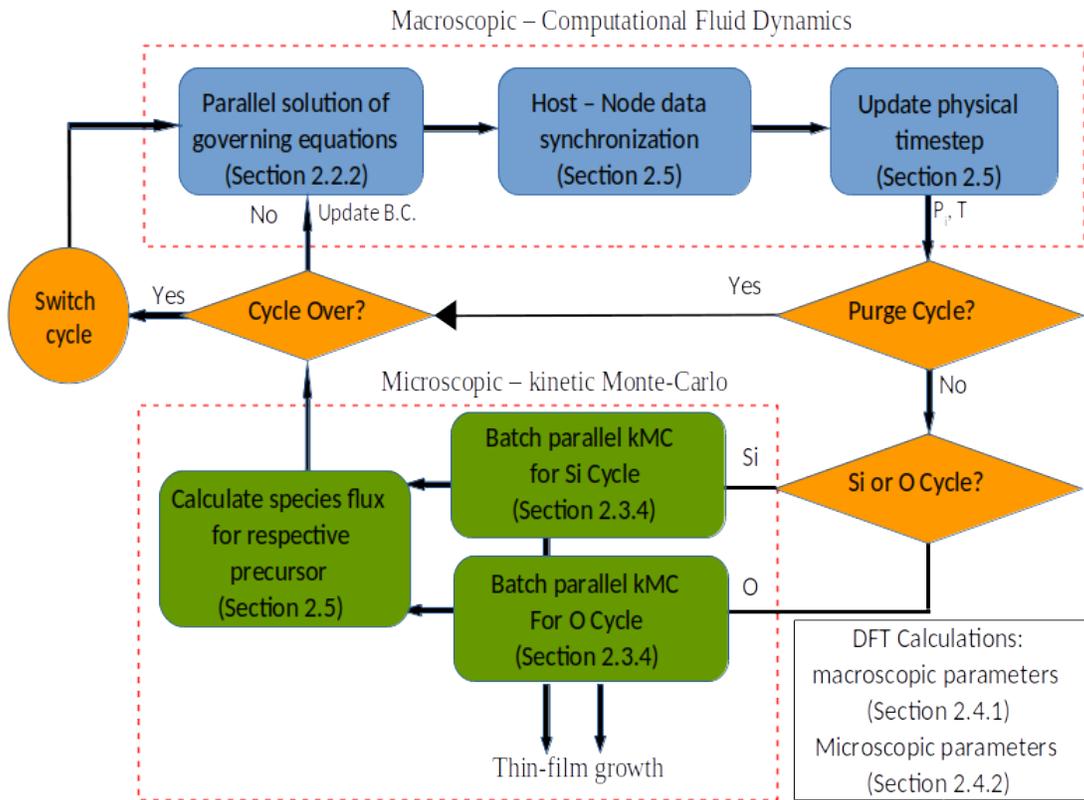
Figure 2.1: Multiscale workflow, parallel processing and information exchange illustration.

2.1.1 Precursor Selection

The choice of the Si precursors in the SiO_2 ALD process is one of the most important factors that determines the growth rate and the quality of SiO_2 thin-films. In recent studies, aminosilane-based precursors have gained significant popularity because the H-N hydrogen bonds formed between the precursor particles and the substrate surface significantly lower the reaction energy barrier and hence improve SiO_2 deposition [62]. As demonstrated in [55], we choose BTBAS as the Si half-cycle precursor because of its fast growth rate (1.4 Å ~ 2.1 Å per cycle), sufficient experimental and industrial data, and detailed theoretical reaction mechanisms [62, 63]. For the oxygen precursor, we choose ozone (O_3) among the common candidates for thermal ALD of oxide films, which is extensively studied in the laboratory and utilized in the industry [63–65].

13

2.1.2 Macroscopic CFD Model

Deposition of SiO_2 thin-films in an industrial ALD reactor is intimately related to the physical and chemical phenomena that govern the gas-phase. In order to obtain an accurate characterization of velocity, concentration and temperature profile, the coupled continuity, energy, and mass balance equations need to be solved in 3D domain defined by the proposed reactor geometry. Computational fluid dynamics is specifically tailored to tackle such tasks. In this work, we utilize ANSYS Fluent 19.2 software package, a commercial CFD simulation software, to construct the reactor geometry, perform meshing and solve for the gas-phase profile [66].

2.1.2.1 Reactor Design and Meshing

Various thermal ALD reactor geometries are used in industry. In our model, we adopt a cylindrical showerhead reactor chamber based on the design of EmerALD XP by ASM International [60]. Specifically, the reactor chamber is scaled to accommodate for processing the up-to-date industrial 300 mm wafer, with the exact dimension shown in Figure 2.2 (a) [67]. Precursor gas is introduced from the top of the reactor, labeled 1, and develops uniformly in the upstream region, labeled 2. Then, precursor gas flows through the showerhead holes, labeled 3, falls onto the substrate surface, labeled 4, and then exits through the bottom outlet that is connected to a vacuum pump, labeled 5. The nominal showerhead design contains 10 mm-diameter showerhead holes, uniformly spaced in a rectangular array across the showerhead region, while the optimized designs of showerhead and upstream regions will be discussed in Section 2.2. The 3D ALD reactor geometry is constructed in AUTOCAD, a 3D computer-aided design software. It is important to capture the 3D detail because the showerehad design is complicated and not azimuthally symmetric for a 2D axisymmetric characterization. Next, the resulting geometry is exported to the ANSYS Design Modeler for spatial discretization, which is necessary for the hybrid meshing mentioned below. The discretized components are then recombined to a single fluid body after sectioning in order to ensure zone conformity [66].

In CFD, the governing transport PDEs need to be solved in smaller subdomains made of

14

(a)

(b)

1

2

300 mm

30 mm

30 mm

3

4

3
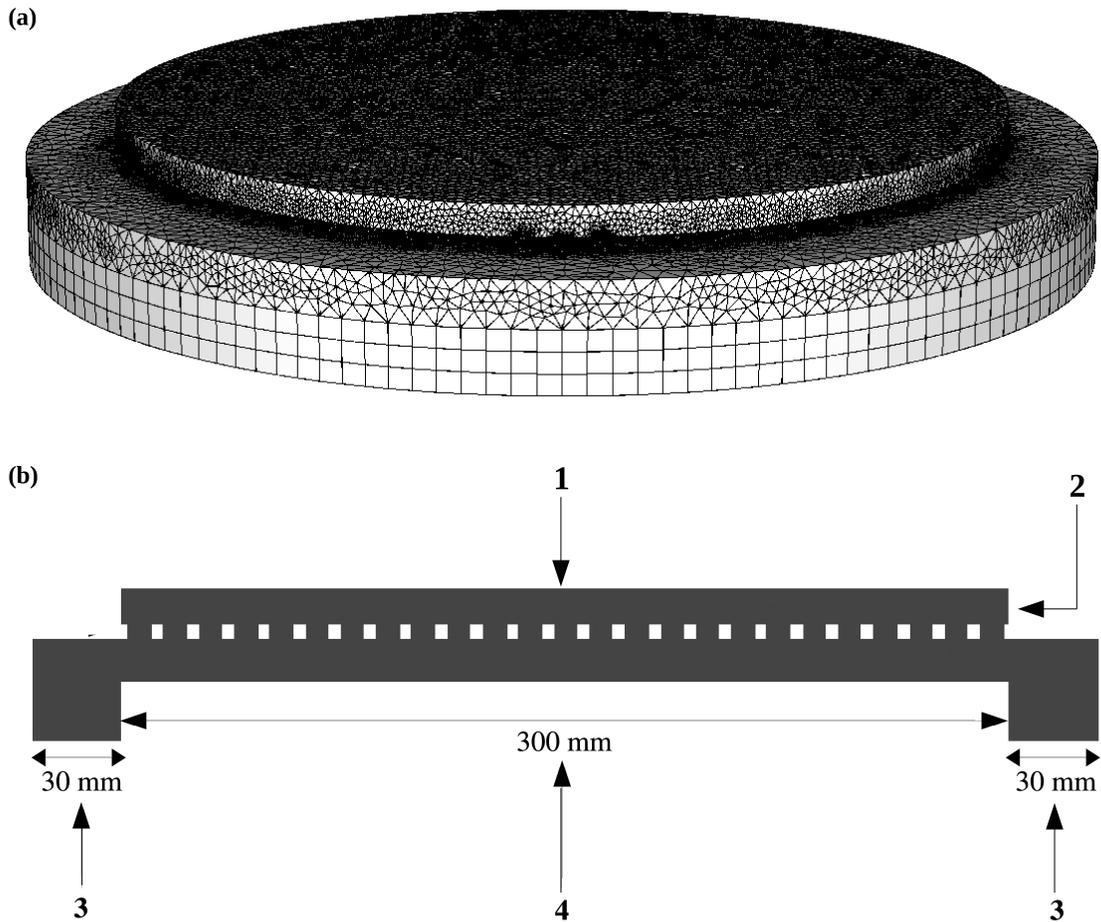
Figure 2.2: Mesh and *x*-cross-section geometry of the nominal reactor. The gas inlet is labeled
1. The upstream region is labeled 2, where the gas-profile gradually develops. The showerhead
holes is labeled 3. The substrate surface is labeled 4, and the exits through the bottom outlet that
is connected to a vacuum pump, labeled 5.

geometric primitives, split from the major geometry. The collection of those geometry primitives

is called a mesh. Because the governing equations will be discretized and solved inside each of

these subdomains using the finite volume method, the meshing resolution and quality are crucial

for the overall CFD accuracy [68]. Two meshing strategies are generally adopted: structured

mesh and unstructured mesh. While unstructured mesh is a more staight-forward approach and

is faster in implementation, it leads to longer simulation computation time due to the geometric

irregularity and the large number of cells in the resulting mesh. On the contrary, performing a

fully structured meshing on our reactor geometry may improve the computational efficiency but is

labor and time-consuming. Any adaptive structural optimization of the reactor design will require a complete manual re-mesh. Therefore, in this section, a hybrid mesh is chosen as a balance between the two approaches. Using ANSYS Meshing, we adopt a multi-zone method to discretize the single body. Structured mesh with hexahedral elements is used for the majority of reactor geometry that has simpler overall geometry. By contrast, unstructured meshing is performed on the showerhead and substrate surface regions, which have high aspect ratio and more predefined geometric irregularities. Patch independent tetrahedral meshing method with high relevance is adopted to discretize such areas with good precision to obtain good cell quality for computational accuracy and efficiency.

Mesh independence test is carried out, where the smallest mesh produced for an invariant simulation result is chosen so that the best computational efficiency can be achieved without influencing the simulation accuracy. Meshes containing 2 to 2.5 million cells are proved to produce mesh independent results depending on the specific reactor geometries. The resulting mesh is demonstrated to have a good quality, which is characterized by the overall quality, skewness and aspect ratio. Table 2.1 shows the acceptable range of each parameter with $*$ being the desired value, and the corresponding quality of our meshes from the nominal design to the final optimized design, respectively. The data shows that the min/max bounds of each mesh quality metric lie within the acceptable range and the average mesh metrics lie close to the ideal values. The three factors and their recommended value ranges are suggested by the manufacturer ANSYS Inc. [66, 69]. Other optimized designs have shown similar quality.

Table 2.1: Cell quality of the nominal and final design hybrid reactor mesh

	Acceptable Range	Nominal Bound; Average	Final Bound; Average
quality	0 - 1*	0.34; 0.79	0.33; 0.79
skewness	0* - 0.95	0.69; 0.29	0.74; 0.29
aspect ratio	1* - 100	5.41; 1.99	6.49; 1.98

2.1.2.2   Gas-Phase Transport Model

In the macroscopic domain, a transient model is applied in CFD using ANSYS Fluent pressure-based solver. The built-in transport models and several user defined functions (UDFs) are implemented to apply ALD process specifications on the materials, transport phenomena, as well as operating and boundary conditions.

Three gas species are used in the Fluent simulation. Argon is the purging gas in between Si-Cycle and O-cycle. BTBAS and ozone are the Si-cycle and O-cycle precursors, respectively, which are involved in the surface reactions that will be introduced in more detail in Section 2.1.3.2. The other reaction by-products are proved not important in the convention-dominated gas-phase profile, and thus, it is not included in the CFD. Argon, ozone are selected from the Fluent Database, while BTBAS is defined with DFT-calculated and literature values: the molecular weight equals to 174.36 g/mol; viscosity is approximated with a linear fit of 5.582×10^{-6} kg/m-s at 398K and 6.824×10^{-6} kg/m-s at 600 K [70]; the specific heat (C_p) is characterized as a piecewise-polynomial fit with coefficients A = -4.1113, B = 1.0243, C = -0.0005 under Fluent input format [66]; standard enthalpy is determined to be -267.84 kcal/mol, and the standard entropy is determined to be 126.79 cal/mol-K. Viscosity, enthalpy, entropy and C_p are calculated using DFT and thermodynamic principles, which will be discussed in Section 2.1.4. The mixture of gas species is defined using the Fluent Mixture Template [66]. The mixture is assumed to be an incompressible ideal gas mixture, appropriate for a laminar system with Mach number smaller than 0.3, which is valid for all inlet flow rates and operating pressures considered in this model [71]. All other mixture properties like C_p, thermal conductivity, viscosity, mass diffusivity, and density are under ideal-gas mixing law, appropriate in a low pressure and high temperature thermal ALD reactor.

Momentum, mass, and energy transports are important in simulating the gas-phase velocity, temperature and concentration profile and in predicting the boundary profile at the wafer surface, which is used for the microscopic simulation. As mentioned before, ANSYS Fluent solves the coupled Navier-Stoke equations to obtain an accurate profile for the reactor dynamics. The

following general flow field equations are given in the tensor form for Fluent calculations [66,72]:

$$\frac{\partial(\rho\vec{v})}{\partial t} + \nabla \cdot (\rho\vec{v}\vec{v}) = -\nabla P + \nabla \cdot (\bar{\bar{\tau}}) + \rho\vec{g} + \vec{F} \tag{2.1}$$

$$\bar{\bar{\tau}} = \mu\left[(\nabla\vec{v} + \nabla\vec{v}^T) - \frac{2}{3}\nabla \cdot \vec{v}I\right] \tag{2.2}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = \nabla(k\nabla T - \Sigma h\vec{J} + (\bar{\bar{\tau}}\vec{v})) + S_h \tag{2.3}$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla \cdot (\rho\vec{v}Y_i) = -\nabla \cdot \vec{J}_i + R_i + S_i \tag{2.4}$$

$$\vec{J}_i = -\rho D_{m,i}\nabla Y_i - D_{T,i}\frac{\nabla T}{T} \tag{2.5}$$

where ρ is the density of the gas-phase species mixture, $\vec{v}$ is the velocity of gas-phase species mixture, P is the static pressure, T is the fluid temperature, $\vec{g}$ is the component of gravity on the direction of the fluid flow, $\bar{\bar{\tau}}$ is the stress tensor, μ is the fluid viscosity, E is the fluid internal energy, I is the unit tensor, R_i and S_i are the reactions and mass transfer source terms of species i, respectively, S_h is heat transfer source term, $\vec{J}_i$ is the diffusive flux of species i, and $D_{m,i}$ and $D_{T,i}$ are the mass and heat diffusivities of species i.

The operating pressure is set to be 133 Pa, which lies in the typical pressure range (100~500 mTorr) of thermal ALD processes [5]. In the CFD model, the inlet of the reactor is defined as a mass-flow-inlet at 600 K and four UDFs are created to define the precursors half-cycle duration and fluxes, and the purging durations and fluxes according to experimental and industrial procedures [67,73]. The gas inlet temperature is specified to be 473.15 K according to the industrial standard [60]. The outlet is defined as a pressure-outlet, simulating a 12 M^3/h rotary vacuum pump, and a UDF is used to specify a 1×10^{-5} mbar vacuum for the purging [4]. The substrate plate temperature is maintained at 600 K with a simulated heater. The substrate surface, where the microscopic and macroscopic domain intersect, is defined as a mass-flow-outlet. Four UDFs are utilized to provide gas-phase boundary information for the microscopic model and to define the CFD boundary conditions, accounting for the consumption of precursors and the generation of side

products from microscopic kMC deposition. The term S_i appearing in Equation 4 accounts for the mass flow generation and consumptio also believe that it is important tn in the user defined sources or sinks, while the enthalpy exchange of the term S_h in Equation 3 is demonstrated to have minimal influence on the wafer temperature, and thus, it is not considered in our work. The information exchange and the interaction between the microscopic and macroscopic domains will be explained in more details in section 2.1.5.

Based on the mesh structure introduced in the previous section, the finite volume method is utilized by ANSYS Fluent to solve the governing equations implicitly [68]. Specifically, for each time step, ANSYS Fluent adopts the pressure-based coupled algorithm to solve simultaneously the momentum, continuity, energy, species, and other transport equations. Calculations are iterated until the residual convergence criterions are satisfied.

The calculation time step and the courant number selection are also critical to the CFD simulation accuracy [74]. Courant number is a quantitative description of the number of computational cells that the fluid moves through in one time step, where the time step is the unit discretization of the real physical time. The Courant-Friedrichs-Lewy (CFL) condition, a common criteria for finite-difference approximation of PDEs, relates the time step to the interval lengths in spatial coordinate and the speed of information traveling in the physical space [75]:

$$C = \frac{u\Delta t}{\Delta x} \tag{2.6}$$

where C is the selected courant number, Δx is the minimum cell length and u is the average flow velocity.

Traditional direct numerical PDE analysis allows a maximum courant number of 1, while the advanced CFD software allows convergence under higher courant number [75]. In particular, the Fluent pressure-based solver is less sensitive to courant number [66]. In our model, a courant number of 50 is used, which proved to be able to converge the simulation efficiently, successfully and accurately. Using the above equation, time steps can be determined for the BTBAS half-cycle,

19

purge, and ozone half-cycle depending on the different inlet flow rates, respectively. Time step sizes of 0.05 s and 0.01 s are used under 100 sccm and 1330 sccm BTBAS inlet flow rates, respectively. It was found that smaller time step sizes do not lead to significant change of the simulation accuracy, while bigger time steps lead to inaccurate results.

2.1.3 Microscopic Surface Model

Although the macroscopic CFD modeling determines the wafer surface boundary conditions (i.e., temperature and pressure of the surface chemistry), it does not provide any information about the behavior of the microscopic deposition regime. The microscopic simulation captures the structural detail of SiO_2 thin-film and the reaction mechanisms, which is modeled using 3D kMC algorithm. The 3D lattice assumption, detailed mechanism and kinetics for surface physical and chemical reactions, and the kMC formulation will be discussed in the following subsections.

2.1.3.1 Structural Assumptions of Deposited SiO_2

The α-Quartz SiO_2 thin-film has a local SiO_4 structure close to tetrahedron and a trigonal space group of $P3_121$. As demonstrated in [55], a 3D triangular model is constructed as an approximation of the actual α-Quartz crystal structure. This approximation, Figure 2.3 (a), is a close resemblance of the real lattice, Figure 2.3 (b), and preserves the validity of the thin-film model. A final lattice dimension is 1200×1200 sites per layer with height dependent on the number of cycles simulated. This lattice is large enough to be size-independent while computationally efficient [76, 77].

2.1.3.2 Reaction Mechanism

For each full deposition cycle in the ALD process, there are two half-cycles, each using a specific precursor species to deposit the desired element onto the film. As discussed in Section 2.1.1, BTBAS and ozone are chosen for the SiO_2 deposition simulation. The reaction mechanism using these two precursors was reported by [63] and is explained in detail below.

Figure 2.3: (a) Top view of a 5-layer 12×12-site miniature demonstration of the full kMC simulation lattice. The five layers and the species on the lattice are shown using different colors and symbols, respectively. The first (bottom) layer, labeled red, contains the base Si atoms. The second layer, labeled black, contains oxygen atoms or hydrogenated oxygens. The third layer, labeled yellow, contains the species from the first silicon half-cycle: Si is the neighbour-binding silicon, Si! is the self-binding silicon, and PsP and CsP are the physisorbed and chemisorbed precursors, respectively. The fourth layer, labeled green, contains the species from the first oxygen half-cycle: O and OH are the oxygen atoms and hydroxyl groups. The fifth (top) layer, labeled blue, contains physisorbed ozones (PO1 and PO2), which remain to be oxidized. (b) Top view. (c) Side view.
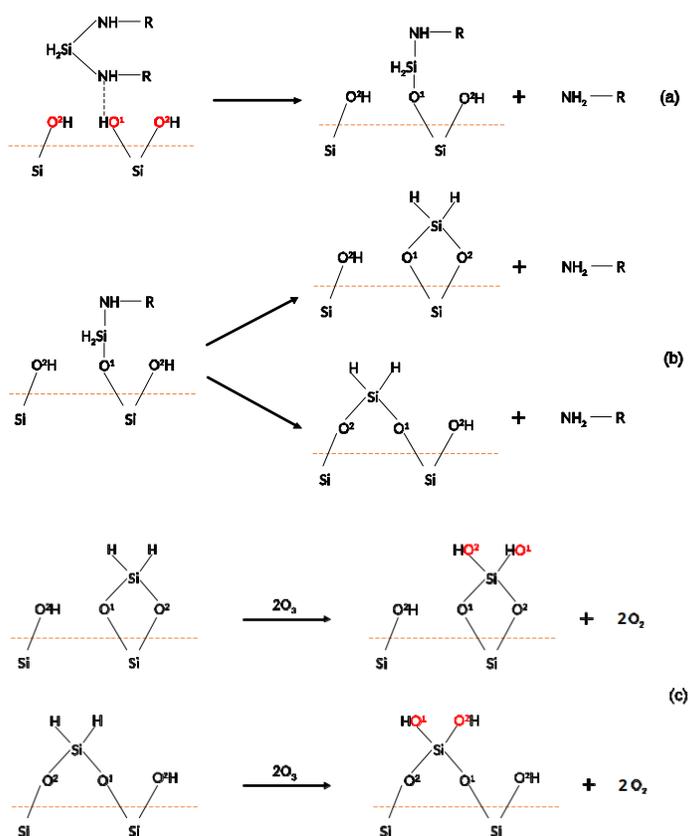
Figure 2.4: (a) First dissociative chemisorption step of BTBAS. (b) Second dissociative chemisorption step of BTBAS under self-binding and neighbour-binding mechanisms. (c) Oxidation of self-binding and neighbour-binding SiH_2 with ozone.

The first half-cycle is referred to as the Si-Cycle, which contains physisorption, abstraction and a two-step dissociative chemisorption. A hydroxylated $SiO_2(001)$ substrate surface is defined as the basis, shown in Figure 2.3 (c). According to [63], the two oxygen atoms in a SiO_2 cell have different electronegativities. The more electronegative oxygen atom, denoted as O^1, is more reactive and is therefore more likely to be electrophilicly attacked by precursor particles than the less electronegative oxygen atom, denoted as O^2. Therefore, as shown in Figure 2.4 (a), the precursor particle is first physisorbed onto the O^1-type hydroxyl group through a strong H-bond to form the reactant. Then, the physisorbed precursor goes through the first dissociative chemisorption step, forming a monoamine intermediate and releasing one of the two aminoethyl

groups. Next, the remaining aminoethyl group electrophilicly attacks an adjacent O^2-type hydroxyl group, which can be either from the neighbour Si atom, i.e., neighbour-binding route, or from the same substrate Si atom, i.e., self-binding route, as shown in Figure 2.4 (b). The former reaction pathway retains the original surface orientation, resulting in a thermodynamically favorable structure, whereas the latter, which is more kinetically favorable, causes a deviation from the (001) surface orientation and leads to defect formation. After the electrophilic attack, the other aminoethyl group is released from the surface structure and another O-Si bond is formed. The remaining two H atoms from the Si atom then become the new substrate surface. The competition of kinetic and thermodynamic favorability is crucial in explaining the structural non-uniformity of SiO_2. Therefore, both reaction pathways and their reverse reactions are incorporated in our kMC model, and the reaction kinetics will be explained in more details in the next section.

The second half-cycle is referred to as the O-Cycle, which contains the ozone physisorption, abstraction and surface oxidation. The oxidation steps of self-binding and neighbour-binding H-Si groups are shown in Figure 2.4 (c). Once the surface is partially/fully chemisorbed, both terminating H atoms will be oxidized by ozone to hydroxyl group (-OH), which are utilized in the next Si-cycle.

2.1.3.3   Relative Rate Determination

The reaction kinetic rates, discussed in the previous section, are improtant for kMC event selection. The physisorption is a gas-surface athermal barrierless process, and the Collision Theory, as expressed in the equation below, is generally used to determine the rate constant for such processes:

$$r_{phs} = \frac{p_i}{RT} \sqrt{\frac{8k_bT}{\pi m_i}} s_{c,i} N_a \sigma \tag{2.7}$$

where r_{phs} is the physisorption reaction rate, p_i is the partial pressure of the species i, R is the ideal gas constant, T is the temperature, k_b is the Boltzmann constant, m_i is the molecular weight of species i, $s_{c,i}$ is the sticking coefficient of the species i at given temperature, N_a is the Avogadro

number, and σ is the average area per surface site. Although the sticking coefficient of BTBAS is not reported in previous works, we obtain its value through an analogy with the sticking coefficient of bis(diethylamino)silane (BDEAS) because of the structural and electronic similarity [78].

On the contrary, chemisorption, abstraction and oxidation are thermodynamically activated kinetic reactions, which are generally described by the transition state theory (TST) [79]. Assuming quasi-equilibrium is achieved between the complex and the reactant, the reaction rate can be estimated using the thermodynamic properties of the transition state complexes, which are computed using DFT. Thus, the reaction rate equation can be formulated with standard Arrhenius-type rate law as follows:

$$r_{rxn,i} = A_i \exp\left(\frac{-E_{a,i}}{k_b T}\right) \tag{2.8}$$

where $r_{rxn,i}$ is the reaction rate of the ith thermodynamically activated reaction, $E_{a,i}$ is its activation energy for the transition state complex, and A_i is its pre-exponential factor, which is determined as follows:

$$A_i = f_i^{\text{TST}}\left(\frac{k_b T}{h}\right) \tag{2.9}$$

where k_b is the Boltzmann constant, T is the temperature, and f_i^{TST} is the ratio of the vibrational partition function between the transition state complex and the reactant, calculated with DFT. In this work, all DFT calculations are performed using the Gaussian09 software, which will be illustrated in more details below. The resulting parameters are collected in [55] and the associated nomenclature is explained in Section 2.1.3.4.

2.1.3.4   Kinetic Monte Carlo Algorithm

As mentioned in the Introduction, a first-principles MD simulation is too computationally demanding to be feasible for the scale of system discussed in this work [27–29]. Thus, we adopt an n-fold hybrid kMC algorithm in the framework proposed by earlier works [24, 80, 81]. kMC is a stochastic algorithm that uses the kinetic rate information and uniformly distributed random

numbers to determine event execution and system time evolution. Specifically, we define an event set as a collection of all events that have comparable rates. A total rate, r_{total}, is defined as:

$$r_{total} = \sum_{i=1}^{N} r_i \tag{2.10}$$

where r_i represents the respective rate of each event within an event set, which consists of total N events. Then, each rate is normalized with respect to the associated total rate to derive its relative probability. The normalized indicator of the ith event, $l_i \in (0,1]$, can be interpreted as the sum of the normalized probabilities of the first i events:

$$l_i = \frac{\sum_{j=1}^{i} r_j}{r_{total}}, i = 1,...,N \tag{2.11}$$

This indicator is then used for event selection via a uniformly distributed random number selection, $\gamma_1 \in (0,1]$. If γ_1 falls in the interval of normalized indicators l_{i-1} to l_i, the ith event will be selected for execution.

The transient behavior of the model is characterized by the time evolution scheme proposed by the kMC algorithm, where the amount of time for each event is governed by using another random number, $\gamma_2 \in (0,1]$. Starting from a given time, the simulation time clock is advanced by Δt for the chosen event, where Δt is given by the following equation:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{2.12}$$

Therefore, the total rate for O-Cycle is computed as follows:

$$r_{o,total} = r_{o,phs} + r_{o,des} + r_{o_a,f} + r_{o_b,f} \tag{2.13}$$

where $r_{o,phs}$ is the rate of ozone physisorption, $r_{o,des}$ is the rate of ozone desorption, and $r_{o_a,f}$ together with $r_{o_b,f}$ are the oxidation rates of the chemisorbed species attached to a

neighbour-binding silicon. The oxidation rate of the chemisorbed species attached to a self-binding silicon is orders of magnitude higher than that of a neighbour-binding silicon. Therefore, it is considered instantaneous and deterministic, and thus, omitted in the O-Cycle kMC selection. Similarly, the total rate for Si-Cycle is:

$$r_{si,total} = r_{btbas,phs} + r_{btbas,des} + r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} \qquad (2.14)$$

where $r_{btbas,phs}$ and $r_{btbas,des}$ are the physisorption and desorption rates of BTBAS, respectively. $r_{neigh,f}$ and $r_{neigh,r}$ are the forward and reverse rates of the neighbour-binding dissociative chemisorption, respectively, and $r_{self,f}$ and $r_{self,r}$ are the forward and reverse rates of the self-binding dissociative chemisorption, respectively. The reaction rate of first chemisorption step $r_{si,chem}$ is orders of magnitude higher than those of other events. Therefore, it is considered instantaneous and deterministic, and thus, omitted in the Si-Cycle kMC selection.

For the O-Cycle, the rates of all considered reactions are comparable and can be modeled with the standard n-fold kMC algorithm. However, for the Si-Cycle, in order to simulate the realistic behavior of reaction kinetics, we need to consider surface reaction events separately from physisorption events for the following two reasons: First, surface reaction events are formulated and compared differently from physisorption events since surface species concentrations need to be considered to correctly describe the competition between the thermodynamic and kinetic favorability of competing pathways. Second, physisorption rates are an order of magnitude lower than surface reaction rates according to the DFT calculation, which means that the model will be saturated by surface reactions events if the events are not allocated properly. Thus, a decoupled kMC scheme is proposed to partition the entire Si-Cycle events into two event sets: adsorption events containing only physisorption events, and surface reaction events containing the remaining events. The partitioned total rates, $r_{si,ads}$ and $r_{si,rxn}$, are then defined as follows:

$$r_{si,rxn} = r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} + r_{btbas,des} \qquad (2.15)$$

$$r_{si,ads} = r_{btbas,phs} \tag{2.16}$$

Additionally, in order to apply the decoupling scheme, we first compute a ratio, $J_{si,ads}$ as the ratio of the adsorption rate versus the total rate, which is derived as follows:

$$J_{si,ads} = \frac{r_{si,ads}}{r_{si,total}} = 1 - J_{si,rxn} \tag{2.17}$$

Therefore, for a total assigned duration, t_{total}, adsorption events are pre-allocated with a duration of $t_{total} \cdot J_{si,ads}$, and the remaining time is assigned to surface reaction events. Next, during the allocated time period for surface reactions, the normalized event indicator under the competition of reaction pathways and directions is calculated by the concentration-weighted reaction rates as follows:

$$l_{si,i} = \frac{\sum_{j=1}^{i} r_{rxn,j} R_j}{\sum_{k=1}^{N} r_{rxn,k} R_k}, \quad i = 1, ..., N \tag{2.18}$$

where $l_{si,i} \in (0, 1]$ represents the normalized indicator of the ith event in the surface reaction event set, $r_{rxn,j}$ is the un-weighted chemical reaction rate for the jth event calculated from Equation 2.8, R_j is the number of reactants for each surface reaction, and N is the total number of events in the Si-Cycle surface reaction event set. The normalized indicators are then used to execute the event selection following the same approach performed in the standard kMC algorithm. In Section 2.2, it is demonstrated that this decoupling scheme achieves desired accuracy.

2.1.4 DFT and Thermodynamic Calculations

The reaction mechanism along with the associated activation energies are reported for BTBAS and ozone for SiO_2 [63]. However, many fundamental thermodynamic properties of BTBAS are yet to be investigated. Since it is important for us to capture those thermodynamic properties to reproduce the correct behavior in both the macroscopic CFD model and the microscopic kMC model, DFT calculations are carried out using Gaussian09 software packages and parallel computational work with Linda worker is adopted [61, 82–84].

2.1.4.1 Density Functional Theory

The solution of Schrödinger equation of many-body systems, as shown below, provides a first-principles understanding of the material properties [85]:

$$\widehat{H}\Psi(r_i, R_I) = E\Psi(r_i, R_I) \tag{2.19}$$

where Ψ is the particle wave function, r_i is the position of each electron, R_i is the position of each nuclei, E is the ground state energy, and $\widehat{H}$ is the Hamiltonian operator:

$$\widehat{H} = \widehat{T}_{nuc}(R) + \widehat{V}_{coulomb} + \widehat{H}_{elec}(r, R) \tag{2.20}$$

where $\widehat{T}_{nuc}(R)$ is the nucleic kinetic energy, $\widehat{V}_{coulomb}$ is the coulombic potential between nuclei and electrons, and $\widehat{H}_{elec}(r, R)$ is the electronic Hamiltonian.

However, the exact solution to the Schrödinger equation of a many-body system is extremely difficult to obtain and its complexity scales rapidly with system size. Therefore, many efforts have been put to approximate the solution of the Schrödinger equation. DFT is one of those approaches, which, instead of focusing on each electron, solves for the overall electron density, $n(\mathbf{r})$ [85]. According to the Hohenberg and Kohn theorem, the ground state energy of the system is an unique functional of electron density [86]. Therefore, by assuming the energy functional and minimizing the overall energy, the true electron density of the system can be obtained. The total energy functional consists of the non-interacting kinetic energy, electron-nuclei potential, the approximation of the full interaction (also known as the hartree energy), and the exchanges and correlations of the system [87]. The exact form of the exchange-correlation functional is not known and hence must be approximated. Over the past few decades, a lot of efforts have been made to propose highly-accurate energy functionals, with one of the most popular being the B3LYP (Becke, 3-parameters, Lee-Yang-Parr) functional [88, 89]. After having a descriptive energy functional, the overall system energy is broken down into the energies of pseudo spin-orbitals, known as

the Kohn-Sham orbitals. The energy of each spin-orbitals is solved iteratively by reconciling the differences between the guessed electron density with the computed electron density from the proposed energy functional [90]. The final computation results for each Kohn-Sham orbitals are combined together to be an accurate electron density, which can then be used to compute a variety of material properties.

2.1.4.2 Macroscopic Parameters Determination

As mentioned in Section 2.1.2.2, BTBAS is not in the Fluent material database, and many physical and chemical properties that are crucial to thermodynamic calculations are not reported in previous works. Thus, its thermodynamic/physical parameters, including specific heat C_p, viscosity μ, standard state enthalpy of formation $\Delta_f H^o(M, 298K)$, and standard state entropy $S^o(298K)$, need to be computed for the gas-phase transport model. To determine those parameters, DFT calculations using the Gaussian-4 (G4) theory are performed. Based on ab-initio molecular-orbital theory, G4 is a precise computation method for calculating molecular energies of compounds composed of elements in the first three rows of the periodic table. Thus it is an appropriate and accurate method in determining the thermodynamic properties of BTBAS [91].

The method utilizes a sequence of single point energy calculations using appropriate basis set, spin-orbit correction and advanced higher level correction. The detailed calculation steps and energy corrections are discussed in literature [91]. Using G4 method, standard state entropy $S^o(298K)$ and specific heat C_p of BTBAS under various temperatures can be directly calculated. The standard state enthalpy of formation $\Delta_f H^o(M, 298K)$ for the BTBAS molecule can be calculated from:

$$
\begin{aligned}
\Delta_f H^o(M, 298K) =& \Delta_f H^o(M, 0K) + H_M^o(298K) - H_M^o(0K) \\
& - \sum_{atoms} x(H_X^o(298K) - H_X^o(0K))
\end{aligned}
\tag{2.21}
$$

where $H_M^o(298K)$ is the molecular enthalpy at standard state and $H_M^o(0K)$ is the molecular enthalpy at 0 K. The difference between $H_M^o(298K)$ and $H_M^o(0K)$ accounts for the temperature correction

on molecular enthalpy. Similarly, $\sum_{atoms} x(H^o_X(298K) - H^o_X(0K))$ is the enthalpy temperature correction of all atomic elements, given in [92], where X refers to each type of element and x is the number of element X. In addition, $\Delta_f H^o(M, 0K)$ is the enthalpy of formation of BTBAS molecule at 0 K, which is defined as:

$$\Delta_f H^o(M, 0K) = \sum_{atoms} x\Delta_f H^o(X, 0K)$$
$$- \left(\sum_{atoms} x\varepsilon_o(X) - \varepsilon_0(M) - \varepsilon_{ZPE}(M) \right) \tag{2.22}$$

where X refers to the type of each element, x is the number of X atom, $\Delta_f H^o(X, 0K)$ is the enthalpy of formation of element X at 0 K. Moreover, zero-point energy of molecule $\varepsilon_{ZPE}(M)$, total energy of the molecule $\varepsilon_0(M)$, and total energy of each atom $\varepsilon_0(X)$ can be found directly from G4 method [92].

In addition, viscosity μ is estimated from the molecular theory of Newtonian gases, considering pure gas composed of rigid, non-attracting molecules, which is valid for low density scenario [70]:

$$\mu = \frac{2}{3\pi} \frac{\sqrt{\pi m k T}}{\pi d^2} \tag{2.23}$$

where m is the molecular mass, k is the Boltzmann constant, T is the temperature, and d is the molecular diameter of the approximated rigid ball of BTBAS.

2.1.4.3 Microscopic Parameters Determination

For the surface reaction kinetics, many properties are difficult to measure experimentally yet essential to the accurate deposition simulation, including vibrational partition. As discussed in detail in [55], in the Si-Cycle, to calculate properties of BTBAS and its reaction kinetics with SiO_2 lattice, the hydroxylated surface $3 \times 3 \times 1$ SiO_2 lattice is constructed and optimized to an acceptable energy minimum with basis set 6-31G+dp accuracy level [61, 63, 88, 89, 93]. Subsequently, the TS complex is obtained from the optimized BTBAS molecule and a $3 \times 3 \times 1$ SiO_2 surface lattice, using optimization to transition state (Opt=TS) method with the same level of basis set

accuracy. For the O-cycle, the same calculation procedure is implemented for the ozone molecule, H-Si surface lattice and physisorbed ozone. Finally, in order to perform a precise vibrational frequency calculation, the G4 method is adopted. The vibrational partition function along with other important thermodynamic properties are reported in [55].

2.1.5   Automated Workflow and Parallelization

In order to improve computational efficiency, parallelization and distribution of the calculations are adopted in the macroscopic and microscopic domains of the ALD reactor model, respectively. In the gas-phase domain, we utilized 36 cores on the UCLA Hoffman2 Cluster where the parallel computation is implemented by the Fluent Message Passing Interface (MPI) [66]. The number of cores is chosen according to the parallel processing benchmark and the communication overhead between cores and the host node. Then, the output from the gas-phase model is passed to the microscopic model through an automatic workflow. To accurately represent the entire wafer surface, the substrate domain is spatially split into 36 sub-regions uniformly, each having a corresponding kMC model. Each kMC model runs in a serial manner, while the message passing interface distributes 36 distinct kMC models onto 36 idle cores on the cluster and manages the synchronization of all kMC models.

The interaction and methodology that connect the microscopic and the macroscopic domain are also crucial to the multiscale CFD model. In this study, the automated workflow adopted is shown in Figure 2.1. The green region enclosed by dashed line belongs to the microscopic domain (kMC), the blue region belongs to the macroscopic domain (gas-phase), and the yellow region and the associated arrows belong to the cross-linkage. The simulation starts in the macroscopic domain at $t = 0$ where mass and energy balances are solved in the flow field in a time step Δt. Next, after the reconciliation of data between each computation node, the surface boundary conditions (e.g., temperature, partial pressure and time step) are collected for each sub-region and transferred to the microscopic domain. If the system is currently in a purge cycle, nothing will be done in the microscopic domain and it would continue with the macroscopic computation for the next

31

time step. Otherwise, based on whether the system is in O-cycle or Si-cycle, the corresponding kMC algorithm will run to solve for the surface structure, reactions and precursor usage for a time interval Δt. After the computation, the results of all kMC sub-regions are synchronized, and then the averaged species fluxes for each sub-region are reported back to the macroscopic domain to update the corresponding boundary conditions for the next time step. The physical time of the system will be updated to $t + \Delta t$, according to an implicit Euler scheme [66]. If the current half-cycle has reached its assigned time limit at the end of the microscopic domain, the workflow scheme will inform the system to switch to the next half-cycle. This workflow is executed until the desired cycle number is achieved.

2.2 Multiscale CFD Simulation Results and Chamber Design

The results presented below represent the transient behavior of the thermal ALD reactor at operating conditions of P = 133 Pa and constant surface heating at T = 600 K, which are chosen according to the referenced experimental works and conventional industrial standard operating pressure and temperature of SiO_2 thermal ALD [94]. The discussion mainly focuses on the BTBAS half-cycle, because the flow field of the purge and the ozone half-cycle behave similarly, and the major economic and processing time concerns are associated with BTBAS usage and throughput. The inlet BTBAS flow rate is one of the most important operating parameters, which varies from 100 sccm to 1330 sccm, and the half-cycle times are correspondingly adjusted according to the inlet flow rate, ranging from 5 s to 1 s, respectively. The maximum Knudsen number in all in1let flow rates is 0.096, an order of magnitude smaller than unity, which indicates that the continuum assumption in the flow conservation equations is valid and continuity-based CFD models can be utilized for the gas-phase simulation [95]. In the following sections, first, the nominal reactor design is tested and compared to literature and industrial inlet flow rates as a validation of the CFD model. Then, according to the simulation results, the non-ideal flow regime in the nominal design is identified, based on which the optimization of the showerhead and the design of reactor

upstream configuration are carried out, respectively. Finally, the individual design improvements are integrated and adjusted to determine a final optimized design of the thermal ALD reactor.

2.2.1 Nominal Geometry

The final coverage in all inlet conditions from 100 sccm to 1300 sccm are around 96% ~ 98%, corresponding to a growth per cycle (GPC) of 1.74 - 1.76 Å per cycle. The wafer cannot reach complete 100% full coverage because of insufficient cycle-time, competing reaction mechanism and equilibrium, but the GPC lies inside the range of the reported value of 1.4 - 2.1 Å per cycle, and this result is also consistent with the previously reported standalone kMC model results of 1.8 Å [55, 62]. At an inlet flow rate of 1330 sccm, the gas-phase development takes 0.5 s to reach steady-state and, as shown in Figure 2.5, the entire wafer reaches steady-state coverage at 1.06 s, which is very close to the experimentally determined cycle time interpolated at the same operating conditions and reactor design, which is 0.931 s [73]. In the industrial process, the precursor flow rate is lower due to economic concern, and a typical choice of flow rate is within 50 to 100 sccm, along with a standard half-cycle time of 5 s. For our model, at an inlet flow rate of 100 sccm, the gas-phase development has reached steady-state in 5.8 s cycle time, with an average wafer coverage close to the desired unity value. Also, the importance of the interconnection between microscopic and macroscopic domain is demonstrated. Several trials of standalone CFD and consecutive standalone kMC simulation are carried out, with no updated boundary surface precursor consumption from kMC to CFD on the wafer surface. A transient differences of 10 ~ 20% in surface partial pressure, deposition rate and cycle-time are observed. These significant deviations show that the application of the interchange of information and the multiscale workflow is critical to the accuracy of ALD process simulation. Therefore, our multiscale CFD model is a good representation of the realistic full thermal ALD reactor, and the nominal reactor design at 100 sccm inlet flow rate will be used as a base case for reactor design optimization below.

As shown in Figure 2.5, the transient thin-film growth in the peripheral region (radius ¿ 120 mm) of the wafer is the fastest, and the transient thin-film growth in the central region (radius ¡ 70
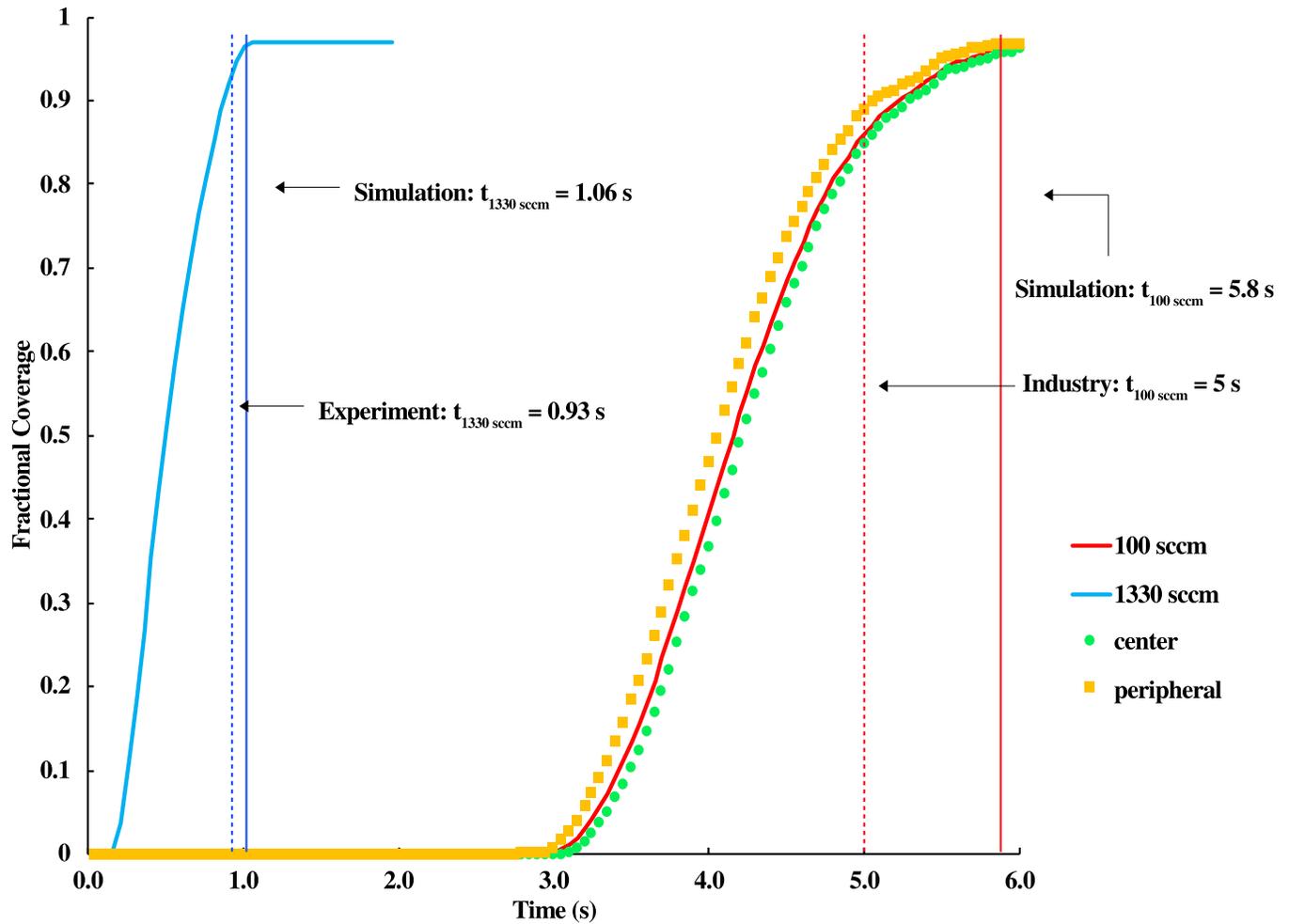
Figure 2.5: Transient deposition rates at inlet flows of 100 sccm and 1330 sccm. The *y*-axis is the fractional coverage, and the *x*-axis is the deposition time. The dashed blue and red lines are the experimental or industrial referenced cycle times at 1330 sccm and 100 sccm, respectively. The vertical blue and red lines are the simulation steady-state cycle times at 1330 sccm and 100 sccm, respectively. The blue and red solid curves are average deposition rates at 1330 sccm and 100 sccm, respectively. The yellow and green data points are the deposition rates of the peripheral and the center region, respectively.

mm) is the slowest. The maximal transient deposition rate discrepancy at a single time step is more than 400%. Since the surface temperature is maintained to be uniform by a surface heater, this large transient non-uniformity in the deposition speed is directly caused by the uneven distribution of BTBAS partial pressure on the wafer surface. This partial pressure non-uniformity is associated with the velocity profile in the reactor upstream. Because the gas flow outlet is located around the wafer, the higher local dynamic pressure difference speeds up the flow near the reactor side wall.

BTBAS
Mole fraction
1.0

0.1   (a)

Velocity
Magnitude
(m/s)
0.355

0   (b)
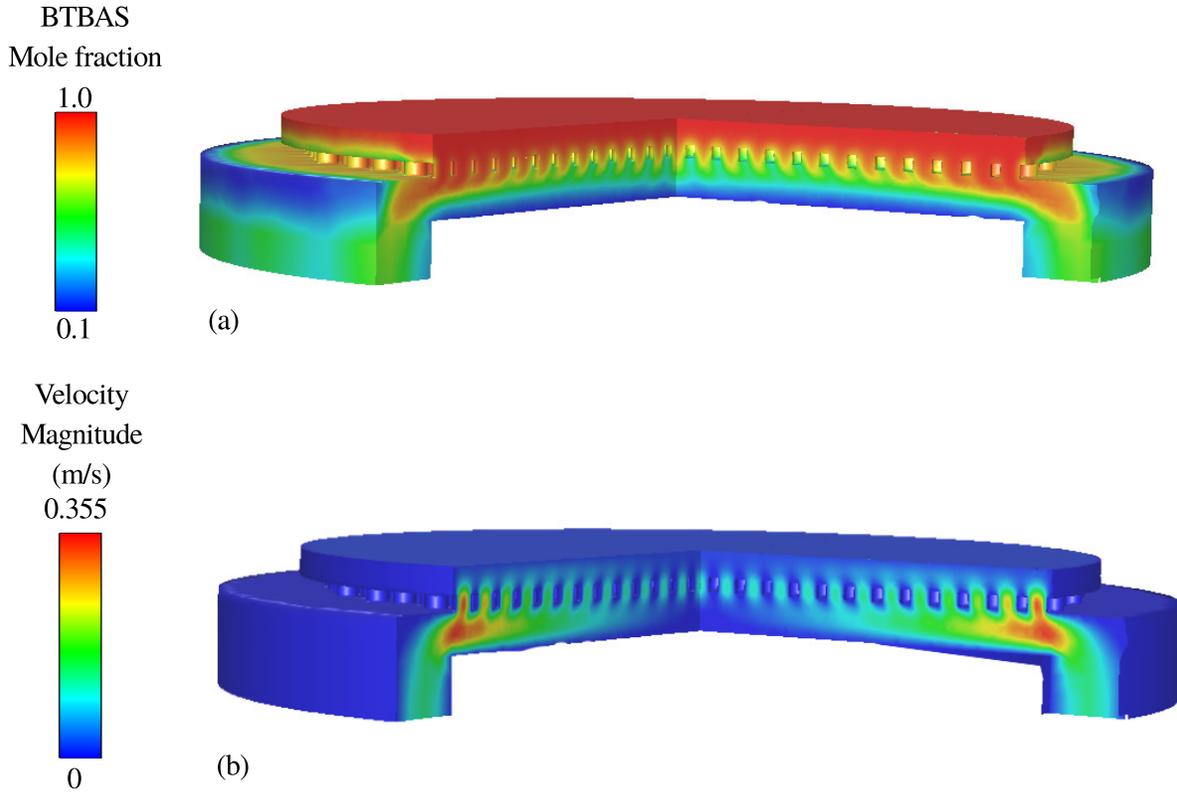
Figure 2.6: BTBAS mole fraction distribution and velocity flow field of the nominal design for inlet flow rate at 100 sccm. (a) a 3D view of the BTBAS development in the entire reactor scale at half standard Si-cycle time (2.5 s). (b) a 3D view of the velocity profile in the entire reactor scale at half standard Si-cycle time (2.5 s).

Thus, the flow velocity close to the reactor side wall is much higher and the gas flow stagnates in a temporary dead zone in the central region, shown in Figure 2.6 (b). As a result, BTBAS particles reach the outer part of the wafer much faster than the central region, as shown in Figure 2.6 (a). Therefore, even though the total pressure does not vary much on the wafer surface, the partial pressure of the precursor is lower in the center of the wafer region.

2.2.2   Improved Reactor Chamber Design

In this section, we will discuss the progressive optimization of the ALD reactor geometry. As shown in the previous section, the initial attempt to use a nominal geometry of the showerhead ALD reactor design fails to provide a uniform and fast flow field development. The non-uniformity

issue is reduced using the radially adaptive adjustment of the size and position of showerhead holes. Also, an actual inlet is added to the reactor model, and then the upstream region is modified as a volume adjusting horn, described by US Patent [60], which decreases the upstream deadzone volume and guides the precursor stream towards more uniform distribution downstream. The gas-phase simulation adopting each design improvement is analyzed and a final design that adjusts and combines those two aspects of optimization will be examined.

2.2.2.1   Showerhead Design Development

In the nominal design of the showerhead, shown in Figure 2.7 (a), there are 449 showerhead holes uniformly spaced in an orthogonal manner throughout the entire showerhead panel, which is 340 mm in diameter. Each showerhead hole has a diameter of 10 mm, and the radial distance between two showerhead holes is about 4 mm. However, as demonstrated in the previous section, this evenly distributed showerhead panel is not able to generate a uniform flow field because of the overall reactor spatial geometry.

In order to resolve the non-uniformity in the mass flow rate profile throughout the showerhead, the size and distribution of the showerhead holes can be redesigned so that the mass flow rate in the central region and the peripheral regions are balanced. First, the relationship between hole size and mass flow rate must be understood. For an ideal fluid flowing through a frictionless hole, a smaller hole diameter would speed up the fluid to keep the volumetric flow rate constant. However, a realistic flow field with complex geometry demonstrates that there is a large dependence on friction. According to the Hagen-Poiseuille equation, which is valid for describing the friction of a small-diameter tube under a flow regime with Re ¡ 20, a smaller diameter would induce a much larger friction loss:

$$S = \frac{32\mu v}{gD^2} \tag{2.24}$$

where S is the hydraulic slope (head loss due to friction), μ is the dynamic viscosity, v is the velocity and, g is the gravitational acceleration, and D is the diameter of each showerhead hole. This friction loss lowers the local differential pressure, hence slowing down the flow through the

hole. Thus, the actual volumetric flow rates through the smaller holes are much lower than those of the bigger holes. Nevertheless, the dynamic pressure across the entire wafer is complicated and the exact showerhead hole size and distribution are adjusted run-to-run according to the CFD simulation results.

A modified radially distributed showerhead design shown in Figure 2.7 (b) is developed to improve the nominal reactor geometry. In total, 233 showerhead holes are deployed throughout the panel, where the diameter of showerhead holes in the central region is 20 mm and the diameter radially decreases to 10 mm at the periphery. As shown in Figure 2.8 (a), the modified showerhead allows a more radially uniform flow field distribution across the showerhead holes with all the other operating conditions being the same. In addition, Figure 2.8 (b) and (c) demonstrate that the partial pressure on the substrate surface has less radial variation and is able to reach almost unity everywhere at steady-state. Thus, a larger showerhead hole indeed leads to a higher mass flow in the corresponding downstream region. However, this modification has a disadvantage that the substrate species distribution has an azimuthal pattern, which was resulted from the linearly arranged showerhead holes at each 60 degree angle. Therefore, this method of showerhead optimization will be further adjusted for the final design of the ALD reactor after incorporating the upstream modification.

2.2.2.2 Upstream Geometry Improvement

In the nominal design, the inlet region is assumed to be a fully developed flow field. However, in a realistic industrial ALD reactor, the precursor gas enters the reactor chamber through a narrow gas inlet that is around 10 mm in diameter, located close to the showerhead, in order to reduce the time and precursor usage for developing a steady flow profile [60]. Thus, instead of assuming a well-developed inlet region, we add the actual small inlet at the reactor top in our model to reflect realistic industrial scenario. Although the smaller and lower inlet can reduce the time of gas profile development, the sudden expansion leads to undesirable flow turbulence around the inlet and a confined vertical gas flow as shown in Figure 2.9 (a).
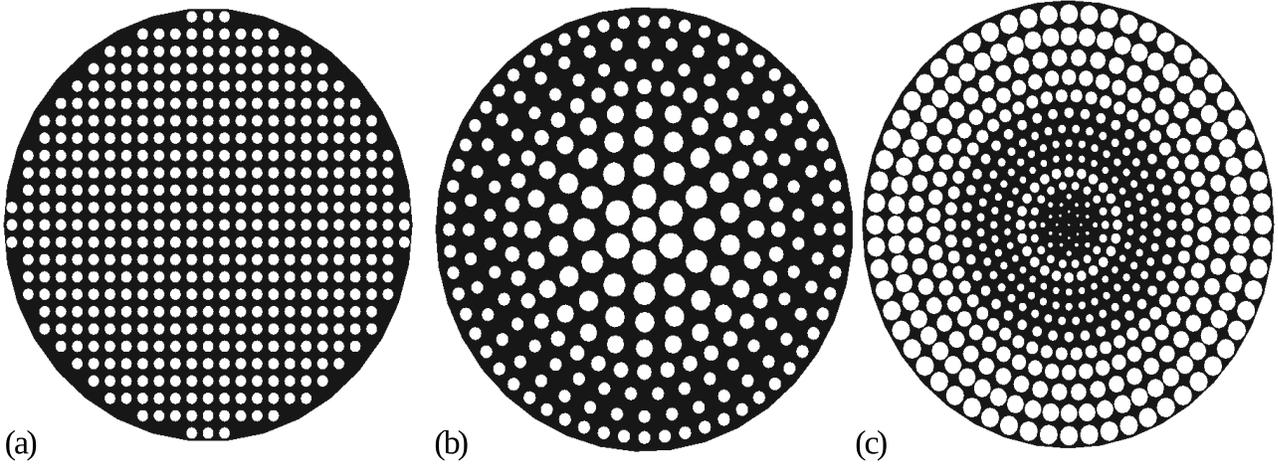
Figure 2.7: Nominal and modified showerhead designs. (a) shows the industrial standard orthogonal showerhead distribution. (b) shows the radially modified showerhead design for the well-developed inlet reactor. (c) shows the radially optimized showerhead design for the final reactor design with a small inlet and a modified upstream.

A reasonable solution is to add a physical component, a volume adjusting horn, in the dead zone of the upstream region after the small inlet to reduce the discharge coefficient and to guide the velocity and pressure profiles towards the peripheral region, as shown in Figure 2.9 (b) and (d) [60]. The simulation results from Figure 2.9 (a) and (b) demonstrate that the inlet turbulence is reduced to a great extent, preventing possible damage of the deposited film. In addition, Figure 2.9 (c) and (d) show that the gas stream tends to flow vertically after it enters the reactor chamber, and the horn-shaped upstream configuration successfully generates more radial momentum so that region far away from the center can have an increased mass flow. Thus, this canopy-like upstream chamber wall facilitates the radial motion of gas stream, reduces undesirable inlet disturbance and allows for more uniform velocity and precursor partial pressure development.

2.2.2.3   Integrated Optimal ALD Chamber Design and Evaluation

According to the individual optimizations introduced in the previous sections, the transient flow non-uniformity could be mitigated by changing the sizes of showerhead holes as well as their locations. In general, larger holes lead to higher volumetric flow rate. The actual small inlet represents the realistic industrial reactor geometry, and reduces gas-phase development time. The
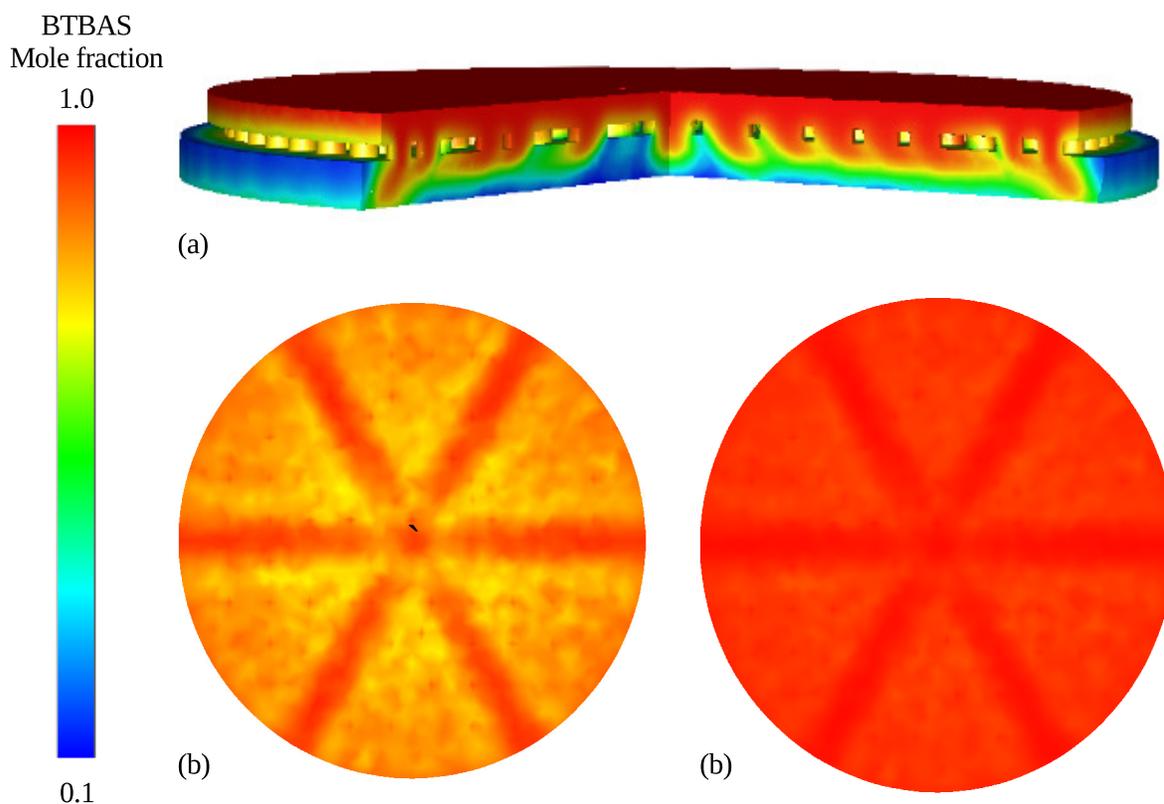
Figure 2.8: BTBAS mole fraction distribution of the radially adjusted showerhead design for 100 sccm inlet flow rate. (a) shows a 3D view of the modified showerhead region. (b) and (c) show the transient profile and the steady-state profile of the substrate surface, respectively.

canopy-like upstream chamber decreases dead zone volume and enhances radial flow. Therefore, in order to reduce the ALD cycle time, the final design should consider all aforementioned geometry optimizations and fine-tune them to achieve the optimal result.

The final ALD reactor design combines the small inlet, the modified upstream region introduced in the last section, shown in Figure ̊ffig:upstream2 (b), and the optimized showerhead holes distribution, shown in Figure 2.7 (c). Although it is demonstrated that a confined flow can lead to a faster gas-phase development, the high velocity ($>$30 m/s) of the confined gas stream could potentially cause showerhead fouling and damage to the deposited film and long-term contamination of the showerhead holes. The high speed stream, possibly combined with radicals from plasma treatment, will have enough bombardment energy that is destructive to the deposited
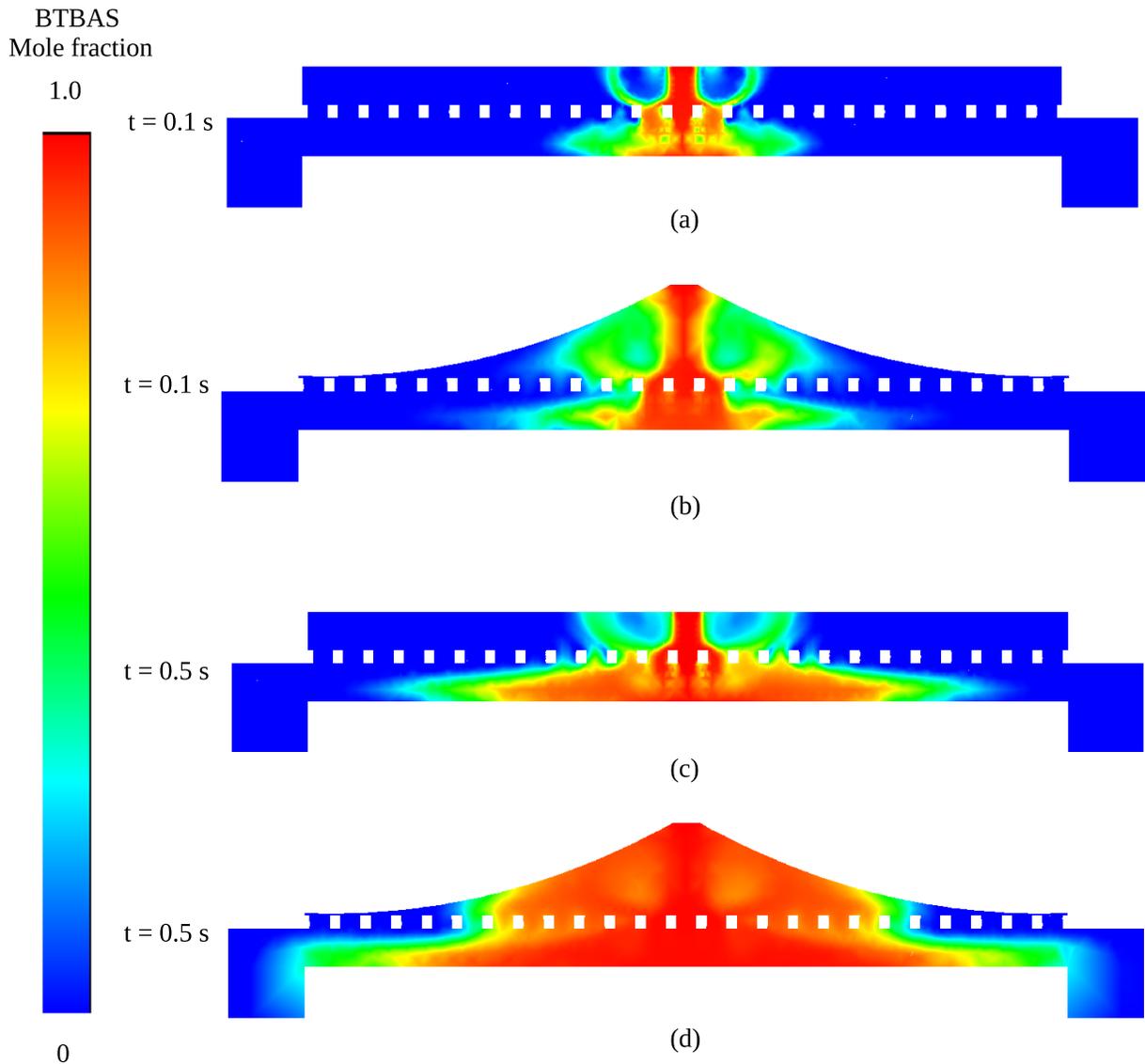
Figure 2.9: BTBAS mole fraction comparison of the modified upstream design and the sudden small inlet design for 100 sccm inlet flow rate. (a) and (c) show the sudden inlet design at 0.5 s and 1 s respectively. (b) and (d) show the modified upstream design at 0.5 s and 1 s, respectively

wafer surface. Therefore, in this design, the central region of the showerhead panel contains small holes that are 5 mm in diameter to prevent the direct mass flow onto the substrate surface and enhance radial dispersion. The holes immediately around the central region are much bigger in order to reduce the turbulence and to compensate for the low downward dynamic pressure caused by the turbulence, and the outer showerhead holes are radially increased according to the velocity

profile.
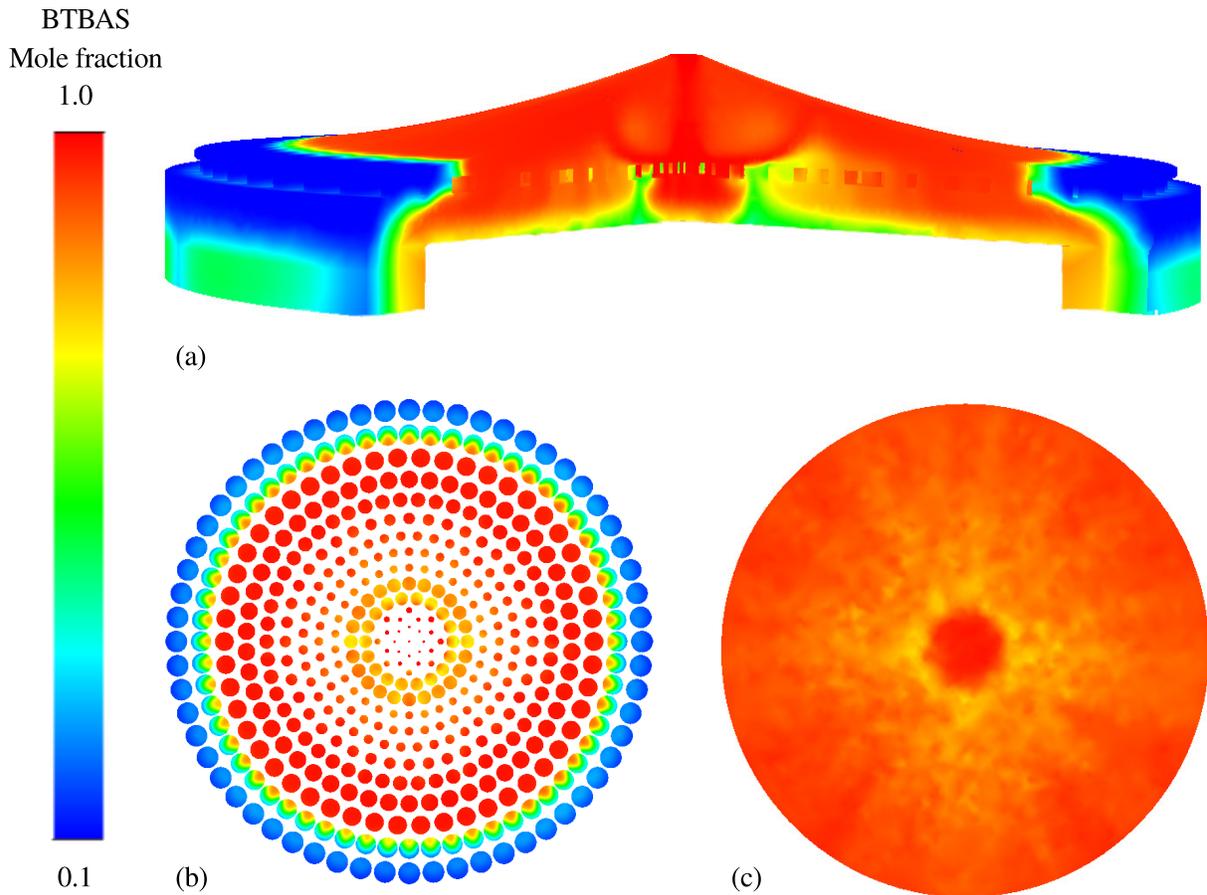
Figure 2.10: BTBAS mole fraction distribution of the final reactor design at 1.25 s for 100 sccm inlet flow rate. (a) shows the 3D view of the entire reactor. (b) shows the *z*-cross-section view at the middle of the showerhead. (c) shows the wafer surface.

As shown in Figure 2.10 (a), the species mass transfer is much more uniform radially across the entire wafer and is causing less turbulence than that in Figure 2.9 (d). The small showerhead holes in the middle of the showerhead successfully reduce the gas velocity right below the inlet area, while the horn-shaped upstream and the bigger holes, together, create higher differential pressure gradient towards the side of the reactor, hence increasing the radial velocity. As a result, the species development across the showerhead holes and wafer surface are acceptably uniform, as shown in Figure 2.10 (b) and (c). In addition, it is shown in Figure 2.11 that the optimized final design leads

41

to a 39.6% shorter finishing time for BTBAS half-cycle.
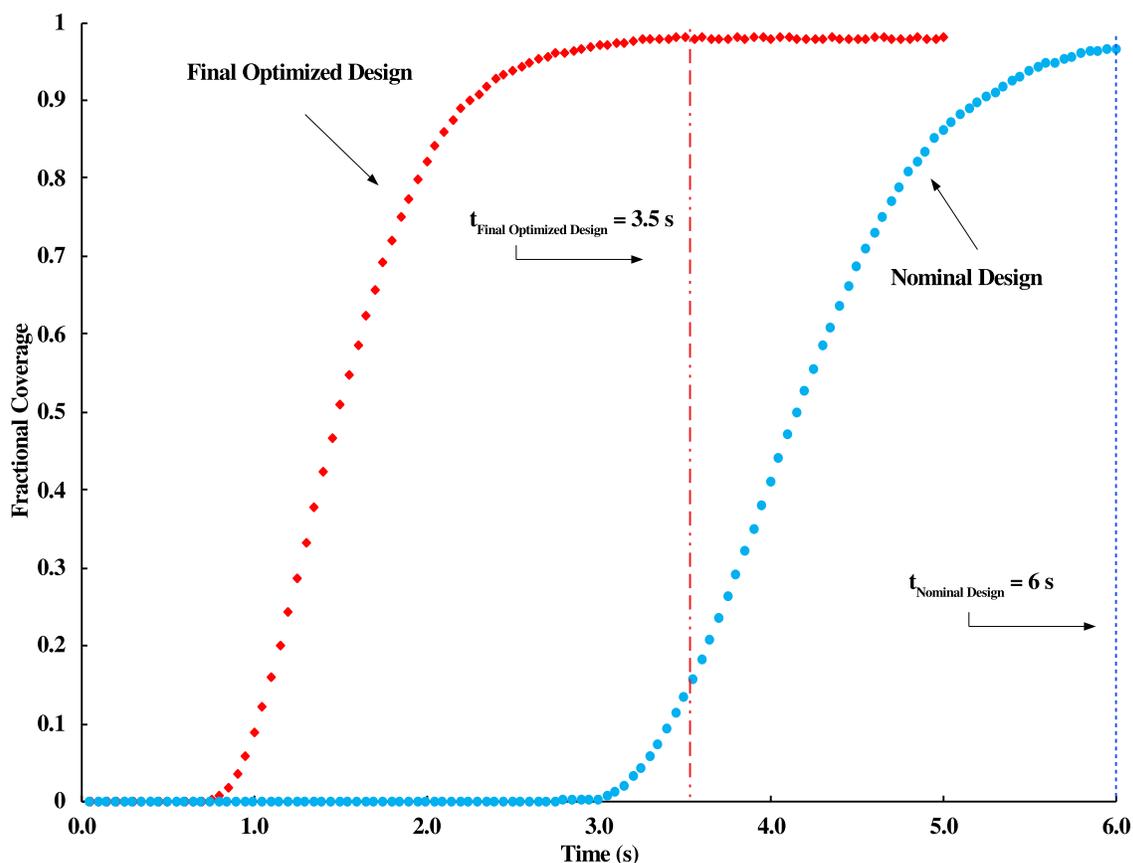
Figure 2.11: Transient deposition rate comparison of the nominal design and the final optimized design. The *y*-axis is the fractional coverage, and the *x*-axis is the deposition time. The red data points and dashed line are the transient fractional coverage and the steady-state cycle time of the optimized reactor design, respectively. The blue data points and dashed line are the transient fractional coverage and the steady-state cycle time of the nominal reactor design, respectively.

2.3 Conclusion

In this work, we developed a first-principles-based multiscale 3D CFD model of thermal ALD of SiO_2 thin-films using BTBAS and ozone as precursors. Application of this model for a nominal reactor geometry design sucessfully reproduced experimental and industrial GPC and half-cycle time at 100 sccm and 1330 sccm precursor inlet flow rates. Based on the transient non-uniformity discovered in the nominal design, geometry optimizations on the showerhead hole

size and distribution and upstream region were performed and their influence on the gas-phase profile and microscopic deposition were investigated. Two geometry optimizations were adjusted and combined for a final optimized design of the ALD showerhead reactor. The resulting optimal chamber design successfully reduced the gas-phase profile development time and created radial momentum that guided the gas stream to the less concentrated region. As a result, under the same industrial standard operating conditions, the optimized geometries reduced the half-cycle time by 39.6%, while maintaining a GPC of 1.76 Å per cycle. Moreover, The multiscale simulation methodology developed in this section can serve as a guidance model to other ALD systems where gas-phase transport and surface deposition need different simulation methods with information exchange. With the correct reaction kinetics for the microscopic film growth model and thermodynamic data of the reactants and products for the macroscopic gas transport model, such multiscale model can realistically represent a variety of ALD processes. In our future work, we are going to delve into the connections between the operating conditions and the cycle-time needed and to build a data-driven model to characterize such complex relationship. This model will systematically reflect the influences of the operating temperatures, pressures and flow rates, and also will help experimental and industrial works select and control ALD parameters more effectively and efficiently.

Chapter 3

Integration of Feedback Control and Run-to-Run Control for Plasma Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films

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. For example, the transistors in recent memory devices like 3D NAND (not-and) 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].

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 [96]. 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 [97].

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 [98]. 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 [99]. Thus, experimental works have only been able to investigate a limited range of precursor selections and operating conditions. Furthermore, because of the large physical dimension of the batch ALD systems,

45

they are much more susceptible to batch-to-batch variation and process disturbances compared to single wafer systems [102, 103]. 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 [73]. Several researches have investigated control schemes for single-wafer and small-scale CVD and ALD systems [49, 50, 53, 104]. For example, [54] implemented feed-back control scheme for plasma enhanced CVD system to enhance the film thickness uniformity. [105] utilized an ion beam sculpting feedback control to fabricate non-porous thin silicon nitride membranes using ALD coated pores. [106] 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 [107, 108]. 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 [56].

In this chapter, a multiscale CFD model and an integrated on-line control run-to-run scheme are constructed for a furnace ALD process of SiO_2 thin-film production with Bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors, as shown in Figure 3.1. 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.

3.1   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. Specifically, the ALD process of 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 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.

This section is structured as follows. 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.

3.1.1   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 [66].
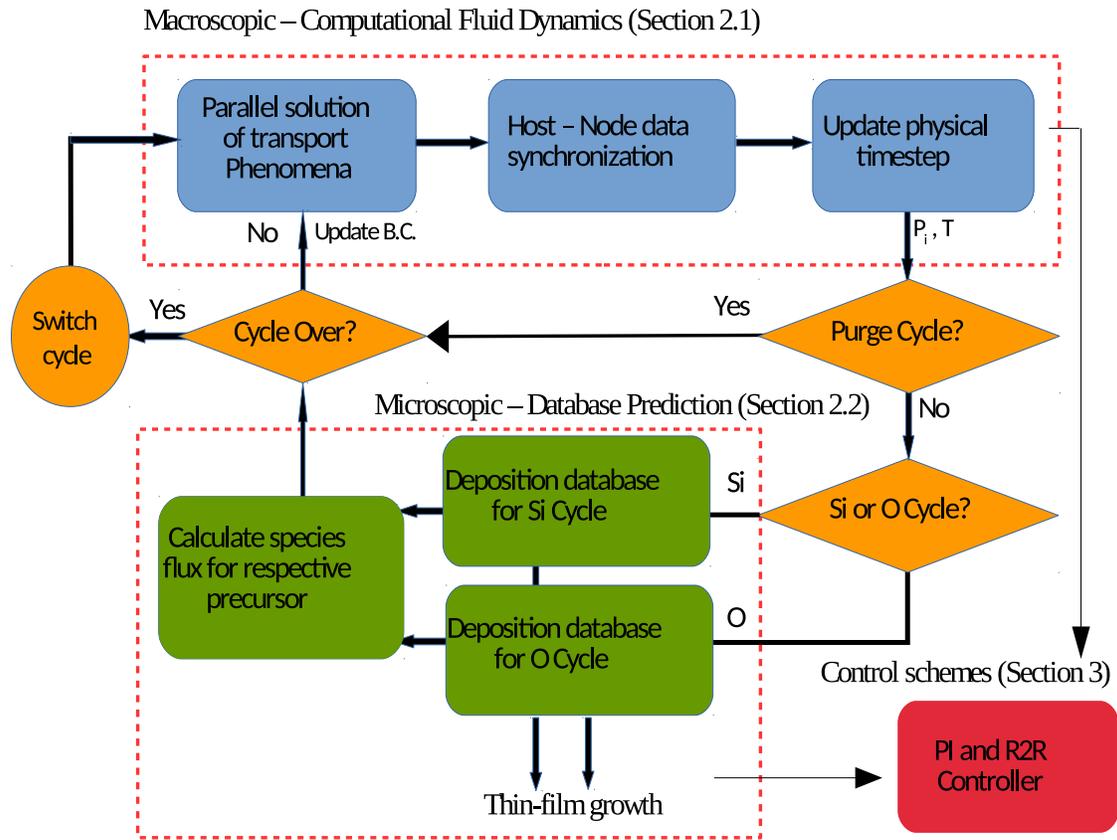
Figure 3.1: Multiscale CFD simulation workflow, parallel processing, information exchange illustration, and control schemes.

3.1.1.1 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 3.2 (a), which is demonstrated to have better performance than a parallel furnace platform or a spatial platform [73]. 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

48

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 3.2 (b), can be achieved using AUTOCAD, a computer-aided design software.

(a)　　　　　　(b)
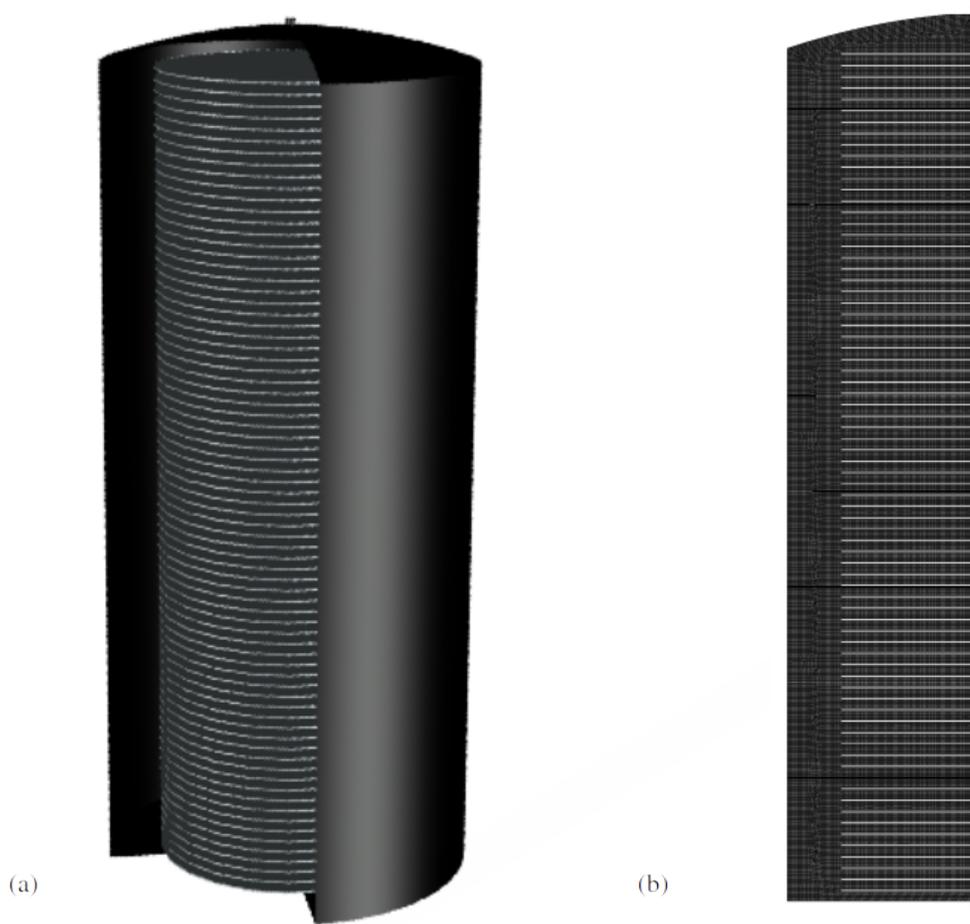
Figure 3.2: Cylindrical furnace ALD reactor geometry and the 2D-axisymmetric meshing.

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. 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 3.1. 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 [66, 69].

Table 3.1: Mesh quality metrics

	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

3.1.1.2   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 3.2 (b) 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 [109].

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 [74]:

$$C = \frac{u\Delta t}{\Delta x} \tag{3.1}$$

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 Δx is the minimum cell length. It is demonstrated that a commercial CFD software allows convergence with high Courant number around 50 [75]. The transport of species, energy, and momentum predicts the

50

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 [66, 72]:

$$\vec{J}_i = -\rho D_{m,i} \nabla Y_i - D_{T,i} \frac{\nabla T}{T} \tag{3.2}$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla \cdot (\rho \vec{v} Y_i) = -\nabla \cdot \vec{J}_i + R_i + S_i \tag{3.3}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = \nabla(k \nabla T - \Sigma h \vec{J} + (\bar{\bar{\tau}} \vec{v})) + S_h \tag{3.4}$$

$$\frac{\partial(\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla P + \nabla \cdot (\bar{\bar{\tau}}) + \rho \vec{g} + \vec{F} \tag{3.5}$$

$$\bar{\bar{\tau}} = \mu \left[(\nabla \vec{v} + \nabla \vec{v}^T) - \frac{2}{3} \nabla \cdot \vec{v} I \right] \tag{3.6}$$

where $\vec{J}_i$ is the diffusive flux of species i, ρ 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, $\bar{\bar{\tau}}$ 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, μ 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 [37]. 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 [110]:

$$D_{ij} = 0.00188 \frac{\left(T^3 \left(\frac{1}{M_{w,i}} + \frac{1}{M_{w,j}}\right)\right)^{\frac{1}{2}}}{P_{abs} \sigma_{ij}^2 \Omega_D} \tag{3.7}$$

where P_{abs} is the local absolute pressure, σ_{ij} is the arithmetic average of the individual rigid elastic sphere diameter, Ω_D is the diffusion collision integral, which is a function of the Lennard-Jones parameter ε that represents the intermolecular interaction.

3.1.2 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 [55]. 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 [101]. Thus, in this chapter, we adopt the developed deposition rate database to simulate the ALD surface reaction, instead of running the computationally demanding kMC models.

3.1.3   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) [66]. 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 3.1, 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, Δt, 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 $t + \Delta t$ [66].

3.1.4 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 [98] and in [99] utilized empirical knowledge and a rudimentary gas-phase CFD computation to analyze the transport time scale. Specifically, the time scale results are compared to the 2D asymmetric multiscale CFD model in this section 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 is 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 3.3 (a), which is chosen at a random process time for illustration purpose. The downward transfer in the vertical region besides the wafers is 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 is diffusion-only in the radial direction because the velocity is approaching zero. In addition, under a rotary vacuum pump setting of 12 M^3/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 is higher in the reactor because of the static pressure, reaching around 500 Pa, as shown in Figure 3.3 (b). The flat-out orange region is a result of the diffusion dominant mass transfer. In the majority part of the reactor, the velocity magnitude is 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 are 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 reaches the edges of upper wafers under the diffusion across the side of the
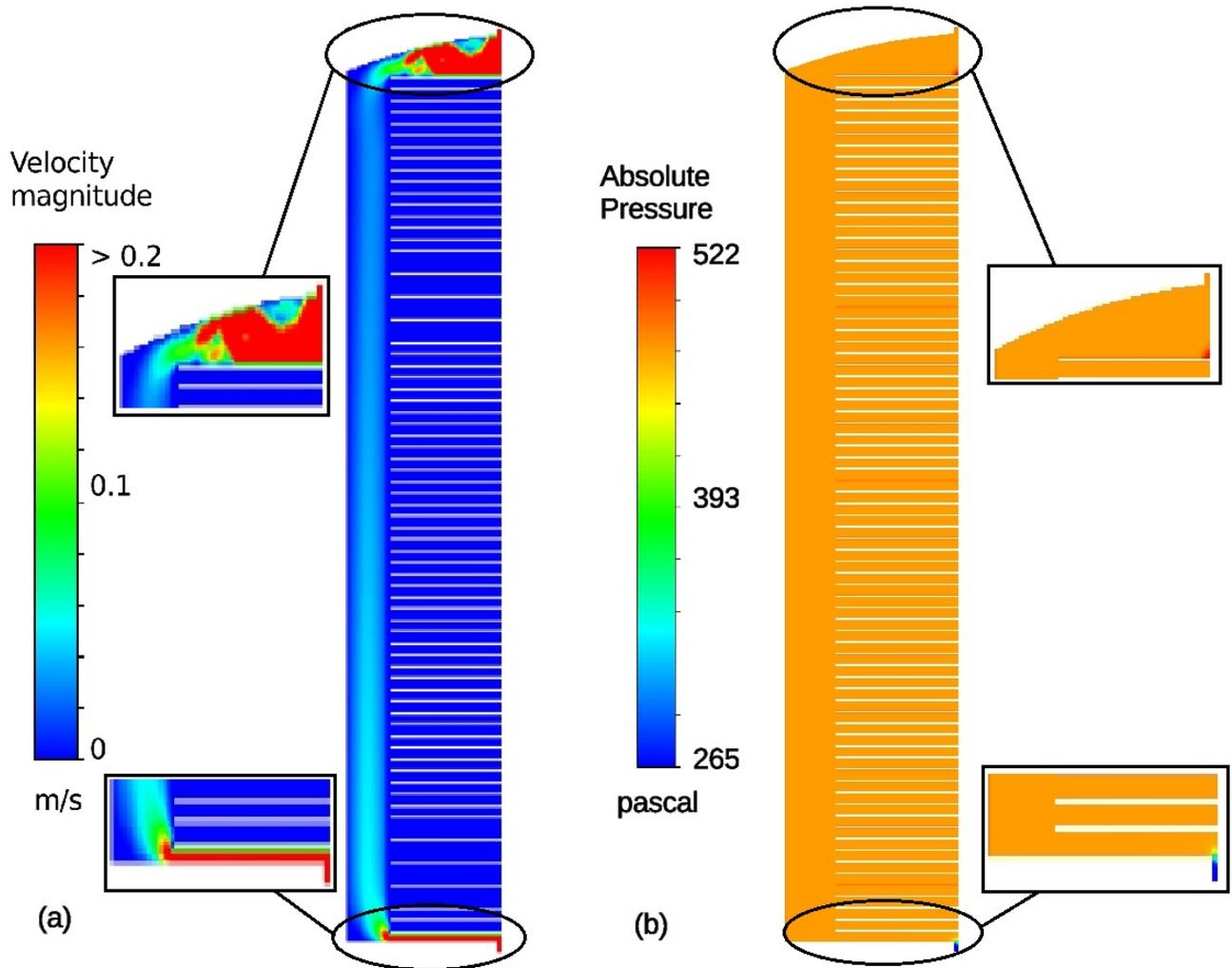
Figure 3.3: CFD velocity and pressure profiles at a randomly chosen time for simulation illustration (t = 60 s) at an operating pressure of 266 Pa. (a) Velocity magnitude. A range of velocity magnitudes is represented by colors, with the red characterizing velocities with magnitude higher than 0.2 m/s. (b) Absolute pressure profile.

furnace, and finally reaches the center of the bottom wafers under diffusion into the wafer, as shown in Figure 3.4 (a). In the meantime, as the precursor comes into the reactor and in between the wafers, Ar is purged out concurrently, as shown in Figure 3.4 (b). The corresponding dynamic gas-phase BTBAS partial pressure development profiles at the top, middle and bottom wafer surfaces are shown in Figure 3.5 (a). Additionally, the transient film coverage developments at the middle and the peripheral wafer regions are shown in Figure 3.5 (b). The gas-phase dynamics and
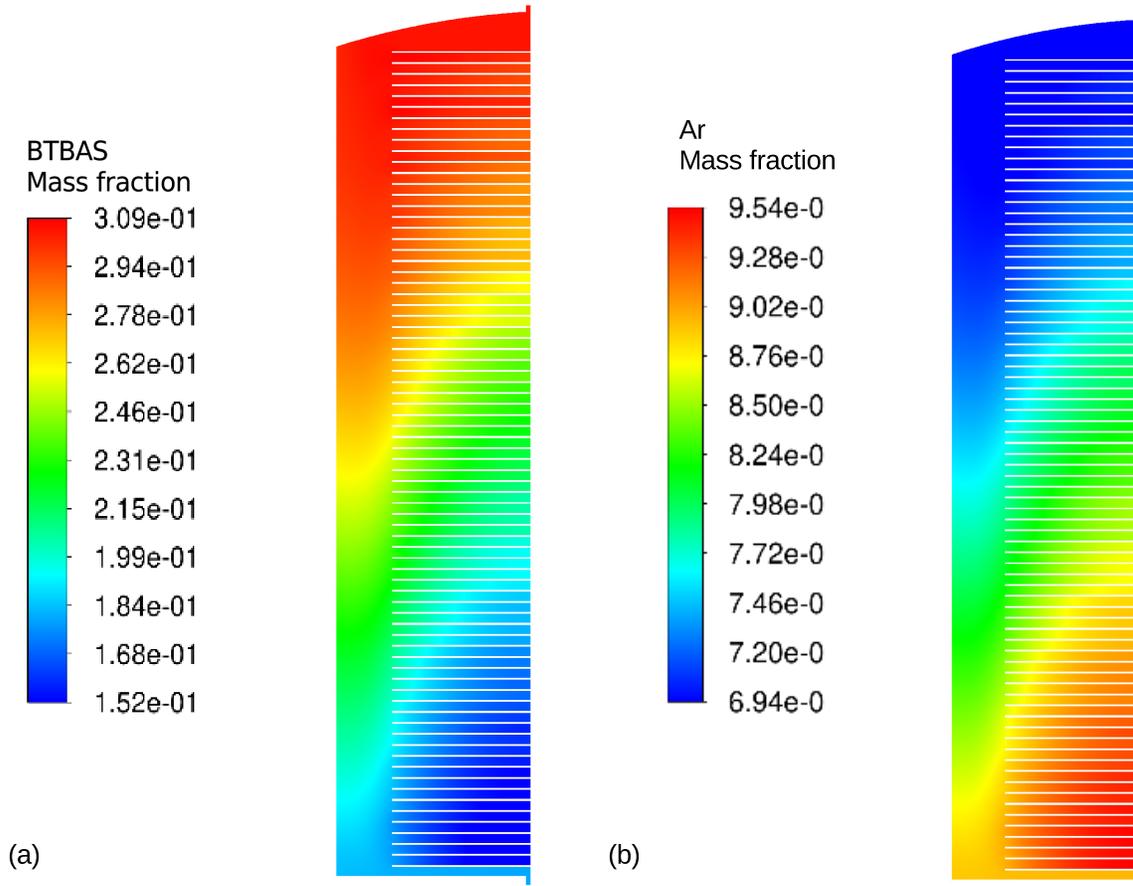
Figure 3.4: CFD species transport profiles at a randomly chosen time for simulation illustration (t = 60 s) at an operating pressure of 266 Pa. (a) BTBAS mass fraction profile. (b) Ar mass fraction profile.

the surface deposition at different reactor locations are demonstrated to be similar to the reported profiles in [98]. In addition, the half-cycle time, at which all the wafers reach full coverage, ranges 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 [73]. 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 are taken as measured observables as discussed in the next section for the formulated

control schemes.
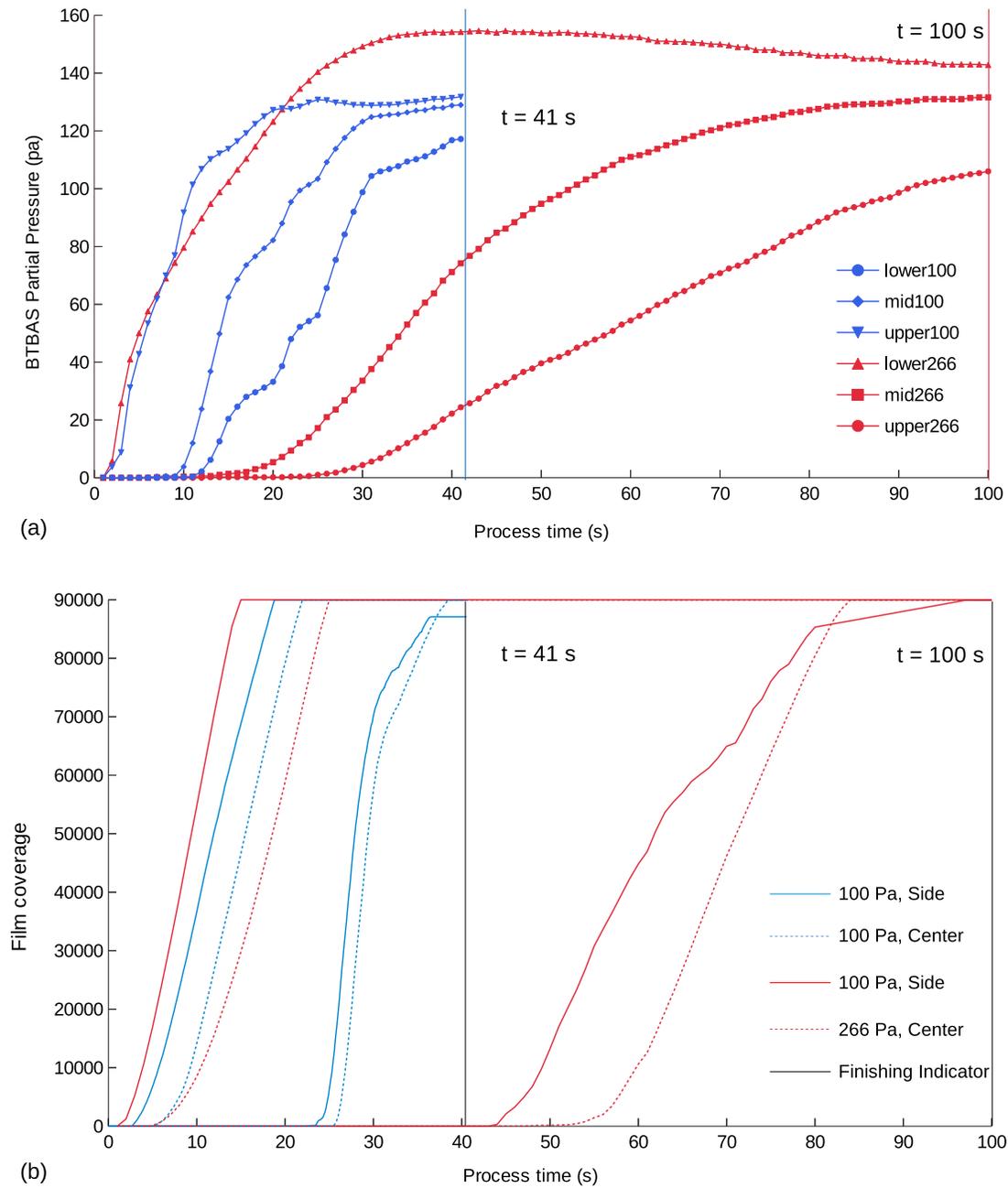
Figure 3.5: (a) BTBAS partial pressure evolution at various reactor levels. (b) Film coverage profile on the top and bottom substrates, which are located on the left side and right side of figure, respectively.

3.2 Control schemes

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.

3.2.1 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, [111] 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.

3.2.1.1 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, τ_{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 3.6.
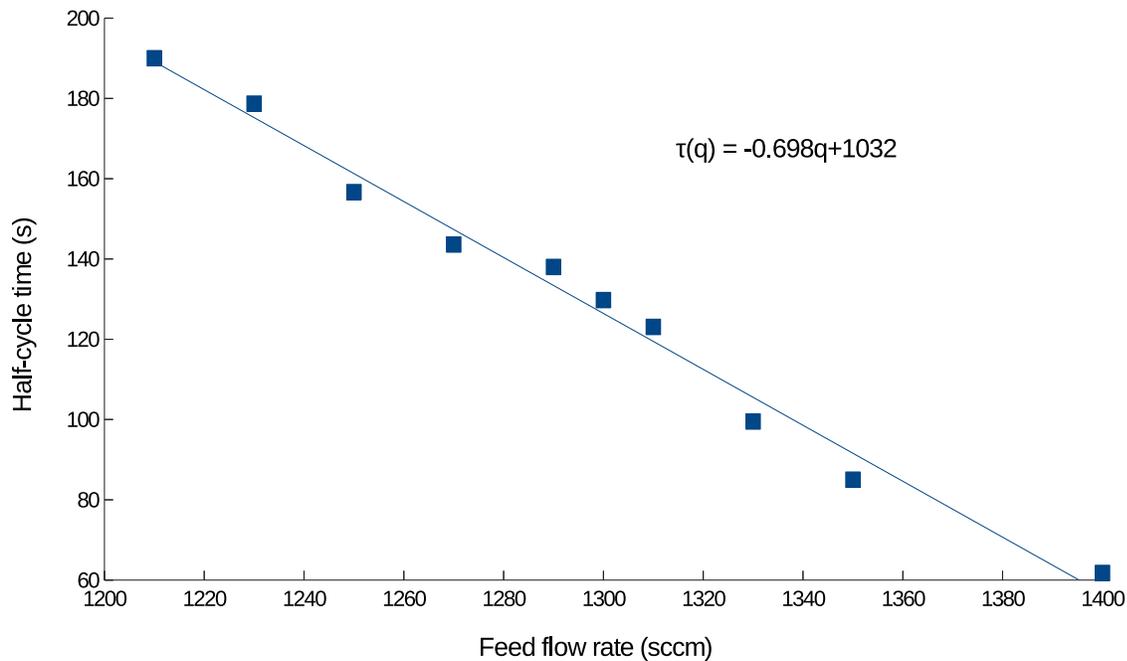
Figure 3.6: Relationship between the half-cycle time and the feed flow rate. The blue dots represent the half-cycle time from the simulation and the blue line represents the regression fitting in the operating region.

3.2.1.2 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 section, 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:

$$\varepsilon_{k+1} = (1-\lambda)\varepsilon_k + \lambda(q_{sp} - q_{cfd})$$
$$q_{k+1} = q_k + \varepsilon_{k+1}$$
(3.8)

where ε_k and ε_{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.

3.2.1.3 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 3.7, 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

60
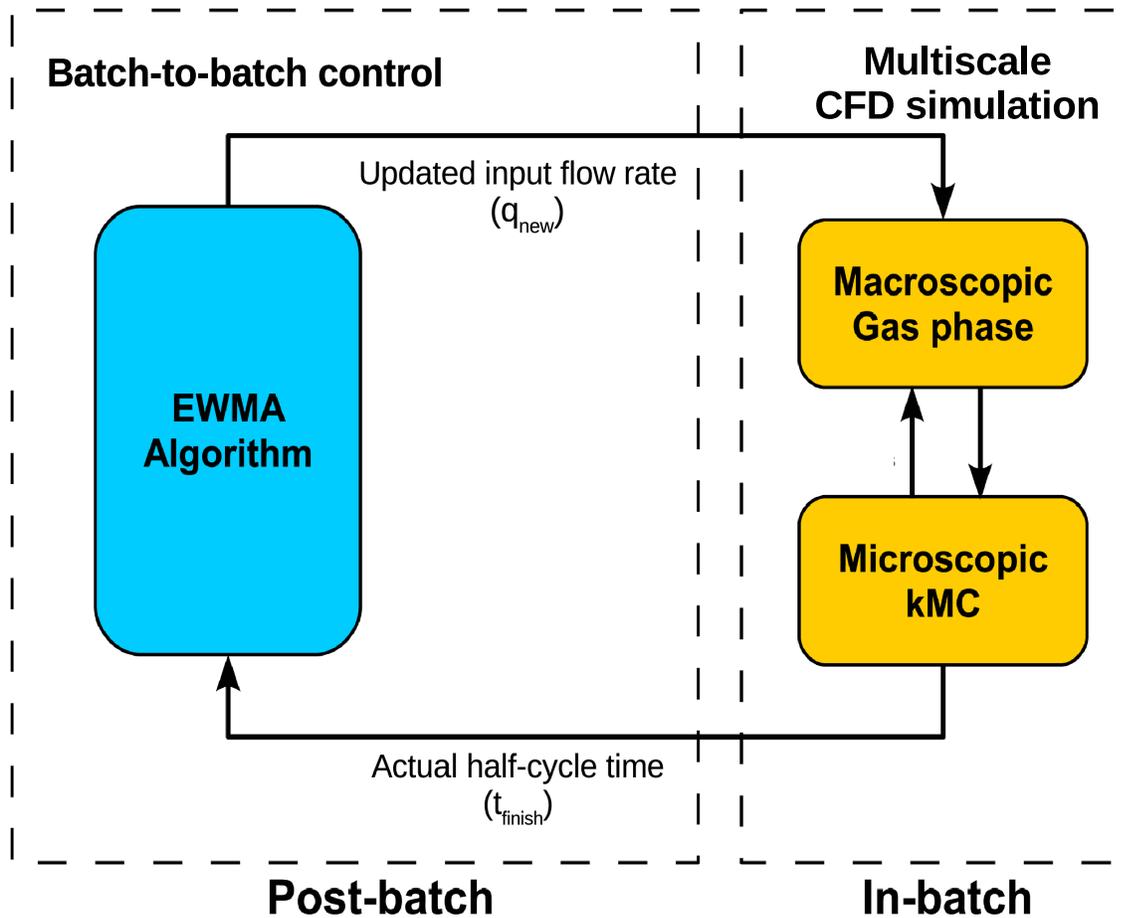
Figure 3.7: Schematic illustration of the closed-loop system under the post-batch R2R controller integrated with the multiscale CFD furnace ALD reactor model.

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 3.8, the upward (green) and downward (yellow) triangular data points represent the R2R closed-loop system with learning factors λ 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 12 and 25 batches, respectively.
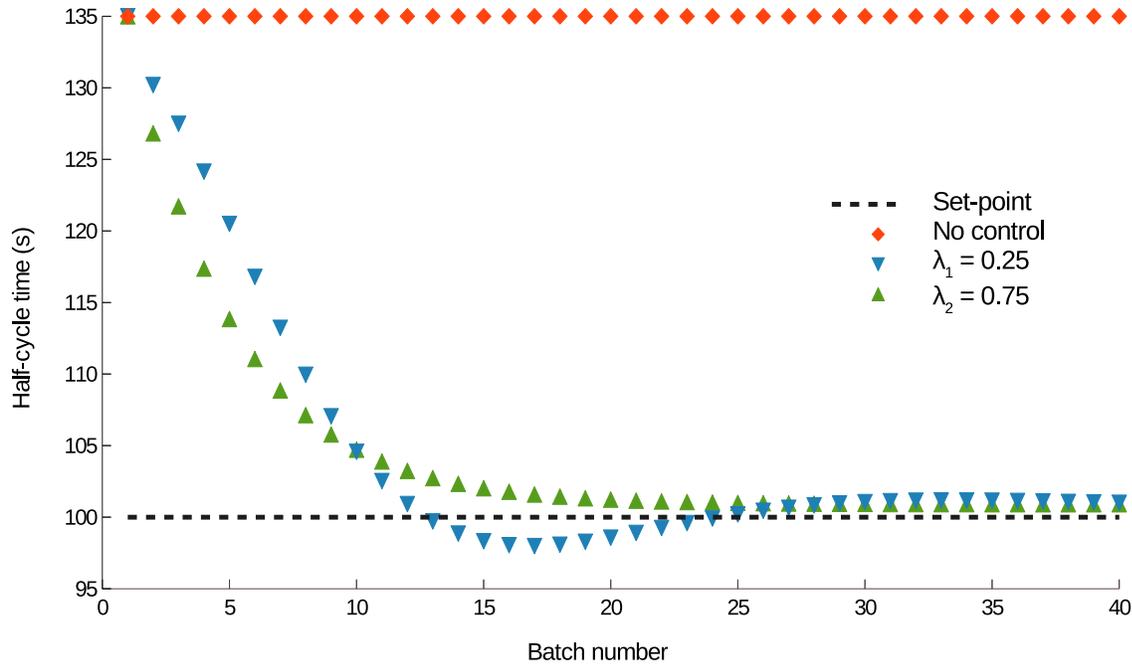
Figure 3.8: Closed-loop half-cycle time correction via an R2R controller under vacuum pump variation.

3.2.2 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. 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 chapter, a PI control scheme is formulated for the furnace ALD system to counteract process disturbances.

3.2.2.1 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 [112, 113]. This gas sensing technique has been proven to be compatible for the ALD process [114]. Also, the chosen precursor BTBAS is a perfectly flammable precursor, suitable for the aforementioned measurement devices [115].

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:

$$u(t) = K_C \left(e(t) + \frac{1}{\tau_I} \int_0^t e(t^{'}) dt^{'} \right) \tag{3.9}$$

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

3.2.2.2 PI controller tuning

In order to determine the appropriate PI controller parameters K_C and τ_I, Cohen-Coon tuning rules, the standard PI tuning rule appropriate for non-interactive controller, is utilized to provide

an initial estimate for the tunning 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 $\tau_i = 28.5$.

3.2.2.3 Closed-loop operation results under PI control
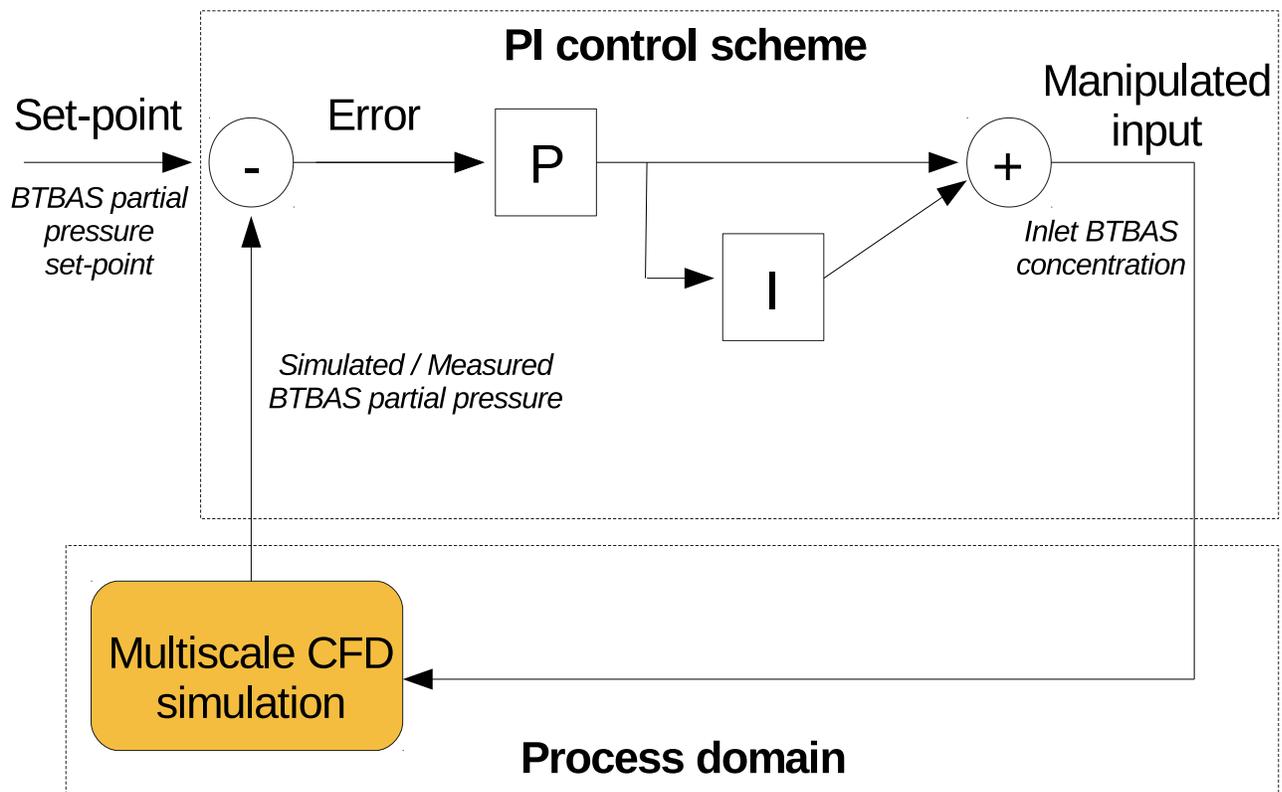
Figure 3.9: Schematic illustration of the closed-loop system under the PI controller and the multiscale CFD model.

With the calculated set of PI controller parameters, the closed-loop control system is

constructed as shown in Figure 3.9. 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 3.10 (a), 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 3.10 (b). This BTBAS partial pressure valve allows full coverage to be achieved for all wafers but leads to a variation of the half-cycle time.

(a)

(b)
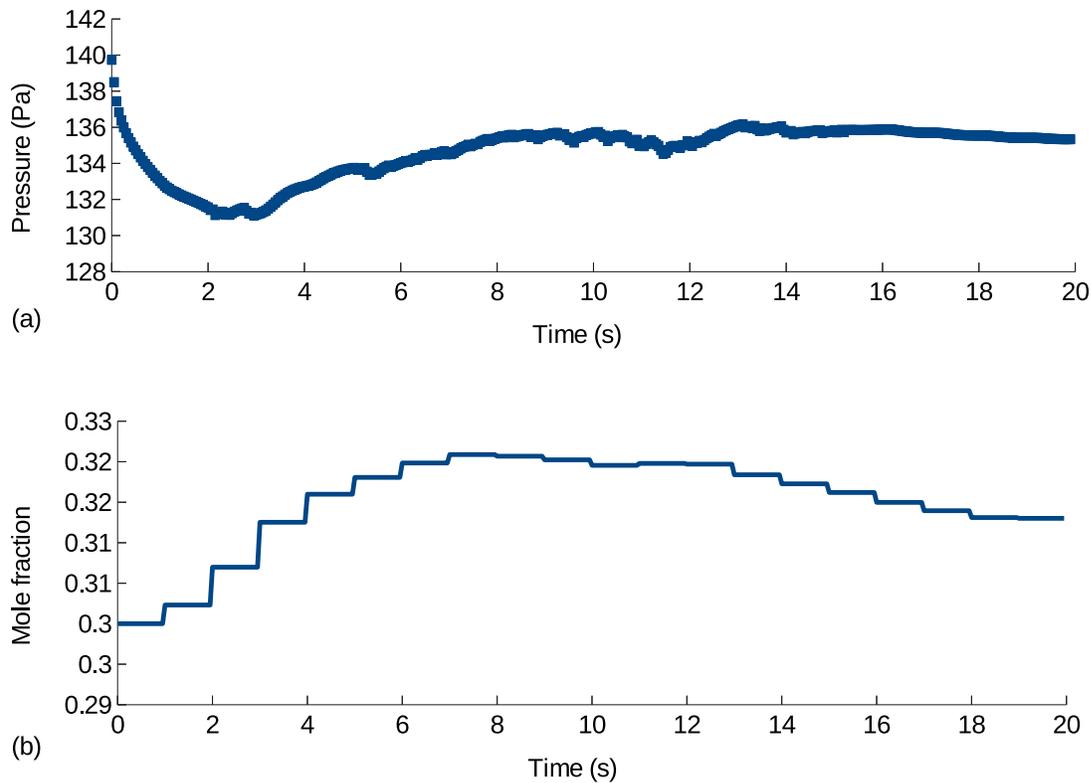
Figure 3.10: (a) Closed-loop BTBAS precursor partial pressure set-point tracking under vacuum pump disturbance, driving the operating pressure to be 100 Pa. The *x*-axis is the process time starting when the PI controller is initially triggered. The *y*-axis is the BTBAS partial pressure at the gas sensor position. (b) Manipulated input profile. The *y*-axis is the manipulated inlet BTBAS mole fraction.

3.2.3 Integrated R2R/PI control implementation

After the PI and R2R control schemes are separately tuned and validated, the two controllers are 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 is implemented with the same activation criterion. At the end of the first batch, the R2R controller is 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 is again introduced in the CFD simulation, which drives the equivalent operating pressure of the reactor to be at 500 Pa. For the real-time feedback control, it is demonstrated that a reduction of inlet BTBAS mole fraction is required to reduce the amount of precursor needed at this pressure. As shown in Figure 3.11, a series of control actions with a sampling time of 1 s are carried out according to the PI control scheme, which reduces the inlet feed mole fraction to 0.226 and drives the precursor partial pressure to the set-point and stabilizes the process output after 18 s. Also, the controller successfully reduces 23% of the precursor usage in the first batch.

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

$$\tau(q) = -0.11q + 322.16 \tag{3.10}$$

According to the reactor response under the updated feed mole fraction, a half-cycle time of 130 s is chosen to be the set-point for the R2R control, and a corresponding q_{sp} is 1700 sccm. As shown in Figure 3.12, the diamond (red) data points represent the simulated half-cycle time under the existing pressure disturbance, and the downward (yellow) and upward (green) triangular data points represent those under the R2R control schemes with a learning factor λ of 0.25 and 0.75, respectively. It is demonstrated that under the proposed R2R control scheme, the input feed flow rate gradually decreases, leading to an increased half-cycle time, which eventually reaches the set-point. Also, as shown in Figure 3.12, both learning rates allow convergence towards the

half-cycle set-point, but a learning factor $\lambda = 0.75$ allows faster convergences and less overshoot. Besides, without the R2R control scheme, the performance of the batch ALD reactor is negatively affected by the disturbance, resulting in a 5.7% waste of precursors.

(a)

(b)
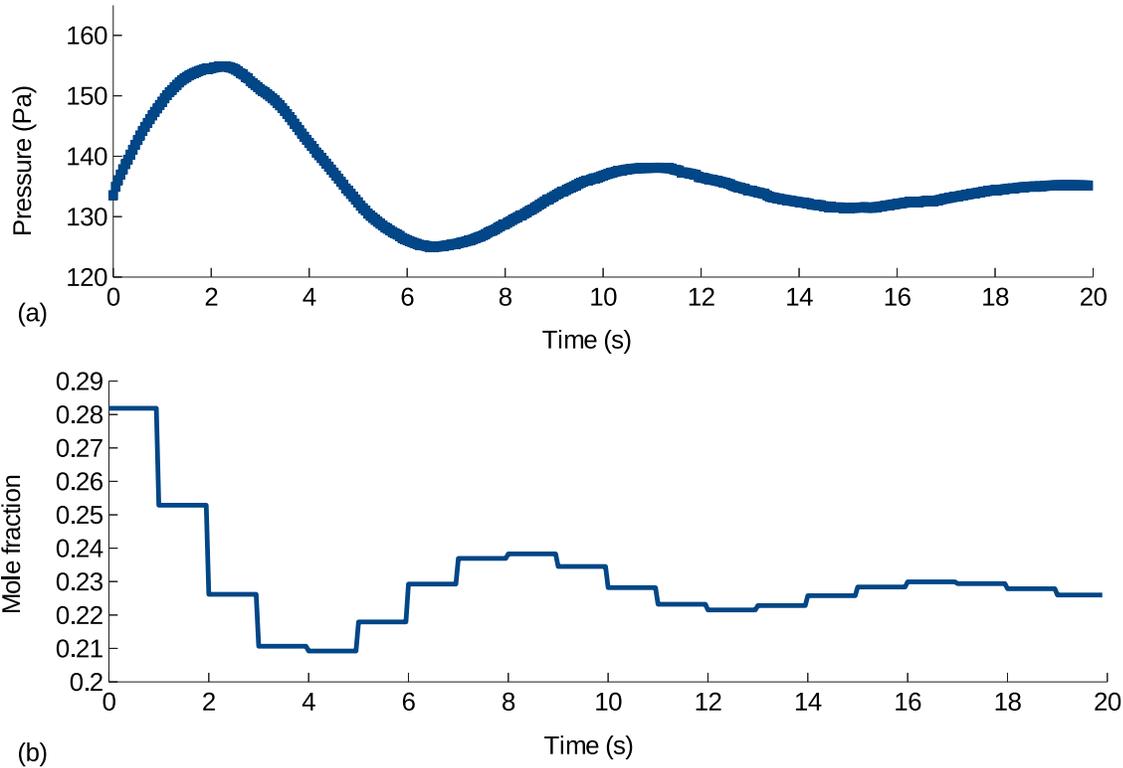
Figure 3.11: (a) Closed-loop BTBAS precursor partial pressure set-point tracking under a disturbance in the vacuum pump, driving the operating pressure to be 500 Pa. The *x*-axis is the process time starting when the PI controller is initially triggered. The *y*-axis is the BTBAS partial pressure at the gas sensor position. (b) Manipulated input profile. The *y*-axis is the manipulated inlet BTBAS mole fraction.

3.3 Conclusion

In this section, 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 SiO_2 thin-film ALD using BTBAS and ozone as precursors. Firstly, the multiscale CFD model was built and validated, which
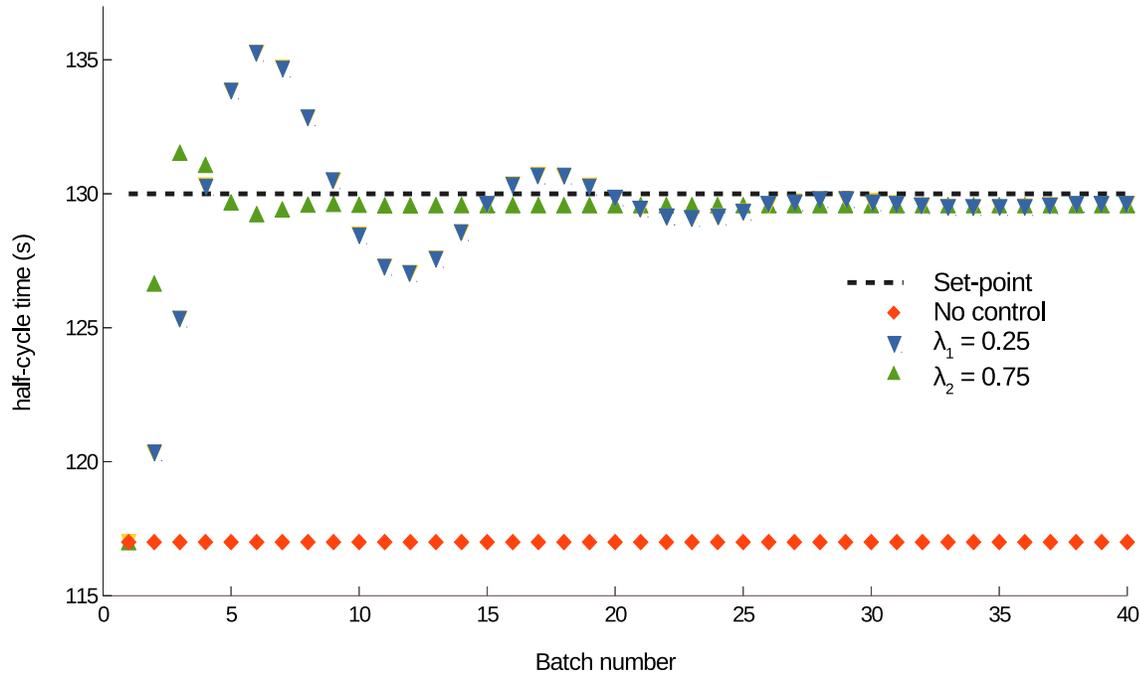
Figure 3.12: Closed-loop half-cycle time correction via an R2R controller under the updated feed precursor mole fraction and the vacuum pump variation.

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%.

Chapter 4

Multiscale Computational Fluid Dynamics Modeling and Reactor Design of Plasma-Enhanced Atomic Layer Deposition

4.1 Introduction

In the semiconductor device fabrication, atomic layer deposition (ALD) is developed and utilized to enhance the deposition conformity in the high aspect ratio structures and to make the production of extremely-thin oxide films possible for the metal-oxide-semiconductor field-effect transistors (MOSFETs) [3]. To miniaturize the NAND (Not-And)-type flash memory devices with the pace of Moore's Law, companies and experimental labs have been constantly striving to reduce the dimension of the oxide films (e.g., SiO_2, Si_3O_4 and Al_2O_3) [1]. However, the leakage current through the oxide film significantly increases due to quantum tunneling as the thickness of the traditional oxides becomes lower than 20 Å [121]. Further miniaturization is thus restricted by the lowering device reliability and the spiking power consumption. As a result, the endeavor of ensuring large electric potential barriers with a thin oxide film has been extensively investigated, and high dielectric constant (high-κ) materials are proposed as the solution [122]. Nevertheless,

the surface deposition reactions of high-κ materials often have high activation energy barriers, thus leading to low-throughput, high operating costs and challenges for temperature sensitive material using ALD [10]. As an alternative to the high thermal energy, plasma-enhanced atomic layer deposition (PEALD) has been invented to overcome the high energy barrier, by benefiting from the high energy free radicals and ions species [11].

Despite the experimental efforts made to explore and understand the PEALD process, the high operating costs, the process complexity, and the lack of appropriate monitoring techniques restrict the exploration of a variety of operating conditions and the development of a comprehensive input-output relationship between operating parameters and film product properties. The hardware of PEALD is complicated and expensive, which involves adequate chamber design and precise gas flow controllers. In the plasma generation chamber, a complex RF power source and sophisticated pumping systems are required [60,123]. In the main reactor, the influence of the complex gas-phase particle transport phenomena on the microscopic deposition profile has not been characterized [124]. On the wafer surface, the reaction details are also not fully understood, which is crucial to the deposition rate and the film property [60]. The ability of the in-situ analysis techniques like quart crystal microbalance (QCM) and spectroscopic ellipsometry are limited [57]. Also, microstructure analysis methods like scanning tunneling microscopy (STM) and scanning electron microscope (SEM) are film destructive [12]. As a result, accurate first-principles based modeling of the plasma reaction, macroscopic particle transport, and surface deposition are beneficial to both experimental and industrial work.

In this chapter, a multiscale CFD simulation of the PEALD process with TDMAHf and O_2 plasma as precursors is developed, with the first-principles thermodynamic parameters calculated from the density functional theory (DFT) calculation. Specifically, the multiscale model combines a 2D axisymmetric plasma reaction domain, a 3D macroscopic CFD domain for gas-phase transport, and a 3D microscopic kMC domain for surface deposition. In particular, the plasma reaction domain incorporates a 2D multi-physics model that accounts for the interaction of AC circuits, magnetic field, species transport, and ionic reactions for plasma generation simulation.

Connected to the plasma boundary, a macroscopic gas-phase domain adopts a transient 3D CFD model to solve for the coupled species, momentum, energy transport, and the heavy particle reactions in the complex PEALD main reaction chamber. For the microscopic surface deposition domain, we adopt the previously developed 3D kMC algorithm by [125] to capture the surface reaction mechanism and the structure detail of the HfO_2 thin-film. DFT calculations are utilized to accurately characterize the key parameters of all domains, where thermal conductivity, viscosity, and heat capacities affect the gas-phase CFD model, and the activation energies and surface structure detail are critical to the microscopic domain. Furthermore, to implement the coupling of three respective simulation models, message passing interface (MPI) and parallel computation are used to link and solve all domains simultaneously.

4.2   PEALD Process Description and Multiscale CFD Modeling

This section addresses the construction of the 3D multiscale computational fluid dynamics (CFD) model that simulates the PEALD of HfO_2 thin-film. The workflow is shown in Figure 4.1. First, the construction of the CFD model for the ICP plasma generation in the remote plasma chamber is explained in detail. Next, the gas-phase CFD model for transport phenomena in the main ALD reactor chamber is elaborated. After that, the previously developed kMC model for the microscopic surface film deposition is briefly reviewed. Finally, the information exchange through the automated workflow between the three models and the parallel computing architecture with the Message Passing Interface (MPI) scheme is presented.

4.2.1   Plasma Chamber CFD Model

The plasma chamber design and the tuning of operating parameters critically influence the plasma chemistry as well as the final substrate deposition. As discussed in Section 4.1, the remote plasma system has been commonly adopted in the recent industrial manufacturing processes. The remote plasma reactor design enhances the performance of thermal ALD with the high-energy radicals
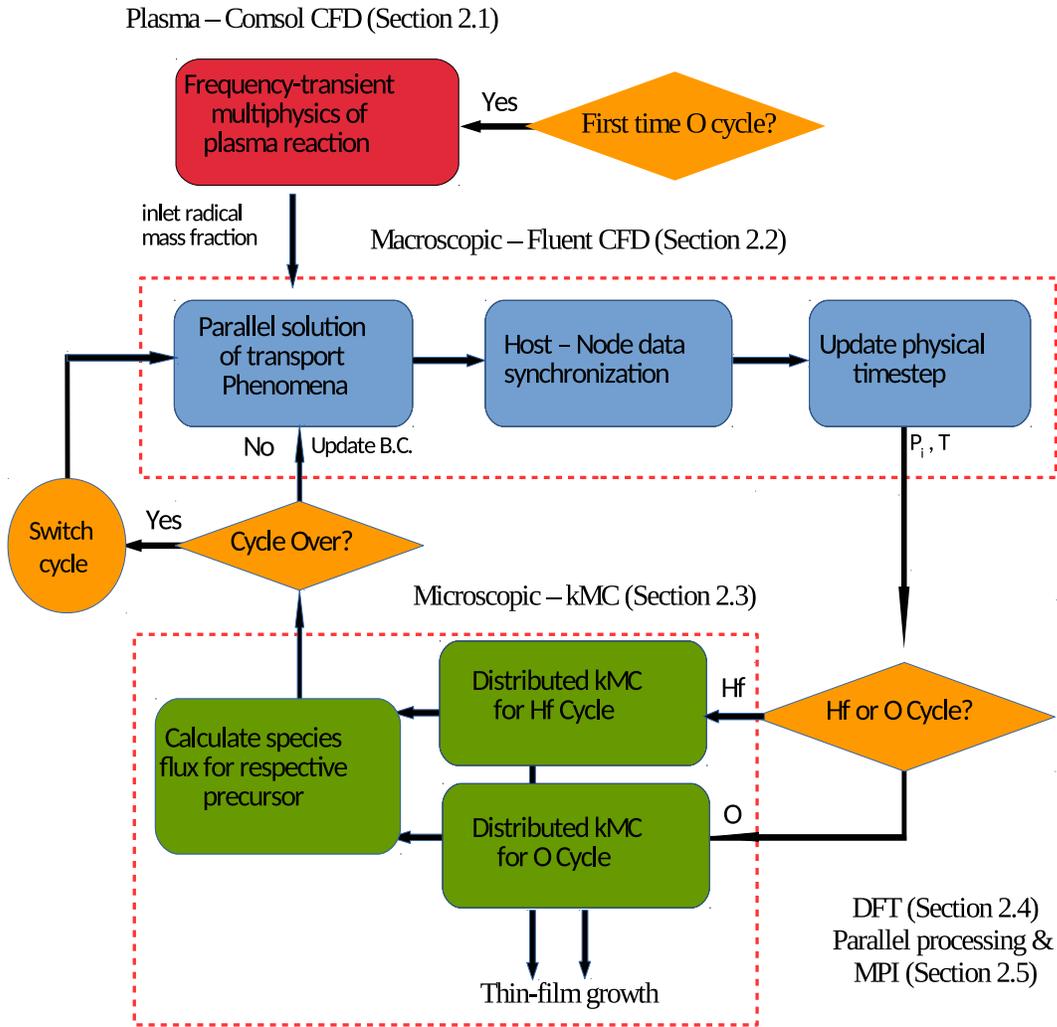
Figure 4.1: Multiscale workflow, parallel processing and information exchange illustration.

and avoids the surface damage caused by the direct ionic plasma sputtering on the substrate. Two kinds of plasma sources are often used: the capacitively coupled plasma (CCP) and the inductively coupled plasma (ICP). For PEALD application, the ICP plasma source, which generates plasma using the alternative current (AC) induced magnetic field, is often utilized due to its capability to deliver high density and high purity plasma. To simulate the coupled physics, we take advantage of the built-in plasma module in COMSOL Multiphysics software (Version 4.3). It is noteworthy that the plasma generation simulation is decoupled from the gas transport in the main reactor chamber due to the difference in model timescales. Specifically, compared to the transport timescale in

the main reactor chamber, which is typically around a couple of seconds, the timescale associated with plasma generation is extremely short, which is typically around 10^{-3} s. Therefore, it is reasonable to neglect the interaction between the plasma generation and the subsequent reactor transport. The simulation geometry we select is the Gaseous Electronics Conference (GEC) cell, which was introduced by the National Institute of Standards and Technology (NIST) at the 1988 GEC projects, aiming at providing a standard experimental and modeling platform for electronic and plasma study. The outlet is modified to fit the connection to the reactor main body, which is shown in Figure 4.2.
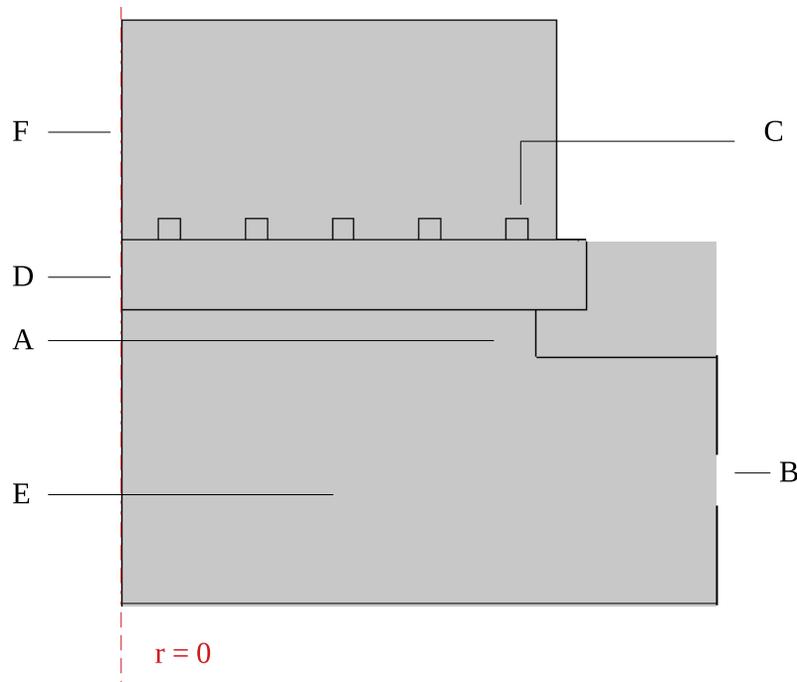
Figure 4.2: Axis-symmetric geometry of the plasma reaction chamber. The gas inlet is labeled as A, the gas outlet is labeled as B, the coil circuits are labeled as C, the circuit wall, which is made of quartz, is labeled as D, the reaction chamber is labeled as E, and the relevant ambient atmosphere region is label as F. The red dashed line on the left is the symmetry axis. Note that exit is only one side of the reactor.

For the ICP power source, the power of AC generator is set to 2000 W, and the current flows through the five fold copper coil at a radio-frequency (RF) of 13.56 MHz. The AC current creates a changing electric field $\vec{E}$ and a changing current density $\vec{J}$, which lead to the generation of the

magnetic field given by the Ampere-Maxwell equation:

$$\vec{\nabla}(\varepsilon_f \vec{B}) = \vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \qquad (4.1)$$

where ε_f is the electric permeability of materials, $\vec{B}$ is the magnetic field, ε_0 is electric permeability of free space and t is the time.

The plasma chamber is constantly filled with argon, and when the oxygen cycle is about to start, the oxygen valve is opened to introduce an oxygen flow and the AC circuit is activated. The Ar/O$_2$ plasma generation is governed by the the reaction sets shown in Figure 4.3, Figure 4.4 and Figure 4.5. Three types of reactions are defined and simulated in COMSOL: the electron impact reactions, which involve the participation of electrons, the heavy particle reactions, which involve only atomic and molecular species, and the surface reactions, which occur at the plasma chamber surfaces. The electron impact reactions are further classified into four categories including elastic, ionization, excitation, and attachment reactions, and their rates can be accurately captured by the reaction collisional cross-section data. Heavy particle reactions include charge exchange, recombination, and quenching and the rates of heavy particle reactions can be characterized by Arrhenius-type equations using appropriate rate constants. In addition, the rates of surface reactions can be computed by the sticking coefficients on the wall-surface material. The important reaction parameters, collisional cross-section data and sticking coefficients can be found in [31], [32] and COMSOL documentations [126]. The species involved in the reactions include: O(^{3}P), or O, which is the ground state atomic oxygen; O(^{1}S), which is the first excited atomic oxygen; O(^{1}D), which is the second excited atomic oxygen; O$_2$($^3\Sigma_g^-$), or O$_2$, which is the ground state molecular oxygen; O$_2$($^1\Delta_g$), which is the first excited singlet molecular oxygen; O$_2$($^1\Sigma_g^+$), which is the second excited singlet molecular oxygen; Ar, which is the ground state argon, and Ar*, which is the excited argon. For ionic species, $^-$ and $^+$ refer to the negative and positive charge respectively. Finally, e stands for the electron.

To simulate the interaction of the physical processes, we combine the simulation of fluid flow

Reaction	Formula	Type
1	$e + O_2 => e + O_2$	Elastic
2	$e + O_2 => O + O^-$	Ionization
3	$e + O_2 => e + O_2(^1\Delta_g)$	Excitation
4	$e + O_2 => e + O_2(^1\Sigma_g^+)$	Excitation
5	$e + O_2 => e + 2O$	Excitation
6	$e + O_2 => e + O + O(^1D)$	Excitation
7	$e + O_2 => e + O + O(^1S)$	Excitation
8	$e + O_2 => 2e + O_2^+$	Ionization
9	$e + O_2(^1\Delta_g) => e + O_2(^1\Delta_g)$	Elastic
10	$e + O_2(^1\Delta_g) => e + 2O$	Excitation
11	$e + O_2(^1\Delta_g) => 2e + O_2^+$	Ionization
12	$e + O_2(^1\Sigma_g^+) => e + O_2(^1\Sigma_g^+)$	Elastic
13	$e + O_2(^1\Sigma_g^+) => e + 2O$	Excitation
14	$e + O_2(^1\Sigma_g^+) => 2e + O_2^+$	Ionization
15	$e + O => e + O$	Elastic
16	$e + O => e + O(^1D)$	Excitation
17	$e + O => e + O(^1S)$	Excitation
18	$e + O => 2e + O^+$	Ionization
19	$e + O(^1D) => e + O(^1S)$	Excitation
20	$e + O(^1D) => 2e + O^+$	Ionization
21	$e + O(^1S) => 2e + O^+$	Ionization
22	$e + Ar => e + Ar$	Elastic
23	$e + Ar => e + Ar^*$	Excitation
24	$e + Ar => 2e + Ar^+$	Ionization
25	$e + Ar^* => 2e + Ar^+$	Ionization
26	$e + O_2 + O_2 => O_2 + O_2^-$	Attachment
27	$e + O_2^+ => 2O$	Attachment
28	$e + O_2^+ => O + O(^1D)$	Attachment

Figure 4.3: Electron impact reactions.

(laminar flow module), heat transfer (heat transfer in fluid), magnetic field (AC/DC module), and plasma reaction (plasma reaction) together with COMSOL Multiphysics function. A frequency-transient model is selected to investigate the plasma reaction and generation, which outputs the ion and radical compositions throughout the plasma chamber. Because of the comparatively fast development time compared to the species transport, a steady-state profile can be chosen as the constant output for the inlet profile of the subsequent domain.

75

Reaction	Formula	Type
1	$O_2(^1\Delta_g) + O_2 => O_2 + O_2$	Quenching
2	$O_2(^1\Delta_g) + O => O_2 + O$	Quenching
3	$O_2(^1\Delta_g) + O(^1D) => O_2 + O$	Quenching
4	$O_2(^1\Delta_g) + O_2(^1\Delta_g) => O_2 + O_2(^1\Sigma_g^+)$	Quenching
5	$O_2(^1\Sigma_g^+) + O_2 => O_2 + O_2$	Quenching
6	$O_2(^1\Sigma_g^+) + O_2 => O_2(^1\Delta_g) + O_2$	Quenching
7	$O_2(^1\Sigma_g^+) + O => O_2 + O$	Quenching
8	$O_2(^1\Sigma_g^+) + O => O_2 + O(^1D)$	Quenching
9	$O(^1D) + O_2 => O + O_2$	Quenching
10	$O(^1D) + O => 2O$	Quenching
11	$O(^1D) + O_2 => O_2(^1\Sigma_g^+) + O$	Quenching
12	$O(^1D) + O_2 => O_2(^1\Delta_g) + O$	Quenching
13	$O(^1S) + O_2 => O + O_2$	Quenching
14	$O(^1S) + O => 2O$	Quenching
15	$2O + O_2 => 2O_2$	Recombination
16	$3O => O + O_2$	Recombination
17	$O^+ + O_2 => O + O_2^+$	Charge exchange
18	$O^- + O => O_2 + e$	Ionization
19	$O^- + O_2 => O + O_2^-$	Charge exchange
20	$O^- + O^+ => 2O$	Recombination
21	$O^- + O_2^+ => 3O$	Recombination
22	$O^- + O_2^+ => O_2 + O$	Recombination
23	$O_2^- + O_2^+ => 2O2$	Recombination
24	$O^- + Ar^+ => O + Ar$	Recombination
25	$O_2 + Ar^* => O_2 + Ar$	Quenching
26	$O_2 + Ar^* => 2O + Ar$	Quenching
27	$O_2 + Ar^+ => O_2^+ + Ar$	Charge exchange
28	$O + Ar^* => O + Ar$	Quenching
29	$O + Ar^+ => O^+ + Ar$	Charge exchange

Figure 4.4: Heavy particle reactions.

4.2.2 Macroscopic CFD Model

The surface deposition profile of the HfO_2 thin-films is closely related to the gas-phase species transport in the main reactor chamber. In this work, we build a 3D ALD reactor model and use ANSYS Fluent 2019R3, a commercial CFD software, to solve for the coupled energy, momentum, and mass transport phenomena. The reactor geometry design, the meshing, and the CFD model

76

Reaction	Formula	Type
1	$O_2(^1\Delta_g) => O_2$	Surface
2	$O_2(^1\Sigma_g^+) => O_2$	Surface
4	$O_2^+ => O_2$	Surface
5	$O(^1D) => O$	Surface
6	$O(^1S) => O$	Surface
7	$O^+ => O$	Surface
8	$2O => O_2$	Surface
9	$Ar^+ => Ar$	Surface
10	$Ar^* => Ar$	Surface

Figure 4.5: Plasma surface reactions.

details are discussed in the following sections.

4.2.2.1 Reactor Design and Meshing

There are a variety of ALD reactor designs developed by the semiconductor manufacturing companies. In this chapter, we adopt the ALD reactor design developed in [37], which is originally based on the EmerALD XP by ASM International [60]. Specifically, the reactor chamber design is optimized to achieve a uniform gas transport profile and is scaled to accommodate the commercial 300-mm diameter wafer, which is shown in Figure 4.6(a) [67]. Precursor gases (both hafnium precursor and oxygen plasma) are introduced from the top of the reactor from a 1-cm diameter inlet with the argon carrier gas, and the gas-profile is first developed in the upstream region, which features a horn-shaped design to allow fast and uniform gas distribution. Then, the gas stream flows downward and the downstream gas-profile is tested either with or without a showerhead panel, which contains optimized hole sizes and hole distribution, as shown in Figure 4.6(b). Finally, the gas stream uniformly approaches and reacts with the substrate surface, and the unreacted precursor and carrier gas exit from the bottom outlet, which has an inner and outer diameter of 30 cm and 36 cm, respectively.

Moreover, in order to utilize the geometry for CFD computation, a fine mesh needs to be generated for the reactor geometry, where the reactor domain is dissected into geometric primitives.

77
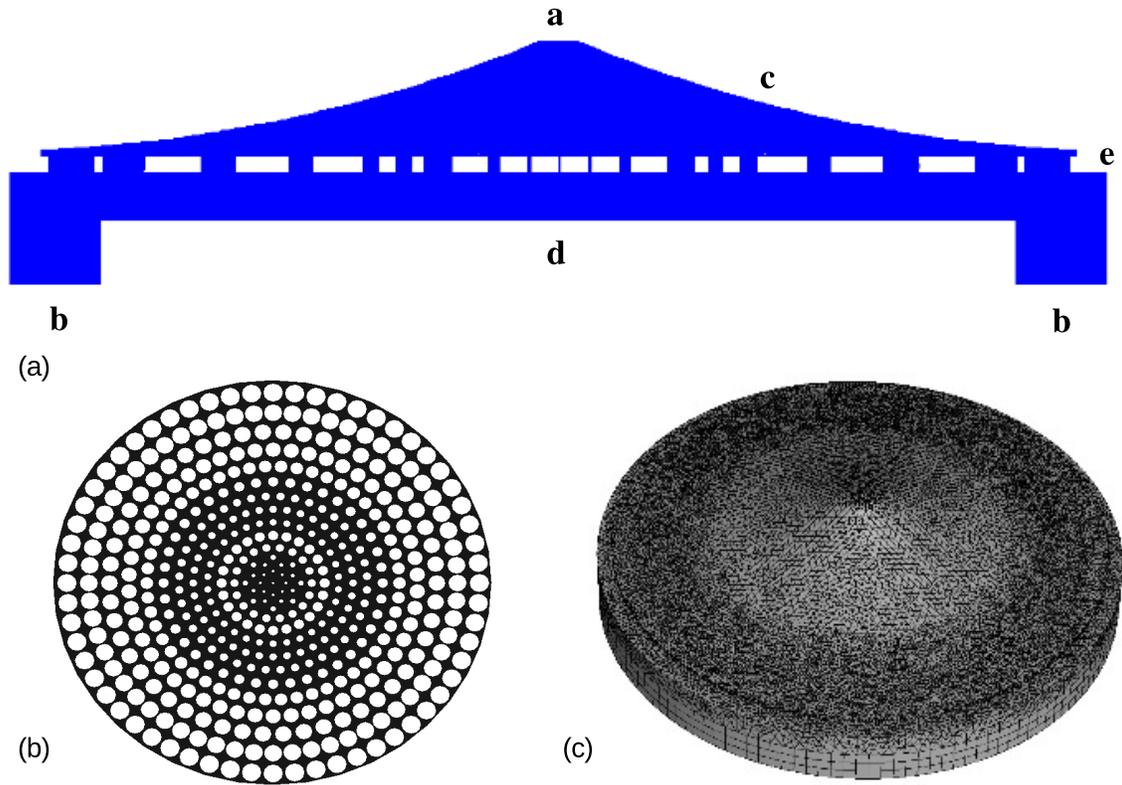
Figure 4.6: Reactor *y* cross-section view, showerhead design and hybrid mesh. The precursor/plasma inlet is labeled **a** and the outlet is labeled as **b**. The reactor upstream is labeled **c**, where the gas-profile develops. The substrate surface is labeled **d**. The showerhead panel is labeled **e**.

Adopting the hybrid meshing technique developed in [37], the reactor geometry is decomposed into two zones using structured and unstructured meshes, respectively. As a result, a good overall cell quality can be guaranteed, and the number of cells can be minimized. To balance the computation efficiency and accuracy, as shown in Figure 4.6(c), a final mesh of around 1.2 million cells is determined from the mesh independence test, where respective quality metrics are shown in Table 4.1. The results show that the quality of all cells lie within the acceptable quality range and each average metric is close to the optimal values suggested by the Fluent manual [66].

4.2.2.2 Gas-Phase Transport Model

In this section, the simulation setup in the main reactor gas chamber using ANSYS Fluent is discussed in detail. A pressure-based solver is utilized, which is appropriate for low pressure and

Table 4.1: Cell quality of the hybrid reactor mesh (∗ symbolizes the desired value).

	Acceptable Range	Bound; Average
quality	0 - 1*	0.14; 0.82
skewness	0* - 0.95	0.87; 0.24
aspect ratio	1* - 100	1.85; 1.12

near-room temperature systems. Twelve gas-phase species are incorporated in the model: argon is used as carrier and purging gas; TDMAHf is the Hf-cycle precursor; ground state molecular oxygen and atomic oxygen, together with their excited and ionic states, are the O-cycle precursors, as specified in the last section. The by-products from the surface reactions are ignored because of the limited contribution to the fluid flow from the relatively small quantities being generated. While the thermodynamics and kinetics properties of the oxygen-related species and argon can be found in the Fluent Database, TDMAHf is defined using literature and previously obtained values from density functional theory (DFT) [8, 125, 127]. The molecular weight of TDMAHf equals 354.8 g/mol, while its heat capacity, viscosity, and thermal conductivity are defined with linear temperature dependency, which will be discussed in further detail in Subsection 4.2.4. A base operating pressure is set to be 60 Pa, which is achieved with constant pumping and a flow of 300 sccm Ar. The operating pressure in each half cycle may increase because of the additional precursor in each cycle. The chosen operating pressure is within the general pressure range of 100 to 500 mTorr for typical PEALD and ALD processes [5, 128]. In addition, CFD boundary conditions are defined according to the PEALD operations from experimental works [129]. The profile of the reactor inlet is defined using several user-defined functions (UDF) in Fluent, which characterize the inlet flow rates, temperature, the duration of each half-cycle and purging steps, and species mass fluxes. In particular, the inlet is defined with the equivalent profile of a gas-delivery bubbler at 348.15 K, which is a typical temperature that leads to the vapor pressure of 1 Torr for the TDMAHf [130]. The outlet is also defined with UDF, which simulates a 12 m^3/h vacuum pump to drive the existing pressure to be 1×10^{-6} mTorr in the actual reactor outlet [10]. The substrate surface is the interface between the microscopic and macroscopic domains. UDFs are again adopted to gather gas-phase surface temperature and pressure information that is used

by the microscopic model and also to define the boundary conditions for the substrate surface, which accounts for the precursor consumption rate and surface temperature. In addition, both volumetric and surface reaction sets are defined. For most surface reactions, shown in Figure 4.5, a sticking coefficient of 1 and a high rate constant is used in the Arrhenius equation to simulate the instantaneous de-excitation and de-ionization. For atomic oxygen recombination, a sticking coefficient of 0.17, reported in [32], is simulated using the source UDF. For plasma reactions, only heavy particle reactions, as shown in Figure 4.4, are included in the gas-phase model as the number density of electrons is negligible due to the long distance from the remote plasma chamber to the main reactor chamber.

With the appropriate boundary conditions applied on the inlet, outlet and the substrate surface, reaction kinetics and the momentum, mass, and energy transport equations are solved to obtain the gas-phase velocity, concentration, and temperature profile. Surface temperature and pressure within the boundary layer can be extracted from the computed gas-phase profile, and provide information for the microscopic computation. To effectively solve the Navier-Stoke equations and other transport equations simultaneously, the coupled solver in ANSYS Fluent is used. The following flow field and reaction rate equations, shown in the tensor form, are solved in each time step [66]:

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + P)) = \nabla(k\nabla T - \Sigma h\vec{J} + (\bar{\bar{\tau}}\vec{v})) + S_h \tag{4.2}$$

$$\frac{\partial(\rho\vec{v})}{\partial t} + \nabla \cdot (\rho\vec{v}\vec{v}) = -\nabla P + \nabla \cdot (\bar{\bar{\tau}}) + \rho\vec{g} + \vec{F} \tag{4.3}$$

$$\bar{\bar{\tau}} = \mu\left[(\nabla\vec{v} + \nabla\vec{v}^T) - \frac{2}{3}\nabla \cdot \vec{v}I\right] \tag{4.4}$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla \cdot (\rho\vec{v}Y_i) = -\nabla \cdot \vec{J}_i + R_i + S_i \tag{4.5}$$

$$\vec{J}_i = -\rho D_{m,i}\nabla Y_i - D_{T,i}\frac{\nabla T}{T} \tag{4.6}$$

$$\vec{R}_i = M_{w,i}\sum_{r=1}^{N_r} R_{i,r} \tag{4.7}$$

where ρ is the density of the gas-phase species mixture, E is the fluid internal energy, $\vec{v}$ is the

velocity of gas-phase species mixture, P is the static pressure, k is the thermal conductivity, T is the fluid temperature, $\overline{\overline{\tau}}$ is the stress tensor, h is the sensible enthalpy $\overrightarrow{J}$ is the mixture diffusion flux, S_h is heat transfer source term, $\overrightarrow{g}$ is the component of gravity on the direction of the fluid flow, F is the external force, μ is the fluid viscosity, I is the unit tensor, Y_i is the mass fraction of species i, $\overrightarrow{J}$ is the diffusive flux, R and S are the reactions and mass transfer source terms, respectively, and D_m and D_T are the mass and heat diffusivities. For the total reaction rate $\vec{R}$, M_w is the molecular weight, N_R is the number of reactions involved for the species, and R_r is the individual rate of generation/consumption, which is governed by the rate constant specified in the UDFs. For all the above variables, subscript i refers to species i.

Based on the mesh developed in the previous section, the finite volume method is utilized by ANSYS Fluent to implicitly solve the aforementioned equations [68]. Specifically, at each time step, calculations for all transport equations are iterated until all residual convergence conditions satisfy the respective predefined criterion. The chosen calculation time step is also critical to the CFD simulation efficiency and accuracy [74]. According to the Courant-Friedrichs-Lewy (CFL) condition, a criterion on finite-difference approximation of PDEs, the time step size can be related to the spatial interval lengths Δx, the information traveling speed u and the Courant number C, which is the number of computational cells that the fluid information can pass through in each time step, [75]:

$$\Delta t = \frac{C \Delta x}{u} \tag{4.8}$$

Advanced CFD software allows convergence under higher courant number for faster CFD computation [75]. In particular, a courant number of 50 is used in our pressure-based solver to efficiently arrive at an accurate solution. According to Equation 4.8, the appropriate time step sizes can be determined for each half cycle and the purging steps based on the different gas flow velocities.

4.2.3　Microscopic Surface Model

In addition to the plasma chamber and the gas-phase transport through the main reactor, the reactions on the substrate surface need to be characterized. Utilizing a previously determined 3D kMC algorithm, the surface deposition model describes the detail structure of the HfO_2 thin-film and the important reaction mechanisms. The assumed 3D lattice, reaction kinetics and pathway of the surface physical and chemical reactions, and the kMC algorithm are developed in [125] and discussed in the following subsections.

4.2.3.1　Structural Characterization of the Deposited HfO_2

HfO_2 takes on a variety of crystal lattice structures, depending on the substrate temperature. At a low operating temperature of the PEALD, the monoclinic phase alike is considered in the model [131], as shown in Figure 4.7(a). For the starting surface, the (111) direction is considered, and surface treatments using piranha solution and dilute HF solution leads to a clean hydroxyl-terminated configuration, as shown in Figure 4.7(b). In order to reduce the computational and conceptual demand, 3D triangular on-lattice approximation, developed in [55], slightly changes the Hf-O-Hf Hf to 90 degrees, which successfully preserves the important structural features. An example of the resulting simulation lattice is shown in Figure 4.7(c), which contains exemplary icons of most associated species. PsHf represents the physisorbed precursor, CsHf represents the chemisorbed precursors, HfL2 represents the deposited Hf terminated with two methylamino ligands, and PsO1 and PsO2 represent two distinct physisorbed oxygens. We can see that this approximated lattice resembles the real lattice structure and gives rise to a realistic cycle repetition pattern. In order to be size-independent, the simulation lattice is determined to be $1200 \times 1200 \times$(the number of cycles simulated) [76].

4.2.3.2　Reaction Mechanism

During each half cycle, the associated precursors transported from the macroscopic reactor scale will react and deposit the desired layer of element onto the film. For the Hf half-cycle, TDMAHf

	0	1	2	3	4	5	6	7	8	9	10	11
0	O1	[ ]	O2	Hf	O1	[ ]	O2	HfL2	O1	[ ]	O2	HfL2
1	[ ]	[ ]	[ ]	O2H	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
2	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
3	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
4	O1H	[ ]	O2H	[ ]	CsHf!	[ ]	O2H	[ ]	O1H	[ ]	O2H	[ ]
5	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
6	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
7	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
8	O1H	[ ]	O2H	[ ]	PsHf!	[ ]	O2H	[ ]	O1H	[ ]	O2H	[ ]
9	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
10	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
11	[ ]	[ ]	[ ]	O1H	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	PsO1

Layer Number — ● First ● Second ● Third
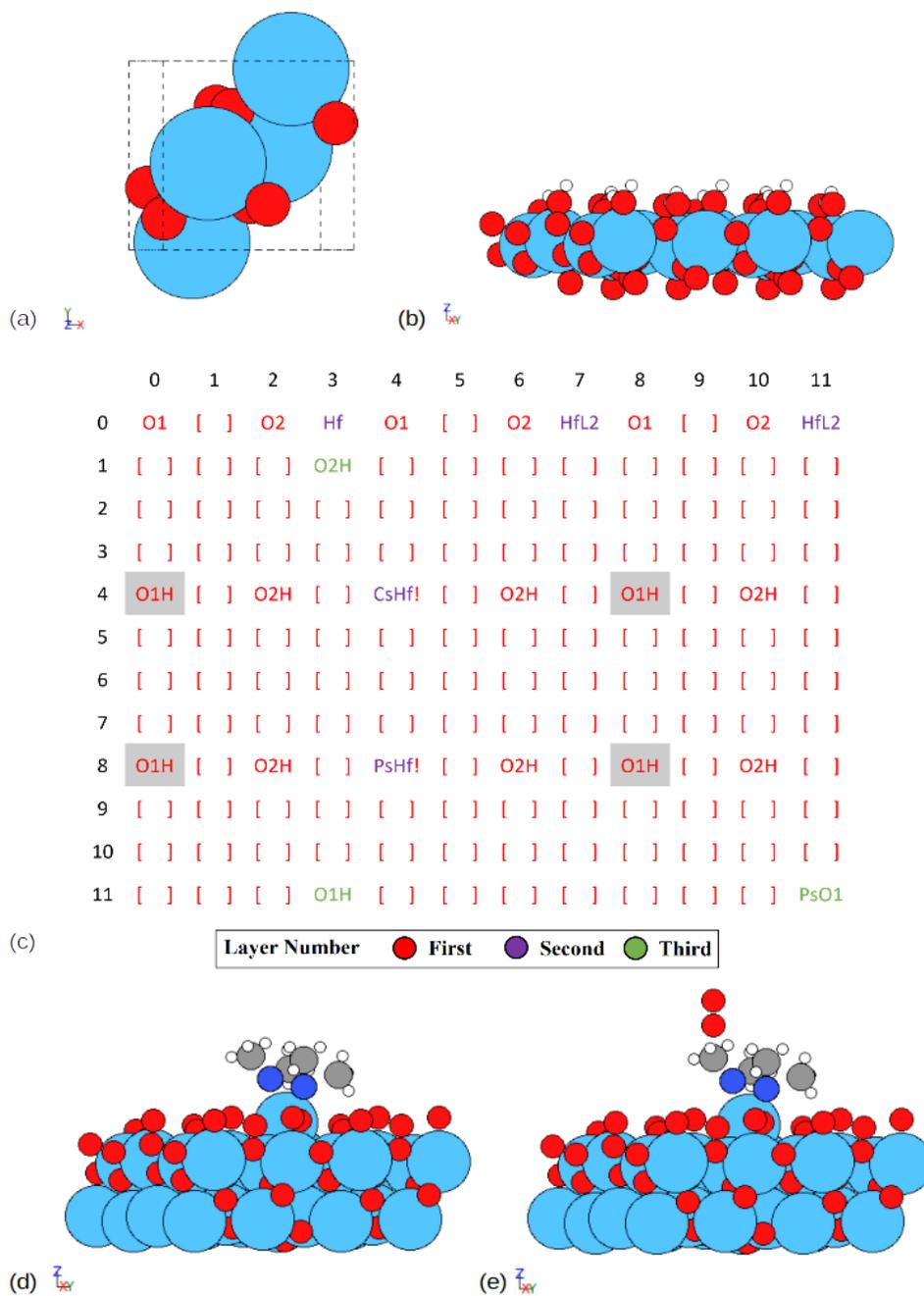
Figure 4.7: Surface slab and approximated lattice. (a) The HfO_2 unit cell. (b) The clean hydroxyl-terminated HfO_2 slab. (c) The approximated lattice used in the simulation. (d), (e) The HfL2-terminated slab with and without O_2 physisorption, respectively.

remains as the precursor, and for the O half-cycle, atomic oxygen, oxygen gas, and their radicals are the major reaction species, as demonstrated in the result of the macroscopic scale simulation. A key reaction set was selected from the detailed mechanism reported in [8] and [127], with

simplifications on the proton diffusion, ligand rotation, and other features, which dramatically reduce the computational demand, making industrial-sized wafer-scale simulation feasible while preserving the reaction mechanism fidelity. The reaction mechanisms are shown in the Figure 4.8, where the black arrows denote the reaction pathways, and the red arrows denote possible proton diffusion, where Hs denote the possible protons.

In the Hf-cycle, the reaction set contains physisorption, two-step dissociative chemisorptions, and desorption, which is shown in Figure 4.8(a). The precursor (TDMAHf) particles that are in contact with the hydroxyl-terminated surface first physisorb onto the more electronegative hydroxyl group, as shown in Figure 4.8(1). Subsequently, the precursor goes through the first chemisorption step, where the hydrogen atom from the attacked hydroxyl group is transferred to the precursor and release a dimethylamine (DMA) group, as indicated by Figure 4.8(2). Next, as indicated by Figure 4.8(3), another DMA group is released after reacting with the neighbor hydroxyl group, anchoring the Hafnium atom onto the substrate surface. The resulting slab is terminated by the remaining two DMA groups, and the surface terminal species transfer back some of the ligand protons to the neighbor sites.

In the O-Cycle, both the ground state particles and their radicals have an important contribution to the deposition rate as demonstrated in our previous work and in experimental studies [132]. The high energy of radicals, $^1O_2$ and ^3O, easily overcome the energy barrier, thereby instantaneously oxidizing the surface DMA groups and generating ethanimine compounds or the nitroxyl and ethane as byproducts as shown in Figure 4.8(4) and Figure 4.8(5). For the neutral species, $^3O_2$, the competition of the surface reactions is important to be characterized. First, the $^3O_2$ is physisorbed on to the substrate as a ring-shaped complex with the terminating DMA, as shown in Figure 4.8(6). This physisorbed complex may desorb from the surface, as shown in Figure 4.8(7), or move forward and oxidize, producing nitroxyl and ethane as byproducts in each step and eventually regenerates the OH-terminated slab, as shown in Figure 4.8(8).
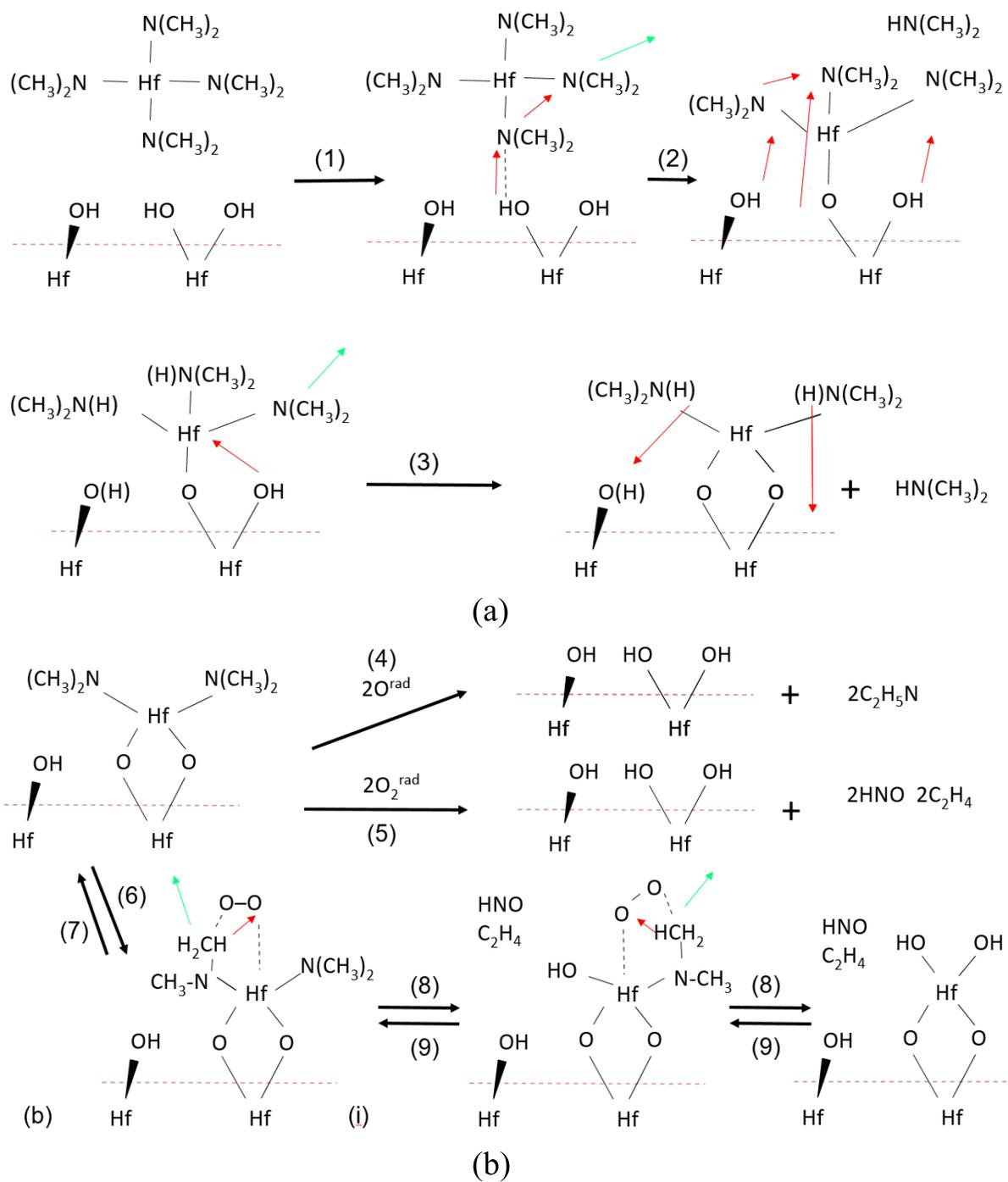
Figure 4.8: Illustration of precursor surface reaction mechanisms.

4.2.3.3 Relative Rate Determination and Kinetic Monte-Carlo Algorithm

The reaction kinetics of the pathways discussed in the previous section is crucial to the kinetic Monte-Carlo (kMC) method. Two ways of computing reaction rates are typically adopted to

characterize the reactions involved in our model. The dissociative chemisorption, oxidation and desorption are thermodynamically activated reactions, which can be described by the general Arrhenius type equation from the transition state theory (TST) as follows [79]:

$$r_{rxn} = A\exp\left(\frac{-E_a}{RT}\right)$$ (4.9)

where E_a is the activation energy of the transition state complex, A is the pre-exponential factor, R is the gas constant, and T is the substrate temperature.

In contrast, collision theory can be used to describe the gas-surface athermal barrierless processes like physisorption r_{phs}, which is governed by the the partial pressure p of the precursor:

$$r_{phs} = \frac{p}{RT}\sqrt{\frac{8RT}{\pi m}} s_c N_a \sigma$$ (4.10)

where m is the molar weight of the precursor, R is the gas constant, σ is the unit cell surface area, N_a is the Avogadro number, and s_c is the sticking coefficient, reported in [133].

The modified n-fold hybrid kMC algorithm, developed in [55], is applied to determine the reaction selection and time evolution to simulate the deposition profile and the structural growth, which is dependent on the total reaction rate r_{total}:

$$r_{total} = \sum_{i=1}^{N} r_i$$ (4.11)

where r_i represents the respective rate of each event within a N events set. For the event selection, a random number $\gamma_1 \in (0,1]$ is used to locate the event according to its normalized indicator, $l_i \in (0,1]$, which is defined as the sum of the normalized probabilities of the previous events:

$$l_i = \frac{\sum_{j=1}^{i} r_{unweighted}^{j} C_j}{\sum_{k=1}^{N} r_{unweighted}^{k} C_k}, \quad i = 1,...,N$$ (4.12)

where the reaction rate $r_{unweighted}$ will be adjusted with respect to the available reacting agents for

each event C, which is explained in detail in [125]. The i-*th* event is chosen for execution if the random number lies in between l_{i-1} to l_i.

The time evolution is calculated using a second random number $\gamma_2 \in (0, 1]$:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{4.13}$$

The event times are added to the simulation time clock until the predefined half-cycle time is reached.

4.2.4 DFT and Thermodynamic Calculations

In order to correctly implement the proposed first-principles based method, many fundamental thermodynamic properties of TDMAHf need to be captured for the microscopic kMC model and the macroscopic CFD model. Thus Density functional theory (DFT) calculations are adopted using Gaussian09 software packages and molecular visualization toolbox ASE [61, 84]. DFT is a solution to the Schrödinger equation, which calculates the electron density by approximating an exchange-correlation functional for the ground state energy of the system.

For the macroscopic model, TMDAHf is not defined in the Fluent material database, and the majority of the physical and chemical properties are not reported in previous works, including the thermal conductivity k, the specific heat C_p, and the viscosity μ. To estimate those parameters, the Hartree-Fock (HF) method is adopted with the Los Alamos National Laboratory 2-double-z (LanL2DZ) basis, which is a reasonable basis set for atoms like Hf considering the accuracy and computational speed [134]. Specific heat C_p is directly given by the DFT calculation, while viscosity μ can be calculated according to the molecular theory of Newtonian gas [70]:

$$\mu = \frac{2}{3\pi} \frac{\sqrt{\pi m k_b T}}{\pi d^2} \tag{4.14}$$

where k_b is the Boltzmann constant, m is the molecular mass, d is the rigid ball diameter of TDMAHf molecule, and T is the absolute temperature. In addition, thermal conductivity k is

calculated with the Chapman-Enskog relation [135]:

$$k = f \mu C v \tag{4.15}$$

where f is the correlation constant that has the value of 2.5.

For the microscopic model, the desorption rate of oxygen gas is an important factor that influences the neutral specie contribution to the deposition, which is investigated in our microscopic modeling investigation [125]. A single-Hf(NCH$_3$)$_2$-terminated HfO$_2$ slab is created using python Catkit package and optimized using CatGen as shown in Figure 4.7(d). The molecular oxygen gas is then positioned onto the terminating -NCH$_3$ group to mimic O$_2$ desorption, as shown in Figure 4.7(e). Adopting the transition state calculation, the resulting slab is optimized for an initial guess of the configuration and the activation energy of the desorption reaction can be further obtained using the trial-and-error method [136–138].

4.2.5 Automated Workflow and Parallelization

It is not only important to ensure the successful simulation of each scale, but also to make sure the connections between all domains are effective and accurate. The automated workflow, shown in Figure 4.1, provides an overview of how the solution of each domain is relayed in this work. The blue region represents the gas-phase domain (macroscopic), the green region represents the surface deposition domain (microscopic), the red region represents the plasma generation domain, and the arrows and the yellow region represent the cross-linkage through the message passing interface (MPI). The simulation starts at the Hafnium half-cycle (Hf-Cycle), during which the plasma generation simulation is not triggered. The macroscopic model in Fluent starts to solve for the mass and energy balances of TDMAHf and Ar for a prescribed time step Δt. The substrate surface is dissected into 36 sub-regions, and the surface temperature and partial pressure conditions are collected for each sub-region. These information on surface temperature and pressure is then transferred to the microscopic domain. Under each sub-region in the microscopic domain, a

microscopic kMC model will be triggered to perform the surface reaction simulation and determine the evolution of surface structure within the given interval Δt. After the kMC computations are completed and synchronized, the precursor consumption information will be used to calculate the species flux for each of the sub-region and reported back to the macroscopic domain. Next, system simulation clock will be advanced to $t + \Delta t$, and the boundary conditions of the macroscopic model will be updated according to calculated fluxes [66]. The above steps are repeated until the end of the Hf-cycle, and the time selection scheme will switch the conditions to the subsequent purge cycle, where residual TDMAHf can potentially continue the physisorption until purged out of the reactor. During the oxygen half-cycle (O-Cycle), before the execution of the gas-phase CFD model, the plasma generation model in COMSOL will be triggered to calculate the plasma reactions and to provide the inlet information (e.g., mass fraction plasma species and temperature) for Fluent, which is used in all subsequent O-Cycles. Also, the volumetric and surface reaction will be turned on in Fluent for the gas-phase CFD model during O-Cycles. The overall workflow will be iterated until the predefined cycle number has been arrived.

To enhance the computational efficiency, job distribution and parallelization are realized for the PEALD microscopic and macroscopic domain calculation, respectively. For the macroscopic domain, because of the high computational demand of the fine mesh and the volumetric reactions, 84 computing cores each equipped with 8 GB or 16 GB RAM on the UCLA hoffman2 cluster are used. The number of used cores is based on the resource available on the cluster, the communication overhead, and the the parallel computing benchmark. These job parallelization is managed using the Fluent Message Passing Interface (MPI) [66]. On the other hand, the local microscopic kMC models are distributed onto the idle computing cores, where 36 kMC models run simultaneously and the results are gathered and synchronized.

4.3 Multiscale CFD Simulation Results and Operating Condition Exploration

In this section, we validate the performance of the constructed CFD model by first comparing each component of the multiscale model (the plasma generation, the main chamber gas-phase transport, and the microscopic deposition) with experimental observations and individual scale simulations. Then, the collective influence of plasma generation and gas-phase transport on the deposition of the substrate will be investigated.

4.3.1 Plasma Generation

The plasma generation model in the RF excited remote plasma chamber is tested with an inlet volumetric flow rate of 100 sccm of Ar and O_2 mixture, in which Ar has a mole fraction of 0.3 and O_2 has a mole fraction of 0.7. The coil has a power of 2000 W and the plasma chamber pressure is set to be 2.67 Pa. Using a frequency-transient simulation, steady-state can be reached at 3×10^{-4} s, which is demonstrated to be much faster than the gas-phase transport as discussed in the Section 4.2.1. The spatial distributions of the electron and species densities at steady-state are shown in Figure 4.9.

The power distribution inside the plasma chamber is demonstrated to have its maximum value close to the coils. The resulting electron number density is in the range of 10^{17} $1 \cdot m^{-3}$, which is close to the value reported in [32] and [139] at similar operating pressure and power, and the exact values differ because of the chamber geometry and coil layout. Also, as shown in Figure 4.9(b), we can see that the maximum electron density lies at the center of the generation chamber as expected. Due to the larger reactor geometry and the horizontally distributed coils simulated in this chapter, the power used in this model is higher than in some other research works. For example, a power of 1000 W, leading to a capacitive power deposition of 0.1 - 1 W/m^3 was reported by [139]. Whereas in our model, a power of 2000 W is used, leading to a capacitive power deposition of 1.6 W/m^3. It is demonstrated that at higher power, the gas heating is more significant, which leads to a higher
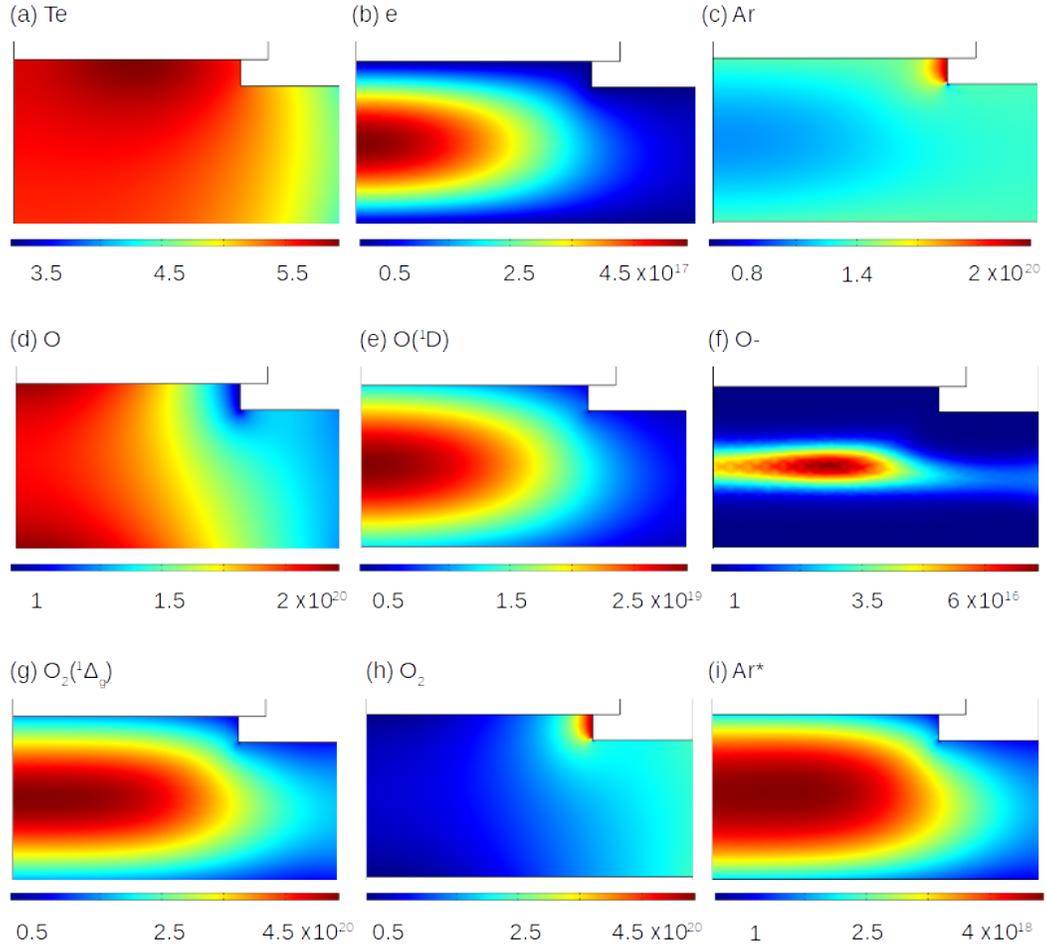
Figure 4.9: Electron temperature and number densities of important plasma species.

electron temperature, T_e. On the other hand, the electron temperature decreases with a higher chamber pressure as observed in [139]. Thus, with a reasonable extrapolation based on the trend of power, we can estimate that the expected electron temperature lies in the range of 2 - 5 V, which successfully correlates the result observed in our simulation model, as shown in Figure 4.9(a).

The number densities of O_2 and its excited species are shown in Figure 4.9(h) and (g). For O_2, the maximum number density is at the inlet and the majority of O_2 are excited at the center of the reactor to the activated species: $O_2(^1\Delta_g)$ and $O_2(^1\Sigma_g^+)$. Away from the center, O_2 reappears from the recombination of atomic oxygen and excited oxygen radicals. Due to the additional surface recombination at the chamber wall, more O_2 can be found near the wall. The number density of the ground state O_2 and its excited derivatives are demonstrated to be at the same order of

magnitude as in shown [32] near the center plasma generation region.

The number densities of atomic oxygen-derived species (O, O(^{1}D), and O$^-$) are shown in Figure 4.9(d) and (e) and (f). The ground state atomic oxygen exists all the way to the outlet but its ionic and excited derivatives are contained at the reactor inlet. The high power adopted in this model enhances the ionization process as well as the conversion from O_2 to O and its radicals. However, because of the high activation energy of the atomic oxygen radicals, which are twice as big as the molecular oxygen radicals, they are more prone to return to ground states. Also, the ions are quickly de-ionized as soon as they enter the region with low electron density. This quick drop in the number densities of atomic oxygen ions and radicals has also been observed in [32].

It can be seen that the final product of the plasma generation domain is composed mostly of neutral atomic oxygen radicals, a small fraction of charged species, and a reasonable combination of ground state molecular oxygen and other excited species. The final molar fractions at the outlet region are: Ar, 2.80×10^{-1}; O_2, 5.39×10^{-1}; O, 1.65×10^{-1}; O(^{1}D), 8.00×10^{-3}; Ar*, 2.74×10^{-3}; $O_2(^1\Delta_g)$, 2.67×10^{-3}; $O_2(^1\Sigma_g^+)$, 6.36×10^{-4}; O_2^+, 3.46×10^{-5}; O$^+$, 2.86×10^{-5} and O$^-$, 2.44×10^{-5}. After the plasma stream is introduced into the main reactor chamber, further recombination reactions will take place to reduce the plasma energy and minimize the impact of the charged species.

4.3.2   Main Reactor Gas Transport

The gas-phase CFD model is tested over a wide range of precursor inputs, and it is demonstrated to successfully capture the macroscopic transport phenomena and reactions happening in the reactor chamber. The maximum Knudsen number in the reactor chamber under all operating conditions is around 0.05, which is much smaller than 0.1. Therefore, the continuum hypothesis is guaranteed at all parts of the reactor, and the Navier-Stokes equation can be safely used to solve for the fluid flow profile. Moreover, the maximum local Reynolds number is less than 200 and the fluid flow is entirely laminar. Therefore, turbulent modeling is not required in this simulation [95].

First, we will investigate the reaction and transport in the main reactor chamber during the

92
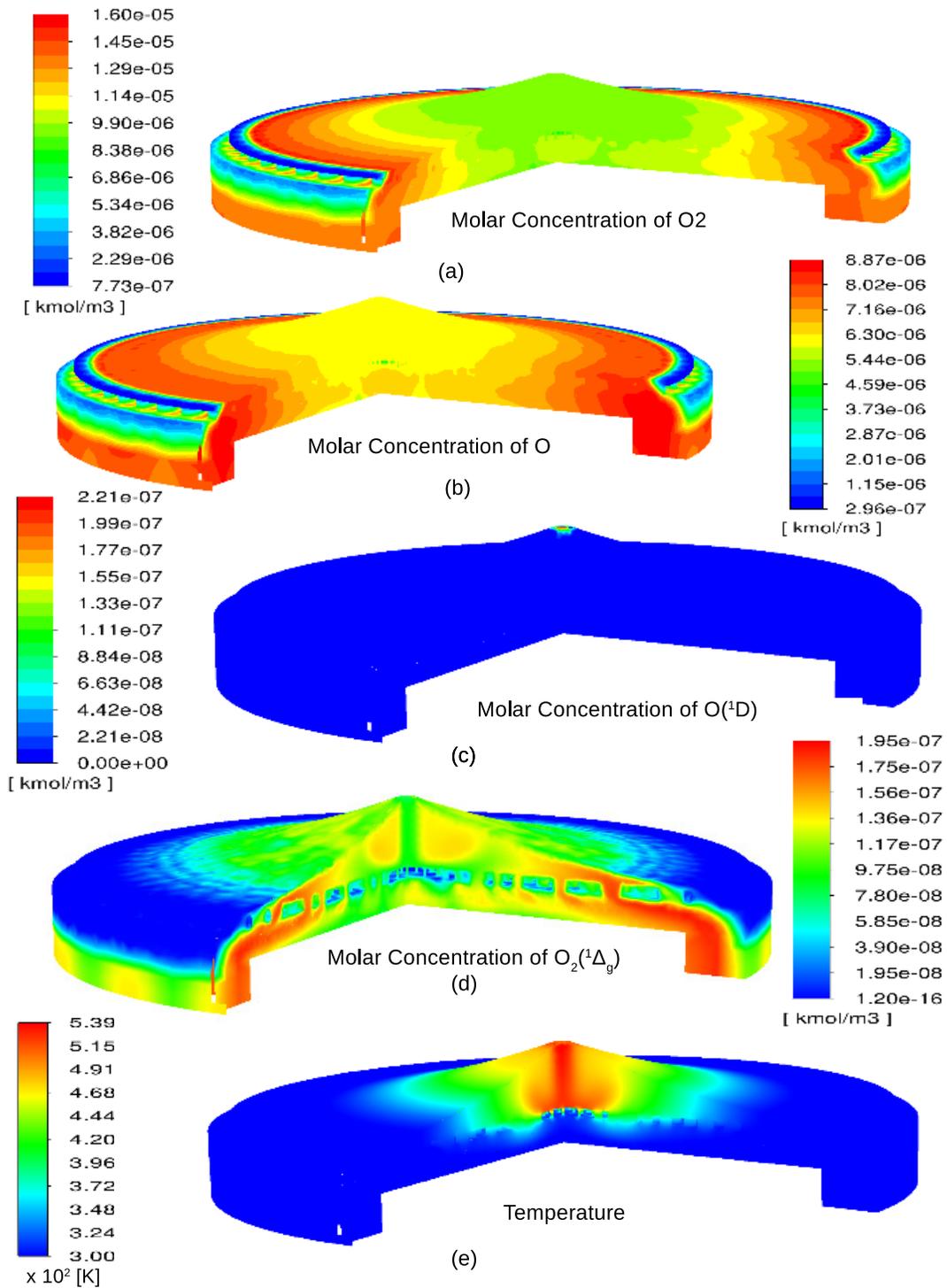
Figure 4.10: Molar concentration of important species in the gas-phase CFD.

O-cycle, given a total of 450 sccm input from the plasma generation in COMSOL. The molar concentrations (number densities) of the important species in the reactor geometry are shown

in Figure 4.10, and the results are compared with respect to the established simulation models for plasma generation [32]. As shown in Figure 4.10(a), the oxygen gas molar concentration at the substrate surface and the outlet is about 1.1 to 2 times of the corresponding concentration at the inlet. This increase in the concentration is majorly contributed by two types of reactions: the de-excitation/de-ionization reactions of radicals and ions, and the gas-phase and surface recombination reactions of atomic oxygen. We can see from Figure 4.10(c) that, for the radical species, $O_2(^1\Delta_g)$, its concentration remains fairly constant across the reactor despite its participation in the reactions that regenerate O_2, which mostly happened before crossing the showerhead. This is due to the fact that $O_2(^1\Delta_g)$ can be replenished from $O_2(^1\Sigma_g^+)$ in the downstream. On the other hand, the concentration of the other radical, $O_2(^1\Sigma_g^+)$, decreases rather fast and vanishes before the showerhead. Similarly, due to the high energy associated, the concentrations of positive and negative ions decrease in a similar trend to $O_2(^1\Sigma_g^+)$. Thus, nearly no ionic species is observed near the substrate.

As shown in Figure 4.10(b), the atomic oxygen has a slight increase in concentration throughout the reactor. The rates of de-excitation and de-ionization of the atomic oxygen radicals and ions are extremely fast and cause those species to vanish right after entering the inlet. In particular, the singlet atomic oxygen species, according to the [32], recombine at a high rate in the main reactor gas transport, and will thus not be considered for the surface reaction. Meanwhile, a lot of ground state atomic oxygen is recombined into oxygen gas when it is in contact with the showerhead and the upstream reactor walls, which can be seen from the light regions and the sudden decrease of the concentration profile across the showerhead panel. However, it is demonstrated that not as much atomic oxygen is recombined to form oxygen gas as in [32]. This is because the reactor pressure is higher, given the larger precursor flow and the compact showerhead reactor geometry, which leads to a smaller recombination coefficient. Still, the ratio of number density between the upstream and the top of the wafer are in reported ranges.

In addition, it can be seen in Figure 4.10(e) that the region above the middle of the showerhead has a noticeably higher temperature than the rest of the reactor. This is caused by the huge

extent of reaction in that region. In particular, charged particles like O^-, O^+, and O_2^+ are the major contributors to the heating due to their participation in the de-ionization reactions and the high-velocity kinetic impact, as demonstrated in [10]. As the gas species travel across the showerhead, the heat is less prone to transport to the downstream and is spread to the side of the reactor. Thus, the temperature on the substrate stays around 300 K, which is optimal for the temperature-sensitive materials and is much less likely for significant precursor decomposition to happen, which could lead to undesirable multiple-layer deposition reactions and surface island formation.

To further demonstrate the necessity of the showerhead design, we compare the gas-phase profile inside the reactor with and without a showerhead. For the reactor without a showerhead, the physics is axisymmetric and can be represented by a 2D axisymmetric model. The comparison of the velocity profile between the two models is shown in Figure 4.11(a) and (b). Both velocity profile demonstrate that the middle of the reactor has a very high velocity under a high inlet flow rate of 450 sccm of argon and oxygen mixture. Without the showerhead, the velocity can still reach around 40 m/s at the substrate surface. In most applications, this high velocity is benign because the amoutn of ionic species that reach substrate is substantially reduced. However, it has been demonstrated that in the PEALD of HfO_2 on graphene, defects due to ion species can still be observed even when remote oxygen plasma is used [140]. The observed high velocity in the center of the reactor may still lead to ion bombardment energy that is destructive to sensitive surfaces. Showerhead design is a solution to this problem. Even when the velocity at the inlet region is on the scale of a few hundred meters per second, the showerhead successfully blocks the fast traveling vertical gas flow in the center with the small showerhead holes, as shown in Figure 4.11(b). Therefore, by having a showerhead panel, film damage through plasma distance and high inlet flow rate is minimized even for sensitive substrate surfaces. It is worth mentioning that, the inclusion of the showerhead indeed slows down the vertical transport speed of the inlet species to the substrate surface, as shown in Figure 4.11(c) and (d). Nevertheless, by optimizing the design of the showerhead panel, as shown in Figure 4.6(d), the radial flow of the precursor
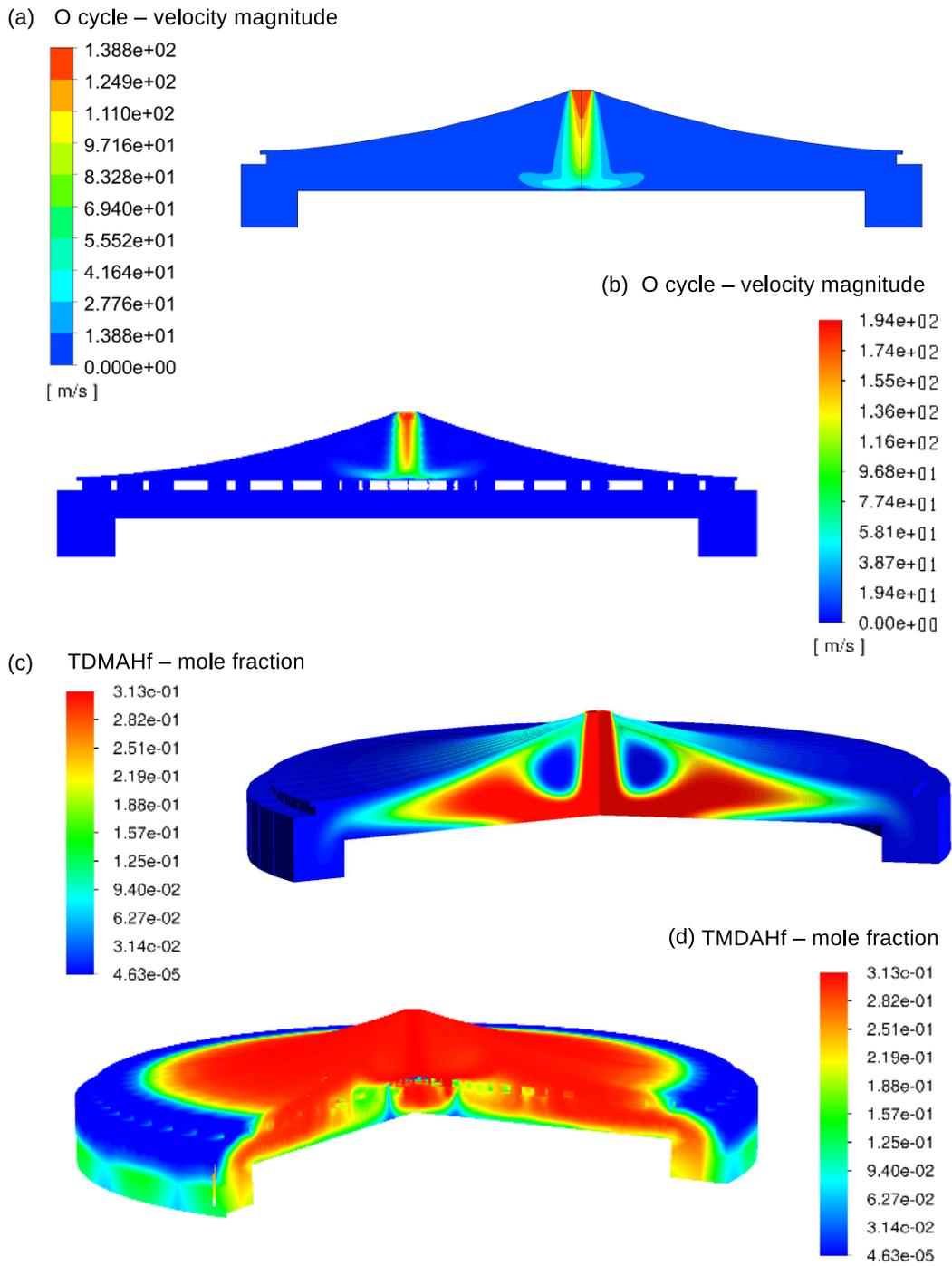
Figure 4.11: Comparison of velocity profile and species transport between (a), (b) 2D and (c), (d) 3D models.

is dramatically increased by assigning increasing diameter for the showerhead holes in the radial direction. Thus, with this enhanced guidance of gas flow towards the peripheral of the wafer, a much more uniform precursor distribution can be achieved on the substrate surface. Therefore, in order to capture the correct gas-phase profile, a full 3D model is helpful due to the showerhead design. However, the 2D axisymmetric model is less computationally demanding and can be used for an initial estimate for a wide range of operating conditions. In addition, experimental works, such as [10], have demonstrated the strong dependency of thickness non-uniformity on the plasma pressure and chamber size. In our model, a similar maximum surface temperature around 500 K is observed under the corresponding plasma generation pressure, which validates the neutral heat transfer and the charged particle diffusion. However, we can see that the showerhead really endeavors to minimize the influence of the plasma operating parameters to the film uniformity. Thus, it allows more freedom for the plasma generator to be tailored for number density without the need to worry about the uniformity and temperature negative effect.

4.3.2.1 Surface Deposition

The microscopic simulation of the surface deposition has been elaborated in [125], where various operating parameters (operating temperature, precursor partial pressure, and etc.) are evaluated for their overall effects on the deposition rate by assuming a fixed surface profile for operating conditions. Under this assumption, a saturated GPC of around 1.17 Å per cycle is found to be achieved for the deposition of HfO_2 given sufficient precursor dosages at all time. Moreover, a range of different fixed operating conditions has been tested to obtain the critical precursor dosage that ensures both saturated GPC and low precursor waste. Despite the insights provided by the previous work on the overall trend, in a realistic deposition process, the precursor partial pressures and temperature are changing and are heavily influenced by the gas-phase species transport. Therefore, in this section, we look at the effect of time-dependent gas-phase profile on the microscopic surface deposition and the further validation of the deposition model based-on realistic gas-phase transport.

Studies were performed to derive the trend of pressure development using the 2D axisymmetric model, which is used to test the microscopic kMC model performance. Since the effect of steric hindrance is established in the Hf-Cycle, the preliminary studies focus on Hf-Cycle. The influence of partial pressure development on the fraction of sites that are physisorbed is demonstrated in Figure 4.12. Specifically, under a high inlet flow rate of 300 sccm of TDMAHf and Ar, the precursor partial pressure saturates around 2.5 s, and the surface physisorption can finish at 2.275 s. By contrast, under a low inlet flow rate of 150 sccm TDMAHf and Ar, the precursor partial pressure does not achieve full saturation until the end of Hf-cycle, and the surface physisorption takes around 3 s. In addition, the temperature influence on the deposition rate is similar to what has been learned in [125]. Moreover, our studies demonstrate that given sufficient precursor dosages, the surfaces are guaranteed to reach saturation coverage (limited by the steric hindrance), and under those operating conditions, a linear growth profile can be observed for multi-cycle deposition as reported in the experimental works.

4.3.2.2   Integrated ALD Chamber Design and Evaluation

After each simulation scale is validated, all model components are integrated for the final multiscale analysis as indicated in Section 4.2.5. Based on literature references and preliminary calculations from individual models, an operating baseline is tested as shown in Figure 4.14, which consists of a 5 s TDMAHf pulse at 87.25 Pa, followed by a 5 s oxygen plasma pulse at a pressure of 90 Pa, while each precursor pulse is separated by 2 s of purging and pumping down to a base pressure of 60 Pa. The base operating pressure is achieved with a constant 300 sccm flow of argon carrier gas, and the effective precursor molar flow rate for Hf-Cycle and O-Cycle is 137 sccm and 150 sccm, respectively. Unlike the experimental GPC measurement using ellipsometry and X-ray photoelectron spectroscopy (XPS), the simulation model does not directly report the height of the deposited thin-film because of the lattice approximation adopted. However, GPC can be calculated using the coverage and lattice information. A GPC of 1.17 Å/Cycle can be achieved given enough precursor dosage, which is similar to the experimental
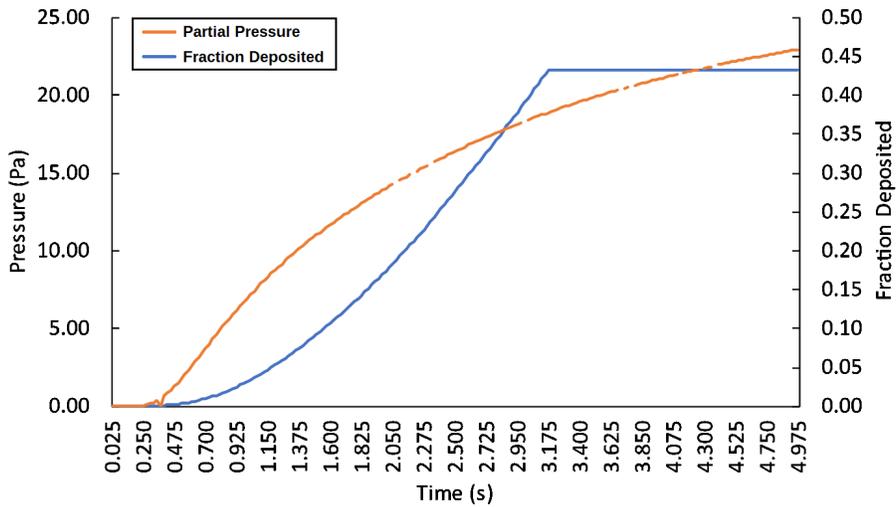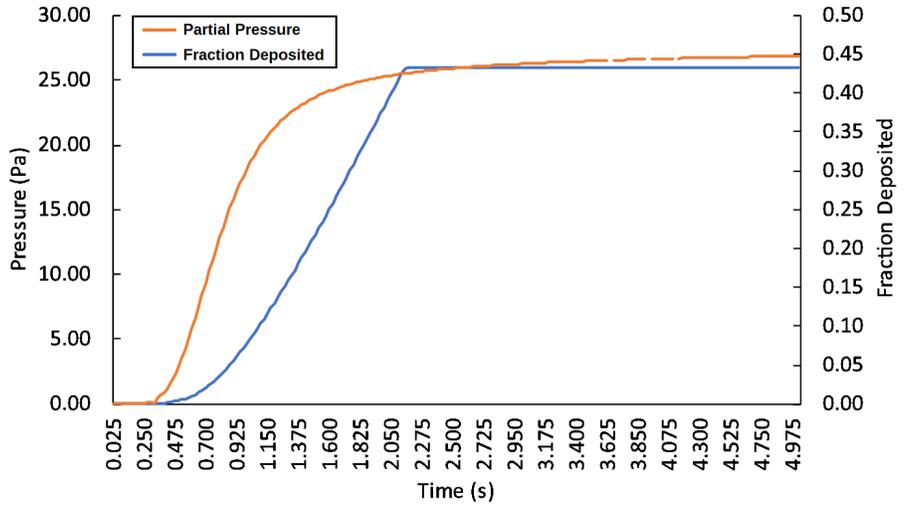
Figure 4.12: Pressure development effect on Hf-cycle physisorption.

results with comparable operating conditions [128, 141, 142]. Also, the simulation model shows a linear growth for multi-cycle deposition, as observed in the experimental works by [14] and [132]. A summary of the assumptions and the general results of the developed multiscale CFD model is included in Figure 4.13.

In addition to the overall growth, it is important to investigate the spatial difference in the deposition profile. Non-uniformity is often encountered in large wafer production, but the lack of low-cost and straightforward in-situ monitoring techniques makes it difficult to find out the

Assumptions	
Plasma Domain	Decoupled completely from other domains
Main-Chamber Domain	Incompressible ideal gas in continuum regime
Surface Domain	Depends only on surface pressure, temperature and configuration

Results	
Plasma Domain	Steady-states are achieved in a short timescale ($\sim 10^{-3}$ s)
Main-Chamber Domain	Realistic profiles are obtained under typical feed flow rates (~ 150 sccm)
Surface Domain	An overall growth per cycle (GPC) of 1.17 Å/Cycle is observed

Figure 4.13: Summary of the assumptions and results of each domain in the multiscale CFD model.
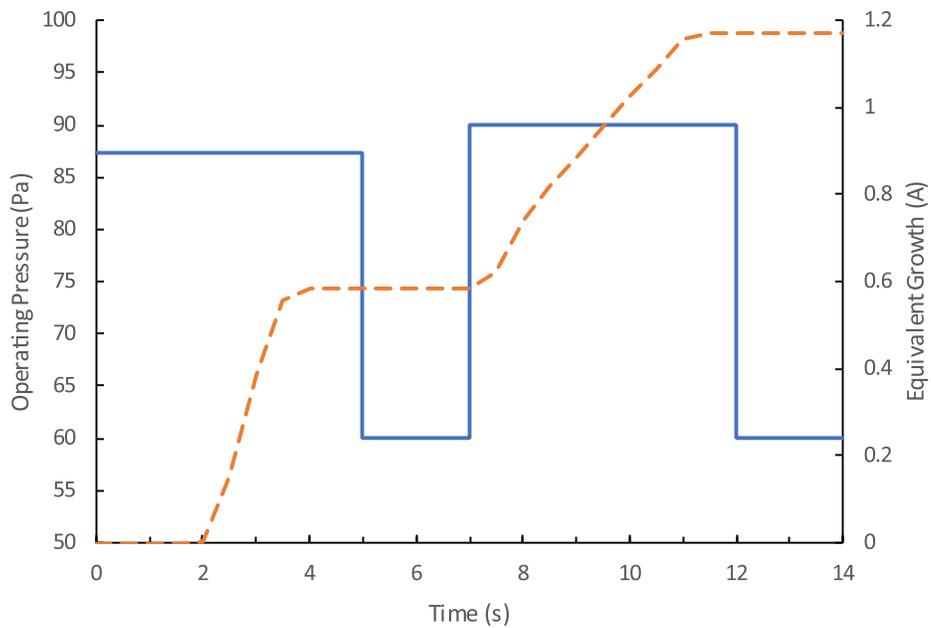
Figure 4.14: Operating baseline and average deposition rate. *x*-axis is time, left *y*-axis is the operating pressure and right *y*-axis is the effective average growth over the entire substrate surface. The orange dashed line represents the change in the effective average thickness and the solid blue line represent the change in operating pressure.

problem during the manufaturing process. With the multiscale CFD model, we can analyze the gas-phase flow profile and the surface deposition profile to reduce non-uniformity through reactor geometry optimization and operating parameter adjustment. The major non-uniformity often occurs along the radial direction. Therefore, the analysis on the entire substrate is divided into three zones according to their radius: central region (0 to 50 mm), middle region (50 to 100 mm), and peripheral region (100 to 150 mm). The deposition development profiles at those three regions

are shown in Figure 4.15. With the aid of the showerhead and the horn-shaped upstream, the flow of the precursors is guided in the radial direction, the transient film non-uniformity is substantially reduced, and the final influence of this non-uniformity is eliminated within the prescribed cycle time for both cycles.
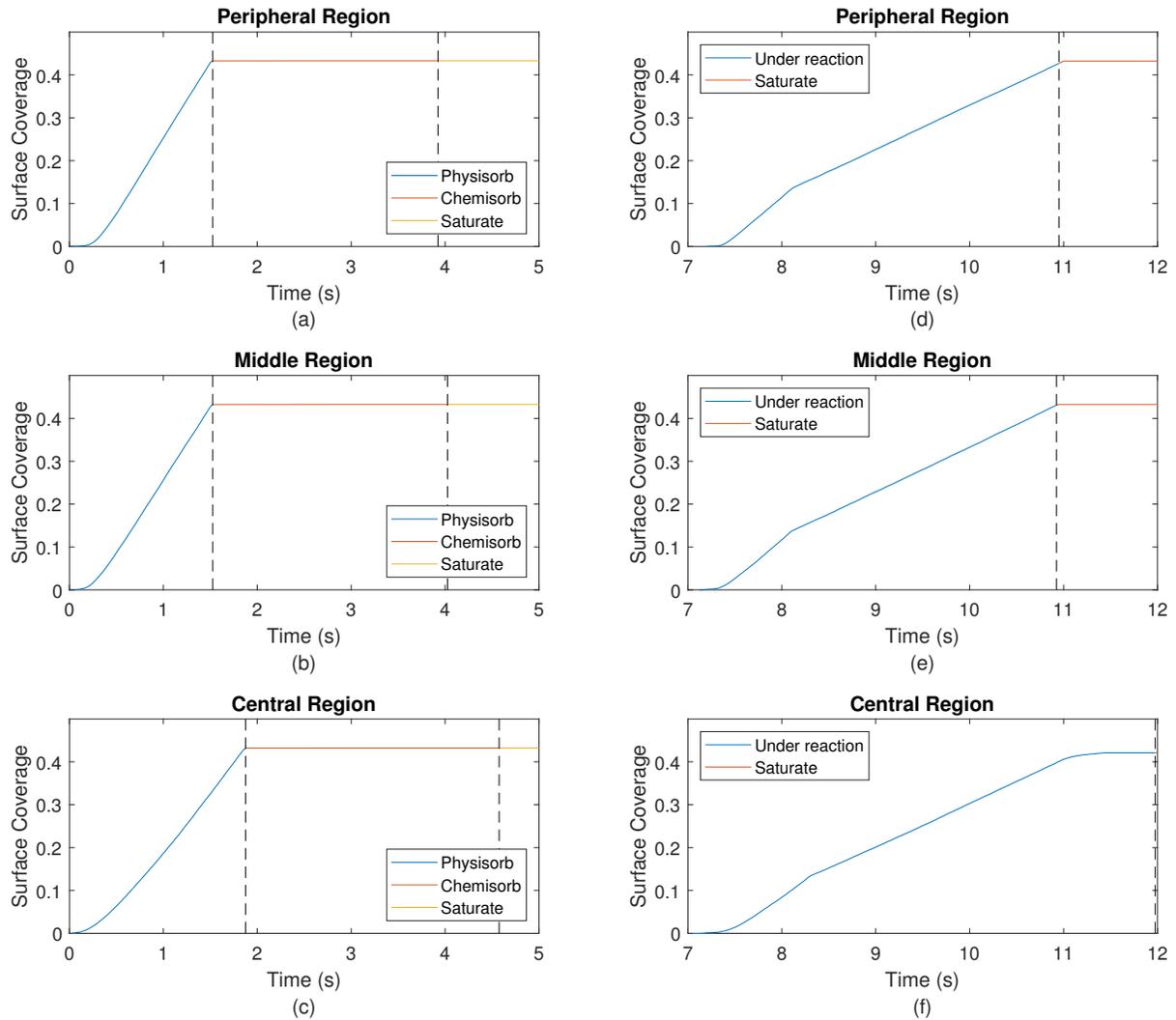
Figure 4.15: (a), (b), (c) Hf-Cycle deposition profiles of the peripheral, middle, and central regions of the substrate, respectively. The blue line represents the time duration for TDMAHf physisorption, the red line represents the time duration for TDMAHf chemisorption and the orange line represents the time duration after the surface is saturated. (d), (e), (f) O-Cycle deposition profiles of the peripheral, middle, and central regions of the substrate. The blue line represents the time duration for O-Cycle deposition reaction and the red line represents the time duration after the surface is saturated.

Utilizing the multiscale simulation model, we have investigated a potential way of reducing the

required cycle time by effectively utilizing initial period of the purge step for deposition. Purge step has always been solely used to clear out the remaining precursor, assuming that surface deposition is finished within each precursor half-cycle time. This is because the amount of precursors remaining on the wafer during purging cannot be accurately measured and the resulted contribution to the deposition is not well understood. Nevertheless, the developed multiscale CFD model is able to capture the exact amount of precursor existing on the surface and the associated rate of surface reactions. From simulation results, we observe that the surface partial pressure of both precursors could retain a decent level at the initial time steps, and the middle region, which is demonstrated to have the slowest deposition rate, has even longer residual time. Therefore, these remaining precursor will still lead to a high enough physisorption rate, thus maintaining the equilibrium for the subsequent forward chemisorptions. A case study on O-Cycle is performed using the baseline operating temperature and pressure, with a reduced O-Cycle time of 4 s. The overall surface deposition at the end of O-Cycle is found to be 97.4% of the saturation coverage, and the center has a lower value of 87%. After we shift to the purging, the overall wafer is found to retain 50% remaining precursors at 0.2 s, and the center is found to retain 42 % remaining precursors at 0.4 s, which will continue to contribute to the surface physisorption. At the end of the purge step, all precursors are cleared out and the entire surface reaches full coverage. Thereby, the O-cycle time can be reduced by 20 %, which is also observed in Hf-cycle. As a result, it is demonstrated that the remaining precursor is able to provide a significant contribution to the deposition in both Hf-cycle and O-cycle purging.

4.4   Conclusion

In this chapter, we have constructed a multiscale CFD model for the plasma-enhanced atomic layer deposition of HfO$_2$ using TDMAHf and oxygen plasma. The remote plasma generation model and the macroscopic gas-phase CFD model are constructed and coupled with the previously developed microscopic kMC model through an automated workflow. Both macroscopic domains were

individually validated against experimental and existing single scale simulation data. Specifically, the detailed plasma profile was analyzed in both the plasma domain and the gas-phase domain, in which the species transport and plasma reactions across the reactor chamber were correctly captured. The multiscale model was validated and utilized to identify the effects of plasma generation power, operating pressure, operating temperature, and inlet flow rates on the deposition of HfO_2 thin-films. The model output was demonstrated to be consistent with the overall trend reported from literature and industrial specifications. Furthermore, using the multiscale CFD model, a set of baseline operating parameters were determined to produce highly uniform HfO_2 thin-film within conventional industrial half-cycle time, and the transient non-uniformity across the entire substrate was successfully reduced under the baseline operating conditions. In addition, the model investigated the possibility of reducing the total operation time by at least 14.3% from efficiently taking advantage of the remaining precursors at the beginning of the purge steps. This work provides a general-purpose tool to explore the optimal operating conditions under which a thin-film can be produced with guaranteed quality with a significantly reduced process operation time. Using this model, a further process optimization strategy can be determined, thereby greatly increasing production throughput.

Chapter 5

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

5.1 Introduction

The stringent quality requirement of PEALD demands the operating conditions to be consistent throughout the entire deposition process. Therefore, disturbances to the reactor operations must be effectively evaluated in various deposition processes to ensure the final film quality. Traditionally, feedback control systems have been used to eliminate the effect of disturbances in real-time. The performance of the feedback control has been extensively proved in various processes, including the application of plasma enhanced chemical vapor deposition (PECVD) [54], which is very similar to PEALD. In the feedback control system the measurement of the control variable is constantly fed back to the controller. The controller then compares the deviation of the measured value with the set point value and sets it as the process error. To offset the error, the controller changes the

manipulated variable through an actuator, such as pressure regulating valves or flow control valves.

Three types of feedback controllers are commonly used for the real-time closed-loop system: the proportional (P) controller, the proportional-integral (PI) controller, and the proportional-integral-derivative (PID) controller. The integral (I) mode is essential to ultimately eliminate the residual error. In the present work, PI control is used instead of PID control, as PI control allows for a desired closed-loop response without offset while PID control may lead to sensitivity to sensor noise due to the derivative term. Specifically, the PI controller adjusts the inlet mass flow rate to maintain the desired partial pressure of the TDMAHf precursor. A typical pressure sensor that can be used for the feedback control system considered in this chapter is AST20HA pressure transducer manufactured by American Sensor Technologies, Inc., and a common choice for the flow rate actuator is the SmartTrak50 mass flow controller from Sierra Instruments. Additionally, the real-time measurement of the molar fraction can be achieved by using the HPR-20 real-time GC gas analysis system manufactured by Hiden Analytics, of which the result can be used to compute the partial pressure of the precursor.

Nevertheless, due to the limitation on control actions and the relatively short process time-scale of the PEALD, a feedback controller alone, even an advanced controller like a model predictive controller (MPC), may not be able to fully counter the effect of disturbances. As a result, similar to [116], a run-to-run (R2R) controller is used along side the PI controller to adjust process parameters across multiple batches that can not be adjusted in real-time. Specifically, the R2R controller adjusts the TDMAHf molar fraction of the inlet stream to ensure the full coverage through the mixing ratio of carrier gas and precursor flow rate. The adjustment is performed based on a empirically determined input-output relationship between the inlet TDMAHf molar fraction and the final film coverage, and the control action can be also actuated through the aforementioned mass flow controller. Using the developed multiscale CFD model, the parameters of the PI controller and the R2R controller are identified and tuned, and both controllers are integrated to ensure the full coverage on the wafer under disturbances.

5.2 Control Scheme

The feedback controller-only results and the integrated feedback-R2R controller results are presented in the following sections.

5.2.1 Disturbances and Open-Loop Simulation Results

The first disturbance considered in this chapter is the variation in the operating pressure. The PEALD process is typically operated near vacuum and the surface reaction rate is very sensitive to pressure change. Therefore, even a small deviation in the operating pressure can cause a significant effect on the wafer coverage. Such deviation can be induced by the malfunctioning of the vacuum pump. In the scenario where the vacuum pump experiences a power surge, a sudden increment in pump power could result in the reduction of the operating pressure. In this chapter, a case study is performed when the operating pressure of 100 Pa, which is the standard operating condition, is reduced to 30 Pa.

Another possible disturbance involves the temperature of the gas feeder, which is usually in the form of a bubbler. The temperature variation inside the bubbler can cause a considerable fluctuation of the TDMAHf molar fraction due to the rapid change in the precursor vapor pressure, which is highly dependent on temperature. A standard operating temperature of 340 K is chosen for the bubbler. In the case study, a disturbance of 10 K drop from the standard operating temperature is explored, which can be commonly seen in the bubbler application [117]. In addition, it is assumed that the inlet stream temperature is still maintained at 340 K through additional heating devices, which is not characterized in this model, so that the bubbler temperature disturbance does not change the thermal environment of the microscopic surface reactions. The open-loop simulation results are shown in Figure 5.1, where the impact of the two disturbances can be seen. The coverage of the wafer under the pressure disturbance and the temperature disturbance are 62% and 0.7%, respectively, when the PEALD process is operated without any control action.
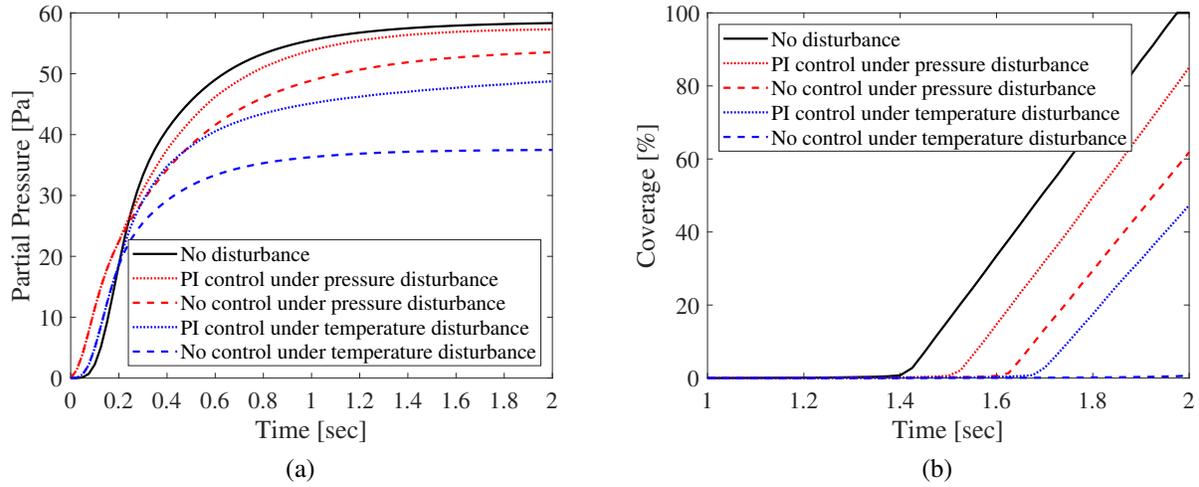
Figure 5.1: The development of process variables over process time for various operating conditions and controller actions. The black solid line represents the result without disturbance. The red dotted line represents the result with disturbance and with control action from the PI controller. The red dashed line represents the result with disturbance in pressure but without control action from the PI controller. The blue dotted line represents the result with disturbance and with control action from the PI controller. The blue dashed line represents the result with disturbance in temperature but without control action from the PI controller. (a) Precursor partial pressure over process time. (b) Film coverage over process time.

5.2.2 Feedback Control Formulation and Tuning

As mentioned in Section 5.2, a PI controller is formulated to regulate the process by adjusting the inlet mass flow rate to bring the control variable, i.e., the TDMAHf partial pressure, to the set point. The PI control algorithm is described below:

$$u(t) = K_c(e(t) + \frac{1}{\tau_I} \int_0^t e(t)\,dt\ ) \tag{5.1}$$

where t is the process time, K_c and τ_I are the controller gain and time constant, respectively, $u(t)$ is the process input, which is the inlet mass flow rate, and $e(t)$ is the error, which is the deviation between the sensed partial pressure and the set point. The partial pressure set point of 58.3 Pa is chosen for the PI controller according to the steady-state partial pressure of TDMAHf under standard operating condition.

It is also noteworthy that the online monitoring of the precursor partial pressure can be achieved

using a pressure sensor and a real-time gas chromatography (GC) device [118]. The pressure sensor is able to measure the total pressure in the reactor via a pressure sensor and generate the corresponding electric signal that is transmitted to the PI controller. Meanwhile, the small amount of gas in the reactor is continuously injected to the GC device to measure the TDMAHf molar fraction. After all, the TDMAHf partial pressure is calculated by multiplying the total pressure by the molar fraction. Then, the PI controller sends the control signal to a flow control valve in accordance with the partial pressure deviation from the set point. The control signal causes the actuator of the valve to move its modulating element and thus, the PI controller is able to adjust the inlet mass flow rate. There are several types of actuators, among which the electric actuator can be used for the PEALD process due to higher level of precision and energy efficiency. The location of the sensor should be carefully selected so that the accurate TDMAHf partial pressure can be obtained since the partial pressure, unlike the total pressure, varies upon the sensor location. The imprecise measurement of the TDMAHf partial pressure can result in invalid control action, which can cause a negative impact on the film quality. In this chapter, the pressure sensor is installed near the edge of the wafer substrate, along the precursor flow path, to measure the precise TDMAHf partial pressure on the surface as shown in Figure 5.2. The control action is executed every 0.2 s due to the sampling limitation of the GC device where the composition analysis is assumed to be implemented every 0.2 s.

In order to ensure the performance of the PI controller, its tuning parameters, K_c and τ_I, must be optimized. First, an approximated process curve is obtained under a step change input through the first order process plus dead-time (FOPDT) approach, which is widely used to identify the relationship between process time and process output [119]. As shown in Figure 5.3 (a), the approximation from the FOPDT model is able to accurately capture the features of the real process behavior. After the process is characterized, the tuning parameters are determined through pre-formulated tuning methods. A variety of tuning strategies are available and among which the most popular choices are Cohen-Coon (C-C) method and the internal model control (IMC) method. Both methods are tested and the results are demonstrated in Figure 5.3 (b). As
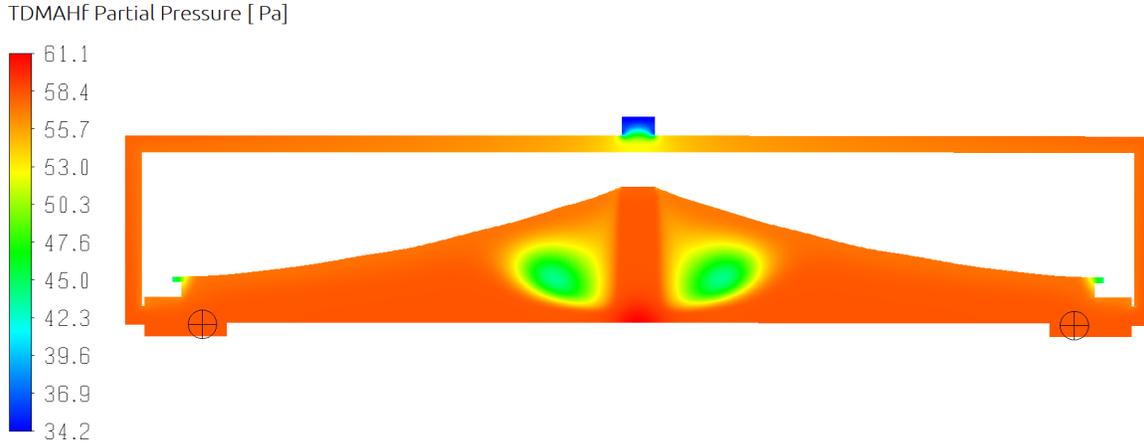
Figure 5.2: Multiscale computational fluid dynamics (CFD) model with the TDMAHf partial pressure profile at the end of the batch at the standard conditions. The ⊕ indicates the location of the pressure sensor.

indicated by Figure 5.3 (b), the response using the C-C method is 8.6% faster, and the overshoot is approximately 1.0% lower than those from the IMC method. Therefore,the tuning para meters, K_c and τ_I, are determined using the C-C method, which are obtained to be 2.37×10^{-7} and 0.9113, respectively.

5.2.2.1 Closed-Loop Results under Feedback Controller

The PI controller using the tuning parameters from Section 5.2.2 is evaluated under the previously stated disturbances where the operating pressure drops to 30 Pa from 100 Pa and the bubbler temperature drops to 338 K from 348 K. According to Figure 5.1 (b), the coverage on the wafer under the pressure disturbance and the temperature disturbance remarkably increases up to 93% from 62% and 47% from 0.7%, respectively. It is demonstrated that the PI controller noticeably diminishes the impact of the disturbances. Nevertheless, the PI control does not completely eliminate the effect of the disturbances. In other words, the PI controller does not guarantee the partial pressure set point of TDMAHf within 2 s, which is the run time determined in [120]. As can be seen in Figure 5.1 (a), the TDMAHf partial pressures under the pressure and temperature disturbances are obtained at the end of the batch as 57.3 Pa and 48.8 Pa, respectively, which are below the set point of 58.3 Pa, thus resulting in an incomplete thin-film deposition. Consequently,
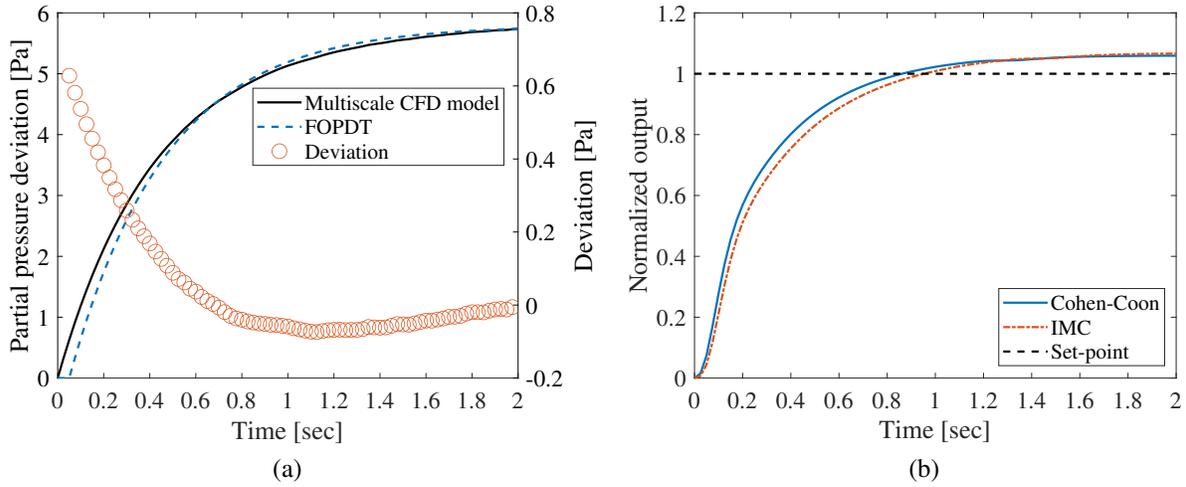
Figure 5.3: (a) Error of partial pressure versus the process time. The black solid line denotes the result from the numerical trials using multiscale CFD model. The blue dashed line denotes the process curve fitted using the the FOPDT model approximation and the red circled line shows the deviation from the multiscale CFD model. (b) Normalized partial pressure versus the process time. The black dashed line denotes the set point, which is at 100%, the blue solid line denotes the results using the parameters determined from the Cohen-Coon method, and the red dotted-dashed line denotes the results using the parameters determined from the IMC tuning method.

the result of the closed-loop PI controller reveals that though the film coverage notably increases, the single control action of the PI controller is not enough to cope with the disturbances. Thus, it is concluded that other control actions must be used to achieve the full coverage on the wafer.

5.2.3 Run-to-Run (R2R) Control Formulation and Parameter Determination

Although the PI controller effectively adjusts the inlet flow rate to counter the effect of disturbances on the partial pressure, the limited range of control action fails to provide enough driving force to ensure the quality of the deposited film. As a result, other measures must be taken to improve the film quality by accounting for the differences in partial pressure. Other than manipulating the inlet flow rate, another efficient way to compensate for partial pressure disturbances is to adjust the TDMAHf molar fraction. One option is to construct another real-time feedback controller to control the partial pressure by changing the molar fraction. However, the interference between the

inlet molar fraction and flow rate in real-time may negatively impact the controller performance. Therefore, a run-to-run (R2R) control methodology, which adjusts the manipulated input after each batch instead of in real-time, is adopted in this study.

A variety of R2R control algorithms have been proposed in the past, among which two are the most popular: the exponentially weighted moving average (EWMA), and the predictor-corrector controller (PCC), which is also known as the double-EWMA (d-EWMA). The EWMA algorithm is based on a single-input-single-output (SISO) linear regression model, also known as the input-ouput relationship, which is defined by the following equation:

$$y_t = \alpha + \beta u_{t-1} \tag{5.2}$$

where y_t is the output of the process at batch number t, α is the offset, β is the process gain, and u_{t-1} is the input (i.e., the manipulated variable) at batch number $t-1$. At the end of every batch, the input is estimated in an iterative manner and expressed as below:

$$u_t = \frac{T - a_t}{\beta} \tag{5.3}$$

where T is the target value of the output, and the EWMA estimate of the process output a_t is updated by the following equation:

$$a_t = \lambda(y_t - bu_{t-1}) + (1 - \lambda)a_{t-1} \tag{5.4}$$

where λ denotes the weight factor.

On the other hand, the PCC (d-EWMA) algorithm is the expanded version of the EWMA. It has two estimates of the process outputs, a_t and d_t, and it uses a modified linear regression model as below:

$$y_t = \alpha + \beta u_{t-1} + \eta_{t-1} \tag{5.5}$$

where η_{t-1} is the process disturbance and can be assumed to be a white noise. Similar to the

EWMA algorithm, the input of PCC algorithm is estimated by the following equation:

$$u_t = \frac{T - a_t - d_t}{\beta} \tag{5.6}$$

where the estimates of the process output a_t and d_t are updated by the following equations:

$$a_t = \lambda (y_t - bu_{t-1}) + (1 - \lambda) a_{t-1} \tag{5.7}$$

$$d_t = \omega (y_t - bu_{t-1} - a_{t-1}) + (1 - \omega) d_{t-1} \tag{5.8}$$

where λ and ω denote the weight factors.

To formulate the EWMA and the PCC-based R2R controller, the input-output relationship between the inlet TDMAHf molar fraction and the coverage needs to be derived. By running the multiscale CFD simulation under different inlet TDMAHf molar fractions (the input), the corresponding coverage (the output) is recorded to determine the input-output relationship. As shown in Figure 5.4, the linear relationship between the TDMAHf molar fraction and the coverage on the wafer was obtained in the practical operation region in which α and β in equation (5.3) and (5.5) are given as -1.2797 and 6.6775, respectively. The range of the operating TDMAHf molar fraction is limited due to the low efficiency and high operating cost. Thus, the upper-limit of the TDMAHf mole fraction is set to be 0.40. The operating TDMAHf mole fraction, 0.345, is chosen so that the full coverage can be achieved within 2 s. After every batch, the parameters a_t and d_t in equation (5.4), (5.7), and (5.8) are updated by the weighting factors λ and ω. It is also worth mentioning that, despite the wide acceptance and superior performance of linear input-output relationships, the nonlinear relationship between TDMAHf molar fraction and film coverage can potentially be characterized by a sigmoidal-like nonlinear input-output relationship. However, the asymptotic convergence of the sigmoid function may cause a slow achievement of the desired film quality, or even cause an offset in the final achievable coverage. Therefore, although there may be benefits adopting a nonlinear input-output relationship, this chapter will focus on the use of the

linear input-output relationship.
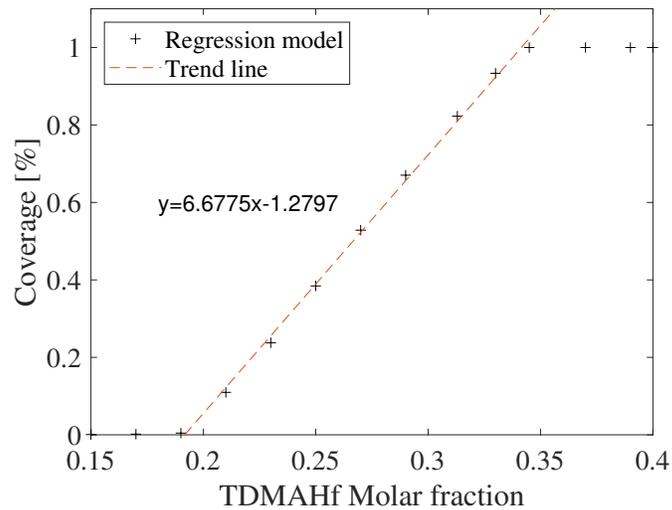
Figure 5.4: The input-output relationship between the final film coverage and the precursor molar fraction derived from linear regression used in both EWMA and PCC-based R2R controller. The *y*-axis is the final coverage and the *x*-axis is the inlet TDMAHf molar fraction.

5.2.3.1 Results under Run-to-Run Controller

As indicated in Figure 5.5, the PCC-based R2R controller shows faster respond compared to the EWMA-based R2R controller under the disturbances. As shown in the Figure 5.5 (a), the EWMA and PCC-based R2R controller reach the final coverage under the pressure disturbance at batch number 4 and 3, respectively. Likewise, the EWMA and PCC-based R2R controller achieve the final coverage under the temperature disturbance at the batch number 5 and 4, respectively as can be seen in Figure 5.5 (b). Nonetheless, the deposited film coverage is not fully achieved, where a 93 % coverage is reached under the pressure disturbance and a 62 % coverage is reached under the temperature disturbance for the two types of R2R controllers because the updated molar fraction hits the upper limit of the operating rage. The results of the R2R controllers under the disturbances reveals that an additional control action should be taken to ensure the quality of the thin-film.
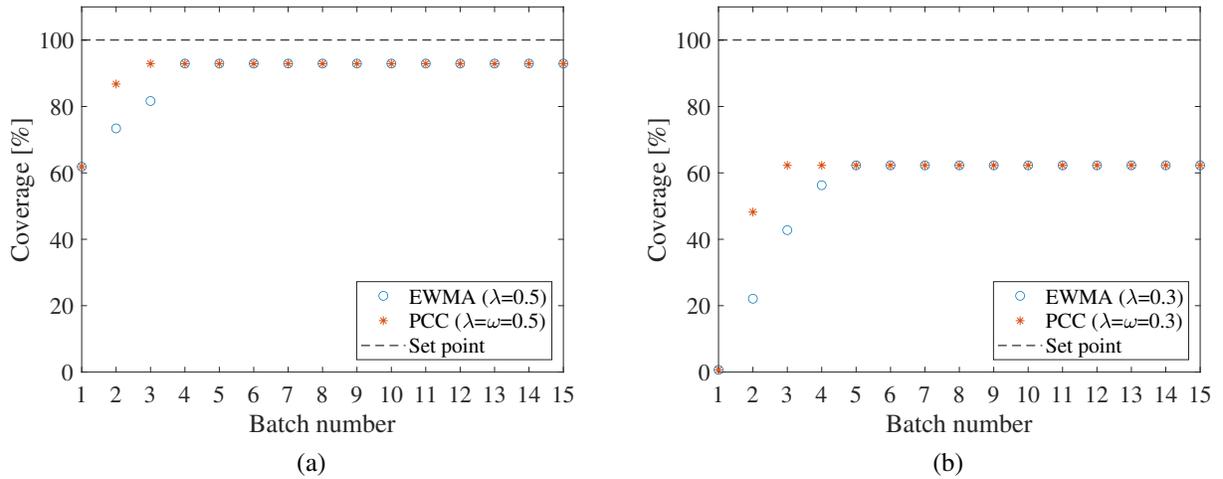
Figure 5.5: The final film coverage over fifteen batches under disturbances with only R2R control. The black dashed line represents the film coverage set point, which is at full coverage. The blue dots denote the result using the EWMA algorithm, and the red asterisks denote the result using the PCC algorithm. (a) Under pressure disturbance. (b) Under temperature disturbance.
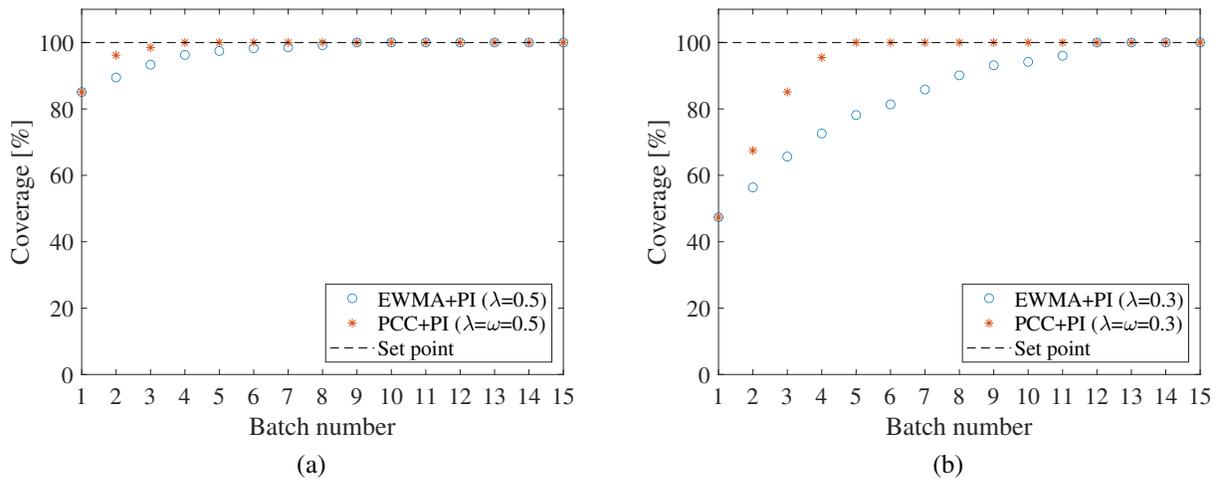
Figure 5.6: The final film coverage over fifteen batches under disturbances with the integrated PI-R2R control. The black dashed line represents the film coverage set point, which is at full coverage. The blue dots denote the result using EWMA as the R2R algorithm, and the red asterisks denote the result using PCC as the R2R algorithm. (a) Under pressure disturbance. (b) Under temperature disturbance.

5.2.4 Integration of Feedback and Run-to-Run Control

As described in the previous sections, even when the parameters of the PI controller and the R2R controller are well tuned, independently and when used in isolation they are not able to completely
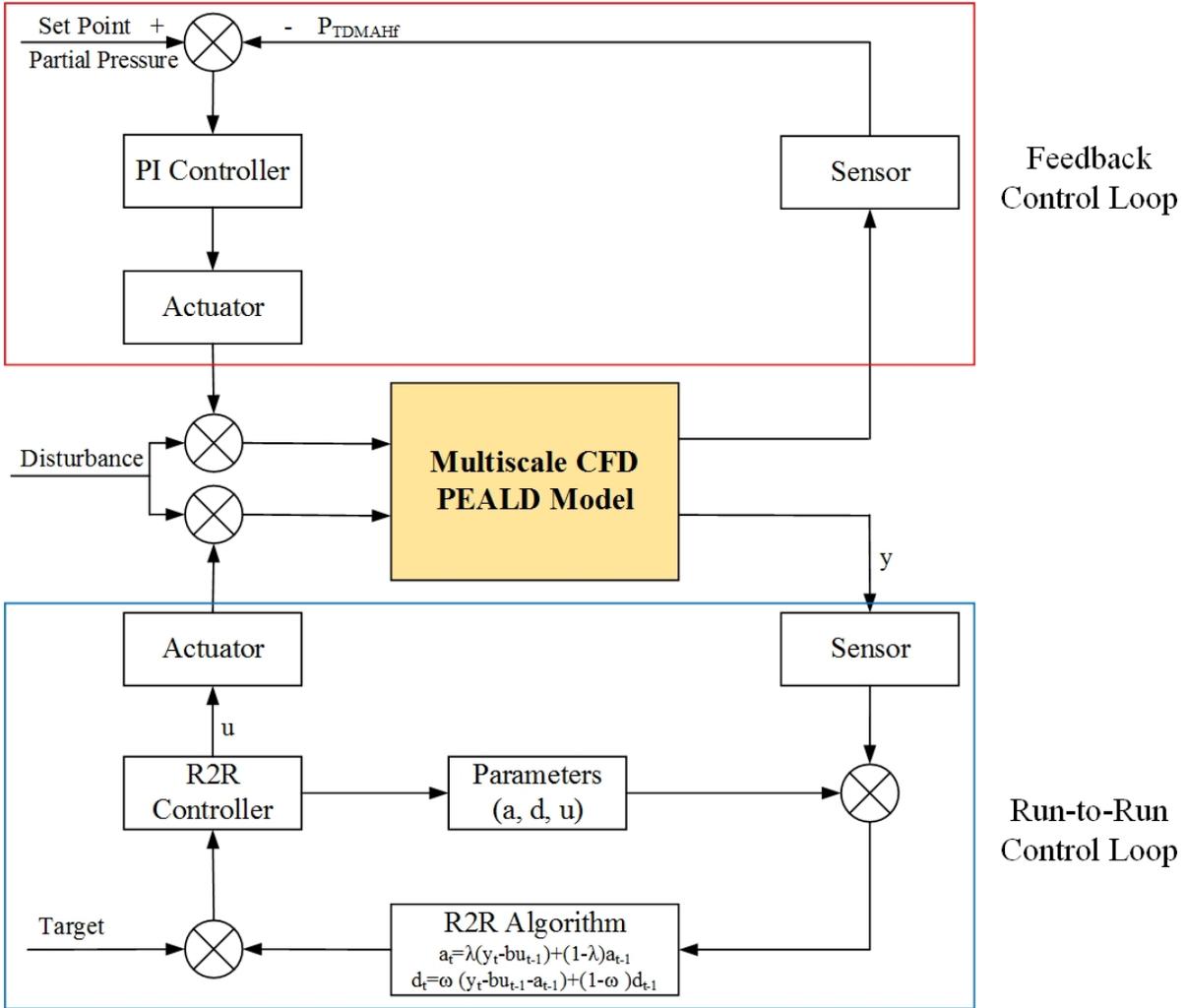
Figure 5.7: Operating principle diagram of the feedback controller and the run-to-run (R2R) controller. The red box denotes the feedback control loop, and the blue box denotes the R2R control loop.

reject the effect of the disturbances that can be potentially encountered in the PEALD process. Therefore, the mutual effort of both PI controller and of the R2R controller is required to ensure the robustness of the deposition process against disturbances. Specifically, the controller workflow is shown in the operating principle diagram as in Figure 5.7 for both the feedback controller (PI) and the R2R controller (EWMA & d-EWMA). A number of simulation runs using the multiscale CFD model have been carried out to test the performance of the integrated PI-R2R controller. First, the PI controller, formulated in section 5.2.2, is used to adjust the inlet mass flow rate during the deposition process. Then, after the first batch has finished, the film coverage is recorded and

transferred to the R2R controller. Before the next run, the R2R controller computes the TDMAHf molar fraction of the inlet stream through either the EWMA or the PCC algorithms, of which the performance will be evaluated. The updated TDMAHf molar fraction is then used as the inlet condition for the next run to further eliminate the effect of the disturbances. The same disturbances mentioned in Section 5.2.1 are used to verify the performance of the integrated PI-R2R controller.

The results of the integrated PI-R2R controller using both EWMA and PCC algorithms in the R2R controller are shown in Figure 5.6. As shown in Figure 5.6 (a), the EWMA-based and the PCC-based integrated PI-R2R controllers achieve full coverage on the surface at batch number 9 and 4, respectively, under the same pressure disturbance. Likewise, the EWMA-based and the PCC-based integrated controller reach full coverage at batch number 12 and 5, respectively, under the temperature disturbance as shown in Figure 5.6 (b). Therefore, it is proved that the integrated controller successfully ensures the full coverage of the thin-film under various disturbances. Furthermore, as noted in Section 5.2.3.1, the PCC-based integrated PI-R2R controller shows faster response than the EWMA-based PI-R2R controller. In particular, the deposition under the regulation of the PCC-based PI-R2R controller achieves full coverage about 60% faster than the EWMA-based integrated PI-R2R controller. Thus, the strategy of integrating the PI controller and R2R controller is validated and can largely improve the process robustness.

5.3 Conclusion

In this chapter, the previously developed multiscale CFD model was used to design and evaluate a control strategy that utilizes an integrated PI-R2R controller which adopts both feedback control and R2R control. The designed controller was investigated from both open-loop and closed-loop simulations in the presence of two types of disturbances. To eliminate the effect of the disturbances, first, the PI controller was constructed to ensure the desired TDMAHf partial pressure by adjusting the inlet mass flow rate. Next, the R2R controller was set up to manipulate the inlet TDMAHf molar fraction according to the derived input-output relationship at the beginning

116

of every batch to further ensure the complete deposition and the film conformity. To validate the necessity and the effect of the integrated controller, both the PI and the R2R controller were evaluated independently. These evaluation results demonstrated that individual control actions were not sufficient to guarantee full coverage, and an additional control effort was required. Thus, the R2R controller and the PI controller were integrated to fully eliminate the effect of the disturbances. A corresponding case study using the multiscale CFD model validated that the integrated PI-R2R controller was able to reject the disturbances and achieves the conformity of atomic layer deposition. Furthermore, the results demonstrated that the PCC algorithm used in the R2R controller was able to improve the run-to-run control action and respond approximately 60 % faster than the EWMA algorithm. In the future, research adopting more advanced control method, such as model predictive model using the developed data-driven model, is recommended to optimize the control workflow and maximize control efficiency.

Chapter 6

Conclusion

This dissertation provides a number of rigorous methods for the design of multiscale CFD modeling for the ALD and PEALD processes and the construction of associated control schemes.

Specifically, in Chapters 2 and 4, we have constructed a first-principles-based multiscale CFD model for atomic layer deposition of SiO_2 thin-films using BTBAS and ozone and the plasma-enhanced atomic layer deposition of HfO_2 using TDMAHf and oxygen plasma. Application of the ALD model for a nominal reactor geometry design sucessfully reproduced experimental and industrial GPC and half-cycle time at 100 sccm and 1330 sccm precursor inlet flow rates. Based on the transient non-uniformity discovered in the nominal design, geometry optimizations on the showerhead hole size and distribution and upstream region were performed and their influence on the gas-phase profile and microscopic deposition were investigated. Two geometry optimizations were adjusted and combined for a final optimized design of the ALD showerhead reactor. The resulting optimal chamber design successfully reduced the gas-phase profile development time and created radial momentum that guided the gas stream to the less concentrated region. The remote plasma generation model and the macroscopic gas-phase CFD model are constructed and coupled with the previously developed microscopic kMC model through an automated workflow. The multiscale model was validated and utilized to identify the effects of plasma generation power, operating pressure, operating temperature, and inlet

flow rates on the deposition of HfO_2 thin-films.Using the multiscale CFD model, a set of baseline operating parameters were determined to produce highly uniform HfO_2 thin-film within conventional industrial half-cycle time, and the transient non-uniformity across the entire substrate was successfully reduced under the baseline operating conditions. Moreover, these multiscale simulation methodology developed can serve as a guidance model to other ALD systems where gas-phase transport and surface deposition need different simulation methods with information exchange. With the correct reaction kinetics for the microscopic film growth model and thermodynamic data of the reactants and products for the macroscopic gas transport model, such multiscale model can realisically represent a variety of deposition processes.

In Chapters 3 and 5, feedback and run-to-run control schemes were constructed for a furnace ALD system and an industrial scale PEALD system. In particular, 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. Substrate deposition profiles at various wafer locations and their relationships to the half-cycle time were recorded. Using this information, 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. 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. Similarly, a similar but more comprehensive control strategy was designed for PEALD system which adopts both feedback control and R2R control. The designed controller was investigated from both open-loop and closed-loop simulations in the presence of two types of disturbances. To eliminate the effect of the disturbances, first, the PI controller was constructed to ensure the desired TDMAHf partial pressure by adjusting the inlet mass flow rate. Next, the R2R controller was set up to manipulate the inlet TDMAHf molar fraction according to the derived input-output

relationship at the beginning of every batch to further ensure the complete deposition and the film conformity. To validate the necessity and the effect of the integrated controller, both the PI and the R2R controller were evaluated independently. These evaluation results demonstrated that individual control actions were not sufficient to guarantee full coverage, and an additional control effort was required. Thus, the R2R controller and the PI controller were integrated to fully eliminate the effect of the disturbances. It was demonstrated that the integrated PI-R2R controller was able to reject the disturbances and achieves the conformity of PEALD. Furthermore, the results demonstrated that the PCC algorithm used in the R2R controller was able to improve the run-to-run control action and respond faster than the EWMA algorithm. In the future, research adopting more advanced control method, such as model predictive model using the developed data-driven model, is recommended to optimize the control workflow and maximize control efficiency.

Bibliography

[1] H. Nalwa, Handbook of Thin Films, Vol. 1, Academic Press, Burlington, 2002.

[2] S. M. George, A. W. Ott, J. W. Klaus, Surface chemistry for atomic layer growth, The Journal of Physical Chemistry 100 (1996) 13121–13131.

[3] K. Schuegraf, M. C. Abraham, A. Brand, M. Naik, R. Thakur, Semiconductor logic technology innovation to achieve sub-10 nm manufacturing, IEEE Journal of the Electron Devices Society 1 (2013) 66–75.

[4] T. Kääriäinen, D. Cameron, M.-L. Kääriäinen, A. Sherman, Atomic layer deposition: principles, characteristics, and nanotechnology applications, John Wiley & Sons, 2013.

[5] S. M. George, Atomic layer deposition: An overview, Chemical Reviews 110 (2009) 111–131.

[6] C. M. Tanner, Y.-C. Perng, C. Frewin, S. E. Saddow, J. P. Chang, Electrical performance of Al_2O_3 gate dielectric films deposited by atomic layer deposition on $4H$-SiC, Applied Physics Letters 91 (2007) 203510.

[7] T. R. B. Foong, Y. Shen, X. Hu, A. Sellinger, Template-directed liquid ALD growth of TiO_2 nanotube arrays: Properties and potential in photovoltaic devices, Advanced Functional Materials 20 (2010) 1390–1396.

[8] M. Shirazi, S. D. Elliott, Atomistic kinetic Monte Carlo study of atomic layer deposition derived from density functional theory, Journal of Computational Chemistry 35 (2014) 244–259.

[9] K. Ishikawa, K. Karahashi, T. Ichiki, J. P. Chang, S. M. George, W. Kessels, H. J. Lee, S. Tinck, J. H. Um, K. Kinoshita, Progress and prospects in nanoscale dry processes: How can we control atomic layer reactions?, Japanese Journal of Applied Physics 56 (2017) 06HA02.

[10] J. Joo, S. M. Rossnagel, Plasma modeling of a PEALD system for the deposition of TiO_2 and HfO_2, Journal of Korean Physical Society 54 (2009) 1048.

[11] Y. Won, S. Park, J. Koo, S. Kim, J. Kim, H. Jeon, Initial reaction of hafnium oxide deposited by remote plasma atomic layer deposition method, Applied Physics Letters 87 (2005) 262901.

[12] M. C. Schwille, T. Schössler, J. Barth, M. Knaut, F. Schön, A. Höchst, M. Oettel, J. Bartha, Experimental and simulation approach for process optimization of atomic layer deposited thin films in high aspect ratio 3D structures, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 35 (2017) 01B118.

[13] K. Kukli, M. Ritala, M. Schuisky, M. Leskelä, T. Sajavaara, J. Keinonen, T. Uustare, A. Hårsta, Atomic layer deposition of titanium oxide from TiI_4 and H_2O_2, Chemical Vapor Deposition 6 (2000) 303–310.

[14] X. Liu, S. Ramanathan, A. Longdergan, A. Srivastava, E. Lee, T. E. Seidel, J. T. Barton, D. Pang, R. G. Gordon, ALD of hafnium oxide thin films from tetrakis (ethylmethylamino) hafnium and ozone, Journal of the Electrochemical Society 152 (2005) G213–G219.

[15] S. J. Yun, J. W. Lim, J.-H. Lee, PEALD of zirconium oxide using tetrakis (ethylmethylamino) zirconium and oxygen, Electrochemical and Solid-State Letters 7 (2004) F81–F84.

[16] X. Liu, S. Ramanathan, T. E. Seidel, Atomic layer deposition of hafnium oxide thin films from tetrakis (dimethylamino) hafnium (TDMAH) and ozone, MRS Online Proceedings Library Archive 765.

[17] X. Shi, H. Tielens, S. Takeoka, T. Nakabayashi, L. Nyns, C. Adelmann, A. Delabie, T. Schram, L. Ragnarsson, M. Schaekers, et al., Development of ALD $HfZrO_x$ with TDEAH/TDEAZ and H_2O, Journal of the Electrochemical Society 158 (2011) H69–H74.

[18] V. R. Rai, V. Vandalon, S. Agarwal, Surface reaction mechanisms during ozone and oxygen plasma assisted atomic layer deposition of aluminum oxide, Langmuir 26 (2010) 13732–13735.

[19] K. Kanomata, P. Pansila, B. Ahmmad, S. Kubota, K. Hirahara, F. Hirose, Infrared study on room-temperature atomic layer deposition of TiO_2 using tetrakis (dimethylamino) titanium and remote-plasma-excited water vapor, Applied Surface Science 308 (2014) 328–332.

[20] T. K. Won, H. Nominanda, S.-M. Cho, S. Y. Choi, B. S. Park, J. M. White, S. Anwar, J. Kudela, Thin film deposition using microwave plasma, uS Patent 8,883,269 (Nov. 11 2014).

[21] H.-T. Jeon, I.-H. Kim, S.-H. Kim, C.-W. Chung, S.-K. Lee, Apparatus for generating remote plasma, uS Patent App. 11/703,621 (Aug. 23 2007).

[22] M. Nayhouse, J. S.-I. Kwon, P. D. Christofides, G. Orkoulas, Crystal shape modeling and control in protein crystal growth, Chemical Engineering Science 87 (2013) 216–223.

[23] M. Ikegawa, J. Kobayashi, Deposition profile simulation using the direct simulation Monte Carlo method, Journal of the Electrochemical Society 136 (1989) 2982–2986.

[24] M. Crose, W. Zhang, A. Tran, P. D. Christofides, Multiscale three-dimensional CFD modeling for PECVD of amorphous silicon thin films, Computers & Chemical Engineering 113 (2018) 184–195.

[25] S. D. Elliott, J. C. Greer, Simulating the atomic layer deposition of alumina from first principles, Journal of Materials Chemistry 14 (2004) 3246–3250.

[26] S. Rasoulian, L. A. Ricardez-Sandoval, Uncertainty analysis and robust optimization of multiscale process systems with application to epitaxial thin film growth, Chemical Engineering Science 116 (2014) 590–600.

[27] C. C. Battaile, D. J. Srolovitz, Kinetic Monte Carlo simulation of chemical vapor deposition, Annual Review of Materials Research 32 (2002) 297–319.

[28] J. C. Rey, L. Cheng, J. P. McVittie, K. C. Saraswat, Monte Carlo low pressure deposition profile simulations, Journal of Vacuum Science & Technology A 9 (1991) 1083–1087.

[29] A. Dkhissi, A. Estève, C. Mastail, S. Olivier, G. Mazaleyrat, L. Jeloaica, M. Djafari-Rouhani, Multiscale modeling of the atomic layer deposition of HfO$_2$ thin film grown on silicon: How to deal with a kinetic Monte Carlo procedure, Journal of Chemical Theory and Computation 4 (2008) 1915–1927.

[30] H. C. M. Knoops, E. Langereis, M. C. M. Van De Sanden, W. M. M. Kessels, Conformality of plasma-assisted ALD: physical processes and modeling, Journal of The Electrochemical Society 157 (2010) G241–G249.

[31] C. Lee, M. Lieberman, Global model of Ar, O$_2$, Cl$_2$, and Ar/O$_2$ high-density plasma discharges, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 13 (1995) 368–380.

[32] S. Tinck, A. Bogaerts, Computer simulations of an oxygen inductively coupled plasma used for plasma-assisted atomic layer deposition, Plasma Sources Science and Technology 20 (2011) 015008.

[33] C. Corr, S. Gomez, W. Graham, Discharge kinetics of inductively coupled oxygen plasmas: experiment and model, Plasma Sources Science and Technology 21 (2012) 055024.

[34] D. I. Gerogiorgis, B. Ydstie, Multiphysics CFD modelling for design and simulation of a multiphase chemical reactor, Chemical Engineering Research and Design 83 (2005) 603 – 610.

[35] E. I. Epelle, D. I. Gerogiorgis, A multiparametric CFD analysis of multiphase annular flows for oil and gas drilling applications, Computers & Chemical Engineering 106 (2017) 645 – 661.

[36] D. Pan, T. Li, T. Chien Jen, C. Yuan, Numerical modeling of carrier gas flow in atomic layer deposition vacuum reactor: A comparative study of lattice boltzmann models, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 32 (2014) 01A110.

[37] Y. Zhang, Y. Ding, P. D. Christofides, Multiscale computational fluid dynamics modeling of thermal atomic layer deposition with application to chamber design, Chemical Engineering Research and Design 147 (2019) 529–544.

[38] C. A. Murray, S. D. Elliott, D. Hausmann, J. Henri, A. LaVoie, Effect of reaction mechanism on precursor exposure time in atomic layer deposition of silicon oxide and silicon nitride, ACS Applied Materials & Interfaces 6 (2014) 10534–10541.

[39] J. M. Lee, J. H. Lee, Approximate dynamic programming-based approaches for input-output data-driven control of nonlinear processes, Automatica 41 (2005) 1281–1288.

[40] F. G. Djurabekova, R. Domingos, G. Cerchiara, N. Castin, E. Vincent, L. Malerba, Artificial intelligence applied to atomistic kinetic Monte Carlo simulations in Fe–Cu alloys, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 255 (2007) 8–12.

[41] C. Nicolas, M. Lorenzo, Calculation of proper energy barriers for atomistic kinetic Monte Carlo simulations on rigid lattice with chemical and strain field long-range effects using Artificial Neural Networks, The Journal of Chemical Physics 132 (2010) 074507.

[42] S. Rasoulian, L. A. Ricardez-Sandoval, A robust nonlinear model predictive controller for a multiscale thin film deposition process, Chemical Engineering Science 136 (2015) 38–49.

[43] P. Mhaskar, A. Garg, B. Corbett, Modeling and Control of Batch Processes: Theory and Application, Springer Science & Business Media, 2018.

[44] G. Kimaev, L. A. Ricardez-Sandoval, A comparison of efficient uncertainty quantification techniques for stochastic multiscale systems, AIChE Journal 63 (2017) 3361–3373.

[45] H. Singh Sidhu, P. Siddhamshetty, J. Kwon, Approximate dynamic programming based control of proppant concentration in hydraulic fracturing, Mathematics 6 (2018) 132.

[46] Y. Wang, F. Gao, F. J. Doyle, Survey on iterative learning control, repetitive control, and run-to-run control, Journal of Process Control 19 (2009) 1589–1600.

[47] J. S.-I. Kwon, M. Nayhouse, G. Orkoulas, D. Ni, P. D. Christofides, A method for handling batch-to-batch parametric drift using moving horizon estimation: Application to run-to-run mpc of batch crystallization, Chemical Engineering Science 127 (2015) 210–219.

[48] J. S.-I. Kwon, M. Nayhouse, G. Orkoulas, D. Ni, P. D. Christofides, Run-to-run-based model predictive control of protein crystal shape in batch crystallization, Industrial & Engineering Chemistry Research 54 (2015) 4293–4302.

[49] S. Rasoulian, L. A. Ricardez-Sandoval, Robust multivariable estimation and control in an epitaxial thin film growth process under uncertainty, Journal of Process Control 34 (2015) 70–81.

[50] S. Rasoulian, L. A. Ricardez-Sandoval, Stochastic nonlinear model predictive control applied to a thin film deposition process under uncertainty, Chemical Engineering Science 140 (2016) 90–103.

[51] S.-K. Oh, J. M. Lee, Iterative learning model predictive control for constrained multivariable control of batch processes, Computers & Chemical Engineering 93 (2016) 284 – 292.

[52] D. Chaffart, L. A. Ricardez-Sandoval, Robust dynamic optimization in heterogeneous multiscale catalytic flow reactors using polynomial chaos expansion, Journal of Process Control 60 (2017) 128–140.

[53] D. Chaffart, L. A. Ricardez-Sandoval, Optimization and control of a thin film growth process: A hybrid first principles/Artificial Neural Network based multiscale modelling approach, Computers & Chemical Engineering 119 (2018) 465–479.

[54] A. Armaou, P. D. Christofides, Plasma enhanced chemical vapor deposition: Modeling and control, Chemical Engineering Science 54 (1999) 3305–3314.

[55] Y. Ding, Y. Zhang, K. Kim, A. Tran, Z. Wu, P. D. Christofides, Microscopic modeling and optimal operation of thermal atomic layer deposition, Chemical Engineering Research and Design 145 (2019) 159–172.

[56] J. Elam, M. Groner, S. George, Viscous flow reactor with quartz crystal microbalance for thin film growth by atomic layer deposition, Review of Scientific Instruments 73 (2002) 2981–2987.

[57] S. Pittal, P. G. Snyder, N. J. Ianno, Ellipsometry study of non-uniform lateral growth of ZnO thin films, Thin Solid Films 233 (1993) 286–288.

[58] T. J. Dalton, W. T. Conner, H. H. Sawin, Interferometric real-time measurement of uniformity for plasma etching, Journal of the Electrochemical Society 141 (1994) 1893–1900.

[59] J. Bühler, F. Steiner, H. Baltes, Silicon dioxide sacrificial layer etching in surface micromachining, Journal of Micromechanics and Microengineering 7 (1997) R1.

[60] C. S. Lee, M. S. Oh, H. S. Park, Plasma enhanced atomic layer deposition (PEALD) equipment and method of forming a conducting thin film using the same thereof, uS Patent 7,138,336 (2006).

[61] A. Frisch, A. B. Nielsen, A. J. Holder, Gaussview user manual, Gaussian Inc., Pittsburgh, PA 556.

[62] M. L. O′Neill, H. R. Bowen, A. Derecskei-Kovacs, K. S. Cuthill, B. Han, M. Xiao, Impact of aminosilane precursor structure on silicon oxides by atomic layer deposition, The Electrochemical Society Interface 20 (2011) 33–37.

[63] B. Han, Q. Zhang, J. Wu, B. Han, E. J. Karwacki, A. Derecskei, M. Xiao, X. Lei, M. L. O′Neill, H. Cheng, On the mechanisms of SiO_2 thin-film growth by the full atomic layer deposition process using bis(t-butylamino)silane on the hydroxylated $SiO_2(001)$ surface, The Journal of Physical Chemistry C 116 (2011) 947–952.

[64] T. Nishiguchi, H. Nonaka, S. Ichimura, Y. Morikawa, M. Kekura, M. Miyamoto, High-quality SiO_2 film formation by highly concentrated ozone gas at below 600°C, Applied Physics Letters 81 (2002) 2190–2192.

125

[65] G. Prechtl, A. Kersch, G. S. Icking-Konert, W. Jacobs, T. Hecht, H. Boubekeur, U. Schröder, A model for Al_2O_3 ALD conformity and deposition rate from oxygen precursor reactivity, in: Proceedings of the International Electron Devices Meeting, Washington, DC, USA, Vol. 810, 2003.

[66] A. Fluent, Ansys fluent theory guide 15.0, ANSYS, Canonsburg, PA.

[67] S. W. King, Plasma enhanced atomic layer deposition of SiN_x:H and SiO_2, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 29 (2011) 041501.

[68] R. Eymard, T. Gallouët, R. Herbin, Finite volume methods, Handbook of numerical analysis 7 (2000) 713–1018.

[69] A. Tran, A. Aguirre, H. Durand, M. Crose, P. D. Christofides, CFD modeling of a industrial-scale steam methane reforming furnace, Chemical Engineering Science 171 (2017) 576–598.

[70] R. B. Bird, W. E. Stewart, E. N. Lightfoot, Transport phenomena, John Wiley & Sons, 2007.

[71] D. F. Young, B. R. Munson, T. H. Okiishi, W. W. Huebsch, A brief introduction to fluid mechanics, John Wiley & Sons, 2010.

[72] Q. A. Acton, Chemical Processes-Advances in Research and Application: 2012 Edition: ScholarlyBrief, ScholarlyEditions, 2012.

[73] E. Granneman, P. Fischer, D. Pierreux, H. Terhorst, P. Zagwijn, Batch ALD: Characteristics, comparison with single wafer ALD, and examples, Surface and Coatings Technology 201 (2007) 8899–8907.

[74] R. Courant, K. Friedrichs, H. Lewy, On the partial difference equations of mathematical physics, IBM Journal of Research and Development 11 (2) (1967) 215–234.

[75] C. A. d. Moura, C. S. Kubrusly, The Courant-Friedrichs-Lewy (CFL) Condition: 80 Years After Its Discovery, Birkhäuser Basel, 2012.

[76] J. Huang, G. Hu, G. Orkoulas, P. D. Christofides, Dependence of film surface roughness and slope on surface migration and lattice size in thin film deposition processes, Chemical Engineering Science 65 (2010) 6101–6111.

[77] J. Huang, G. Hu, G. Orkoulas, P. D. Christofides, Dynamics and lattice-size dependence of surface mean slope in thin-film deposition, Industrial & Engineering Chemistry Research 50 (2010) 1219–1230.

[78] M. C. Schwille, T. Schössler, S. Florian, M. Oettel, J. W. Bartha, Temperature dependence of the sticking coefficients of bis-diethyl aminosilane and trimethylaluminum in atomic layer deposition, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 35 (2017) 01B119.

[79] R. D. Cortright, J. A. Dumesic, Kinetics of heterogeneous catalytic reactions: Analysis of reaction schemes, Advances in Catalysis 46 (2001) 161–264.

[80] Y. Lou, P. D. Christofides, Feedback control of surface roughness of GaAs(001) thin films using kinetic Monte Carlo models, Computers & Chemical Engineering 29 (2004) 225–241.

[81] P. D. Christofides, A. Armaou, Y. Lou, A. Varshney, Control and optimization of multiscale process systems, Springer Science & Business Media, 2008.

[82] N. R. Council, et al., Beyond the molecular frontier: challenges for chemistry and chemical engineering, National Academies Press, 2003.

[83] M. Fermeglia, S. Pricl, Multiscale molecular modeling in nanostructured material design and process system engineering, Computers & Chemical Engineering 33 (2009) 1701–1710.

[84] Y. Liu, H. Liu, Development of 3D polymer DFT and its application to molecular transport through a surfactant-covered interface, AIChE Journal 64 (2018) 238–249.

[85] F. Jensen, Introduction to computational chemistry, John Wiley & Sons, 2017.

[86] T. Gilbert, Hohenberg-Kohn theorem for nonlocal external potentials, Physical Review B 12 (1975) 2111.

[87] O. V. Gritsenko, P. R. T. Schipper, E. J. Baerends, Exchange and correlation energy in density functional theory: Comparison of accurate density functional theory quantities with traditional Hartree-Fock based ones and generalized gradient approximations for the molecules Li_2, N_2, F_2, The Journal of Chemical Physics 107 (1997) 5007–5015.

[88] C. Lee, W. Yang, R. G. Parr, Development of the colle-salvetti correlation-energy formula into a functional of the electron density, Physical Review B 37 (1988) 785.

[89] A. D. Becke, Density-functional thermochemistry. iii. the role of exact exchange, The Journal of Chemical Physics 98 (1993) 5648–5652.

[90] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138.

[91] L. A. Curtiss, P. C. Redfern, K. Raghavachari, Gaussian-4 theory, The Journal of Chemical physics 126 (2007) 084108.

[92] J. W. Ochterski, Thermochemistry in gaussian, Gaussian Inc (2000) 1–19.

[93] V. Mankad, P. K. Jhu, First-principles study of water adsorption on α-SiO_2(110) surface, AIP Advances 6 (2016) 085001.

[94] M. Putkonen, M. Bosund, O. M. Ylivaara, R. L. Puurunen, L. Kilpi, H. Ronkainen, S. Sintonen, S. Ali, H. Lipsanen, X. Liu, Thermal and plasma enhanced atomic layer deposition of SiO_2 using commercial silicon precursors, Thin Solid Films 558 (2014) 93–98.

[95] N. M. Laurendeau, Statistical thermodynamics: fundamentals and applications, Cambridge University Press, 2005.

[96] P. Poodt, D. C. Cameron, E. Dickey, S. M. George, V. Kuznetsov, G. N. Parsons, F. Roozeboom, G. Sundaram, A. Vermeer, Spatial atomic layer deposition: A route towards further industrialization of atomic layer deposition, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 30 (2012) 010802.

[97] R. Cook, R. Stevens, P. Schwartz, C. Tejamo, V. Nilsen, G. Ormonde, A. Paranjpe, S. Nag, M. Patten, Substrate carrier for parallel wafer processing reactor, uS Patent App. 10/966,245 (Sep. 1 2005).

[98] A. Lankhorst, B. Paarhuis, H. Terhorst, P. Simons, C. Kleijn, Transient ALD simulations for a multi-wafer reactor with trenched wafers, Surface and Coatings Technology 201 (2007) 8842–8848.

[99] G. Dingemans, B. Jongbloed, W. Knaepen, D. Pierreux, L. Jdira, H. Terhorst, Merits of batch ALD, ECS Transactions 64 (2014) 35–49.

[100] Z. Deng, W. He, C. Duan, B. Shan, R. Chen, Atomic layer deposition process optimization by computational fluid dynamics, Vacuum 123 (2016) 103–110.

[101] Y. Ding, Y. Zhang, Y. M. Ren, G. Orkoulas, P. D. Christofides, Machine learning-based modeling and operation for ALD of SiO_2 thin-films using data from a multiscale CFD simulation, Chemical Engineering Research and Design 151 (2019) 131–145.

[102] S. Lu, Y. Chang, D. Sun, V. Dang, M. X. Yang, A. Chang, A. N. Nguyen, M. Xi, Valve control system for atomic layer deposition chamber, uS Patent 6,734,020 (May 11 2004).

[103] S. Shinde, A. Sonar, Y. Sun, Advanced process control for furnace systems in semiconductor manufacturing, in: ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference, IEEE, 2013, pp. 275–279.

[104] S. Rasoulian, L. A. Ricardez-Sandoval, A robust nonlinear model predictive controller for a multiscale thin film deposition process, Chemical Engineering Science 136 (2015) 38–49.

[105] P. Chen, T. Mitsui, D. B. Farmer, J. Golovchenko, R. G. Gordon, D. Branton, Atomic layer deposition to fine-tune the surface properties and diameters of fabricated nanopores, Nano Letters 4 (2004) 1333–1337.

[106] W.-J. He, H.-T. Zhang, Z. Chen, B. Chu, K. Cao, B. Shan, R. Chen, Generalized predictive control of temperature on an atomic layer deposition reactor, IEEE Transactions on Control Systems Technology 23 (2015) 2408–2415.

[107] B. Burnak, N. A. Diangelakis, E. N. Pistikopoulos, Towards the grand unification of process design, scheduling, and control—utopia or reality?, Processes 7 (2019) 461.

[108] M. Rafiei, L. A. Ricardez-Sandoval, New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability, Computers & Chemical Engineering 132 (2020) 106610.

[109] A. J. Chorin, Numerical solution of the Navier-Stokes equations, Mathematics of Computation 22 (1968) 745–762.

[110] H. Grad, Principles of the kinetic theory of gases, in: Thermodynamik der Gase/Thermodynamics of Gases, Springer, 1958, pp. 205–294.

[111] M. Crose, J. S.-I. Kwon, A. Tran, P. D. Christofides, Multiscale modeling and run-to-run control of PECVD of thin film solar cells, Renewable Energy 100 (2017) 129–140.

[112] K. Jousten, S. Putzke, J. Buthig, Partial pressure measurement standard for characterizing partial pressure analyzers and measuring outgassing rates, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 33 (2015) 061603.

[113] A. Fedele, M. Pedone, R. Moretti, T. Wiersberg, R. Somma, C. Troise, G. De Natale, Real-time quadrupole mass spectrometry of hydrothermal gases from the unstable pisciarelli fumaroles (campi flegrei): Trends, challenges and processes, International Journal of Mass Spectrometry 415 (2017) 44–54.

[114] M. Leskelä, M. Ritala, Atomic layer deposition (ALD): from precursors to thin film structures, Thin Solid Films 409 (2002) 138–146.

[115] S. Kher, S. Nguyen, P. Narwankar, S. Tandon, S. Jumper, V. Sermona, Precursors and hardware for CVD and ALD, uS Patent App. 11/847,158 (Mar. 13 2008).

[116] M. Crose, W. Zhang, A. Tran, P. D. Christofides, Run-to-run control of PECVD systems: Application to a multiscale three-dimensional CFD model of silicon thin film deposition, AIChE Journal 65 (7) (2019) e16400.

[117] M. S. Gaffney, C. M. Reaves, R. S. Smith, A. L. Holmes, S. P. DenBaars, Control of iii–v epitaxy in a metalorganic chemical vapor deposition process: impact of source flow control on composition and thickness, Journal of Crystal Growth 167 (1) (1996) 8 – 16.

[118] D. O. Sparkman, Z. E. Penton, F. G. Kitson, Chapter 2 - gas chromatography, in: Gas Chromatography and Mass Spectrometry (Second Edition), second edition Edition, 2011, pp. 15 – 83.

[119] D. R. Coughanowr, S. E. LeBlanc, Chapter 18 - controller tuning and process identification, in: Process Systems Analysis and Control, McGraw-Hill Education, 2009, pp. 408–418.

[120] Y. Ding, Y. Zhang, H. Y. Chung, P. D. Christofides, Machine learning-based modeling and operation of plasma-enhanced atomic layer deposition of hafnium oxide thin films, Computers & Chemical Engineering 144 (2021) 107148.

[121] S.-J. Jeong, Y. Gu, J. Heo, J. Yang, C.-S. Lee, M.-H. Lee, Y. Lee, H. Kim, S. Park, S. Hwang, Thickness scaling of atomic-layer-deposited HfO_2 films and their application to wafer-scale graphene tunnelling transistors, Scientific Reports 6 (2016) 20907.

[122] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, G. Ghibaudo, Review on high-k dielectrics reliability issues, IEEE Transactions on Device and Materials Reliability 5 (2005) 5–19.

[123] H. Profijt, S. Potts, M. Van de Sanden, W. Kessels, Plasma-assisted atomic layer deposition: basics, opportunities, and challenges, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 29 (2011) 050801.

[124] Y. Zhang, Y. Ding, P. D. Christofides, Integrating feedback control and run-to-run control in multi-wafer thermal atomic layer deposition of thin films, Processes 8 (2020) 18.

[125] Y. Ding, Y. Zhang, G. Orkoulas, P. D. Christofides, Microscopic modeling and optimal operation of plasma enhanced atomic layer deposition, Chemical Engineering Research and Design 159 (2020) 439–454.

[126] C. Multiphysics, Introduction to COMSOL multiphysics$^®$, COMSOL Multiphysics, Burlington, MA, accessed Feb 9 (1998) 2018.

[127] H. Jeon, Y. Won, The reaction pathways of the oxygen plasma pulse in the hafnium oxide atomic layer deposition process, Applied Physics Letters 93 (12) (2008) 124104.

[128] S. X. Lao, R. M. Martin, J. P. Chang, Plasma enhanced atomic layer deposition of HfO_2 and ZrO_2 high-k thin films, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 23 (2005) 488–496.

[129] S. Heil, J. Van Hemmen, C. Hodson, N. Singh, J. Klootwijk, F. Roozeboom, M. Van de Sanden, W. Kessels, Deposition of TiN and HfO_2 in a commercial 200 mm remote plasma atomic layer deposition reactor, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 25 (2007) 1357–1366.

[130] D. M. Hausmann, E. Kim, J. Becker, R. G. Gordon, Atomic layer deposition of hafnium and zirconium oxides using metal amide precursors, Chemistry of materials 14 (2002) 4350–4358.

[131] N. Kumar, B. P. A. George, H. Abrahamse, V. Parashar, S. S. Ray, J. C. Ngila, A novel approach to low-temperature synthesis of cubic HfO_2 nanostructures and their cytotoxicity, Scientific Reports 7 (2017) 1–14.

[132] J. Provine, P. Schindler, J. Torgersen, H. J. Kim, H.-P. Karnthaler, F. B. Prinz, Atomic layer deposition by reaction of molecular oxygen with tetrakisdimethylamido-metal precursors, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 34 (2016) 01A138.

[133] R. Dorsmann, C. R. Kleijn, A general correction to surface reaction models based on reactive sticking coefficients, Chemical Vapor Deposition 13 (2007) 91–97.

[134] C. E. Check, T. O. Faust, J. M. Bailey, B. J. Wright, T. M. Gilbert, L. S. Sunderlin, Addition of polarization and diffuse functions to the LANL2DZ basis set for p-block elements, The Journal of Physical Chemistry A 105 (2001) 8111–8116.

[135] A. L. Garcia, B. J. Alder, Generation of the Chapman–Enskog distribution, Journal of computational physics 140 (1998) 66–70.

[136] P. J. Hay, W. R. Wadt, Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals, The Journal of Chemical Physics 82 (1) (1985) 299–310.

[137] P. J. Hay, W. R. Wadt, Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg, The Journal of Chemical Physics 82 (1) (1985) 270–283.

[138] W. R. Wadt, P. J. Hay, Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi, The Journal of Chemical Physics 82 (1) (1985) 284–298.

[139] C. Lee, D. Graves, M. Lieberman, D. Hess, Global model of plasma chemistry in a high density oxygen discharge, Journal of the Electrochemical Society 141 (1994) 1546–1555.

[140] H. C. Knoops, T. Faraz, K. Arts, W. M. Kessels, Status and prospects of plasma-assisted atomic layer deposition, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 37 (2019) 030902.

[141] S. Choi, J. Koo, H. Jeon, Y. Kim, Plasma-enhanced atomic-layer deposition of a HfO_2 gate dielectric, Journal of the Korean Physical Society 44 (2004) 35–38.

[142] P. K. Park, J.-S. Roh, B. H. Choi, S.-W. Kang, Interfacial layer properties of HfO_2 films formed by plasma-enhanced atomic layer deposition on silicon, Electrochemical and Solid-State Letters 9 (2006) F34–F37.