UNIVERSITY OF CALIFORNIA

Los Angeles

First-Principles and Multiscale Modeling for Design and Operation

of Atomic Layer Processing

A dissertation submitted in partial satisfaction of the

requirement for the degree Doctor of Philosophy

in Chemical Engineering

by

Sungil Yun

2023

ABSTRACT OF THE DISSERTATION

First-Principles and Multiscale Modeling for Design and Operation

of Atomic Layer Processing

by

Sungil Yun

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2023

Professor Panagiotis D. Christofides, Chair

There has been growing demand for high performance nanochips due to the Fourth Industrial

Revolution. However, semiconductor manufacturing leaders are struggling to meet the need due

to the current challenges and requirements, such as displacement of vertically stacked architec-

tures and stricter film quality. Despite leveraging revolutionary technology such as atomic layer

deposition (ALD), the development of more advanced techniques are necessary to address these

challenges. Recently, atomic layer etching, as a counterpart of ALD, and area-selective atomic

layer deposition, as a bottom-up nanopatterning technique, have recently gained momentum due

to their potential abilities to satisfy the stringent quality specification and assemble nanostructures

in a self-aligned manner. Thus, in this dissertation, a series of *in silico* research using multiscale computational fluid dynamics modeling have been performed to provide comprehensive understanding and insight to industry with respect to operation and robust control in order to further miniaturize the dimension of chips and precisely stack 3D features on semiconductors.

The dissertation of Sungil Yun is approved.

Carlos Morales-Guio

Samanvaya Srivastava

Laurent Pilon

Panagiotis D. Christofides, Committee Chair

University of California, Los Angeles

2023

Contents

List of Figures

xix

List of Tables

Acknowledgements

First and Foremost, I would like to praise and thank God for his unfailing love, mercy and faithful guidance to achieving this degree throughout all the challenging moments of completing this dissertation. All praises are to Jesus Christ who died for me and is risen from the death and to Holy Spirit whose wisdom and almighty power has been with me. Surely your goodness and love will follow me all the days of my life. Amen.

I would like to express my deep gratitude to Professor Christofides, my research advisor, for his valuable contributions and support. He has provided encouragement and constructive recommendations that immensely broaden my perspective. Besides my advisor, I would like to thank the doctoral committee: Professor Morales-Guio, Professor Srivastava, and Professor Pilon for their insightful comments and expertise. Further, I would like to extend my gratitude to Professor Orkoulas at Widener University for his constructive advice and inspirations. Without their guidance and persistent supervision, I could never have completed my dissertation. In addition, my special thanks are extended to my colleagues as the co-authors Matthew Tom, Henrik Wang, and Feiyang Ou who have provided insight that greatly assisted the research.

Most of all, I am forever grateful to my wife, Sharon for her enduring love and all the sacrifices that she has made on my behalf when undertaking the doctoral work. She has always provided countless love and devotion. Also, I am thankful to my parents, Hyun-sook and Young-oong and my sisters, Jieun and Sol for their endless love, unwavering support, and valuable prayers. Last but not least, my special thanks go to my beloved two children, Anna (Sojeong) and Daniel (Geon) who have also motivated and been proud of me.

Curriculum Vitae

Education

M.S. in Chemical and Biomolecular Eng., UCLA, USA   Jun 2019 - Jun 2021

GPA: 3.9/4.0, Advisor: P.D. Christofides

Thesis: Integration of feedback control and run-to-run control for plasma enhanced atomic layer deposition hafnium oxide thin films

B.S. in Chemical and Biomolecular Eng., Sogang Univ., Korea   Feb 2006 - Feb 2012

GPA: 4.17/4.5 (Summa Cum Laude)

Professional Experiences

Korea Gas Corporation, Korea   Jan 2012 - Present

Project administration for gas pipeline projects
Process engineering

Honors & Awards

Korea Gas Corporation, Korea   Sep 2019 - Jun 2021

Scholarship to earn M.S. degree at UCLA

Publications

1. *S. Yun, H. Wang, M. Tom, F. Ou, G. Orkoulas, P.D. Christofides, "Multiscale CFD Modeling of Area-Selective Atomic Layer Deposition: Application to Reactor Design and Operating Condition Calculation", Coatings, 13, 558, 2023*

2. *D. Richard, M. Tom, J. Jang, S. Yun, P.D. Christofides, "Quantifying Transport and Electrocatalytic Reaction Processes in a Gastight Rotating Cylinder Electrode Reactor via Integration of Computational Fluid Dynamics Modeling and Experiments", Electrochimica Acta, 440, 141698, 2023*

3. *M. Tom, S. Yun, H. Wang, F. Ou, G. Orkoulas, P.D. Christofides, "Machine Learning-Based Run-to-Run Control of a Spatial Thermal Atomic Layer Etching Reactor", Comp. & Chem. Eng., 1084044, 2022*

4. *S. Yun, F. Ou, H. Wang, M. Tom, G. Orkoulas, P.D. Christofides, "Atomistic-Mesoscopic Modeling of Area-Selective Thermal Atomic Layer Deposition", Chem. Eng. Res. & Des., 188, 271-286, 2022*

5. *S. Yun, M. Tom, G. Orkoulas, P.D. Christofides, "Multiscale Fluid Dynamics Modeling of Spatial Thermal Atomic Layer Etching", Comp. & Chem. Eng., 163, 107861, 2022*

6. *S. Yun, M. Tom, G. Orkoulas, P.D. Christofides, "Multivariable Run-to-Run Control of Thermal Atomic Layer Etching of Aluminum Oxide Thin Films", Chem. Eng. Res. & Des., 182, 1-12, 2022*

7. *S. Yun, M. Tom, F. Ou, G. Orkoulas, P.D. Christofides, "Multiscale Computational Fluid Dynamics Modeling of Thermal Atomic Layer Etching: Application to Chamber Configuration Design", Comp. & Chem. Eng., 161, 107757, 2022*

8. *S. Yun, M. Tom, J. Luo, G. Orkoulas, P.D. Christofides, "Microscopic and Data-Driven Modeling and Operation of Thermal Atomic Layer Etching of Aluminum Oxide Thin Films", Comp. Eng. Res. & Des., 177, 96-107*

9. *S. Yun, Y. Ding, Y. Zhang, 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, 2021*

Chapter 1

Introduction

1.1   Motivation

Atomic layer deposition (ALD) is a technique for depositing materials on wafer surfaces, which has been extremely investigated and has made a significant contribution to the production of fin field-effect transistors (FinFETs). In this method, also known as self-limiting reactions, two or three precursor pulses are introduced sequentially in between purging steps to deposit films layer-by-layer. Despite the development of the advanced deposition technique and technological innovation, the semiconductor industry has been facing several challenges. one of the challenges is that the downscaling, or shriking, of chips is limited by unwanted mobility loss and short channel effects [93], which makes it unfeasible to produce chips with fin widths of 5 *nm* or below. Another issue is that the misalignment of features in the 3D structures is a bottleneck in the miniaturization process when vertically stacking features in nanochips [71, 70, 111]. This displacement is attributed to the top-down fabrication method, which requires a series of manufacturing steps such as deposition, lithography, and etching [43]. In the process of repeating deposition and etching, the

inconsistency of stacking structures may occur, which deforms stacking features and interferes with the flow of electric currents, thereby increasing the likelihood of chip malfunctioning [78, 101]. Thus, the threshold of sub 5-*nm* and the misalignment issue must be resolved in order to satisfy the specification of quality and enhance nanochip performance,

Additionally, the Fourth Industrial Revolution's increased demand for high performance semi-conductors is placing a strain on the world's chip manufacturers. Moreover, global chip shortages are being accelerated by sectors related to autonomous vehicle, cloud computing, the internet of things (IoT), and AI (artificial intelligence). However, conventional deposition processes are time-consuming due to long purge steps to maintain self-limiting nature, which is not viable for high-volume production. Therefore, innovative technology and novel manufacturing methodology are essential to overcome current challenges.

1.2   Background

Atomic layer etching (ALE) has gained momentum as a promising method for further minia-turization, which is able to overcome the physical limit of 5 *nm* for size reduction [69]. As a coun-terpart of ALD, ALE is an etching technique in which two or three precursors are used to remove a monolayer of the substrate, resulting in ultra-conformal thin films. With the aforementioned characteristic of ALE, it is capable of precise thickness control; therefore, enabling commercial-izing gate-all-around (GAA) transistors where vertically stacked features, the so-called nanowires or nanoribbons, are surrounded by transistor gate materials. GAA transistors are recognized as next generation transistors in that they have greater power efficiency and outstanding electrostatic

properties. In order to fabricate more advanced nanochips such as GAA transistors, ALE is an essential process and has to be fully characterized.

Area-selective atomic layer deposition (ASALD) is a bottom-up fabrication technique, which does not include repeating processes of deposition, lithography, and etching steps in contrast to traditional top-down nanopatterning. In this bottom-up fabrication, films are built up solely on the target area of the wafer surface due to chemoselectivity, allowing for self-aligned assembly [5, 70]. To prevent ALD growth on the undesirable region, ASALD utilizes an inhibitor to form protective layers on the region, known as the nongrowth area (NGA). Owing to the blocking layers on the NGA, precursors selectively react and adsorb on the growth area (GA). Meanwhile, the deactivated NGA by protective layers does not allow the precursor molecules to adsorb on the surface. In this perspective, the area-selective ALD can enable self-aligned fabrication by allowing precursors to deposit only on the GA. Another benefit is that this technique is able to reduce production costs and minimize hazardous chemicals by eliminating lithography and etching steps [30, 98]. It is of importance to select appropriate inhibitors to accomplish successful area-selective ALD. Among several inhibitors, small-molecule inhibitors (SMIs) have been proposed to improve deposition selectivity for the following reasons [113, 74]: SMIs can be readily integrated to various vapor-phase processes in that many of most manufacturing processes are based on vapor phase. The desorption of SMIs during other deposition steps is relatively insignificant. Even if the protective layers degrade, the selectivity can be preserved by adding a fixing step.

Despite the benefits and technical viability of ALE and ASALD processes mentioned above, there has been lack of research on these subjects. In addition, it is challenging to understand the reaction mechanisms at the atomistic level and the effects of fluid dynamics on the processes

3

via *in vitro* studies due to the fact that in situ monitoring and measurements are restricted, and experimental procedures are time and cost-consuming. In the meantime, *in silico* research can measure numerous properties and provide high-fidelity multiphysics data in a cost-effective manner. Moreover, numerical analysis of computational fluid dynamics (CFD) modeling can enhance its merits by incorporating additional modeling at different scales. Thus, this dissertation employs a multiscale modeling approach to characterize reaction mechanisms and fluid dynamics. First, an atomistic-mesoscopic model is developed for surface kinetics of ALE and ASLAD using a kinetic Monte Carlo (kMC) method. Next, diverse reactor designs for cross-flow type reactors and spatial reactors are constructed to investigate the fluid dynamics effects on film quality, which is simulated in Ansys Fluent. Then, the developed atomistic-mesoscopic model is integrated with the macroscopic model with different reactor configurations to build multiscale models. Finally, the developed multiscale models are used to study operation and control for those processes.

1.3   Dissertation Objectives and Structure

This dissertation presents the integration of atomistic-mesoscopic modeling to macroscopic modeling using computational fluid dynamics (CFD) to characterize both ALE (atomic layer etching) and ASALD (area-selective atomic layer deposition) and provide understanding of design and operation of the processes, and offer new insights for further industrialization. The objectives of this dissertation are summarized as follows:

1. To develop an atomistic-mesoscopic surface model for atomic layer etching (ALE) of Al_2O_3 films and area-selective atomic layer deposition (ASALD) of SiO_2 films by using a kinetic

Monte Carlo (kMC) model and a 2-D stochastic surface model in conjunction with a data-driven model.

2. To construct and optimize a variety of reactor designs using multiphysics numerical modeling.

3. To present a multiscale modeling framework by integrating aforementioned models for diverse applications to provide an insight for optimizing operation and process, and to formulate robust run-to-run control systems using various regression models and algorithms.

The structure of this work is as follows: **Chapter 2** discusses the formulation of the surface model for ALE of Al_2O_3 thin films and the data-drive model, including the electronic structure calculations, known as *ab initio* DFT (density functional theory). **Chapter 3** describes the multiscale numerical modeling by combining the developed surface model to collect the flow patterns and coverage profiles and to optimize process operation. **Chapter 4** provides the multivariable run-to-run (R2R) control system, which is constructed by two individual R2R control algorithms. A novel approach to enhance the control performance is presented and validated from the multiscale simulations of the integrated model for ALE in Chapter 3. **Chapter 5** covers the discussion of the sheet-to-sheet spatial reactor design employing dynamic mesh to capture the motion of the substrate. The key parameters that have great impacts on film quality such as gap distance, substrate velocity, and feed flow rate are explored in great detail. **Chapter 6** and **Chapter 7** elaborate area-selective atomic layer deposition (ASALD) of SiO_2 thin films. Specifically, the atomistic-mesoscopic surface model using a kMC method, similar to ALE in Chapter 2, is illustrated in Chapter 6. Also, a 2D surface model using Van der Waals radii is simulated to characterize the

protective layer on Al_2O_3, which is validated with experimental references. A rotary spatial atomic

layer deposition (SALD) is optimized to offer better film uniformity and high-throughput. Finally,

Chapter 8 provides a summary of the results of the dissertation.

Chapter 2

Microscopic and Data-Driven Modeling and Operation of Thermal Atomic Layer Etching of Aluminum Oxide Thin Films

2.1   Introduction

In this work, microscopic and data-driven models are developed for the thermal ALE process of aluminum oxide thin films. First, DFT (Density Functional Theory)-based electronic structure calculations are conducted using Quantum ESPRESSO (QE), which contains a database of pseudopotentials to obtain the activation energies for species not published in the literature. Then, these activation energies are used as input parameters in a microscopic model using the kinetic Monte Carlo (kMC) algorithm, which provides the progression of the etching process over time. The developed microscopic model is validated via comparison with available experimental results [61]. Nevertheless, the kMC simulation is a time-consuming task so its use in generating the

7

etch time data set over a wide range of operating conditions may be ineffective for real-time cal-culations. Therefore, a feed-forward artificial neural network (FNN)-based deep learning model is established to determine the relationship between operating conditions and the etch time by using training data sets generated from the kMC simulation. FNNs have recently been utilized to cor-relate and fit the results of kMC data. [25] utilized a feed-forward Bayesian regularized artificial neural network (BRANN) to characterize the relationship between input, temperature and pres-sure, and output, half-cycle time, using kMC data for plasma enhanced atomic layer deposition of HfO_2. [24] utilized both FNN and BRANN to fit kMC data produced by simulating thermal atomic layer deposition of SiO_2. Therefore, the FNN model is used to identify the optimal operating con-ditions for the thermal ALE of aluminum oxide. The data-driven FNN model offers significant economic benefits since it is costly to conduct many experiments over a wide range of operating conditions to determine the optimal operating conditions.

2.2   DFT (Density Functional Theory) Calculation

In order to build a microscopic model for the thermal ALE of aluminum oxide, all of the pos-sible reaction pathways and their kinetic parameters should be listed and determined. However, it is a time-consuming and practically unnecessary task to simulate the entirety of the intermediate reactions involved and these parameters are not known in general and must be calculated using theoretical concepts and simulation. To simplify the reaction network, critical reaction steps that have slow kinetics and low thermodynamic spontaneity are chosen to define the calculated process time while reactions that occur spontaneously are neglected. Then, this work models crystal struc-

tures of Al_2O_3, which are approximated by developing their 3D lattice models in Python. Next, the structural optimization of potential geometries, which are segregated from unstable structures, is performed based on the DFT (Density Functional Theory) method. The open-source computation package, Quantum ESPRESSO (QE), is used to optimize the geometries of the molecular structures and to perform other electronic structure calculations that are pertinent to computing the thermodynamic properties of the molecular structure. QE has been widely used to compute fundamental quantum properties on the basis of first-principles DFT calculations [88]. With the optimized structures computed from QE, reaction mechanisms are identified and established. Finally, the nudged elastic band (NEB) method is introduced to compute the activation energies that are later substituted into the kinetic Monte Carlo (kMC) algorithm, described in Section 2.3, to model the etching process.

2.2.1   Aluminum Oxide Lattice Modeling

Building a lattice structure plays an important role in atomistic modeling because kinetic parameters and thermodynamic properties vary in accordance with crystal structures and their orientations. Therefore, it is of paramount importance to build a suitable lattice model at the atomistic level. There are a number of crystal structures of aluminum oxide [92], which can be found in different conditions. Among them, θ-Al_2O_3 ($\bar{2}$ 0 1) was found to grow along the ($\bar{2}$ 0 1) orientation stacked on Si (1 0 0) through the atomic layer deposition (ALD) process under annealing [12]. Thus, θ-Al_2O_3 ($\bar{2}$ 0 1), which is designated as A0 in this study, is used for atomistic modeling as shown in Table 2.2. BURAI, a GUI (Graphical User Interface) system of the QE package, is applied to visualize the geometries studied in this work. Table 2.1 provides the lattice

9

Table 2.1: Lattice parameters of monoclinic θ-Al$_2$O$_3$.

Lattice parameter	Experimental value
a (Å)	11.85
b (Å)	2.90
c (Å)	5.62
α (°)	90.00
β (°)	103.83
γ (°)	90.00

parameters for monoclinic θ-Al$_2$O$_3$ from experimental results [124].

2.2.2 SCF (Self-Consistent Field) Calculation

The electronic structure calculation is performed using pseudopotentials that are built in the QE simulation packages. The pseudopotentials assume that the core ions are frozen and only the valence electrons are involved in the potentials. In other words, the pseudopotentials dramatically decrease the number of plane waves so that the Schrödinger equation becomes solvable. The time-dependent Schrödinger equation is defined as:

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian, Ψ is the wave function, and E is the energy of the system. The Hamiltonian is described as follows:

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V} \tag{2.2}$$

where $\hbar$ is the the Planck constant divided by 2π, m is the mass of the particle, ∇^2 is the Laplacian operator in Cartesian coordinates, and $\hat{V}$ is the potential energy. There are three pseudopotentials in general: Norm-conserving (NC), Ultrasoft (US), and Projector Augmented Wave (PAW) pseudopotentials. In this work, PAW pseudopotentials are used for their computational efficiency [88].

It is necessary to perform the self-consistent field (SCF) calculation to obtain quantum parameters that can facilitate accurate calculation results, such as the kinetic energy cut-off (ecutwfc), the kinetic energy cut-off for charge density and potential (ecutrho), and the k-points. First, the optimization of the kinetic energy cut-off (ecutwfc) is performed to limit the number of plane waves for computational efficiency. Figure 2.1a shows the convergence of the total energy at various ecutwfc. The total energy decreases as the ecutwfc increases. However, the total energy does not change when ecutwfc is greater than 50. Next, the optimization of the kinetic energy cut-off for charge density and potential (ecutrho) is computed as shown in Figure 2.1b. Unlike the ecutwfc, the total energy increases as the ecutrho increases. The total energy does not change significantly when ecutrho is greater than 200. Lastly, the total energy calculations are carried out with different k-points. The convergence of the total energy is achieved when the number of k-points is 4 as shown in Figure 2.1c. Thus, a kinetic energy cut-off (ecutwfc) of 50, a kinetic energy cut-off for charge density and potential (ecutrho) of 200, and a k-point value of 4 are used as the input parameters for all the electronic structure calculations in this work.

2.2.3 Structural Optimization and Reaction Mechanisms

Due to the difficulty of modeling the entirety of the intermediate reactions, the reaction mechanisms for Steps A and B can be selected based on the critical reaction paths that have longer

11
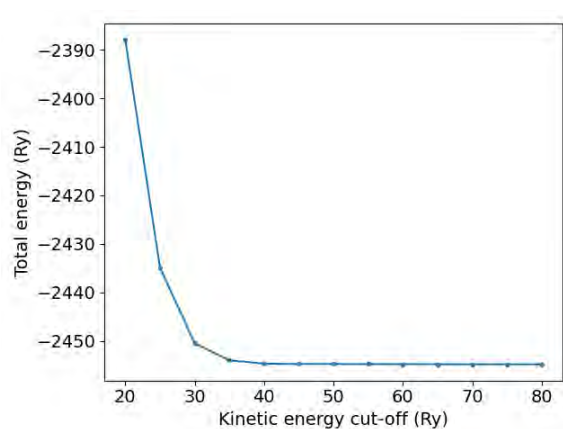
(a) The kinetic energy cut-off (ecutwfc) optimization of θ-Al$_2$O$_3$ ($\bar{2}$ 0 1).
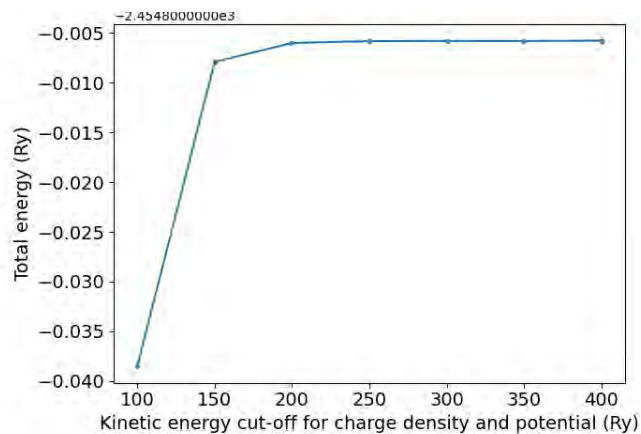
(b) The kinetic energy cut-off for charge density and potential (ecutrho) optimization of θ-Al$_2$O$_3$ ($\bar{2}$ 0 1).
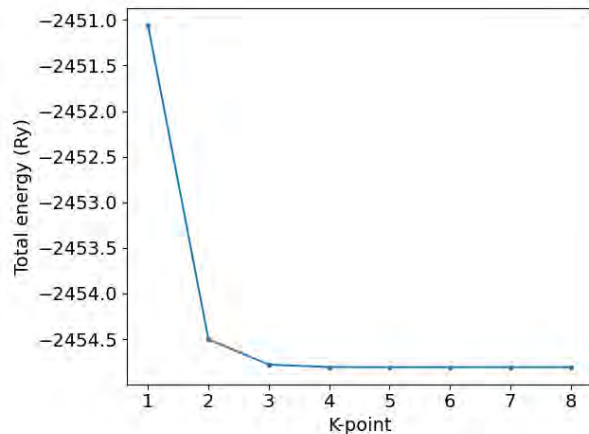
(c) The k-point optimization of θ-Al$_2$O$_3$ ($\bar{2}$ 0 1).

Figure 2.1: The total energy converges when ecutwfc is 50 (a), ecutrho is 200 (b), and the number of k-points is 4 (c), respectively.

reaction times, which are directly proportional to the reaction rate constant and hence the activation energy. An overall process time will essentially be dependent on slow reaction paths rather than fast reaction paths. Thus, neglecting the reaction paths that occur at a fast or instantaneous rate has negligible effect on the overall process time. Therefore, it is essential to study critical reaction paths for the kMC simulation. Finding the critical paths, which affect the overall process

time, can be achieved through structural optimization. If a structure can be optimized from the electronic structure calculation, the optimized structure can be assumed to be at a minimum along with the reaction path, which implies that the reaction for the formation of the structure is stable and relatively slow enough to influence the overall reaction time. On the other hand, if a structure is not optimized, it can be assumed that this structure is thermodynamically unstable and is essentially nonexistent such that this reaction path may not be considered as a critical path.

Quantum ESPRESSO (QE) also provides various packages, including the PWscf (Plane-Wave self-consistent field) program, to optimize crystal structures using DFT. The lattice parameters of θ-Al_2O_3 in Table 2.1 and the parameters computed from scf calculation in Section 2.2.2 are applied in PWscf calculation as input parameters. The results from the PWscf calculation for possible crystal structures are optimized and reaction mechanisms are proposed, which are shown in Tables 2.2 and 2.3

The first half-cycle, hereby referred to as Step A, for the etching of Al_2O_3 involves the addition of gaseous HF, which binds to the surface Al and O atoms [38]. This crucial step involves the addition of a reactive species that prevents further permeation beyond the surface of the substrate; thus, the process is said to experience transport-limited phenomena [52]. Following the modification step, the second half-cycle, hereby referred to as Step B, utilizes a secondary precursor to convert the surface from Step A into a volatile species. For the Al_2O_3, trimethylaluminum (TMA) is commonly used to convert AlF_3 into the volatile species, dimethylaluminum fluoride (DMAF) [38]. It has been proposed that there is a possibility that dissociative adsorption exists of HF, hydrogen diffusion, and surface reaction including the desorption of H_2O in the reaction mechanism of Step A [80]. In this paper, the surface reaction is divided into a surface reaction to form

an H_2O molecule and a desorption of the H_2O, which is obtained from the structural optimization. The reaction mechanism is summarized below and visualized in Table 2.2.

(1) A0 $\leftrightarrow$ A1: An HF molecule is dissociatively adsorbed onto the substrate in which the H atom is attached to an O atom and the F atom binds to an Al atom. The reverse reaction might occur but can be negligible due to the order of magnitude difference in the forward and reverse reaction rates.

(2) A1 $\leftrightarrow$ A2: The adsorbed H atom can freely diffuse on the substrate as the neighboring site is vacant.

(3) A3 $\leftrightarrow$ A4: When two H atoms are neighboring each other, one of the two H atoms can shift and react with the adjacent OH group, thus forming an H_2O molecule. For the reverse reaction, the H_2O molecule attached on the surface can also decompose and synthesize to the former state.

(4) A4 $\rightarrow$ A5: Finally, the H_2O molecule can be released. The reverse reaction might be possible; however, it is assumed that the reverse reaction does not occur for the existence of the stop condition specified in the simulation.

Table 2.2: Reaction mechanism for Step A.

Reaction pathway		E_a (eV)	
		Forward	Reverse
A0 $\Longleftrightarrow$ A1		NA*	2.02
A1 $\Longleftrightarrow$ A2		0.98	0.71
A3 $\Longleftrightarrow$ A4		1.28	0.76

Continued on next page

Reaction pathway		E_a (eV)	
		Forward	Reverse
A4 $\Longrightarrow$ A5		0.88	–

*The rate constants for the adsorption and physisorption is calculated by Collision Theory.
**Red, gray, green, and white atoms represent O, Al, F, and H, respectively.

The proposed mechanism for Step B involves the conversion of the AlF_3 surface into a volatile layer. It has been proposed to use trimethylaluminum (TMA) as the secondary precursor that modifies the AlF_3 into the volatile species, dimethylaluminum fluoride (DMAF) [60]. Despite this proposal, detailed mechanisms, especially reaction intermediates, are not known in a kinetic manner. Thus, this work proposes the reaction mechanisms for Step B in detail by comparing the thermodynamic feasibility of the intermediate molecules developed from QE in Section 2.2.2. There are two reaction sites (aluminum fluorides) in the unit cell of B0, which are visualized in Table 2.3. The left AlF_2 is referred to as l-AlF_2 and the right AlF_3 is referred to as r-AlF_3. A proposed mechanism for Step B is discussed below. The proposal is elaborated with l-AlF_2 first (from molecules B1 to B3) and followed by r-AlF_3 (from molecules B4 to B7).

(1) B0 $\leftrightarrow$ B1: Initially, a molecule of TMA binds to an F atom on one of the two reaction sites. It is also noted that l-AlF_2 is dissociated from one of the F atoms, which becomes attached

16

to the neighboring Al atom. Here, it is assumed that the binding is non-regiospecific, that is to say, no adsorption site (left or right) predominates during the first TMA adsorption. As a result, B0 ↔ B4 has an equal probability of occurring initially.

(2) B1 → B2: The attached TMA molecule on the l-AlF$_2$ site undergoes a ligand-exchange that occurs with an F atom and a CH$_3$ molecule being exchanged. Then, the product, DMAF, is removed from the surface. It is reasonably assumed that DMAF is extremely volatile that the reverse reaction is negligible.

(3) B2 → B3: The second adsorption of TMA on the l-AlF$_2$ site can occur. Unlike B0 ↔ B1, The physisorbed structure is not able to be optimized due to its electronic instability. Thus, it can be assumed that the adsorption of TMA and the desorption of DMAF occur instantaneously. The Al(CH$_3$)$_2$ molecule on the l-AlF$_2$ site of B3 cannot be detached without binding to the neighboring F atom. However, according to the electronic structure calculation, the Al(CH$_3$)$_2$ molecule is not able to react with the F atom in the presence of the r-AlF$_3$ site. Therefore, the Al(CH$_3$)$_2$ molecule can combine with the neighboring F atom and then, can be withdrawn in the same way as B6 → B7 until the r-AlF$_3$ is removed.

(4) B0 ↔ B4: A molecule of TMA binds to an F atom on r-AlF$_3$ site. The attached TMA can also be desorbed on the surface. However, it can also be negligible due to the order of magnitude difference between the forward and reverse reaction rates.

(5) B4 → B5: The TMA and AlF$_3$ undergo a ligand-exchange, and the volatile leaving group, DMAF, is formed. It is also assumed that DMAFs are extremely volatile that the reverse

17
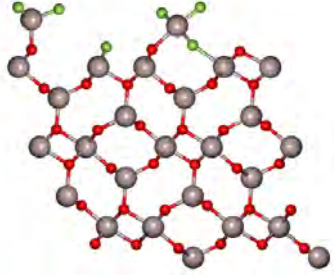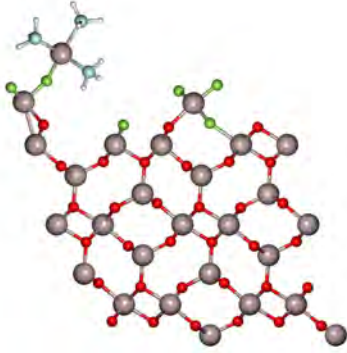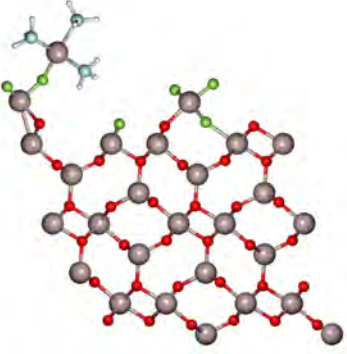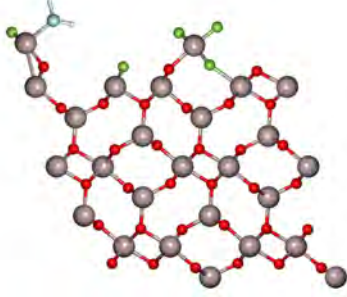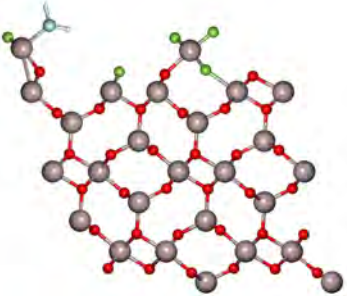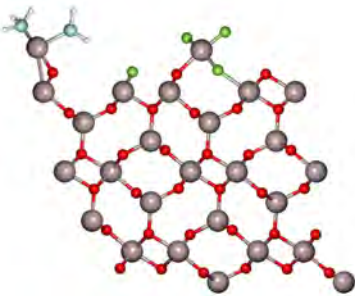
reaction does not occur in the same manner as B1 $\rightarrow$ B2.

(6) B5 $\rightarrow$ B6: The second adsorption of TMA occurs and the product, DMAF (from TMA), is released immediately in the same manner as B2 $\rightarrow$ B3.

(7) B6 $\rightarrow$ B7: By increasing the temperature, the volatile DMAF molecule bound to the substrate is desorbed and an Al atom is removed from the Al_2O_3 substrate. The neighboring l-F atom attaches to the l-AlF$_2$ site to form l-AlF$_3$ so that the l-AlF$_3$ can be removed through further reactions. This reaction mechanism is the propagation step that will contribute to l-AlF$_3$ removal if TMA is adsorbed to the l-AlF$_2$ before r-AlF$_3$ is removed. Thus, additional species in place of the l-AlF$_3$ site include AlF(CH$_3$), AlF$_2$(CH$_3$), Al(CH$_3$)$_2$, and AlF(CH$_3$)$_2$. In Table 2.3, molecule B7 generalizes the latter list of possible molecules in place of l-AlF$_3$. It is assumed that the left and right AlF$_3$ reactions proceed in multiple pathways that are randomized using the kMC method, which is described in Section 2.3, instead of a sequential reaction path that may be implied in Table 2.3.
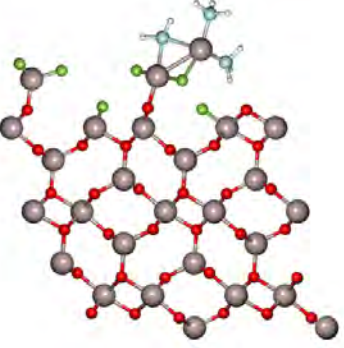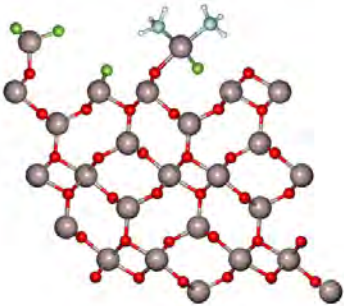
Table 2.3: Reaction mechanism for Step B.

Reaction pathway	E_a (eV)	
	Forward	Reverse

Continued on next page
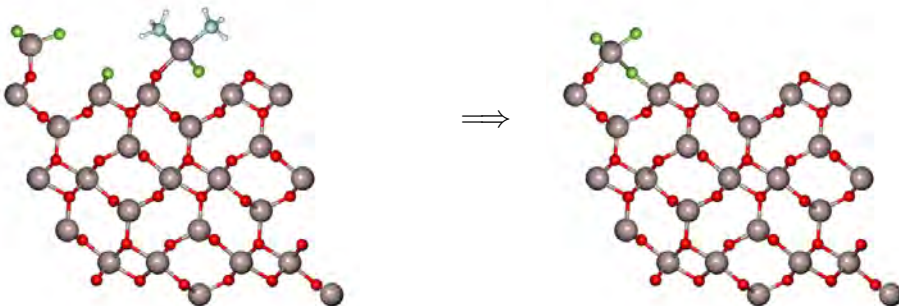
Reaction pathway		E_a (eV)	
		Forward	Reverse
B0 $\iff$ B1		NA*	42.27
B1 $\implies$ B2		1.45	–
B2 $\implies$ B3		NA*	–

19

Reaction pathway		E_a (eV)	
		Forward	Reverse
B0 $\Longleftrightarrow$ B4		NA*	21.29
B4 $\Longrightarrow$ B5		0.82	–
B5 $\Longrightarrow$ B6		NA*	–

Reaction pathway	E_a (eV)	
	Forward	Reverse
B6 $\implies$ B7	1.12	–

*The rate constants for the adsorption and physisorption is calculated by Collision Theory.
**Red, gray, green, blue, and white atoms represent O, Al, F, C, and H, respectively.

2.2.4   NEB (Nudged Elastic Band) Calculation

To build the kinetic Monte Carlo (kMC) based microscopic model, all of the rate constants involved in the reaction mechanisms should be provided. To be specific, the activation energies are required so that the rate constants can be derived from the Arrhenius equation or other kinetic theories. The nudged elastic band (NEB) calculation is carried out to compute the activation energies, which is enabled by the PWneb (Plan-Wave nudged elastic band) package built in QE. The NEB method is widely used for calculating reaction pathways and activation energies [102]. In addition to the parameters used for the PWscf calculations in Section 2.2.2, when optimized images of the initial and final structures are known in Section 2.2.3, the NEB method calculates activation energies by determining a minimum-energy path (MEP) between the reactants and products. The computed activation energies for each reaction path are shown in Tables 2.2 and 2.3.

The rate constants can be calculated by Collision Theory and Transition State Theory-based Arrehenius equations. For adsorption reactions, the rate constant can be calculated by Collision Theory:

$$k_{ads} = \frac{PA_{site}}{\sqrt{2\pi m k_B T}} \tag{2.3}$$

where P is the partial pressure of the precursor, A_{site} is the area of the single site, m is the mass of the atom or the molecule, k_B is the Boltzmann constant, and T is the temperature. However, an HF is dissociatively adsorbed on the reaction site. The rate constant for dissociative adsorption is expressed as below:

$$k_{d,ads} = \frac{2PA_{site}\sigma}{Z\sqrt{2\pi m k_B T}} \tag{2.4}$$

where Z is the coordination number, and σ is the sticking coefficient. The sticking coefficients of HF and TMA are given as 0.15 [33] and 0.02 [97], respectively.

For the desorption, the surface, and the diffusion reactions, the rate constants can be given by the Arrehenius equation:

$$k = \nu \exp\left(\frac{-E_a}{RT}\right) \tag{2.5}$$

where ν is the pre-exponential factor and E_a is the activation energy. The pre-exponential factor is defined by Transition-State Theory as follows:

$$\nu = \frac{k_B T}{h}\frac{Q^{\ddagger}}{Q} \tag{2.6}$$

where h is the Planck constant, $Q^{\ddagger}$ is the partition function of the transition state, and Q is the partition function of the reactant. Though the pre-exponential factor is calculated from the par-

tition functions of the reactants, the ratio of the partition functions can be approximated as 1 for simplicity [48]. Thus, this approximation is applied to this work.

2.3   Kinetic Monte Carlo Simulation of Etching Process

The kinetic Monte Carlo (kMC) is a computational sampling work in the basis of the randomness, which has been widely used in various fields. The computational works for the atomic layer deposition (ALD) have been widely performed in different materials [24, 112, 114] and the performance of the kMC has been approved. There are various kMC algorithms such as the variable step size method (VSSM), the random selection method (RSM), and the first reaction method (FRM) [48]. Among them, the VSSM has been widely used for the kMC simulation, which was developed by Bortz, Kalos, and Lebowitz (so-called BKL). The kMC simulation for the thermal ALE process of Al_2O_3 using VSSM computes the progress of the etch process over time and the etch time where a single layer of the aluminum oxide is completely removed. The kMC simulation is performed according to the following procedure:

- First, a list of all possible reaction paths across all the reaction sites is prepared so that the total of the rate constants, k_{total}, is calculated as the sum of all the rate constants.

$$k_{total} = \sum_{i=1}^{N} k_i \tag{2.7}$$

   where k_i is the rate constant of the reaction i, and N is the number of the possible reaction paths.

- Then, a reaction path is chosen at a single reaction site using the algorithm defined as:

$$\sum_{i=1}^{q-1} k_i \leq \gamma_1 k_{total} \leq \sum_{i=1}^{q} k_i \tag{2.8}$$

where q represents the reaction, q, and $\gamma_1 \in (0, 1]$ is the first random number for the reaction selection. The reaction selection is performed at each reaction site with a random number that is generated for every reaction site. If the value of $\gamma_1 k_{total}$ falls between $\sum_{i=1}^{q-1} k_i$ and $\sum_{i=1}^{q} k_i$, reaction q is selected for the reaction site.

- When the reaction selection is completed across the reaction sites, a time interval is computed as follows:

$$\Delta t = \frac{-\ln \gamma_2}{k_{total}} \tag{2.9}$$

where γ_2 is the second random number for the time progression.

- Finally, the system time changes, $t \rightarrow t + \Delta t$.

The system time of the VSSM (Variable Step Size Method) does not depend on the lattice size. However, it is governed by the sum of the rate constants with respect to all of the possible reaction paths in the system. After the simulation for all the reaction sites in the system is performed, the k_{total} is updated. In other words, the disabled reaction paths are removed from the list of the reactions, and the newly enabled reactions are included in the list. If the k_{total} is updated, the system repeats the aforementioned simulation steps until the stop condition (i.e., full etching) is fulfilled.

2.4 Feed-Forward Artificial Neural Network Model

The half-cycle time is the most critical consideration when designing an industrial ALE process. The kMC model, initially proposed by [10], received increasing attention in molecular dynamic simulation due to its capability of generating a probabilistic model to mimic the randomness of molecular movement in the natural reaction. Therefore, the kMC simulation can provide a reliable reference to the half-cycle time. However, the kMC simulation is highly computationally demanding, especially for the process invoking large-scale particles and multi-step reactions, and on the other hand, it is a discontinuous open-form simulation. Thus, it may not be utilized directly as the supervising model for both the process operation and experimental decision-making tasks. A classical regression model, such as a polynomial regression model, may be considered to characterize the relationship between the inputs (pressure, temperature) and output (half-cycle time). Nevertheless, it may not be applicable due to the non-linearity and the randomness caused by the kMC simulation. Thus, significant efforts may be required for the classical regression analysis. On the other hand, a feed-forward artificial neural network (FNN) model, as one of the robust data-driven deep learning models, can be trained easily regardless of the non-linearity while preserving fidelity and accuracy. Thus, FNN models are formulated in this work.

An artificial neural network (NN) is designed to portray a biological neuron, which is capable of perceiving and understanding the components that define objects, patterns, or concepts. From a mathematical perspective, artificial NNs collect and analyze data and then perform a regression analysis on the data to process their behavior at various conditions. The structure of the NN for this thermal ALE process with temperature and pressure defined as inputs to the NN in order to study

their effect on the output, cycle time. The FNN model can be expressed in the following form:

$$Y = F_{NN}(X) = \begin{cases} h_j^{[1]} & = \sigma^{[1]}(\sum_{i=1}^{q} \omega_{ji}^{[1]} x_i + b^{[1]}) \\ h_j^{[2]} & = \sigma^{[2]}(\sum_{i=1}^{q} \omega_{ji}^{[2]} h_i^{[1]} + b^{[2]}) \\ y_j & = \sigma^{[o]}(\sum_{i=1}^{q} \omega_{ji}^{[o]} h_i^{[o]} + b^{[o]}) \end{cases} \qquad (2.10)$$

where $b^{[k]}$ and $\omega_{ji}^{[k]}$ are the biases and weights, respectively, connecting the ith input from the prior layer to the jth neuron in the kth layer, $j = 1, \ldots, q$, $k = 1, \ldots, o$. The notation q is the number of neurons in the kth layer and it is not necessary to be constant for each layer, and o stands for the output layer. $X = [x_1, \ldots, x_n] \in \mathbf{R}^n$ and $Y = [y_1, \ldots, y_m] \in \mathbf{R}^m$ are the input vectors containing the operating states of the process and the corresponding output vector. $\sigma^{[k]}(\cdot)$ denotes the activation function taking outputs from the prior layer in the calculation of hidden neurons $h_j^{[k]}$ in the kth layer.

In this work, as shown in Figure 2.2, two two-input-single-output FNN models are developed with two hidden layers to apply to Steps A and B for nonlinear regression, respectively. Tensor-Flow's Keras, which is an API (Application Programming Interface) widely used to build and train deep-learning models, is used to build the FNN models. Both FNN models have two neurons in the input layer and one neuron in the output layer representing the pressure and temperature as the causation and the half-cycle time as the consequent. 80 % of the data points of the kMC simulation is chosen randomly for training the FNN models while the remaining kMC data points are used for the model evaluation. The models for Steps A and B contain 50 neurons each in the first and second hidden layers, which were selected due to the complexity of their reaction mechanisms.

The exponential linear unit (ELU) equation is utilized for all layers, as defined in Equation
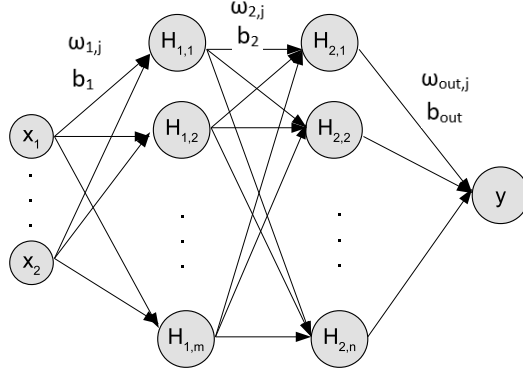
Figure 2.2: Two-input-one-output FNN model with two hidden layers.

(2.11), in order for the activation function to have effective gradient propagation and smoother

prediction.

$$
ELU(z) = \begin{cases} z & \text{for} \quad z > 0 \\ \alpha(e^z - 1) & \text{for} \quad z <= 0 \end{cases}
\tag{2.11}
$$

where α is a positive constant. Subsequently, the FNN is trained in terms of minimizing the

following mean square error function:

$$
Loss(X, \Omega) = \frac{1}{N_D} \sum_{i=1}^{N_D} [\hat{y} - F_{NN}(X, \Omega)]^2
\tag{2.12}
$$

where $\Omega = [\omega_{11}^{[1]}, \ldots, \omega_{q\hat{q}}^{[o]}, b^{[1]}, \ldots, b^{[o]}]$ is the weight vector containing all the weights and biases to

be optimized. N_D denotes the number of data points in the training data set, and $\hat{y}$ is the reference

output value. The optimum weight vectors Ω^* are calculated by adopting the "Adam" optimizer.

2.5 Simulation Results

2.5.1 Validation of Microscopic Kinetic Monte-Carlo Model

A number of the kMC simulations with the lattice size of 50×50 through 1500×1500 were performed to investigate their dependence on lattice size. There was no significant disparity among different-sized lattice models, which is supported by [45]. In this work, the 300×300 lattice is applied to the microscopic model. Figure 2.3 plots the lattice at different times in the course of the simulation, which visualizes how the etch process takes place on the reaction sites. The lattice simulation begins with the blue color, which represents the surface of the aluminum oxide substrate. As the reaction progresses, the color of each reaction site turns yellow, which indicates that the surface is fluorinated. Eventually, the lattice completely turns yellow, which indicates full fluorine coverage and that the half-cycle is complete.
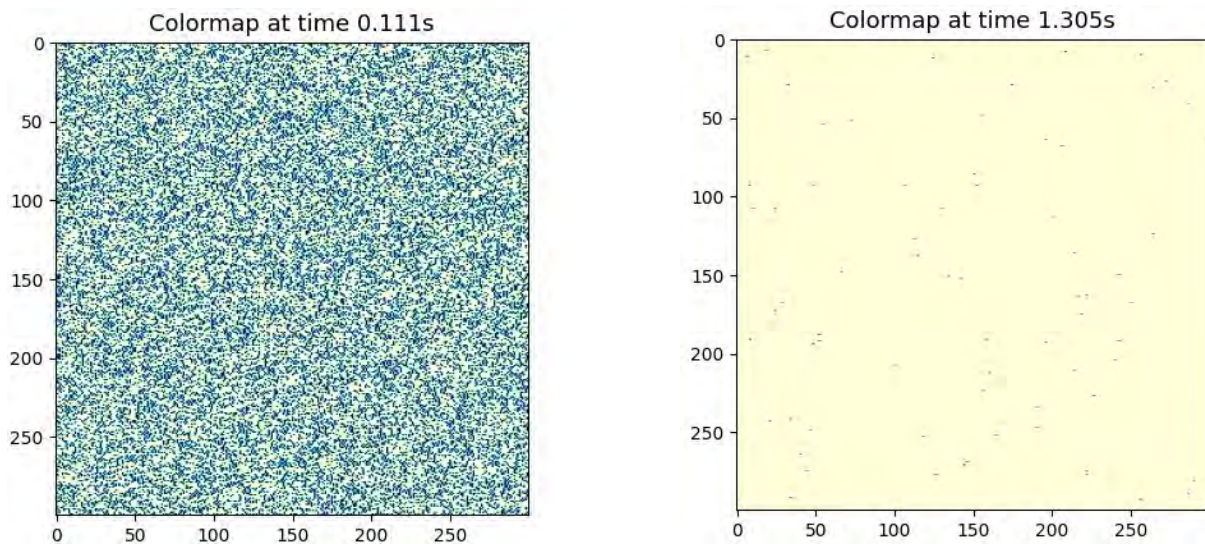
Figure 2.3: Film structure from the kMC simulations for Step A. The simulation begins with blue colors (unetched film). When the simulation progresses, a reaction site that is etched turns yellow. The left map shows the lattice at $0.111\ s$ and the right one shows the lattice at $1.305\ s$.

The progression of the half-cycle over time for both steps is shown in Figure 2.4. A monolayer of Al_2O_3 is fully fluorinated as AlF_3 at 1.38 s, and then, the modified thin film of AlF_3 from Step A is completely etched at 2.38 s. The microscopic model of this work can be validated by the experimental results [61]. From the results of Lee's work, nearly self-limiting behavior was observed after 1.0 s of HF exposure for Step A, and then there was no significant mass change of the substrate after 1.5 s. The half-cycle time of 1.38 s calculated from this kMC simulation lies within this range of times. For Step B, the experimental results indicate that there is no considerable mass change of the substrate after 2.25 s of TMA exposure, and thus, it can be regarded that the fluorinated layer is fully etched at 2.25 s. Likewise, the half-cycle time of 2.38 s from the kMC simulation is comparable to the experimental half-cycle time of 2.25 s. Therefore, it is demonstrated that the developed microscopic model successfully characterizes the thermal ALE process of aluminum oxide.
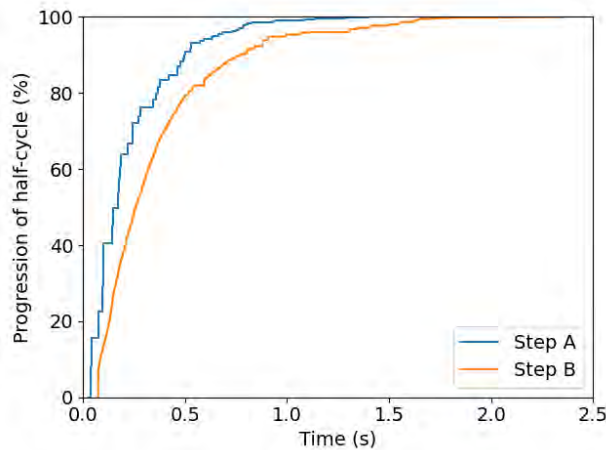
Figure 2.4: Reaction progression of half-cycles for both steps over time with an operating temperature of 573 K. The full progression is achieved at 1.38 s for Step A and 2.38 s for Step B, respectively.

The developed microscopic model does not provide the etch per cycle (EPC). Nevertheless, it

is capable of plotting the thickness of the wafer over the number of cycles by using the experimental data. The EPC was observed as 0.46 Å/cycle [38], which is employed to graph the thickness over the number of cycles as shown in Figure 2.5. The initial thickness is assumed to be 100 Å. The microscopic model can also provide the mass change of each half-cycle. Figure 2.6a provides the mass change over time for 3 cycles in which the mass change of 13 ng/cm^2 and -29 ng/cm^2 are applied for HF and TMA, respectively. The purge time is set to 30 s for Steps A and B. During Step A, the mass change increases as 3 oxygen atoms are released, meanwhile, 6 fluorine atoms are attached onto the surface. After the purge time of 30 s, a AlF$_3$ monolayer is etched, thus leading to the reduction in the mass change. Figure 2.6b reveals that the lower the temperature, the longer it would take for 3 cycles to run. These results are expected because the rate constants are calculated by the Arrhenius equation.
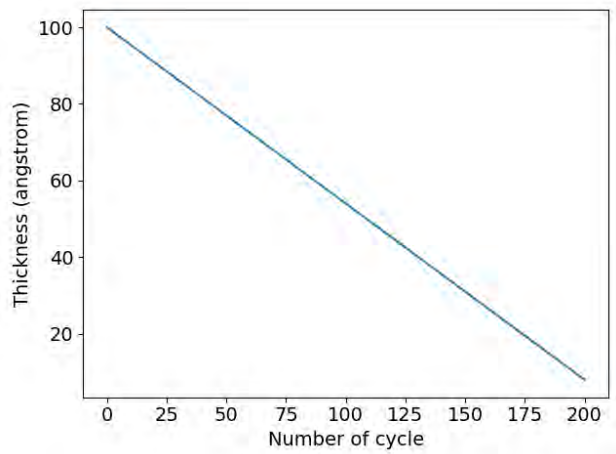
Figure 2.5: Film thickness with respect to number of cycles. It is assumed that the initial thickness is 100 $Å$. An etch thickness per cycle of 0.46 $Å/cycle$ is applied.
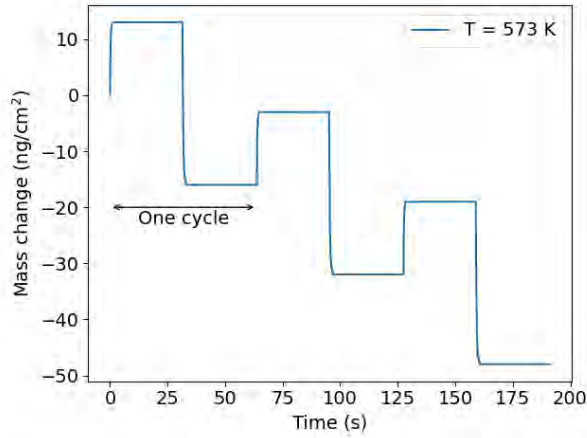
(a) Mass change over time at T = 573 K.
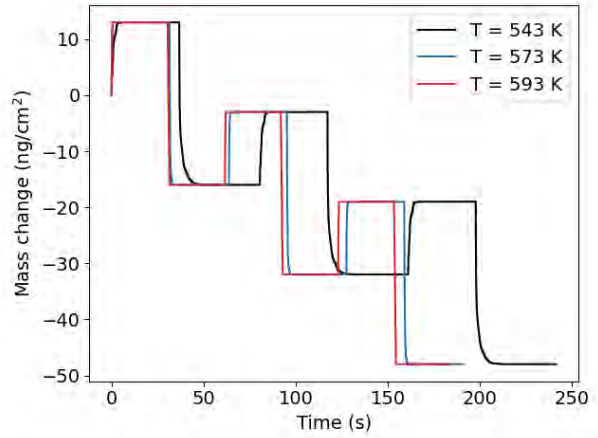
(b) Mass change over time for three different temperatures.

Figure 2.6: Mass change over time during 3 cycles with purge time of 30 s.

2.5.2 FNN Model Validation and Prediction

In addition to training a data-driven deep learning model, data sets should be generated for various operating conditions. As shown in Figures 2.7a and 2.7b, the kMC model is executed with a pressure range of 10 to 100 Pa in intervals of 5 Pa and with a temperature range of 533 to 598 K in intervals of 5 K. For generating data sets, the stop condition of the kMC simulation is set to 99.9 % of the full etch progression. The kMC computational time for a single data point (i.e., half-cycle time) depends on the input parameters, but it took about 20 minutes on average with 64 GB memory. The data sets were collected from an average of 10 simulations for each operating condition so that the randomness of the kMC simulation yields a lesser effect on the results. The data sets for both half-cycles were generated and were employed for the training and validation of the feed-forward artificial neural network (FNN) model discussed in Section 2.4. With respect to the training of the FNN models, it took less than 5 minutes on average with the same processor for the kMC simulation. However, once the model was trained, it took less than a second to predict a

31

data point, which is sufficient for real-time operational parameter calculation since a batch run for

a half-cycle typically takes less than 5 seconds in the semiconductor industry. In other words, the

FNN models can be used for real-time operational control for the semiconductor industry.
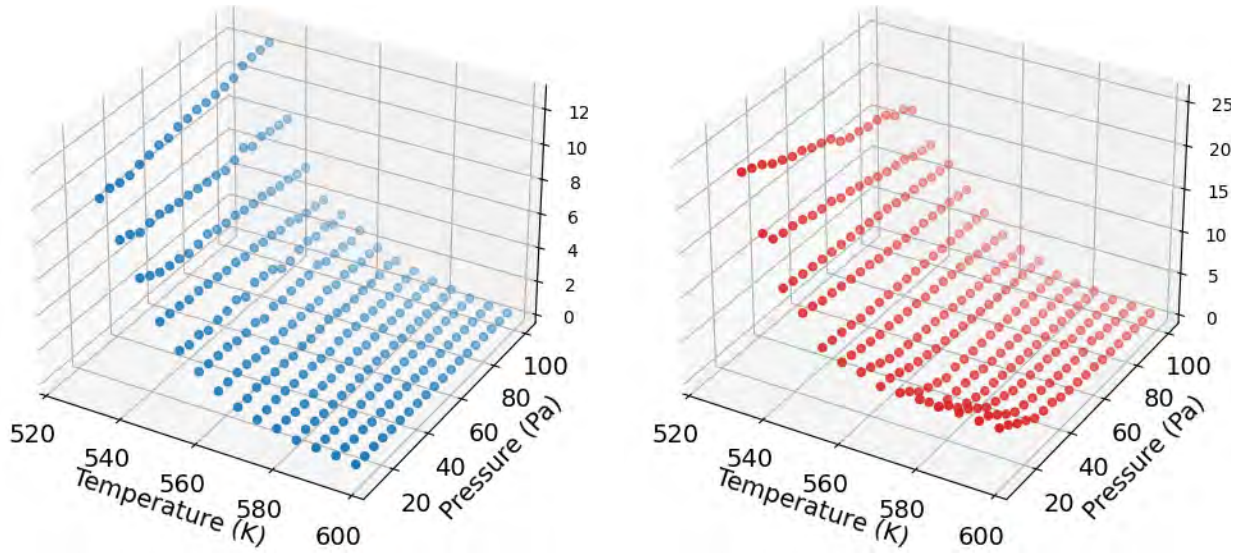
(a) Data set for Step A generated from the microscopic model.

(b) Data set for Step B generated from the microscopic model.

Figure 2.7: Data points collected for Step A and Step B at various operating conditions.

The results from the FNN model indicate that there is an agreement between the operating

conditions and the half-cycle time. The comparisons between the data from the kMC model and

the predicted data from the FNN model are represented in Figure 2.8. A mean squared error of

0.0856 % and 0.175 % for the sample test size data set (20 % of the data points from the kMC

data set) was calculated for Steps A and B, respectively, which indicates that the regression model

is characteristic of the data from the kMC model. The scatter plots in Figures 2.8a and 2.8c

demonstrate that the FNN data are representative of the kMC data. The accuracy of the predictions

is visualized in the line plots as shown in Figures 2.8b and 2.8d, which shows a majority of data

points in the predicted time region. Therefore, the results of the FNN model verifies that the trained

FNN model accurately provides the half-cycle time without demanding kMC computational work.

As shown in Figures 2.9a and 2.9c, the contour plots depict the half-cycle times at various temperature and pressure ranges. The objective of these graphs is to determine whether the given operating conditions are feasible or not in an industrial setting. The threshold half-cycle times for the feasible operating conditions are defined as $1.5\ s$ and $2.5\ s$ for Steps A and B, respectively, as aforementioned in Section refsec:val and visualized in Figures 2.9b and 2.9d. This is reasonable for the steady-state microscopic kMC modeling. If a gas transport time-scale domain of 2 to 3 s is considered, the overall process time including the half-cycle time in the microscopic domain and the gas transport time in the macroscopic domain can be less than the industrial half-cycle time of $5\ s$.

The contour plots for both steps show different patterns. Figure 2.9a reveals that Step A is independent on pressure as opposed to Figure 2.9c for Step B, which indicates that the process is pressure-dependent on the TMA reagent. In other words, the half-cycle time does not vary significantly in different operating pressures, which is unsurprising due to HF molecules being readily adsorbed to the surface from a kinetic perspective. However, Step B is pressure-dependent, which is caused by the kMC model equipped with the kinetic mechanisms of Step B. If a TMA molecule adsorbs and remains on the reaction site, the secondary TMA cannot react with the surface molecule due to the instability and the steric hindrance effect. In addition, no matter how low the operating pressure is, the first TMA adsorption occurs due to the so-called no-rejection rule of the kMC model. Thus, the secondary adsorption increases with increasing pressure.
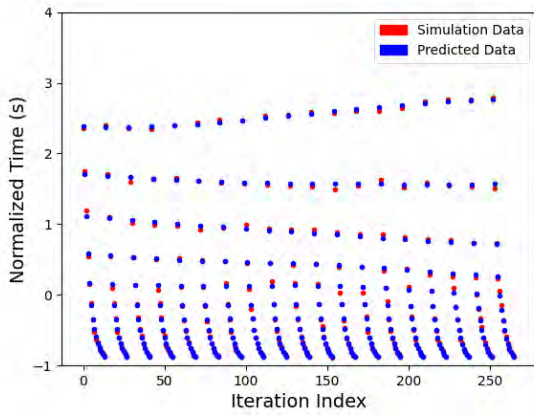
(a) Step A: Scatter Plot.
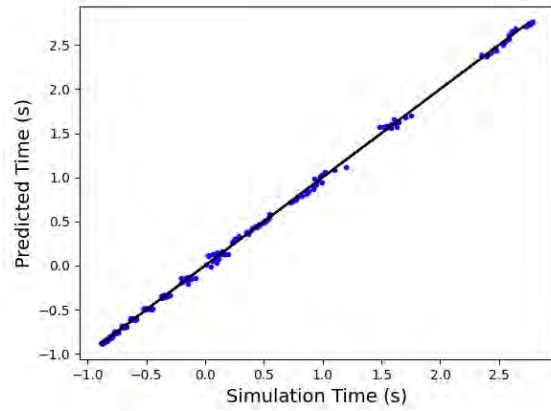
(b) Step A: Line Plot.
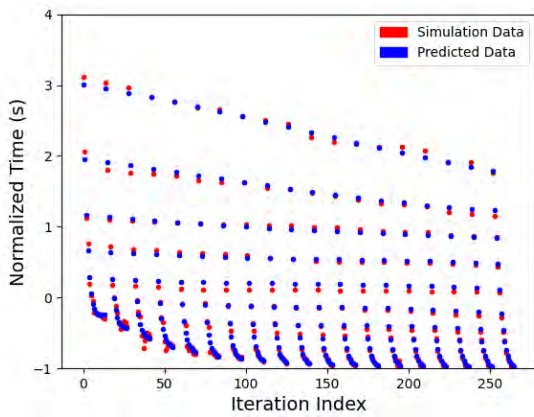
(c) Step B: Scatter Plot.
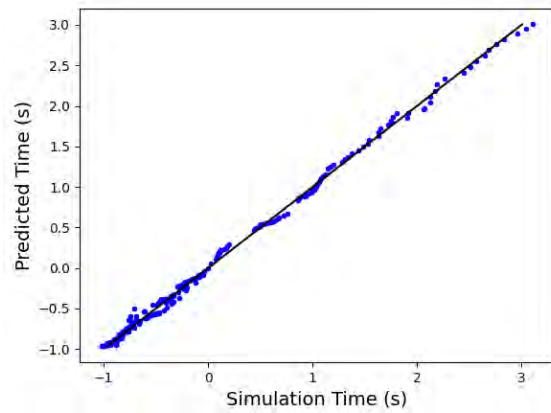
(d) Step B: Line Plot.

Figure 2.8: Comparison of the kMC simulation data and of the predicted data calculated by the FNN for Steps A and B. Scatter (a) and line (b) plots of Step A represent a mean squared error of 0.0856 %, which were calculated from 20 % of the data points of the full kMC data set for Step A. Scatter (c) and line (d) plots of Step B represent a mean squared error of 0.175 %, which were calculated from 20 % of the data points of the full kMC data set for Step B.
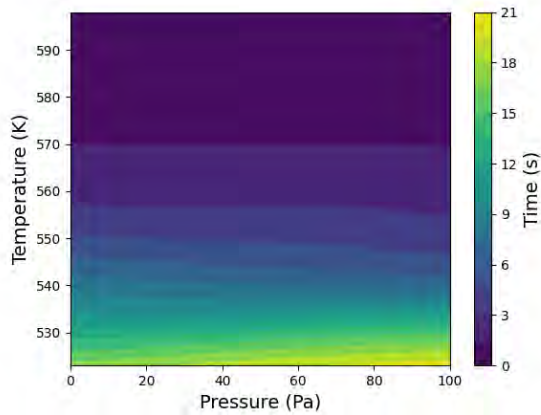
(a) Step A: Contour Plot.
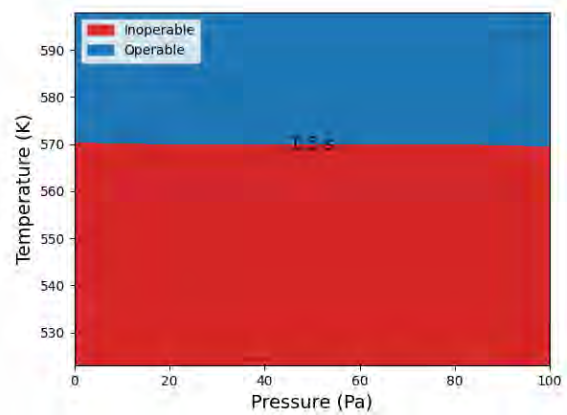
(b) Step A: Operational Feasibility Plot.
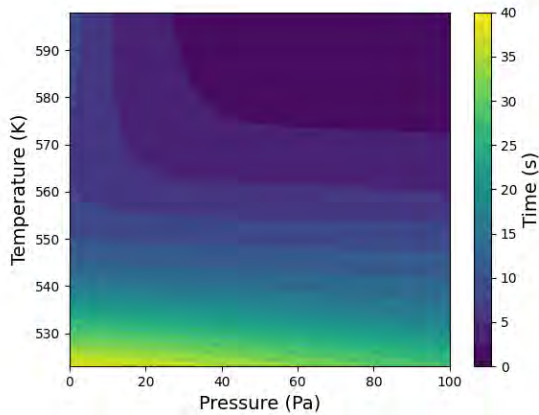
(c) Step B: Contour Plot.
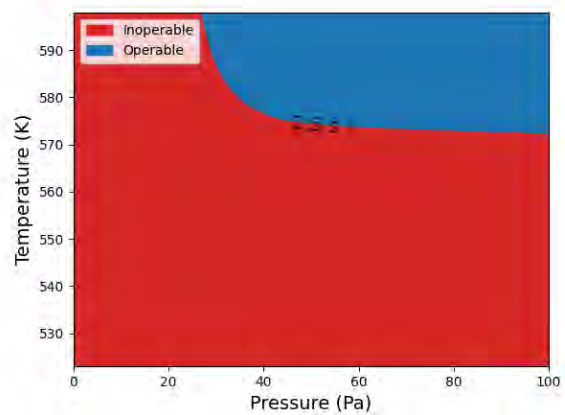
(d) Step B: Operational Feasibility Plot.

Figure 2.9: Comparison of the projected half-cycle time zones at differing operating conditions for temperature and pressure of the reagents. For Step A, a time contour plot (a) and an operational feasibility plot (b) with a feasible time range of 1.5 s are shown. For Step B, a time contour plot (c) and an operational feasibility plot (d) with a feasible time range of 2.5 s are shown.

2.6   Conclusion

In this work, microscopic and data-driven models were developed for the thermal atomic

layer etching (ALE) process of aluminum oxide. First, the kinetic mechanisms for HF and TMA

were proposed and their activation energies were calculated by DFT (Density Functional Theory)-

based electronic structure calculations, which were performed by Quantum ESPRESSO. The kinetic mechanisms and parameters were then used in a microscopic model using the kinetic Monte Carlo (kMC) algorithm. The kMC model was validated with available experimental results. Subsequently, the kMC model was used to generate data points for various operating conditions, which were used to train a computationally efficient feed-forward artificial neural network (FNN) model that can be used to determine optimal operating conditions in real-time. From the development of the reaction network in this work, the design of a reactor that can yield the desired etching results can be developed. Future work aims to amplify this research into macroscale modeling of a ALE reactor chamber by accounting for the effects of fluid dynamics, heat transfer, and mass transport phenomena.

Chapter 3

Multiscale Computational Fluid Dynamics Modeling of Thermal Atomic Layer Etching: Application to Chamber Configuration Design

3.1 Introduction

As a proposal to fabricate sub-5 nm nodes, a gate-all-around (GAA) approach has been pursued, which may one day become a potential successor to FinFETs, resulting in faster speed and greater power efficiency [62]. GAA transistors use vertically stacked nano-sheets or nano-wires instead of fins in FinFETs so nano-sheets are covered on all sides by the gate. The GAA technology has been predicted to reach an era of sub-5 nm thickness. Nevertheless, it has been difficult to commercialize GAA technology. In addition to ALD, atomic layer etching (ALE), as a counter

part of ALD, has emerged as an essential process for GAA-based nano-chip production. ALE is an etching process in which the substrate is exposed to sequential precursor pulses to remove a monolayer of the substrate in each etching cycle. ALE is a relatively new technique, and therefore, it has not been fully investigated in both empirical and computational ways. In order to completely understand the ALE process and make it possible to optimize the process configuration design, it is essential to fully develop a multiscale computational fluid dynamics (CFD) model for ALE processes.

A number of studies using a computational fluid dynamics (CFD) approach for atomic layer deposition processes have been carried out since 2010. [85] and [99] carried out CFD simulations for atomic layer deposition processes. An area-selective deposition process from a CFD point of view has been studied [22]. [20] and [122] performed multiscale CFD simulations for plasma enhanced chemical vapor deposition (PECVD) and plasma enhanced atomic layer deposition (PEALD), respectively. Their research, however, is limited to simulations for understanding atomic layer processes but not for reactor design and optimization. [121] recently proposed an optimized showerhead design for top injection reactors. Despite the progress made on the research for these cross-flow reactors, with their strengths and drawbacks being generally described by [41], there has not been any quantitative comparison of reactor design performance via multiscale CFD-based modeling. Thus, this work is aimed to evaluate different types of cross-flow reactors for thermal ALE and to characterize their features and performances using multiscale CFD modeling. Several factors including film uniformity and reduction in process etching time will be investigated to determine the optimal reactor design.

Specifically, in this work, a multiscale computational fluid dynamics (CFD) model for ther-

mal atomic layer etching of aluminum oxide (Al_2O_3) thin films is developed. Initially, a previously developed microscopic model based on the kinetic Monte Carlo (kMC) algorithm is adopted for the microscopic surface domain to describe the etching process [116] to capture the nature of the surface etching reactions at the atomic level. Next, a 3D CFD macroscopic model using ANSYS Fluent 2021R2, as a commercial CFD software, is established for the gas-phase domain in which mass, momentum, and energy transport are considered. Lastly, the microscopic and the macroscopic models are combined to fully characterize the thermal atomic layer etching of aluminum oxide thin films and used to evaluate four reactor chamber designs.

3.2 Multiscale CFD Modeling for thermal ALE

3.2.1 Background and Overall Modeling Framework

Experiments only permit data to be obtained from limited locations in the system that is equipped with sensors, and despite having these sensors, the amount of data collected experimentally may not gather the complete information of the system under various operating conditions. Meanwhile, 3D computational fluid dynamics (CFD) modeling based on the principles of fluid dynamics and transport phenomena allows one to obtain engineering data without any physical experiments for a considerably inexpensive cost. Moreover, the CFD simulation can be performed at various operating conditions, enabling one to ascertain an empirical model with a greater collection of data. It is, however, limited to provide the atomistic reaction information from a microscopic point of view even if macroscopic CFD modeling offers extensive data in terms of mass, momentum, and energy transport. To overcome this issue in this work, a 3D multiscale CFD model is built

by combining a macroscopic CFD model with a previously developed microscopic model [116] of the etching process based on a kinetic Monte Carlo (kMC) algorithm, thus resulting in providing a comprehensive understanding for the thermal ALE process of aluminum oxide thin films. [122] and [114] have important and timely articles on multiscale CFD modeling for plasma enhanced atomic layer deposition (PEALD). The authors provided valuable insight of the PEALD process of hafnium oxide thin films and used these models to study real-time control. Despite their efforts, their multiscale CFD modeling lacked a degree of accuracy since they did not consider the consumption of reactants and the production of products, which would affect the pressure distribution of the system. To address this issue, in this work, the heterogeneous surface reactions in the 3D CFD model are established in accordance with the reaction mechanisms of the microscopic model so that the etching of the surface species on the substrate can be simulated. Those reactions clearly have an impact on the momentum, energy, and mass transport in the gas-phase domain, in which the two precursors are consumed and the products of water and dimethylaluminum fluoride are yielded and transported from the substrate to the gas-phase domain. Pressure and temperature at different locations on the substrate are calculated at every time step and transferred to the microscopic model to calculate the etching progression in an atomistic level, of which detailed descriptions are provided in the following sections.

3.2.2 Microscopic Modeling

The thermal atomic layer etching (ALE) of aluminum oxide is driven by two reaction steps (Step A and Step B) using sequential and self-limiting thermal reactions that are each followed by purge steps. Hydrogen fluoride (HF) and trimethylaluminum [TMA, $Al(CH_3)_3$] are involved

to remove the Al_2O_3 surface layer. In Step A (Modification cycle), HF exposure fluorinates the Al_2O_3 surface and forms AlF_3 on the substrate. During Step B (Etching cycle), TMA exposure facilitates ligand-exchange reactions and modifies the AlF_3 surface into a volatile layer composed of dimethylaluminum fluoride [DMAF, $AlF(CH_3)_2$]. Following Step A and Step B, a purge gas, N_2 is used to remove any byproducts produced and remaining precursors during a purge time. The schematic of the thermal ALE of aluminum oxide is illustrated in Figure 3.1 and the overall reaction can be described by

$$Al_2O_3 \text{ (s)} + 6HF \text{ (g)} + 4Al(CH_3)_3 \text{ (g)} \rightarrow 6AlF(CH_3)_2 \text{ (g)} + 3H_2O \text{ (g)} \qquad (3.1)$$
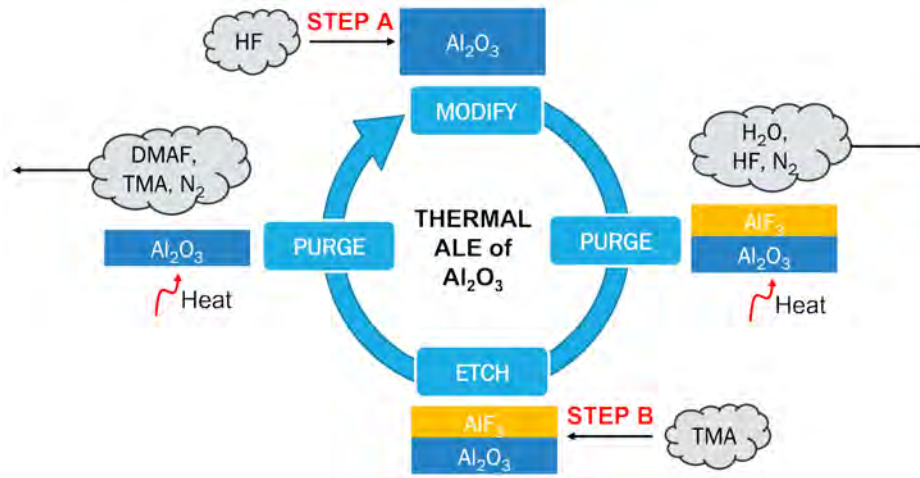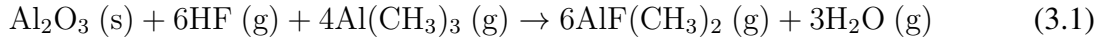
Figure 3.1: The thermal ALE cyclical process for Al_2O_3. The process begins with Step A, in which HF fluorinates the surface of the substrate and modifies the surface producing AlF_3. Following Step A is a purge step to remove H_2O vapor and residual HF. Next, Step B consists of the etching cycle to convert the modified AlF_3 layer into the volatile species DMAF using the reagent, TMA. The cycle concludes with another purging step to remove trace TMA and DMAF produced during the etching cycle. The addition of heat allows for complete vaporization of volatile species.

The microscopic model for the thermal ALE process of aluminum oxide thin films is formulated based on the variable step size method (VSSM) known as the kinetic Monte Carlo (kMC)

algorithm, of which the detailed description was given in [114]. In this previous work, θ-Al$_2$O$_3$ ($\bar{2}$ 0 1) for the aluminum oxide structure was employed and approximated to a 300×300 lattice model. After modeling the surface, DFT (Density Functional Theory) calculations were performed to investigate all critical reaction steps that have significant impacts on the overall surface reaction time and to estimate their kinetic parameters.

In this work, the wafer is divided into twelve regions to spatially simulate the microscopic model to obtain realistic and accurate etching data across the entire wafer surface, which is presented in Figure 3.2a. Pressure and temperature data at the twelve wafer regions are extracted from the CFD model at each time step and substituted into the kMC algorithm. All reaction rate constants are obtained from temperature and pressure by using Collision Theory and Transition-State Theory. Finally, the sum of the rate constants (k_{total}) is calculated by

$$k_{total} = \sum_{i=1}^{N} k_i \tag{3.2}$$

where k_i is the reaction rate constant of the reaction i, and N is the number of reaction pathways. For the reaction selection of a single reaction site on the wafer, a specific reaction can be randomly chosen as follows:

$$\sum_{i=1}^{j-1} k_i \leq \gamma_1 k_{total} \leq \sum_{i=1}^{j} k_i \tag{3.3}$$

where j represents the reaction j and $\gamma_1 \in (0, 1]$ is the first random number for the reaction selection. The reaction selection is implemented at every reaction site in which a random number is generated for each reaction site. If the value of $\gamma_1 k_{total}$ lies between $\sum_{i=1}^{j-1} k_i$ and $\sum_{i=1}^{j} k_i$, the reaction j is chosen for the reaction site. Otherwise, no reaction occurs at the reaction site. Once

the reaction selection task for every reaction site is completed, the system clock evolves with a time interval determined as follows:

$$\Delta t = \frac{-\ln \gamma_2}{k_{total}} \qquad (3.4)$$

where γ_2 is the second random number used for the time evolution ($\gamma_2 \in (0, 1]$).
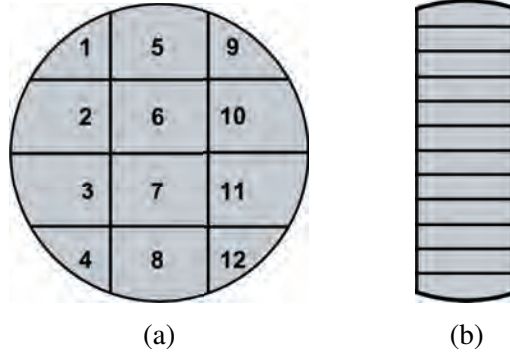
Figure 3.2: (a) Twelve substrate regions for microscopic simulations. (b) Twelve substrate positions for the investigation of the flow distribution.

3.2.3 Macroscopic Gas-Phase Modeling for Four Reactor Configurations

For macroscopic modeling, four different types of reactors are constructed and their performances are evaluated with respect to two metrics: film uniformity and etching speed (the specific designs considered are discussed in greater detail in the next subsection). The performance of a reactor is closely related to how fast a film on the wafer is deposited or etched. The magnitude of the deposition and etching rates is largely dependent on the type of fluid flow (laminar or turbulent), which is directly related to the precursor flow rate. For instance, [87] demonstrated that higher precursor flow rates resulted in higher deposition and etching rates, but the reduction in their process operating times was not significant. In addition to the deposition/etching rate, the film etching uniformity across the wafer is another key factor to evaluate reactor designs. Despite

43

the fact that self-limiting behaviors have been reported in atomic layer processes [67], the spatial film etching uniformity could be degraded due to the non-uniform distribution of precursors [27] in the gas-phase above the wafer, and this could compromise the integrity of the etched product. For instance, turbulent flow can disrupt the uniformity of the fluid flow, thus undermining film quality of the wafer. For this research, it is desirable to consider several reactor configurations that can introduce inherent reactor resistance to turbulent flow and maintain a laminar flow profile for operating flow rate regime to improve film etching quality. Therefore, these two aforementioned factors, etching rate and uniformity, are considered to compare the performances of the four reactor configurations considered in this work.

Reactor Chamber Designs

There are two general types of reactors for single-wafer systems: the top injection reactor and the cross-flow reactor [41]. The top injection reactor with distributors enables precursors to be uniformly injected above the wafer leading to highly uniform etching of the thin films. On the other hand, the cross-flow reactor has a smaller height of a few mm so that the gas displacement time is minimized. The cross-flow model also maximizes the lateral convective flow across the wafer. In this work, the cross-flow reactor is adopted to reduce the process and purge time since HF has a long residence time, which may remove self-limiting behavior resulting in spontaneous chemical vapor etching [60]. These cross-flow reactors are constructed with a feed source from one end of the reactor and the output source on the opposing end of the reactor to induce mass transport from one end of the wafer to the opposing end of the wafer. However, for conformal thin film etching, it is essential to obtain uniform flow profiles across the surface of the wafer [27]. Therefore, different

distributors are employed on the cross-flow reactors to optimize the flow profiles of the precursor. Then, the modified reactors are compared to the simplest reactor geometry that has no distributor to determine if the distributor is effective in improving the performance with respect to a reactor design without distributors.

Specifically, in this work, four types of reactor chambers are created and their performances are evaluated by multiscale CFD simulations. First of all, the typical geometry (G0) is developed, which is a cylindrical-shaped chamber with a 500 mm outer diameter and 10 mm height as shown in Figure 3.3a. A wafer of 300 mm diameter is placed at the center of the bottom face of the chamber where an inlet of 20 mm diameter and an outlet of 40 mm diameter are located on the bottom face. Based on G0, a multi-inlet geometry (G1), a showerhead geometry (G2), and an inclined plate geometry (G3) are proposed. The multi-inlet geometry (G1) is constructed with three inlets in place of one inlet, in which each inlet has the same diameter as that of G0, but the total feed flow rate is divided evenly for all three inlets, with the total flow rate summing to the same inlet flow rate as that of G0. G1 is visualized in Figure 3.3b. As can be seen in Figure 3.3c, the showerhead geometry (G2) is constructed similarly to G0 but includes a showerhead divider of 2-mm thickness that distributes the inlet flow to achieve a uniform flow. The size of the pores is 4-mm in diameter for a total of 63 pores that are distributed into two rows. Lastly, the inclined plate geometry (G3) is developed with an arch-shaped inclined plate with 2-mm thickness between the inlet and the wafer surface with five degrees of deviation from the horizontal as shown in Figure 3.3d. The inclined plate is used to guide the center flow to regions near the edges in order to reach a uniform flow pattern for the film uniformity. There are several assumptions that are applied to the operation of the reactor:

1. There are carrier gas manifolds in the upstream facility so that the precursors and N_2 are well-mixed when introduced into the reactor.

2. Other geometric objects such as sensors and mechanical structures are ignored.

3. The temperature of the substrate is maintained through a PID (proportional-integral-derivative) controller at a desired set-point.

4. The operating pressure is controlled and maintained via the vacuum pump in the downstream facility.

5. The flow through the reactor is characterized as laminar flow.

The above assumptions are implemented into the boundary conditions that would generate the mesh for the multiscale computational fluid dynamics simulation.

(a) G0

(b) G1

(c) G2

(d) G3
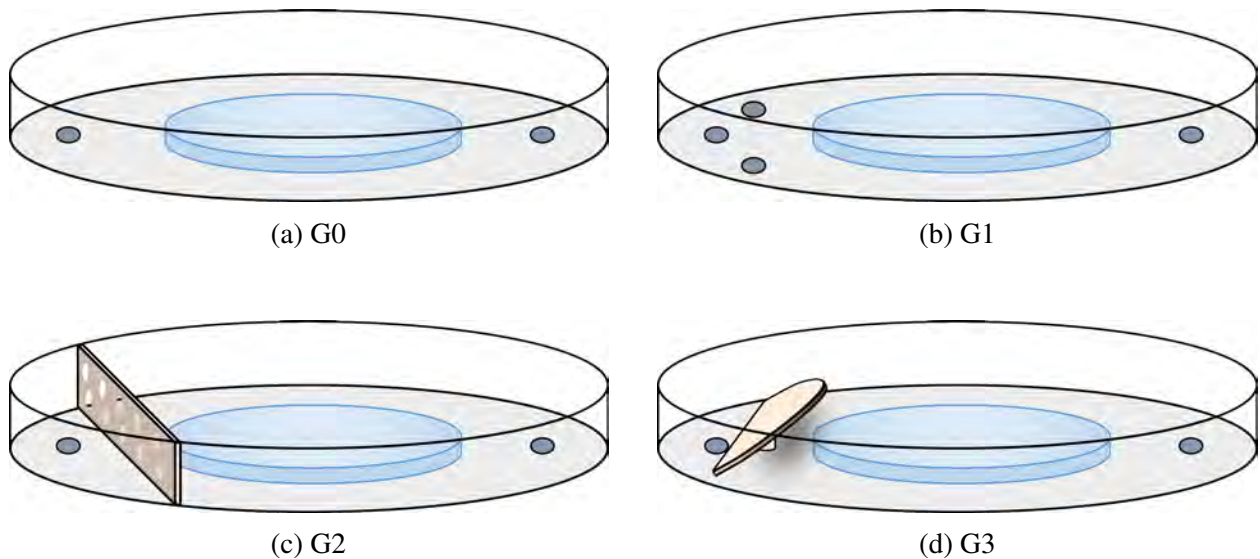
Figure 3.3: Schematic diagrams of the ALE reactors: (a) typical, G0; (b) multi-inlet, G1, (c) showerhead, G2; and (d) inclined plate, G3. The input(s) (dark gray) are located on the left-hand side of the reactor and the output (dark gray) is located on the right-hand side of the reactor. The wafer is presented in blue.

Meshing

The characteristics of the mesh for each reactor geometry will play a substantial role in the convergence, accuracy, and stability of the numerical solutions that will be calculated. Meshing Mode, an application of ANSYS Fluent, is used to construct the mesh for the reactors described in Section 3.2.3. An acceptable mesh can be determined by mesh quality criteria in accordance with the standards outlined by [4] as shown in Table 3.1.

Table 3.1: The mesh quality acceptability criteria range and mesh parameters calculated from ANSYS Fluent for various reactor geometries. For orthogonality, the minimum value is presented on the left and the average value is presented on the right.

Quality Indicator	Orthogonality	Skewness	Aspect Ratio	Number of Cells
Criteria	$0.001 \sim 1^*$	$0^* \sim 0.95$	$1^* \sim 8$	N/A
G0	0.130/0.727	0.271	2.060	266,291
G1	0.152/0.727	0.272	2.036	273,210
G2	0.002/0.741	0.252	1.525	1,550,322
G3	0.207/0.738	0.261	2.977	574,414

*Desired value for ideal mesh quality.

Specifically, Table 3.1 shows the key indicators used for analyzing the mesh quality. The quality of the mesh depends on the geometry of the cells and the boundary conditions used to define the overall geometry of the mesh. In this work, hybrid meshes, consisting of mixed element types, are generated to substantially reduce the computation time but still maintain acceptable mesh quality. Prism layers are utilized to resolve the boundary regions and tetrahedral cells are employed as a rudimentary element in the reactor chambers.

Among the factors that affect the mesh quality, skewness; orthogonality; aspect ratio; and res-

olution; are considered for the evaluation of the developed mesh structures for the various chamber geometries. The skewness of a cell is defined as the measure of the difference between a cell's geometry with that of an equivalent equilateral geometry of the same volume of the actual cell. The equilateral skewness for the tetrahedral mesh is calculated from the following equation:

$$Skewness = \frac{optimal\ cell\ size - cell\ size}{optimal\ cell\ size} \quad (3.5)$$

The optimal cell size is defined as the size of an equilateral cell with the same circumradius. Thus, if the cell size is approximately equal to the optimal cell size, an ideal skewness of 0 is obtained. A low skewness is desirable to obtain an accurate and stable solution. The orthogonality is defined as the minimum value of all of the cells of the mesh from the following equation:

$$\frac{\vec{A}_i \cdot \vec{c}_i}{|\vec{A}_i||\vec{c}_i|} \quad (3.6)$$

where $\vec{A}_i$ is the area vector of a face and $\vec{c}_i$ is the vector from the centroid of the cell to the centroid of the adjacent cell. An ideal mesh has an orthogonality that is close to unity. For tetrahedral cells, the orthogonal quality is the minimum of the orthogonality of all cells in the mesh. Due to the variance in the orthogonality for each cell, the minimum orthogonality of all the cells should be greater than 0.001. The aspect ratio is another important indicator, which is a measure of the stretching of a cell. The aspect ratio is calculated as the ratio of the maximum to the minimum of the normal distance between the face centroids and the cell centroids, and the distance between the nodes and the centroid. A uniform aspect ratio is desirable for regions where the flow field varies greatly and an ideal aspect ratio is equivalent to unity for equilateral cells. Lastly, the resolution has

a significant contribution to how critical regions are calculated and directly affects the total number of cells used to describe the mesh. The resolution of the mesh is a measure of the distribution of cells in particular regions of the mesh geometry and is measured in terms of coarseness or fineness. Meshing mode contains a tool that produces adaptive sizing that automatically generates regions where the mesh is finer at the boundary regions and coarser in regions away from the boundary. It is important to obtain high resolution, especially in critical regions where boundary layers change dramatically in their behavior, specifically, wall-fluid boundaries, which will affect the accuracy of the computed numerical solution. Thus, the following equation, which is derived from the Blasius approximate solution for laminar flow over a flat plate, can be employed for determining the meshing in the flow domain near the walls:

$$y_p \sqrt{\frac{u_\infty}{\nu x}} \leq 1 \qquad (3.7)$$

where y_p is the distance to the wall from the adjacent cell centroid, u_∞ is the free stream velocity, ν is the kinematic viscosity of the fluid, and x is the distance along the wall from the starting point of the boundary layer.

The consideration of the aforementioned factors affecting mesh quality leads to the development of the meshes for each reactor configuration, which are visualized in Figure 3.4. The results from the meshing process for all reactor configurations, which are listed in Table 3.1, indicate that all reactor geometries are within the acceptability criteria for the average values. This also implies that the meshes built via Fluent's Meshing Mode would have reliable computed results. Mesh independence studies were also carried out for all reactor designs to ensure that the simulation results

are independent of mesh structure.

(a) G0

(b) G1

(c) G2

(d) G3
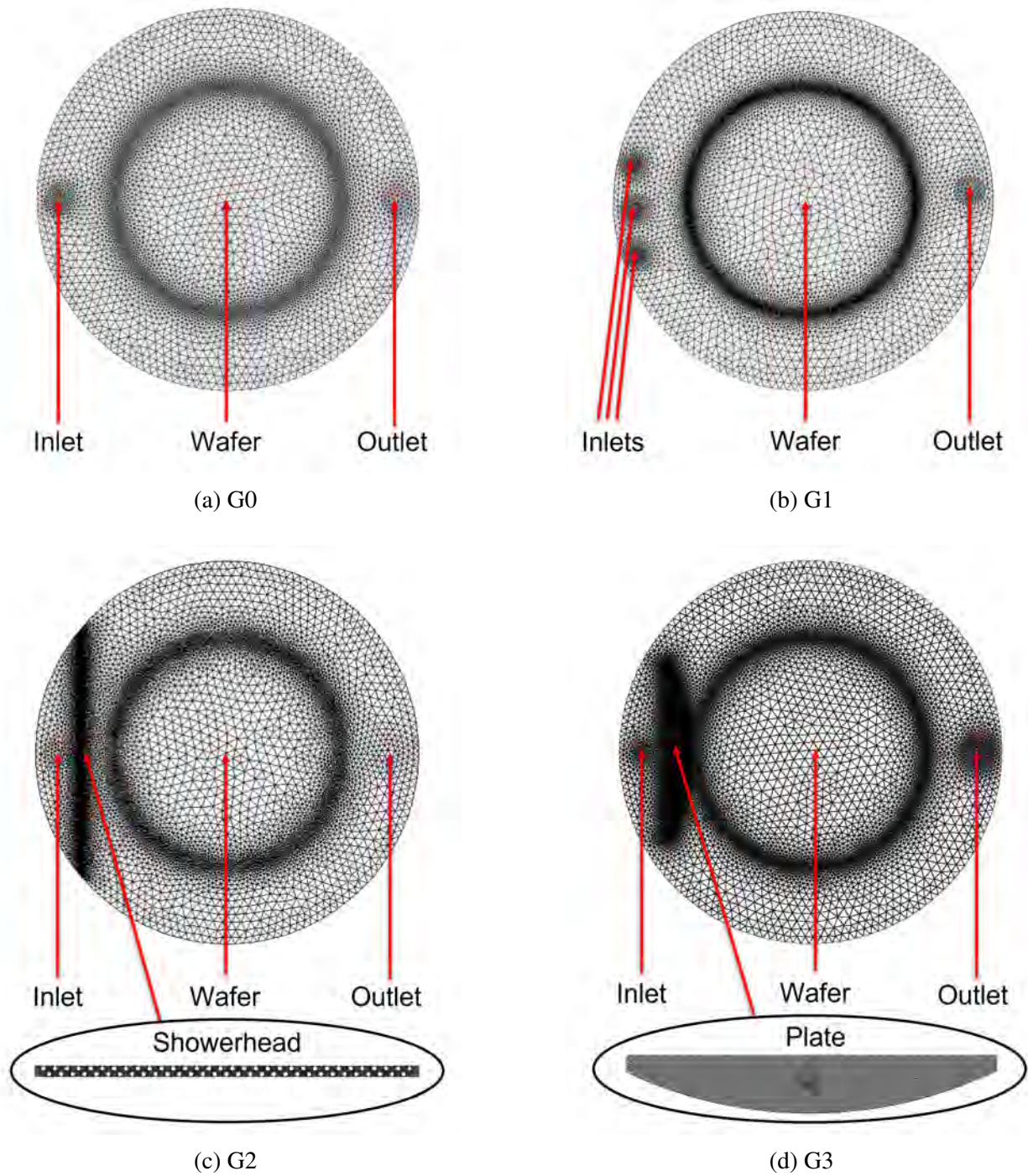
Figure 3.4: Meshes for each of the reactors produced from ANSYS Fluent's Meshing Mode: (a) the typical reactor, (b) the multi-inlet reactor with three inlets, (c) the showerhead reactor, and (d) the inclined plate reactor.

Thermophysical Property Calculation

Thermophysical data are required for the materials used in the etching process and are employed in the computational fluid dynamics (CFD) simulations. However, some species produced during the etching process have limited thermophysical data in the open literature. One of these species, dimethylaluminum fluoride (DMAF), has little to no available experimental data, thus computational chemistry calculations via the open-source thermochemistry simulation software, Quantum ESPRESSO (QE), are utilized to calculate thermophysical parameters including the standard enthalpy, the standard entropy, and the specific heat [39, 7] for DMAF. Despite the accessibility of QE, there are limitations for calculating other thermophysical parameters including thermal conductivity and viscosity. For this work, these parameters are determined by adopting the parameters from a chemically similar molecule, dimethylaluminum chloride (DMACl), to DMAF [26], because chlorine and fluorine exhibit similar chemical behaviors as halogens. Lastly, the density of DMAF is calculated by assuming that the species behaves as an ideal gas due to the ambient environment of the reactor having low pressure and high temperature operating conditions, which are summarized in Table 3.3.

Quantum ESPRESSO contains several packages that are required for calculating thermophysical data and these programs must be run sequentially. First, the PWscf (Plane-Wave self-consistent field) program is used to calculate the electronic properties and optimize the atomic positions of the molecule, which are modeled using Density Functional Theory (DFT) and PAW (Projector Augmented Wave) pseudopotential data. Next, the PHonon package is used to calculate the dynamical matrix of the phonons. Various programs are dedicated to the building of the dynamical matrix (PH

program), to the solving of the interatomic force constants (IFC) from the dynamical matrix (Q2R program), and to calculating the eigenvalues of the dynamical matrix, which are the vibrational frequencies of the molecule (MATDYN program), by employing the finite displacement method and the density functional perturbation theory.

Lastly, the QHA (Quasi-Harmonic Approximation) package including the Partial Phonon DOS (Density of States) program is used to calculate atom projected density of states, the Mean Square Displacement program in the QHA package is utilized to calculate the deviation of the atom with respect to a reference position caused by the vibration of the atoms, and lastly, the FQHA (Fractional Quasi-Harmonic Approximation) program combines the results produced from the latter-mentioned programs to calculate the thermophysical properties including entropy, enthalpy, Helmholtz free energy, and specific heat at constant volume as functions of temperature. The results from the phonon calculation are displayed in Table 3.2. The formulation and derivation of the equations to solve the vibrational frequencies of the dynamical matrix and to calculate the thermophysical properties using statistical thermodynamics are discussed in greater detail by [6] and [107].

Three-Dimensional Computational Fluid Dynamics Simulation

Facilitated by the Hoffman2 Cluster at UCLA, the Fluent computations are implemented with 24 parallel central processing units (CPU) with 16 GB memory for each core processor so that the parallel processing splits the gas-phase domain into multiple partitions to improve the computation efficiency. There are two solver technologies available in Fluent: pressure-based and density-based. The pressure-based solver has been traditionally used for incompressible and mildly com-

Table 3.2: Thermophysical material properties of DMAF specified in ANSYS Fluent.

Thermophysical Parameter	Value	Units
Standard Enthalpy of Formation*	-499.290	kJ/mol
Standard Entropy of Formation*	196.421	J/(mol·K)
Specific Heat at Constant Pressure*	123.633	J/(mol·K)
Thermal Conductivity†	0.07268	W/(m·K)
Viscosity†	0.01100	kg/(m·s)

*Parameters calculated from Quantum ESPRESSO.
†Property data of DMACl [26].

pressible flow. The thermal ALE is operated at an extremely low pressure in a single-phase flow, and thus, the pressure change of the mixture is negligible. In addition, the feed composition of the precursor is low, therefore the density and pressure change of the species are negligible for this simulation. Therefore, the pressure-based solver is applicable for this work. The operating conditions of the reactor are listed in Table 3.3. In addition, under the pressure-based solver, the coupled algorithm is used to significantly decrease the convergence time, in which the momentum equation and the pressure-based continuity equation are solved simultaneously. Transient analysis for comprehensive computational fluid dynamics (CFD) modeling of thermal ALE is performed with a time step of 0.025 s with 200 iterations under transport phenomena. The conservation equations for mass and momentum are written as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m \tag{3.8}$$

$$\frac{\partial(\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot (\overline{\overline{\tau}}) + \rho \vec{g} + \vec{F} \tag{3.9}$$

where ρ is the density of the mixture, $\vec{v}$ is the velocity of the mixture, S_m is the mass transfer source term, p is the static pressure, $\bar{\bar{\tau}}$ is a rank two stress tensor that is symmetric, $\rho\vec{g}$ is the gravitational body force, and $\vec{F}$ is the external body force. In addition, the conservation of energy is described by

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = -\nabla(\Sigma h_j J_j) + S_h \qquad (3.10)$$

where E is the internal energy, h_j is the sensible enthalpy of species j, J_j is the diffusion flux of species j, and S_h is the heat transfer source term.

The inclusion of precursor consumption from Steps A and B to the multiscale 3D CFD simulation is employed to generate a more realistic flow profile. The consumption of the precursors and generation of products are calculated from the reaction rate constants determined by the modified Arrhenius equation that is defined as follows:

$$k_j = A_j T^{\beta_j} e^{-E_{A,j}/RT} \qquad (3.11)$$

In the above equation, k_j is the reaction rate constant for reaction j, T is the ambient temperature of the reactor, β_j is the temperature exponent for reaction j, $E_{A,j}$ is the activation energy for reaction j, and R is the ideal gas constant. For simplicity, the temperature exponent, β_j, would be declared 0 for the simulation. Due to the large number of reactions and the lack of thermophysical data for numerous species obtained from prior microscopic research from [116], Eq. (3.1) will be simplified into two surface reaction steps that are defined below:

$$\text{Al}_2\text{O}_3 \text{ (s)} + 6\text{HF (g)} \rightarrow 2\text{AlF}_3 \text{ (s)} + 3\text{H}_2\text{O (g)} \qquad (3.12)$$

$$2\text{AlF}_3 \text{ (s)} + 4\text{Al}(\text{CH}_3)_3 \text{ (g)} \rightarrow 6\text{AlF}(\text{CH}_3)_2 \text{ (g)} \tag{3.13}$$
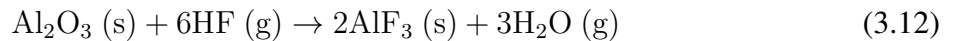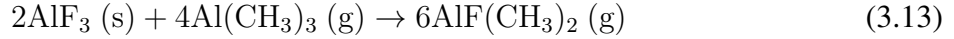
First, the gaseous precursor, HF, physisorbs onto the surface of Al_2O_3 to produce AlF_3 and water vapor when reacting under high temperatures. Subsequently, the gaseous species, TMA, chemisorbs onto the AlF_3 surface forming the volatile species DMAF. After defining these reactions, operating conditions are needed to fully define the system.

Table 3.3 shows the operating conditions of the multiscale 3D CFD simulations presented below. The operating pressure is set to be 133 Pa and the temperature is maintained at 573 K; thus, the material thermophysical properties are expected to be constant, hence, the data calculated in Table 3.2 is not required to account for the temperature dependence. As previously mentioned in Section 3.2.3, the temperature of the surface can be maintained through a control system that measures the temperature in real time while the operating pressure is controlled by discharging effluent through a vacuum pump. A constant flow of 150 sccm of N_2 gas is used to carry hydrogen fluoride (HF) and trimethylaluminum (TMA) into the reactor. The operating conditions are defined by a user-defined function (UDF) implemented in ANSYS Fluent, in which the operating conditions are automatically adjusted according to the cyclical operation.

3.3 Simulation Results and Reactor Design Evaluation

The multiscale computational fluid dynamics (CFD) simulations are first performed for each reactor model with an HF flow rate of 150 sccm and a TMA flow rate of 70 sccm to validate the 3D multiscale CFD model. Next, the results from the multiscale CFD modeling of the four reactors, G0 through G3, are discussed to observe which reactor design achieves a better distribution of

Table 3.3: Operating conditions for the Thermal ALE process.

Parameter	Value	Units
Operating Pressure	133	Pa
Operating Temperature	573	K
N_2 flow rate	150	sccm
HF half-cycle	2.0	s
1st Purge	5.0	s
TMA half-cycle	3.0	s
2nd Purge	5.0	s

precursor flow for the film uniformity and faster half-cycle times for Steps A and B. Finally, the reactor that produces the best performance is selected and simulated at different precursor flow rates to be compared with the typical type reactor (G0) in terms of efficiency and effectiveness.

3.3.1 Simulation Results of Multiscale CFD Modeling and Validation

In the multiscale CFD simulations, the flow is assumed to be laminar in the fluid dynamics point of view as discussed in Section 3.2.3. The assumption is validated by the contours of the Reynolds number at 0.025 s of half-cycle time elapsed at standard reactor operating conditons, which are illustrated in Figures 3.5 and 3.6. It is observed that the highest Reynolds number is localized at the precursor injection region for Steps A and B, with the largest Reynolds number among all reactor configurations being 1.30 for Step A and 2.59 for Step B with both values obtained from the showerhead reactor. Consequently, no turbulent regime is observed throughout all of the reactors due to the atmospheric operating condition which provides more control over the flow pattern. As a result, the flow through the reactor is characterized by laminar behavior and

supports the assumption made in Section 3.2.3.

As shown in Figure 3.2a, the wafer is divided into twelve parts to calculate the etching progression for the microscopic simulations since the process data varies with location on the surface. Dividing the wafer into twelve sections enables one to collect more accurate and plausible numerical solutions than simulations with averaged pressure and temperature for the whole wafer. The results of multiscale computational fluid dynamics (CFD) simulations for the reactors are provided in Table 3.4. For all four reactors, the half-cycle times for Step A are calculated as 1.414 s through 1.446 s and the half-cycle times for Step B are also calculated as 2.498 s through 2.542 s using the kinetic Monte Carlo (kMC) method. The half-cycle times for both steps were computationally determined to be 1.38 s and 2.38 s, respectively, in the previously developed microscopic model [116] where the simulations were performed under ideal conditions without the influence of transport phenomena effects in the gas phase, which was supported by the experimental data from [61]. It is obvious that the half-cycle times from the multiscale CFD modeling are delayed due to the inclusion of mass transport as it takes some time for the wafer to be saturated by the precursors unlike the microscopic model at the steady-state. Furthermore, the consideration of the consumption of precursor species contributes to the slower process time. Hence, it is demonstrated that the overall multiscale CFD modeling including the thermophysical data is successfully developed.

3.3.2   Comparison of Reactor Designs

The central region (sections 5 through 8 in Figure 3.2a) of the wafer, as shown in Figure 3.2b, is divided into 12 substrate positions (labeled in increasing order from the top to the bottom of the
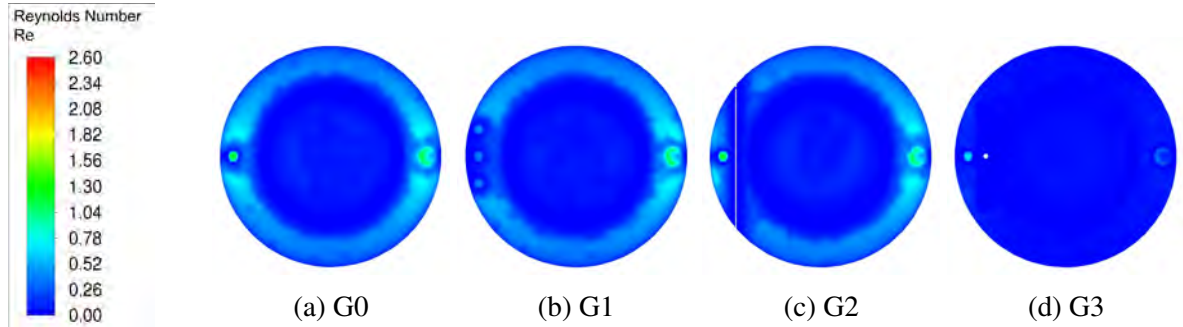
Figure 3.5: Contours of Reynolds Number for Step A of various reactor configurations at 0.025 s in the standard condition in Table 3.3
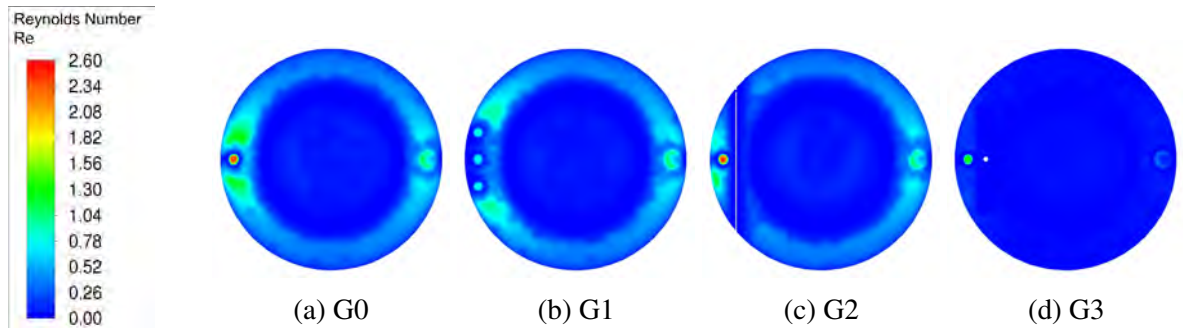
Figure 3.6: Contours of Reynolds Number for Step B of various reactor configurations at 0.025 s in the standard condition in Table 3.3

Table 3.4: Half-cycle times determined by the kMC simulation of the multiscale CFD model.

Reactor	Step A	Step B
G0	1.437 s	2.514 s
G1	1.446 s	2.542 s
G2	1.436 s	2.528 s
G3	1.414 s	2.498 s

figure) to calculate the pressure of the precursors at each part so that the analysis of film uniformity is carried out. Figure 3.7 shows the flow patterns of the four reactors over substrate position, which exactly agrees with the pressure contours of the precursor in Figure 3.8. The pressure contours of HF precursor are also consistent with the pressure contours of TMA in Figure 3.9, thus the type

58

of species in the thermal ALE cycle of Al_2O_3 plays a limited role in affecting the flow profile. As shown in Figure 3.7a, the flow of the typical reactor (G0) is formed in a circular shape due to the isotropic flow. The largest pressure deviation (i.e., the difference between the maximum and minimum pressures) is given as 15.2 Pa at 0.2 s and the half-cycle times are 1.437 s for Step A and 2.514 s for Step B.

Figure 3.7b indicates that the flow of the reactor with three inlets (G1) is more evenly distributed than that of G0, and consequently, G1 has less pressure deviation than G0, of which the largest value is 11.4 Pa at 0.2 s. This is also shown in Figure 3.8b. However, the half-cycle times (1.446 s for Step A and 2.54 s for Step B) are greater than those of G0 due to the lower precursor velocity as can be seen in Table 3.4. Despite the input being distributed through three inlets, the flow appears to migrate more towards the outlet of the wafer, leading to an uneven flow profile.

Figure 3.7c reveals that the reactor with the showerhead (G2) improves the flow pattern when compared to that of G0 and G1, and the largest pressure deviation is calculated to be 8.5 Pa at 0.2 s. Also, with time progression, the uniformity of the flow improves faster, hence, the pressure deviation decreases, compared to G0 and G1. In the initial stages of flow development, the parabolic pressure profile mentioned in the discussion of G0 is also displayed with that of G2 in Figure 3.8c and Figure 3.9c. It is observed that the showerhead serves to decrease the amount of precursor in central regions of the wafer, and therefore G2 has better uniformity compared to G0 and G1. The half-cycle times (1.436 s for Step A and 2.528 s for Step B) are similar to the half-cycle times of G0 due to the flow resistance of the showerhead divider. Thus, the addition of the showerhead divider improves the uniformity, but marginally improves the half-cycle time for Step A and slightly worsens the half-cycle time for Step B.

59

As shown in Figure 3.7d, the reactor with the inclined plate (G3) has the most uniform flow pattern at every time step, in which the largest pressure deviation is 2.9 Pa at 0.2 s. Table 3.4 shows that G3 has the least half-cycle times (1.414 s for Step A and 2.528 s for Step B) among the reactors despite the inclined plate acting as a flow resistance. Nevertheless, the results indicate that the effect of the uniformity outweighs the flow resistance, thus leading to the expediting of the etching process. Therefore, it is concluded that G3 may be able to give better film quality and thickness control than the other types of reactors. The complete etching cycle of Al_2O_3 for G3 is displayed in Figure 3.10. The inclined plate reactor (G3) shows the best performance in terms of the film uniformity and etching speed. Figure 3.10 shows the pressure of the two precursors and the etching progression over time, which is provided from the multiscale CFD simulation. The cycle consists of an HF dose of 2 s, an N_2 purge of 5 s, a TMA dose of 3 s, and an N_2 purge of 5 s.

3.3.3   Efficiency of the Inclined Plate Reactor

Multiscale CFD simulations are performed previously for different feed flow rates for the typical reactor (G0) and the reactor with the inclined plate (G3). The comparison of the half-cycle times of Steps A and B and the annual feed consumption of the precursor species, HF and TMA, are displayed in Figure 3.11. The annual feed consumption is calculated by assuming that 96 cycles of etching are conducted daily and that half-cycle times for each feed flow rate remain constant with each cycle in a single wafer system. The estimates for the precursor consumption are calculated for a single wafer. The results for Step A in Figure 3.11a indicate that the half-cycle time for Step A for G3 is consistently faster compared to that of G0, thus, the amount of precursor needed to ensure complete coverage is less than that of G0. By utilizing G3, at least 1.3×10^3

60
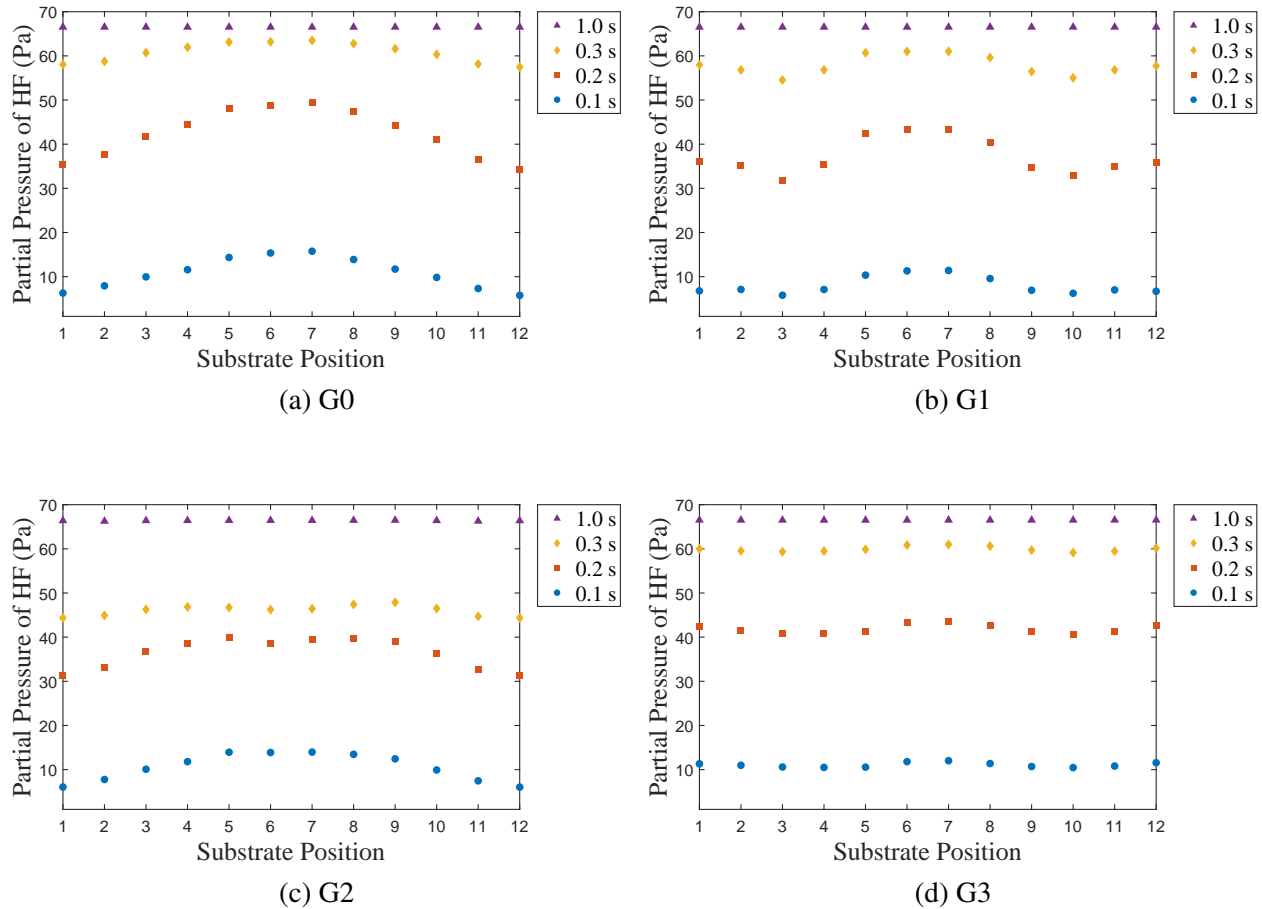
Figure 3.7: Centerline HF pressure data for each reactor at various times for an HF feed flow rate of 150 sccm. The substrate position is numbered starting from the top of the divided wafer in Figure 3.2b to the bottom.
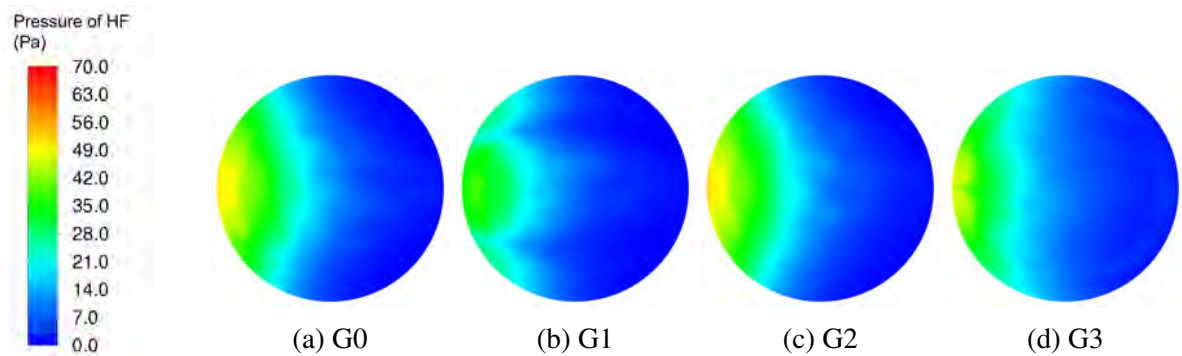
Figure 3.8: Contours of pressure of HF on the surface of the wafer for a Step A process time of 0.1 s and for an HF feed flow rate of 150 sccm.
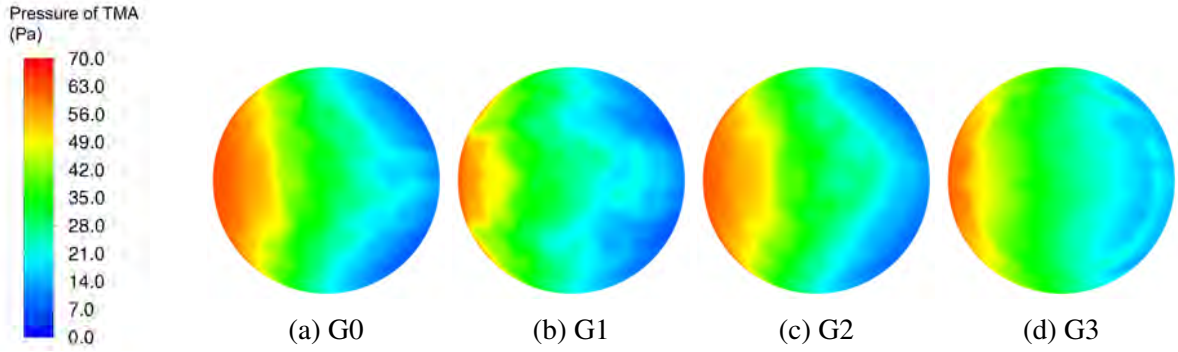
Figure 3.9: Contours of pressure of TMA on the surface of the wafer for a Step B process time of 0.2 s and for a TMA feed flow rate of 70 sccm.
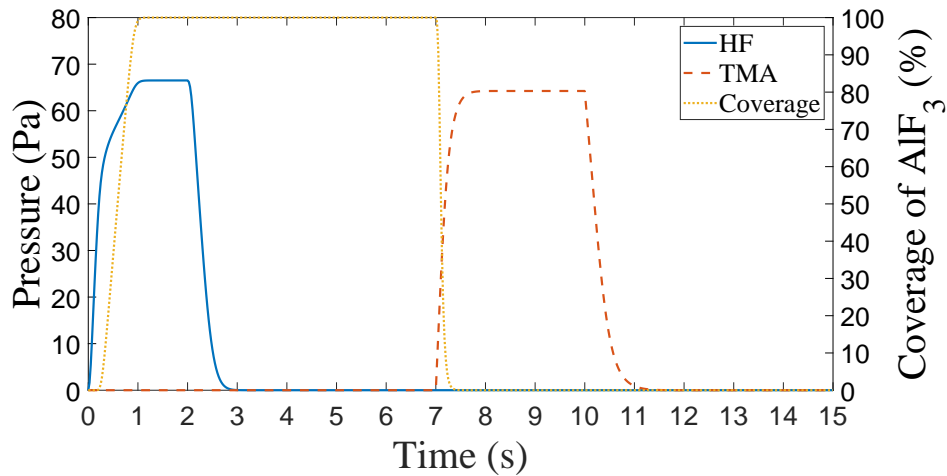
Figure 3.10: Complete cycle of G3 displaying the pressure of HF and TMA and coverage of AlF_3 for an HF and TMA feed flow rate of 150 sccm and 70 sccm, respectively. The blue solid line and the orange dashed line indicate the pressure of HF and TMA over time, respectively. The yellow solid line shows the coverage of AlF_3. The AlF_3 is formed in Step A and etched in Step B.

$std\ cm^3$ and at most $5.5 \times 10^4\ std\ cm^3$ of HF can be saved for the range of the simulated flow rates with 600 sccm flow rate achieving the greatest amount of precursor that could be saved. The quantitative results from Step B are displayed in Figure 3.11b where faster etching results for Step B are observed for G3, thus, lesser TMA is needed to achieve complete etching of a mono-layer of surface substrate. Adopting the G3 model could save at least $6.5 \times 10^2\ std\ cm^3$ and at most $6.7 \times 10^4\ std\ cm^3$ of TMA for the simulated range of feed flow rates. The greatest amount of

TMA saved for G3 in comparison to G0 occurs with a flow rate of 600 sccm. Consequently, the benefits of utilizing the inclined plate reactor not only include reduced Modification (Fluorination) and etching times but also a lesser amount of precursors is needed when compared to the typical reactor. Despite being able to maximize the amount of precursor saved for the 600 sccm flow rate, the reduction in half-cycle times for Steps A and B suggest that increasing the precursor flow rate does not reduce the half-cycle times significantly, and thus, it may be preferable to operate under laminar-like conditions with a lower magnitude of flow rate as discussed by [87].

(a) Step A	(b) Step B
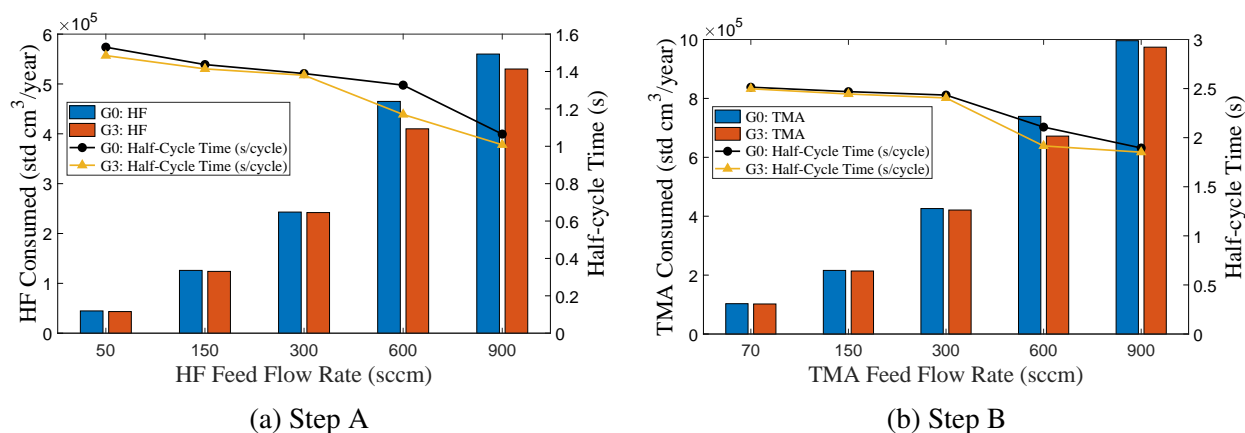
Figure 3.11: Process time and consumption of HF and TMA per year comparison between the CFD simulations of the typical reactor (G0) and of the inclined plate reactor (G3). The blue and orange bar indicate the consumption of the precursors for G0 and G3, respectively, for a single reactor. The black and yellow solid lines indicate the half-cycle time of G0 and G3 for both steps, respectively.

3.4   Conclusion

The thermal atomic layer etching (ALE) process of Al_2O_3 was simulated using a multiscale 3D computational fluid dynamics (CFD) model to investigate the impact of gas-phase transport phenomena on the etching process across a wafer. The CFD simulation was performed first by

constructing various reactor geometries (typical, multi-inlet, showerhead, and inclined plate) and meshing these geometries was completed until the quality criteria were met using ANSYS software. Next, ANSYS Fluent was used to perform the CFD simulation with the inclusion of precursor consumption via reactions corresponding to the microscopic model of the etching process. The phonon calculations to obtain thermophysical data were carried out prior to CFD calculations with some materials requiring the use of computational chemistry software, Quantum ESPRESSO (QE), to calculate the thermophysical data through electronic calculations. Lastly, the process data exported at every time step from Fluent were used in the kinetic Monte Carlo (kMC) microscopic model to determine the half-cycle times for Steps A and B for each reactor geometry. The results of the multiscale CFD model were validated by the experimental results from [60]. It was found that the inclined plate reactor produces a desirable distribution of precursor to the wafer and has faster cycle times for both Steps A and B compared to the other three reactor geometries studied.

Chapter 4

Multivariable Run-to-Run Control of Thermal Atomic Layer Etching of Aluminum Oxide Thin Films

4.1 Introduction

Due to the recent Covid-19 pandemic, there has been a transition from the real world to the digital world or the so-called "metaverse." As telecommuting increases, the demand for a wider range of electronic devices has surged alongside it. This trend also comes with a growing requirement for more robust and reliable computing power, which can only be obtained from high-performance semiconductors. In addition, the autonomous vehicle market is greatly expanding, and more vehicles are integrating additional safety and convenience features that require even more semiconducting materials. The combination of all of these factors has resulted in a growing demand for these valuable, ultra-high-performance semiconductors from various industries dur-

ing this Fourth Industrial Revolution. However, the semiconductor industry's fabrication output has not been able to meet this growing demand, and with the uncertainty of the Covid-19 pandemic, shortages are becoming more and more prevalent in the semiconductor market [109]. Thus, a more effective and efficient method for producing these semiconductors must be established to meet the growing consumer demand for these materials. Despite the pressure to continuing miniaturizing semiconductors, the continued shrinking of the fin dimensions of the fin field-effect transistors (FinFETs) has been obstructed by obstacles associated with the 5 nm node size [93]. Thus, Gate-all-around (GAA) transistors have become the most promising competitor to replace FinFET technology, as they are able to enter the sub-5 nm era.

Atomic layer etching (ALE), considered one of the most advanced etching techniques in semiconductor fabrication, is anticipated to be one of the key processes necessary to overcome this miniaturization issue [69]. ALE is an etching process that uses sequential precursor pulses in between purge steps to achieve self-limiting reaction behavior that etches away the substrate surface monolayers. Due to this behavior, ALE has received great attention as a promising etching process that is able to usher in the sub-5 nm era. Many researchers have investigated and demonstrated the technical viability of ALE processes with various materials such as Si [1], SiO_2 [75], MoS_2 [56], Si_3N_4 [64], and Al_2O_3 [61]. In particular, high dielectric materials, which can output a higher computing speed with a lower leakage current [51], have received tremendous attention. Recently, [61] has validated the self-limiting behavior of aluminum oxide in a thermal ALE process. These processes are particularly noticeable since they can ensure high uniformity and conformity of the thin film, thus resulting in ultra-smooth thin films. Nevertheless, the thermal ALE process of aluminum oxide has not been extensively studied, and thus, the process has not been fully characterized. The

thermal ALE of Al_2O_3 has been investigated microscopically [116] and from a multiscale modeling framework [120] that provides a multiscale perspective on the process, which was validated from the experimental results of [61].

Despite the advances made in ALE modeling, an optimal control scheme and guideline to perform process control for the ALE system are still lacking. For example, prior research has been conducted in the feedback control methods for silicon [21]. However, one of the most notable control methods in the semiconductor industry is run-to-run (R2R) control, which has the form of a batch system where a process recipe is modified in between the "batches" according to a predefined relationship between the inputs and outputs. Despite the fact that R2R control has been gradually incorporated into semiconductor fabrication facilities, the lack of quantitative and qualitative process information and nonlinear dynamic behaviors are some of the many reasons why semiconductor manufacturers are skeptical about integrating R2R control into ALE processes. In addition, a single R2R algorithm may not be sufficient to cover all possible disturbances such as process drift, shift, and other kinds of variability [81]. To overcome this issue, a combination of batch process control and feedback control has recently emerged in the semiconductor manufacturing industry [13]. The conjunctive strategy of R2R and feedback control has been studied by [114] and [123]. As another combined strategy, effective multi-algorithms for R2R control have emerged as robust, stable, and optimal control schemes in the semiconductor manufacturing industry [77]. In this research, multiple algorithms for multivariable R2R control are developed to ensure the conformity and uniformity of thin films and to adapt to disturbances that may impede the standard operating conditions in the context of thermal ALE of aluminum oxide thin films.

4.2   Multiscale CFD Modeling of Thermal ALE

The thermal atomic layer etching (ALE) for aluminum oxide (Al_2O_3) thin films can be simulated by multiscale computational fluid dynamics (CFD) modeling, which is an integrated form of microscopic and macroscopic modeling. This multiscale model is conducted similarly to prior research by [57] and [19]. Based on the kinetic Monte Carlo (kMC) method, the microscopic modeling composes of a general simulation of the kinetic components of the reaction under ideal conditions and is simulated through randomness to generate a kinetic model of the surface. For the gas transport domain, the previously developed inclined plate reactor is used and CFD simulations are performed using ANSYS Fluent 2021R2, which is a widely used commercial CFD software. For the surface domain, the process data from the CFD model is exported and transferred to the microscopic model to calculate the coverage or etching fraction in the atomistic point of view.

In the previous research [120], the multiscale CFD simulation included the influence of transport phenomena effects to various reactor models, which were examined for their effectiveness in delivering a uniform precursor flow pattern to the substrate surface and for the time to achieve complete coverage and etching. It was revealed that the inclined plate reactor performed optimally when comparing the types of reactors that were studied. Therefore, in this work, the multiscale CFD model of the inclined plate reactor design is adopted to develop and optimize a Multivariable R2R control system.

4.2.1 Microscopic Surface Modeling

The thermal atomic layer etching (ALE) of Al_2O_3 uses two precursor pulses in a sequential manner. Thus, there are two reaction steps in between the purge steps: Step A and Step B. In Step A, the wafer is exposed to hydrogen fluoride (HF), which fluorinates the aluminum oxide on the surface to produce an AlF_3 layer. Next, a purge step sweeps the remaining HF in the reactor to avoid undesired reactions and to guarantee self-limiting behavior. Then, the fluorinated surface (AlF_3) undergoes a ligand-exchange and is converted into a volatile species, dimethylaluminum fluoride [DMAF, $AlF(CH_3)_2$], when trimethylaluminum (TMA), $Al(CH_3)_3$, is supplied during Step B. The volatile layer, DMAF, is desorbed and thus, a monolayer of the substrate surface is removed. Lastly, another purge step removes the remaining TMA and residual products. A cyclical operation consists of the aforementioned steps, which is repeated until the desired film thickness is achieved. The overall reaction is described by

$$Al_2O_3 \text{ (s)} + 6HF \text{ (g)} + 4Al(CH_3)_3 \text{ (g)} \rightarrow 6AlF(CH_3)_2 \text{ (g)} + 3H_2O \text{ (g)}$$
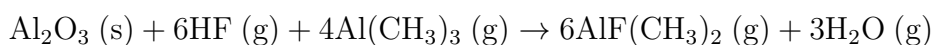
A microscopic surface model for etching of Al_2O_3 thin films was described in the previous work [116]. A brief description of the microscopic model is presented here. θ-Al_2O_3 ($\bar{2}$ 0 1) was found on Si(1 0 0) through the atomic layer deposition (ALD) process under annealing [12]. Thus, θ-Al_2O_3 ($\bar{2}$ 0 1) is used as the preferred lattice structure. A 300×300 lattice is applied to the microscopic model in which etch reactions take place. The surface kinetics in the atomistic level is modeled using the kinetic Monte-Carlo (kMC) method, in particular the variable step size method (VSSM) that is often called the n-fold way or BKL, which refers to the algorithm developed by

Bortz, Kalos, and Lebowitz [48]. The kMC algorithm was popularized by [40] who integrated the Monte Carlo algorithm into chemical kinetics. The kMC method is widely used to simulate individual reactions on the microscopic scale including the dependence on the lattice structure [68]. In the kMC algorithm, the total rate constant, k_{total}, is an important parameter that selects a reaction on the reaction site and calculates the time progression, which is computed as follows:

$$k_{\text{total}} = \sum_{i=1}^{N} k_i \tag{4.1}$$

where k_i is the reaction rate constant of the reaction i, and N is the number of reactions. After all reactions are defined and the total reaction rate constant is calculated, a random number, $\gamma_1 \in (0, 1]$, is chosen to determine the reaction on the site by using the criterion as follows:

$$\sum_{i=1}^{j-1} k_i \leq \gamma_1 k_{\text{total}} \leq \sum_{i=1}^{j} k_i \tag{4.2}$$

where j represents the reaction j. Finally, an additional random number, $\gamma_2 \in (0, 1]$, is generated to compute the time interval defined as:

$$\Delta t = \frac{-\ln(\gamma_2)}{k_{\text{total}}} \tag{4.3}$$

Detailed kinetic mechanisms for the fluorination and ligand-exchange reaction as well as their kinetic parameters, which were calculated using electronic structure optimization methods and Density Functional Theory, can be found in the research of [116].

4.2.2 Macroscopic Modeling

In the previous work, as shown in Figure 4.1a, a 3D computational fluid dynamics (CFD) model for the inclined plate reactor design was developed and evaluated by [120]. The optimized inclined plate reactor showed the uniform precursor distribution while preserving the fastest cycle time among other geometries. Thus, in this paper, the inclined plate reactor is adopted to formulate the control scheme. The reactor configuration consists of a cylindrical-shaped chamber with a 500 mm diameter and 10 mm height, a round-shaped inlet of 20 mm diameter, and a round-shaped outlet of 40 mm diameter. A substrate of 300 mm is placed at the center of the reactor. To optimize the precursor distribution, an arch-shaped inclined plate with 2 mm thickness and $5°$ angle from the horizontal is equipped near the inlet. Meshing Mode, a feature of ANSYS Fluent 2021R2, was utilized to generate the mesh for the plate reactor. Tetrahedral cells were adopted to reduce simulation time while maintaining the accuracy of the numerical calculations. Mesh quality criteria ranges recommended from [4] including orthogonality, skewness, aspect ratio, and resolution were standards that were integrated into the development of the plate mesh, which is illustrated in Figure 4.1b. The characteristics of the plate mesh as well as the quality criteria calculated from ANSYS Fluent 2021R2 are discussed in greater detail by [120].

ANSYS Fluent 2021R2 is utilized to conduct numerical computational fluid dynamics (CFD) calculations of the fluid flow and to simulate the surface kinetics of HF and TMA half-cycle in the ALE process. The pressure-based solver under transient mode in ANSYS Fluent is used with time step of 0.025 s and 200 iterations per time step. The coupled algorithm is used to decrease the computation time. The pressure-based solver in ANSYS Fluent solves mass and momentum

(a)   (b)
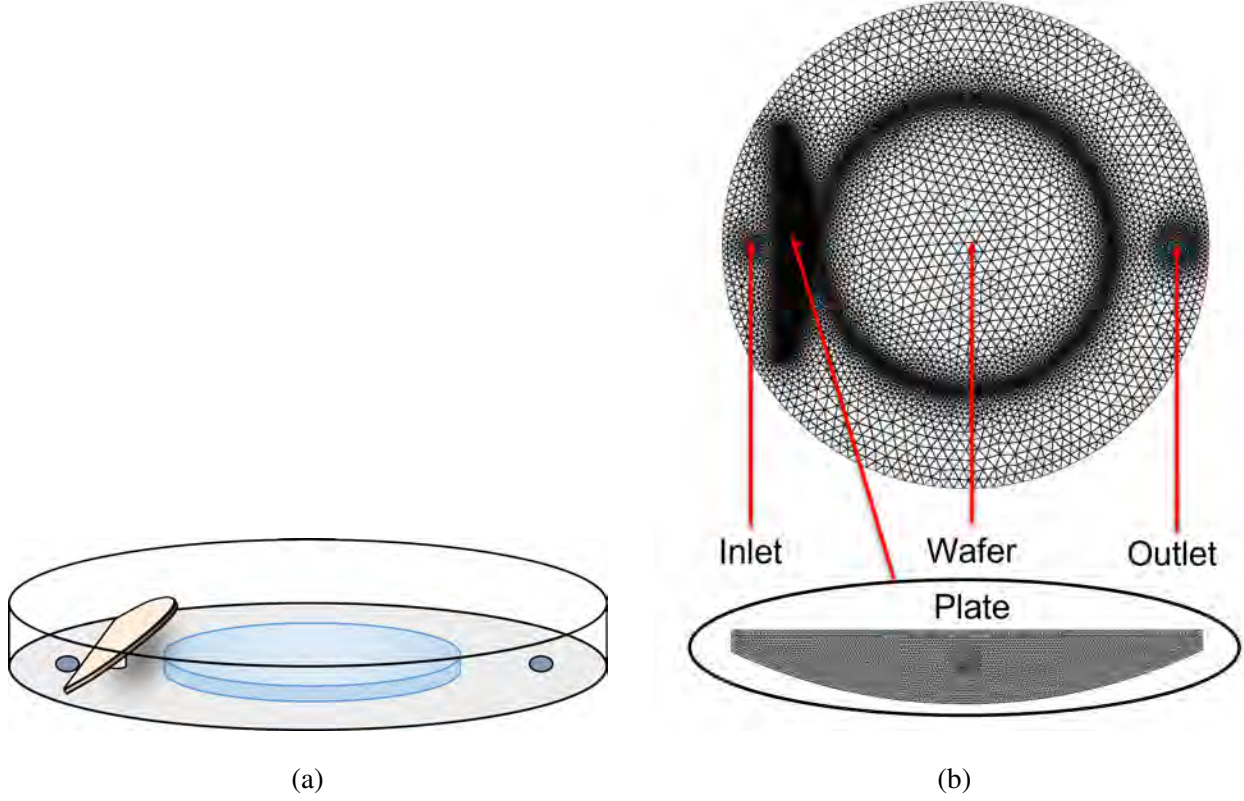
Figure 4.1: A schematic diagram of the inclined plate reactor (a) and the inclined plate reactor mesh generated from ANSYS Fluent (b) are illustrated from [120].

conservation equations, which are expressed as follows:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m \tag{4.4}$$

$$\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot (\overline{\overline{\tau}}) + \rho \vec{g} + \vec{F} \tag{4.5}$$

where ρ is the density of the mixture, $\vec{v}$ is the velocity of the mixture, S_m is the mass transfer source term, p is the static pressure, $\overline{\overline{\tau}}$ is a symmetric rank two stress tensor, $\rho \vec{g}$ is the gravitational body force, and $\vec{F}$ is the external body force. In addition, ANSYS Fluent solves the conservation

72

of energy equation which is described by:

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = -\nabla(\Sigma h_j J_j) + S_h \tag{4.6}$$

where E is the internal energy, h_j is the sensible enthalpy of species j, J_j is the diffusion flux of species j, and S_h is the heat transfer source term. To ensure more accurate results, in the pressure-based coupled solver, the second-order upwind scheme is used for spatial discretization, leading to higher-order accuracy through a Taylor series expansion. It is also required to calculate gradients to compute the values of scalars at the cells on the mesh. To calculate the gradients, the Incomplete Lower Upper (ILU) method is employed, which requires more computing power but offers better smoothing properties for the pressure-based coupled solver. Based on these solver settings, aforementioned equations are solved iteratively at a given time step. Once the solution satisfies the convergence criteria, the solver advances to the next time step.

To generate a more realistic profile, the consumption of precursors and generation of products are included in the CFD model by determining the reaction rate from the modified Arrhenius equation that includes the temperature dependence of the pre-exponential factor, which is defined as follows:

$$k_j = A_j T^{\beta_j} e^{-E_{A,j}/RT} \tag{4.7}$$

In the above equation, k_j is the reaction rate constant for reaction j, T is the temperature on the surface, β_j is the temperature exponent for reaction j, $E_{A,j}$ is the activation energy for reaction j, and R is the ideal gas constant. Both half-cycle reactions are developed in ANSYS Fluent, which

are described below:

$$Al_2O_3 \text{ (s)} + 6HF \text{ (g)} \rightarrow 2AlF_3 \text{ (s)} + 3H_2O \text{ (g)}$$

$$2AlF_3 \text{ (s)} + 4Al(CH_3)_3 \text{ (g)} \rightarrow 6AlF(CH_3)_2 \text{ (g)}$$
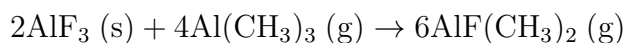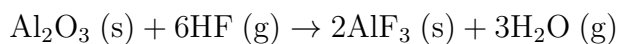
The monitoring of the surface temperature in real time is necessary to ensure that the temperature does not decrease, especially for Step A, which requires higher operating pressures for HF at lower temperatures [116]. Typically, the surface temperature is maintained to guarantee film quality, and thus, a PI (proportional-integral) controller is assumed to work appropriately in this process. In addition, the cyclical operation is carried out through a user-defined function (UDF) in which operating conditions and boundary conditions are specified. The detailed description for the macroscopic modeling can be found in the previous work of [120].

4.3 Multivariable R2R Control Formulation

Various algorithms have been employed in existing run-to-run (R2R) control algorithms for chemical vapor deposition (CVD) and chemical mechanical polishing (CMP) processes including the exponentially weighted moving average (EWMA), predictor-corrector control (PCC), and optimizing adaptive quadratic controller (OAQC) methods. For example, [114] utilized the EWMA and PCC methods on PEALD of HfO_2 thin films to compare their effectiveness and [19] developed an R2R controller using the EWMA algorithm for thin film Si-H solar cells. The EWMA controller, which is a linear approximation model-based controller, is widely utilized and inte-

grated into semiconductor fabrication processes due to its versatility to compensate for process shift, drift, and noise among other disturbances [15]. Run-to-Run controllers are adjusted through multiple batches by using statistical process control (SPC) and by tuning the controllers using engineering process control (EPC) to provide a new recipe (i.e., input) for the next batch run [28].

In this work, an EWMA-based multivariable R2R control scheme is developed using two inputs (the process time and the precursor flow rate) and two outputs (the coverage or etching fraction and the precursor partial pressure). In order to tune an EWMA-based R2R controller, the input-output relationship must be established first. A multi-input-multi-output (MIMO) model may be used to formulate the multivariable R2R controller. However, in this work, the process is locally approximated by two single-input-single-output (SISO) linear regression models, with each one of them being defined by an equation of the following form:

$$y_t = \alpha + \beta u_{t-1} \tag{4.8}$$

where y_t represents the output of the process for batch run t, α is the bias, β is the process gain, and u_{t-1} is defined as the recipe (or input variable) in between batch run $t-1$ and batch run t. The process parameters, α and β, are determined using the standard least squares method where α is the y-intercept and β is the slope of the linear regression model. R2R controllers are modeled under an evolutionary operation mode [11] that combines statistical results to improve efficiency and increase productivity. In other words, the control system works to adapt to changes in the process environment and tune the manipulated variables to sustain standard output conditions from the R2R

75

controller. The first step of the control work is to update the model, which is expressed below:

$$a_t = \lambda \left(y_t - b u_{t-1} \right) + (1 - \lambda) \, a_{t-1} \tag{4.9}$$

where λ represents the weight factor that is responsible for the translation of the regression model, a_t is the updated bias, and b is the process gain that is the same as β in Eq. (4.8). Typically, λ is chosen to be $0.1 \sim 0.3$ [13]. After the model is updated, the new recipe for the next batch run is computed by the following expression:

$$u_t = \frac{T - a_t}{\beta} \tag{4.10}$$

where u_t is the new recipe for the next batch run and T represents the target or desired value of the output variable.

The stability and robustness of the EWMA-based R2R controllers have been studied; however, it has been reported that the EWMA and PCC (also known as double-EWMA) methods underperform for nonlinear systems due to the linear model-based R2R algorithm [81, 77]. Nevertheless, the EWMA method is able to be used for nonlinear systems if linear regression models are well-developed for fitting nonlinear responses, which is discussed in this work. In this paper, two different R2R controllers are formulated and integrated, resulting in the Multivariable R2R control system in which each controller is modeled using a single-input-single-output (SISO) regression model as shown in Figure 4.2. The first R2R (denoted as R2R-1) tracks the reaction progression (reflected by the coverage or etching fractions) to adjust the process time. Typically, a quartz crystal microbalance (QCM) is used to monitor the etching rate through mass changes

along the surface of the thin film substrate [60] in real time, which allows the etching and coverage

fractions to be computed. The second R2R (denoted as R2R-2) manipulates the flow rate of the

precursors by monitoring the precursor partial pressure deviation from the desired partial pressure

at the standard operating conditions for the developed inclined plate reactor outlined in Table 4.1.

The following sections discuss the tuning methodology of R2R-1 and R2R-2 in greater detail.
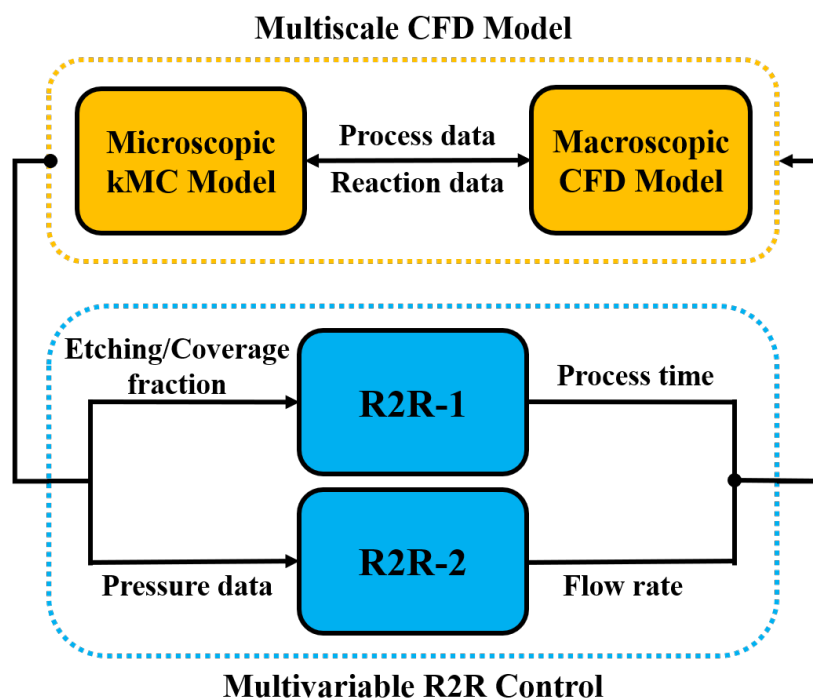
Figure 4.2: Multivariable run-to-run control of the inclined plate reactor.

4.3.1 R2R-1 Modeling and Tuning

The schematic process diagram shown in Figure 4.3 illustrates the Multivariable R2R control

system. R2R-1 is implemented to adjust the valve opening time (i.e, process time) for precursor

injection to the reactor to achieve complete AlF_3 coverage and etching for Steps A and B respec-

tively. Multiscale computational fluid dynamics (CFD) simulation data are collected at standard

Table 4.1: Standard operating conditions for the multiscale CFD simulation.

Standard Condition	Step A	Step B
Precursor Flow Rate (sccm)	150	70
Operating Pressure (Pa)	133	133
Temperature (K)	573	573
N_2 Flow Rate (sccm)	150	150
Process Time (s)	1.1	2.0

operating conditions for Steps A and B for the inclined plate reactor, which are summarized in Table 4.1. Process or valve opening times of 1.1 s and 2.0 s in Table 4.1 reflect the ideal time for the Al_2O_3 thin film substrate to reach complete coverage or etching respectively under the standard operating conditions. In this work, it is assumed that the time until the valve fully opens is negligible, and thus, the dynamics of the valve do not interfere with the process dynamics. In other words, R2R-1 only adjusts the process time; however, the flow rate is only adjusted by the upstream valve as shown in Figure 4.3, which is controlled by R2R-2. The R2R-1 controller is approximated to a single-input-single-output (SISO) model in which the process time (t) for the precursor injection serves as the recipe ($u = t$ in Eqs. 4.12 and 4.13) and the etch or coverage fraction (f) of AlF_3 serves as the output variable ($y = f$ in Eqs. 4.12 and 4.13). It is of paramount importance to determine a reasonable solution to the SISO regression model with low variance along the regression line since the solution has a great impact on the control work [77]. In this work, multiple regression models for R2R-1 are developed and evaluated using the standard least squares method between the result data and the solution of the SISO model. First, the standard linearization of the multiscale CFD simulation results for Steps A and B are presented in Figure 4.4.
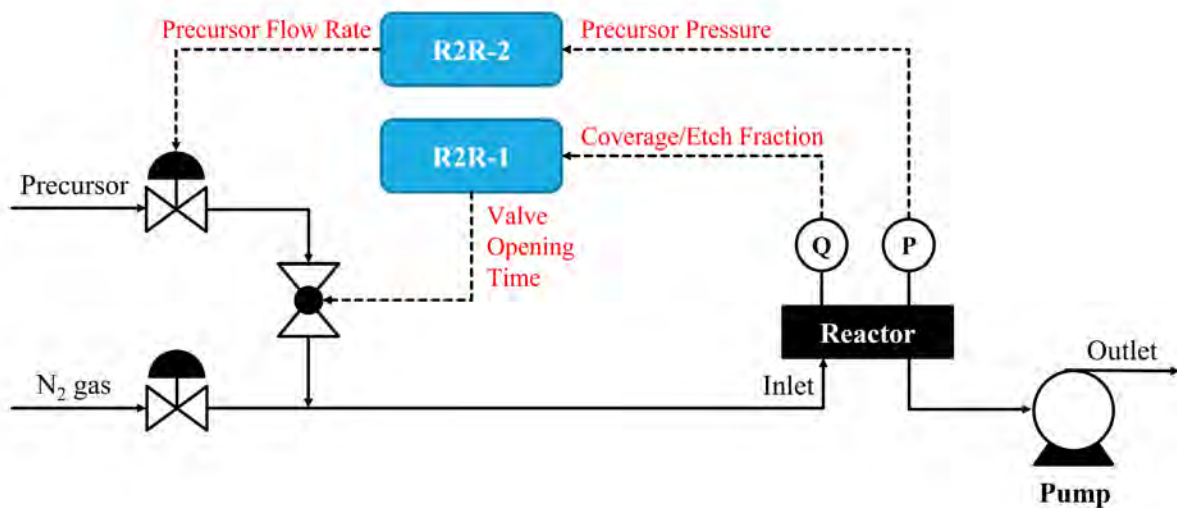
Figure 4.3: The Multivariable Run-to-Run control system.

(a)

(b)
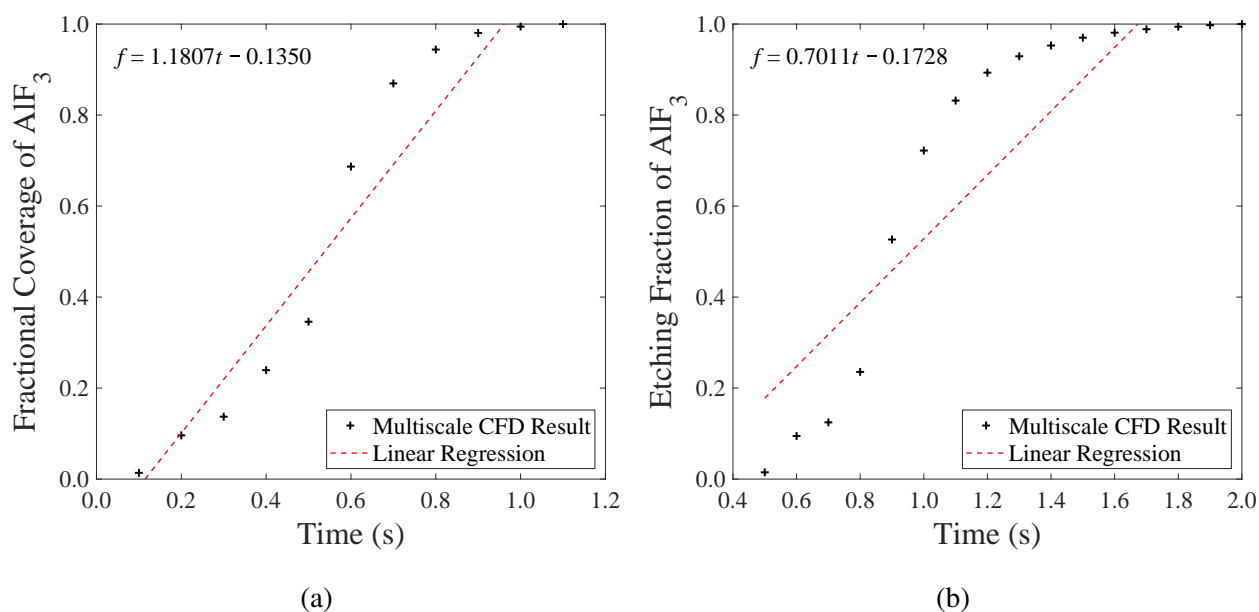
Figure 4.4: The input-output relationship between the fractional coverage of AlF_3 and the precursor valve opening time (a) for Step A and the etching fraction of AlF_3 and the precursor valve opening time (b) for Step B, which is derived from a standard linear regression model for the EWMA-based R2R controller, R2R-1. For Steps A and B, the R^2 values from Table 4.2 indicate marginal linear behavior for Step A and a lack of linear behavior for Step B.

Figure 4.4 reveals that the linear regression model does not fit the simulation data for Steps A and B due to their nonlinear process responses and the larger variance of the data from the regression line, thus the linear regression of the entire data set may be difficult to accomplish. To improve the least squares values and thereby strengthen the fit of the regression model to the multiscale CFD results, several techniques are used to achieve a better fit of the model data: piecewise and modified median-effect. For the piecewise regression model, the data curve is divided into two groups to generate two linear piecewise plots that have an intersection point (t_p, f_p) at a time of, t_p, which is expressed as follows:

$$f_p = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} \tag{4.11}$$

where α_1 and α_2 are the biases for the two plots, respectively, and β_1 and β_2 are the process gains, respectively. A conditional loop is necessary to ensure the correct piecewise regression model to use, which is based on the intersection point, f_p. The piecewise regression models are presented in Figure 4.5 and displays a marginal improvement of linearity for both precursor injection steps.

As another alternative technique to improve the curve fitting of the multiscale CFD results, the median-effect equation is adopted from prior research [16] and is applied to the formulation for the coverage and etching fraction progression with time. The median-effect equation is suitable for the EWMA method because of its logistic-like function behavior, which exemplifies the multiscale CFD data trend and its ability to transform both the coverage (or etching) fraction and the process time into logarithmic forms. The median-effect method can also be modified into a linearized function in the form of Eq. (4.8), which will be discussed later in this section. The median-effect
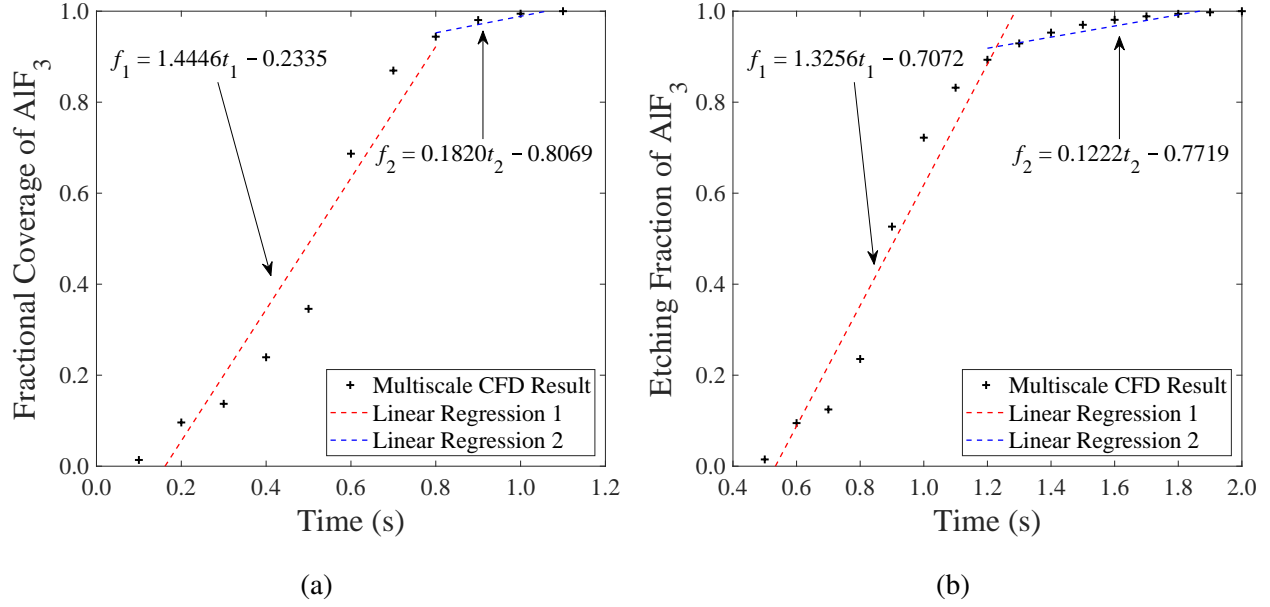
Figure 4.5: The input-output relationship between the fractional coverage of AlF_3 and the precursor valve opening time (a) for Step A and the etching fraction of AlF_3 and the precursor valve opening time (b) for Step B, which is derived from a standard linear regression model divided into linear piecewise functions for the EWMA-based R2R controller, R2R-1. The R^2 values from Table 4.2 indicate a marginal to moderate linear relationship of multiscale CFD data.

equation is defined as follows:

$$\frac{f}{1-f} = \left(\frac{t}{\eta}\right)^{\delta} \tag{4.12}$$

where f is the coverage or etching fraction, t is the process time, and the constants, η and δ, are obtained by regressing the median-effect equation into a linearized form to obtain the slope and y-intercept. It is also notable that f can be solved algebraically, which demonstrates the practicality of integrating the median-effect equation into the EWMA model. In this research, additional tuning constants (γ and ϵ) are introduced to reduce the variance of the data to generate a "modified" median-effect equation, which is defined below.

$$\frac{\gamma f}{1-\gamma f} = \left(\frac{t+\epsilon}{\eta}\right)^{\delta} \tag{4.13}$$

81

In particular, γ and ϵ are adjustable parameters for modifying the median-effect equation in Eq. (4.12) and are determined by tuning the factors until a desirable R^2 value of the linear regression is obtained. The original median-effect equation from [16] is obtained by declaring $\gamma = 1$ and $\epsilon = 0$ in Eq. (4.13). The parameter, γ, is bounded such that $\gamma \in (0, 1]$, and is employed to shift the upper horizontal asymptote of the median-effect regression while ϵ is utilized for translating the regression line along the ordinate direction. The linearized form of the modified median-effect equation exemplified by [17] is derived below:

$$\ln\left(\frac{\gamma f}{1 - \gamma f}\right) = \delta \ln(t + \epsilon) - \delta \ln(\eta) \tag{4.14}$$

Plotting the left-hand side of Eq. (4.14) as a function of $\ln(t + \epsilon)$ can generate a linearized plot to determine the values of η and δ through linear regression where $\beta = \delta$ and $\alpha = -\delta \ln(\eta)$ from Eq. (4.8). The regression models of the modified median-effect and median-effect equations for Steps A and B are presented in Figure 4.6. For Step A, values of $\gamma = 0.99$ and $\epsilon = 0.35$ are used and for Step B, values of $\gamma = 1.0$ and $\epsilon = 0$ are used, and hence, Step B is represented by the original median-effect model. The linear modified median-effect regression model in Figure 4.6a shows that the transformed multiscale CFD dataset exhibits a more linear relationship compared to that of the linear and piecewise regression models. The regression model also correlates the data from the multiscale CFD simulation reasonably, which is observed in Figure 4.6b. The modified median-effect is not required for Step B since the median-effect (i.e., $\gamma = 1$ and $\eta = 0$) has an R^2 value close to unity and the multiscale CFD results have low variance. The results from the median-effect regression model for Step B are displayed in Figures 4.6c and 4.6d. The trans-

formation of the multiscale CFD data improves the linearization of both the linear (Figure 4.4b) and piecewise (Figure 4.5b) regression models, which is validated by an R^2 value of 0.9854 and exhibits a strong least squares fit. In the simulations, a given output variable is logarithmized to compute a logarithmized recipe for the next batch run, while a logarithmized recipe is calculated from Eq. (4.14) and the antilogarithmic form is used to simulate the process.

In conclusion, the SISO regression models of R2R-1 controllers for both half-cycles are obtained using three different regression methods and the parameters are presented in Table 4.2. As a result, the modified median-effect method provides better regression models for the SISO regression and overcomes the nonlinear relationship of the multiscale CFD data. The performances of the three models are evaluated in Section 4.4.

Table 4.2: A comparison of the standard linear, piecewise, and modified median-effect regression model parameters that are calculated from the standard least squares method for Steps A and B.

R2R	Half-Cycle	Regression Model	R^2	Process gain	Bias
R2R-1	Step A	Linear Model	0.9251	1.1807	−0.1350
		Piecewise-1	0.9366	1.4446	−0.2335
		Piecewise-2	0.8688	0.1820	0.8069
		Modified Median-Effect	0.9740	7.5727	1.5184
	Step B	Linear Model	0.8019	0.7011	−0.1728
		Piecewise-1	0.9520	1.3256	−0.7072
		Piecewise-2	0.8531	0.1222	0.7719
		Modified Median-Effect	0.9854	7.3701	0.8523
R2R-2	Step A	Linear Model	0.9651	0.8286	−145.6438
	Step B	Linear Model	0.9398	1.2388	−74.8118
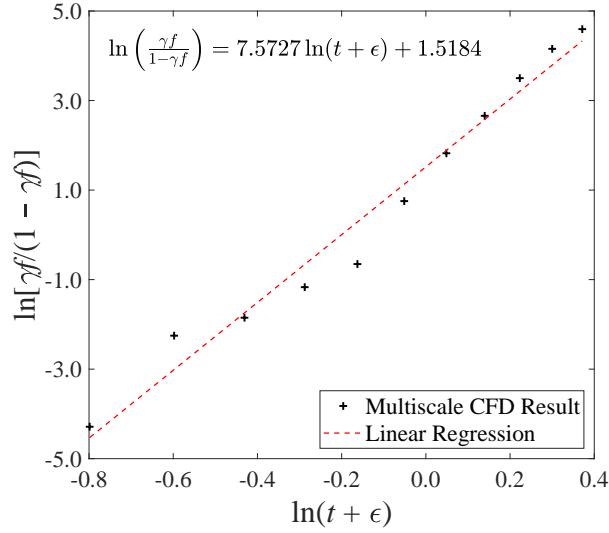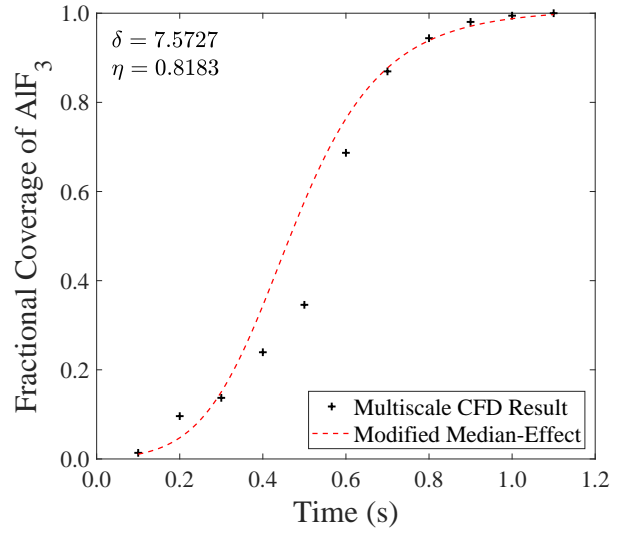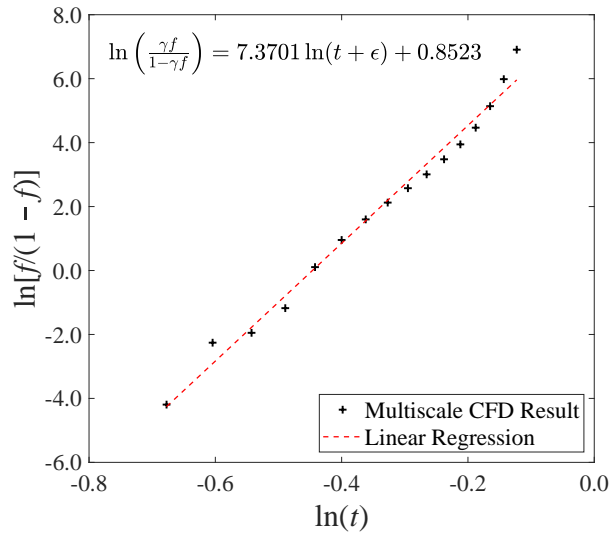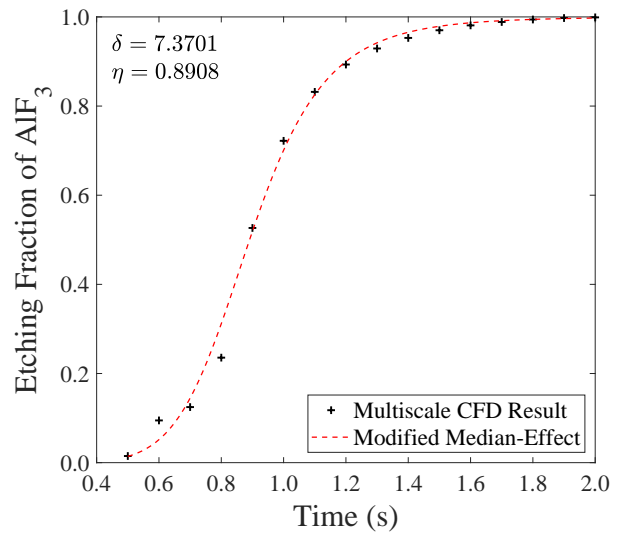
Figure 4.6: The input-output relationship with the logarithmized multiscale CFD data and logarithmized time from the modified median-effect equation for Step A (a) with $\gamma = 0.99$ and $\epsilon = 0.35$ and for Step B (c) with $\gamma = 1.00$ and $\epsilon = 0$ for the EWMA-based R2R controller, R2R-1. The R^2 values in Table 4.2 indicate a strong linear relationship for Steps A and B. The multiscale CFD results with the standard modified median-effect equation are presented in (b) and (d) for Steps A and B, respectively.

4.3.2 R2R-2 Modeling and Tuning

R2R-2 is developed to minimize the precursor partial pressure deviation from the ideal partial pressure at the standard operating condition within the reactor by adjusting the precursor flow rate to the reactor. The R2R-2 controllers for both half-cycles are also approximated to SISO models with the precursor flow rate, V, serving as the recipe ($u = V$), and the sum of the partial pressure deviation at various times, E_s, is defined as the output variable ($y = E_\mathrm{s}$). The controlled variable for R2R-2 is expressed as the following:

$$E_\mathrm{s} = \sum_{i=1}^{4} C_i \left[P_\mathrm{m}(t_i) - P_\mathrm{d}(t_i) \right] \tag{4.15}$$

where $P_\mathrm{m}(t_i)$ refers to the measured partial pressure at time t_i for four times of 0.1 s, 0.2 s, 0.4 s, and 0.6 s, and $P_\mathrm{d}(t_i)$ refers to the desired partial pressure obtained from the standard operating pressure condition that is outlined in Table 4.1 without any disturbances. C_i represents a coefficient that is selectively chosen to improve the linearity of the recorded pressure deviation at the four times. A pressure measuring device such as a pressure sensor can be implemented to monitor the pressure within the reactor at various times to ensure that the sum of the partial pressure deviations is reduced to prevent further loss of control of the system and degradation of the substrate materials. Therefore, R2R-2 can effectively function with the sum of pressure deviations as the output parameter, y_t, and as the precursor (HF and TMA) flow rates serving as the recipe, u_t. As illustrated in Figures 4.7a and 4.7b, the SISO regression models fit the multiscale CFD data well using the coefficients of 1.0 for all times for Step A and 1.0 for all times for Step B except for 2.0 for the first timestep ($C_1 = 2.0$) for Step B. The solutions of the SISO model have R_2 values of

0.9651 and 0.9398 for Steps A and B, respectively.
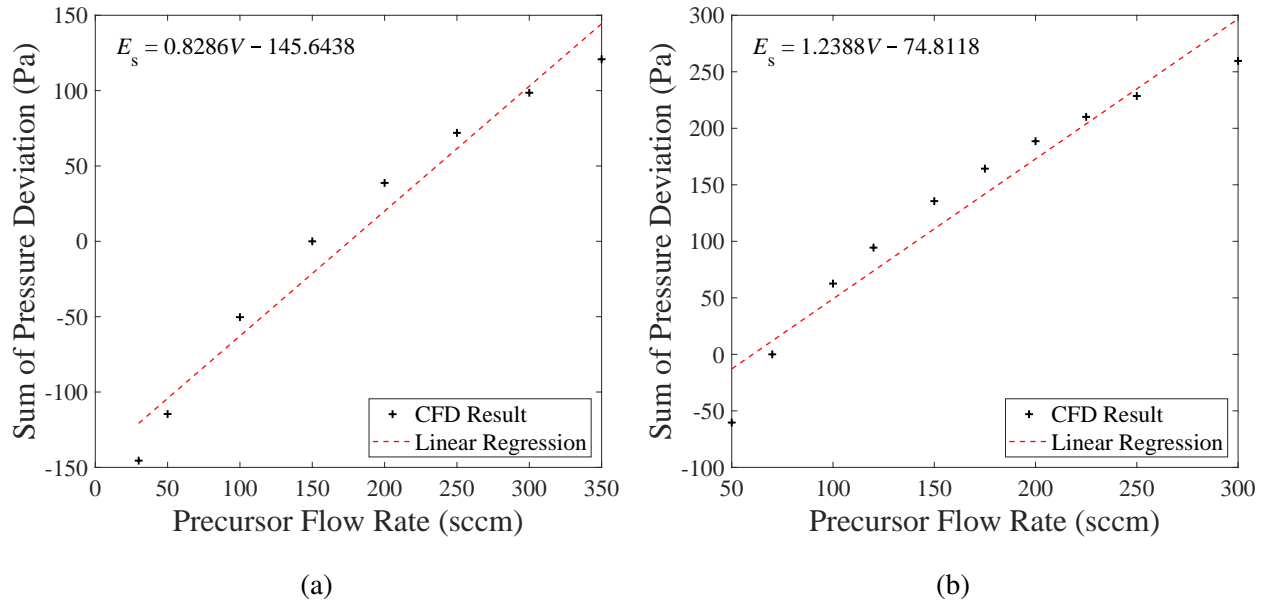
Figure 4.7: The input-output relationship between the sum of partial pressure deviations and the flow rate for Step A (a) and Step B (b), which is derived from a standard linear regression model for R2R-2. The R^2 values from Table 4.2 indicate a moderate linear relationship.

4.4 Simulation Results and Discussion

Semiconductor manufacturing processes can be subject to some unmeasurable drifts or shifts from wall deposition and equipment aging in the semiconductor industry. In this work, two different disturbances are simulated to evaluate the performance of the multivariable R2R control system. The first disturbance is a shift (referred to as a "kinetic disturbance"), which is able to be driven by cyclical operations, machine maintenance or changes in process settings [77]. It can be implemented simply by multiplying the reaction constants by a factor that is less than 1. The other disturbance (referred to as a "pressure disturbance") is simulated by reducing the operating pressure, which can be caused by a malfunction of the vacuum pump. In particular, R2R-2

86

is designed to deal with any pressure deviation from the standard conditions. A threshold value of 0.999 is set as the target for the fractional coverage for Step A and of the etching fraction for Step B. The multiscale computational fluid dynamics (CFD) simulation is performed by using 24 parallel computing processors with 384 GB memory on a compute cluster. Simulation time for a half-cycle takes half an hour on average. The batch-to-batch calculation of the control action is not demanding compared to the multiscale CFD simulation, and thus, it is considered negligible.

Initially, the R2R-1 controllers for both half-cycles are simulated under a kinetic disturbance using the multiscale CFD model to evaluate the regression models discussed in Section 4.3.1. After a regression model with a better control performance is selected, the Multivariable R2R control system is simulated, evaluated, and then compared with the single R2R control system (R2R-1 and R2R-2) under the two disturbances. As shown in Figure 4.8, the modified median-effect regression model outperforms the linear and piecewise regression models under a kinetic disturbance. The kinetic disturbance is modeled such that the reaction rate constants are multiplied by a shift factor of 0.7 in the kMC model when process time is updated, leading to a reduction in reaction rates. The median-effect regression models reach the target for Step A and Step B at batch runs of 8 and 9 respectively. However, the other two models do not reach the target value until a batch run of 30. One of the major concerns of process control is the possibility that the controllers may dramatically overshoot the target threshold. However, Figure 4.8 reveals that the modified median-effect models not only reach the target in a reduced number of batch runs, but also when they approach the target, the controllers add sufficiently small increments to the new recipe to prevent the possibility of overshooting the target. In other words, controller response overshoots and rapid control actions are not allowed in the process due to the best curve fit, and thus, the median-effect

87

regression model is selected for the R2R-1 controllers.

(a)

(b)
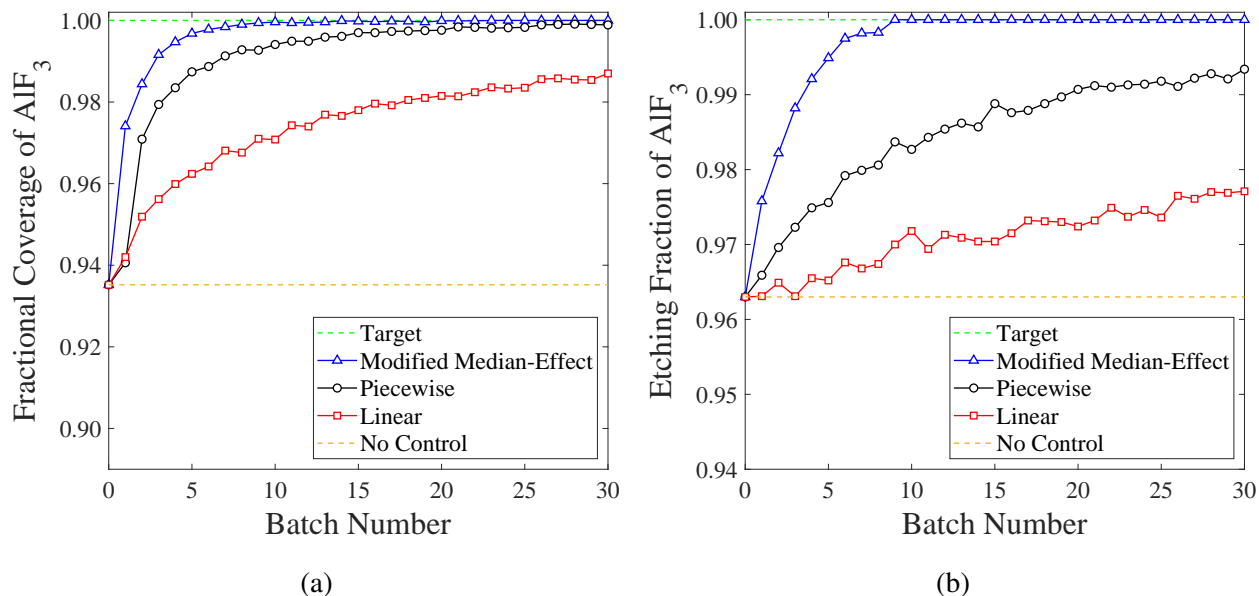
Figure 4.8: Comparison of the responses for various regression methods of R2R-1 under the presence of a kinetic disturbance for Steps A (a) and B (b). The weight factors (λ) of 0.3 and 0.1 are used for Steps A and B, respectively.

The results of the Multivariable R2R and individual R2R control systems under the two shift disturbances for Steps A and B, respectively, are presented in Figure 4.9. To model the kinetic disturbance, the factors of 0.55 and 0.7 are multiplied to the reaction rate constants for Step A and Step B, respectively, and the pressure disturbance is simulated by reducing the operating pressure of the reactor to 40 Pa from 133 Pa. For Step A, both individual R2R configurations reveal that the effects of the disturbances are not able to be removed within 15 batch runs as shown in Figure 4.9a. For Step B, as illustrated in Figure 4.9b, R2R-1 is able to mitigate the impact of the disturbances within 7 batch runs. However, R2R-2 is not able to eliminate the effects of the disturbances within 15 batch runs. In contrast to individual R2R controllers, the multivariable R2R controller is able to successfully mitigate the effect of the disturbances within batch number 5 for Step A in Figure 4.9a

and within 3 batch runs for Step B in Figure 4.9b. As a result, the multivariable R2R configuration of R2R-1 and R2R-2 performs more efficiently since the lesser number of batch runs is required to approach the complete coverage or etching, which is represented by the target line.
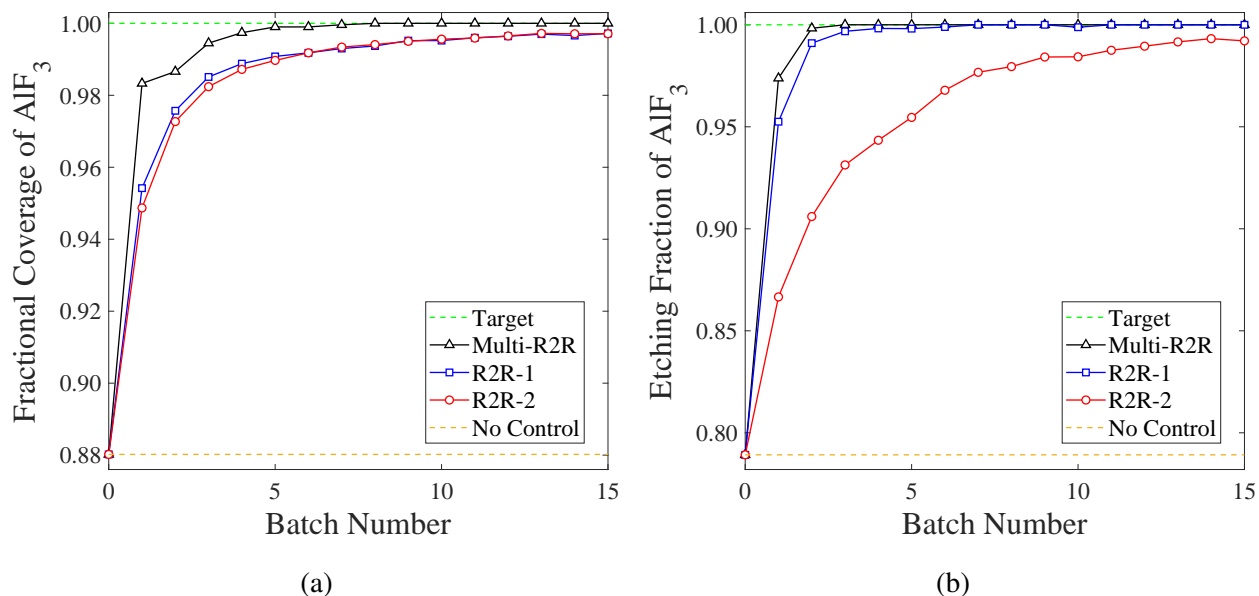
Figure 4.9: Comparison of the responses of various R2R control systems under the presence of a kinetic and pressure disturbance for Steps A (a) and B (b). The weight factor (λ) of 0.3 is chosen for all case studies. R2R-1 and the multivariable R2R system are simulated with the modified median-effect regression model.

The multivariable R2R control algorithm is also efficient in regard to adjusting the input variables: process time and precursor flow rate. Figure 4.10 shows the corresponding input values from the control work in Figure 4.9. For both half-cycles, R2R-1 continuously increases the process time to compensate for the effect of the disturbances, as can be seen in Figures 4.10a and 4.10c. However, In Step B, R2R-1 does not update the recipe after batch run 9 since it achieves a full etching fraction. R2R-2 also continues to increase the flow rate to get the target as shown in Figures 4.10b and 4.10d. In this work, there is no limit to the precursor flow rate even though there are valve opening constraints in the semiconductor industry. As predicted, the multivariable R2R

algorithm manipulates the two input values such that both parameters (process time and flow rate) are significantly less than those of R2R-1 and R2R-2. Therefore, it is demonstrated that the multivariable R2R control system exhibits a stronger performance compared to that of the individual R2R-1 and R2R-2 control systems.

4.5 Conclusion

In this work, the previously developed multiscale computational fluid dynamics (CFD) model for the thermal atomic layer etching (ALE) of aluminum oxide thin films was employed to build a multivariable R2R control system using an integration strategy. First, two individual R2R controllers (R2R-1 and R2R-2) were formulated based on the single-input-single-output (SISO) regression method using the exponentially weighted moving average (EWMA) algorithm. To significantly increase the linearity of the SISO model, a novel regression method, the modified median-effect, was implemented and compared with standard linearization and piecewise linearization. The modified median-effect method outperformed the other regression models and demonstrated the best fit of the multiscale CFD data using the EWMA method for the nonlinear system. R2R-1 was designed to adjust the valve opening time of precursor release to the reactor by measuring the coverage or etching fractions on the wafer, while R2R-2 was formulated to maintain the desired partial pressure of the precursor by manipulating the precursor flow rate into the reactor. Kinetic and pressure disturbances, which are industrially-relevant disturbances, were introduced to the system to determine the effectiveness of various R2R control algorithms. Consequently, this study substantiates that the multivariable R2R control scheme is able to successfully achieve complete
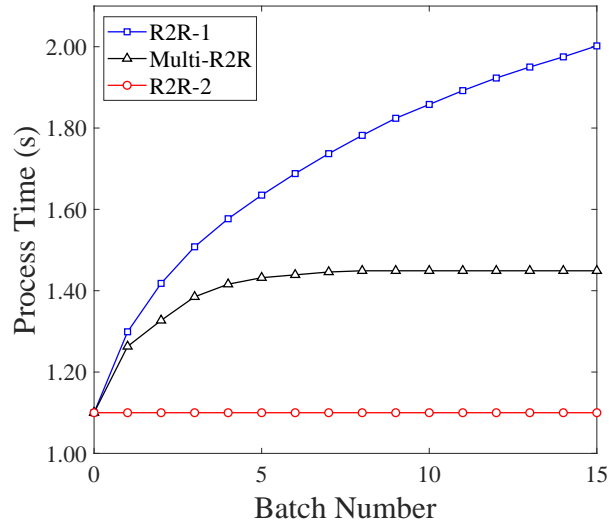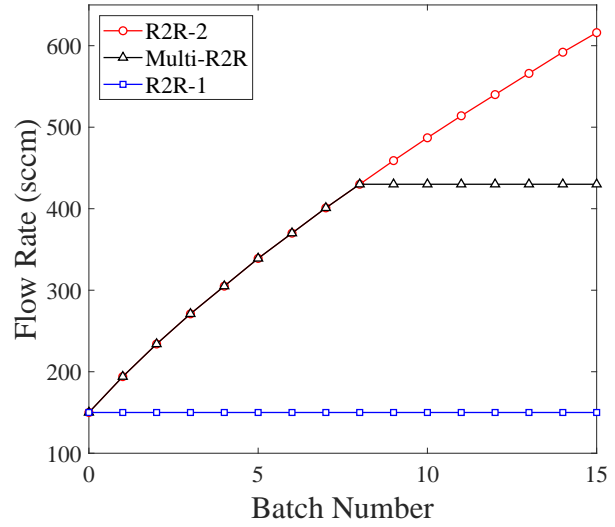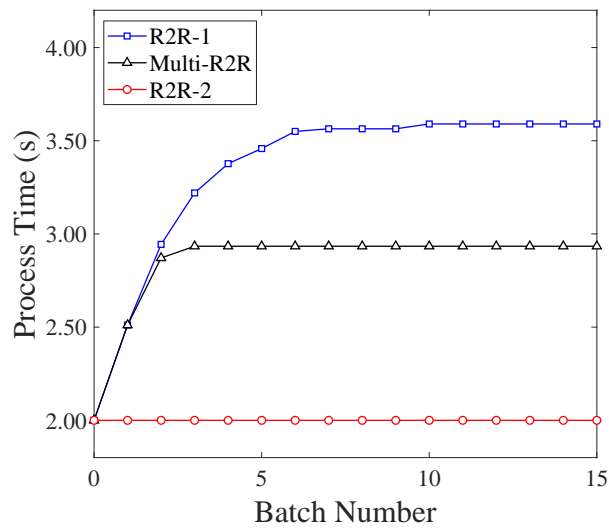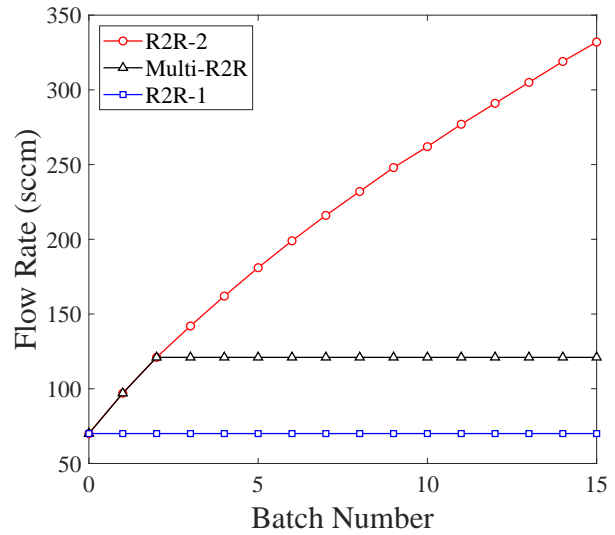
Figure 4.10: Progression of the adjustments made to the recipes (process time, precursor flow rate) in the presence of a kinetic and pressure disturbance through various EWMA-based R2R control systems for Steps A (a-b) and B (c-d).

coverage and etching and overcome the effects of disturbances in the least number of batch runs

compared to that of the individual R2R control schemes implemented only one at the time.

Chapter 5

Multiscale Computational Fluid Dynamics Modeling of Spatial Thermal Atomic Layer Etching

5.1 Introduction

Thermal atomic layer deposition (ALD) is a premature process that is gradually being integrated into the semiconductor production industry and is known to be effective at depositing layers of material on the surfaces of semiconductor substrates. Thermal ALD uses gaseous reactants called precursors that are introduced in sequential pulses. The precursors make modifications to the surface of the substrate and introduce a depositing film to generate the thin film layer [52]. Meanwhile, thermal atomic layer etching (ALE), a bifurcation of ALD, is a process that etches the substrate material by integrating gaseous reactant pulses. The reactant pulses convert the surface of the substrate into an active reaction layer and etch the modified surface layer [52]. For the ther-

mal ALD and ALE processes, the precursor pulses are separated by purging steps with a purge gas to flush the residual products from the reaction chamber. The applicability of ALD and ALE to semiconductor fabrication is based on the notion that these processes exhibit self-limiting behavior. This characteristic behavior prevents precursor adsorption beyond the substrate surface where the surface layer exemplifies a transport-limited barrier. In other words, the processes are restricted to a single atomic layer (monolayer) of the substrate material that can be deposited or etched for each cycle [53]; thus providing substantial controllability over the thickness of the substrate in the Ångstrom scale.

Despite the capabilities of ALD and ALE in precise substrate thickness control, conventional ALD and ALE processes are time-consuming and counterproductive, making these processes undesirable for some applications that require high productivity. Conventional ALD and ALE require long purging steps to ensure self-limiting behavior is achieved. To overcome this obstacle, rapid atomic layer [125] and spatial atomic layer processes for ALD [90, 31] have recently emerged as alternative methods for achieving high deposition or etching rate, compared to conventional ALD and ALE, in a reduced cycle time. However, for rapid atomic layer etching, researchers observed a degradation in the film roughness [125]; thus spatial atomic layer processes are considered in this work as they are more reliable in maintaining thin film conformance [63]. In spatial atomic layer deposition processes, the precursors are continuously dosed in physically separated reaction zones between the purging regions [79] in order to permit precursor reactions in sequential steps. For the purposes of this paper, spatial atomic layer deposition (SALD) and spatial atomic layer etching (SALE) are referred to as spatial thermal ALD and spatial thermal ALE, respectively. The substrate is transferred through each region at a constant velocity and maintains a presence in the

93

precursor regions that optimize the amount of material that is deposited onto the surface of the substrate. A schematic interpretation of the differences between conventional and spatial ALD/ALE processes are presented in Figure 5.1. Figure 5.1a resembles a conventional type ALD/ALE reactor where the substrate is stationary and consists of a single inlet and outlet. In a typical reactor for ALD/ALE, precursors are introduced separately into the reactor chamber, which is purged after each precursor injection. Figure 5.1b shows the patented sheet-to-sheet (S2S) SALD/SALE reactor where the substrate moves between adjacent zones that are injected with inert gases or precursor species. There have been developments in SALD and SALE reactor design. The original design of the spatial ALD was developed by [106], which was classified as a roll-to-roll type. Afterwards, rotating [100, 2, 105] and sheet-to-sheet models [34] have been developed.

(a)   (b)
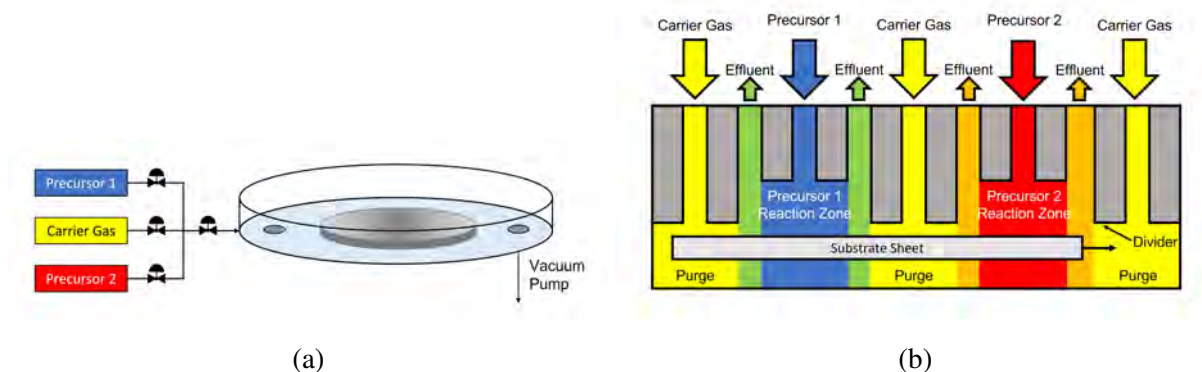
Figure 5.1: Schematic diagrams of a conventional ALD/ALE reactor (a) and a spatial ALD/ALE reactor (b). Conventional ALD/ALE processes feed precursors in sequential steps while the substrate remains stationary in contrast to spatial ALD/ALE processes, which feed precursors continuously in separate zones while the substrate sheet moves at a constant velocity through each region.

Although ALD and ALE processes are practical for the fabrication of thin-film wafers, these processes have not been fully investigated and characterized for integration into industry. To conduct experiments in a laboratory setting is costly, time-consuming, and difficult to study un-

der controlled environments due to the spontaneous nature of the reactions, which occur semi-instantaneously. Thus, a computational simulation approach can be conducted to reflect laboratory-setting experiments to collect a diverse data set that can improve the characterization of these ALD and ALE processes in a cost-effective and timely manner. For instance, [85] and [99] have accomplished computational fluid dynamics (CFD) modeling of ALD processes. To improve upon the CFD modeling, adopting a combined microscopic simulation using density functional theory (DFT) and kinetic theory with a macroscopic computational fluid dynamics (CFD) simulation, a "Multiscale" model of a laboratory-setting experiment can be conducted. Previous research has already been conducted on multiscale CFD simulation for chemical vapor deposition (CVD) and ALD processes. For example, [20] and [122] have performed a 3D multiscale CFD simulation for plasma-enhanced chemical vapor deposition (PECVD) and plasma-enhanced atomic layer deposition (PEALD), respectively, but their research is limited in the integration of the microscopic surface domain and the macroscopic gas-phase domain. Prior multiscale simulation was conducted in a two-step process such that CFD simulation of the macroscopic gas phase was run to the final flow time to extract the steady-state process parameters, which were later imported into the microscopic model to calculate the deposition rates. This method unreliably connects the microscopic and macroscopic domains but was previously adopted due to the complexity of the numerical simulation for a complete multiscale simulation, resulting in increased computation time. Also, prior research from [120] on the thermal ALE of aluminum oxide thin films adopted a method to include constant species consumption and generation terms in the CFD model to produce a stronger connection between the macroscopic and microscopic domains. However, this prior research was limited in computational efficiency due to the number of nodes specified in the surface substrate

95

mesh, which prevented the development of a fully autonomous multiscale CFD model that updates the source generation and consumption terms through each multiscale loop. Thus, this work makes enhancements to the previously developed multiscale CFD model by simplifying the number of nodal regions on the surface of the substrate to allow this connection between the microscopic and macroscopic domains to exist.

Several control and quality parameters are used to determine the optimality of the reactor configuration and process operation. Several studies [86, 18, 66] have been conducted on the gap distance between the top surface of the substrate to the bottom surface of the injection dividers, the purge gas flow rate, the precursor flow rates, the substrate velocity, and the vacuum pump pressure. The role of the gap distance serves to generate a finite boundary layer to prevent the intermixing of the precursor species, while the remaining parameters are crucial to the kinetics of the half-cycle reactions and to the self-limiting behavior of the half-cycle reactions. The afore-mentioned properties will be used to determine their impact on the quality of the substrate, which is determined by the amount of etching per cycle (EPC) and the surface uniformity of the sub-strate after the etching process. Thus, this research aims to advance from a previously developed microscopic model for the ALE of aluminum oxide (Al_2O_3) thin films [116] and to develop a mul-tiscale CFD model of a SALE reactor using an multiscale computing method that improves the algorithms from [120, 20, 122] and optimize the process operation by studying the effects of the aforementioned parameters on the quality of the substrate film.

5.2   Multiscale CFD modeling of SALE

Prior research on computational fluid dynamics (CFD) modeling of spatial atomic layer deposition (SALD) has been conducted to discuss the features of SALD and to optimize the process [22, 86, 18, 66, 23, 83]. The aforementioned projects have investigated the design of SALD reactors by examining the effects of the substrate velocity [22, 83], the gap distance [22, 86], the pitch distance between adjacent injectors [18], the precursor flow rate [86, 18, 23], and the purge gas flow rate [23]. These investigations provide invaluable information for SALD; however, they do not discuss the surface kinetics from the microscopic perspective. In addition, the relationship of the dynamic effects of the moving substrate on the film uniformity has not been investigated as conformity is an essential parameter to discuss the quality of the finished product. Thus, a complete survey of a spatial reactor configuration is required to include all technical information while accounting for the multiscale point of view to provide theoretical guidance toward further operation and control of a SALD/SALE reactor and process. Therefore, this research aims to build a multiscale CFD model by integrating a microscopic model for the surface kinetics and a macroscopic model for the transport phenomena effects to offer a comprehensive understanding and to investigate key parameters to optimize the reactor configuration and operation.

The combination of the microscopic surface domain and the macroscopic gas-phase domain generates the so-called multiscale computational fluid dynamics (CFD) model. For the microscopic surface domain, aluminum oxide thin films are modeled on the basis of a kinetic Monte Carlo (kMC) algorithm that is initiated using a Python script code. For the macroscopic gas phase domain, ANSYS Fluent 2021R2 is used to simulate the two-dimensional (2D) CFD of the gaseous

97

species. The schematic concept of the multiscale CFD model is illustrated in Figure 5.2 where input parameters including the substrate velocity, precursor (HF and TMA) flow rates, and vacuum pressure are specified to the macroscopic CFD simulation to calculate the etching per cycle. In the multiscale CFD model, Fluent first simulates the gas-phase domain at every time step to generate the process data (surface pressure and temperature) on the substrate surface. Then, the process data is transferred to the microscopic model, which simulates the surface reactions with a lattice model defined as a 300×300 grid and updates the surface boundary conditions in Fluent based from the result of the kMC simulation. This cyclical operation is executed in a Linux shell script such that the macroscopic and microscopic simulations are carried out consecutively using 36 compute cores with 192 GB memory through a computer cluster network with each computation of the multiscale CFD simulation taking 2 hours on average.

5.2.1 Microscopic surface domain

In the spatial atomic layer etching (SALE) of aluminum oxide thin films, two half-cycle reactions occur at different locations where the precursor species, hydrogen fluoride (HF) and trimethylaluminum [$Al(CH_3)_3$, TMA], are injected continuously at constant flow rates. First, the top surface of the substrate is fluorinated when the substrate is transported through the hydrogen fluoride reaction zone. The HF half-cycle produces a modified surface layer composed of aluminum fluoride (AlF_3). It is assumed that the adsorption occurs in a self-limiting behavior such that the HF precursor cannot permeate beyond the surface of the substrate. Next, an adjacent vacuum port exhausts the remaining HF and the byproduct, H_2O, via a vacuum pump. An inert zone using a barrier gas, N_2, is also integrated into the reactor configuration to remove the residual HF
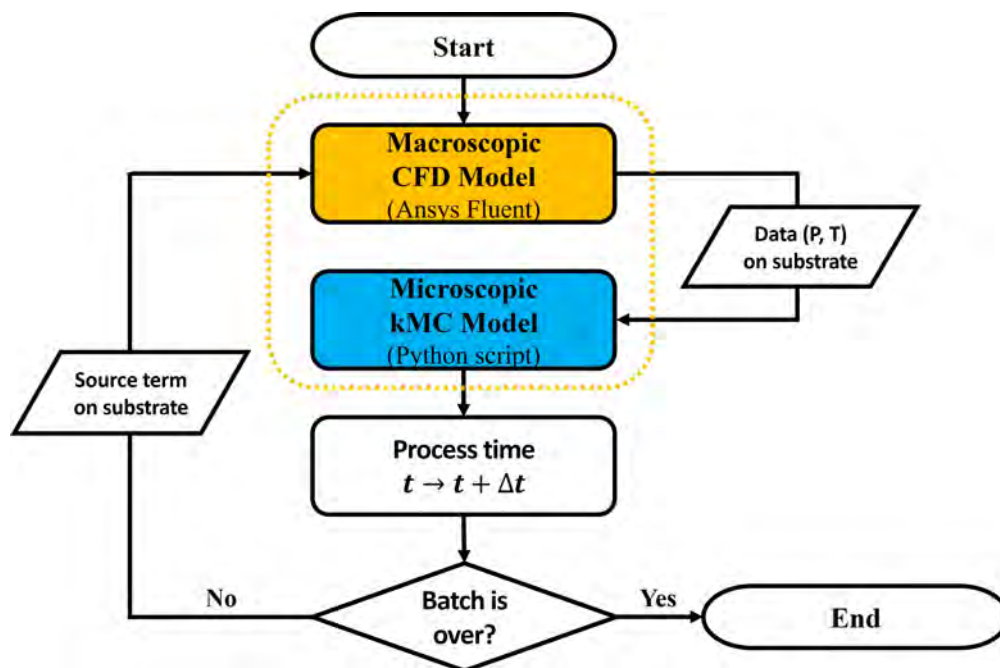
Figure 5.2: A process diagram of the multiscale computational fluid dynamics (CFD) workflow. The macroscopic model using ANSYS Fluent runs with a time step size of 0.0005 s and transfers precursor partial pressure and temperature data on 18 nodes of the substrate to the microscopic model. The microscopic model simulates the surface kinetics on the substrate using the kMC algorithm. The kMC code calculates the source terms (consumption and generation of species), which enables the macroscopic model to consume or generate the corresponding amount of species on a surface area-average on 18 nodal regions of the substrate. Also, the kMC script calculates the time evolution, Δt, which allows the macroscopic model to compute the exact number of time steps to run the CFD simulation. Lastly, the macroscopic and microscopic simulations are connected by a Linux shell script, resulting in an multiscale CFD simulation, which is halted until the substrate reaches the end of the reactor.

on the surface of the substrate. Following the purging step, a volatile species, dimethylaluminum fluoride [DMAF, $AlF(CH_3)_2$], is produced following a ligand-exchange reaction between the second precursor, TMA, and the AlF_3 surface layer. In the next vacuum line, the residual TMA and DMAF are withdrawn from the reaction chamber to maintain the selectivity of DMAF. Finally, an adjacent inert zone containing N_2 gas pushes the trace TMA to the vacuum port to maintain self-limiting behavior. Unlike conventional thermal atomic layer etching (ALE), SALE achieves

a cyclical operation of thermal ALE as the substrate moves under an injection assembly. The schematic diagram of the thermal ALE process is summarized in Figure 3.1 while the overall ALE reaction is described by Reactions 3.1, 3.12, and 3.13.

The kinetic Monte Carlo (kMC) method is a stochastic computer algorithm that calculates numerical results of surface reactions through random sampling of the surface active sites to mimic the naturality of the real world. The practical usage of the kMC method serves to bridge the microscopic and macroscopic regimes. Kinetic Monte Carlo (kMC) simulations for atomic layer processes have been widely used [116, 25, 24, 68, 35, 103] for surface reaction kinetics. In this research, the developed kMC-based microscopic model by [116], validated with the experimental work from [61], is adopted to characterize the surface reactions on Al_2O_3 thin films. Table 5.1 shows the kinetic mechanism and the activation energies for ALE of Al_2O_3 and a detailed discussion of the kinetic mechanism is further examined in [116]. In this research, the coverage and etching fractions are used to express the progression of the first and second half-cycle reactions, respectively, of the ALE process are calculated by the kMC simulation, which assumes that the half-cycle reactions occur in self-limiting behavior such that the surface of the substrate, which is modeled as a 3D lattice. As an output of the kMC simulation, the etching per cycle (EPC) is computed by multiplying the etching fraction by the EPC of 0.46 Å/cycle, which has been reported by [61] whose experimental conditions were similar to the boundary conditions that would be defined in the macroscopic CFD model.

Kinetic parameters are an integral part of the input of the kMC simulation. A common approach to compute the reaction rate constant is through transition-state theory (TST), which is defined in Eq. (2.5). The reaction rate constant, k, is dependent on the pre-exponential factor, ν,

the activation energy, E_A, the universal gas constant, R, and the temperature, T, of the reaction. The activation energies are calculated by the nudged elastic band (NEB) method, which was previously performed [116]. The pre-exponential term depends on the quantum partition functions for the transition state ($Q^\ddagger$) and reactant (Q) species and is calculated from the following expression:

$$\nu = \frac{k_B T}{h} \frac{Q^\ddagger}{Q} \tag{5.1}$$

where k_B is the Boltzmann constant, T is the temperature of the reaction, and h is the Planck constant. For simplicity, the ratio of the transition state and reactant partition functions can be approximated to unity [48]. Desorption, surface reaction, and diffusion rates are computed by Eq. (2.5) and (5.1). However, for the adsorption process, collision theory (CT) is used, which is based on Maxwell-Boltzmann statistics.

$$k_{d,ads} = \frac{2 P A_{site} \sigma}{Z \sqrt{2 \pi m k_B T}} \tag{5.2}$$

where P is the partial pressure of the adsorption species, A_{site} is the area of a single surface active site, σ is the sticking coefficient, Z is the coordination number, m is the mass of the adsorption species, and k_B is the Boltzmann constant. Based on the aforementioned reaction theories, the total rate constant, as a fundamental element of the kMC algorithm, is calculated as follows:

$$k_{total} = \sum_{i=1}^{N} k_i \tag{5.3}$$

where k_i is the reaction rate constant of reaction i and N is the number of reaction pathways. The

kMC computation is performed as follows, which is repeated until the termination condition is achieved.

1. The sum of the rate constant, k_{total}, is computed from the list of all possible reaction paths across the lattice model.

2. An event is chosen to fulfill the following equation.

$$\sum_{i=1}^{m-1} k_i \leq \gamma_1 k_{total} \leq \sum_{i=1}^{m} k_i \tag{5.4}$$

where m is the reaction m and $\gamma_1 \in (0, 1]$ is the first random number for the event selection.

3. The secondary random number, $\gamma_2 \in (0, 1]$, is selected to evolve the system clock with a time interval defined as follows:

$$\Delta t = \frac{-\ln \gamma_2}{k_{total}} \tag{5.5}$$

5.2.2 Macroscopic CFD domain

The effects of transport phenomena within the spatial reactor are investigated through computational fluid dynamics (CFD) software to compute spatial and time-dependent variables within the reactor. Spatial atomic layer etching (SALE) reactors with different gap distances are developed to investigate the intermixing of the precursors and the effect of the velocity of the substrate on thin film quality based on a two-dimensional (2D) multiscale CFD simulation with a dynamic mesh. First, WorkBench 2021R2, one of the ANSYS packages, is used to construct the mesh with triangular elements for five reactors (denoted as R1 through R5) whose mesh quality is quantified

Table 5.1: Activation energies for rate-determining intermediate reactions for the thermal ALE of Al_2O_2 using HF and TMA precursors, which were calculated from Density Functional Theory via Quantum Espresso by [116].

Reaction pathway	Activation energy (eV)
$-O-O-AlF_x- + HF_{(g)} \rightarrow -O-OH-AlF_{x+1}-$	NA[†]
$-O-OH-AlF_{x+1}- \rightarrow -O-O-AlF_x- + HF_{(g)}$	2.02
$-O-OH-AlF_x- \rightarrow -OH-O-AlF_x-$	0.98
$-OH-O-AlF_y- \rightarrow -O-OH-AlF_y-$	0.71
$-OH-OH-AlF_y- \rightarrow -O-H_2O-AlF_y-$	1.28
$-O-H_2O-AlF_y- \rightarrow -OH-OH-AlF_y-$	0.76
$-H_2O-AlF_y- \rightarrow -AlF_y- + H_2O_{(g)}$	0.88
$-AlF_2- + Al(CH_3)_{3(g)} \rightarrow -AF-FAl(CH_3)_3-$	NA[†]
$-AlF-FAl(CH_3)_3- \rightarrow -AlF_2- + Al(CH_3)_{3(g)}$	42.27
$-AlF-FAl(CH_3)_3- \rightarrow -AlF(CH_3) + AlF(CH_3)_{2(g)}$	1.45
$-AlF(CH_3)- + Al(CH_3)_{3(g)} \rightarrow -Al(CH_3)_2- + AlF(CH_3)_2$	NA[†]
$-AlF_3- + Al(CH_3)_{3(g)} \rightarrow -AlF_2-FAl(CH_3)_3$	NA[†]
$-AlF_2-FAl(CH_3)_3- \rightarrow -AlF_3- + Al(CH_3)_{3(g)}$	21.29
$-AlF_2-FAl(CH_3)_3- \rightarrow -AlF_2(CH_3)_2- + AlF_2(CH_3)_{2(g)}$	0.82
$-AlF_2(CH_3)_2- \rightarrow AlF_2(CH_3)_{2(g)}$	1.12

[†] The rate constant is calculated by Collision Theory.

[‡] x can be 0 through 2 and y can be 1 through 3.

in Table 5.2. The mesh quality plays a significant role in the convergence, numerical solution accuracy, and stability of the CFD computation. As shown in Table 5.2, the mesh quality of all reactor designs built in this research is considered acceptable by the criteria defined by [4]. A more detailed discussion of the mesh quality parameters is given in [120].

In order to simulate the moving substrate, a dynamic mesh is employed, in which two dynamic mesh update methods are used: diffusion-based smoothing and remeshing. The diffusion-based

smoothing method adjusts the mesh through each time step as the substrate moves, which is defined by the following equation:

$$\nabla \cdot (\gamma \nabla \vec{u}) = 0 \tag{5.6}$$

where γ is the diffusion coefficient and $\nabla \vec{u}$ is the mesh displacement velocity. The diffusion coefficient is calculated as a function of the boundary distance.

$$\gamma = \frac{1}{d^\alpha} \tag{5.7}$$

where d is the normalized boundary distance and α is the diffusion parameter. The boundary-distance-based diffusion method allows the absorption of the mesh motion and preserves the mesh quality around the boundary of the motion. For α, a range of 0 through 2 is practical and 1.5 is chosen in this work. As another mesh update method, the remeshing method is adopted, which is modeled on local cells and allows the mesh to be updated locally with an acceptable skewness criterion so that the mesh quality is maintained through each time step. In the local cell remeshing mode, the maximum cell skewness is set to 0.7, which is recommended for 2D simulations by [4].

Each reactor is designed with a length of 160 *mm*, in which 9 injection or vacuum ports (denoted as an injection assembly) of 10-*mm* width are located at the top of the reactor. Each port has a 20-*mm* height and 20-*mm* pitch (that is, the distance between the centers of two adjacent ports). The injection assembly can be modularized and extended by adding extra injection assemblies as needed. For R5, the substrate is placed in the leftmost purge (N_2) zone as illustrated in Figure 5.3. For simplicity, in this work, a 5-*mm* long substrate is employed; however, the length of the substrate can be enlarged. The multiscale CFD simulation is performed using reactor configurations

104

R1 through R4 to investigate the effects of the gap distance, the N_2 flow rate, and the precursor (HF and TMA) flow rates on precursor separation. Then, the reactor, R5, is used to study the effects of a moving substrate and of the outflow vacuum pressure on the film quality.

In the CFD simulation, the mass, momentum, and energy conservation equations are solved on the basis of the coupled pressure-based solver in ANSYS Fluent, which are described by:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m \tag{5.8}$$

$$\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot \left(\overline{\overline{\tau}}\right) + \rho \vec{g} + \vec{F} \tag{5.9}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = -\nabla(\Sigma h_j J_j) + S_h \tag{5.10}$$

where ρ is the density of the mixture, $\vec{v}$ is the velocity of the mixture, S_m is the mass transfer source term, p is the static pressure, $\overline{\overline{\tau}}$ is a symmetric rank two stress tensor, $\rho \vec{g}$ is the gravitational body force, $\vec{F}$ is the external body force, E is the internal energy, h_j is the sensible enthalpy of the species j, J_j is the diffusion flux of species j, and S_h is the heat transfer source term. To solve these partial differential equations, the finite volume method is employed in the discretized time domain. To specify the consumption of HF and TMA and the generation of H_2O and DMAF from the surface kinetics, the source term (S_m) for each species is defined in a user-defined function (UDF) and the source terms are updated every time step, which are computed from the microscopic kMC simulation. The operating conditions, temperature and pressure, and boundary conditions, precursor and purge flow rates, are also defined in the UDF.

The combination of the microscopic kMC model and the macroscopic CFD model is made possible by defining convergence criteria and iteration conditions during the numerical simulation

for the macroscopic gas-phase domain. Default convergence thresholds and a time step size of 0.0005 *s* are defined to prevent the risk of divergent solutions in order to maintain the connection between the microscopic kMC model and the macroscopic CFD model. The number of time steps is determined by the microscopic kMC model, which updates the flow time as well as the mass source generation and consumption terms for each species in the half-cycle reactions. At time $t = 0$, the SALE reactor will be in steady-state conditions such that all gaseous species will have reached equilibrium conditions, which is determined by initially simulating the reactor with the steady-state solver. Once the steady-state conditions are obtained, the transient solver is used to propagate the movement of the substrate through the reaction chamber.
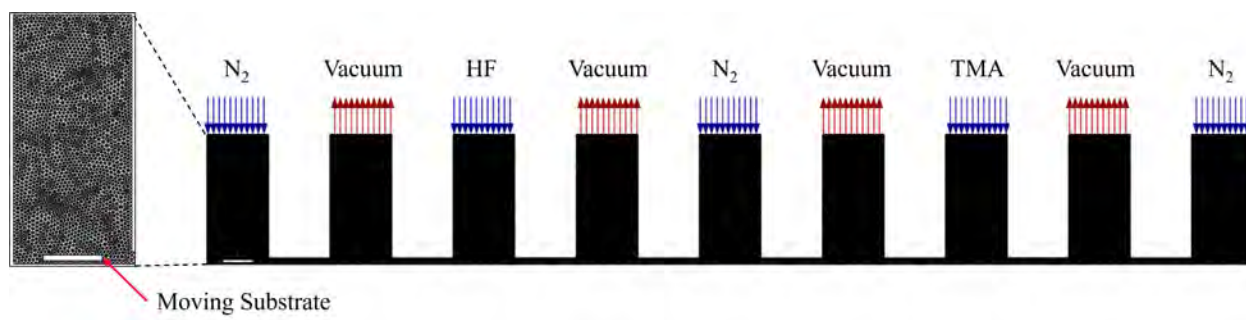
Figure 5.3: A 2D side view of the dynamic mesh for the spatial reactor design with 0.25 *mm* gap distance.

5.3 Simulation results and discussion

The effects of parameters (gap distance, purge gas flow, precursor flow, substrate velocity, and vacuum pressure) are studied for the optimization and operation of the sheet-to-sheet (S2S) spatial reactor design. Table 5.3 summarizes the list of case studies that are carried out by adjusting the aforementioned parameters. First of all, the steady-state simulation mode in Fluent is

Table 5.2: Minimum to maximum mesh quality parameter ranges for various sheet-to-sheet reactor configurations with the overall quality determined by criteria from [4].

No.	Gap (*mm*)	Skewness	Orthogonal quality	Aspect ratio	Overall quality*
†R1	10.00	$0 \sim 0.528$	$0.682 \sim 0.964$	$1 \sim 2.297$	Very good
†R2	1.00	$0 \sim 0.674$	$0.565 \sim 1.000$	$1 \sim 2.899$	Very good
†R3	0.50	$0 \sim 0.660$	$0.588 \sim 1.000$	$1 \sim 2.747$	Very good
†R4	0.25	$0 \sim 0.682$	$0.553 \sim 1.000$	$1 \sim 2.991$	Very good
‡R5	0.25	$0 \sim 0.525$	$0.671 \sim 1.000$	$1 \sim 2.316$	Very good

* Determined by [4]
† No substrate to investigate the effects of precursor intermixing.
‡ Optimized reactor design with a substrate.

used to run the computational fluid dynamics (CFD) simulation to analyze the effects of the gap distance, the N_2 flow rate, precursor (HF and TMA) flow rates by using reactor configurations R1 through R4 without a substrate in Table 5.2. The effects of those parameters on precursor separation are discussed in the following sections. Then, on the basis of the results of the aforementioned simulation, the optimal reactor parameters are selected and used to investigate the effects of the substrate velocity and vacuum pressure on the film quality, which are simulated using the transient simulation mode in Fluent. In this paper, the term, "cycle," is used even if the spatial atomic layer etching reactor is not operating in a cyclical operation. Thus, "a cycle" by definition occurs when a substrate passes through the 9 ports of the injection assembly. There are several assumptions that underlie the simulations: (1) there are manifolds in the upstream facility so the precursors and N_2 are well-mixed and injected through the injection ports, (2) the temperature of the substrate surface is maintained by a PID (proportional-integral-derivative) controller at 573 *K*, (3) the velocity of the substrate is constant in each simulation, (4) a vacuum pump, which is not constructed in the reac-

tor configuration, is implemented to exhaust gases through the vacuum ports, and (5) a physical

conveyor belt that moves the substrate is ignored, since the research focuses on the fluid regime

between a substrate surface and the bottom of an injection assembly.

Table 5.3: Various operating conditions for multiple sheet-to-sheet reactor configurations to examine their effects on precursor separation and thin film quality.

Case	Temperature	Pressure (Pa)		Gap distance	Flow rate ($sccm$)		
(No.)	(K)	Operating	Vacuum	(mm)	N_2	HF/N_2	TMA/N_2
1	573	300	-100	10.00	100	100/50	60/100
2	573	300	-100	1.00	100	100/50	60/100
3	573	300	-100	0.50	100	100/50	60/100
4	573	300	-100	0.25	100	100/50	60/100
5	573	300	-100	10.00	500	100/50	60/100
6	573	300	-100	10.00	1,000	100/50	60/100
7	573	300	-100	10.00	2,000	100/50	60/100
8	573	300	-100	10.00	100	200/50	120/100
9	573	300	-100	10.00	100	400/50	240/100
10	573	300	-100	10.00	100	800/50	480/100

5.3.1   Effect of the gap distance

The separation of precursors is highly dependent on the gap distance between the top of the

substrate surface and the bottom of the divider walls enclosing the injection assembly, as exemplified in Figure 5.1b. Thus, the separation of the precursors is primarily determined by a gap

distance, which serves to generate effective N_2 zones so that precursor species are not able to intermix. Unsuccessful precursor separation would result in undesirable etching and spontaneous

etching [86]; therefore, an effective gas barrier design is needed. In this research, reactor configurations R1 through R4 with 10.00 *mm*, 1.00 *mm*, 0.50 *mm*, and 0.25 *mm* gap distances (shown in Table 5.3) are used to investigate the effect of the gap distance on precursor separation. Simulations are carried out under the operating conditions defined as Case 1 through Case 4 in Table 5.3. As shown in Figure 5.4, precursor separation improves with decreasing gap distance, which is visualized by decreasing the shaded area under the mole fraction curves of HF and TMA and is consistent with the findings of the results of [22, 86]. An insignificant difference is observed between reactors with a gap distance of less than 1 *mm*. Figures 5.4a and 5.4b reveal that a smaller gap distance also marginally increases the mole fraction of the precursors in the half-reaction zones, resulting in a higher adsorption rate and an effective etching reaction.

The pressure difference is a driving force for gases to flow from a high pressure field to a low pressure field. For the reactor configuration, R1, of 10.00 *mm* gap distance, no pressure gradient is observed around the gap. Therefore, due to the lack of a strong driving force, most of the precursors are exhausted through the vacuum outlet when the precursors pass through the gap, but some of the precursors are able to cross over to the N_2 enriched zone and intermix with one another. However, in the reactor, R4, with a gap distance of 0.25 *mm*, reasonably large pressure gradients of 75 $\sim$ 113 *Pa* are observed between the gap and a vacuum outlet, resulting in the strong suction of the precursor through the vacuum outlet. Consequently, no trace of the precursor is allowed to pass through a vacuum outlet. Furthermore, the compression associated with a smaller gap distance increases the etch rate as the precursor adsorption is proportional to the precursor partial pressure according to collision theory in Eq. (5.2).
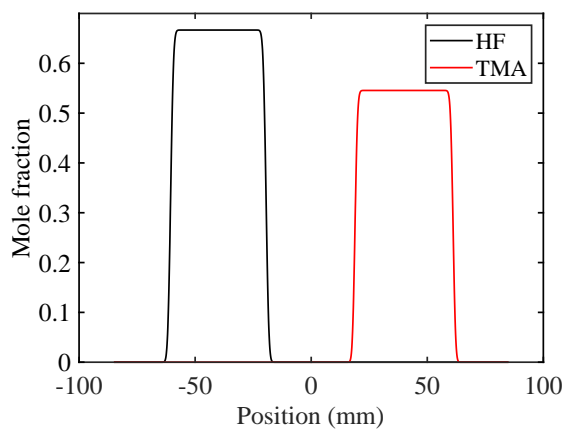
Figure 5.4: Line plots illustrating the effects of the gap distance on precursor separation, of which operating condition is (a) Case 1, (b) Case 2, (c) Case 3, and (d) Case 4 as described in Table 5.3. The shaded area indicates the precursor intermixing zone.

5.3.2 Effect of the purge gas flow

The separation of the two half-reaction zones is also controlled by the amount of purge gas flow (N_2); thus, enough purge gas is required to prevent the precursors from intermixing in the N_2-enriched zones. As discussed in Section 5.3.1, the SALE reactor with a gap distance less than 1.00 mm shows effective precursor separation. Thus, computational fluid dynamics (CFD)

simulations are performed with reactor, R1, which has a gap distance of 10.00 *mm* to show the noticeable differences with varying N_2 flow rates. The operating conditions of the simulations are Case 1 and Cases 5 through 7 in Table 5.3. The results of the simulations are presented in Figure 5.5. The precursor intermixing improves as the N_2 flow rate increases. Due to the higher velocity of N_2 flow, additional precursor is carried out through a vacuum outlet. However, one of the drawbacks of increasing the N_2 flow is that it dilutes the half-reaction zones, leading to lower mole fractions of the precursors in the half-reaction zones and thus, reducing the half-cycle reaction rates. Consequently, if enough purge gas flow rate is integrated with a small gap distance, the effective precursor separation is obtained.

5.3.3   Effect of the precursor flow

The precursor flow rate is a crucial parameter in the design and optimization of SALD or SALE reactors. The CFD simulations with 4 different precursor flow rates (Cases 1, 8, 9, and 10 shown in Table 5.3) are performed to study the effect of precursor flow rates on precursor separation. When the flow rates of the precursors increase, the mole fractions of the precursors increase in the half-reaction zones, which accelerate the HF adsorption and TMA ligand-exchange reactions on the surface, as presented in Figure 5.6. Also, Figure 5.6 illustrates that the precursor intermixing in the half-reaction zones is marginally affected by changes in the precursor flow rate. However, in the N_2 zone, precursor intermixing increases with increasing precursor flow rate. Compared to Section 5.3.2, increasing the N_2 flow rate is a more efficient method than decreasing the precursor (HF and TMA) flow rates to enhance precursor separation.

(a) N$_2$ flow rate of 100 $sccm$

(b) N$_2$ flow rate of 500 $sccm$

(c) N$_2$ flow rate of 1,000 $sccm$

(d) N$_2$ flow rate of 2,000 $sccm$
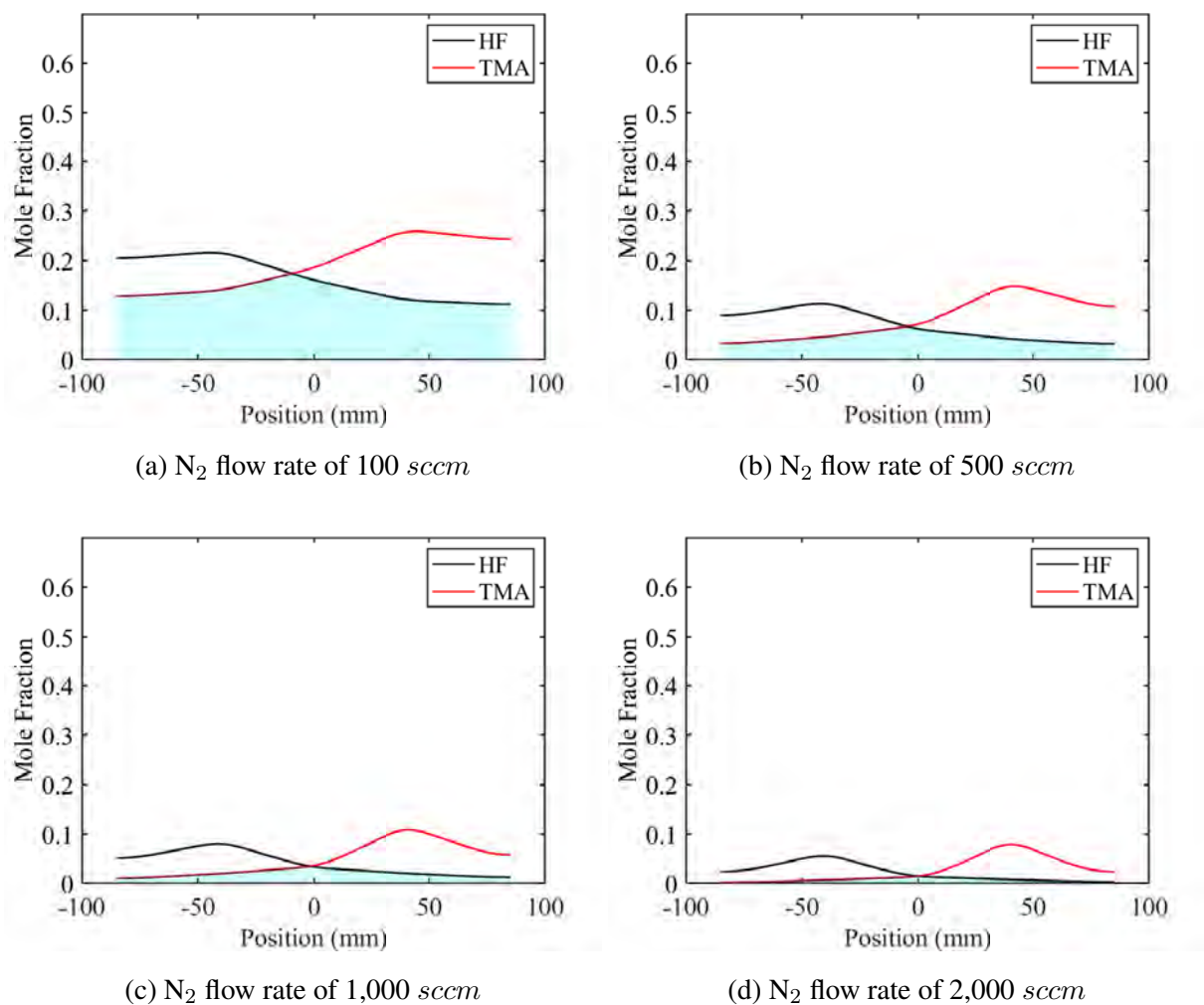
Figure 5.5: Line plots illustrating the effects of the N$_2$ flow rate on precursor separation, of which operating condition is (a) Case 1, (b) Case 5, (c) Case 6, and (d) Case 7 as described in Table 5.3. The shaded area indicates the precursor intermixing zone.

5.3.4 Effect of the substrate velocity

The effect of substrate velocity is explored using the R5 reactor configuration with a moving substrate, which is built on the basis of previous sections. As described in Section 5.3.1, a gap distance less than 1.00 mm ensures effective precursor separation; however, 0.25 mm is chosen to maximize precursor separation in the multiscale computational fluid dynamics (CFD) simulation.

(a) HF 100 *sccm* and TMA 60 *sccm*

(b) HF 200 *sccm* and TMA 120 *sccm*

(c) HF 400 *sccm* and TMA 240 *sccm*

(d) HF 800 *sccm* and TMA 480 *sccm*
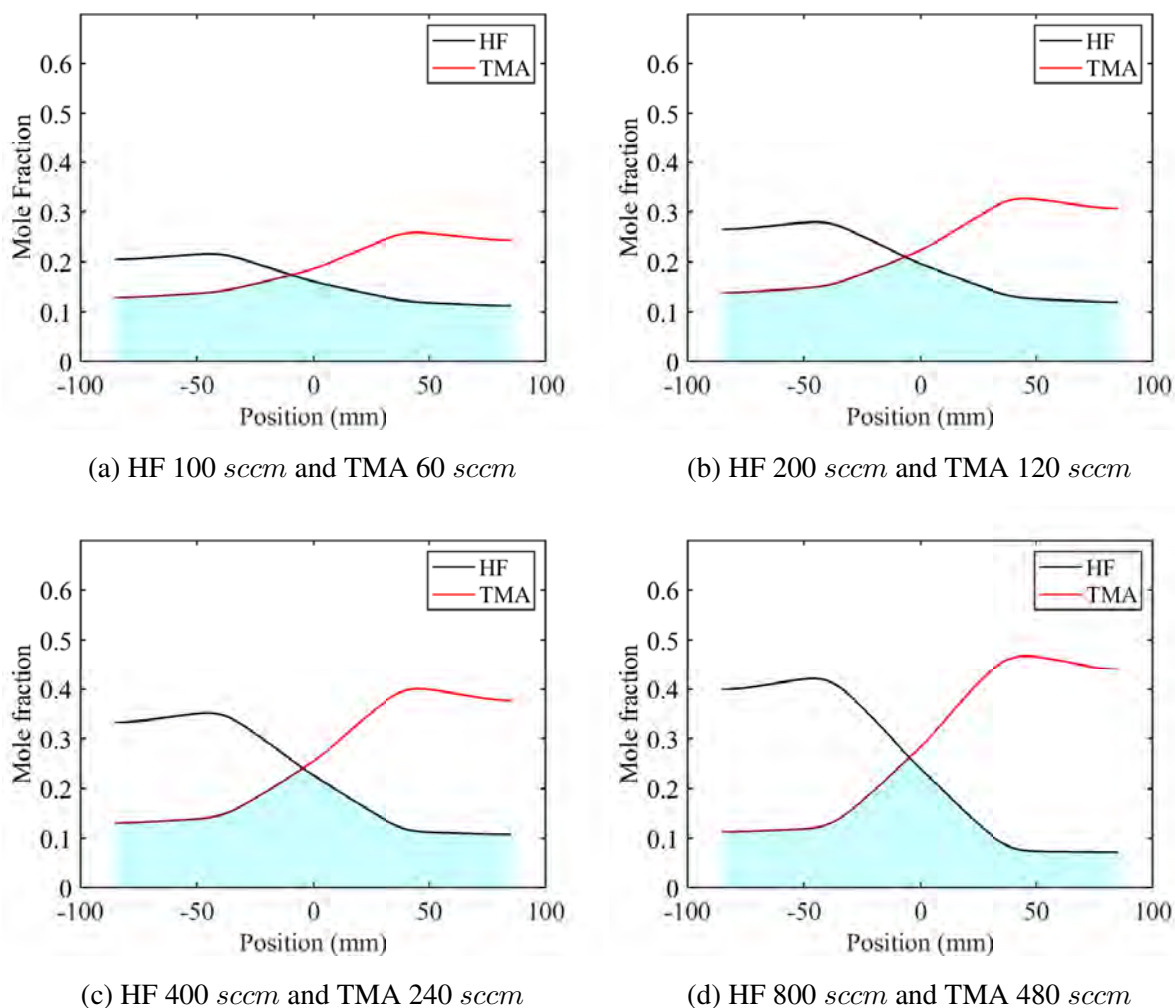
Figure 5.6: Line plots illustrating the effects of the precursor flow rate on precursor separation, of which operating condition is (a) Case 1, (b) Case 8, (c) Case 9, and (d) Case 10 as described in Table 5.3. The shaded area indicates the precursor intermixing zone.

As described in Case 4, a N_2 flow rate of 500 *sccm*, a HF/N_2 flow rate of 100/50 *sccm*, and a TMA/N_2 flow rate of 60/100 *sccm* are selected for the multiscale CFD simulation. Taking into account the thickness of the substrate, the height of the reactor chamber is 1.00 *mm*; however, a gap distance of 0.25 *mm* is preserved when the substrate moves under an injection assembly. Figures 5.7 and 5.8 illustrate the pressure contours of the species on a moving substrate. As pictured in Figure 5.7b and 5.8b, H_2O and DMAF are generated on the surface by the collaboration

of the kMC-based microscopic model and the CFD-based macroscopic model.
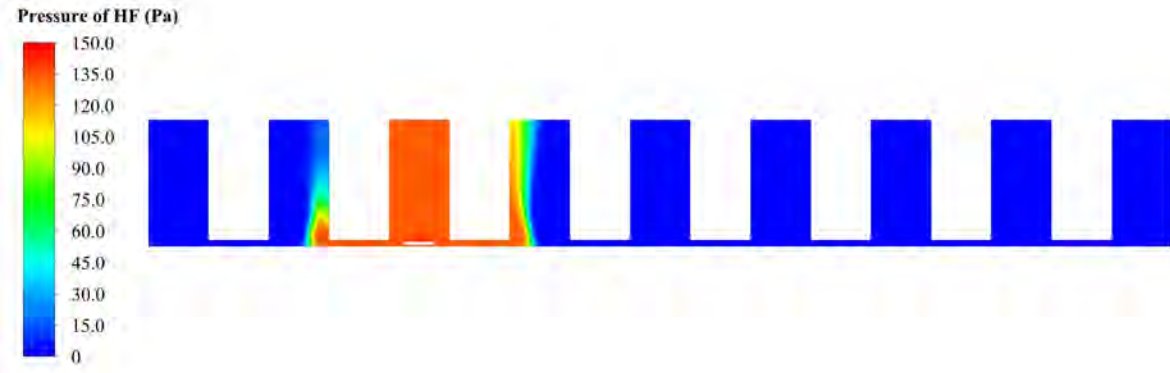
(a)
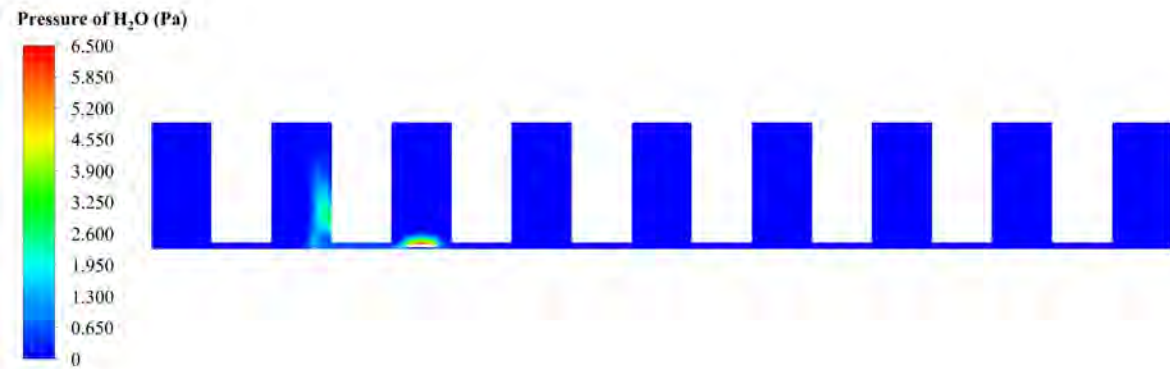
(b)

Figure 5.7: Pressure contours of HF (a) and H_2O (b) in the HF injection region for a substrate velocity of 80 *mm/s* and vacuum pressure of -100 *Pa* after 0.5 *s* of process time.

The substrate velocity is an important factor to control the film quality as the exposure time of the precursors to the substrate surface is directly related to the substrate velocity. Figure 5.9a shows the etching per cycle (EPC) versus the substrate velocity. Greater EPC is observed at lower substrate velocities, which is accordant with the experiment by [100] and the CFD results by [22, 83]. The reported etching rate of 0.46 Å/cycle is obtained at velocities less than 40 *mm/s*. The Al_2O_3 etching rate linearly reduces with increasing velocity for substrate velocities greater than 40 *mm/s*. This behavior can be used to determine the optimal velocity for etching Al_2O_3 films. Film

114
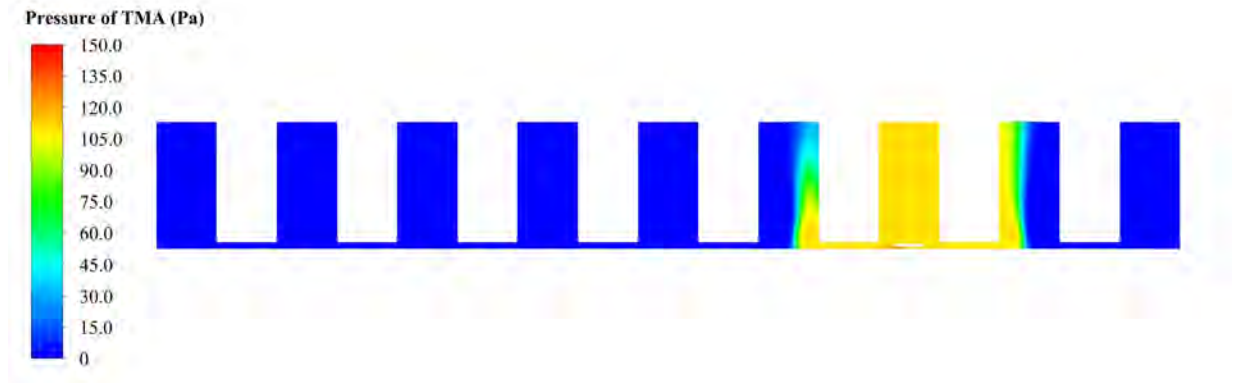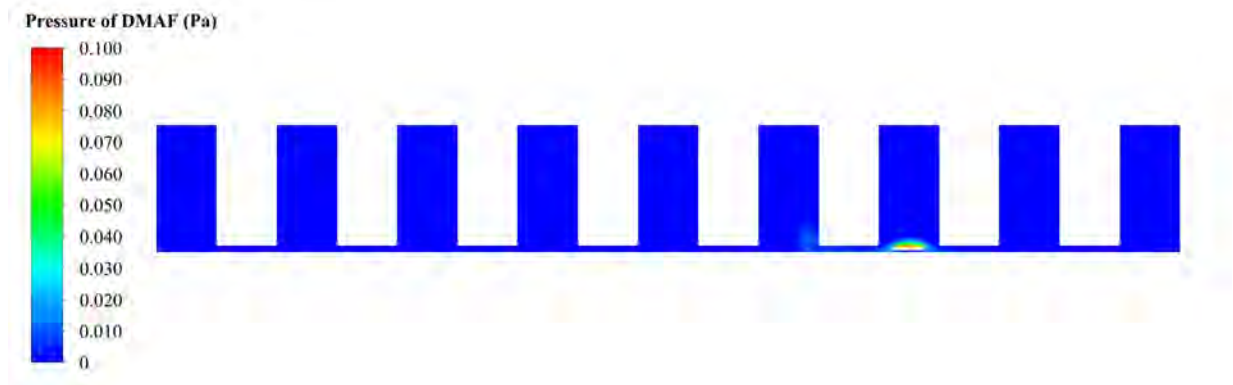
Figure 5.8: Pressure contours of TMA (a) and DMAF (b) in the TMA injection region for a substrate velocity of 80 *mm/s* and vacuum pressure of -100 *Pa* after 1.5 *s* of process time.

uniformity is highly required to produce conformal thin films in the semiconductor industry. Thus,

a multiscale CFD simulation is performed on each node of the mesh of the surface to explore the

film uniformity; therefore, 18 kinetic Monte Carlo (kMC) simulations are performed. As seen in

Figure 5.9b, the uniform etch rate on the surface is observed with velocities of 80, 120, and 160

mm/s. The film conformance is quantitatively indicated by the highly low standard deviations of

etched thickness for each cycle on 18 nodes of the surface mesh, which are 0.010, 0.007, and 0.002

Å/cycle for 80, 120, and 160 *mm/s* substrate velocities, respectively. In particular, the substrate has

the desired uniformity at 80 *mm/s* while a substrate velocity of 160 *mm/s* has marginal degradation

of the uniformity of the film. The results imply that the short exposure time attributed to higher velocities has some surface roughness that is able to disrupt conformal etching reactions on the surface.
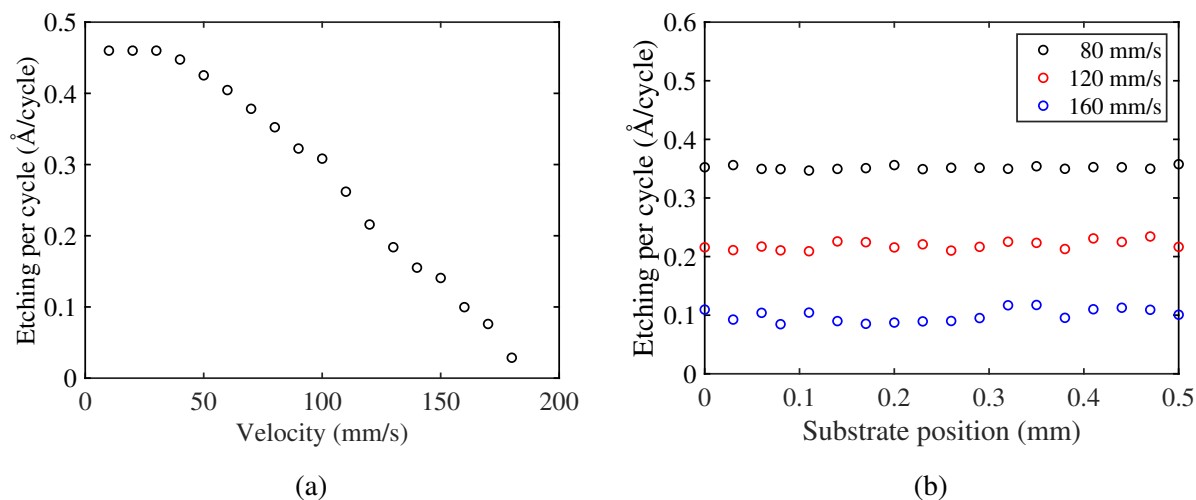
Figure 5.9: (a) The effect of the substrate velocity on the etching per cycle. (b) The effect of the moving substrate on etching per cycle to illustrate the uniformity of the etch.

5.3.5 Effect of vacuum pumping

The vacuum pressure is another factor that is considered in this discussion since the vacuum pressure directly affects the distribution of gaseous species within the reactor and is also responsible for preventing overaccumulation of feed into the reactor. The vacuum pressure is defined as an outflow pressure boundary condition in ANSYS Fluent, which is varied to determine its effects on the EPC of the substrate. The relationship between EPC and vacuum pressure for variable substrate velocities is further examined in Figure 5.10, which indicates that high vacuum pressure magnitudes lead to a decrease in EPC and the vacuum pressure effects on EPC for lower substrate velocities are insignificant. Several factors contribute to the aforementioned results, including in-

creased vacuum pressure, which decreases the partial pressure of precursors within the reaction injection zones and increasing substrate velocity results in decreased substrate exposure time to the precursors and ultimately decreased coverage and etching fractions on the substrate.
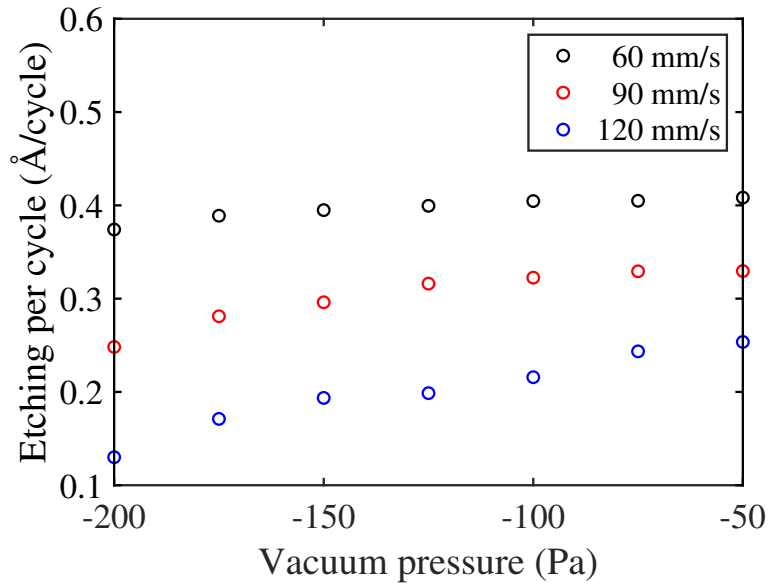
Figure 5.10: The vacuum pressure versus etching per cycle relationship for several substrate velocities.

5.4   Conclusion

An optimal spatial atomic layer etching (SALE) reactor configuration was developed for thermal atomic layer etching (ALE) of Al_2O_3, which was determined by analyzing the fundamental parameters that are directly related to the conformance of the quality of the substrate film. The quality of the substrate film is based on an etch rate and substrate film uniformity obtained from the reactor configuration. An multiscale computational fluid dynamics (CFD) model was used to fabricate the relationship between the fundamental characteristics of the SALE reactor design.

The multiscale CFD model was constructed by integrating a kMC-based microscopic model with a CFD-based macroscopic model. The multiscale CFD simulation revealed that the lower gap distances ensure effective separation of the precursors. Furthermore, an increase in N_2 flow rate and reduction of the HF/TMA flow rate also prevent the intermixing of the precursors. However, when a sufficiently small gap distance is determined, the effects of the flow rates of the precursors and N_2 are minimized. Therefore, a gap distance is the main factor in reducing N_2 and precursor flow rates while preserving effective separation of the precursors. To investigate the effects of the moving substrate on Al_2O_3 film quality, the dynamic mesh method was used to simulate the moving substrate in the optimized SALE reactor. Implementing a higher substrate velocity to the reactor configuration revealed a lower etching rate due to the reduced exposure time of the substrate. In addition, the larger vacuum pump resulted in a lower etching rate due to its strong precursor suction. As a result, the developed SALE reactor design and operating conditions were able to produce conformal ultrathin films.

Chapter 6

Atomistic-Mesoscopic Modeling of Area-Selective Thermal Atomic Layer Deposition

6.1 Introduction

There are several approaches to creating a protective layer that allow for selective deposition. The most popular approach is to use self-assembled monomers (SAMs) as inhibitors to produce effective barriers. SAMs, such as alkylsilanes, have large aliphatic chains on their tails that block or delay ALD nucleation on the NGA while the head groups of the SAMs are connected to the NGA. Thus, SAMs act as growth inhibitors and allow film deposition to occur only on the GA of the substrate surface. The applications of ASALD using various SAMs have been investigated in several works [14, 104, 46, 76, 42]. Another advantage is that SAMs are readily removed by acetone and demineralized water after the ALD process is completed. Despite the aforementioned

benefits of the SAMs, the ALD growth selectivity is limited to just a few nanometers of thickness due to potential defects in the SAMs. Furthermore, it is difficult and time-consuming to prepare defect-free SAMs [30]. In addition, most SAMs, which are based on wet chemistry, are difficult to integrate into vapor phase processes such as ALD. These issues make it challenging for SAM-based ASALD to yield high-aspect-ratio dielectric films.

To overcome the drawbacks of SAM-based ASALD, small-molecule inhibitors (SMIs) have been proposed. Unlike SAMs, which use a large inert functional group to physically block ALD growth, the reactive moieties of SMIs play a paramount role in preventing precursor adsorption onto the NGA [113]. ASALD based on SMIs effectively blocks precursors through a combination of chemical passivation and steric shielding [74]. The degradation of the blocking layer created by SMIs during the deposition processes is insignificant in contrast to that of the polymeric blocking layer of SAMs [73]. Since SMIs can be delivered through the vapor phase by adding a fixing step, the application of SMIs can be easily implemented to further improve selectivity. The protective layer formed via SMI-based ASALD allows surface destructive co-reactants such as oxygen plasma and ozone to be applicable as oxidants. An ABC-type area-selective ALD [71] has been developed based on this motivation and is also integrated into this work as discussed below.

The reaction mechanism for ASALD is composed of three sequential steps designated by Steps A, B, and C. In Step A, inhibitors selectively adsorb onto the NGA and form a blocking layer to prevent precursors from adsorbing onto the NGA during the entire ALD process. In Step B, precursors restrictively adsorb onto the GA due to the deactivation of the top layer of the NGA. Lastly, an oxidant oxidizes the top layer of the GA to complete a cycle of the deposition process. During the dosing of the precursor and oxidant, there is the possibility that some inhibitor

molecules might desorb from the NGA, which contributes to a loss in selectivity. Thus, the correction step (Step A) is repeatedly introduced into the entire reaction cycle to re-saturate the protective layer. The ABC-type ASALD is schematically illustrated in Figure 6.1.
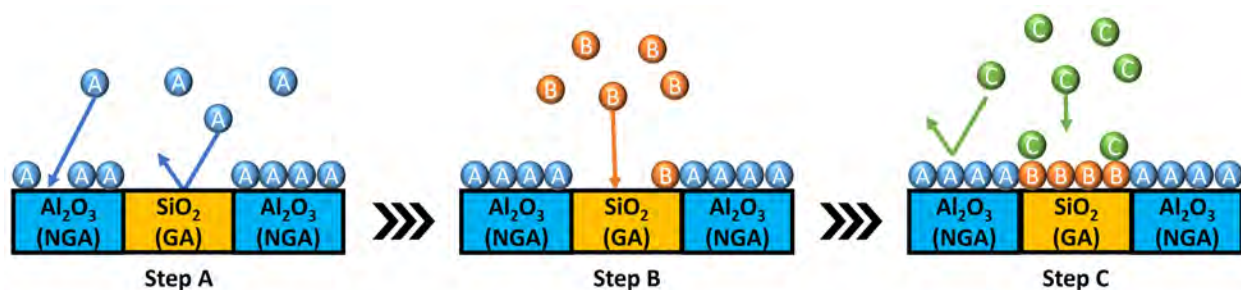
Figure 6.1: Schematic illustration of the ABC-type SMI-based ASALD. The molecules A, B, and C correspond to the inhibitor, precursor, and oxidant, respectively.

Despite the merits mentioned above, there is little computational or experimental research on the mesoscopic behavior of SMI-based ASALD. In particular, prior research [71, 73] demonstrated the technical viability of ASALD using SMIs, which was mainly performed from an experimental perspective. However, these previous works have not studied the surface kinetics for the entire cycle at the atomistic level in great detail. Thus, it is necessary to study how inhibitors selectively prohibit film growth on the NGA substrate with SMI-based ASALD in great detail. In this work, a kinetic Monte Carlo (kMC) study is performed to characterize ASALD of SiO_2 (GA)/Al_2O_3 (NGA) using small-molecule inhibitors (SMIs). Past studies have successfully employed the kMC method to simulate atomic layer etching processes [120, 116, 118, 117, 119] to characterize the stochastic behavior of thin-film surface kinetics, which will similarly be integrated into this work. Acetylacetone (Hacac), bis(diethylamino)silane (BDEAS), and ozone (O_3) are selected as a SMI, precursor, and oxidant, respectively. Density functional theory (DFT) calculations are carried out to conduct an investigation into the atomistic behavior of all reaction steps. In addition, to visualize

121

the steric effects of Hacac, a model that simulates the surface coverage for Hacac is developed. The structure of this work is as follows: Section 6.2 describes the methodology of mesoscopic modeling, Section 6.2.1 illustrates the ASALD process in details, Section 6.2.2 describes the calculations to define the atomistic behavior via DFT, Section 6.2.3 discusses the steric effects, Section 6.3 describes the results of the kMC study and its validation, and Section 6.4 provides a summary of this work.

6.2   Atomistic-mesoscopic modeling

6.2.1   Surface kinetics

The purpose of the ASALD process is to achieve a selective thin-film metal-oxide deposition that occurs solely on the growth area (GA) of β-SiO$_2$ (1 0 1) by integrating a protective surface of the non-growth area (NGA) of α-Al$_2$O$_3$ (0 0 0 1). The ASALD process consists of three sequential reactions (Steps A, B, and C) that are situated between purging steps, which are conducted in a cyclical manner. This work will also assume that all reactions are of bimolecular type, which only occur in elementary steps; thus, the reactions will proceed sequentially. Step A is composed of an inhibition-adsorption reaction in which the inhibitor selectively binds to Al$_2$O$_3$. Step B is a precursor modification reaction that uses bis(diethylamino)silane (BDEAS) to produce a modified surface layer on SiO$_2$, which possesses a self-limiting nature. Step C is an oxidation process that introduces ozone (O$_3$) onto the modified surface layer, thereby depositing a monolayer. These three sequential reactions are components of the so-called ABC-type reaction cycle, which is elucidated in this section.

Inhibitors are designed to reduce the occurrence of a reaction, which makes inhibitor selection an issue of great importance for ASALD processes. Ideally, the inhibitor for an ASALD process must be chosen such that the inhibitor selectively reacts with the NGA; thus, this implies that there needs to be a substantial difference in the magnitude of the activation energy barriers between the NGA and GA, which is characterized by chemoselective behavior. Inhibitors must also have a low decay rate so that the protective layer can sustain cycles of Steps B and C without requiring numerous doses of inhibition treatment between the deposition steps (Steps B and C). Thus, this work employs a small molecular inhibitor (SMI) with acetylacetone (Hacac) to fabricate an inhibition layer on the NGA of Al_2O_3, which exploits the conjugated structure of the tautomerized enol isomer to deprive the active oxygen sites on the NGA of subsequent adsorption reactions (Steps B and C). Hacac can exist in a keto isomer that can undergo a tautomerization process to produce a stable enol isomer composed of conjugated π bonds, which is the driving force that allows Hacac to bind to an Al atom on the NGA.

The keto and enol structural isomers that are formed through resonance allow the two resonance structures to coexist in basic and acidic forms, respectively. However, the stability of the enol form is more preferential and stable than the keto structure in the gas phase [32]. This work first characterizes the surface of Al_2O_3 as being uniformly composed of hydroxyl-terminated ligands (Al-OH) in a vicinal diol structure. This hydroxylation of the Al_2O_3 surface promotes basicity and interacts with the acidic Hacac, giving rise to a thermodynamically feasible neutralization

reaction [71]. The reaction between the hydroxyl-terminated Al and Hacac molecule produces

H_2O vapor as a byproduct, which is defined in the following reaction:

A1   Monodentate (A3) $+$ H_2O (Step A-1)

Monodentate (A3)   Chelate (A4)   (Step A-2)

In Step A-1, an enol Hacac molecule withdraws its acidic hydrogen proton from the hydroxyl

group, which is donated to the hydroxyl-terminated Al atom to produce a H_2O leaving group. The

deprotonated enol Hacac molecule adsorbs onto the Al site to form a monodentate structure. Then,

in Step A-2, the deprotonated Hacac adsorbate forms a protective six-member cyclic ring, also

called its chelate form, through a chelation process by adsorbing onto the surface Al atom with

the nonadsorbed keto group. The delocalized electrons in the conjugated π bonds of the chelate

ring is illustrated with a dotted line [36]. Also, note that the wavy bonds from the Al atoms do

not necessarily imply that the terminal is a single bond as shown in the reaction mechanisms;

instead, they are used to represent the ligand structure that is unimportant to the mechanism and

to illustrate that the reactions are occurring on the surface of the substrate. Although Hacac is

intended to selectively adsorb only to Al-OH active sites, there is also potential for Hacac to adsorb onto Si-OH active sites that may be generated during the hydroxylation step.
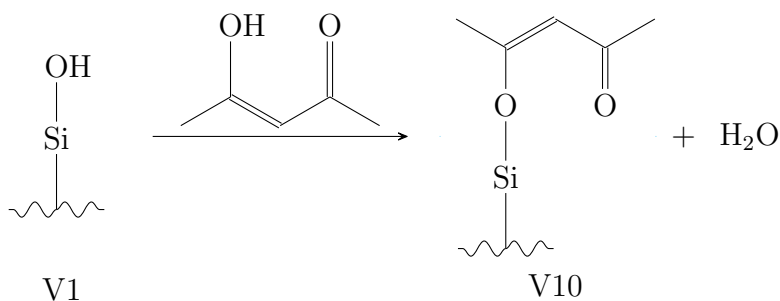
This undesirable reaction is the main source for nonuniform deposition in the SiO_2 growth region and can impede substrate quality and conformity by deflecting precursor adsorbates in Steps B and C of the ASALD process.

Precursors are intended to undergo surface reactions on the GA while having inherent properties that prevent further permeation beyond the substrate surface. Typically, bulky precursors are integrated into this modification step to ensure that a transport-limited boundary is established. The precursor, bis(diethylamino)silane (BDEAS), is used as an adsorbate on SiO_2 to produce Si lattice sites. The adsorption of BDEAS on the GA exemplifies self-limiting behavior such that single atomic monolayers of Si are deposited and byproducts are easily volatized. This Step B reaction mechanism is simplified by the following overall reaction:

(Step B-1)

H_2Si–N(ethyl)(ethyl) ... O–Si / OH–Si → V4 (H–Si–H with two O–Si) + HN(ethyl)(ethyl) (Step B-2)

V3   V4   DEA
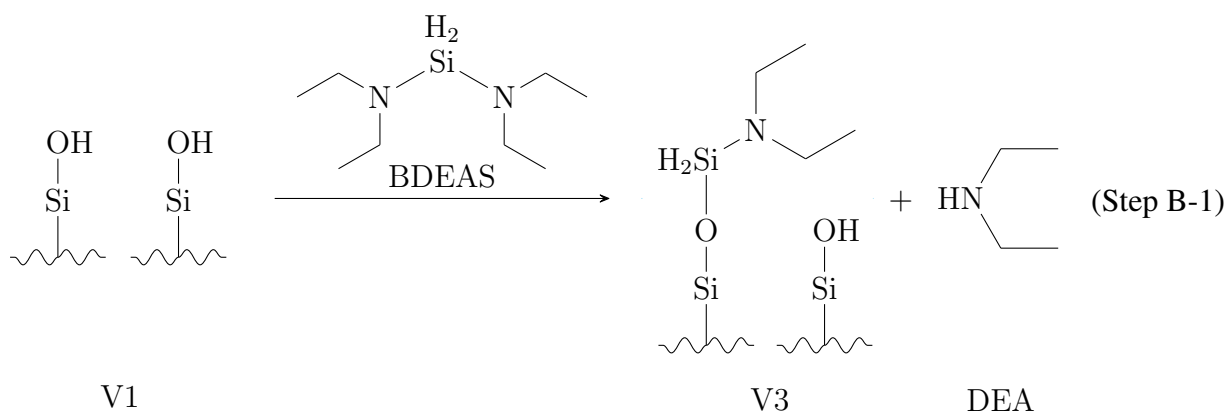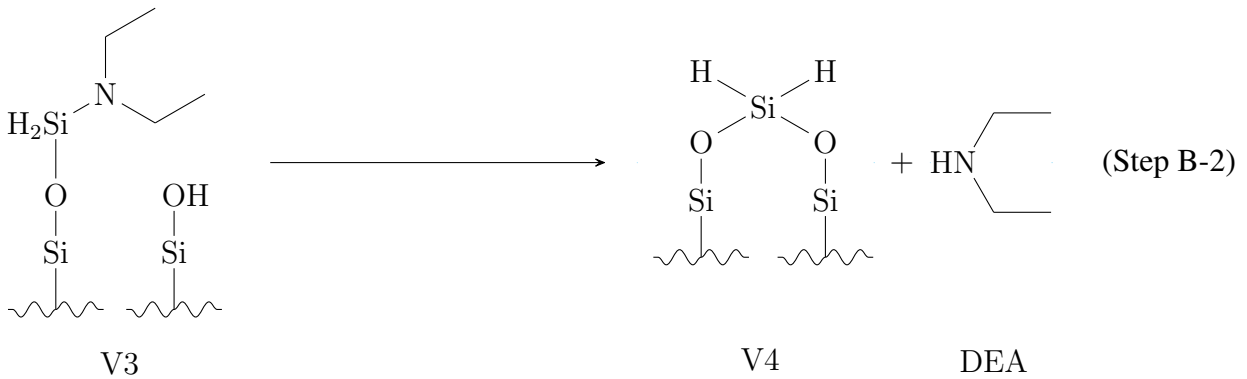
In Step B-1, BDEAS first adsorbs to an active hydroxyl surface site while simultaneously with-drawing a molecule of diethylamine (DEA). It is assumed that the hydroxylated SiO_2 surface consists of a homogeneous composition of vicinal hydroxyl ligands. Following the initial adsorption of BDEAS in Step B-2, the adsorbate binds to an adjacent hydroxyl surface site while simultaneously removing an additional DEA molecule, which results in a deposited monolayer of silicon atoms.

Following Step B, an oxidant, ozone, is introduced into the ASALD process to convert the hydride-terminated ligands of the newly deposited Si surface to hydroxyl-terminated ligands through an oxidation process. Ozone, which exhibits both electronegative and electropositive characteristics, performs a simultaneous and spontaneous ligand-exchange mechanism that substitutes a hydroxyl group for the hydride-terminated group while producing a leaving group of O_2, oxygen gas. This oxidation step is incredibly spontaneous to the point where the reaction can occur simultaneously for both hydride-terminated groups on the Si atom. The result completes the final monolayer of SiO_2 deposition and concludes the cycle of the ASALD process. The overall reaction

for Step C is described as follows:

(Step C-1)

V4 → V6

(Step C-2)

V6 → V8
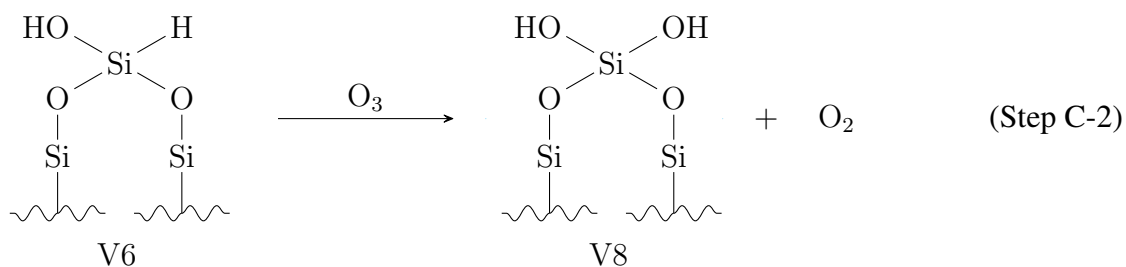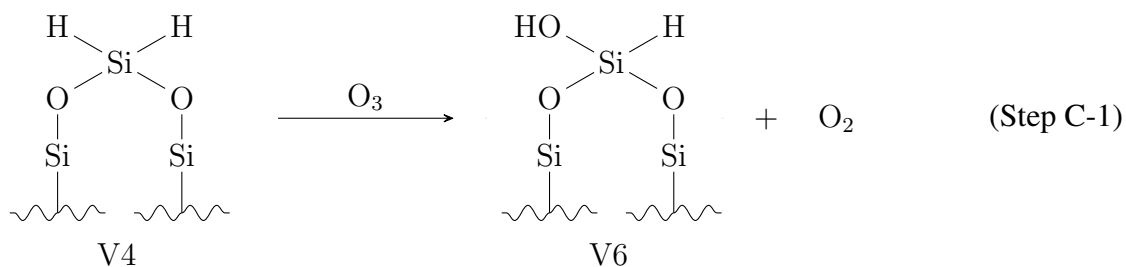
The resulting surface layer becomes composed of a homogeneous array of geminal hydroxylated ligands that can be recrystallized through an annealing cycle after the ASALD process [47] to reform the initial vicinal diol surface structure. It is, however, worth mentioning that after each step (Steps A, B, and C) in the ASALD process, a purging cycle is introduced with an inert gas such as nitrogen gas, N_2, to ensure that the reagent species (Hacac, BDEAS, and O_3) do not intermix and disrupt the self-limiting behavior, which is essential to ensuring high surface uniformity.

6.2.2   DFT calculation

Electronic structure calculation framework

With the reactions defined in Section 6.2.1, kinetic rate constants for each reaction must be defined for these elementary reaction steps. Mesoscopic simulation requires the knowledge of important reaction parameters to compute such reaction rate constants including the activation

energies, the pre-exponential factors, vibrational frequencies, and sticking coefficients. These parameters may be difficult to measure experimentally and generally require various computation methods that combine *ab initio* molecular dynamics, first principles quantum mechanics, and statistical chemistry. The open-source software for materials modeling, Quantum ESPRESSO (QE), is used in this work to calculate such kinetic parameters, which are needed to define the mesoscopic model. QE uses theoretical concepts including density functional theory (DFT), which computes the energy of phononic interactions using pseudopotential data that define each element by wavelengths to facilitate the computations [39, 8]. Previous works have employed QE for computing electronic structure optimization as well as for performing computations of kinetic parameters [120, 116, 88, 117]. This work will integrate DFT computations via QE to compute an optimal (minimal electronic energy) electronic structure for each species produced in each step of the ASALD process.

The initial stage of the electronic structure optimization process consists of modeling the atomic structures using an open-source, graphical user interface (GUI) program that supports QE called BURAI. The surface crystal structures of α-Al_2O_3 and β-SiO_2 are constructed by defining a Bravais lattice of (0 0 0 1) [71] with a trigonal crystal and vicinal (1 0 1) [95] with a triclinic crystal structure, respectively. Also, lattice model unit cells for α-Al_2O_3 and β-SiO_2 are constrained within a $2 \times 2 \times 2$ supercell. Precursor and byproduct species including BDEAS, DEA, H_2O, O_3, and O_2 are developed using a free crystal lattice that is self-determined using a unit cell size that contains the entirety of the molecular structure. This method prevents divergence by defining boundary conditions for the "particle-in-a-box" model. Projector Augmented-Wave (PAW) pseudopotential data are used in conjunction with the Plane-Wave Self-Consistent Field (PWscf)

package. In the program, the "relaxed" calculation method is utilized to optimize electronic forces that allow the translation of atoms in space to generate an electronic structure of minimum electronic energy. The use of the aforementioned pseudopotential data facilitates the solving of the position and time-dependent Schrödinger equation, which is defined by the following expression:

$$\hat{H}\psi(\overrightarrow{x},t) = E\psi(\overrightarrow{x},t) \ \text{where} \ \hat{H} = \frac{\hbar}{2m_e}\nabla^2 + \hat{V}(\overrightarrow{x},t) \tag{6.1}$$

where $\hat{H}$ is the Hamiltonian operator, which is a function of the sum of the kinetic and potential energies that are defined by the reduced Planck constant, $\hbar$, the particle mass, m_e, the Laplacian operator, ∇, and the potential energy, $\hat{V}$, respectively. Also, E refers to the total electronic energy of the system, and $\psi(\overrightarrow{x},t)$ is the wave function that depends on the displacement vector, $\overrightarrow{x}$, and time, t. A complete summary of the variables and their definitions used throughout this work is provided in Table 6.1. Self-determined parameters including the *degauss*, the kinetic energy cutoff for wave functions (*ecutwfc*), the kinetic energy cutoff for charge density and potential (*ecutrho*), and the k-points are determined through the SCF (self-consistent field) computation in QE and summarized in Table 6.2. The optimized electronic structures exported from QE and compiled in BURAI are visualized in Figure 6.2.

Computation of kinetic parameters

Various first principles theoretical concepts derived from quantum mechanics and chemistry are integrated into this work to compute reaction rate parameters that include the reaction rate constant, activation energy, and pre-exponential factor. The aforementioned parameters are fun-

damental for the mesoscopic modeling via the kinetic Monte Carlo method, which is discussed in Section 6.2.4. This section examines the methodological approaches to computing these kinetic parameters for adsorption and nonadsorption reactions that are described in Section 6.2.1.

Adsorption reactions occur during the initial bombardment of the inhibitor (Step A), precursor (Step B), and oxidant (Step C) on the surfaces of the substrate. These adsorption reactions can be modeled as bimolecular reactions through Maxwell-Boltzmann statistics and Collision Theory (CT). The probability of a successful adsorption reaction is modeled based on a sticking coefficient factor, σ, which is deduced from experimental works. The following equation provides the definition of CT:

$$k_{ads} = \frac{PA_{site}\sigma}{Z\sqrt{2\pi m k_B T}} \tag{6.2}$$

where k_{ads} is the reaction rate constant for the adsorption reaction, P is the pressure of the gaseous reagent (Hacac, BDEAS, or O_3), A_{site} is the surface area of a single active site, Z is the coordination number of the gas, m is the atomic mass of the gas, k_B is the Boltzmann constant, and T is the absolute temperature of the ambient environment. This work defines a sticking coefficient, σ, for Hacac as 1.0×10^{-4} [37], BDEAS as 2.0×10^{-5} [97], and O_3 as 4.5×10^{-5} [58]. It is noteworthy that the sticking coefficient of Hacac was determined from surface adsorption using precursor species for ALD and was assumed to be similar in magnitude to that of Hacac adsorption onto Al_2O_3.

A majority of reactions discussed in this work involve chemisorption, desorption, and other surface reactions that cannot be computed via CT. Nonadsorption reaction rate constants can be

calculated using the Arrhenius equation, which is defined by the following equation:

$$k_{nonad} = \nu \exp\left(-\frac{E_A}{RT}\right)$$ (6.3)

where k_{nonad} is the reaction rate constant for the nonadsorption reaction, ν is the pre-exponential factor, E_A is the activation energy, R is the universal gas constant, and T is the absolute temperature of the reaction. One challenge arises with the computation of the pre-exponential factor term, which is modeled on the frequency of collisions where the reactants overcome the activation energy barrier. Thus, Transition State Theory (TST) is often employed to reduce the computational requirements to collect such frequency data and is described by the following expression that computes the pre-exponential factor:

$$\nu = \frac{k_B T}{h} \frac{Q^{\ddagger}}{Q}$$ (6.4)

and $Q^{\ddagger}$ and Q are the products of the electronic, rotational, translational, and vibrational partition functions for the transition state species and reactant, respectively. To reduce the computational requirements for evaluating the partition functions, the ratios between the transition state and reactant partition functions are often simplified to unity [48]. Thus, the pre-exponential factor is largely dependent on the temperature of the reaction. Lastly, the activation energy of the reactions should also be established in Eq. (6.3), which are computed using the nudged elastic band (NEB) method in this work. QE contains a NEB package that is designated for carrying out such computations that essentially construct a minimum energy path between reactants and products. The NEB method requires optimized structures, which are obtained from the PWscf package described in

Section 6.2.2. With the optimized structure data, the NEB computation requests a self-determined number of electronic steps that computes an energy for each step to ultimately determine the activation energy barrier. This work defines seven electronic steps to reduce the computational demand while establishing a robust procedure for computing an accurate activation energy barrier to establish the reaction rate constants. Figure 6.2 illustrates the results produced from the application of *ab initio* and first principles modeling concepts as well as simulation results generated from QE. The activation energies essential for the mesoscopic model are shown in Figure 6.2. It is notable that the energy on the *y*-axis does not denote an absolute energy, but rather the energy difference between the starting structure and the ending structure in each step. Geometries for the reaction paths on Al_2O_3 and on SiO_2 are numbered as **A1** through **A4** and **V1** through **V10**, respectively, and they are summarized in Table 6.3.

6.2.3   Steric hindrance

Fundamentals of steric effects

Steric effects play a substantial role in the frequency of adsorption reactions discussed above by blocking other molecules from approaching the surface, thus hindering specific surface reactions. The regioselective adsorption of the SMI, Hacac, and bulky precursor, BDEAS, depends on how much space is available for these reactions to occur. In other words, Hacac and BDEAS cannot be densely packed on the surface due to their large molecular size, and the bulky adsorbates that have already bonded to the surface can prevent surrounding molecules from interacting with the substrate surface because of steric effects. Thus, the repulsive effect that results from the steric hindrance of the bulky reagents used in ASALD must be investigated and accounted for to

132

(a) Hacac cycle on Al_2O_3

(b) Hacac cycle on SiO_2

(c) BDEAS cycle on SiO_2

(d) O_3 cycle on SiO_2
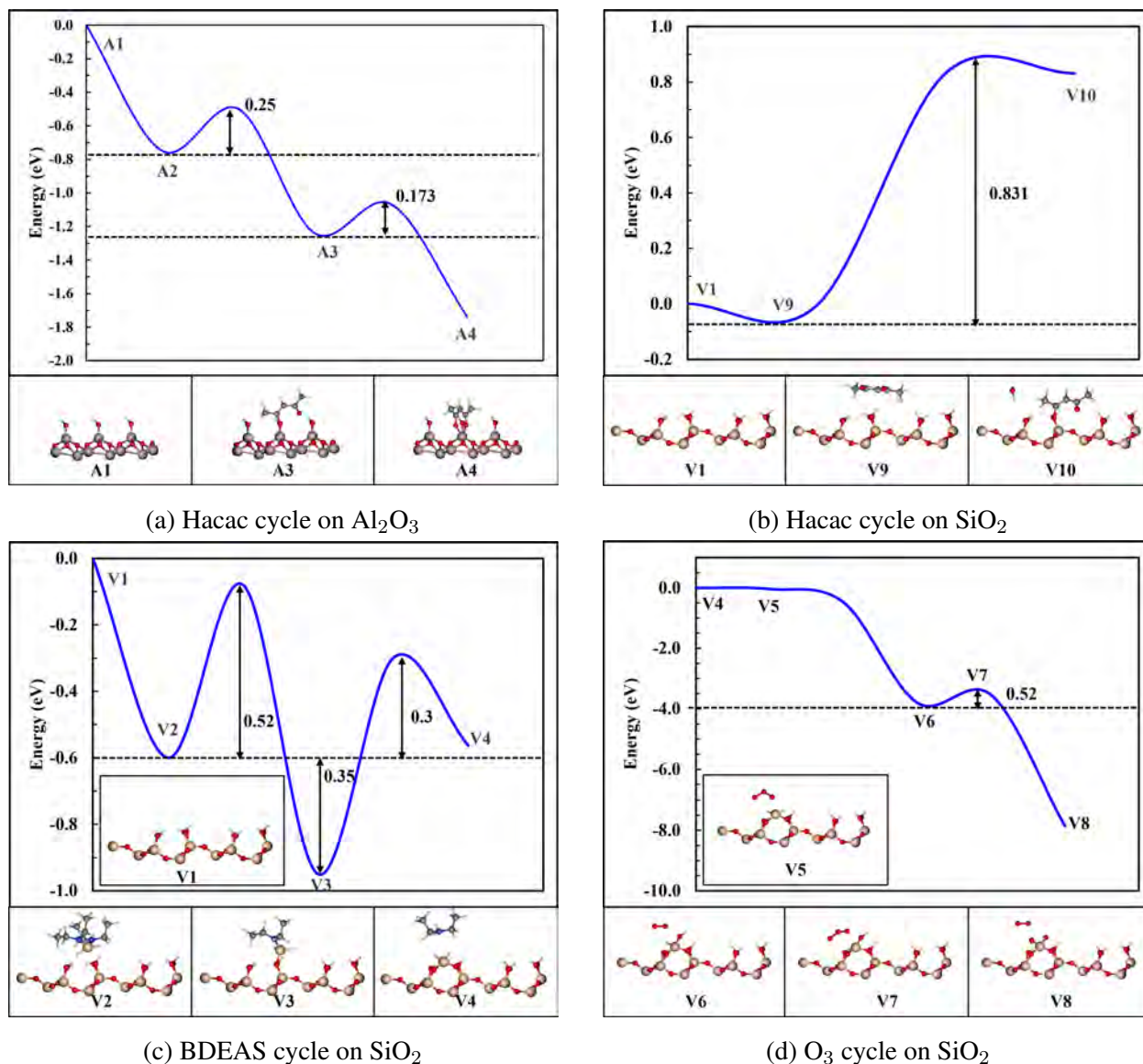
Figure 6.2: Minimum energy paths from the DFT calculations for the Hacac adsorption on (a) Al_2O_3 and (b) SiO_2, for the BDEAS adsorption on (c) SiO_2, and for the O_3 adsorption on (d) SiO_2. Color code for atoms: aluminum, dark gray; silicon, light brown; oxygen, red; hydrogen, white; nitrogen, dark blue; carbon, gray. The paths for the Hacac adsorption from **A1** to **A3** was sourced from [71]. Specifically, **A2**, **V2**, **V5** and **V7**, and **V9** indicate the physisorption reactions for Hacac on the NGA, BDEAS on the GA, ozone on the GA, and Hacac on the GA, respectively.

accurately depict the kinetic behavior of the surface reactions. The computational methods used to

simulate the steric effects in conjunction with the kinetic Monte Carlo model are elucidated in this

133

section.

There are several studies that have reproduced the repulsive effects of steric hindrance in surface adsorption reactions through computational approaches [112, 44, 49, 54]. For example, [44] considered the molecular interactions of Si-S, Si-Se, and Si-Te bonds between the substrate and bulky organic functional ligands in adsorption reactions, which were simulated in a two-dimensional (2D) stochastic model. In this model, the adsorbates were represented by disks defined by their van der Waals radii, which are defined as the minimum approach distance between two atoms that are represented as solid spheres. The van der Waals radii of some organic functional groups are presented in [94], which are applied in the simulation for Step A and summarized in Table 6.4. This section will discuss the steric hindrance involved in the Hacac adsorption on Al_2O_3 in Step A and the BDEAS adsorption on SiO_2 in Step B (due to the small van der Waals radius of ozone, steric hindrance effects are not expected in Step C). In regard with the BDEAS adsorption, a detailed review of the open literature indicates that it is difficult to find either experimental or simulation studies that discuss the steric hindrance effects taking place in the ALD process with BDEAS as precursor. Prior work presented important results from macroscopic measurements and observations [71, 95], but there is no mesoscopic explanation for the phenomena from a steric hindrance perspective. Thus, this surface modeling work accounting for steric effects provides a potential approach to reveal how steric hindrance influences the ALD nucleation of BDEAS.

Hacac adsorption on Al_2O_3

Hacac is a bulky molecule compared to the active reaction sites on α-Al_2O_3, which are defined as OH ligands. Thus, the deprotonated Hacac adsorbates on the Al_2O_3 surface can block

other Hacac molecules from adsorbing onto neighboring reaction sites. There are two molecular configurations for deprotonated Hacac that has adsorbed onto Al_2O_3: monodentate (**A3**) and chelate (**A4**), as shown in Figure 6.2a and described in Table 6.3. The reaction (Step A-2) from the monodentate to the chelate configuration is exothermic and has a low activation energy barrier as illustrated in Figure 6.2a, which implies that the monodentate ligand can easily overcome the activation energy barrier of 0.173 eV and transform into the chelate molecule through a chelation process. As a result, the monodentate structure spontaneously converts to the chelate structure from a kinetics perspective, which is in agreement with [71]. However, [73] reported that part of deprotonated Hacac exists as monodentates on the Al_2O_3 surface through IR inspection.

To investigate the Hacac adsorption process in greater detail, this work constructs a 2D model with a 100×100 surface grid, which has been proposed by [44]. In the grid, all active sites, denoted by small black dots in Figure 6.7, are randomly searched to determine their possibility for Hacac adsorption. In the simulation, the chelate adsorption is performed first and then followed by the monodentate adsorption due to the difference in magnitude between the sizes of the two molecules. A horizontal chelate adsorbate exists on the surface as revealed in Figure 6.3. The chelate structure has two methyl, CH_3, functional groups with a van der Waals radius of 2.0 Å and a spacing of 4.8 Å between the two functional groups, which is denoted as an empty capsule with two circles in Figure 6.7. Conversely, a monodentate adsorbate is placed diagonally on the surface as shown in Figure 6.3. To simulate the adsorption of deprotonated Hacac monodentates onto the 2D grid, it is assumed that the CH_3 functional group does not affect other adsorption reactions at adjacent reaction sites due to its high position. Instead, the free CO functional group is solely responsible for the steric repulsions of other molecules. Therefore, in the 2D model,

135

monodentate structures have CH_3 and CO functional groups with van der Waals radii of 2.0 and 1.7 Å, respectively, and a spacing of 3.8 Å; they are depicted as two conjoined circles in blue in Figure 6.7. The simulation procedure is as follows:

1. The chelate adsorbate is placed on the site throughout the grid if there is no steric hindrance. The orientation of the adsorbate is stochastically selected from 360° with intervals of 10°.

2. Next, the monodentate adsorption selection is performed, filling the empty sites with adsorbates in viable orientations in the same manner with the chelate adsorption.

Even after Step 1 (chelate adsorption) is performed, vacant active sites still exist due to the steric hindrance of the chelate configurations. Thus, Hacac molecules can nonetheless still adsorb onto the surface in the monodentate configuration. To reduce large deviations attributable to the stochastic nature of the algorithm, 100 simulations are carried out, and the surface densities are averaged. In addition, the standard deviation is calculated from the set of computations to evaluate the developed model for steric effects.

A prior study developed a 2D surface model to investigate the effects of steric hindrance on Hacac adsorption on Al_2O_3 [73]. The surface model simulated chelate and physisorbed monodentate configurations; however, the model lacked chemisorbed monodentate products on the active reaction sites. To advance the model proposed by [73] and to obtain the final surface structure, both chemisorbed chelate and monodentate molecules are considered in this research. This improvement enables the model to obtain more precise and realistic surface information, which provides a deeper explanation of the hinderance effect. The approach of this paper, which is adopted from [44] and advanced from [73], can be extended to other surface kinetics in which the surface can

be approximated into a 2D lattice model simply with the calculations of van der Waals radii of adsorbed molecules. The simulation results from the surface model of the Step A (Hacac) cycle are discussed in Section 6.3.1. After the simulation, the molecular density on the surface is calculated by Eq. (6.5).

$$\rho = \frac{n}{D^2 \cdot s \cdot x} \tag{6.5}$$

where ρ is the molecular density in $molecules/nm^2$, n is the number of molecules on the 100×100 grid, D is the dimension of the grid, which is 100 in this simulation, s is the distance between two adjacent sites, and x is the distance between two sites in the horizontal direction, which is calculated by the following equation:

$$x = \frac{\sqrt{3}}{2} s \tag{6.6}$$
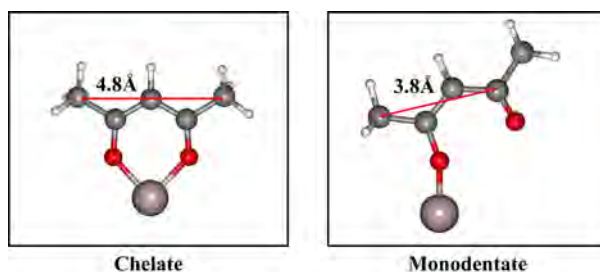
Figure 6.3: The intervals between two functional groups of the chelate and the monodentate are 4.8 and 3.8 Å, respectively. For the monodentate configuration, the CO group is considered as the group that is potentially able to hinder neighboring adsorption reactions. It is assumed that the CH_3 group at the far right of the monodentate product does not contribute to steric effects on adjacent reaction sites. Color code for atoms: aluminum, dark gray; oxygen, red; hydrogen, white; carbon, gray.

BDEAS adsorption on SiO$_2$

A BDEAS adsorbate is capable of preventing other BDEAS molecules from adsorbing on the GA, SiO$_2$. Figure 6.5 illustrates the top view of the GA surface, which is expressed by Geometry **V1** in Table 6.3, and it consists of 8 active reaction sites, which are shown as OH functional groups. If a BDEAS molecule adsorbs onto one of these sites (Geometry **V2**), the two functional groups of diethylamine (DEA) may protect the two neighboring reaction sites from the physisorption of other BDEAS molecules as shown in Figures 6.4a and 6.4b. When the physisorbed BDEAS reacts with the OH group by releasing H$_2$O, the BDEAS molecule turns into a DEA-deprived BDEAS (Geometry **V3**) on the GA as illustrated in Figure 6.4c. However, the remaining DEA group may block an adjacent site so that additional BDEAS physisorptions on the blocked site are hindered. When forming SiH$_2$ (Geometry **V4** in Table 6.3) by discharging the DEA group from Geometry **V3**, the adsorbate must bind to two adjacent sites on the surface. There are 5 potential reaction sites from which SiH$_2$ can be formed on the GA. In Figure 6.5, the distances between Sites 1 and the 4 adjacent reaction sites, denoted as Sites 2 through 5, are calculated from the optimized structure computed through electronic structure calculations. The distances between Sites 1 and 2, Sites 1 and 3, Sites 1 and 4, Sites 1 and 5, and Sites 1 and 6 are 5.03, 4.01, 4.02, 6.00, and 5.92 Å, respectively. It is reasonable to assume that Site 1 preferentially binds with Sites 3 and 4 to deposit SiH$_2$ films (Geometry **V4**) due to these sites being the closest, leading to those binding reactions to have the lowest activation energy. Therefore, the reactions between Sites 1 and 2, Sites 1 and 5, and Sites 1 and 6 can be reasonably ignored in the kinetic Monte Carlo (kMC) simulation. According to the periodical and symmetrical nature of the surface grid, Site 3 can only react with Sites 1

138

or 2. In other words, the reactions are constrained in the "bicolumn" system from a modeling perspective.

(a)

(b)

(c)
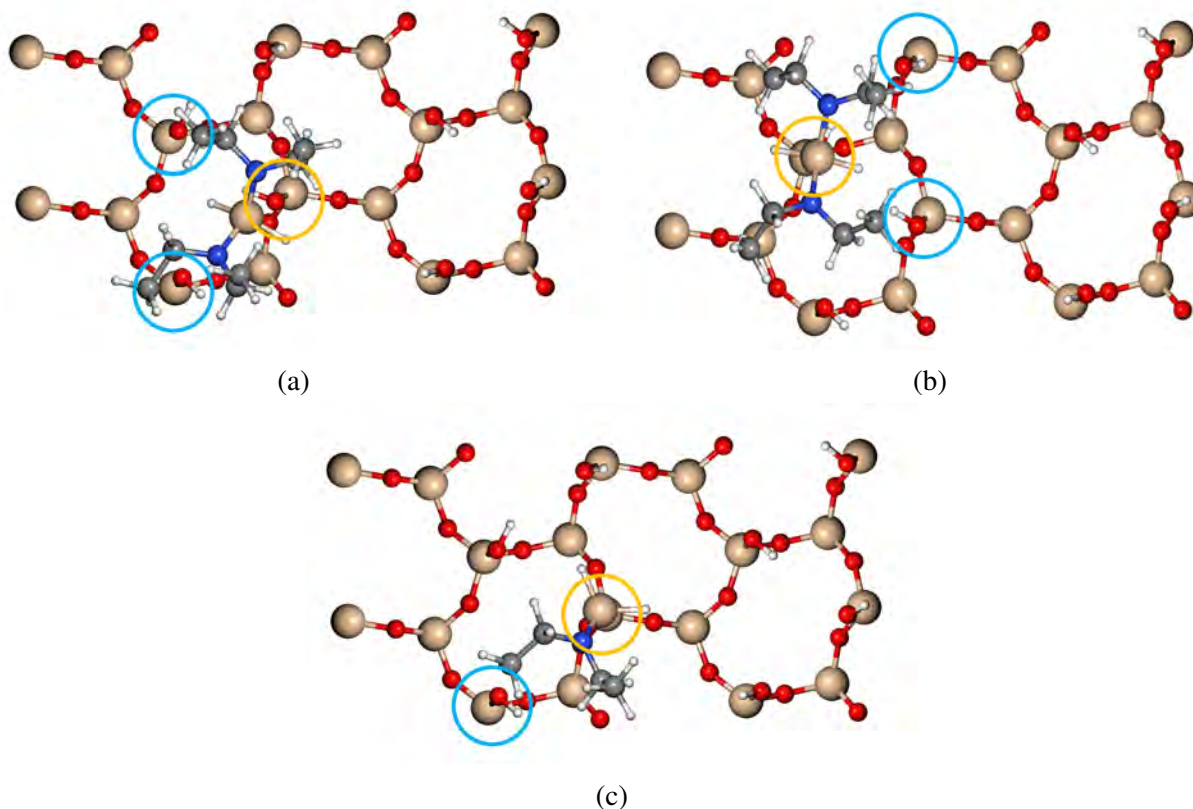
Figure 6.4: Top view of (a) Case 1 for Geometry **V2** (b) Case 2 for Geometry **V2** (c) Geometry **V3**. The blue circled reaction sites are hindered and deactivated by the current reaction site in yellow. Color code for atoms: silicon, light brown; oxygen, red; hydrogen, white; nitrogen, dark blue; carbon, gray.

In addition to the steric effects, another issue to consider is that some reaction sites can be deactivated through isolation effects when two adjoining sites are already occupied. As a result, it is nearly impossible to reach full coverage for Step B (BDEAS cycle on the GA). In the experimental data of the BDEAS adsorption from [95], coverages of 94% and 86% were observed both within and outside the studied ALD window, respectively. In this mesoscopic model, the deactivated sites are ignored when calculating the surface coverage in the kMC simulation so that the
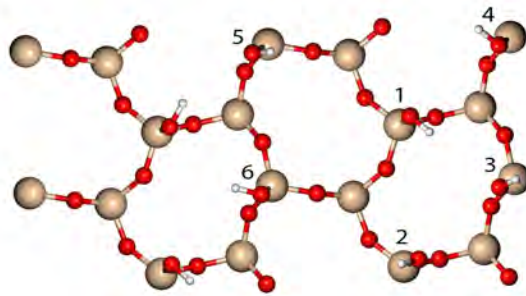
Figure 6.5: Top view of a OH terminated SiO_2 surface, expressed by Geometry **V1** in Table 6.3. Sites 2 through 5 are numbered with Site 1 as the center. Color code for atoms: silicon, light brown; oxygen, red; hydrogen, white.

simulation stops appropriately. With this modification, the computation model can reach 100% coverage by disabling the deactivated sites. On the basis of these steric and kinetic constraints, the kinetic Monte Carlo (kMC) model is built to simulate the surface kinetics, including the blocking effect of adsorbed molecules and the interactions between reaction sites. The detailed description of the kMC model is discussed in Section 6.2.4 in depth.

6.2.4 Kinetic Monte Carlo simulation

The conformity of the substrate is dictated by various quality specifications including the uniformity and the amount of deposition on the growth area (GA). However, the latter presents a challenge for modeling in the mesoscopic surface domain due to the disorganization of the reaction kinetics. While it is theoretically possible to simulate the complete reaction mechanism on the surface of a 200 or 300 mm substrate, such a task is extremely computationally demanding to the point of infeasibility. As a popular approach to simulate surface kinetics at the mesoscopic level, statistical Monte Carlo (MC) methods use random sampling in a constrained domain to accurately estimate the effects of a much larger stochastic system. Such atomistic-mesoscopic simulation

using MC methods is applicable to nanoscale material systems and minimizes the variability of timescale prediction as described by [72]. These algorithms are the most reliable when applied to complex simulations that maintain their probability distribution. Additionally, the kinetics of multiple reaction systems will evolve as the system progresses; thus, there must also be a temporal element to the algorithm. The reaction systems occur at predetermined reaction sites on the surface of the wafer substrate, which indicates that the kinetic Monte Carlo (kMC) simulation method is an appropriate algorithm to integrate into this work [48]. There are various kinetic Monte Carlo (kMC) algorithms that have been investigated; however, this work applies the variable step size method (VSSM), also called the Gillespie algorithm, which was advanced by Bortz, Kalos, and Lebowitz [10]. This algorithm is ideal because it is more computationally efficient when there are numerous reaction pathways [103], which is exactly the case for the ASALD reactions. The VSSM method converts the surface structure of the substrate into a matrix of identifiable numbers, which is illustrated in Figure 6.6. The basic formulation of the VSSM algorithm can be separated into 5 steps as described in [116, 103]: 1) identification of possible reactions, 2) rate constant summation, 3) reaction selection, 4) time step evolution, and 5) process continuation.
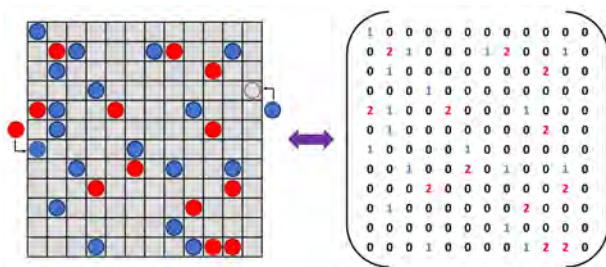
Figure 6.6: Conversion of the mesoscopic surface kinetics into a matrix of identifiable numbers. The blue and red circles in the lattice, corresponding to 1 and 2 in the matrix, indicate different reaction statuses.

The first step of the VSSM algorithm is to identify all possible reactions for each active site,

which is identified by conditions in the simulation. Next, all the possible reaction rate constants are summed, which establishes the bounded conditions for the kMC simulation and is calculated from the following equation:

$$k_{total} = \sum_{i-1}^{N} k_i \tag{6.7}$$

where k_{total} is the sum of all reaction rate constants, k_i is the reaction rate constant for the potential reaction i, which was evaluated in Section 6.2.2, and N is the total number of possible reaction pathways. Secondly, the algorithm employs a randomly generated number, $\gamma_1 \in (0, 1]$, to stochastically select a reaction, j, that fulfills the following criterion:

$$\sum_{i=1}^{j-1} k_i \leq \gamma_1 k_{total} \leq \sum_{i=1}^{j} k_i \tag{6.8}$$

where j is the index of the reaction in N that satisfies the equation. As a reaction path is determined depending on the product of γ_1 and k_{total}, it is observed that the larger the individual rate constant, the more likely it is to be chosen. Of note, the random number is computed using the Python "*random*()" function from the *random* package. Then, the following algorithm is used to determine a time interval in which no reactions occur [48].

$$\Delta t = \frac{-\ln \gamma_2}{k_{total}} \tag{6.9}$$

where Δt is the aforementioned time interval and $\gamma_2 \in (0, 1]$ is another randomly generated number between 0 and 1 that is independent of γ_1. Finally, the simulation is conditionalized with an ending criterion to determine the continuation or termination of the simulation. If the former is chosen,

the algorithm is repeated starting from the first step of the VSSM algorithm as the time progresses from $t_{old} \rightarrow t_{new} + \Delta t$. In addition, k_{total} is recalculated with each iteration of the algorithm since the possible reaction pathways across the substrate surface change as the system evolves. For the simulation of ASALD, the kMC simulation is used to estimate the time necessary for the surface of the wafer substrate to reach complete, 100%, coverage; thus, the terminal condition for the simulation is defined for the latter. The simulation is conducted using Python scripting language in serial processing and various packages dedicated to the randomization of the simulation and for carrying out numerical computations.

For generality, the procedural steps of the atomistic *ab initio* quantum mechanics and mesoscopic simulations are summarized as follows:

1. Elementary reaction pathways are defined and limited to rate-determining reactions from the electronic structure calculations of the surface structures.

2. *Ab initio* quantum mechanics simulations are simulated via Quantum ESPRESSO for individual species generated for surface reactions.

3. Reaction rate constants are calculated using CT and the Arrhenius Equation for adsorption and nonadsorption reactions, respectively.

4. The kMC method considering hinderance effects is exploited following the characterization of the kinetics parameters.

6.3 Simulation results and discussion

In this work, surface modeling results for the simulation of Hacac adsorption on the NGA, described in Section 6.2.3, are discussed. In addition to the surface modeling, the temperature and pressure dependence of the area-selective atomic layer deposition (ALD) of SiO_2/Al_2O_3 are investigated in a pressure range of 10 to 500 Pa with a temperature range of 423 to 573 K. As mentioned above, SiO_2 is defined as the growth area (GA); on the other hand, Al_2O_3 is considered to be the non-growth area (NGA) where the film deposition is not required. It is also noted that the ALD process has a self-limiting behavior in that only a single layer of SiO_2 is deposited on the top layer of the surface. Thus, the kinetic Monte Carlo (kMC) simulations considering steric effects are performed until each step in the ABC cycle reaches full coverage. It is noted that the kinetic interaction between SiO_2 and Al_2O_3 on the boundary is ignored in this study as described in Section 6.2.4. Thus, two separate lattices for SiO_2 and Al_2O_3 of 100×100 are simulated for surface kinetics. A number of simulations with a lattice size of 100 through 900 were performed to explore size-dependence of the kMC simulation. It was observed that there was no significant effect of the lattice size on the simulation results while requiring higher computing power with increasing lattice size, which is also supported by previous studies [45, 116]. To minimize the effect of stochastic sampling, 100 kMC computations are performed and averaged to obtain a data point for the process time under the operating conditions. To provide a benchmark, the computation time for the entire cycle is about 30 min for 100 trials using a constant temperature of 523 K and pressure of 300 Pa, while simulated using similar computational resources.

6.3.1 Surface modeling for protective layer

Figure 6.7 visualizes the surface of Al_2O_3 where chelate and monodentate products are randomly distributed at different angles. Due to the small molecular size, chelate molecules occupy the space between monodentate molecules. The surface density of each chelate and monodentate adsorbate is $1.45 \pm 0.02 \ molecules/nm^2$, and $0.35 \pm 0.03 \ molecules/nm^2$, respectively. $1.80 \pm 0.01 \ molecules/nm^2$ is thus calculated to be the combined surface density from the simulation. The fraction of monodentate configurations on the surface is $19.37 \pm 1.40\%$. The reported experimental value of the monodentate fraction on the surface by IR inspection is $20 \pm 5\%$ [73], and thus, the simulation result conforms with the experimental value. However, the molecular density simulated from the model developed in this work is lower than the density from [73] for the chelate configuration, which is $1.7 \pm 0.1 \ molecules/nm^2$. The deviation of the molecular density is attributed to differences in the atomic distances calculated in Section 6.2.2 using Quantum ESPRESSO and the atomic distances evaluated by [73] who reported a lower atomic distance.

6.3.2 Selectivity of ASALD

Figure 6.8 illustrates the selectivity of Hacac adsorption on Al_2O_3 (NGA) as opposed to SiO_2 (GA) in Step A (Hacac cycle) at $T = 423 \ K$ and $P = 400 \ Pa$. No Hacac adsorption is observed on the NGA, which is also detected throughout the entire operating window, which is in agreement with the experimental results [71, 73]. The selective Hacac adsorption forms a protective layer on the NGA, leading to SiO_2 being deposited on the GA in a selective manner. The selectivity of ASALD can be influenced by two factors: chemical passivation and steric hindrance. In contrast

145
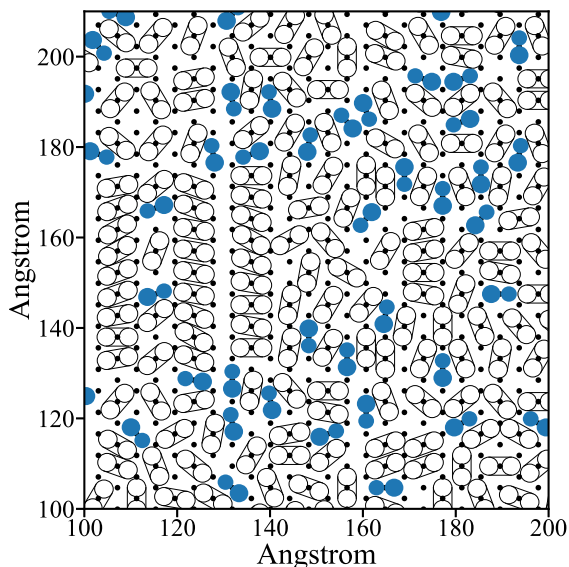
Figure 6.7: The adsorption pattern of Hacac chelate (represented by unfilled geometry) and monodentate (represented by filled geometry) configurations on Al_2O_3 sites (blue dots) from the 2D stochastic simulation model. The empty capsule with two circles in black and the two conjoined blue circles symbolize monodentate and chelate molecules, respectively.

to the GA, the NGA is chemically deactivated through the adsorption of Hacac as a result of the difference in the magnitude of the activation energies of the adsorption. This chemoselectivity is supported by the DFT calculations as discussed in Section 6.2.2. As shown in Figures 6.2a and 6.2b, the Hacac adsorption on SiO_2 is endothermic and has a high activation energy barrier of 0.831 *eV*. Meanwhile, the Hacac adsorption on Al_2O_3 is exothermic and has a low activation energy of 0.25 *eV*. Therefore, the adsorption predominantly occurs on Al_2O_3, resulting in the localized formation of a protective layer on the NGA. In addition to the chemoselective deposition behavior, the regioselective behavior is accounted for by the inclusion of steric repulsion effects. As discussed in Sections 6.2.3 and 6.3.1, complete Hacac coverage is not possible on the NGA since some reaction sites are sterically hindered by neighboring deprotonated Hacac adsorbates. As

a result, the blocked reaction sites are physically deactivated. In conclusion, the NGA is deactivated

both chemically and sterically, leading to high effective selectivity for ASALD.
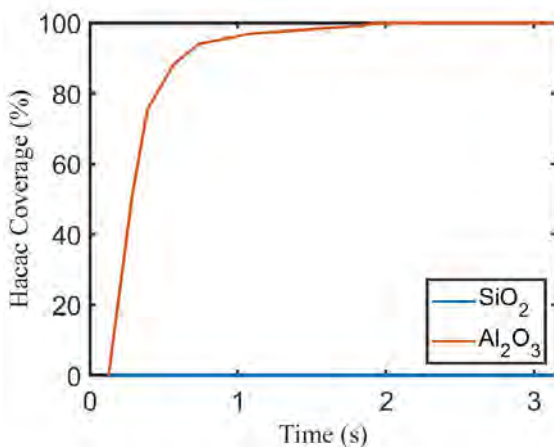
Figure 6.8: Hacac coverage versus time at $T = 423\ K$ and $P = 400\ Pa$. The red and blue solid lines denote the Hacac adsorption on Al_2O_3 and SiO_2, respectively.

6.3.3 Impact of operating conditions

The simulation results in the studied operating window ($423\ K \leq T \leq 573\ K$ and $10\ Pa \leq P \leq 500\ Pa$) from the mesoscopic model based on the kMC simulation with steric hindrance are shown in Figure 6.9. 800 data points are collected with a temperature and pressure interval of $10\ K$ and $10\ Pa$, respectively. Depending on the operating temperature and pressure, the overall process time significantly varies. Specifically, a Hacac dose of more than $3\ s$ is calculated for saturation at a pressure of below $100\ Pa$. Hacac pulse times from the kMC simulations are comparable to the experiment results, where the Hacac pulse of $5\ s$ per cycle was reported to be used to reach saturation [71]. In the BDEAS cycle, a BDEAS dosage time of $2.3\ s$ at $P = 400Pa$ is computed to reach full coverage, which is consistent with prior experimental research from [73]. In addition, [95] reported that dosing BDEAS for around $2\ s$ was required for saturation on SiO_2

in a typical ALD operating window.

Figure 6.10 shows the process time dependence on the operating temperature and pressure on 2D graphs by presenting isobaric lines produced from Figure 6.9. As expected, a higher operating pressure causes a shorter process time for all steps due to the higher pressure accelerating the physisorption reactions, which can be explained by the kMC algorithm. Surface reactions and desorption reactions, which are based on the Arrhenius equation, are mostly dependent on temperature. On the contrary, the pressure only affects physisorption reactions. Therefore, high pressure solely has a substantial impact on the physisorption step. Specifically, when pressure is increased from 100 to 500 Pa, the process time decreases by a factor of 5.03 for Step A, 4.89 for Step B, and 5.01 for Step C throughout the temperature window. Therefore, the operating pressure linearly affects the process time for all steps, which indicates that the physisorption step is the key component in determining the process time.

As shown in Figure 6.10, changing the temperature does not cause any significant impact on the process time in the aforementioned range. Due to the randomness, the process time slightly fluctuates with increasing temperature. Although an increase in temperature accelerates all surface reactions based on the simulation algorithm, Figure 6.10 reveals that the process time is rarely reliant on temperature. Prior research reported that the deposition of BDEAS was nearly constant [95]. With regard to the kMC algorithm, the reaction rate constants of the surface reactions are based on the Arrhenius equation in Eq. (6.3) increase with increasing temperature. However, increasing temperature decelerates the physisorption reaction rate. With the independence of temperature in the ALD growth process, the reduction in the physisorption reaction rates and the increase in the surface reaction rates with increasing temperature offset each other.

148

(a) Step A (Hacac cycle)

(b) Step B (BDEAS cycle)

(c) Step C (O$_3$ cycle)
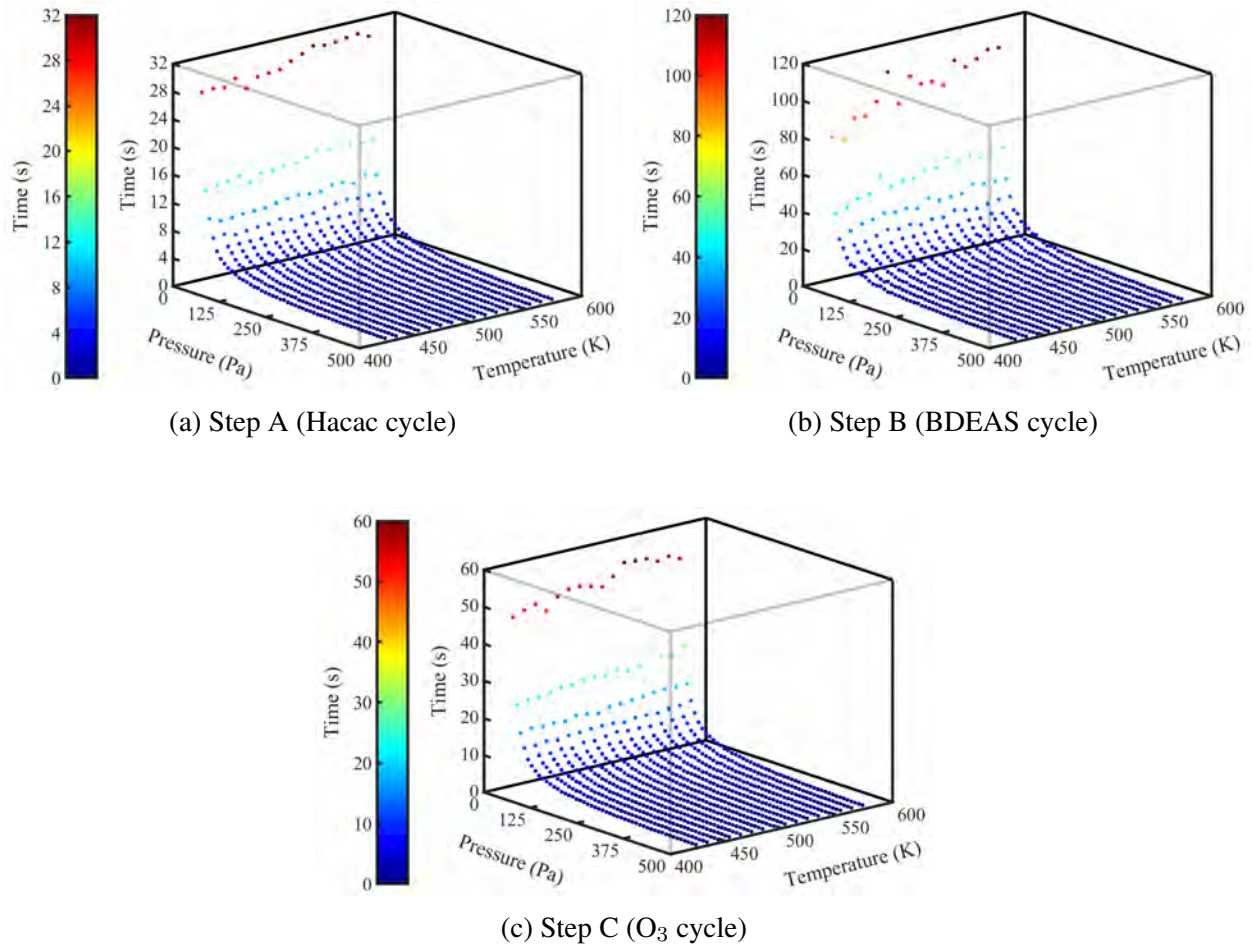
Figure 6.9: Scatter charts for process time collected from kMC simulations under the operating window ($423 \leq T \leq 573\ K$ and $10 \leq P \leq 500\ Pa$) in (a) Step A, (b) Step B, and (c) Step C, respectively.

The stochastic behavior of the kMC algorithm may present a variation in dosage times for reaching full coverage. Thus, 100 simulations of the same condition were simulated and averaged to compensate for the randomness and deviation in the dosage times. Histograms depicting the distribution of the dosage times for Steps A, B, and C, are presented in Figure 6.11 and illustrate that most dosage times computed were within a single standard deviation, σ, from the average, μ, dosage time. Therefore, the stochastic behavior has a moderate impact on the precision of the dosage times, which is the reason why averaging the dosage time is an effective method for
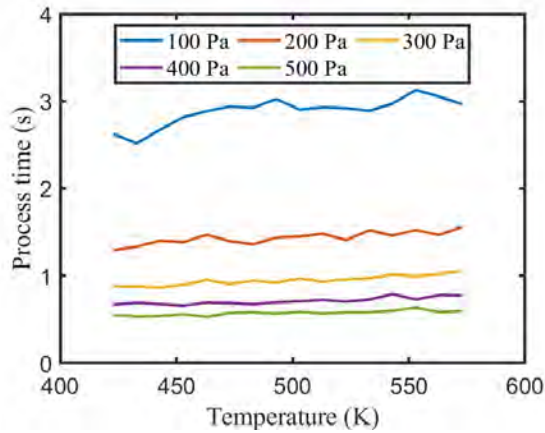
(a) Step A (Hacac cycle)
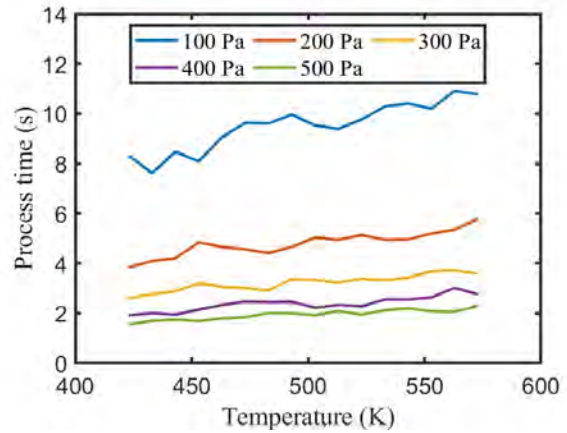
(b) Step B (BDEAS cycle)
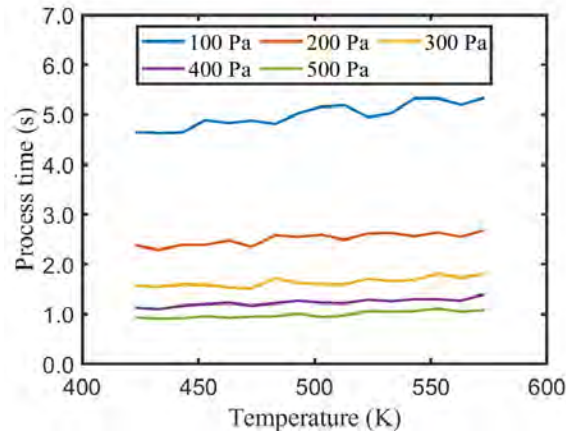
(c) Step C (O$_3$ cycle)

Figure 6.10: 2D plots of process time as a function of temperature with different pressures in (a) Step A, (b) Step B, and (c) Step C, respectively.

producing consistent results.

6.4 Conclusion

In this work, a combined atomistic and mesoscopic model considering the effects of chemoselectivity and regioselectivity (steric hindrance) based on a kinetic Monte Carlo (kMC) simulation was developed to investigate the surface kinetics of area-selective atomic layer deposition
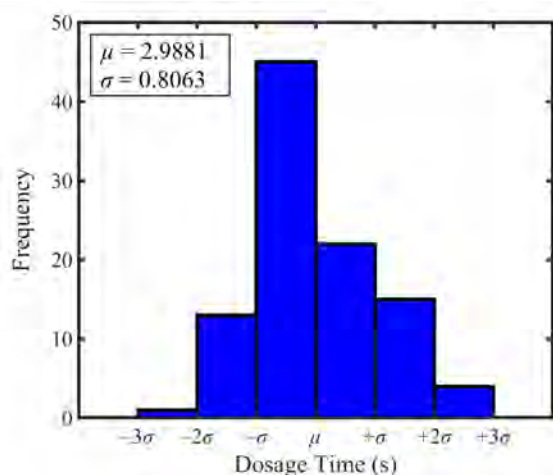
(a) Step A (Hacac cycle)
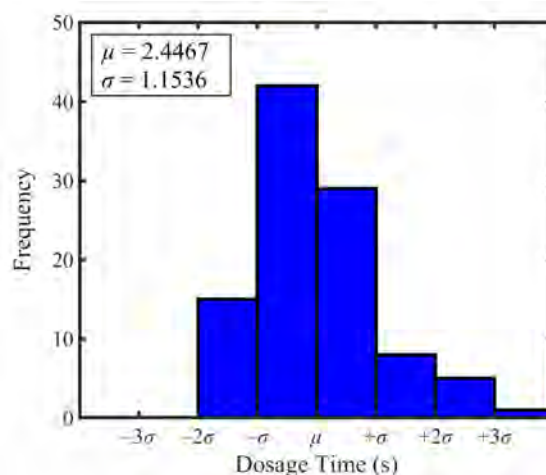
(b) Step B (BDEAS cycle)
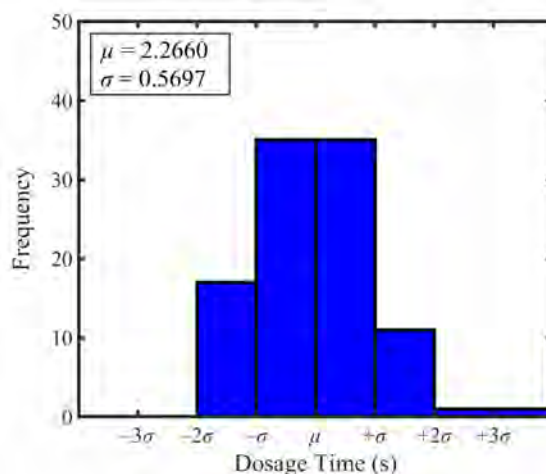
(c) Step C (O_3 cycle)

Figure 6.11: Histograms depicting the distribution of dosage times for 100 iterations while at constant operating conditions for (a) Step A ($T = 523\ K, P = 100\ Pa$), (b) Step B ($T = 523\ K, P = 350\ Pa$), and (c) Step C ($T = 523\ K, P = 200\ Pa$).

(ASALD) of SiO_2/Al_2O_3. Acetylacetone (Hacac), bis(diethylamino)silane (BDEAS), and ozone were used for Steps A through C as small molecular inhibitors (SMI), precursors, and oxidants, respectively. Density functional theory (DFT) calculations including the nudged elastic band (NEB) method were performed to obtain kinetic parameters for the basis of the kinetic mechanisms. This study revealed that the inhibitor forms a protective layer on the non-growth area (NGA), Al_2O_3, as

a preferential reaction path as opposed to the growth area (GA), SiO_2, by visualizing the surface coverage of Hacac, demonstrating the chemoselectivity of the Hacac adsorption. Additionally, the regioselective behavior of bulky species, Hacac and BDEAS, was simulated, which determined the preference for particular configurations (chelate configuration) that would provide the least steric hindrance to maximize the surface coverage. It was also observed that regioselective dependence of BDEAS adsorption prevented the precursor from adsorbing onto the Hacac protective layer as a consequence of steric repulsion effects. In addition, the temperature and pressure dependence on the ALD growth on the GA was discussed. The ALD growth was linearly dependent on pressure, while there was no impact of temperature on the ALD growth on the GA. This simulation work was supported and demonstrated by several experimental studies as discussed in Section 6.3. Therefore, the developed mesoscopic model greatly exemplifies the surface kinetics at the atomic level, providing insight into ASALD as a brand-new technology for bottom-up nanofabrication. This research can be integrated with not only other experimental research, but also a computational fluid dynamics model to optimize the industrial reactor design and the operating conditions of the ASALD process in future research.

Table 6.1: Summary and description of variables.

Variable	Definition
A_{site}	Surface area of an active reaction site
D	Dimension size of a $D \times D$ 2D grid
E	Total electronic energy
E_a	Activation energy
h	Planck constant
$\hbar$	Reduced Planck constant
$\hat{H}$	Hamiltonian operator
k	Reaction rate constant
k_{ads}	Reaction rate constant of adsorption reaction
k_{nonad}	Reaction rate constant of nonadsorption reaction
k_B	Boltzmann constant
k_{sum}	Sum of the reaction constants
m	Atomic mass of adsorption species
m_e	Atomic mass of particle
n	Number of molecules
P	Adsorption species partial pressure
Q	Partition function for the reactant
$Q^{\ddagger}$	Partition function for the transition state
R	Universal gas constant
s	Distance between two adjacent active sites
t	Reaction time progress
T	Operating temperature (absolute)
$\hat{V}$	Potential energy
x	Horizontal distance between two active sites
$\overrightarrow{x}$	Displacement vector
Z	Coordination number of adsorption species
Δt	Time interval
γ_1, γ_2	Random numbers where $\gamma_1, \gamma_2 \in (0, 1]$
ν	Pre-exponential factor
ρ	Molecule density
σ	Sticking coefficient for adsorption species
ψ	Time and position-dependent wave function

Table 6.2: Self-determined parameters for DFT calculations.

Material	Variable	Value
SiO_2	*degauss*	0.02
	$ecutwfc^\dagger$	50 *Ry*
	$ecutrho^\ddagger$	200 *Ry*
	k-points	2 2 2
Al_2O_3	*degauss*	0.01
	$ecutwfc^\dagger$	50 *Ry*
	$ecutrho^\ddagger$	200 *Ry*
	k-points	3 3 2

[†] Kinetic energy cutoff for wave functions.
[‡] Kinetic energy cutoff for charge density and potential

Table 6.3: Geometry description for reaction paths

Substrate	Step	Number	Description
Al_2O_3	Step A	A1	OH-terminated Al_2O_3
		A2	Hacac physisorption on Al_2O_3
		A3	Monodentate configuration (deprotonated Hacac on Al_2O_3)
		A4	Chelate configuration (deprotonated Hacac on Al_2O_3)
SiO_2	Step A	V1	OH-terminated SiO_2
		V9	Hacac physisorption on SiO_2
		V10	Monodentate configuration (deprotonated Hacac on SiO_2)
	Step B	V2	BDEAS physisorption on SiO_2
		V3	DEA-deprived BDEAS adsorbate on SiO_2
		V4	SiH_2 adsorbate on SiO_2
	Step C	V5	O_3 physisorption on SiH_2
		V6	H-Si-OH on SiO_2
		V7	O_3 physisorption on H-Si-OH
		V8	Final configuration ($Si(OH)_2$)

Table 6.4: Van der Waals radii of functional groups in Hacac molecule

Functional group	Van der Waals radius (Å)
$-CH_3$	2.0
$-CH_2-$	2.0
$C\!=\!O$	1.7

Sourced from [94]

Chapter 7

Multiscale CFD Modeling of Area-Selective Atomic Layer Deposition: Application to Reactor Design and Operating Condition Calculation

7.1   Introduction

Recently, there has been growing interest in spatial atomic layer deposition, which is a potential substitute for conventional ALD processes in that it is capable of high-volume manufacturing [90]. This high throughput capability is made possible by the injection of reagents into separate reaction zones within the spatial reactor that are isolated with an inert gas curtain and vacuum zones. Figure 7.1 depicts the schematic diagram of a sheet-to-sheet spatial ALD reactor configuration where the substrate is transferred between reaction zones that are isolated by inert gases (gas

curtain) by a conveyor belt. Due to the spatially separated reaction zones, no lengthy purging steps are required to maintain the self-limiting nature of the reactions. In other words, spatial ALD is able to save a significant amount of purging time, resulting in a high throughput ALD process. There have been various proposals of different spatial ALD reactor configurations. The original concept of a spatial reactor as a roll-to-roll reactor was first proposed in 1977 [106], of which applications were limited to flexible substrates. Afterward, a few sheet-to-sheet designs for solar photovoltaic technology and other applications have been developed and commercialized [105, 34]. Recent *in silico* research into multiphysics simulations have focused on sheet-to-sheet designs to provide valuable insight into a route for further industrialization. For instance, several works have explored the effects of the gap distance, which is the separation distance between the dividers and surface of the substrate in Figure 7.1, on the flow dynamics [86] as well as reactor operating conditions including pumping pressure and substrate velocity [18, 66]. Furthermore, ref [22] examined how precursor flow rates and other process operating parameters impact on transport phenomena and surface reactions. Additionally, prior works [117, 108] investigated the effect of the substrate velocity on film uniformity and surface kinetics using a three-dimensional multiscale computational fluid dynamics model. Despite the potential of high-volume production, the sheet-to-sheet reactors weaken in terms of versatile applications since it is impossible to avoid necessitating larger equipment with longer substrates compared to standardized wafers. As another approach, rotary designs have been experimented on a laboratory scale, and they have demonstrated the operational capability to provide uniform film growths [100, 89]. Furthermore, the dimensions of rotary designs are comparable to that of stationary ALD reactors. Nevertheless, a few studies have carried out experiments to evaluate the rotary design [100, 91]. Moreover, aforementioned *in silico* stud-

ies have solitarily focused on the sheet-to-sheet type reactors, which are significantly different from rotary type reactors in many aspects including transport phenomena, reactor configuration optimization, and control strategies. At this stage, research studies aided by computational fluid dynamics simulations are needed to thoroughly investigate the surface kinetics in conjunction with the fluid dynamics of rotary designs.
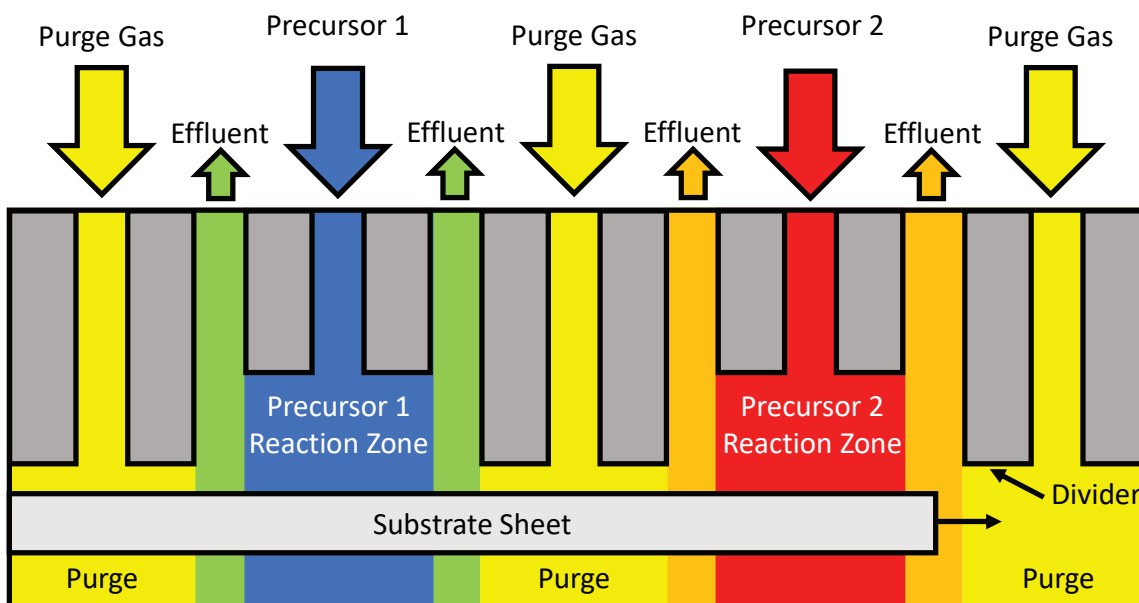
Figure 7.1: Sheet to sheet type reactor design for a two-step, spatial atomic layer deposition process. Reagents are continuously introduced in separate reaction zones, and a plate-shaped substrate is transported via a conveyor belt through each zone.

Therefore, in this study, a computational simulation method, the so-called multiscale model, is presented to qualitatively and quantitatively analyze the features of a rotary spatial ALD reactor design and the ABC-type ASALD process in an operable ALD window. The structure of this paper is as follows: Section 7.2.1 describes the methodology of integrating an atomistic-mesoscopic model and a macroscopic model, Section 7.2.2 illustrates atomistic-mesoscopic modeling with kinetic descriptions, Section 7.2.3 explains the methodology of macroscopic modeling, Section 7.3

158

discusses the results of the multiscale modeling work, and finally, Section 7.4 encapsulates the conclusion of this work.

7.2 Multiscale Modeling

7.2.1 Multiscale Modeling Framework

It is desired to establish a quantitative relationship between the mesoscale surface kinetics on the substrate surface and the macroscopic fluid dynamics of the reactor to provide an effective and reproducible method for accurately simulating the overall process through a multiscale modeling framework [72]. Additionally, multiscale models are capable of modeling through various time and length scales, as illustrated in Figure 7.2. In this work, a multiscale model is developed by integrating an atomistic-mesoscopic and a macroscopic simulation that are conjoined by a data-sharing programming methodology. This new model operates without requiring robust computational resources through appropriate process assumptions and simplifications, and provides an accurate representation of realistic phenomena encountered in industrial settings that have been incorporated into this work.

The atomistic-mesoscopic model of the ABC-type ASALD of SiO_2/Al_2O_3 surfaces with a SMI was first developed by [115], and it will be integrated into the multiscale model of this work. For the purposes of this work, the atomistic-mesoscopic model will be referred to as the mesoscopic model for simplicity. The mesoscopic model conjoins first principles *ab initio* quantum mechanics (e.g., Density Functional Theory and Nudged Elastic Band) and statistical mechanics (e.g., Transition State Theory and Collision Theory) to evaluate reaction rate constants that will
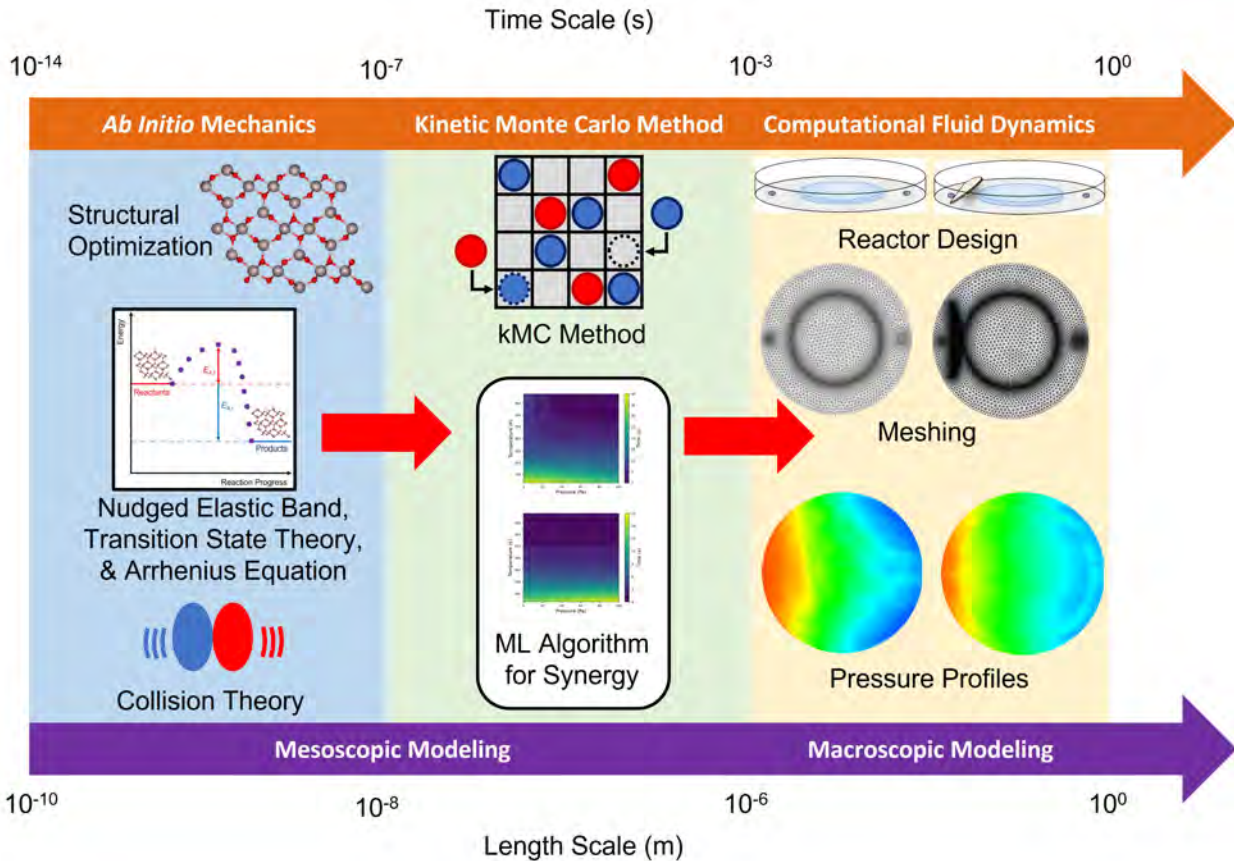
Figure 7.2: Diagram depicting the various length and time scales applicable for the multiscale modeling simulation, which is composed of mesoscopic and macroscopic modeling. Mesoscopic modeling is composed of *ab initio* quantum and statistical mechanics and the kinetic Monte Carlo method, and the macroscopic model consists of reactor modeling and computational fluid dynamics simulations. Prior work [116] has utilized machine learning or regression methods to determine optimal operating conditions to achieve high synergy.

be integrated into a kinetic Monte Carlo algorithm to monitor the surface kinetics. While the mesoscopic simulation records the evolution of surface kinetics, a macroscopic model is needed to examine the continuum mechanics in the ambient fluid conditions. A three-dimensional computational fluid dynamics (CFD) simulation is employed to observe the spatiotemporal behavior of the species and fluid dynamics in a reaction chamber through a rigorous process of reactor optimization, mesh discretization, and tailoring the macroscopic model to the ASALD reactor (e.g.,

160

boundary conditions, user-defined functions, numerical solver methods, and operating conditions). From the development of the mesoscopic and macroscopic domains, the simulations are conjoined by linking the surface partial pressure data evaluated from the macroscopic CFD model to the kMC simulation, which successively evaluates the source accumulation rates of species for a time progression to be defined to the macroscopic CFD model. Thus, a multiscale model is intimately constructed that enables the simultaneous modeling of surface kinetics and continuum mechanics in a spatiotemporal manner, which is depicted by the process diagram in Figure 7.3.
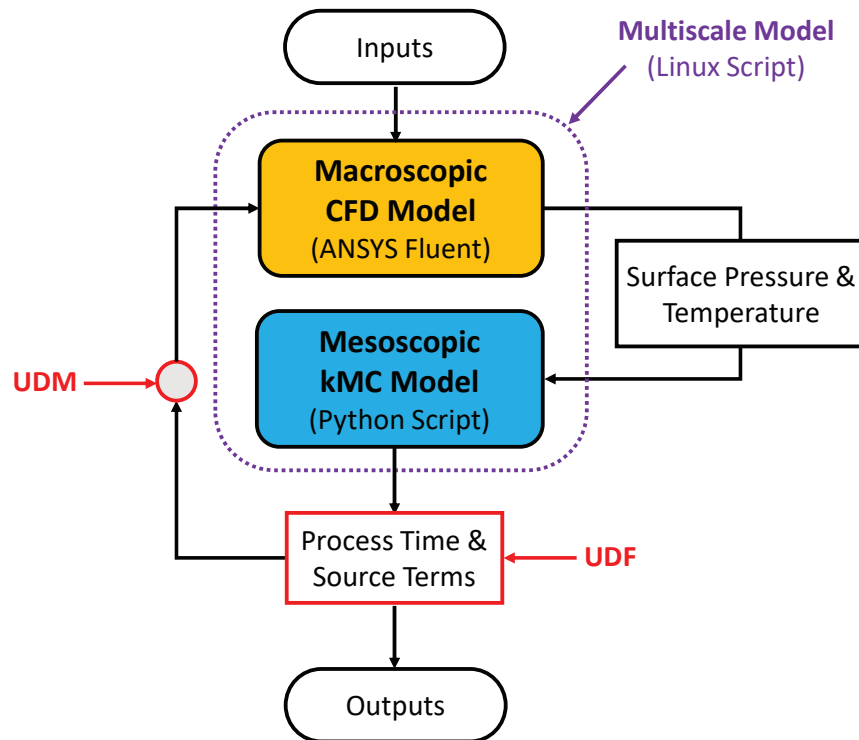
Figure 7.3: Coding architecture of the multiscale simulation using various programming languages to conjoin the mesoscopic and macroscopic simulations. In a Linux Shell script as an interface between the macroscopic and mesoscopic models, two models are simulated while exchanging pressure, temperature, and source generation (or consumption) with each other.

7.2.2 Mesoscopic Model

Surface kinetics

The abundance of undetermined reaction pathways present a challenge for defining the over-all reaction mechanism for each step in the ASALD process. One methodology is to simplify the reaction mechanism to purely consist of rate determining reactions that are also elementary. Thus, the surface chemistries involved in ASALD of SiO_2/Al_2O_3 are selected from experimental works [71, 95]. Below, the proposed reaction pathways for each step in the ABC-ASALD process are summarized. There are three reactions that the mesoscopic model must simulate that together make up one ASALD cycle. The first reaction, Step A, is where acetylacetone (Hacac) adsorbs to Al_2O_3 to act as an inhibitor on the non growth area. This reaction has the following steps:
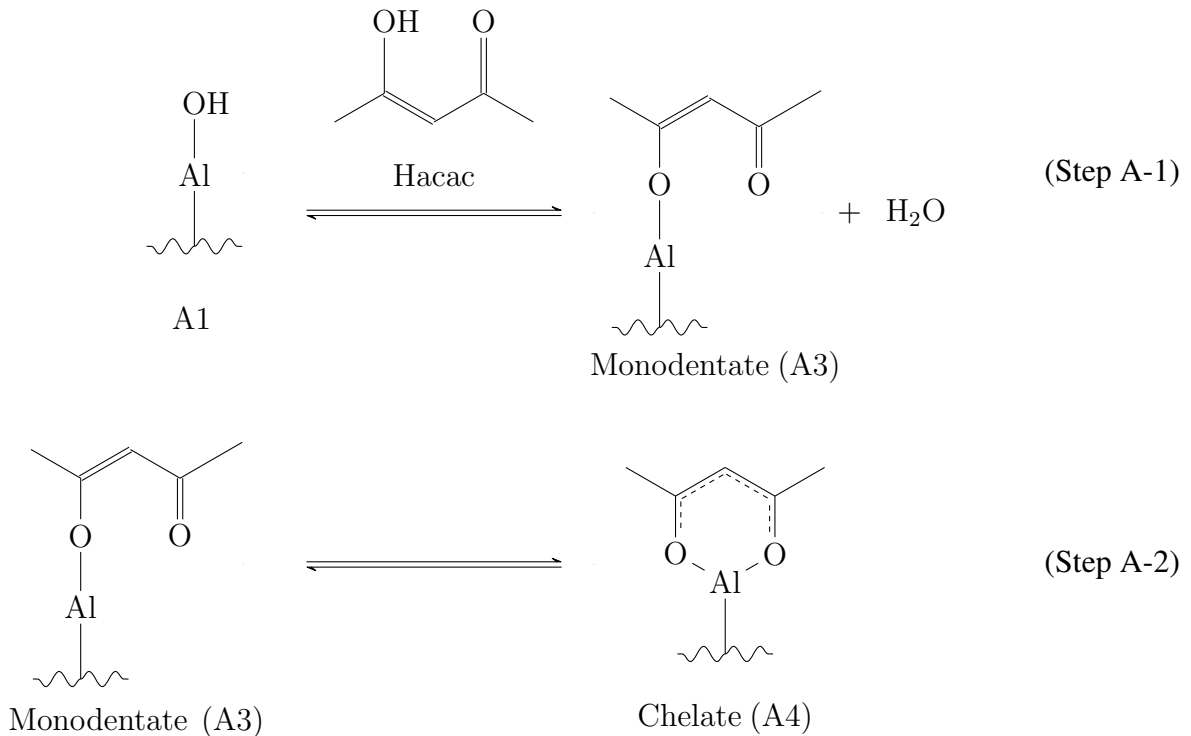
(Step A-1)

(Step A-2)

In Step A-1, an acidic, keto-enol tautomerized Hacac molecule reacts with the basic hydroxyl group on the Al_2O_3 surface, which produces water vapor as a byproduct. Then, in the following step, the monodentate structure of the adsorbed Hacac undergoes a rearrangement mechanism to produce a more stable, chelate form.

The second reaction, Step B, is composed of an adsorption step of is(diethylamino)silane (BDEAS) onto SiO_2 to propagate the growth of a monolayer of SiO_2, which is referred to as the growth area. Analogously to Step A, this reaction also consists of two steps:
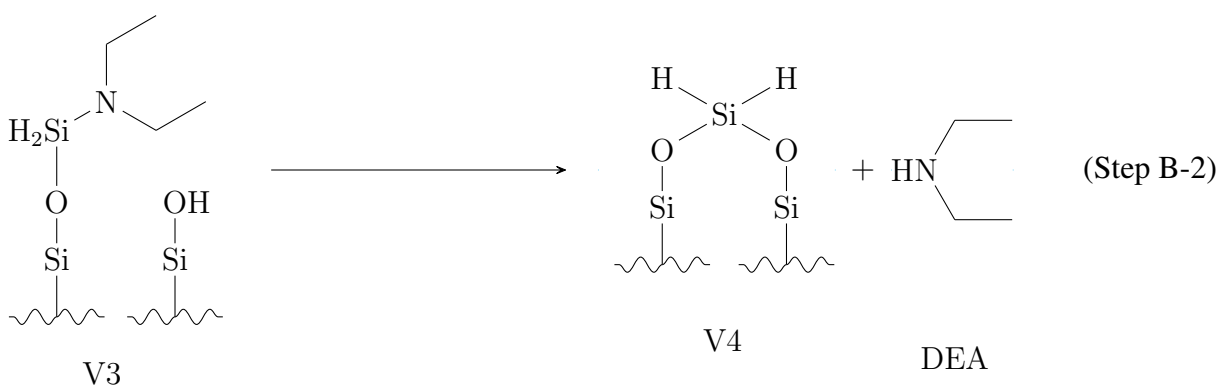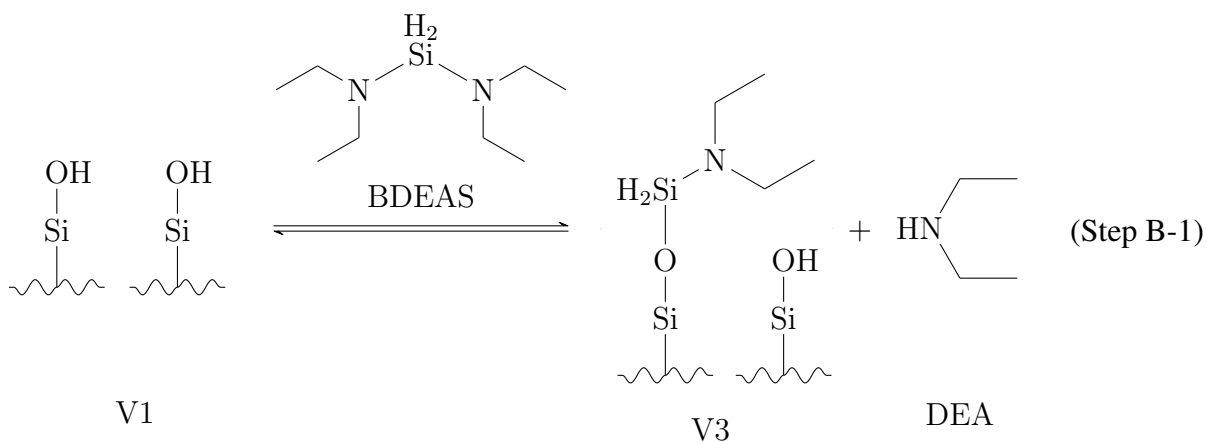
(Step B-1)

(Step B-2)

In Step B-1, the bulky BDEAS molecule adsorbs to one hydroxyl site on SiO_2, which produces one molecule of diethylamine (DEA) as a byproduct. In Step B-2, the BDEAS molecule binds to an adjacent hydroxyl site on SiO_2, which produces a second DEA molecule as a byproduct. After Step B is completed, a monolayer of silicon atoms will have been deposited on the original SiO_2 surface.

Finally, the third reaction, Step C, is where the monolayer of Si is oxidized with O_3 to finish growing the layer of SiO_2. Its reactions have the following form:

$$
\begin{array}{ccc}
\underset{V4}{
\begin{array}{c}
H\diagdown \quad \diagup H \\
Si \\
O \quad \quad O \\
| \quad \quad | \\
Si \quad \quad Si
\end{array}}
&\xrightarrow{\;O_3\;}&
\underset{V6}{
\begin{array}{c}
HO\diagdown \quad \diagup H \\
Si \\
O \quad \quad O \\
| \quad \quad | \\
Si \quad \quad Si
\end{array}}
\;+\; O_2
\end{array}
\qquad \text{(Step C-1)}
$$

$$
\begin{array}{ccc}
\underset{V6}{
\begin{array}{c}
HO\diagdown \quad \diagup H \\
Si \\
O \quad \quad O \\
| \quad \quad | \\
Si \quad \quad Si
\end{array}}
&\xrightarrow{\;O_3\;}&
\underset{V8}{
\begin{array}{c}
HO\diagdown \quad \diagup OH \\
Si \\
O \quad \quad O \\
| \quad \quad | \\
Si \quad \quad Si
\end{array}}
\;+\; O_2
\end{array}
\qquad \text{(Step C-2)}
$$
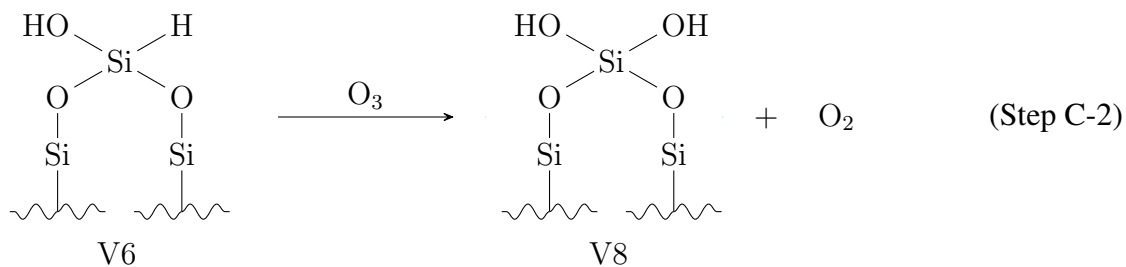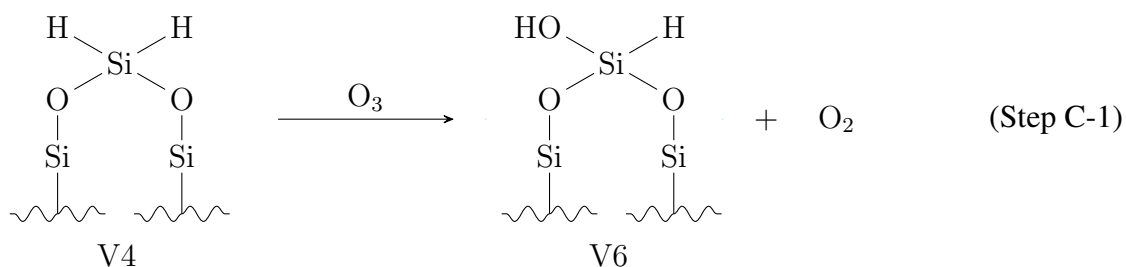
Here, two ozone atoms are consumed per Si atom, leading to a homogeneous array of geminal hydroxylated ligands of SiO_2 with two O_2 atoms as the byproduct. For the kMC algorithm, the kinetic rate constants of all reactions (above) including physisorption and desorption of the reagents must be evaluated as a function of temperature and partial pressures. Detailed descriptions of the

reaction pathways and kinetic parameters can be found in [115].

Quantification of Surface Reactions

Various *ab initio* quantum and statistical mechanics as well as first-principles physical chemistry approaches were integrated into this work to evaluate the reaction rate constants for each reaction. This section discusses the theoretical approaches for computing the reaction rate constants for each reaction, which are categorized into two classes: nonadsorption or adsorption.

For nonadsorption reactions, their reaction rate constants are calculated using the Arrhenius equation, which requires the activation energy and pre-exponential terms to be computed. The first step towards the computing of these reaction rate constants is to generate an optimized crystalline and molecular structure of the reactants and products of the nonadsorption reactions. One underlying assumption of these optimized molecular and crystalline structures is that the surface of the wafer, which is amorphous in nature, is simplified to a homogeneous crystalline structure [92, 12] and is modeled without the presence of crystallographic defects. This procedure has been conducted in prior microscopic and mesoscopic modeling of atomic layer etching and area-selective atomic layer deposition processes to alleviate the computational burden of the simulation and to reduce the complexity of the coding architecture for the kMC simulation while maintaining some structural fidelity [116, 115]. The computations for structural optimization are achieved by employing the open-source molecular dynamics program, Quantum ESPRESSO [96], to perform a structural optimization of the lattice structure of the solid materials and the molecular structure of the gaseous species. This procedure is conducted by finding the minimum, total electronic energy of the species, E_{total}, through tuning the independent parameters $\mathbf{k}$, E_ψ, and E_ρ, which are the

k-points (specified as *automatic*), the kinetic energy cut-off for wave functions (*ecutwfc*), and the kinetic energy cut-off for charge density and potential (*ecutrho*), respectively, in an iterative procedure until a convergence threshold specified to the program is satisfied. The parameters $\mathbf{k}$, E_ψ, and E_ρ, are adjusted until the infimum of the total electronic energy is obtained, which is described below:

$$\inf_{\mathbf{k} \in \mathbb{R}^6, E_\psi \in \mathbb{R}, E_\rho \in \mathbb{R}} E_{\text{total}} \left(\mathbf{k}, E_\psi, E_\rho \right) \tag{7.1}$$

Following the molecular optimization procedure, the nudged elastic band (NEB) method is employed to evaluate the activation energies for the nonadsorption reactions. The NEB approach is a transition-path sampling procedure that identifies saddle points to determine the minimum energy path between the reactant(s) and product(s) by generating a user-defined number of images of the conversion and atomic movement for each molecular structure [50]. Although a higher number of images (intermediate transition states with minimum energy path) is essential for computing molecular properties of high fidelity, such a simulation will require greater computational resources to complete in ample time as well; therefore, a total of nine images were used in a prior work [115] with results that were verified to experimentally recorded data results observed by [71, 73, 95]. Once all the intermediate images are produced along with the saddle point, the maximum energy peak is determined, which is utilized to compute the activation energy for the forward and reverse reactions. With the activation energy, the reaction rate constant for each nonadsorption reaction is

calculated with the Arrhenius equation, which is shown below:

$$k_{nonad} = \nu \exp\left(-\frac{E_{act}}{RT}\right)$$

(7.2)

where k_{nonad} is the reaction rate constant for a nonadsorption reaction, ν is the temperature de-pendent pre-exponential factor, E_{act} is the activation energy of the reaction, R is the universal gas constant, and T is the absolute temperature of the reaction. The calculation of the pre-exponential factor is simplified by assuming that the ratio between the transition state and reactant partition functions is unity [48]. This yields the following equation for the pre-exponential factor ν, which is dependent on the temperature, T, of the reaction:

$$\nu = \frac{k_B T}{h}$$

(7.3)

where k_B is the Boltzmann constant and h is the Planck constant.

For adsorption reactions, it is simpler to calculate their reaction rate constants, which are mod-eled as bimolecular reactions that follow Maxwell-Boltzmann statistics and Collision Theory. The reaction rate constants for adsorbate species, s, are then characterized by the following equation:

$$k_{ads,s} = \frac{P_s A_{site} \sigma_s}{Z_s \sqrt{2\pi m_s k_B T}}$$

(7.4)

where $k_{ads,s}$ is the reaction rate constant for an adsorption reaction, P_s is the partial pressure of the gaseous reagent, s, A_{site} is the surface area of a single active site, σ_s is an experimentally determined sticking coefficient unique to the reagent s, Z_s is the coordination number of the gas

s, m_s is the atomic mass of the gaseous reagent s, k_B is the Boltzmann constant, and T is the absolute temperature of the ambient environment. The sticking coefficient was found for Hacac to be 1.0×10^{-4} [37], that of BDEAS to be 2.0×10^{-5} [97], and that of O_3 to be 4.5×10^{-5} [58]. In [115], these sticking coefficients were validated through comparison to experimental results from [73].

kMC Algorithm

After characterizing the various reaction parameters through *ab initio* quantum and statistical mechanics simulations as described above, the computationally efficient kinetic Monte Carlo (kMC) algorithm is integrated into the mesoscopic model simulation. The kMC algorithm is an approach that uses randomly generated numbers to reflect the stochastic and chaotic behavior of reactions occurring on the atomic scale [110]. This practice is used to model the adsorption and interchanging of various species on the surface of the substrate in this study. The procedure involves the weighting of potential reactions occurring on active sites using a random number, which are discretized into a 300×300 array in the Python programming language, resulting in 90,000 reaction sites in total. Following the random selection of the reaction using the first random number, the duration of the reaction progression is then evaluated through a secondary random number that is independent of the first random number. A summary of the procedures from [48] are described below and illustrated in Figure 7.4:

$$k_{tot} = \sum_{i-1}^{N} k_i \tag{7.5}$$

168
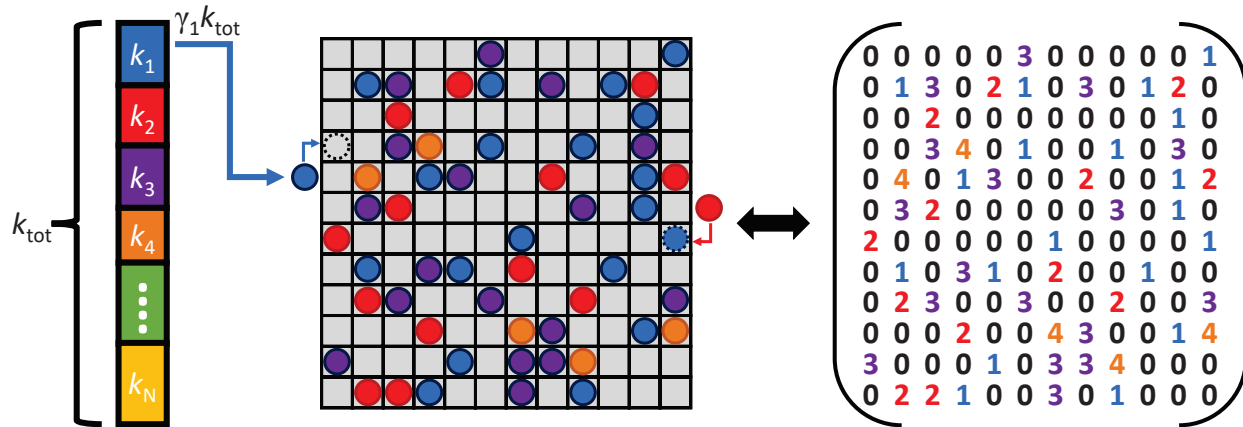
Figure 7.4: Kinetic Monte Carlo simulation on lattice diagram. For each reaction site, an identifiable number is allocated corresponding to a reaction if the selection condition of a certain reaction path is satisfied in Equation (7.6).

where k_{tot} is the sum of all possible reaction rate constants, k_i is the reaction rate constant for the ith potential reaction in the reaction mechanism, and N denotes the total number of possible reaction pathways for the lattice site.

The kMC method establishes a stochastic procedure by randomly assigning a number, $\gamma_1 \in (0, 1]$ to select one of the reaction rate constants that were weighted during the prior summation procedure [110]. Larger reaction rates will have a higher probability for selection due to the weighting procedure, which is also consistent with the nature of kinetics in favoring reaction processes with higher reaction rate constants or lower activation energies.

$$\sum_{i=1}^{j-1} k_i \leq \gamma_1 k_{tot} \leq \sum_{i=1}^{j} k_i \tag{7.6}$$

Above, j represents the reaction index chosen from the N possible reactions. If the above equation is satisfied, the reaction j is selected for the reaction at the site. If the selection procedure for all reaction sites is complete, lastly, the time progression in which the site of reaction is undisturbed

is evaluated using a secondary random number, $\gamma_2 \in (0, 1]$.

$$\Delta t = \frac{-\ln{(\gamma_2)}}{k_{tot}} \tag{7.7}$$

Although the kMC algorithm provides an efficient procedure for predicting the occurrence of a reaction, it alone is not enough to provide an accurate description of the overall kinetics. While activation energy barriers ultimately dictate the directionality of reactions, other factors such as steric hindrance affect the probability of a successful adsorption reaction. The steric hindrance originates from the bulkiness of the Hacac and BDEAS molecules in Steps A and B as the molecules must adsorb to multiple active sites to fully adhere to the SiO_2/Al_2O_3 surface. When these two molecules adsorb to an active site on their respective substrates, the adsorbates are bulky that they inhibit adjacent active sites from properly binding to other molecules, thereby generating atomic vacancies. Thus, the kMC algorithm must be modified to account for these physical limitations by exercising a novel methodology that simulates hindered sites. In a prior study, ref [115] performed this modification to the kMC algorithm by modeling the adsorption of individual Hacac and BDEAS molecules by considering their orientations on the site of adsorption, which is integrated into the multiscale modeling framework in this work.

The kinetic model described above was validated by comparing the *in silico simulation results of Hacac adsorption on Al₂O₃ to reported experimental results of the same reaction. The kinetic model simulated a chelate surface density of 1.45 ± 0.02 molecules/nm² [115] while experimental results determined a chelate surface density of 1.7 ± 0.1 molecules/nm² [73]. This* deviation is attributed to the longer atomic distances calculated with Quantum ESPRESSO com-

pared to the shorter atomic distances observed by [73]. Additionally, the ratio of Hacac chelate to monodentates is also comparable with experimental findings. The kinetic model simulation computed a monodentate fraction of $19.37 \pm 1.40\%$ [115] while the experimental value of monodentate fraction was found to be $20 \pm 5\%$ [73] through infrared (IR) spectroscopy. Furthermore, the computed process time to achieve full coverage for the mesoscopic simulation was determined to be similar to results conducted in experimental laboratory settings [115]; thus, the kMC algorithm is incorporated into the multiscale model with confidence.

7.2.3 Macroscopic Model

Macroscopic Modeling Logistics

The macroscopic model consists of preprocessing and postprocessing procedures that are intended to simulate fluid dynamics for the ASALD process. First, reactor configurations are assembled through computer aided design (CAD) software. The configurations are optimized computationally through a mesh of reasonable quality which ensures lower simulation times and higher substrate quality. Once the reactor model is fully developed, the simulation is tailored to the ASALD process environment by customizing various solver, physics, and personalized settings such as boundary conditions, numerical solver methods, convergence criteria, user-defined functions, materials, fluid dynamics conditions, and remeshing methods. This section will discuss the various settings that are integrated into the macroscopic computational fluid dynamics (CFD) simulation.

Ansys Workbench (2022R2), a simulation integration platform, is used to access various applications for constructing reactors through computer aided design (CAD) software, performing

discretized meshing, and conducting computational fluid dynamics (CFD) simulation. In Ansys Workbench, Ansys DesignModeler is used to build reactor geometries while the spatiotemporal fluid dynamics of the gas-phase domain inside the designed reactor configurations are simulated using Ansys Fluent.

Reactor Geometry and Meshing

Different rotary type reactor configurations are devised, which originate from the industrial patent of Applied Materials, Inc [65]. Figure 7.5 illustrates the schematic representation of the spatial reactor designs to investigate flow profiles. **R1** has a ring-shaped outlet for each reaction zone and 6 10-mm round inlets (3 for the reagents and 3 for nitrogen), as depicted in Figure 7.5a. Meanwhile, Figure 7.5b shows **R2**, and it is designed with three ring, sector-shaped outlets and 6 round inlets. **R3** features asymmetrical round inlets compared to **R2** as shown in Figure 7.5c. Figure 7.5d reveals that **R4** is equipped with a ring, sector-shaped N_2 inlet for each reaction zone, enclosing a ring, sector-shaped outlet and an asymmetrical annular, sector-shaped reagent inlet. Each reactor is designed with a diameter of 760 mm, in which three wafers with a diameter of 200 mm are mounted on the rotating plate at the bottom of the reactor. The rotating plate spins, causing the wafers to encounter the reaction zones A, B, and C sequentially. The gas distribution assembly consists of differently shaped inlets, outlets, and vacuum ports and is located on top of the reactor. With **R1** through **R4**, the effects of the differences in geometry are qualitatively and quantitatively analyzed to propose an optimal reactor geometry. Finally, with the selected reactor design, process operation with different operating conditions is explored through multiscale simulations.

In this study, the operating pressure of the reactor and the vacuum pressure of the outlets are

(a) **R1**

(b) **R2**

(c) **R3**

(d) **R4**
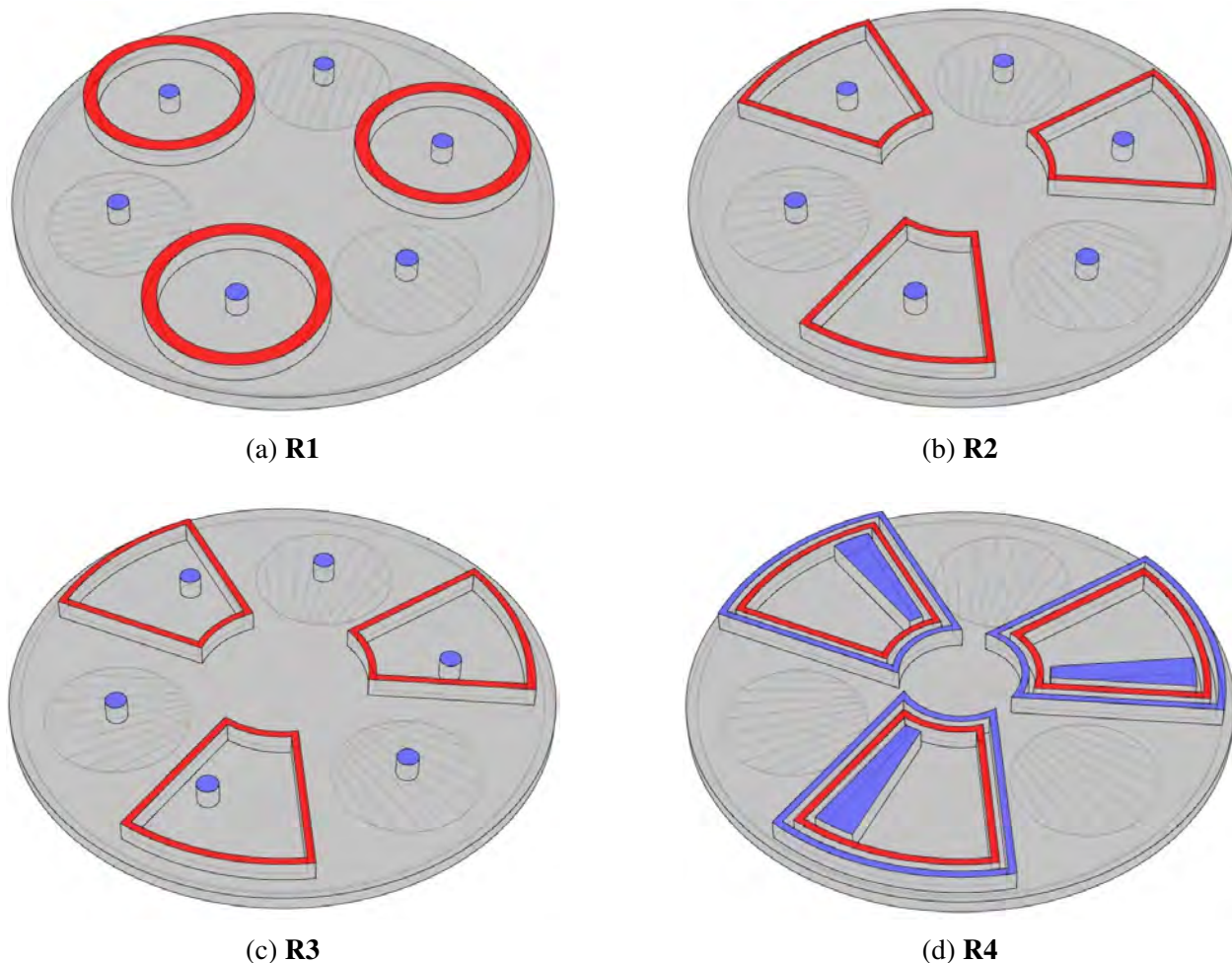
Figure 7.5: Four types of spatial, rotary ASALD reactor models with various inlet and outlet geometries (ring or annular sector) and positioning (symmetric or asymmetric) of reagent inlets. Red and blue areas indicate the outlets and inlets, respectively. 3 wafers are placed at the bottom inside the reactors.

set to 1330 Pa and -150 Pa, respectively, and the temperature is set to 523 K. These conditions are considered typical for silicon oxide deposition [71, 95]. It is assumed that a PID (proportional-integral-derivative) controller regulates the operating temperature on each wafer surface so that the temperature of the wafers is maintained at a constant temperature of 523 K. The three reagents (Hacac, BDEAS, and ozone) are injected through separate inlet ports as a gas mixture with inert gas, N_2. It was reported that there is no significant temperature dependence in the ASALD pro-

cess [115, 95]; therefore, multiscale simulations are performed with a fixed temperature of 523 K while rotation speed is varied in the range of 0.2 to 0.8 rad/s and the mole fractions of the reagents are varied in the range of 0.05 to 0.7 each in Section 7.3.2.

The meshing package on Ansys Workbench is used to generate high-quality meshes. Mesh discretization is conducted to obtain numerical solutions by adopting the finite volume method, and has a significant impact on the accuracy of the solution. Tetrahedral elements are chosen as they are recommended for 3D geometries by providing high mesh convergence. For mesh sizing, proximity, and curvature, refinements are conducted to capture more accurate flow profiles in thin reactor configurations. In addition, the face mesh of the outlets (i.e., vacuum ports) is enhanced through face meshing controls to determine the suitable divisions of the face mesh. To capture the motion of the rotating boundaries, the dynamic mesh model is necessary for modeling flow dynamics spatiotemporally. This feature facilitates the process for updating the volume mesh while preserving the quality of the mesh. The mesh quality of the reactor configurations that are developed in this paper is summarized in Table 7.1. The constructed meshes are qualified for CFD simulations in accordance with the standardized ranges provided by [4].

Table 7.1: Average values of mesh quality factors for various rotary reactor configurations.

No.	* Skewness	** Orthogonal Quality	*** Element Quality	**** Aspect Ratio
R1	0.2536	0.7446	0.8194	1.9136
R2	0.2471	0.7512	0.8244	1.8949
R3	0.2469	0.7514	0.8246	1.8933
[†] **R4**	0.2483	0.7499	0.8239	1.8971

* Excellent (0–0.25), Very good (0.25–0.50), Good (0.5–0.80), Acceptable (0.80–0.94). ** Acceptable (0.15–0.20), Good (0.2–0.69), Very good (0.70–0.95), Excellent (0.95–1). *** A value of 1 represents a perfect regular element. **** Acceptable (<35) depending on mesh configuration. [†] Optimized reactor design with a gap distance of 5 mm.

Characterization of Materials and Macroscopic-Phase Reactions

Ansys Fluent contains a database of material properties that were integrated into the simulation. The consumption (Hacac, BDEAS, and O_3) and generation (H_2O, DEA, and O_2) of various species require the specification of thermodynamic properties to simulate the macroscopic behavior through computational fluid dynamics. However, some thermodynamic properties of the reagents, including Hacac and BDEAS, are limited in literature and experimental data; thus, *ab initio* quantum mechanics simulations are performed to compute the thermophysical parameters (enthalpy of formation, entropy of formation, specific heat at constant pressure) for materials with insufficient data. Quantum ESPRESSO contains programs dedicated to evaluating material properties using density functional theory (DFT) in conjunction with the Quasi Harmonic Approximation (QHA), which is integrated to evaluate the material properties. The computed thermophysical properties were cross-validated with literature results for some properties via the National Institute of Standards and Technology (NIST) database for Hacac and by comparison to similar molecules such as bis(dimethylamino)silane (BDMAS) for BDEAS. Additionally, properties including the thermal conductivity and the viscosities were also defined to each gaseous species. The thermal conductivities of each gas-phase species were approximated to that of air, which is composed of nitrogen and oxygen gas, and is similar to the composition of the species in the reactor. The viscosities of each for species not available in the Ansys Fluent database were obtained through material safety data sheets (MSDS) available online. As summarized in Table 7.2, these thermophysical properties are then defined into the macroscopic simulation on Ansys Fluent for a reference temperature of 273 K.

Table 7.2: Thermophysical material properties of species specified in the gas phase in Ansys Fluent.

Thermophysical Parameter	Hacac	BDEAS	DEA
Standard Enthalpy of Formation [J/kmol]	-3.844×10^8	-6.484×10^8	-7.25×10^7
Standard Entropy of Formation [J/kmol·K]	-3.502×10^5	5.338×10^3	-4.845×10^5
Specific Heat [J/kg·K]	208.2	Polynomial [†]	Polynomial [†]
Thermal Conductivity [†] [W/m·K]	0.0100	0.0100	0.02
Viscosity [†] [kg/(m·s)]	6.00×10^{-4}	1.10×10^{-2}	3.19×10^{-4}

[†] Obtained from NIST or MSDS references. Other parameters were calculated from Quantum ESPRESSO.

In the reactor configuration, highly pure ozone is generated from an ozone-enriched source, which is available in industry [82]. Therefore, it is assumed that ozone is obtained by liquid ozone from an ozone generator and fed to spatial reactors through the O_3 injection port. However, as an oxidizing agent, ozone is used in Step C, which is capable of decomposing in the fluid phase at high operating temperatures in a mechanism that was first proposed in 1957 [9]. Thus, it is important to account for this thermal decomposition of ozone, which is a highly unstable molecule, in the macroscopic CFD simulation.

$$O_3 \rightleftharpoons O_2 + 2O$$

$$O + O_3 \rightarrow O_2$$
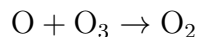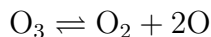
Apart from the surface kinetics established by the mesoscopic model, the volumetric reactions from the ozone decomposition are directly defined in the macroscopic model. All kinetic parameters for the decomposition of ozone were determined by [9]. The consumption of O_3 and generation of O_2 are calculated and included in the mass balance equation for ozone in Equation (7.8).

Numerical Solution Methods

The mass, momentum, and energy conservation equations are solved in the CFD simulation as described below:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m \tag{7.8}$$

$$\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla p + \nabla \cdot (\bar{\bar{\tau}}) + \rho \vec{g} + \vec{F} \tag{7.9}$$

$$\frac{\partial}{\partial t} (\rho E) + \nabla \left(\vec{v} \left(\rho E + p \right) \right) = -\nabla \left(\Sigma h_j J_j \right) + S_h \tag{7.10}$$

where ρ represents the density of the fluid composition, $\vec{v}$ is the velocity of the fluid mixture, S_m denotes the rate of species generation or consumption source term, p is the static pressure of the fluid, $\bar{\bar{\tau}}$ corresponds to the symmetric rank-two stress tensor, $\vec{g}$ is the gravitational acceleration constant, $\vec{F}$ is the external body force, E is the internal energy of the system, h_j symbolizes the sensible enthalpy of species j, J_j represents the diffusion flux of species j, and S_h denotes the heat flux source rate term. Spatiotemporal numerical solutions for each node in the mesh are computed with the pressure-based coupled solver provided by Ansys Fluent, which simultaneously solves the mass and momentum continuity equations, Equations (7.8) and (7.9), with greater computational efficiency and convergence speed compared to the segregated pressure-based solver approach. To model the temporal dynamics, a transient solver method is proposed by adopting a first-order implicit formulation with a fixed 0.001 s time interval. Although higher order methods could have been employed to reduce the global error induced from the numerical method, the first-order implicit method was chosen to lessen the overall computation time of the process while preserving convergence criteria. Lastly, the Gauss-Seidel iterative method is utilized to solve algebraic sys-

tems by default.

Simulation Modeling

Several features are defined to reduce the degrees of freedom by specifying boundary conditions and fluid flow conditions. No-slip and zero diffusion flux boundary conditions are defined to the reactor walls for simplicity and surface of the substrate to resemble the self-limiting behavior of the ASALD process. Various *in silico* modeling works [99, 87, 84] have assumed laminar flow conditions, which are applicable in industry [27]. However, notable disadvantages [55] of laminar flow conditions are irrelevant to this study due to the substrate being small in diameter (200 mm, in Section 7.2.3) and the reactor design is modeled to limit depleted reagents along the perimeter of the reaction zones.

The most notable advantage of Ansys Fluent is its high versatility in the form of User-Defined Functions (UDFs). A UDF is a function that is able to enhance CFD simulations by customizing boundary conditions, defining material properties, and even initializing simulations, for instance. In this study, several UDFs are utilized to define mass flow rates and reagent mole fractions, source terms for species generation and consumption, temperature, the motion of the wafers with respect to the reactor, and allocating indexes in User-Defined Memories (UDMs). UDMs allow Ansys Fluent to find the appropriate nodes for computing reagent consumption and byproduct generation. Additionally, wafers are radially sliced into 10 equiangular sections so that mesoscopic simulations are executed in parallel, as shown in Figure 7.6. Thus, 10 mesoscopic simulations are conducted on each wafer, and the coverage on the 10 sections of a wafer from the mesoscopic simulation is computed as an area-weighted average. Given that 3 wafers are equipped in a spatial reactor, 30

UDMs must be applied to specify the nodes in which the consumption and generation of species take place.

Due to the versatility and advantages mentioned in Section 7.1 of numerical multiphysics modeling, computational fluid dynamics studies are broadly applied for industrial processes under various operating conditions. However, the computation efficiency (e.g., computational burden and time) heavily relies on computing resources. For example, to reduce numerical errors, high-order numerical methods with extremely fine meshes increase computing costs. Therefore, appropriate size of the mesh elements and numerical methods should be employed. In this study, 36-core dual processors with 384 GB of RAM (Random-Access Memory) are utilized to simulate the multiscale CFD model.
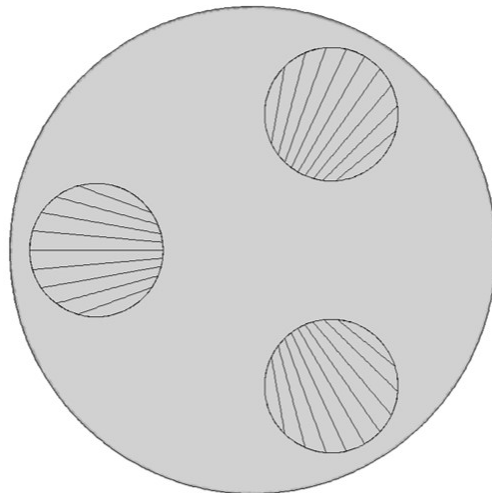
Figure 7.6: Three wafers on the rotating plate. Each wafer has 10 equiangular sections to evaluate 10 surface area-averaged pressures to evaluate an averaged surface coverage, and source generation and consumption rates.

7.3   Results and Discussion

7.3.1   Reactor Optimization

In the design of the reactors for ABC-ASALD processes, the most important criterion considered is the film uniformity, as it is directly related to the quality of semiconductor chips. A reactor with high film uniformity guarantees precise control over the thickness of deposition films when constructing 3D features on nanochips. When discussing film uniformity, there are several other factors that may be evaluated with a more rapid turnaround time than directly measuring film uniformity, such as the reagent exposure time distribution of the substrate and whether there is any intermixing present.

Exposure Time Distribution

The exposure time is defined as the amount of time that a section of the substrate is exposed to the reagents. Depending on reactor geometry and design, the exposure time may vary across the surface of the wafer. Thus, it is vital to examine the distribution of exposure times for various reactors and ensure a tight distribution of exposure times. For a spatial reactor, the exposure time for a specific point is determined not by the gas distribution device, but rather by the shape of the reaction zones where the substrate enters by the rotation of the bottom plate. To investigate the effect that reaction zone configurations have on the exposure time distribution, steady-state and dynamic simulations of **R1**, which has circular reaction zones, and **R4**, which has annular-sectioned reaction zones, were performed with a reactant mole fraction of 0.1 and with an angular velocity of 0.8 rad/s. The results are displayed in Figure 7.7.
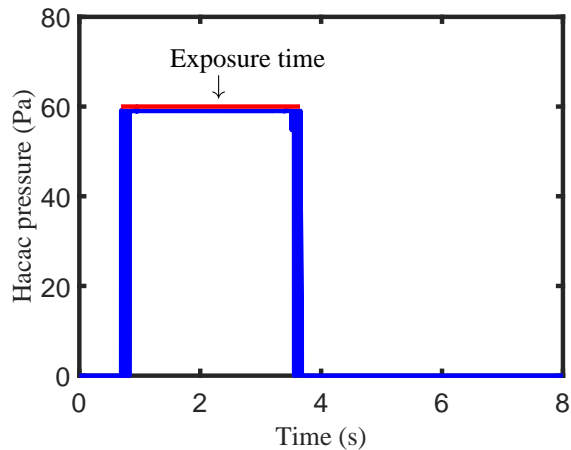
Figure 7.7: Hacac partial pressure over time on a node in the mesh. As the wafer rotates, the pressure rises around 0.8 s to reach the plateau and goes to 0 around 3.6 s. The length of the straight line in red is defined as the exposure time.

To evaluate the exposure time distribution for Hacac, the partial pressure of Hacac was recorded at every node of the meshed wafer surface as it traveled through the Hacac reaction zone. This data was then used to determine how long each wafer node spent saturated with Hacac, which is the exposure time. After being processed, all the nodal data was plotted in Figure 7.8c,d, and the wafer surface contours for all nodes are shown in Figure 7.8a,b. Results suggest that the geometry of the reaction zones for **R1** and **R4**, which are presented in Figure 7.5a,d, respectively, directly affect the distribution of nodal exposure time.

From a physical standpoint, the exposure time distribution is expected to exhibit low variance when all the nodes on the moving wafer experience the same amount of time in the reaction zone. However, the proposed circular spatial reactor differs from a linear sheet-to-sheet reactor in that the wafers are rotating in an angular fashion around the reactor, causing the edge of the wafer that is furthest from the center of the reactor to be traveling at a faster tangential velocity. Thus, to ensure that each point of the wafer spends the same amount of time in the reaction zone, the ideal

reaction zone shape must consist of arc lengths that are solely dependent on their radial distance from the center of the reactor. This condition is best fulfilled by the annular-sectioned reaction zones of reactor **R4**. For reactor **R1**, however, the central region of the substrate is expected to have longer exposure times compared to that of the outer region of the substrate due to the round reaction zone, which is easily demonstrated by examining the point on the wafer that is the furthest from the center of the reactor. This point will only be present in the reaction zone for a short time as the arc length of the reaction zone at that radius is small; thus, all sections of the wafer will not be exposed at the same time.

As illustrated in Figure 7.8c, the bar chart for **R1** shows that the reactor has a wide exposure time distribution, especially when compared to the tightly centered exposure time distribution of **R4**, whose bar chart is found in Figure 7.8d. While the average exposure time for both reactors is around 2.7 s, with **R1** having an average exposure time of 2.717 s and **R4** having an average exposure time of 2.701 s, the two vary greatly in terms of the standard deviation; the standard deviation of **R1** and **R4** were computed as 0.1925 and 0.0079, respectively. This large difference in standard deviation between the two reactors aligns with the predicted result, and it demonstrates that the standard deviation of the exposure times on the wafer is directly correlated to the shape of the reaction zones. To produce films with high uniformity, the reactor must have reaction zones shaped like annular sections as shown in **R4**.

Intermixing

The uniformity of the film produced by a spatial reactor is not solely dependent on the exposure time distribution and the shape of the reactor zones. Another crucial factor is the gas distribu-
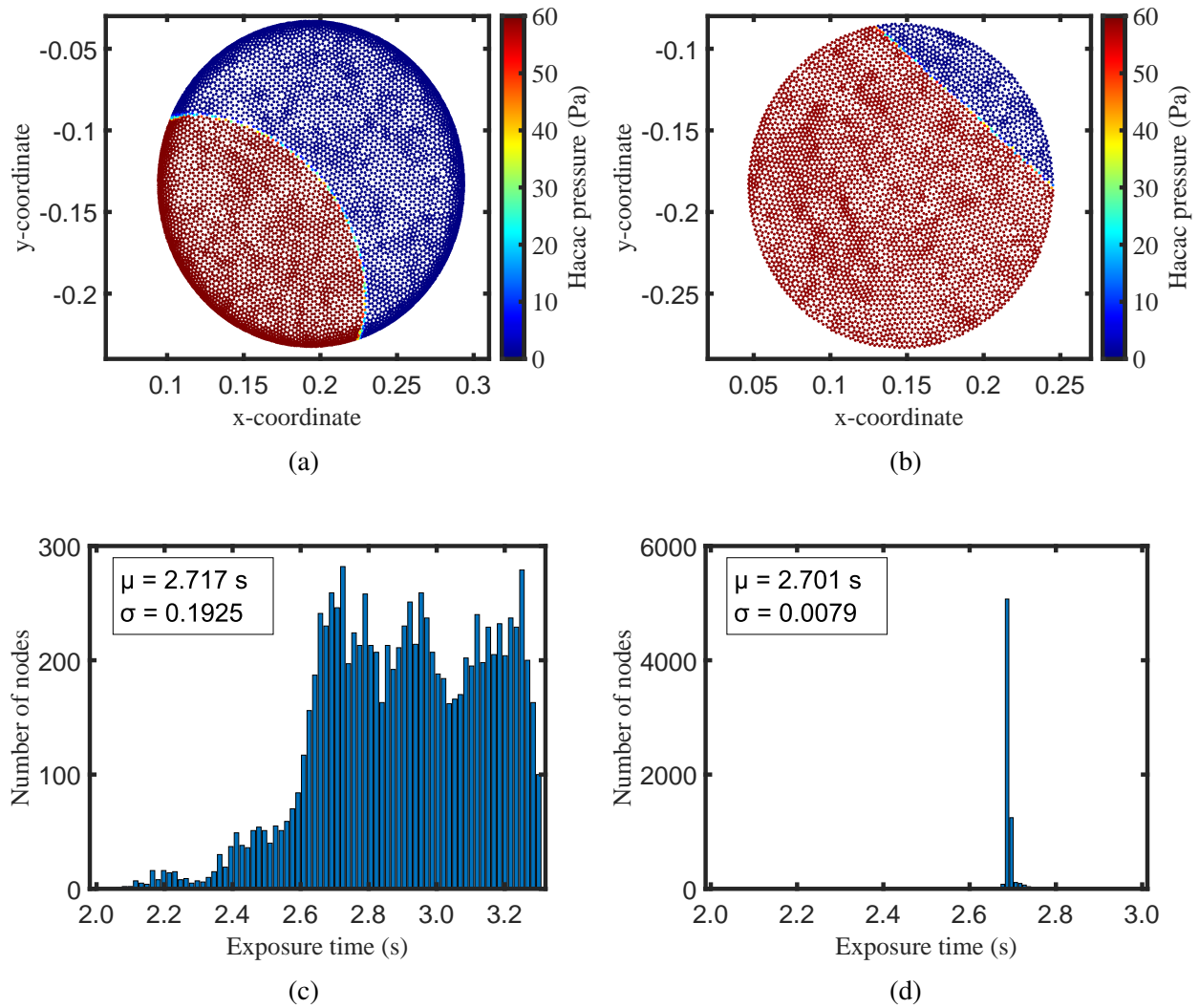
Figure 7.8: Hacac pressure contours across the wafer in (**a**) **R1** at 2 s and (**b**) **R4** at 3 s. The red region in (**a**,**b**) indicates the area that is covered with the reagent, Hacac. Distribution of exposure time for all nodes in the wafer for (**c**) **R1** and (**d**) **R4**. μ and σ represent the average and standard deviation, respectively. There are 10,410 nodes in **R1** and 6,737 nodes in **R4**.

tion across the reaction and purge zones. A nonuniform gas distribution causes the residence time

of the reagents across the wafer to vary, which prevents the reactor from achieving optimal surface

coverage and high film uniformity [27]. Inside spatial reactors, reagents are successfully separated

from the reaction zone by the injection of N_2 gas, thereby establishing a boundary layer that pre-

vents the diffusion of the reagents into the purging region by convection. As the gas flow inside

the reactor is assumed to be laminar in Section 7.2.3, nonuniform gas distribution mainly occurs when reactants cross into the purge zones. This phenomenon is named intermixing, and preventing it is vital to ensure uniform film growth. Intermixing is mainly caused by the substrate physically dragging unreacted reactants into the purge zone, which means that it is correlated to three features of reactor design: the gap distance, the wafer velocity, and the vacuum pressure. As intermixing is a binary criterion in that it is either present or is not, it is best to first determine the domain in which no intermixing is present before identifying other criteria such as fastest computational speed or fastest processing speed to select the optimal reactor design.

The gap distance is defined as the distance separating the wafer and the gas distribution assembly, and it is one of the most important factors in effectuating self-limiting behavior and minimizing reagent intermixing. In this work, the effects of gap distance on intermixing was studied with **R2** at different gap distances of 10 mm, 5 mm, and 1 mm. As shown by N_2 mole fraction contours in Figure 7.9, a lower gap distance inhibits reagent intermixing in the purging regions. This observation of how gap distance affects intermixing has been explored in several works for different spatial reactor configurations [18, 66].

However, despite lower gap distances reducing the intermixing of reagents, a higher number of discretized elements are needed to resolve boundary layers around separating chambers to ensure the accuracy of the numerical model, which carries greater computational burden. Thus, this study will account for a balance in the computational efficiency of the multiscale simulation and effective reagent separation. For the aforementioned reasons, a 5-mm gap distance is integrated into the idealized reactor configuration.

The wafer velocity also affects the intermixing of reagents. Greater angular velocities were

184
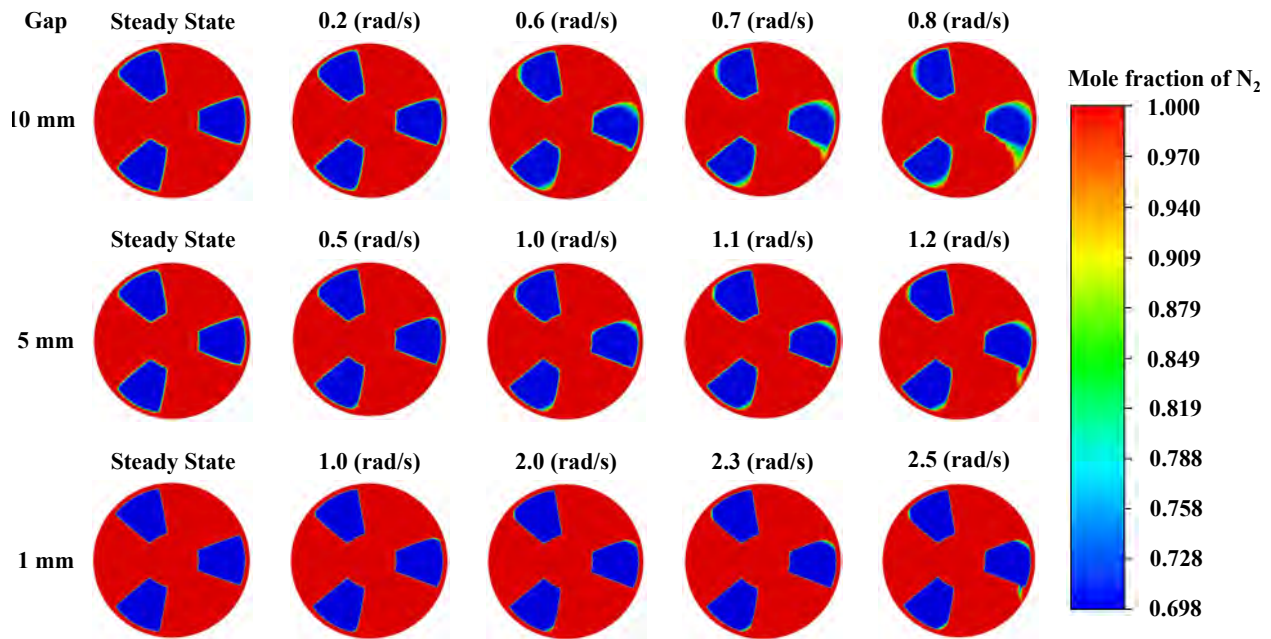
Figure 7.9: Contours of the mole fraction of nitrogen on the plate rotating clockwise in **R2** with gap distances of 10 mm, 5 mm, and 1 mm in various angular velocities, which depict the invasion of N_2 into the reaction zones and reagent runoff into the purge zones at higher rotation speeds.

observed to increase intermixing in response to the push of reagents in the direction of the angular rotation of the wafer. Both Figures 7.9 and 7.10 show how the N_2 mole fraction contours change with the angular wafer velocity. Figure 7.10 in particular, reveals that the invasion of the inert gas (denoted by the ring in orange) at the entry of the reaction zone and the reagent runoff (denoted by the ring in green) at the exit are both conveyed in the direction of the wafer rotation (counter-clockwise). Thus, these instances of intermixing are attributed to the motion of the wafer. As shown in Figure 7.9, with a 5-mm gap distance, the inert gas of N_2 is observed to enter the reaction zone when the angular wafer velocity is above 1.0 rad/s. The reagent enters the purge zone when the angular wafer velocity is above 1.2 rad/s. Thus, to maintain a self-limiting reaction with no

intermixing, the rotation speed must be kept below 1.0 rad/s. Thus, it is imperative to construct a reactor configuration that handles sufficiently high angular velocities to reduce process time while simultaneously producing uniform deposition and promoting effective reagent separation.

To establish effective gas separation and an acceptable process time, it is essential to choose an appropriate rotation speed and gap distance to achieve precise thin film thickness control and surface uniformity. While lower gap distances minimize the effects of reagent intermixing, several operating and reactor design conditions are also considered for minimizing the intermixing of reagents. Thus, an effective balance between the gap distance, (selected to be 5 mm), the vacuum pressure (chosen to be -150 Pa), and the orientation (symmetrical or asymmetrical) of the injection ports are methodically chosen to enable a higher range of substrate angular velocities to be used for the process. For instance, the reactor model **R3** repositions the cylindrical injection ports closer to the substrate entry in the reaction zone as shown in Figure 7.5c, which establishes an asymmetrical inlet configuration. The mole fraction contours of N_2 for reactor model **R3** in Figure 7.10 shows marginal improvement of intermixing, particularly in the reagent runoff from the reaction zone at 1.0 rad/s as opposed to that of **R2**, which was modeled with a symmetrical inlet arrangement. **R4** (as shown in Figure 7.5d), with annular sector-shaped reagent inlets, was proposed to further mitigate intermixing as shown in Figure 7.10. To be specific, the invasion of the inert gas into the reaction zone disappears at both 0.8 and 1.0 rad/s in contrast with **R2** and **R3**, which both still show signs of intermixing at those velocities. Therefore, **R4** is selected as the desired reactor configuration to explore the ASALD process to discuss the quantitative relationship between various operating conditions on the surface coverage in Section 7.3.2.
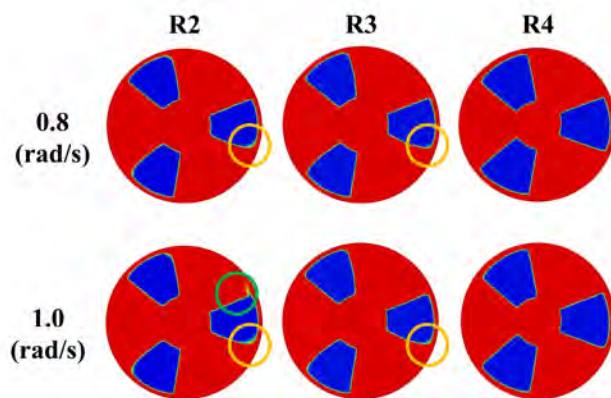
Figure 7.10: Contours of the mole fraction of nitrogen on the plate rotating clockwise in **R2**, **R3** and **R4** with a gap distance of 5 mm in the angular velocities of 0.8 and 1.0 *rad/s*. The legend for the contours is present in Figure 7.9. A ring in orange denotes a nitrogen invasion into the BDEAS reaction zone, and a ring in green denotes the BDEAS runoff from the reaction zone.

7.3.2   Multiscale Simulation Results

The reactor optimization study examines the impact that the gap distance, the injection and outlet port geometry, and the wafer angular velocity have on reagent separation, exposure time distribution, reagent distribution on the substrate surface, and computational efficiency. After considering all the above criteria, reactor model **R4** was designed with a 5-mm gap distance. The next component of this work will examine the effects that operating conditions such as wafer velocity and reagent mole fractions have on the surface coverage for Steps A, B, and C.

The simulations were conducted with a range of rotation speeds, ω, (0.2 rad/s $\leq \omega \leq$ 0.8 rad/s) and a range of reactant species mole fraction, x (0.05 $\leq x \leq$ 0.70) to determine optimal operating conditions to achieve complete surface coverage. For all steps, the surface coverage depends on the rotation speeds and the mole fractions of the reactant species, as illustrated in Figure 7.11. In general, the coverage for each step increases as the mole fraction of the reactant species increases and as the rotation speed decreases. An increase in the mole fraction of the reactant species naturally

187

leads to an increase in the partial pressure of the species, which accelerates the reaction rates for all adsorption reactions according to Collision Theory. Furthermore, a lower rotation speed allows the wafer to be exposed to the reagent for a longer period of time, thus increasing the surface coverage. Specifically, the Hacac step reaches full coverage at 0.2 rad/s with Hacac mole fractions of x_{Hacac} = 0.05, 0.1, and 0.3. However, the coverage decreases with increasing rotation speed. At 0.6 rad/s, full coverage is only achieved with x_{Hacac} = 0.3; for lower Hacac mole fractions, surface coverages of 98.8% and 95.1% were observed. BDEAS and ozone coverage curves also reveal a similar trend. However, the curves for BDEAS appear discontinuous, which is attributed to its proposed kinetic mechanism. An observable limitation of the kMC method is that the algorithm requires a priori knowledge of a reaction mechanism [3] that, if exceedingly simplified, will reduce the accuracy of the kMC model when trying to predict realistic surface kinetics. Figure 7.11b exemplifies a noncompetitive BDEAS adsorption in which desorption and intermediate pathways are not immediately observed after the adsorption reaction takes place. This work utilized a kinetics mechanism determined by [73] with large activation energy differences that were evaluated using the nudged elastic band method by [115]. Such reaction mechanism introduces a spontaneous reaction path where the surface coverage is dictated by the large reaction rate constant differences that generate this noncompetitive kinetics mechanism, where one reaction path dominates. Additionally, the repulsive forces due to the steric effects reduced the versatility of pursuing intermediate or reverse reaction paths, therefore, producing steeper coverage curves.

Noncompetitive kinetics offer insight to the discontinuous curves presented in Figure 7.11; however, the mole fraction gradient along the boundary between the purging and reaction zone presents a challenge for collecting pertinent surface pressure data that will capture the initial ad-
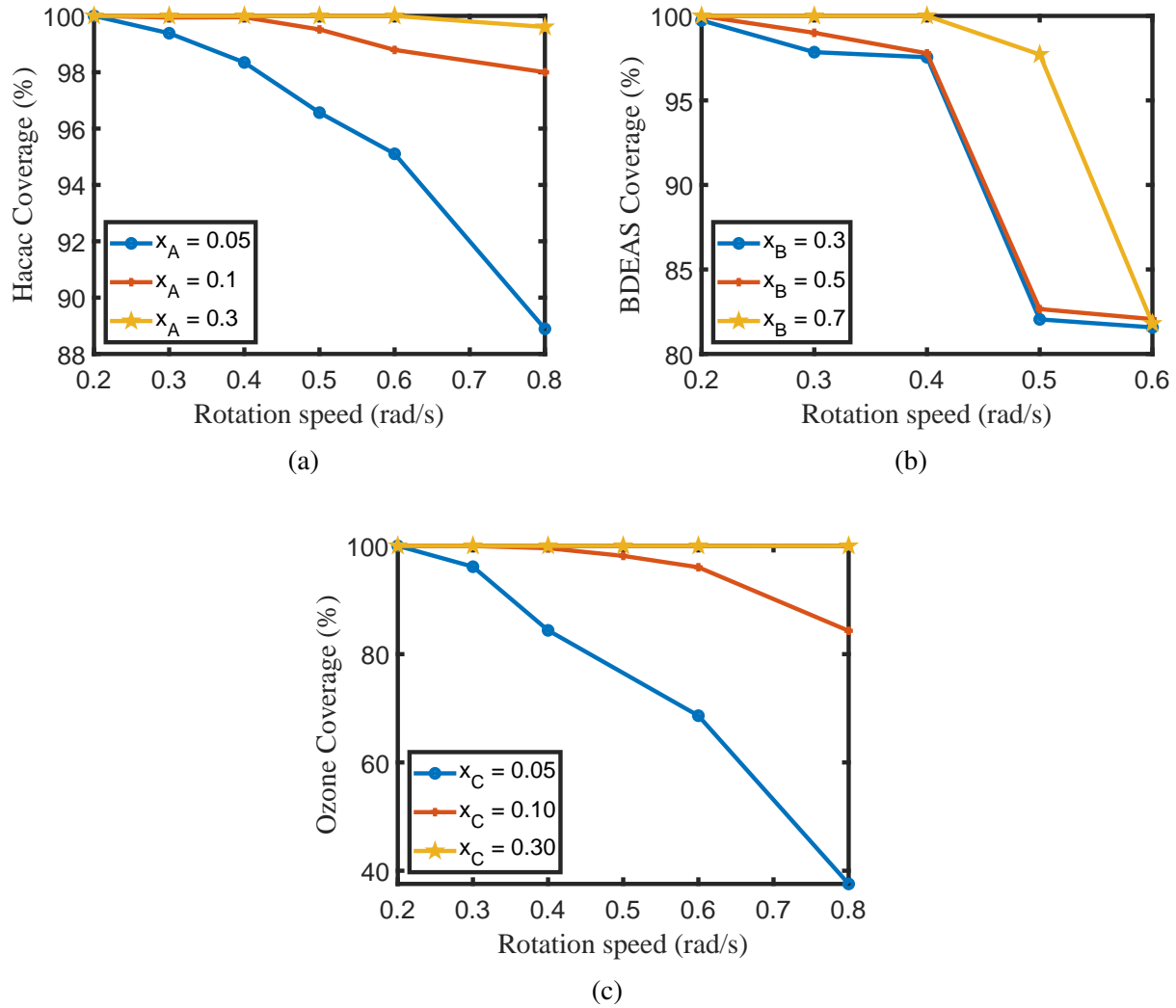
Figure 7.11: Coverage (%) versus rotation speed (*rad/s*) for (**a**) Hacac step, (**b**) BDEAS step, and (**c**) ozone step. x_A, x_B, and x_C represents the mole fraction of Hacac, BDEAS, and ozone, respectively.

sorption step for smaller time and length scales. This concentration gradient is presented in Figure 7.12, which illustrates the lack of nodal data due to the small length scales required to collect additional pressure data. Additionally, at higher angular velocities, the collection of nodal data becomes more challenging due to this small boundary length scale (in μm), which is another reason for the discontinuous curves generated in Figure 7.11. Consequently, the smaller length scale

in micrometers would require smaller time scales (milliseconds) to collect surface coverage data within this boundary layer, as illustrated in Figure 7.2. Thus, the spatiotemporal progression of coverage of the wafer surface would be more difficult to simulate at the reaction zone entrances.
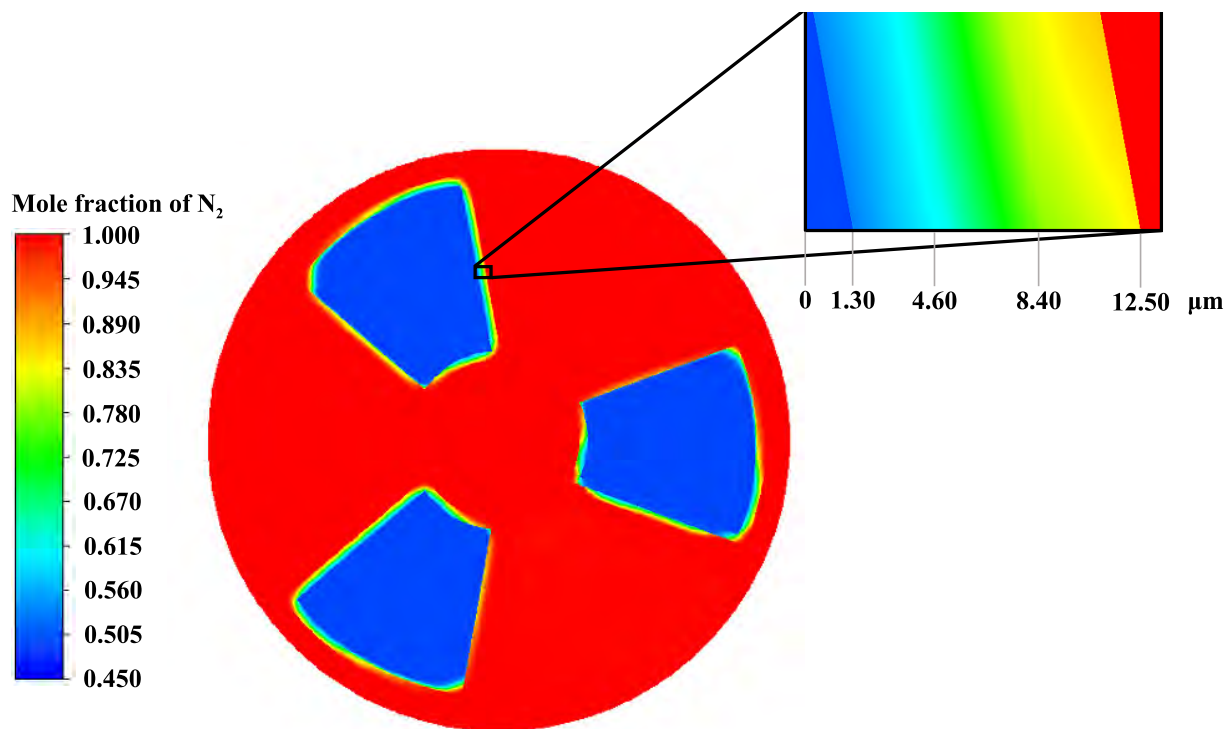
Figure 7.12: Concentration gradient of N_2 mole fraction for **R4** illustrating the small length scales in micrometers for each contour along the reagent-purge boundary region, which limits the availability for studying the temporal progression of surface coverage as depicted in Figure 7.2.

When comparing the multiscale CFD simulation process times to reach full coverage compared to that of the mesoscopic model from prior work [115], the multiscale CFD simulation process times were slower. The integration of the macroscopic model to the mesoscopic model introduced pressure depletion zones that diluted the surface exposure to reagents, thereby reducing the adsorption reaction rates for Hacac and BDEAS. For a constant temperature of 523 K and constant Hacac surface pressure of 130 Pa (Step A), a constant BDEAS surface pressure of 400 Pa (Step B), and a constant O_3 pressure of 130 Pa (Step C), the mesoscopic simulation evaluated

complete coverage times of 2.29 s, 2.28 s, and 3.81 s, respectively [115]. The multiscale CFD

simulation results with similar reactor operating conditions require process times of 3.77 s, 6.99 s,

and 3.72 s to achieve full coverage for Steps A, B, and C, respectively as illustrated in Figure 7.13.

The integration of macroscopic fluid behavior with the mesoscopic model indicates that the reagent

depletion zones and repulsions generated by species intercollisions in the fluid phase contribute to

the substantial increase in the process times, which was observed from a prior multiscale model-

ing work [120]. However, the proposed model, which partitions the wafer into 10 sections and

performs 10 kMC simulations on each section, inaccurately characterizes the surface coverage

progression as a function of time. Ideally, a kMC simulation for each node on the wafer surface

would greatly improve the accuracy of the times to obtain complete surface coverage, but comes at

a computational cost. However, the results are consistent with experimental findings where Hacac

dosage times of 5 s in similar operating conditions were used to ensure complete surface cover-

age [71]. Additionally, BDEAS dosage times of 6 s were employed to maximize the deposition

thickness of the SiO_2 surface [73], although, greater dosage times were observed to contribute

to the nucleation of the oxide film if BDEAS is oversaturated in the reactor. To prevent such a

scenario, the rotary reactor design evacuates unreacted reagent through the purge streams, thereby

maintaining the operating pressure of the reagent within the reaction zone. Lastly, O_3 dosage times

of 2.5 s to 20 s to achieve complete surface coverage [59], which is analogous to the multiscale

computed time.

Figure 7.14 illustrates the contours of species mole fractions simulated with a constant ro-

tation speed of 0.4 rad/s and a mole fraction of 0.3 for Hacac, BDEAS, and ozone. Due to the

fixed rotation speed imposed to the three wafers, the equal areas of each wafer are exposed to each
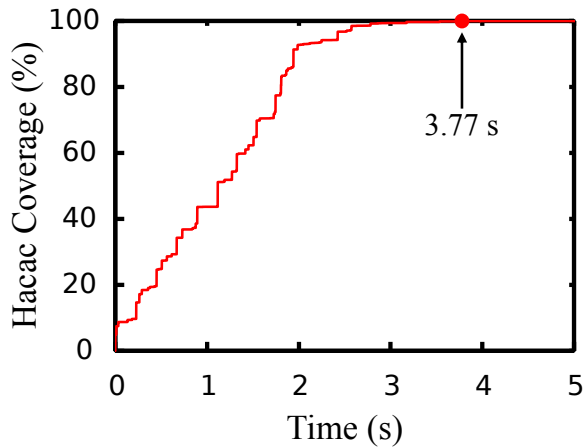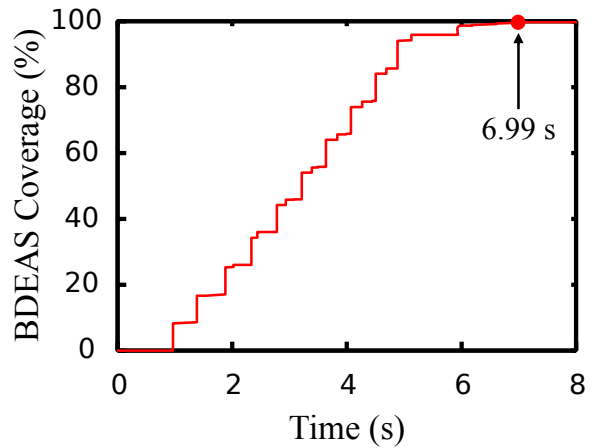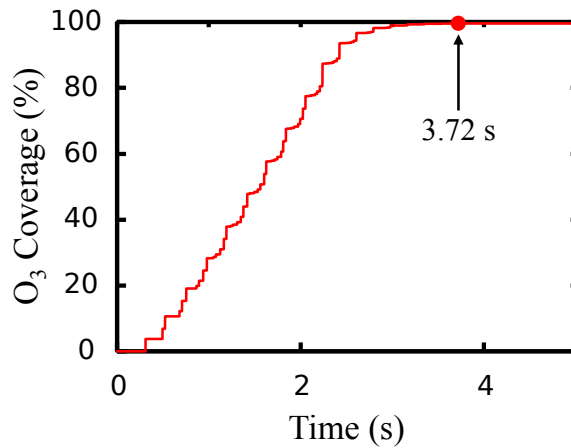
Figure 7.13: Multiscale CFD plots illustrating the surface coverage (%) progression as a function of time for (**a**) 0.1 mole fraction Hacac adsorption, (**b**) 0.5 mole fraction BDEAS adsorption, (**c**) 0.1 mole fraction O_3 adsorption. Evaluated multiscale process times to reach full surface coverage were longer than observed mesoscopic model process times conducted from prior work [115].

reagent as shown in Figure 7.14a. However, the reaction progression for each reaction, which is inspected by measuring the generation of byproducts from the wafer surface, is different as revealed in Figure 7.14b–d. As discussed above, BDEAS adsorption requires higher mole fraction and lower rotation speeds to achieve complete coverage compared with Hacac and ozone adsorp-

tion. Despite the fact that more than half of the wafer area is in the BDEAS-enriched reaction zone (in Figure 7.14c), the byproduct, DEA, is only observed in a small area, which is the area that first encountered the reagent due to the BDEAS adsorption mechanism, and the steric effects discussed in Section 7.2.2. Meanwhile, H_2O produced from Hacac adsorption and O_2 produced from O_3 adsorption are being generated on the region closer to the boundary between the reagent and the inert gas. In other words, Step A and Step C are faster than Step B, which is in agreement with *ab initio* DFT calculation results [115, 95, 29] where the BDEAS adsorption undergoes multiple steps that have higher activation energies than Hacac and ozone. Therefore, the mole fractions of reactants must be carefully selected to optimize the process. When selecting operating conditions, the BDEAS and Ozone steps can be firstly considered since Hacac is covered on the non-growth area at the first cycle and afterward, only the deprived reaction sites are refilled and reinforced. From the multiscale simulations, the mole fractions of 0.1, 0.5, and 0.1 for Hacac, BDEAS, and O_3, respectively, and a rotation speed of 0.2 rad/s can be considered as an optimal operating condition for full coverage.

7.4   Conclusions

An *in silico* multiscale computational fluid dynamics model was developed for an area-selective atomic layer deposition (ASALD) process on the substrate of SiO_2/Al_2O_3 utilizing a spatial reactor design. Three reagents were used to deposit SiO_2 thin films on the growth area (SiO_2) and to create protective layers on the nongrowth area (Al_2O_3): the inhibitor (acetylacetone or Hacac), precursor (bis(diethylamino)silane or BDEAS), and oxidant (ozone) for Steps A
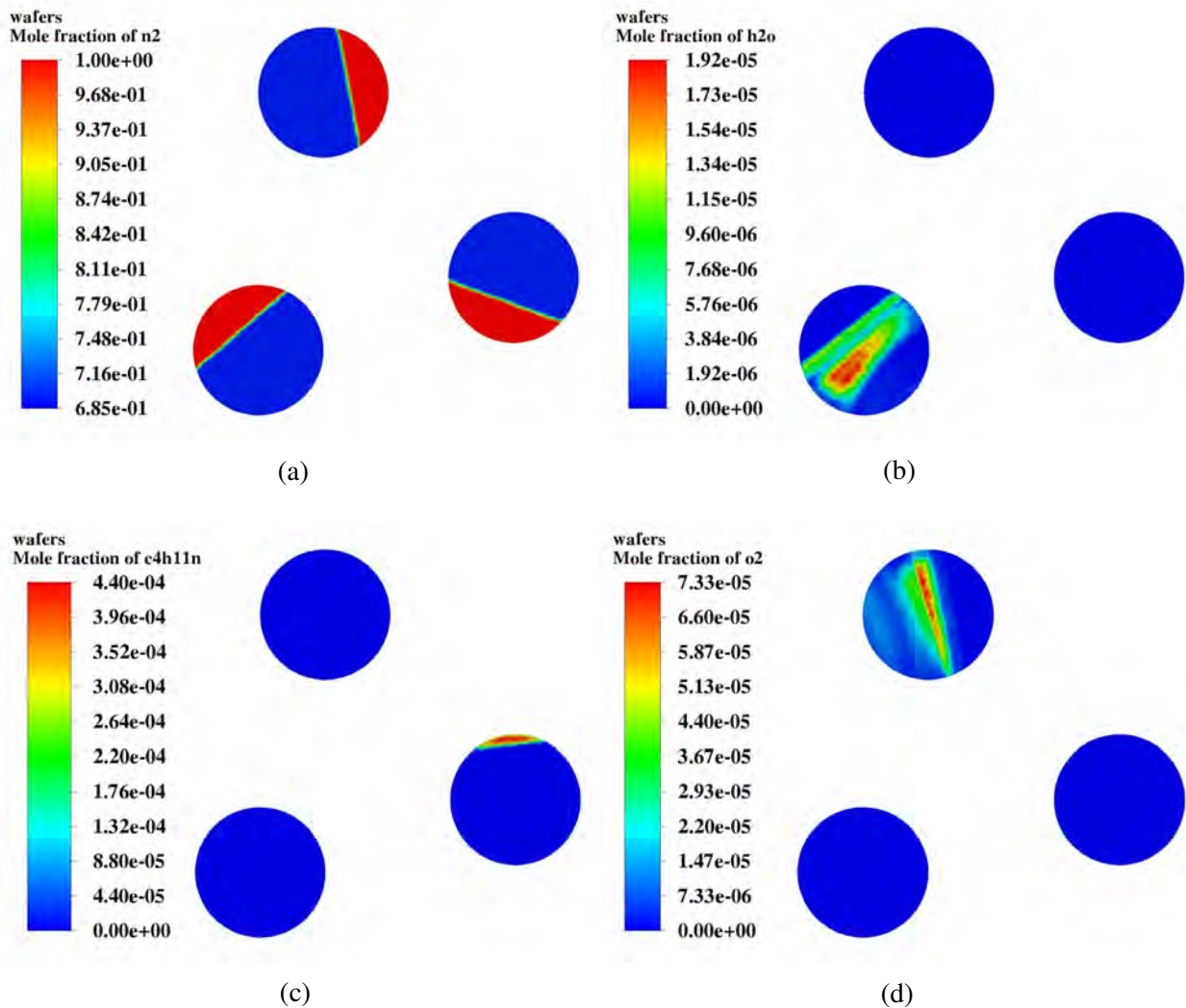
Figure 7.14: Contours of (**a**) Nitrogen mole fraction, (**b**) H_2O mole fraction, (**c**) DEA mole fraction, and (**d**) O_2 mole fraction on the wafers at 2 s with 0.4 rad/s From bottom left counter-clockwise, the wafers in the Hacac, BDEAS, and ozone reaction zones, respectively.

through C, respectively. This model included an atomistic-mesoscopic model using kinetic Monte Carlo (kMC) and a macroscopic model using computational fluid dynamics (CFD). Foremost, a spatial reactor configuration was optimized to ensure uniform reagent exposure time distribution and effective reagent separation to ensure self-limiting behavior was maintained. The gap distance, rotation speed, and reactor geometries were observed to have great influence on transport

behavior of gas species. Lower gap distances and rotational speeds are required to avoid reagent intermixing. In addition, a ring, sector-shaped outlet and N_2 inlet improved reagent separation by removing the reagent runoff and inert invasion; thus creating a low variance gas distribution, leading to uniform exposure to reagents. A more uniform flow was observed with asymmetrical, annular, sector-shaped reagent inlets. After an optimal reactor configuration was built, the multi-scale model was simulated within the operating range of 0.2 rad/s $\leq \omega \leq 0.8$ rad/s with different mole fractions of species for Hacac, BDEAS, and ozone. Sufficiently low angular speeds were required for complete surface deposition by providing longer exposure time to the reagents. A higher mole fraction of species also accelerated the adsorption reactions to reach complete coverage.

It is essential to choose an appropriate mole fraction of each reagent, since all reaction zones rotate at the same velocity. Since the BDEAS reaction takes longer time, a higher mole fraction for BDEAS than the ones for Hacac and ozone is required to obtain full coverage at the same wafer rotation speed. Due to the advantage of this modeling study that can generate a broad set of data at varying operating conditions, this *in silico* work has the potential to expedite reaction scale-up in industry. Therefore, this modeling approach offers insights and data to process engineers to select proper operating conditions for further process development. On top of that, this study presents guidelines to design and optimize a rotary type spatial reactor in an industrial setting. Finally, as described in Section 7.1, further research is needed to design an optimal control system for robust process operation. Specifically, given the complexity of the reaction taking place inside these reactors, advanced control methods for ASALD (e.g., model predictive control, artificial neural network-based run-to-run control) can be developed in a future work to ensure robust process operation.

Chapter 8

Conclusions

This dissertation discusses thermal atomic layer etching (tALE) and area-selective atomic layer deposition (ASALD) with various reactor designs using a multiscale modeling methodology. Those processes were successfully characterized at the atomistic level in conjunction with computational fluid dynamics. Furthermore, run-to-run (R2R) control strategies using various algorithms and regression models were developed to modify process inputs (the so-called "recipe") between batches, thereby eliminating the effects of external disturbances and process drifts.

Chapter 2 described the surface modeling of thermal atomic layer etching on aluminum oxide thin films using hydrogen fluoride (HF) and trimethylaluminum (TMA) as reagents. The kinetic Monte Carlo (kMC) method was used to simulate surface reactions. Conjointly, density functional theory (DFT) calculations were performed to optimize the surface structures with the lowest energies and to find kinetic parameters to facilitate kMC simulations. To find feasible operating conditions based on the industrial half-cycle time, a feedforward artificial neural network (FNN) model was also formulated.

Chapter 3 illustrated multiscale modeling by integrating the surface model, described in **Chapter 2**, with a macroscopic model using a multiphysics software. Four different types of reactors were constructed with different gas distributors to achieve uniform flow profiles. This study revealed that the reactor with the inclined-plate showed the best performance in terms of economics and film uniformity. This approach provided comprehensive understanding for the thermal atomic layer etching process and invaluable insight for reactor design.

Chapter 4 discussed an optimal multivariable run-to-run (R2R) control system for the thermal ALE process using the multiscale model described in **Chapter 3**. The multivariable R2R control was constructed using a multi-input, multi-output (MIMO) model, and its performance was compared with individual R2R control systems. The multivariable R2R control approach perfectly removed the process disturbances.

Chapter 5 investigated spatial atomic layer etching (SALE) using the microscopic surface model discussed in **Chapter 2**. A sheet-to-sheet reactor configuration was developed to study different process conditions and parameters that have impacts on thermal ALE and film conformity in the reactor design. This research is expected to provide a technical guide to industry for further industrialization of spatial atomic layer processes.

Chapter 6 elucidated an atomistic-mesoscopic model for area-selective atomic layer deposition (ASALD) using acetylacetone (Hacac), bis(diethylamino)silane (BDEAS), and ozone (O_3) as a small-molecule inhibitor (SMI), precursor, and oxidant, respectively. The kinetic mechanism of ALD nulceation on silicon oxide (growth areas) and the mechanism for creating protective layers using the inhibitor on aluminum oxide (nongrowth areas) were modeled with a kinetic Monte Carlo (kMC) method considering steric hindrance aided with quantum mechanics. This atomistic-

mesoscopic model characterized regioselective and chemoselective behaviors in depth for ASALD.

Chapter 7 describes a multiscale modeling for ASALD integrated with the atomistic-mesoscopic model in **Chapter 6** using a rotary type spatial reactor design. First, the reactor configuration was optimized to maintain self-limiting nature for film homogeneity. Then, the ASALD process on the substrate of SiO_2/Al_2O_3 with the optimized reactor was simulated with different reagent mole fractions and angular velocities to examine the effects of the operating conditions on film deposition. The developed multiscale model successfully captured the ASALD process and provided the impacts of the operating conditions on film coverage. This multiscale model can be employed to present an appropriate control system for future work.

Bibliography

[1] A. I. Abdulagatov and S. M. George. Thermal atomic layer etching of silicon using O_2, HF, and $Al(CH_3)_3$ as the reactants. *Chemistry of Materials*, 30:8465–8475, 2018.

[2] M. Aghaee, P. S. Maydannik, P. Johansson, J. Kuusipalo, M. Creatore, T. Homola, and D. C. Cameron. Low temperature temporal and spatial atomic layer deposition of TiO_2 films. *Journal of Vacuum Science & Technology A*, 33:041512, 2015.

[3] W. Andreoni and S. Yip. *Handbook of Materials Modeling: Methods: Theory and Modeling*. Springer International Publishing, Cham, 2 edition, 2020.

[4] ANSYS. *Ansys Fluent Theory Guide*. ANSYS Inc., Canonsburg, PA, 2021.

[5] C. Bae, H. Shin, and K. Nielsch. Surface modification and fabrication of 3D nanostructures by atomic layer deposition. *MRS Bulletin*, 36:887–897, 2011.

[6] S. Baroni, P. Giannozzi, and E. Isaev. Density-functional perturbation theory for quasi-harmonic calculations. *Reviews in Mineralogy & Geochemistry*, 71:39–58, 2010.

[7] S. Baroni, P. Giannozzi, and E. Isaev. Thermal properties of materials from ab initio quasi-harmonic phonons. *arXiv preprint arXiv:1112.4977*, 2011.

[8] S. Baroni, P. Giannozzi, and E. Isaev. Thermal properties of materials from ab initio quasi-harmonic phonons. *arXiv preprint arXiv:1112.4977*, 2011.

[9] S. W. Benson and A. E. Axworthy. Mechanism of the gas phase, thermal decomposition of ozone. *The Journal of Chemical Physics*, 26:1718–1726, 1957.

[10] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for monte carlo simulation of ising spin systems. *Journal of Computational Physics*, 17:10–18, 1975.

[11] G. E. P. Box. Evolutionary operation: a method for increasing industrial productivity. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 6:81–101, 1957.

[12] M. Broas, O. Kanninen, V. Vuorinen, M. Tilli, and M. Paulasto-Kröckel. Chemically stable atomic-layer-deposited Al_2O_3 films for processability. *ACS Omega*, 2:3390–3398, 2017.

[13] W. Campbell, S. Firth, A. Toprac, and T. Edgar. A comparison of run-to-run control algorithms. In *Proceedings of the 2002 American Control Conference*, volume 3, pages 2150–2155, Anchorage, AK, 2002.

[14] R. Chen, H. Kim, P. C. McIntyre, D. W. Porter, and S. F. Bent. Achieving area-selective atomic layer deposition on patterned substrates by selective surface modification. *Applied Physics Letters*, 86:191910, 2005.

[15] C.-F. Chien, Y.-J. Chen, C.-Y. Hsu, and H.-K. Wang. Overlay error compensation using advanced process control with dynamically adjusted proportional-integral R2R controller. *IEEE Transactions on Automation Science and Engineering*, 11:473–484, 2014.

[16] T.-C. Chou. Derivation and properties of michaelis-menten type and hill type equations for reference ligands. *Journal of Theoretical Biology*, 59:253–276, 1976.

[17] T.-C. Chou. The mass-action law based algorithms for quantitative econo-green bio-research. *Integrative Biology*, 3:548–559, 2011.

[18] W. Cong, Z. Li, K. Cao, G. Feng, and R. Chen. Transient analysis and process optimization of the spatial atomic layer deposition using the dynamic mesh method. *Chemical Engineering Science*, 217:115513, 2020.

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

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

[21] M. G. Crose, W. Zhang, A. Tran, and 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:e16400, 2019.

[22] C. M. De la Huerta, V. H. Nguyen, J.-M. Dedulle, D. Bellet, C. Jiménez, and D. Muñoz-Rojas. Influence of the geometric parameters on the deposition mode in spatial atomic layer deposition: A novel approach to area-selective deposition. *Coatings*, 9:5, 2018.

[23] Z. Deng, W. He, C. Duan, R. Chen, and B. Shan. Mechanistic modeling study on pro-

cess optimization and precursor utilization with atmospheric spatial atomic layer deposition. *Journal of Vacuum Science & Technology A*, 34:01A108, 2016.

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

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

[26] DIPPR. *DIPPR Project 801*. Design Institute for Physical Properties, AIChE, New York, NY, 2020.

[27] K.-E. Elers, T. Blomberg, M. Peussa, B. Aitchison, S. Haukka, and S. Marcus. Film uniformity in atomic layer deposition. *Chemical Vapor Deposition*, 12:13–24, 2006.

[28] S.-K. S. Fan, B. C. Jiang, C.-H. Jen, and C.-C. Wang. SISO run-to-run feedback controller using triple EWMA smoothing for semiconductor manufacturing processes. *International Journal of Production Research*, 40:3093–3120, 2002.

[29] G. Fang, L. Xu, J. Ma, and A. Li. Theoretical understanding of the reaction mechanism of SiO_2 atomic layer deposition. *Chemistry of Materials*, 28:1247–1255, 2016.

[30] M. Fang and J. C. Ho. Area-selective atomic layer deposition: Conformal coating, sub-nanometer thickness control, and smart positioning. *ACS Nano*, 9:8651–8654, 2015.

[31] T. Faraz, F. Roozeboom, H. C. M. Knoops, and W. M. M. Kessels. Atomic layer etching: what can we learn from atomic layer deposition? *ECS Journal of Solid State Science and Technology*, 4:N5023–N5032, 2015.

[32] M. M. Folkendt, B. E. Weiss-Lopez, J. P. C. Jr., and N. S. True. Gas-phase proton nmr studies of keto-enol tautomerism of acetylacetone, methyl acetoacetate, and ethyl acetoacetate. *The Journal of Physical Chemistry*, 89:3347–3352, 1985.

[33] H. Fontaine, M. Veillerot, and A. Danel. Deposition behavior of volatile acidic contaminants on metallic interconnect surfaces. *Material Science*, 103-104:365–368, 2012.

[34] D. C. Freeman, D. H. Levy, and P. J. Cowdery-Corvan. Method for producing compound thin films, May 3 2010. US Patent 7,858,144 B2.

[35] K. Fu, Y. Fu, P. Han, Y. Zhang, and R. Zhang. Kinetic monte carlo study of metal organic chemical vapor deposition growth dynamics of GaN thin film at microscopic level. *Journal of Applied Physics*, 103:103524, 2008.

[36] J. C. Gamekkanda, A. S. Sinha, J. Desper, M. Đaković, and C. B. Aakeröy. The role of halogen bonding in controlling assembly and organization of Cu(II)-Acac based coordination complexes. *Crystals*, 7:226, 2017.

[37] S. M. George. Atomic layer deposition: An overview. *Chemical Reviews*, 110:111–131, 2010.

[38] S. M. George. Mechanisms of thermal atomic layer etching. *Accounts of Chemical Research*, 53:1151–1160, 2020.

[39] P. Giannozzi. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21:395502, 2009.

[40] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976. ISSN 0021-9991.

[41] E. Granneman, P. Fischer, D. Pierreux, H. Terhorst, and P. Zagwijn. Batch ALD: Characteristics, comparison with single wafer ALD, and examples. *Surface and Coatings Technology*, 201:22–23, 2007.

[42] A. Haider, M. Yilmaz, P. Deminskyi, H. Eren, and N. Biyikli. Nanoscale selective area atomic layer deposition of TiO_2 using e-beam patterned polymers. *RSC Adv.*, 6:106109–106119, 04 2016.

[43] R. G. Hobbs, N. Petkov, and J. D. Holmes. Semiconductor nanowire fabrication by bottom-up and top-down paradigms. *Chemistry of Materials*, 24:1975–1991, 2012.

[44] M. Hu, T. C. Hauger, B. C. Olsen, E. J. Luber, and J. M. Buriak. UV-initiated Si–S, Si–Se, and Si–Te bond formation on Si(111): Coverage, mechanism, and electronics. *The Journal of Physical Chemistry C*, 122:13803–13814, 2018.

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

[46] J. Huang, M. Lee, A. Lucero, L. Cheng, and J. Kim. Area-selective ald of TiO$_2$ nanolines with electron-beam lithography. *The Journal of Physical Chemistry C*, 118:23306–23312, 2014.

[47] K. J. Hughes, A. Dube, M. Sharma, and J. R. Engstrom. Initial stages of atomic layer deposition of tantalum nitride on SiO$_2$ and porous low-κ substrates modified by a branched interfacial organic layer: Chemisorption and the transition to steady-state growth. *The Journal of Physical Chemistry C*, 116:21948–21960, 2012.

[48] A. P. J. Jansen. *An Introduction to Kinetic Monte Carlo Simulations of Surface Reactions*, volume 1. Academic Press, 2012.

[49] L. Jin, Y. Li, Z. Hu, and J. Chu. Full three-dimensional morphology evolution of amorphous thin films for atomic layer deposition. *AIP Advances*, 8:045304, 2018.

[50] H. Jónsson, G. Mills, and K. W. Jacobsen. Classical and quantum dynamics in condensed phase simulations. In B. J. Berne, G. Ciccotti, and D. F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 385–404. World Scientific, 1998.

[51] M. Jurczak, N. Collaert, A. Veloso, T. Hoffmann, and S. Biesemans. Review of FINFET technology. *2009 IEEE International SOI Conference, Foster City, CA, USA*, pages 1–4, 2009.

[52] K. J. Kanarik, T. Lill, E. A. Hudson, S. Sriraman, S. Tan, J. Marks, V. Vahedi, and R. A.

Gottscho. Overview of atomic layer etching in the semiconductor industry. *Journal of Vacuum Science & Technology A*, 33:020802, 2015.

[53] K. J. Kanarik, S. Tan, and R. A. Gottscho. Atomic layer etching: rethinking the art of etch. *Journal of Physical Chemistry Letters*, 9:4814–4821, 2018.

[54] H. G. Kim, M. Kim, B. Gu, M. R. Khan, B. G. Ko, S. Yasmeen, C. S. Kim, S.-H. Kwon, J. Kim, J. Kwon, K. Jin, B. Cho, J. S. Chun, B. Shong, and H.-B.-R. Lee. Effects of Al precursors on deposition selectivity of atomic layer deposition of Al_2O_3 using ethanethiol inhibitor. *Chemistry of Materials*, 32:8921–8929, 2020.

[55] J. Kim, K. Chakrabarti, J. Lee, K.-Y. Oh, and C. Lee. Effects of ozone as an oxygen source on the properties of the Al_2O_3 thin films prepared by atomic layer deposition. *Materials Chemistry and Physics*, 78:733–738, 2003.

[56] K. S. Kim, K. H. Kim, Y. Nam, J. Jeon, S. Yim, E. Singh, J. Y. Lee, S. J. Lee, Y. S. Jung, G. Y. Yeom, and D. W. Kim. Atomic layer etching mechanism of MoS_2 for nanodevices. *ACS Applied Materials & Interfaces*, 9:11967–11976, 2017.

[57] J. S.-I. Kwon, M. Nayhouse, and P. D. Christofides. Multiscale, multidomain modeling and parallel computation: Application to crystal shape evolution in crystallization. *Industrial & Engineering Chemistry Research*, 54:11903–11914, 2015.

[58] G. Lee, B. Lee, J. Kim, and K. Cho. Ozone adsorption on graphene: Ab initio study and experimental validation. *The Journal of Physical Chemistry C*, 113:14225–14229, 2009.

[59] J.-M. Lee, J. Lee, H. Oh, J. Kim, B. Shong, T. J. Park, and W.-H. Kim. Inhibitor-free area-selective atomic layer deposition of SiO_2 through chemoselective adsorption of an amino-disilane precursor on oxide versus nitride substrates. *Applied Surface Science*, 589:152939, 2022.

[60] Y. Lee and S. M. George. Atomic layer etching of Al_2O_3 using sequential, self-limiting thermal reactions with $Sn(acac)_2$ and hydrogen fluoride. *ACS Nano*, 9:2061–2070, 2015.

[61] Y. Lee, J. W. DuMont, and S. M. George. Trimethylaluminum as the metal precursor for the atomic layer etching of Al_2O_3 using sequential, self-limiting thermal reactions. *Chemistry of Materials*, 28:2994–3003, 2016.

[62] Y. Lee, G.-H. Kim, B. Choi, J. Yoon, H.-J. Kim, D. H. Kim, D. M. Kim, M.-H. Kang, and S.-J. Choi. Design study of the gate-all-around silicon nanosheet MOSFETs. *Semiconductor Science and Technology*, 35:03LT01, 2020.

[63] D. H. Levy, S. F. Nelson, and D. Freeman. Oxide electronics by spatial atomic layer deposition. *Journal of Display Technology*, 5:484–494, 2009.

[64] C. Li, D. Metzler, C. S. Lai, E. A. Hudson, and G. S. Oehrlein. Fluorocarbon based atomic layer etching of Si_3N_4 and etching selectivity of SiO_2 over Si_3N_4. *Journal of Vacuum Science & Technology A*, 34:041307, 2016.

[65] N. Li, S. D. Marcus, T. T. Ngo, and K. Griffin. Gas separation control in spatial atomic layer deposition, Jan 2022.

[66] Z. Li, K. Cao, X. Li, and R. Chen. Computational fluid dynamics modeling of spatial atomic layer deposition on microgroove substrates. *International Journal of Heat and Mass Transfer*, 181:121854, 2021.

[67] T. Lill. *Atomic layer processing*. Wiley-VCH, Freemont, CA, 2021.

[68] Y. Lou and P. D. Christofides. Feedback control of growth rate and surface roughness in thin film growth. *AIChE Journal*, 49:2099–2113, 2003.

[69] W. Lu, Y. Lee, J. Murdzek, J. Gertsch, A. Vardi, L. Kong, S. M. George, and J. A. del Alamo. First transistor demonstration of thermal atomic layer etching: InGaAs finfets with sub-5 nm fin-width featuring in situ ale-ald. In *Proceeding of IEEE International Electron Devices Meeting (IEDM)*, pages 39.1.1–39.1.4, San Francisco, USA, 2018.

[70] A. J. M. Mackus, M. J. M. Merkx, and W. M. M. Kessels. From the bottom-up: Toward area-selective atomic layer deposition with high selectivity. *Chemistry of Materials*, 31: 2–12, 2019.

[71] A. Mameli, M. J. M. Merkx, B. Karasulu, F. Roozeboom, W. E. M. M. Kessels, and A. J. M. Mackus. Area-selective atomic layer deposition of SiO_2 using acetylacetone as a chemoselective inhibitor in an ABC-type cycle. *ACS Nano*, 11:9303–9311, 2017.

[72] D. Maroudas. Multiscale modeling of hard materials: Challenges and opportunities for chemical engineering. *AIChE Journal*, 46:878–882, 2000.

[73] M. J. M. Merkx, T. E. Sandoval, D. M. Hausmann, W. M. M. Kessels, and A. J. M. Mackus. Mechanism of precursor blocking by acetylacetone inhibitor molecules during

area-selective atomic layer deposition of SiO_2. *Chemistry of Materials*, 32:3335–3345, 2020.

[74] M. J. M. Merkx, A. Angelidis, A. Mameli, J. Li, P. C. Lemaire, K. Sharma, D. M. Hausmann, W. M. M. Kessels, T. E. Sandoval, and A. J. M. Mackus. Relation between reactive surface sites and precursor choice for area-selective atomic layer deposition using small molecule inhibitors. *The Journal of Physical Chemistry C*, 126:4845–4853, 2022.

[75] D. Metzler, R. L. Bruce, S. Engelmann, E. A. Joseph, and G. S. Oehrlein. Fluorocarbon assisted atomic layer etching of SiO_2 using cyclic Ar/C_4F_8 plasma. *Journal of Vacuum Science & Technology A*, 32:020603, 2014.

[76] F. S. Minaye Hashemi, C. Prasittichai, and S. F. Bent. Self-correcting process for high quality patterning by atomic layer deposition. *ACS Nano*, 9:8710–8717, 2015.

[77] J. Moyne, E. Del Castillo, and A. M. Hurwitz. *Run-to-Run Control in Semiconductor Manufacturing*. CRC Press, 2018.

[78] J. Mulkens, M. Hanna, H. Wei, V. Vaenkatesan, H. Megens, and D. Slotboom. Overlay and edge placement control strategies for the 7 *nm* node using euv and arf lithography. In O. R. W. II and E. M. Panning, editors, *Extreme Ultraviolet (EUV) Lithography VI*, volume 9422, page 94221Q, 2015.

[79] D. Muñoz-Rojas, V. H. Nguyen, C. M. de la Huerta, C. Jiménez, and D. Bellet. Spatial atomic layer deposition. In *Chemical Vapor Deposition for Nanotechnology*, pages 1–25. IntechOpen, 2019.

[80] S. K. Natarajan and S. D. Elliott. Modeling the chemical mechanism of the thermal atomic layer etch of aluminum oxide: a density functional theory study of reactions during HF exposure. *Chemistry of Materials*, 30:5912–5922, 2018.

[81] Z. Ning, J. R. Moyne, T. Smith, D. Boning, E. Del Castillo, J.-Y. Yeh, and A. Hurwitz. A comparative analysis of run-to-run control algorithms in the semiconductor manufacturing industry. In *IEEE/SEMI 1996 Advanced Semiconductor Manufacturing Conference and Workshop. Theme-Innovative Approaches to Growth in the Semiconductor Industry. ASMC 96 Proceedings*, pages 375–381, 1996.

[82] T. Nishiguchi, S. Saitoh, N. Kameda, Y. Morikawa, M. Kekura, H. Nonaka, and S. Ichimura. Rapid oxidation of silicon using UV-light irradiation in low-pressure, highly concentrated ozone gas below 300 °C. *Japanese Journal of Applied Physics*, 46:2835, 2007.

[83] D. Pan. Density functional theory (DFT)-enhanced computational fluid dynamics modeling of substrate movement and chemical deposition process in spatial atomic layer deposition. *Chemical Engineering Science*, 234:116447, 2021.

[84] D. Pan, T. Li, T. Chien Jen, and 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*, 32:01A110, 2014.

[85] D. Pan, T. Li, T.-C. Jen, and 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*, 32:01A110, 2014.

[86] D. Pan, T.-C. Jen, and C. Yuan. Effects of gap size, temperature and pumping pressure on the fluid dynamics and chemical kinetics of in-line spatial atomic layer deposition of Al_2O_3. *International Journal of Heat and Mass Transfer*, 96:189–198, 2016.

[87] P. Peltonen, V. Vuorinen, G. Marin, A. J. Karttunen, and M. Karppinen. Numerical study on the fluid dynamical aspects of atomic layer deposition process. *Journal of Vacuum Science & Technology A*, 36:021516, 2018.

[88] P. Pitriana, T. D. K. Wungu, H. Herman, and R. Hidayat. The computation parameters optimizations for electronic structure calculation of $LiPbI_3$ perovskite by the density functional theory method. *IOP Conference Series Materials Science and Engineering*, 434:012026, 2018.

[89] P. Poodt, A. Lankhorst, F. Roozeboom, K. Spee, D. Maas, and A. Vermeer. High-speed spatial atomic-layer deposition of aluminum oxide layers for solar cell passivation. *Advanced Materials*, 22:3564–3567, 2010.

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

[91] P. Poodt, J. van Lieshout, A. Illiberi, R. Knaapen, F. Roozeboom, and A. van Asten. On the kinetics of spatial atomic layer deposition. *Journal of Vacuum Science & Technology A*, 31:01A108, 2013.

[92] A. B. Rahane, M. D. Deshpande, and V. Kumar. Structural and electronic properties of $(Al_2O_3)_n$ clusters with n = 1-10 from first principles calculations. *Journal of Physical Chemistry C*, 115:18111–18121, 2011.

[93] A. Razavieh, P. Zeitzoff, and E. J. Nowak. Challenges and limitations of CMOS scaling for FinFET and beyond architectures. *IEEE Transactions on Nanotechnology*, 18:999–1004, 2019.

[94] F. M. Richards. The interpretation of protein structures: Total volume, group volume distributions and packing density. *Journal of Molecular Biology*, 82:1–14, 1973.

[95] H. Roh, H.-L. Kim, K. Khumaini, H. Son, D. Shin, and W.-J. Lee. Effect of deposition temperature and surface reactions in atomic layer deposition of silicon oxide using bis(diethylamino)silane and ozone. *Applied Surface Science*, 571:151231, 2022. ISSN 0169-4332.

[96] S. Scandolo, P. Giannozzi, C. Cavazzoni, S. de Gironcoli, A. Pasquarello, and S. Baroni. First-principles codes for computational crystallography in the quantum-ESPRESSO package. *Zeitschrift für Kristallographie - Crystalline Materials*, 220:574–579, 2005.

[97] M. C. Schwille, T. Schössler, F. Schön, M. Oettel, and 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*, 35:01B119, 2017.

[98] S. Seo, B. C. Yeo, S. S. Han, C. M. Yoon, J. Y. Yang, J. Yoon, C. Yoo, H. jin Kim, Y. baek Lee, S. J. Lee, J.-M. Myoung, H.-B.-R. Lee, W.-H. Kim, I.-K. Oh, and H. Kim. Reaction

mechanism of area-selective atomic layer deposition for Al$_2$O$_3$ nanopatterns. *ACS Applied Materials & Interfaces*, 9:41607–41617, 2017.

[99] M. R. Shaeri, T.-C. Jen, and C. Y. Yuan. Reactor scale simulation of an atomic layer deposition process. *Chemical Engineering Research & Design*, 94:584–593, 2015.

[100] K. Sharma, R. A. Hall, and S. M. George. Spatial atomic layer deposition on flexible substrates using a modular rotating cylinder reactor. *Journal of Vacuum Science & Technology A*, 33:01A132, 2015.

[101] R. K. Sharma, R. Gupta, M. Gupta, and R. S. Gupta. Dual-material double-gate SOI n-MOSFET: Gate misalignment analysis. *IEEE Transactions on Electron Devices*, 56:1284–1291, 2009.

[102] D. Sheppard, P. Xiao, W. Chemelewski, D. D. Johnson, and G. Henkelman. A generalized solid-state nudged elastic band method. *Journal of Chemical Physics*, 136:074103, 2012.

[103] M. Shirazi and S. D. Elliott. Atomistic kinetic monte carlo study of atomic layer deposition derived from density functional theory. *Journal of Computational Chemistry*, 35:244–259, 2014.

[104] A. Sinha, D. W. Hess, and C. L. Henderson. Area selective atomic layer deposition of titanium dioxide: Effect of precursor chemistry. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 24:2523–2532, 2006.

[105] W. Sun, Y. Kim, J. Shin, and W. Yang. Shower head of combinatorial spatial atomic layer deposition apparatus, May 3 2017. KR Patent 20,170,025,417 A.

[106] T. Suntola and J. Antson. Method for producing compound thin films, May 3 1977. US Patent 4,058,430.

[107] A. Togo and I. Tanaka. First principles phonon calculations in materials science. *Scripta Materialia*, 108:1–5, 2015.

[108] M. Tom, S. Yun, H. Wang, F. Ou, G. Orkoulas, and P. D. Christofides. Machine learning-based run-to-run control of a spatial thermal atomic layer etching reactor. *Computers & Chemical Engineering*, 168:108044, 2022.

[109] J. Voas, N. Kshetri, and J. F. DeFranco. Scarcity and global insecurity: The semiconductor shortage. *IT Professional*, 23:78–82, 2021.

[110] A. F. Voter. Introduction to the kinetic monte carlo method. In K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga, editors, *Radiation Effects in Solids*, pages 1–23, 2007.

[111] S. Wang, X. Liu, and P. Zhou. The road for 2D semiconductors in the silicon age. *Advanced Materials*, 34:2106886, 2022.

[112] T. Weckman, M. Shirazi, S. D. Elliott, and K. Laasonen. Kinetic monte carlo study of the atomic layer deposition of zinc oxide. *The Journal of Physical Chemistry C*, 122:27044–27058, 2018.

[113] J. Yarbrough, A. B. Shearer, and S. F. Bent. Next generation nanopatterning using small

molecule inhibitors for area-selective atomic layer deposition. *Journal of Vacuum Science & Technology A*, 39:021002, 2021.

[114] S. Yun, 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. *Computers & Chemical Engineering*, 148:107267, 2021.

[115] S. Yun, F. Ou, H. Wang, M. Tom, G. Orkoulas, and P. D. Christofides. Atomistic-mesoscopic modeling of area-selective thermal atomic layer deposition. *Chemical Engineering Research & Design*, 188:271–286, 2022.

[116] S. Yun, M. Tom, J. Luo, G. Orkoulas, and P. D. Christofides. Microscopic and data-driven modeling and operation of thermal atomic layer etching of aluminum oxide thin films. *Chemical Engineering Research & Design*, 177:96–107, 2022.

[117] S. Yun, M. Tom, G. Orkoulas, and P. D. Christofides. Multiscale computational fluid dynamics modeling of spatial thermal atomic layer etching. *Computers & Chemical Engineering*, 163:107861, 2022.

[118] S. Yun, M. Tom, F. Ou, G. Orkoulas, and P. D. Christofides. Multivariable run-to-run control of thermal atomic layer etching of aluminum oxide thin films. *Chemical Engineering Research & Design*, 182:1–12, 2022.

[119] S. Yun, M. Tom, F. Ou, G. Orkoulas, and P. D. Christofides. Multivariable run-to-run control of thermal atomic layer etching of aluminum oxide thin films. *Chemical Engineering Research & Design*, 182:1–12, 2022.

[120] S. Yun, M. Tom, F. Ou, G. Orkoulas, and P. D. Christofides. Multiscale computational fluid dynamics modeling of thermal atomic layer etching: Application to chamber configuration design. *Computers & Chemical Engineering*, 161:107757, 2022.

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

[122] Y. Zhang, Y. Ding, and P. D. Christofides. Multiscale computational fluid dynamics modeling and reactor design of plasma-enhanced atomic layer deposition. *Computers & Chemical Engineering*, 142:107066, 2020.

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

[124] R.-S. Zhou and R. L. Snyder. Structures and transformation mechanisms of the η, γ and θ transition aluminas. *Acta Crystallographica*, B47:617–630, 1991.

[125] D. R. Zywotko, J. Faguet, and S. M. George. Rapid atomic layer etching of Al_2O_3 using sequential exposures of hydrogen fluoride and trimethylaluminum with no purging. *Journal of Vacuum Science & Technology A*, 36:061508, 2018.