UNIVERSITY OF CALIFORNIA

Los Angeles

Microscopic Modeling, Machine Learning-Based Modeling and Optimal Operation of Thermal and Plasma Atomic Layer Deposition

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Chemical Engineering

by

Yangyao Ding

2021

ABSTRACT OF THE DISSERTATION

Microscopic Modeling, Machine Learning-Based Modeling and Optimal Operation of Thermal and Plasma Atomic Layer Deposition

by

Yangyao Ding

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2021

Professor Panagiotis D. Christofides, Chair

Atomic layer deposition (ALD) and plasma enhanced atomic layer deposition (PEALD) are the most widely utilized deposition techniques in the semiconductor industry due to their superior ability to produce highly conformal films and to deposit materials into high aspect-ratio geometric structures. Additionally, plasma enhanced ALD is able to further speed up the deposition process and to reduce the temperature requirement through the utilization of high energy particles. However, ALD and PEALD experiments remain expensive and time-consuming, and the existing first-principles based models have not yet been able to provide solutions to key process outputs that are computationally efficient, which is necessary for on-line optimization and real-time control.

Motivated by the above considerations, this dissertation focuses on addressing these issues for both ALD and PEALD. First, for ALD, the development of key components of a comprehensive simulation framework is presented. The simulation framework integrates first-principles-based microscopic modeling, input/output modeling and optimal operation of thermal atomic layer deposition (ALD) of SiO_2 thin-films using bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. Specifically, we initially utilize Density Functional Theory (DFT)-based calculations for the computation of the key thermodynamic and kinetic parameters, which are then used in

the microscopic modeling of the ALD process. Subsequently, a detailed microscopic model is constructed, accounting for the microscopic lattice structure and atomic interactions, as well as multiple microscopic film growth processes including physisorption, abstraction and competing chemical reaction pathways. Kinetic Monte-Carlo (kMC) algorithms are utilized to obtain computationally efficient microscopic model solutions while preserving model fidelity. The obtained kMC simulation results are used to train Artificial Neural Network (ANN)-based data-driven models that capture the relationship between operating process parameters and time to ALD cycle completion. Specifically, a two-hidden-layer feed-forward ANN is constructed to find a feasible range of ALD operating conditions accounting for industrial considerations, and a Bayesian Regularized ANN is constructed to implement the cycle-to-cycle optimization of ALD cycle time. Extensive simulation results demonstrate the effectiveness of the proposed approaches. The kMC models successfully achieves a growth per cycle (GPC) of 1.8 Å per cycle, which is in the range of reported experimental values. The ANN models accurately predict deposition time to steady-state from the given operating condition input, and the cycle time optimization using BRANN model reduces the conventional BTBAS cycle time by 60%. After developing an efficient simulation framework, a more detailed study on the optimal operation policy is performed using a multiscale data-driven model. The multiscale data-driven model captures the macroscopic process domain dynamics with a linear parameter varying model, and characterizes the microscopic domain film growth dynamics with a feed-forward artificial neural network (ANN) model. The multiscale data-driven model predicts the transient deposition rate from the following four key process operating parameters that can be manipulated, measured or estimated by process engineers: precursor feed flow rate, operating pressure, surface heating, and transient film coverage. Our results demonstrate that the multiscale data-driven model can efficiently characterize the transient input-output relationship for the SiO_2 thermal ALD process using Bis(tertiary-butylamino)silane (BTBAS) as the Si precursor. The multiscale data-driven model successfully reduces the computational time from 0.6 - 1.2 hr for each time step, which is required for the first-principles based multiscale computational fluid dynamics (CFD) model,

to less than 0.1 s, making its real-time usage feasible. The developed data-driven modeling methodology can be further generalized and used for other thermal ALD or similar deposition systems, which will greatly enhance the feasibility of industrial manufacturing processes.

For PEALD, a similar methodology is adopted. First, an accurate, yet efficient kinetic Monte Carlo (kMC) model and an associated machine learning (ML) analysis are proposed to capture the surface deposition mechanism of the HfO2 thin-film PEALD using Tetrakis-dimethylamino-Hafnium (TDMAHf) and oxygen plasma. Density Functional Theory (DFT) calculations are performed to obtain the key kinetic parameters and the structural details. After the model is validated by experimental data, a database is generated to explore a variety of precursor partial pressure and substrate temperature combinations using the kMC algorithm. A feed-forward Bayesian regularized artificial neural network (BRANN) is then constructed to characterize the input-output relationship and to investigate the optimal operating condition. Next, based on an associated work on a comprehensive 3D multiscale computational fluid dynamics (CFD) model for the PEALD process, a 2D axisymmetric reduction of the previous 3D model of PEALD reactors with and without the showerhead design has been adopted to enhance the computational efficiency. Using the derived 2D CFD model, a data-driven model is constructed based on a recurrent neural network (RNN) for process characterization. The developed integrated data-driven model is demonstrated to accurately characterize the key aspects of the deposition process as well as the gas-phase transport profile while maintaining computational efficiency. The derived data-driven model is further validated with the results from a full 3D multiscale CFD model to evaluate model discrepancy. Using the data-driven model, an operational strategy database is generated, from which the optimal operating conditions can be determined for the deposition of HfO_2 thin-film based on an elementary cost analysis.

The dissertation of Yangyao Ding is approved.

Philippe Sautet

Samanvaya Srivastava

Mathieu Bauchy

Panagiotis D. Christofides, Committee Chair

University of California, Los Angeles

2021

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Professor Panagiotis D. Christofides, for his support and encouragement of both my technical work and also of my goals for my life over the years. Professor Christofides sets an example of excellence as a researcher, mentor, instructor, and role model. His mentoring on control theory, process/systems engineering, and life in general has made a profound impact on both my character and career goals. I am also grateful for my family and friends for their encouragement, support, and patience throughout my graduate career and throughout my life. In particular, I am grateful for my parents, Xianjie Ding and Xiaoping Hu.

In addition, I would like to thank all of my colleagues with whom I have worked over the years in the Christofides research group, including Anh Tran, Yi Ming Ren, Marquis Crose, Andres Aguirre, Zhe Wu, Zhihao Zhang, David Rincon, Sungil Yun and Scarlett Chen. I would particularly like to thank Yichi Zhang with whom I have collaborated extensively and spent long hours working on papers together.

I would also like to thank Professor Philippe Sautet, Professor Samanvaya Srivastara and Professor Mathieu Bauchy for serving on my doctoral committee.

Financial support from UCLA Graduate Division Fellowship, the US Department of Energy (DOE) and the US National Science Foundation (NSF) is gratefully acknowledged, and my work could not have been done without this support.

Chapter 2 contains a version of: Ding, Y., Y. Zhang, K. Kim, A. Tran, Z. Wu and P. D. Christofides, "Microscopic Modeling and Optimal Operation of Thermal Atomic Layer Deposition," *Chem. Eng. Res. & Des.*, **145**, 159-172, 2019.

Chapter 3 contains a version of: Ding, Y., Y. Zhang, M. Ren, G. Orkoulas and P. D. Christofides, "Machine Learning-Based Modeling and Operation for ALD of SiO_2 Thin Films Using Data from a Multiscale CFD Simulation," *Chem. Eng. Res. & Des.*,**151**, 131-145, 2019.

Chapter 4 contains a version of: Ding, Y., Y. Zhang, G. Orkoulas and P. D. Christofides, "Microscopic modeling and optimal operation of plasma enhanced atomic layer deposition," *Chem.*

Eng. Res. & Des., **159**, 439-454, 2020.

Chapter 5 contains a version of: Ding, Y., Y. Zhang, H. Y. Chung and P. D. Christofides, "Machine Learning-Based Modeling and Operation of Plasma-Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films," *Comp. & Chem. Eng.*, **144**, 107148, 2021.

Curriculum Vitae

Education

University of California, Los Angeles **Sep. 2013 - July 2017**

B.S., Chemical Engineering **Los Angeles, US**

Journal Publications

1. 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," *Comp. & Chem. Eng.*, **148**, 107267, 2021.

2. **Y. Ding**, Y. Zhang, H. Y. Chung and P. D. Christofides, "Machine Learning-Based Modeling and Operation of Plasma-Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films," Comp. & Chem. Eng., 144, 107148, 2021.

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

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

5. **Y. Ding**, Y. Zhang, G. Orkoulas and P. D. Christofides, "Microscopic modeling and optimal operation of plasma enhanced atomic layer deposition," Chem. Eng. Res. & Des., 159, 439-454, 2020.

6. 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 (1), 18, 18 pages, 2020.

7. **Y. Ding**, Y. Zhang, M. Ren, G. Orkoulas and P. D. Christofides, "Machine Learning-Based Modeling and Operation for ALD of SiO_2 Thin Films Using Data from a Multiscale CFD Simulation," Chem. Eng. Res. & Des., 151, 131-145, 2019.

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

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

Chapter 1

Introduction

1.1 Motivation

The thin-film deposition is one of the most important building blocks in the semiconductor industry. Various deposition techniques, such as epitaxy, chemical vapor deposition (CVD), and physical vapor deposition (PVD), have been developed to deposit high quality thin-films of various materials, e.g., Al_2O_3, Hf_2O_3, RuO_2, SiO_2, etc. [2]. However, the requirements in the production of advanced memory devices have become more and more demanding. For example, the dimensions of new high-k gate dielectrics are under transition to sub-10-nm scale and the associated film thickness is required to be under 30 Å [3, 4]. Also, new transistor designs often involve complex three-dimensional structures rather than two-dimensional planar surfaces, along with the demand for conformal films with a stringent criterion on uniformity and defects. Thus, the atomic layer deposition (ALD) process has been widely adopted by industry to meet the requirements of major design breakthroughs [5]. ALD is a thin-film deposition method originally derived from CVD. The ALD method enables a layer-by-layer film growth with film uniformity at the atomic level, which is more precise and controllable than the traditional CVD approach [6–9]. Therefore, in the field of microelectronics 3D integration, where ultra-thin and highly-conformal films are needed, ALD has gained significant popularity.

Currently, there is a wealth of ALD research on both laboratory and industrial scales [5]. This significant research activity on ALD has led to the discovery of novel precursors and mechanisms which make high throughput film processing possible while allowing various substrate lay-outs [10]. However, experimental and industrial works on ALD remain expensive and time-consuming due to the cost of precursors and ALD-specific equipment, as well as due to the limited throughput [8]. Additionally, the real-time in-situ monitoring of film growth is not possible because the molecular structure can only be understood through methods like scanning electron microscopy (SEM) and scanning tunneling microscope (STM), which are accurate but destructive to the deposited film [11]. Thus, a model for ALD that provides insights on the details of real-time film profile and the overall growth rate can be beneficial to both industrial and research work.

In addition to traditional thermal ALD, there has been a growing number of PEALD works on both laboratory and industrial scales to investigate new types of thin-film materials, novel precursors, and their respective reaction mechanisms which enable improved PEALD operation. High-k dielectric thin-film materials like HfO_2 [12], TiO_2 [13] and ZrO_2 [14] are typically investigated due to their good band alignment to silicon gate, the high thermal stability, and the capability of offering high mobility for charge carriers. A variety of precursor species have been investigated for both the neutral heavy particle pulse and the plasma pulse for the deposition of the aforementioned high-k thin films. Taking HfO_2 as an example, metal-organic precursors are extensively used because of the low activation energy from the H–N hydrogen bonds. There has been a variety of research works for popular Hf-based metal-organic precursors including tetrakis(ethylmethylamino) hafnium (TEMAH) [13], tetrakis(dimethylamino) hafnium (TDMAH) [15], tetrakis(diethylamino) hafnium (TDEAH) [16], and others. For the plasma half-cycle, common oxide precursors involve H_2O plasma [17], O_2 [18] plasma, and O_3 plasma [18]. Despite the various advantages associated with the PEALD process, the experimental studies still face obstacles in operating costs to explore a substantial amount of experimental conditions and in acquiring clear first-principles-based understandings. In the reaction mechanisms domain, the surface precursor and plasma reactions are crucial to the deposition profile, but the detailed

mechanisms remain partially unknown [19]. Facing similar issues as ALD, there has yet to be a satisfying real-time monitoring technique for PEALD to observe the film quality. Therefore, the development of an accurate microscopic surface deposition model is critical to help understand the PEALD process and is the first fundamental step to construct an all-inclusive model to characterize the entire PEALD process. Molecular Dynamics (MD) has traditionally been utilized where an ab initio model tracks the movement of all particles. Nevertheless, such simulations are too computationally expensive for an industrial-scale system [20].

Therefore, for both ALD and PEALD, the development of an efficient simulation framework would be highly valuable and it could greatly reduce the cost to perform operating policy exploration and process condition optimization. Moreover, such a model can be used to configure a process controller to enable both real-time online control and run-to-run control, which makes the process even more robust.

1.2 Background

Both ALD and PEALD use the alternating purge scheme. In an ALD process, a substrate surface is exposed to alternating gas-phase precursor streams such that only one type of reactant is in contact with the substrate surface at each half-cycle. Once in contact, the precursor undergoes self-limiting surface reactions that allow a nearly complete and conformal surface coverage given sufficient exposure time and appropriate reactor conditions. In between the alternating precursor cycles, the reactor is purged with inert gas, ensuring all previously-entered precursors are removed from the chamber prior to the exposure of the film to the next precursor, avoiding undesirable reactions and a decrease in film purity [21]. Similar to ALD, precursors are introduced into the reactor through an alternating purge manner in the PEALD process. However, instead of regular chemical species, high-energy radicals and electromagnetically activated species are used in PEALD to facilitate the reactions.

Because the majority of reactions and dynamics occur on the substrate surface, it is important

to characterize the surface reactions accurately and efficiently. In particular, Molecular Dynamics (MD), and more recently, kinetic Monte-Carlo (kMC) are among the most popular simulation methods [22–24]. An ab initio MD model keeps track of all the particle movements and requires an overwhelming amount of computational resources, making it impossible to perform a simulation on an industrial scale process [20]. However, the kMC method has a crucial advantage in computational efficiency as it tracks a single event at a time in a predefined lattice space. Despite this simplification, the kMC method has been used in deposition models to successfully reproduce realistic profiles [24–26]. Recently, [24] proposed a novel multiscale computational fluid dynamics (CFD) simulation that used a surface microscopic n-fold hybrid kMC model and demonstrated its validity with a PECVD system. Moreover, many groups have shown the validity of using kMC in ALD simulation. For instance, [27] used raw probabilities of reaction and recombination to construct a kMC model for general plasma-enhanced ALD. [8] modeled a small scale Al_2O_3 ALD deposition using kMC based on first-principles analysis.

Still, there are some disadvantages associated with the general kMC models. In particular, the kMC model may be computationally expensive to be implemented in real-time manufacturing edge analytics. Also, the algorithm does not have a closed-form solution, which makes it hard to be incorporated in the model-based optimization and control. In order to deal with those problems, computationally efficient data-driven approaches can be utilized for the deposition processes. For example, [28] used a linear parameter varying model to characterize and optimize a thermal ALD reactor. Nevertheless, traditional statistical and machine learning methods like the ordinary least square regression (OLSR) often fail to reach good enough accuracy because of the complex non-linear input-output relationship in chemical engineering processes. As a result, with the enlightenment of the neural structures and the increasing computational power in recent decades, neural networks (NN) have been extensively researched in the domain of machine learning. A neural network utilizes an interconnected group of artificial neurons, each handling a simple mathematical relationship for the overall complex information processing based on a connectionist approach. Depending on the external and internal information that flows through the

network, the NN structure can be adapted to perform classification or regression on complicated information processing. [29] adopted the NN-assisted model to apply PI and batch-to-batch control in order to deal with the disturbances in the furnace ALD manufacturing. [30] and [31] adopted various advanced neural network algorithms for the thin-film deposition process and incorporated corresponding model-based controls. These previous works have demonstrated the enormous potential of the possible development of sophisticated online optimizations and advanced control schemes for deposition processes [32].

As a result, this dissertation strives to integrate and take advantage of both kMC-based microscopic modeling and machine learning-based data-driven approach to construct a general-purpose simulation framework that accurately and efficiently characterizes the ALD and PEALD process.

1.3 Dissertation Objectives and Structure

This dissertation presents the approaches to formulate a kinetic Monte-Carlo based microscopic model for both the ALD and PEALD process as well as the construction of a data-driven model for efficient optimal operating condition determination.

Specifically, the objectives of this dissertation are summarized as follows:

1. To develop a microscopic surface model to characterize the ALD of SiO_2 and the PEALD of HfO_2 by using the kinetic Monte-Carlo method and perform a preliminary study on the optimal operation of ALD by taking advantage of the feed-forward artificial neural network.

2. To develop a machine learning-based data-driven model construction approach by using developed multiscale computational fluid dynamics models, and to perform a thorough operating policy exploration using the data-driven model.

3. To present a framework for integrating the aforementioned components into a general methodology for the ALD and PEALD applications.

The dissertation is organized as follows. Chapters 2 and 3 discuss the work on traditional

thermal ALD. Specifically, Chapter 2 discusses the formulation of the microscopic surface model of SiO_2 ALD, including the computation of kinetic parameters using density functional theory (DFT), the selection of precursor, the process chemistry, the formulation of kinetic Monte-Carlo (kMC) based simulation method, and the simulation result. Chapter 2 also covers a preliminary study of the optimal operating region using only information from the surface model, where the feasible and unfeasible operating regions are determined using an artificial neural network trained by the microscopic simulation result. On the other hand, in Chapter 3, a dedicated study on exploring the operating domain using machine learning-based model reduction is presented. First, a multiscale computational fluid dynamics (CFD) model is used to collect a database of surface deposition profiles and gas chamber profiles under a large range of simulation conditions. Then, the gas-phase chamber profile is characterized with a linear parameter varying (LPV) model, and the microscopic model is characterized with an artificial neural network (ANN) model. By integrating the LPV model and the ANN model, a data-driven model for the ALD deposition is derived, and it is used to perform process optimization by systematically identifying the optimal operating condition.

Next, Chapters 4 and 5 cover the discussion on the PEALD process. Specifically, Chapter 4 goes over the surface modeling procedures. Similar to ALD, the modeling of the PEALD microscopic domain includes the process chemistry and the details of the kMC method. The simulation results on the PEALD of HfO_2 are also presented and an elementary study on the optimal operating region similar to Chapter 2 is also discussed for PEALD. In Chapter 5, the derivation of an RNN-based integrated multiscale data-driven model is described. Compared to the multiscale data-driven approach adopted in Chapter 3, the integrated data-driven model uses the macroscopic and microscopic data collected from the multiscale CFD model simultaneously to train a recurrent neural network (RNN). The adoption of RNN greatly reduces the required dataset size and enables better performance. Finally, using the RNN-based integrated data-driven model, the efficient exploration of the operating regime is presented.

Finally, Chapter 6 covers a summary of the main results of the dissertation.

Chapter 2

Microscopic Modeling and Optimal Operation of Atomic Layer Deposition

2.1 Introduction

In this chapter, we propose a kMC model to simulate the atomic layer deposition of SiO_2 thin-film on a 3D lattice. Bis(tertiary-butylamino)silane (BTBAS) and ozone are chosen as the precursors, and density functional theory (DFT) is used to obtain thermodynamic and kinetic parameters of the precursors that were not previously reported. Those DFT-calculated parameters such as the intermediate complex activation energies and pre-exponential coefficients crucially affect the microscopic kMC model event selection to reproduce realistic growth rates and structure. This model is also an extension of the previous 2D-lattice kMC models proposed, for example in [24]. Although there are many advantages in using a 2D simplification, including easy setup and computational efficiency, a 3D lattice is required to simulate ALD due to the importance of spatial influence between species. Moreover, the modeling of deposition onto high-aspect ratio (AR) features requires the analysis of edges and corners which is not possible with a 2D geometry [11]. Therefore, adpting a 3D microscopic lattice structure also enables simulating ALD with high AR design in the future. To further improve model performance, we develop a 3D triangular

lattice approximation of real crystal lattice while maintaining important structural characristics. After building the kMC simulation and validating its performance with experimental results, we derive a data-driven model via machine learning techniques to predict the steady-state film growth behavior for cycle-to-cycle optimization. Although kMC simulation can provide information about film growth in real-time, it is computationally expensive to be implemented in a control scheme for a large-scale system such as an entire wafer. Therefore, it is useful to derive a data-driven model that can provide a closed-form solution and can capture key film growth characteristics. Due to the stochastic nature of kMC and the non-linearity involved in the reaction mechanisms, non-linear regression models are applied to capture the input-output relationship. Traditional algebraic input-output models such as the least-squares method are subjected to prediction error and over-fitting error. Therefore, Artificial Neural Networks (ANNs), a more robust and more systematic way of parameter determination for non-linear problem, can be readily tailored to perform such tasks [33]. Specifically, a dense two-hidden-layer feed-forward ANN and a Bayesian Regularized ANN are implemented to find the feasible range of ALD operating conditions and to optimize ALD throughput cycle-to-cycle, respectively. The proposed kMC-model achieves a growth per cycle (GPC) rate of 1.8 Å, which lies in the experimentally reported range of 1.4 Å ~ 2.1 Å per cycle. Extensive simulation results demonstrate the validity of the proposed ANN approach in calculating optimal deposition times with respect to the operating parameters. The resulting model is demonstrated to reduce the industrial conventional cycle time by 60%. Furthermore, the modeling approach developed in this work can serve as a general guideline and be extended to the ALD of other thin-film materials using different precursors and operating conditions.

2.2 ALD Process Description and Modeling

This section focuses on developing a microscopic model that describes the deposition of SiO_2 thin-film via thermal ALD, which captures the structural details, the reaction mechanisms and the growth rate of SiO_2 thin-films. In this section, the approximation of the 3D SiO_2 lattice is

introduced and validated. Then, precursor selection is discussed based on experimental results and thermodynamic data. Subsequently, the reaction mechanism and associated kinetics are discussed in detail, including the DFT calculations of kinetic rate parameters and activation energies by Gaussian09/Gaussview software package [34]. Next, a hybrid n-fold model-specific kMC algorithm is developed to simulate the ALD process. Finally, we present the machine learning algorithms involved with the data-driven analysis on the relationship between operating conditions and cycle completion time.

2.2.1 Structural Assumptions of Deposited SiO_2

Our microscopic model aims to simulate the deposition of α-Quartz SiO_2, which crystallizes in the trigonal crystal system of space group $P3_121$ and has a local SiO_4 structure similar to tetrahedron. Although it may be tempting to use a true α-Quartz lattice structure, such a lattice structure would not be suitable for kMC implementation. The chemical nature of SiO_2 ALD process requires the consideration of multiple reaction pathways, structural geometry and defect generations. A true α-Quartz lattice kMC model would be conceptually complex and computationally challenging [8]. Thus, instead of a fully realistic 3D lattice, a 2D triangular model (i.e., each monolayer is off-shifted from the monolayer below it) adopted in our previous work by [24] is extended to 3D as an approximation of the actual α-Quartz crystal structure. In our model, a bond angle of 90° is assumed for the connectivity between Si and O atom instead of 109.5°. As shown in Figure 5.4 and Figure 2.2, the top view of the simulated lattice closely resembles that of the real lattice, with some angle distortion. This assumption leads to a lattice repetition every four cycles instead of five cycles as appears in the α-Quartz SiO_2. For any silicon atom in the approximated lattice, another silicon atom appears directly above it every four cycles, whereas in a realistic lattice, such pattern repeats every five cycles. However, this simplification does not influence the connectivity of the individual lattice cell or the validity of the model. We will also later demonstrate that our simulation captures accurately the growth rate and defect generation pattern of α-Quartz SiO_2 deposition reported by experimental results. Thus, this 3D triangular model is a valid simplification of the true structure.

A lattice size of 1200×1200 sites per layer is used, which is large enough for the simulation to be size-independent but still computationally efficient as demonstrated by [35, 36], and the height depends on the number of cycles simulated.

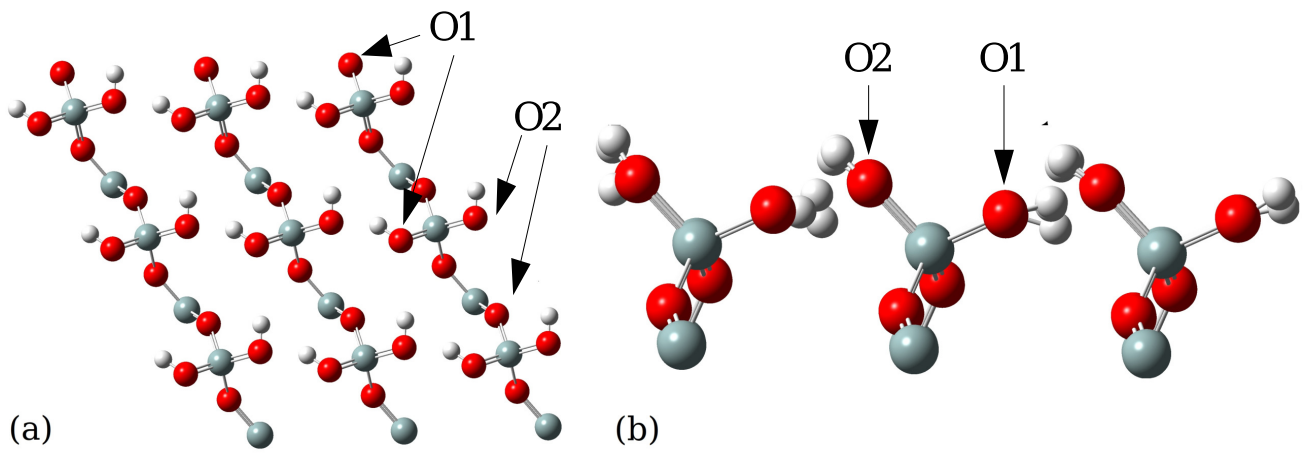


Figure 2.1: (a) Top view of the hydroxylated SiO_2(001) surface. (b) Side view of the hydroxylated SiO_2(001) surface, where O1 is the more electronegative oxygen. The double bonds are due to Gaussian display format, which does not influence the validity of the structure.

2.2.2 Precursor Selection

Surface reactions in the ALD process govern the growth rate and the structural pattern of SiO_2 films. Therefore, the selection of oxygen and silicon precursors is an important topic. In the past, many silicon precursors have been selected and studied to improve the uniformity and the growth rate of SiO_2 deposition. In recent years, aminosilane-based precursors have gained significant popularity because of the low activation energy of the sequential dissociative chemisorption mechanism caused by the H-N hydrogen bonds formed during the adsorption stage. These characteristics lead to high reaction rate and greatly improve the efficiency of SiO_2 deposition [37]. Among those aminosilane precursors, the most popular ones are: bis(tertiary-butylamino)silane (BTBAS), bis(diethylamino)silane (BDEAS), bis(dimethylamino)-silane (BDMAS), tris(dimethylamino)silane (TDMAS) and di(sec-butylamino)silane (DSBAS) [1,38–42]. In order to pick the most favorable precursor for our

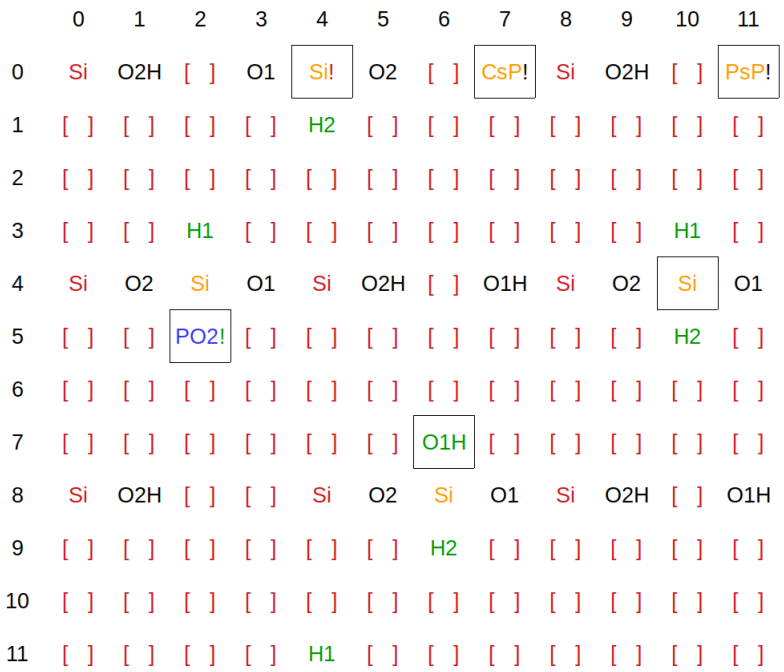


Figure 2.2: Top view of a 5-layer 12×12-site miniature demonstration of the full kinetic Monte-Carlo simulation lattice. The five layers and the species on the lattice are shown using different colors and symbols, respectively. The first (bottom) layer, labeled red, contains the base Si atoms. The second layer, labeled black, contains oxygen atoms or hydrogenated oxygens. The third layer, labeled yellow, contains the species from the first silicon half-cycle: Si is the neighbour-binding silicon, Si! is the self-binding silicon, and PsP and CsP are the physisorbed and chemisorbed precursors, respectively. The fourth layer, labeled green, contains the species from the first oxygen half-cycle: O and OH are the oxygen atoms and hydrogenated oxygens. The fifth (top) layer, labeled blue, contains physisorbed ozones (PO1 and PO2), which remain to be oxidized.

simulation, we account for the following factors: the existence of experimental data (e.g., growth per cycle (GPC) and precursor exposure time), the existence of theoretical data (e.g., reaction mechanism and associated kinetic parameters), and the availability of additional information such as film quality, sticking coefficient and steric hindrance studies. Based on the above considerations, BTBAS is chosen as the Si precursor due to its fast reported growth rate (1.4 Å ~ 2.1 Å per cycle), adequate experimental and theoretical data, and a detailed mechanism available to model the process [37, 43]. With respect to the oxygen precursor, ozone (O_3) is chosen among the common candidates for thermal ALD of oxide films, because ozone is chemically reactive and does not introduce hydrogen-involved side products in the thermal ALD process. Furthermore, ozone is extensively used in the industry and is widely studied in experiments, which makes its major chemical properties and reaction mechanisms readily accessible [43–45].

2.2.3 Reaction Mechanism

A full deposition cycle in the ALD process consists of two half-cycles, each using a specific precursor species to introduce the desired element onto the film. As mentioned above, we choose BTBAS and ozone as the precursors for SiO_2 deposition simulation. The reaction mechanism using these two precursors was reported by [43] and is explained in detail below.

The first half-cycle is referred to as the Si-Cycle, which contains physisorption, abstraction and a two-step dissociative chemisorption. In our model, we picked a fully hydroxylated $SiO_2(001)$ surface as our starting point, shown in Figure 5.4. The silicon precursor, BTBAS, is first physisorbed onto the substrate surface under specific temperature and pressure. According to [43], the two oxygen atoms in a SiO_2 cell have different electronegativities. The more electronegative oxygen atom, denoted as O^1, is more reactive and is therefore more likely to be electrophilicly attacked by precursor particles than the less electronegative oxygen atom, denoted as O^2. Therefore, as shown in Figure 5.5 (a), the precursor particle is first physisorbed onto the O^1-type hydroxyl group through a strong H-bond to form the reactant. Then, the physisorbed precursor goes through the first dissociative chemisorption step, forming a monoamine intermediate and

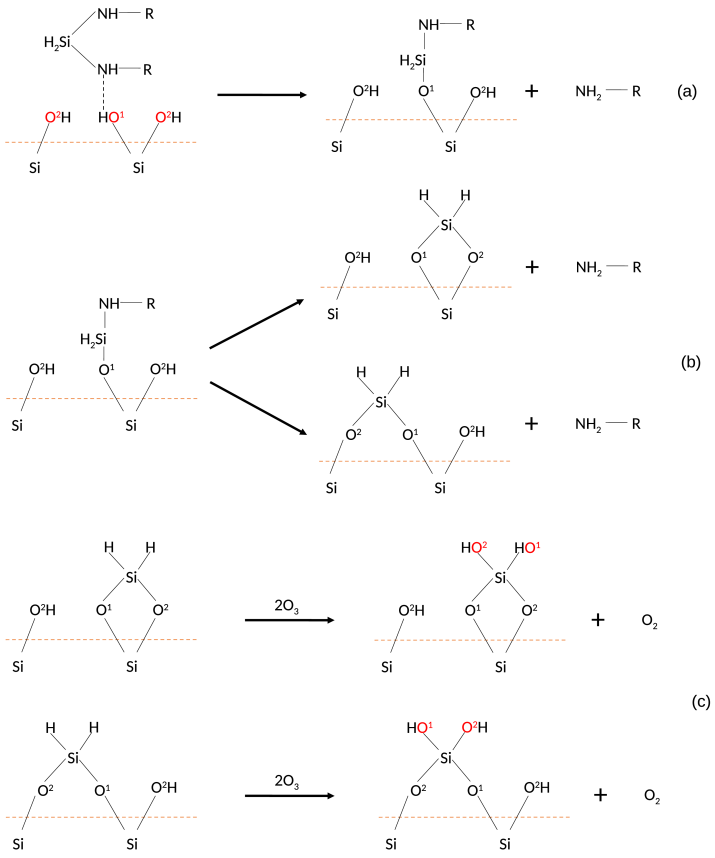


Figure 2.3: (a) First dissociative chemisorption step of BTBAS. (b) Second dissociative chemisorption step of BTBAS under self-binding and neighbour-binding mechanisms. (c) Oxidation of self-binding and neighbour-binding SiH_2 with ozone.

releasing one of the two aminoethyl groups. Next, the remaining aminoethyl group electrophilicly attacks an adjacent O^2-type hydroxyl group, which can be either from the neighbour Si atom, i.e., neighbour-binding route, or from the same substrate Si atom, i.e., self-binding route, as shown in Figure 5.5 (b). The former reaction pathway retains the original surface orientation, resulting in a thermodynamically favorable structure, whereas the latter, which is more kinetically favorable as shown in Table 2.1, causes a deviation from the (001) surface orientation and leads to defect formation. After the electrophilic attack, the other aminoethyl group is released from the surface structure and another O-Si bond is formed. The remaining two H atoms from the Si atom then become the new substrate surface. The competition of kinetic and thermodynamic favorability

is crucial in explaining the structural non-uniformity of SiO_2. Therefore, both reaction pathways and their reverse reactions are incorporated in our kMC model, and the reaction kinetics will be explained in more details in the next section.

The second half-cycle is referred to as the O-Cycle, which contains the ozone physisorption, abstraction and surface oxidation. The oxidation steps of self-binding and neighbour-binding H-Si groups are shown in Figure 5.5 (c). Once the surface is partially/fully chemisorbed, both terminating H atoms will be oxidized by ozone to hydroxyl group (-OH), which are utilized in the next Si-cycle.

2.2.4 Relative Rate Determination

In order to apply the kMC algorithm, we need to compute the kinetic rates of reactions discussed in the previous section. The physisorption of precursor particles onto the substrate surface is a gas-surface reaction. For such athermal or barrierless processes, the Collision Theory, as expressed in the equation below, is generally used to determine the rate constant:

$$r_{phs} = \frac{p_i}{RT}\sqrt{\frac{8k_bT}{\pi m_i}}s_{c,i}N_a\sigma \tag{2.1}$$

where r_{phs} is the physisorption reaction rate, p_i is the partial pressure of the species i, R is the ideal gas constant, T is the temperature, k_b is the Boltzmann constant, m_i is the molecular weight of species i, $s_{c,i}$ is the sticking coefficient of the species i at given temperature, N_a is the Avogadro number, and σ is the average area per surface site. Although the sticking coefficient of BTBAS is not reported in previous works, we obtain its value through an analogy with the sticking coefficient of BDEAS because of structural and electronic similarity [46].

On the contrary, chemisorption, abstraction and oxidation are thermodynamically activated kinetic reactions, which are generally described by the Transition State Theory (TST) [47]. Assuming quasi-equilibrium is achieved between the complex and the reactant, the reaction rate can be estimated using the thermodynamic properties of the transition state complexes, which

are computed using DFT. Thus, the reaction rate equation can be formulated with standard Arrhenius-type rate law as follows:

$$r_{rxn,i} = A_i \exp\left(\frac{-E_{a,i}}{k_b T}\right) \tag{2.2}$$

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

$$A_i = f_i^{\mathrm{TST}}\left(\frac{k_b}{T}\right) \tag{2.3}$$

where k_b is the Boltzmann constant, T is the temperature, and f_i^{TST} is the ratio of the vibrational partition function between the transition state complex and the reactant, calculated with DFT. In this work, all DFT calculations are performed using the Gaussian09 software, which will be illustrated in more details below. The resulting parameters are summarized in Table 2.1 and the associated nomenclature is explained in Section 4.2.4. Since the desorption reactions lead to gas-phase products, the DFT-calculated vibrational partition function ratios of those reactions are small than one, which match the results reported in literature. On the contrary, the other reactions are entirely surface reactions. Therefore, their vibrational partition function ratios all equal to one [47].

Reaction	Activation Energy (kcal/mole)	Vibrational Partition Function Ratio
$r_{si,chem}$	8.9	1
$r_{si,neigh,f}$	20.1	1
$r_{si,neigh,r}$	33.6	1
$r_{si,self,f}$	16.1	1
$r_{si,self,r}$	14.4	1
$r_{si,des}$	17.5	9.56e-8
$r_{o_a,f}$	17.7	1
$r_{o_b,f}$	15.4	1
$r_{o,des}$	9.224	1e-4

Table 2.1: Activation energies and partition function ratios of reactions.

2.2.5 Kinetic Monte-Carlo Algorithm

As mentioned in the Introduction, a first-principles Molecular Dynamic simulation is too computationally demanding to be feasible for the scale of system discussed in this work [20,25,26]. Thus, we adopt an n-fold hybrid kMC algorithm in the framework proposed by earlier works [24, 48, 49]. kMC is a stochastic algorithm that uses the kinetic rate information and uniformly distributed random numbers to determine event execution and system time evolution. Specifically, we define an event set as a collection of all events that have comparable rates. A total rate, r_{total}, is defined as:

$$r_{total} = \sum_{i=1}^{N} r_i \tag{2.4}$$

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

$$l_i = \frac{\sum_{j=1}^{i} r_j}{r_{total}}, i = 1, ..., N \tag{2.5}$$

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

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

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{2.6}$$

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

$$r_{o,total} = r_{o,phs} + r_{o,des} + r_{o_a,f} + r_{o_b,f} \tag{2.7}$$

where $r_{o,phs}$ is the rate of ozone physisorption, $r_{o,des}$ is the rate of ozone desorption, and $r_{o_a,f}$ together with $r_{o_b,f}$ are the oxidation rates of the chemisorbed species attached to a neighbour-binding silicon. The oxidation rate of the chemisorbed species attached to a self-binding silicon is orders of magnitude higher than that of a neighbour-binding silicon. Therefore, it is considered instantaneous and deterministic, and thus, omitted in the O-Cycle kMC selection. Similarly, the total rate for Si-Cycle is:

$$r_{si,total} = r_{btbas,phs} + r_{btbas,des} + r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} \tag{2.8}$$

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

For the O-Cycle, the rates of all considered reactions are comparable and can be modeled with the standard n-fold kMC algorithm. However, for the Si-Cycle, in order to simulate the realistic behavior of reaction kinetics, we need to consider surface reaction events separately from physisorption events for the following two reasons: First, surface reaction events are formulated and compared differently from physisorption events since surface species concentrations need to be considered to correctly describe the competition between the thermodynamic and kinetic favorability of competing pathways. Second, physisorption rates are an order of magnitude lower than surface reaction rates according to the DFT calculation, which means that the model will be saturated by surface reactions events if the events are not allocated properly. Thus, a decoupled

kMC kinetic scheme is proposed to partition the entire Si-Cycle events into two event sets: adsorption events containing only physisorption events, and surface reaction events containing the remaining events. The partitioned total rates, $r_{si,ads}$ and $r_{si,rxn}$, are then defined as follows:

$$r_{si,rxn} = r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} + r_{btbas,des} \tag{2.9}$$

$$r_{si,ads} = r_{btbas,phs} \tag{2.10}$$

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

$$J_{si,ads} = \frac{r_{si,ads}}{r_{si,total}} = 1 - J_{si,rxn} \tag{2.11}$$

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

$$l_{si,i} = \frac{\sum_{j=1}^{i} r_{rxn,j} R_j}{\sum_{k=1}^{N} r_{rxn,k} R_k}, \;\; i = 1, ..., N \tag{2.12}$$

where $l_{si,i} \in (0,1]$ represents the normalized indicator of the i^{th} event in the surface reaction event set, $r_{rxn,j}$ is the un-weighted chemical reaction rate for the j^{th} event calculated from Eq. 5.3, R_j is the number of reactants for each surface reaction, and N is the total number of events in the Si-Cycle surface reaction event set. The normalized indicators are then used to execute the event selection following the same approach performed in the standard kMC algorithm. In Section 4.3.3, it is demonstrated that this decoupling scheme achieves desired accuracy.

2.2.6 DFT and Thermodynamic Calculations

Although the reaction activation energies and mechanisms have already been analyzed for BTBAS by [43] as discussed in Section 4.2.2, many fundamental thermodynamic and kinetic properties of BTBAS have yet to be investigated, including its entropy, enthalpy, vibrational partition and others. Since the above properties are difficult to measure experimentally yet essential to the accurate microscopic simulation of ALD behavior, in this work, we utilize Density Functional Theory (DFT) with Gaussian09 package to compute them [34].

In the Si-Cycle, to calculate properties of BTBAS and its reaction kinetics with SiO_2 lattice, we first need to investigate the configuration of the hydroxylated surface lattice and the structure of the physisorbed BTBAS transition state complex. Specifically, to construct an optimal surface lattice, a generic bulk α-quartz SiO_2 unit cell is modified to generate a desired hydroxylated surface layer. The bulk unit cell is first imported into VESTA 3, which is a 3D visualization program widely adopted to construct crystalline structures [50]. The uppermost layer of Si atoms is removed, leaving two single bonded oxygen atoms per unit cell. Each oxygen atom is terminated with one hydrogen atom, and the new O-H bond is assumed to have the typical bond angle and bond length of 0.98 Å [51]. Then, the hydroxylated unit cell is imported to the Gaussview molecule builder tool. A $3 \times 3 \times 1$ SiO_2 lattice is constructed using the hydroxylated SiO_2 unit cell with Gaussian Periodic Boundary Condition (PBC) cell symmetry replication [34, 52]. A series of optimization steps is carried out, with all atoms other than the surface hydrogen and oxygen atoms fixed during structure optimization [43]. The lattice structure is first optimized using Hartree-Fock (HF) method with basis set 3-21G to obtain an initial guess of the structure. Next, the B3LYP method, a hybrid function Becke′s three-parameter exchange functional (B3) with Lee-Yang-Parr gradient correction functional (LYP) and triple valence plus polarization, is applied to optimize the structure to an acceptable energy minimum with basis set 6-31G+dp accuracy level [53, 54]. Subsequently, an initial guess of the TS complex is obtained by structuring together an optimized BTBAS molecule and a $3 \times 3 \times 1$ SiO_2 surface lattice. Then, the TS2 method in Gaussian09 is used to calculate an optimized TS complex structure that is most energetically favorable. The calculation is carried

out with the same level basis set accuracy as surface lattice structure optimization, plus modified coordinate definition and force constant calculation of every atomic position. The resulting TS structure has one negative vibrational frequency as expected [34]. Similarly, for the O-Cycle reaction, the thermodynamic and kinetic properties of gas-phase ozone molecule, H-Si surface lattice and physisorbed ozone are investigated. The same calculation procedures as Si-Cycle are carried out for the O-Cycle, respectively.

Finally, in order to perform a precise vibrational frequency calculation, Gaussian-4 (G4) theory is adopted. G4 theory is a complex computation method for accurate calculation of molecular energy based on ab initio molecular-orbital theory. It provides thermodynamic results for compounds containing second row (Li-F) and third row (Na-Cl) elements, which is applicable to our reacting molecules. Parallel computational work with Linda worker from Gaussian09 package is carried out to find out all the vibrational frequencies to calculate the ratio f^{TST} [34]. The final vibrational partition function along with other important thermodynamic properties are summarized in Table 2.1.

2.2.7 Artificial Neural Network Model and Non-linear Regression

The half-cycle time plays an important role in both industrial production and experimental studies of ALD, yet the actual time needed for each half-cycle for various experimental conditions of temperature and pressure remains unknown. Specifically, according to kMC simulation results and experimental analysis in SiO_2 thin-film ALD, Si-Cycles require longer time than O-Cycles at high temperature around 600 K, which impacts film coverage and quality [43]. Therefore, it is important to develop a model that can estimate the required Si-Cycle time, given operating temperature and pressure. Although the kMC model can be used to simulate the transient behavior of ALD and provide a reference to cycle time, it is computationally demanding to be applied for multiple-cycle film production in real-time. Moreover, kMC model is not a closed-form model, and thus, cannot be directly utilized for optimization and control. Therefore, instead of using the kMC model to perform real-time optimization, we take advantage of the data-driven modeling approach

and build Artificial Neural Network (ANN) models that correlate the Si-Cycle completion time with operating inlet temperature and pressure, using kMC-generated databases. Compared with traditional input-output models such as the least-squares method, the ANN approach is chosen for its advantages to lowering prediction error and its robustness against over-fitting error. Since it takes a long time to generate a database from the kMC model solutions, especially at low temperature and pressure, two ANN-based models are developed to serve different levels of precision. The first database covers a wider but sparse range of operating conditions, aiming to predict the suitable boundary for operating conditions and to provide a general reference of cycle completion time with acceptable accuracy. However, a higher accuracy is necessary for the real-time control and cycle-time optimization. Therefore, the second database is developed, which focuses on a smaller range with higher resolution. Due to the difference in data ranges, we adopt two levels of regularization to accurately train our neural networks: (1) The standard un-regularzied feed-forward neural network for the feasible range, and (2) The Bayesian Regularized Artificial Neural Network (BRANN) for the optimal range.

For the first database, we develop a feed-forward neural network with regular back propagation method for the nonlinear regression. Specifically, the input layer consists of two neurons, representing the inlet temperature and pressure, respectively. Two hidden layers are constructed, where the first and second layer contain 35 and 30 neurons, respectively. The output layer contains one single neuron, representing half-cycle completion time required to reach steady-state and, if possible, full coverage. For each hidden layer, the Rectified Linear Unit (ReLU) function is used as the non-linear activation function for better gradient propagation and efficient calculation:

$$ReLu(x) = x^{+} = max(0, x) \tag{2.13}$$

Additionally, all layers are densely connected, and the structure of the ANN is optimized via a grid search [55]. The general structure of a feed-forward two-input-single-output neural network with two hidden layers is given in Figure 3.5, where N_{ij} are the input neurons, H_{ij} are

the hidden layer neurons, and N_o is the output neuron. The above ANN structure is constructed with Tensorflow's keras module, a high-level application programming interface (API) designed to build and train a deep-learning ANN model which is widely used in deep learning applications.

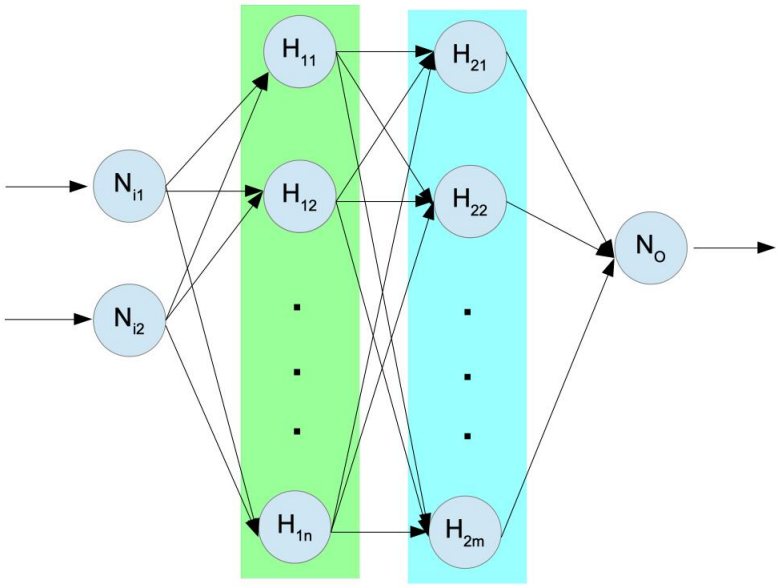


Figure 2.4: Feed-forward Artificial Neural Network with two inputs, two hidden layers, and one output.

The ANN is then trained using the above structure, and the mean square error (MSE) function, $S(w)$, which is typical for regression application, is chosen to be the cost function as follows:

$$S(w) = \frac{1}{N_D} \sum_{i=1}^{N_D} [y_i - (f(x_i, w))]^2 \tag{2.14}$$

where N_D is the number of data samples in the training dataset, y_i is the desired output value, w is the weight vector for all hidden layers, and $f(x_i, w)$ is the predicted value dependent on input x_i and weight w. The proper weight vector is obtained by solving an optimization problem to minimize the cost function $S(w)$ using standard back-propagation. Batch normalization is applied after each hidden layer to avoid saturation and high variance activation, thereby facilitating convergence speed and learning rate [56]. Dropout regularization layers with a rate coefficient of 0.5, which is the typical value for hidden layers, is used to perform model averaging with bagging method,

which enhances the generality of the network and reduces over-fitting. The RMSProp optimizer is adopted for model training, which utilizes normalized gradient from recent iterations by keeping a moving average of the squared gradient for each weight using the following equation:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma) g_t^2 \tag{2.15}$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{E[g^2]_t + \varepsilon}} g_t \tag{2.16}$$

where w_t and g_t are the weight parameter and its gradient at iteration t respectively, $E[g^2]_t$ denotes the running average of g^2 at iteration t, $E[g^2]_{t-1}$ denotes the running average of g^2 at iteration $t-1$, and ε is a smoothing term that avoids division by zero. A learning rate $\eta = 0.01$ and a momentum factor $\gamma = 0.9$ are used, which are the recommended values for RMSprop method [57].

For the second database, our objective is to identify an ANN model that captures the precise input-output relationship over a smaller operating range for real-time cycle-to-cycle optimization. However, the traditional RMSprop mechanism faces a dilemma between model accuracy and over-fitting. Thus BRANN is introduced as an alternative solution, by adding Bayesian regularization to the standard ANN. BRANN is more robust than standard neural networks for precise regression because the Bayesian regularization algorithm converts complex non-linear regression into a rigid regression, which is a well-posed statistical problem. By efficiently turning off the weights that are not relevant in the training process and incorporating Occam's razor principle, BRANN avoids the over-fitting and over-training problems by optimally penalizing excessive complexities in models [58]. In our model, BRANN for the second database is constructed, trained, and implemented using MATLAB machine learning package. Specifically, the input layer and the output layer of the BRANN are constructed using the same approach as the standard ANN above. For the inner structure of the BRANN, one hidden layer with 25 neurons is constructed. The hyperbolic tangent sigmoid function (*tansig*) is used as the activation function for the BRANN.

$$tansig(x) = \frac{2}{1+e^{-2x}} - 1 \tag{2.17}$$

Additionally, the hyperparameters α and β are added to the standard cost function as follows, which are chosen from a uniform random distribution:

$$S(w) = \beta \sum_{i=1}^{N_D} [y_i - (f(x_i, w))]^2 + \alpha \sum_{j=1}^{N_w} w_j^2 \tag{2.18}$$

where N_D is the number of data samples in the training dataset, w is the weight vector for all hidden layers, which is assumed to have a Gaussian distribution, N_w is the number of weight parameters, w_j is the j^{th} entry in the weight vector, y_i is the desired output value, and $f(x_i, w)$ is the predicted value dependent on w and the input x_i. To compute the optimal weight vector w and the continuously updated hyperparemeters α and β, a sequence of optimization problems are solved using the Levenberg-Marquardt algorithm [59]. The Bayesian inference calculation and the construction of the optimization problems are discussed in detail by [60] and [58].

Finally, both databases are generated based on the kMC model using the UCLA Hoffman2 Distributed Cluster. The two ranges of operating conditions are: (1) T = 550 K $\sim$ 700 K (ΔT = 5 K) with P = 80 Pa $\sim$ 160 Pa (ΔP = 2 Pa), and (2) T = 590 K $\sim$ 610 K (ΔT = 0.5 K) with P = 120 Pa $\sim$ 150 Pa (ΔP = 1 Pa). The first set of simulation results is divided into training, validation and test data under the ratio of 8:1:1, and the second set of simulation results is divided into training, validation and test data under the ratio of 7:1.5:1.5. The training dataset is used to determine the model parameters. The validation dataset is used in the training process to validate and improve the training. The testing dataset is randomly chosen from the entire dataset in advance to evaluate the final result and is not used in the training process.

2.3 Simulation Results

The results section is divided into three subsections. First, the kMC model is validated by comparing the simulation results with film growth behavior observed in literature. Then, the neural network models are demonstrated to be successfully developed to capture the relationship between cycle steady-state time and the operating conditions for the two databases covering feasible and optimal ranges. Specifically, the results of the first database cover a wider range of temperature and pressure conditions (feasible operating range), thereby providing a general guideline for the suitable conditions to carry out thermal SiO_2 ALD process, whereas the second database focuses on the range around T = 600 K and P = 133 Pa, typically employed in industry, with a higher resolution and is used for cycle time optimization as discussed in the previous subsection. Finally, the simulation of the multi-layer SiO_2 deposition demonstrates that the average ALD deposition time can be reduced utilizing the BRANN results, potentially allowing a higher industrial throughput.

2.3.1 Validation of Microscopic kMC Model with Experimental Data

The kMC model is validated by observing its behavior under varying temperature for a total of ten ALD cycles. Specifically, the precursor partial pressure is kept constant at 133 Pa, and the operating temperature is varied from 555 K to 625 K with an increment of 20 K for each simulation run. Each half-cycle is assigned 2 seconds, which is sufficient for the cycle to reach steady-state as observed. The same amount of cycle time is reported in experimental work by [37], although the specific value may vary since it also accounted for the time of gas-phase development. The precursor partial pressure on the substrate surface is important in determining physiosorption reaction rate. Therefore, the coupled effect of the gas-phase mass and momentum transfer in the reactor has a significant contribution to the cycle time and film quality, which we will discuss in our future work on the investigation of multi-scale computational fluid dynamics modeling of thermal ALD. Additionally, as mentioned in Section 2.2.7, since the O-Cycle time is much faster than Si-Cycle, we will focus on the Si-Cycle results only.

During the Si-Cycle, the competition between neighbour-binding and self-binding pathways is observed. As discussed in Section 4.2.2, self-binding events are kinetically favorable, while neighbour-binding events are thermodynamically favorable. The designed kinetics in our model is able to capture this behavior, as shown in Figure 2.5. At the beginning of each cycle, self-binding events dominate. As the simulation time goes on, self-binding silicons undergo reverse reaction and decrease, as shown by the solid lines, and neighbour-binding silicons start to form, as shown by the dashed lines. If a cycle is given enough time to develop, the fraction of neighbour-binding silicons will approach unity, while the fraction of self-binding silicons approaches zero. Since the neighbour-binding dominance is thermodynamically driven, as temperature decreases, more self-binding silicons appear in the initial deposition stages and a longer time is required for the lattice to develop into the desirable neighbour-binding-dominant profile, shown by the direction of the dashed arrow in Figure 2.5. This observation is consistent with the calculation and experimental analysis reported in [43].

ALD growth rate is also successfully simulated by the kMC model, which is characterized by the average growth of film thickness per cycle (GPC). Although the kMC model does not report the GPC directly as in experimental approaches, we could use the final coverage information to compute the GPC based on the literature values of SiO_2 unit cell lattice constants: a=5.5407 Å, b=c=4.918 Å [61]. An atom-to-atom measurement is performed in Gaussview to find the relative distance between Si and O atom within a SiO_2 cell, and the ideal layer thickness is calculated using the lattice constants and the atomic radii of oxygen and silicon atoms, which are 0.65 Å and 1.18 Å, respectively [61]. Then, the GPC is inferred through the simulated surface coverage and the ideal layer thickness. The simulated GPC at standard industrial operating condition of 600 K and 133 Pa is shown in Figure 2.6. The GPC varies little with increasing cycle numbers but does show a slightly decreasing pattern as reported in experimental works [3]. The average growth rate over 10 cycles is 1.8 Å per cycle, which is in the range of the SiO_2 GPC of 1.4-2.1 Å per cycle reported by [37].

Although the temperature does not impact GPC when sufficient time is given for each cycle

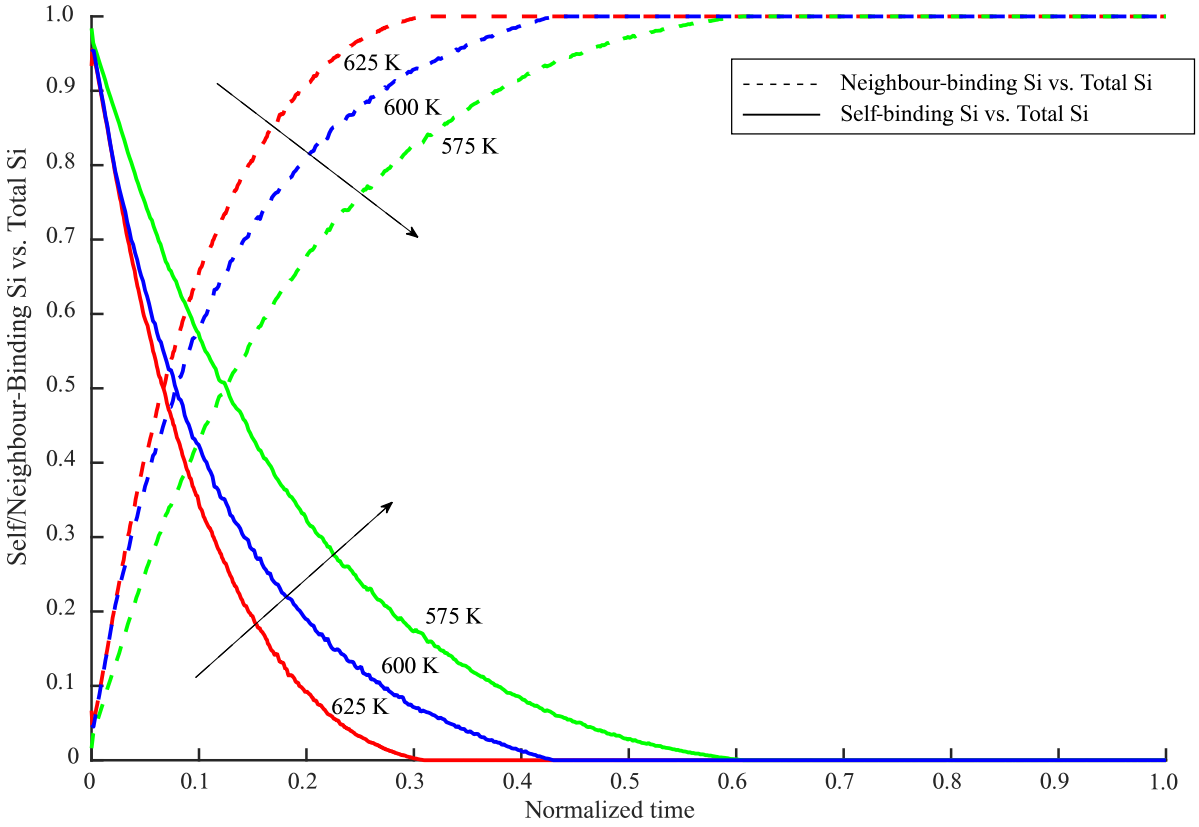


Figure 2.5: Competition between self-binding and neighbour-binding silicons with respect to the normalized cycle time. The dashed lines represent the ratio of neighbour-binding silicons versus total built silicons, and the solid lines represent the ratio of self-binding silicons versus total built silicons. The dashed arrows represent the direction of decreasing temperature.

to reach steady-state and achieve full coverage, it is a crucial factor for the transient deposition rate within each cycle. Figure 2.7 demonstrates that the transient deposition rate increases as temperature increases, and the rate increment is approximately proportional to the temperature increment. This effect is observed in experimental results from the work by [1], where the deposition rate is demonstrated to increase with temperature for unsaturated surfaces. Moreover, it is noteworthy that not all temperatures allow the surface to reach full coverage at steady-state. The selection of an appropriate temperature region will be introduced in the next section.

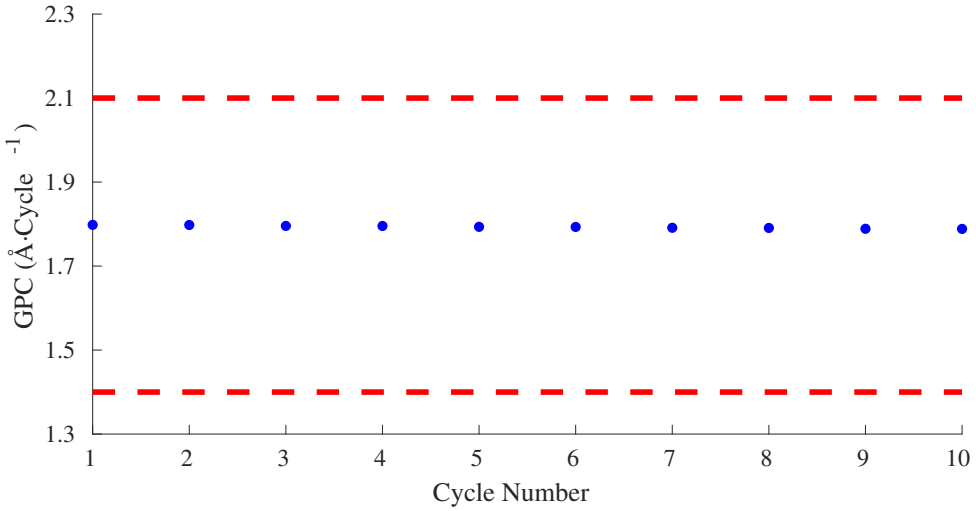


Figure 2.6: Steady-state GPC for the first ten cycles at 600 K and 133 Pa, where the dashed lines represent the upper and lower GPC limits reported in the literature [1].

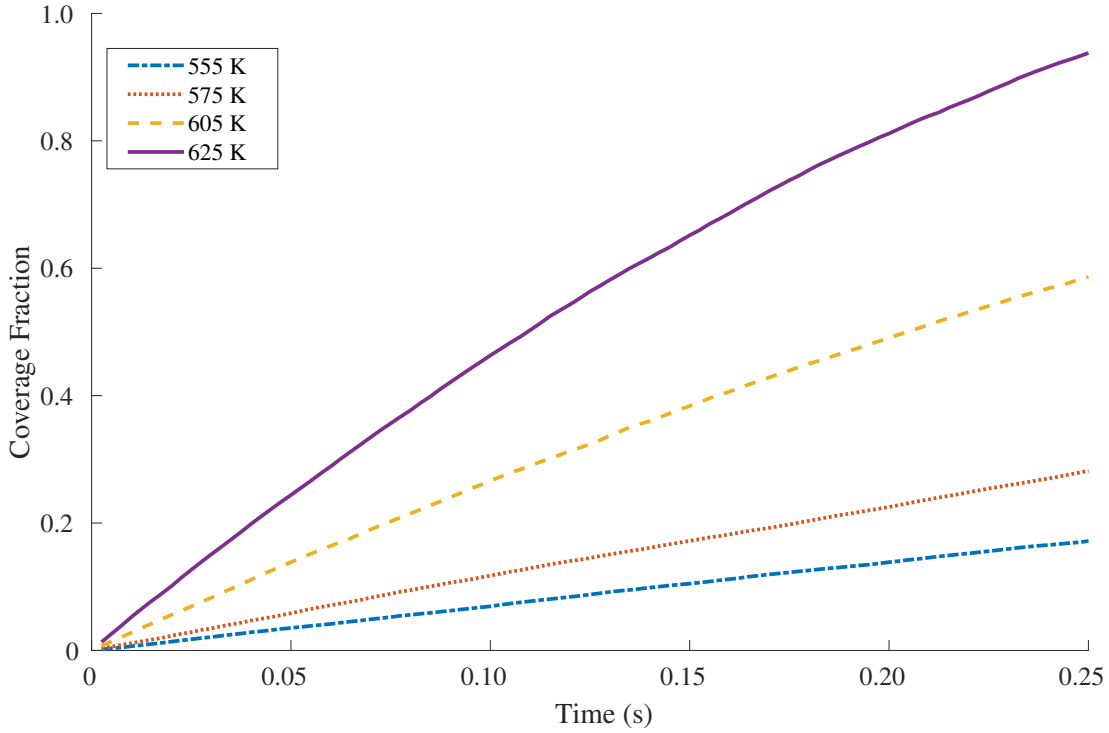


Figure 2.7: Comparison of initial transient deposition profiles for different temperatures at 133 Pa.

2.3.2 ANN Results for Si-Cycle

2.3.2.1 Feasible Operating Regime

In this section, a set of kMC simulations covering a wide range of operating conditions is carried out under fixed temperature and pressure throughout each simulation. The simulation is

terminated either when steady-state is achieved under the given conditions, or when the simulation time exceeds 5 seconds, which is too long to be considered industrially relevant [62]. Due to the stochastic nature of the kMC algorithm, the lattice surface configuration keeps changing at steady-state. However, the overall coverage at steady-state, which is one of the most crucial attributes of ALD processes, will be maintained at a certain value, with fluctuation under 0.5%. Then, the ANN model is developed to capture the relationship between the time to achieve 98% of the final coverage and the operating temperature and pressure. It is noted that, in the ANN model, the time for the system to reach 98% of the final coverage is used instead of the time to reach the final coverage to reduce the noise and inaccuracy involved in the steady-state fluctuations.

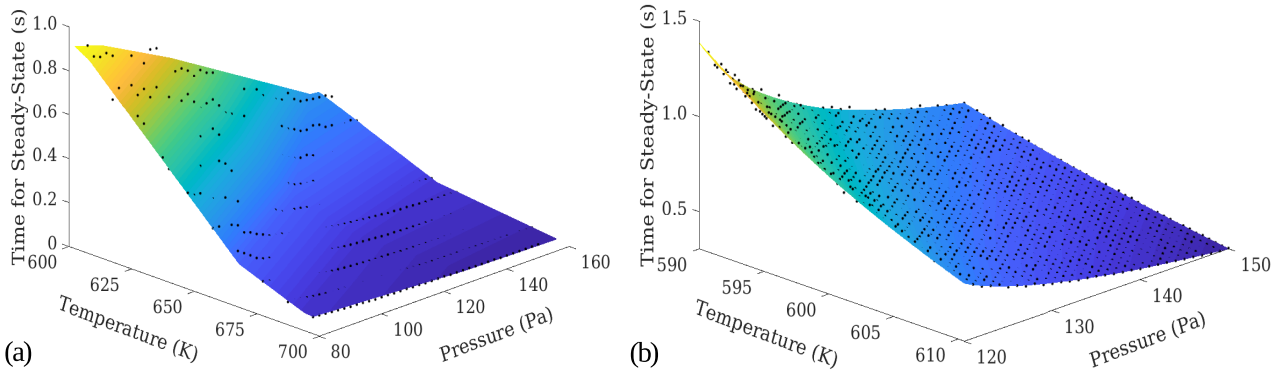


Figure 2.8: Dependence of the time to reach steady-state on the operating conditions, where black markers (dots) represent the training data and the surface represents the fitting result: (a) Large-range operating condition fitting. (b) Small-range operating condition fitting.

Two hidden layers are used for the ANN model and the numbers of neurons are determined to be 35 and 30 for the first and second hidden layer respectively, based on a grid search. A single-hidden-layer feed-forward ANN does not yield a good solution, and the performance cannot be improved by simply adding more neurons since over-fitting error is observed. Therefore, a two-hidden-layer structure is chosen, which is conventionally adopted to capture the exponential-like behaviour in our model. A mean absolute error on the test dataset is reported to be 8.00×10^{-3} s. As shown in Figure 2.8 (a,b), the ANN model achieves desired performance on the test dataset. In Figure 2.9 (a), the error distribution histogram shows a nearly normal distribution

for the error between predicted and actual completion time, with a mean close to zero. Additionally, in Figure 2.9 (b), the R-squared correlation between predicted steady-state time and the simulated time is 0.979, demonstrating that the neural network result has a good resemblance with the actual simulation result.

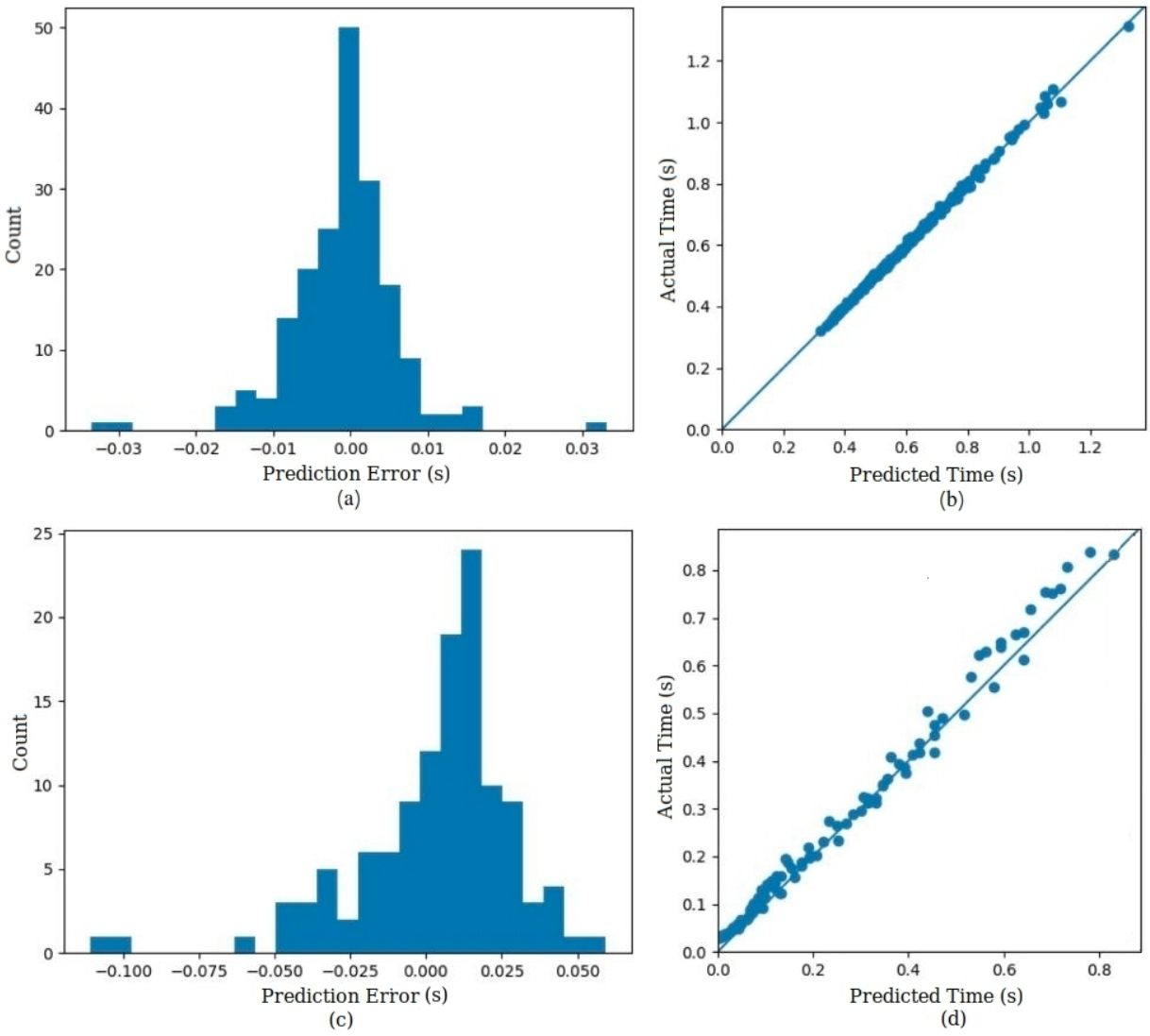


Figure 2.9: Performance of the Artificial Neural Network models: (a) Prediction error distribution histogram for the large-range operating conditions. (b) Correlation accuracy of the predicted time and the actual time for the large-range operating conditions, where the *x*-axis is the predicted time to reach steady-state from the neural network and the *y*-axis is the actual time to reach steady-state from the kMC model. (c) Prediction error distribution histogram for the small-range operating conditions. (d) Correlation accuracy of the predicted time and the actual time for the small-range operating conditions, where the *x*-axis is the predicted time to reach steady-state from the neural network and the *y*-axis is the actual time to reach steady-state from the kMC model.

Moreover, in Figure 2.10, a region of unsuitable operating conditions is identified where either the steady-state is not reached within 4 seconds or the steady-state coverage cannot achieve full coverage. For example, at T = 655 K and P = 86 Pa, although the steady-state is reached at 0.33 seconds, only 95% of the surface is covered. Additionally, at T = 545 K and P = 112 Pa, the surface does not reach the steady-state in 4 seconds. The maximum allowable time to steady-state is set to be 3.5 seconds considering industrial usefulness. Typically, a SiO_2 ALD half-cycle will take up to 5 seconds including the precursor flushing stage, which allows for the substrate surface to be fully saturated with precursor particles. Although the duration of precursor flushing may vary according to reactor geometry, 1 to 1.5 seconds is the usual lower limit allowed in a typical ALD reactor design to prevent damage to reactor caused by high precursor flow rate and pressure [62]. This result demonstrates that in order to achieve a good coverage, the contribution of both temperature and pressure should be accounted for. Since the optimal operating conditions might not be easily achieved or needed, the ANN model mentioned above can be used as an initial guideline to determine the feasible pressure and temperature conditions in experiments and industrial productions to operate the ALD process.

2.3.2.2 Optimal Operating Conditions

In this section, we focus on a narrower range of operating conditions with higher data resolution. For this dataset, one hidden layer with 25 neurons is chosen for the BRANN and the neuron number is determined via a grid search.A single layer BRANN is able to accurately capture the complex non-linearity based on its regularization algorithm. The trained neural network achieves a desired performance with a mean absolute error on the test data of 5.01×10^{-3} s. In Figure 2.9 (d), the statistical analysis shows that the neural network accurately predicts the simulation data. The R-squared correction between the predicted steady-state time and the simulated time is 0.99, characterized by an almost linear correlation curve. Moreover, as shown in Figure 2.9 (c), although limited by the number of data points in the test data, the error distribution is close to a normal distribution, which further validates the model fitness.

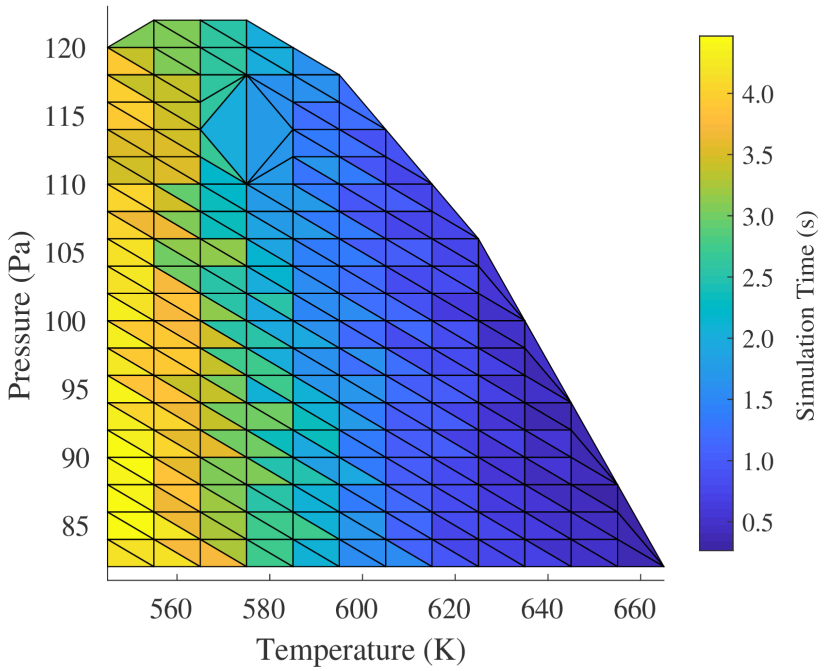


Figure 2.10: Range of operating conditions where the deposited films are not able to reach full coverage at steady-state.

According to the results in Section 2.3.2.1, the operating conditions chosen in this section guarantee full coverage of the newly deposited surface at steady-state within 3.5 seconds. Under these operating conditions, the time necessary to achieve steady-state ranges from 0.31 seconds to 1.33 seconds, and a higher temperature and pressure will reduce the time to reach steady-state. The trained BRANN accurately predicts the time to reach steady-state and full coverage from given pressure and temperature inputs, which will be used for cycle time optimization in the next section.

2.3.3 Cycle Time Optimization for Multi-Cycle SiO_2 ALD

Although the neural network trained for the dataset of the small range of operating conditions in Section 2.3.2.2 is based on single-cycle simulations, it is demonstrated to be applicable for predicting cycle completion time for multi-cycle simulation when all layers are close to full coverage. In a typical industrial setting, cycle time for the entire deposition process is fixed at given operating conditions. However, cycle time can be reduced based on the knowledge of cycle completion time according to temperature and pressure input from BRANN results. Such

optimization is demonstrated by performing two sets of five-cycle simulations. Specifically, five different pairs of temperature and pressure are chosen for each cycle, as shown in Table 4.2. For each cycle in the first set, a fixed duration of 3.5 seconds is given, as discussed in Section 2.3.2.2, to be industrially practical. In contrast, for the second database the steady-state times predicted by the ANN are given based on the respective operating conditions. Within each cycle, temperature and pressure are kept constant. Both sets result in similar deposited five-layer SiO_2 thin-film with almost full coverage (>99.9%). Therefore, by optimizing operating time from cycle to cycle, we can reduce the surface deposition time by 60% using ANN model result, ignoring the gas-phase development time, which will be analyzed in our future work on the multi-scale CFD modeling. In addition, the cycle-time calculation from ANN is almost instantaneous and is much faster than that from the kMC model, which needs approximately one hour to finish a cycle. Moreover, as the film develops thicker, the kMC model will become even slower. This reduction in calculation time is essential for the purpose of cycle-to-cycle optimization and real-time control.

Temperature (K)	Pressure (Pa)	Predicted Time (s)	Original Time (s)
590	120	1.385	3.5
595	125	0.908	3.5
600	130	0.652	3.5
605	135	0.485	3.5
610	140	0.370	3.5

Table 2.2: Temperature, pressure and predicted time for multi-cycle simulation.

2.4 Conclusions

In this chapter, we developed an integrated framework for first-principles-based microscopic modeling, data-driven modeling and optimal operation of thermal atomic layer deposition (ALD) of SiO_2 thin-films using bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. The performance of the 3D kMC model for the SiO_2 ALD process using fundamental chemical properties from DFT calculations was corroborated by experimentally reported data. The thermodynamic-kinetic competing reaction mechanism was reproduced and a GPC of 1.8 Å per

cycle under according operating conditions lied in the range of experimental growth rates. Then, two ANN-based data-driven models for different ranges of operating conditions were constructed. The derived ANN models enabled us to predict the time to achieve steady-state for the film growth at given temperatures and precursor partial pressures. Specifically, the ANN model based on the dataset of large operating range predicted a feasible boundary of operating conditions, and the ANN model based on the dataset of the small range can be utilized in the cycle time optimization to reduce throughput time of the Si-Cycle by 60%, neglecting the gas-phase development time. As a result, the overall approach holds promise for developing an accurate general model for SiO_2 ALD with only precursor thermodynamic and kinetic properties, seeking proper operating conditions while increasing industrial throughput, and significantly saving time and resources that would otherwise have been spent on the testing and manufacturing of physical reaction chambers. Moreover, the generality of the modeling approach in this work makes it possible to extend the current model to the thermal ALD of other thin-film materials with different precursors.

Chapter 3

Machine Learning-Based Modeling and Operation for ALD of SiO_2 Thin Films Using Data from a Multiscale CFD Simulation

3.1 Introduction

In this chapter, a machine learning-based data-driven model for ALD of SiO_2 is developed. Due to the limitation of computational power and the inherent differences in operational length scales, it is not economical or even feasible to develop a model that describes both the bulk gas-phase and the substrate surface processes using a single simulation method. As a result, multiscale models are often proposed and developed in which different simulation methods and techniques are adopted for gas-phase and surface processes, respectively (e.g., [63]). For the gas-phase model, one approach is to develop analytical or numerical solutions of the mass, momentum, and energy conservation equations with the suitable boundary conditions [49]. Nevertheless, the assumptions in these models, particularly when analytical solutions are sought, make the applicability of the

results rather limited as they fail to be applicable to industrial-scale ALD systems, where complex reactor designs are often used to enhance species transport. Therefore, computational fluid dynamics (CFD) modeling may be employed instead and it has been demonstrated to be capable of computing highly accurate gas-phase profiles for complex reactor geometries [24, 64]. On the microscopic scale, surface reactions are the predominant processes that determine film deposition. Molecular Dynamics (MD) has traditionally been used to simulate a variety of molecular-scale microscopic events, but again it cannot be applied to simulate the industrial-size ALD system. More recently, the kMC method, which statistically determines and tracks the probability of each event instead of each particle, is proven to be able to efficiently and accurately represent ALD systems, and thereby it is a good choice for the simulation of the microscopic film-growth activities [22, 24, 27, 65]. Therefore, utilizing and combining distinct methods for the macroscopic (gas-phase) and microscopic (film growth) simulations, we can accurately characterize the ALD system as a whole. Recently, we have investigated the thermal ALD process of SiO_2 adopting BTBAS as the precursor [63]. Specifically, using a standalone 3D surface deposition kMC model, we successfully reported a growth rate that lies within the experimental range (1.4 Å $\sim$ 2.1 Å per cycle); and in [66], we created a 3D multiscale CFD model integrating the CFD and the kMC model, which correctly predicted the half-cycle time needed under different precursor inlet flow rates and reduced the half-cycle time needed by 39.6%, following the ALD chamber geometry optimizations.

Although the multiscale CFD model provides valuable insights to the dynamics of the ALD process, its solution is computationally time-intensive and thus infeasible to be applied in the context of industrial on-line operational optimization. Data-driven modeling adopting machine learning techniques, especially neural networks, has recently received attention in the field of deposition simulations. [67] and [68] have employed an artificial neural network (ANN) to characterize the result of kMC simulation for lattice microscopic structures. [69] and [31] have looked into ANN formulations of multiscale deposition simulation on a 2D geometry. However, no machine learning-based formulation has yet been investigated for an industrial-scale ALD system.

Motivated by these considerations, in this work, we first construct a database using the previously developed multiscale CFD model that incorporates the gas-phase CFD model, the microscopic kMC model, and the multiscale workflow [66]. Then, a multiscale data-driven model is developed, which consists of a linear parameter varying model approximating the gas-phase CFD model and a feed-forward ANN approximating the microscopic kMC model. By fully parameterizing the film deposition rate in terms of input variables that can be manipulated, measured or estimated in real-time, we are able to preserve the model fidelity of key input-output relationships, while greatly enhancing the computational speed. Thereby, the developed multiscale data-driven model may lead to a significant economic benefit by allowing on-line prediction of the minimum cycle-time for full coverage. By contrast, in order to conduct a thorough research on operating conditions with experiments, the precursor alone could cost about 4.08 million dollars, in addition to the operational cost for heating and running the reactor. For some other precursors, such as BDEAS, this number could be doubled due to their smaller production scale.

3.2 Multiscale CFD Modeling, Data Collection and Data-Driven Model Construction

This section summarizes the process description of the SiO_2 thermal ALD deposition, the set-up of the multiscale CFD model, the collection and pre-processing of simulation data, as well as the formulation of the multiscale data-driven model. The following discussion consists of four subsections. First, the macroscopic gas-phase CFD model is introduced, including the geometry development and the tuning of the gas-phase transport model. Next, the microscopic deposition model with the kMC algorithm is discussed. Then, the DFT calculations for the important process-related thermodynamic parameters are briefly reviewed. Subsequently, the multiscale data-driven model, combining a macroscopic linear parameter varying model and a microscopic ANN model, is presented and evaluated. Finally, the simulation data collection procedure adopting the parallel process message passing interface (MPI) scheme and the optimization handling and

noise elimination are introduced.

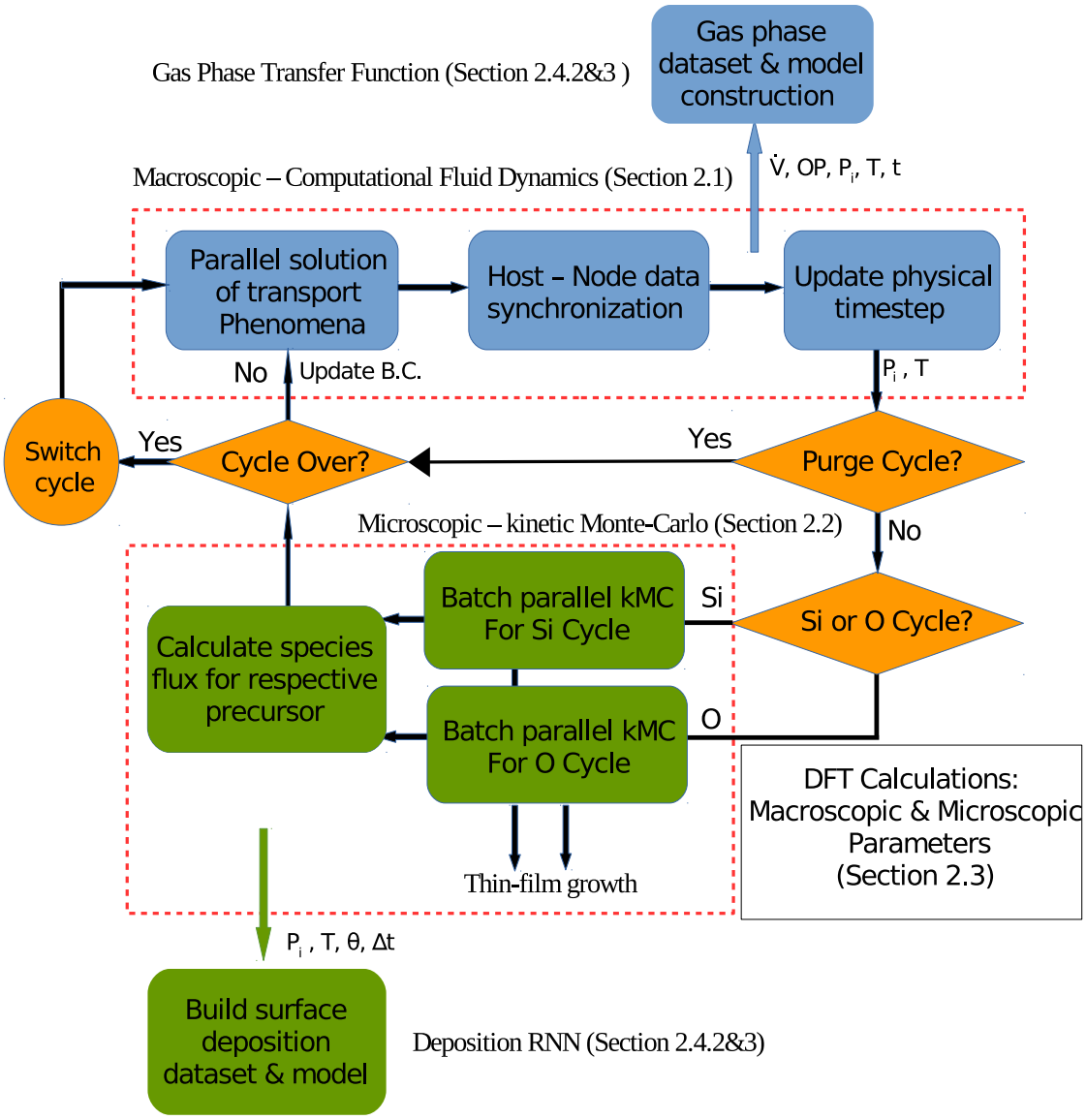


Figure 3.1: Multiscale workflow, parallel processing and information exchange illustration.

Figure 5.1 shows the automated process workflow and serves as a guide for the following sections. The blue region is the gas-phase CFD model domain, the green region is the microscopic kMC model domain, and the orange region is the cross-linking MPI domain. At the beginning of each simulation time step, the CFD simulation is carried out on multiple computational cores in parallel. Next, the reconciled CFD result is passed to the microscopic kMC model, which is

also performed in distributed computation, and its results are fed back to the CFD model as the physical time of the system is updated. The data inter-communication is crucial to the fidelity of the multiscale CFD model and is explained in detail in [66] as well as the data collection in both microscopic and macroscopic domains.

3.2.1 CFD Model for Gas Phase

The ALD deposition profile during the short half-cycle pulse time is closely related to the transient gas-phase transport phenomena. Based on the reactor design of EmerALD XP by ASM International, which is one of the standard industrial cylindrical designs for showerhead reactors, [66] performed a series of geometry optimizations that facilitated the gas profile development and improved distribution and deposition rate spatial uniformity, including a redistribution of the showerhead holes and sizes and a chamber upstream modification, shown in Figure 3.2 (a) and (b). These improvements successfully decreased the half-cycle time required by 39.6%.

Specifically, the optimized reactor chamber accommodates the industrial standard 300 mm wafer [70]. As shown in Figure 3.2, precursor gas comes into the reactor from the small circular inlet, labeled 1, which is 10 mm in diameter. In the upstream region, labeled 2, a horn-shaped upstream design reduces the dead-zone volume in the reactor and leads to a more uniform flow profile. Next, precursor gas flows through the showerhead region, labeled 3, which consists of a showerhead panel specifically tailored for the thermal ALD process [66]. In this CFD simulation, the outlet, labeled 4, omits the final downstream region, which does not contribute to the deposition on the substrate surface, labeled 5. It is important to construct the model in 3D for the ALD reactor simulation because the complex showerhead design is not symmetric in the azimuthal direction and cannot be fully captured by a 2D axis-symmetric model.

The 3D geometry of the ALD reactor is constructed in AUTOCAD and is meshed using the ANSYS meshing tool. As demonstrated in [66], a hybrid mesh, combining structured and unstructured mesh, can produce high mesh quality, reduce computational demand, and save labor time. A final mesh with 3 million cells is chosen based on the mesh independence test [71].

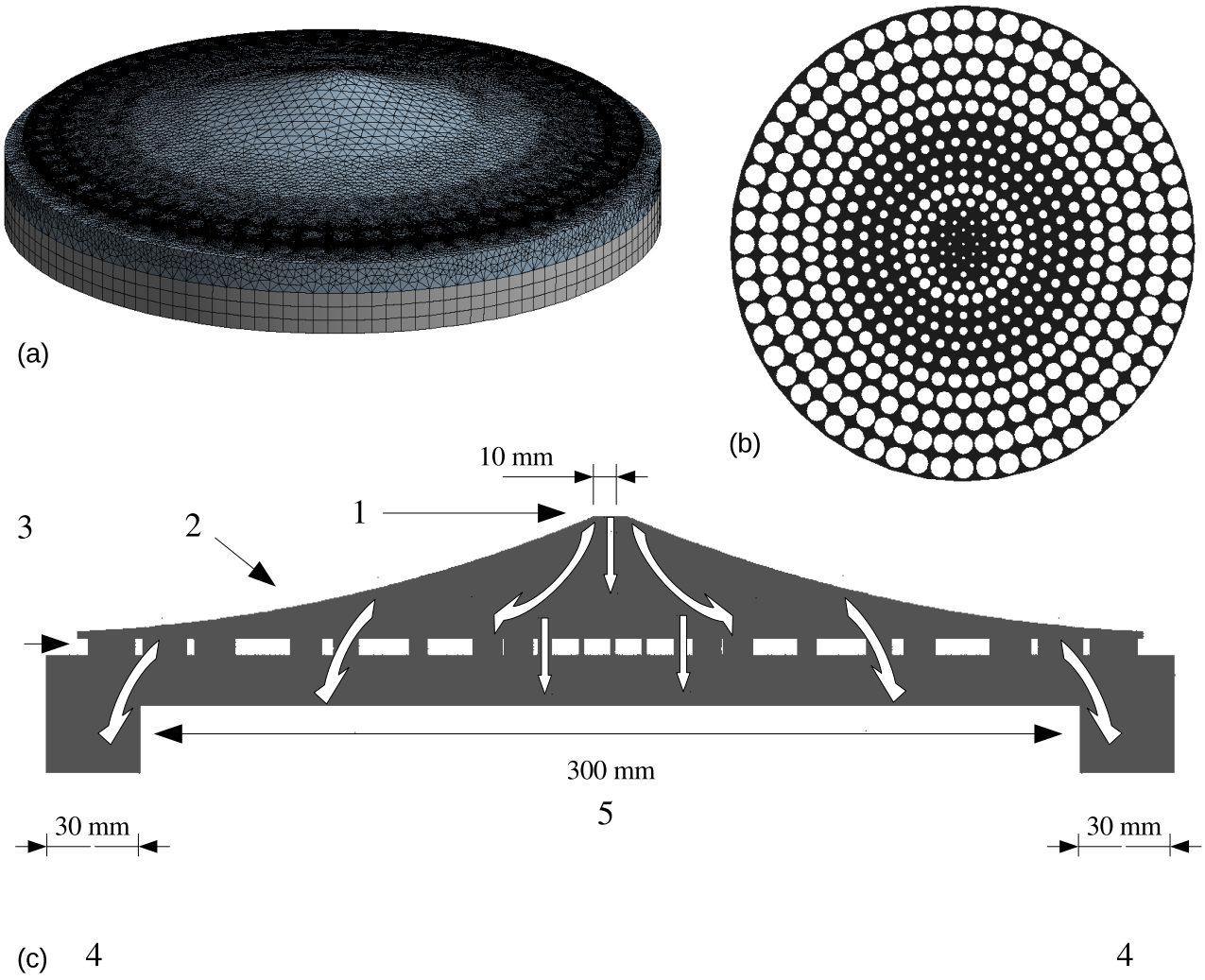


Figure 3.2: (a) Hybrid reactor mesh. (b) Optimized radially distributed showerhead configuration. (c) x-cross-section geometry of the optimized ALD reactor. The gas inlet is labeled 1; the upstream region is labeled 2, where the gas-profile gradually develops; the showerhead holes is labeled 3; the gas outlet is labeled 4, and the substrate surface is labeled 5. The white arrows stand for the overall flow direction.

The meshing quality critically impacts the computational efficiency and accuracy of the CFD simulation, which is generally characterized by the quality, skewness, and aspect-ratio of cell elements shown in Table 3.1. The result shows that the min/max quality of each parameter lies within the acceptable range, and the average value is close to the desired value, denoted by $*$, suggested by the manufacturer ANSYS Inc. [72].

The transient gas-phase CFD calculation is realized using the ANSYS Fluent pressure-based

Table 3.1: Cell quality of the hybrid reactor mesh

	Acceptable Range	Resulted Mesh Bound; Average
quality	0 - 1*	0.33; 0.79
skewness	0* - 0.95	0.74; 0.29
aspect-ratio	1* - 100	6.49; 1.98

solver [72]. Three important gas species, BTBAS, ozone, and argon, are included using the Fluent Mixture Template. The thermodynamic properties of ozone and argon are generated from the Fluent database, while those of BTBAS are acquired from DFT calculations, which will be explained in section 4.2.5. The mixture is defined as an incompressible ideal gas, where Fluent calculates the gas density from ideal gas law using the operating pressure, and applies the assumption of constant density with respect to velocity during the flow field calculation. These assumptions are appropriate for a low pressure and high temperature thermal ALD reactor with Mach number < 0.3 under the maximum inlet flow rate and operating pressure [73]. The operating and time-dependent boundary conditions for the inlet, outlet and substrate surface, accounting for the flow rate, species compositions, temperature, pressure, consumption of precursors from microscopic kMC deposition, are defined using the built-in gas properties and the user-defined functions (UDFs) as explained in [66]. Second-order finite volume methods are used to numerically solve the flow field governing momentum, mass, and energy transports equations in each mesh grid as shown below [62, 72]:

$$\overline{\overline{\tau}} = \mu\left[\left(\nabla\overrightarrow{v} + \nabla\overrightarrow{v}^{T}\right) - \frac{2}{3}\nabla\cdot\overrightarrow{v}I\right] \tag{3.1}$$

$$\frac{\partial(\rho\overrightarrow{v})}{\partial t} + \nabla\cdot(\rho\overrightarrow{v}\overrightarrow{v}) = -\nabla P + \nabla\cdot\left(\overline{\overline{\tau}}\right) + \rho\overrightarrow{g} + \overrightarrow{F} \tag{3.2}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = \nabla(k\nabla T - \Sigma h\vec{J} + (\overline{\overline{\tau}}\vec{v})) + S_h \tag{3.3}$$

$$\vec{J}_i = -\rho D_{m,i}\nabla Y_i - D_{T,i}\frac{\nabla T}{T} \tag{3.4}$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla\cdot(\rho\vec{v}Y_i) = -\nabla\cdot\vec{J}_i + R_i + S_i \tag{3.5}$$

where $\overline{\overline{\tau}}$ is the stress tensor, μ is the fluid viscosity, $\overrightarrow{v}$ is the velocity of gas-phase species mixture, I is the unit tensor, ρ is the density of the gas-phase species mixture, P is the static pressure, $\overrightarrow{g}$ is the acceleration of gravity, $\overrightarrow{F}$ is the force, E is the fluid internal energy, T is the fluid temperature, k is the heat conductivity R_i and S_i are the reactions and mass transfer source terms of species i, respectively, S_h is heat transfer source term, Y_i is the user-defined transport input, $\overrightarrow{J_i}$ is the diffusive flux of species i, and $D_{m,i}$ and $D_{T,i}$ are the mass and heat diffusivities of species i, respectively.

The calculation time step is determined by the Courant-Friedrichs-Lewy (CFL) condition, which relates the discretization time to the speed of information traveling in the physical space and the discrete physical length scale [74]:

$$\Delta t = \frac{C\Delta x}{u} \tag{3.6}$$

where C is the selected Courant number, a quantitative description of the number of computational cells that the fluid moves through in one time step, Δx is the minimum cell length and u is the average flow velocity. Under all experimental conditions, Fluent converges with a Courant number of 50 efficiently and accurately [72].

3.2.2 KMC Model for Film Growth

In addition to the macroscopic CFD model that describes the gas-phase development profile, the microscopic simulation utilizes the kMC method to model the reaction mechanisms and the structural details of the SiO_2 thin-film. Each kMC model has a large enough dimension of 1200×1200 sites per layer as well as a height that equals the number of cycles simulated, which is computationally efficient and yields lattice-size independent results [35, 36]. A 3D triangular lattice is adopted to approximate the true α-Quartz SiO_2 lattice as shown in Figure 5.4 (a) and (b), which is demonstrated to preserve the fidelity of the key features in the microscopic structure [63].

As mentioned in the introduction, a typical ALD process contains two half-cycles. The first half-cycle is the Si-Cycle, which consists of BTBAS physisorption and two-step dissociative

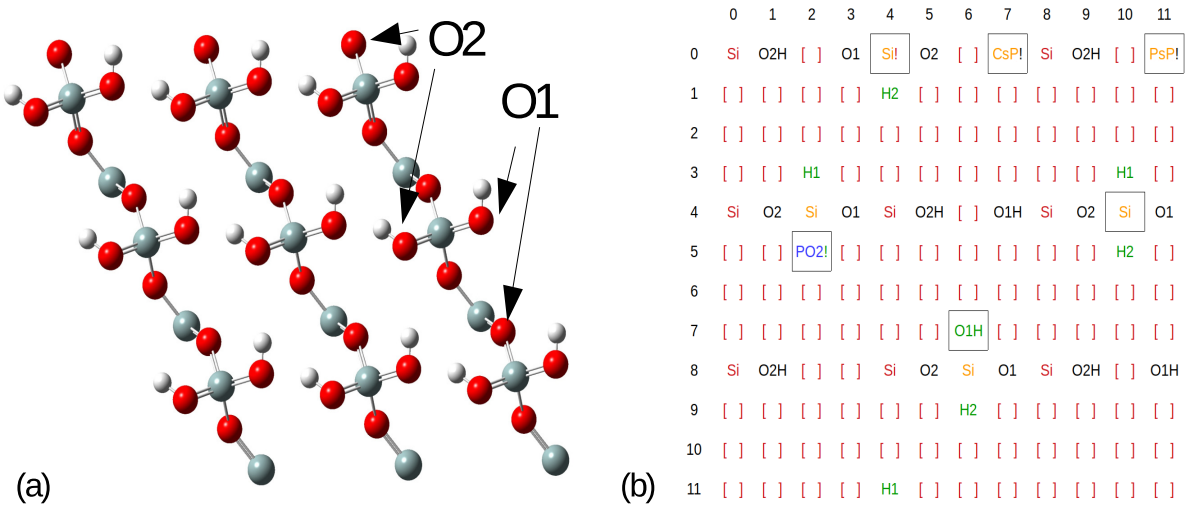


Figure 3.3: (a) Top view of the hydroxylated $SiO_2(001)$ surface, where O1 is the more electronegative oxygen (b) Top view of a 5-layer 12×12-site miniature demonstration of the full kMC simulation lattice. The five layers and the species on the lattice are shown using different colors and symbols, respectively. The first (bottom) layer, labeled red, contains the base Si atoms. The second layer, labeled black, contains bare or hydrogenated oxygen atoms. The third layer, labeled orange, contains the species from the first silicon half-cycle: Si is the neighbour-binding silicon, Si! is the self-binding silicon, and PsP and CsP are the physisorbed and chemisorbed precursors, respectively. The fourth layer, labeled green, contains the species from the first oxygen half-cycle: O and OH are the oxygen atoms and hydroxyl groups. The fifth (top) layer, labeled blue, contains the physisorbed ozones (PO1 and PO2), which are going to oxidize.

chemisorption and abstraction. In the physisorption, the precursor forms a strong H-bond with the more electronegative oxygen atom from the surface O-H groups. Then, the physisorbed precursor sequentially releases two aminoethyl groups to fully attach the Si atom on the substrate, leaving two Si-H bonds as the new surface. It is important to note that, during the second chemisorption, the precursor can either choose an O-H from the same or neighboring Si, leading to self-binding and neighbor-binding final product, respectively. The second reaction pathway is more kinetically favorable, but the first reaction pathway retains the original surface orientation. In addition, the chemisorbed precursor can potentially go through the reverse reaction and get abstracted from the surface. Similarly, as shown in Figure 5.5 (c), the second half-cycle is the O-cycle consists of ozone physisorption, ozone abstraction and two individual surface oxidation reactions. After the O-cycle, the surface will again be terminated by hydroxyl groups. The detailed reaction mechanism can be

found in [43].

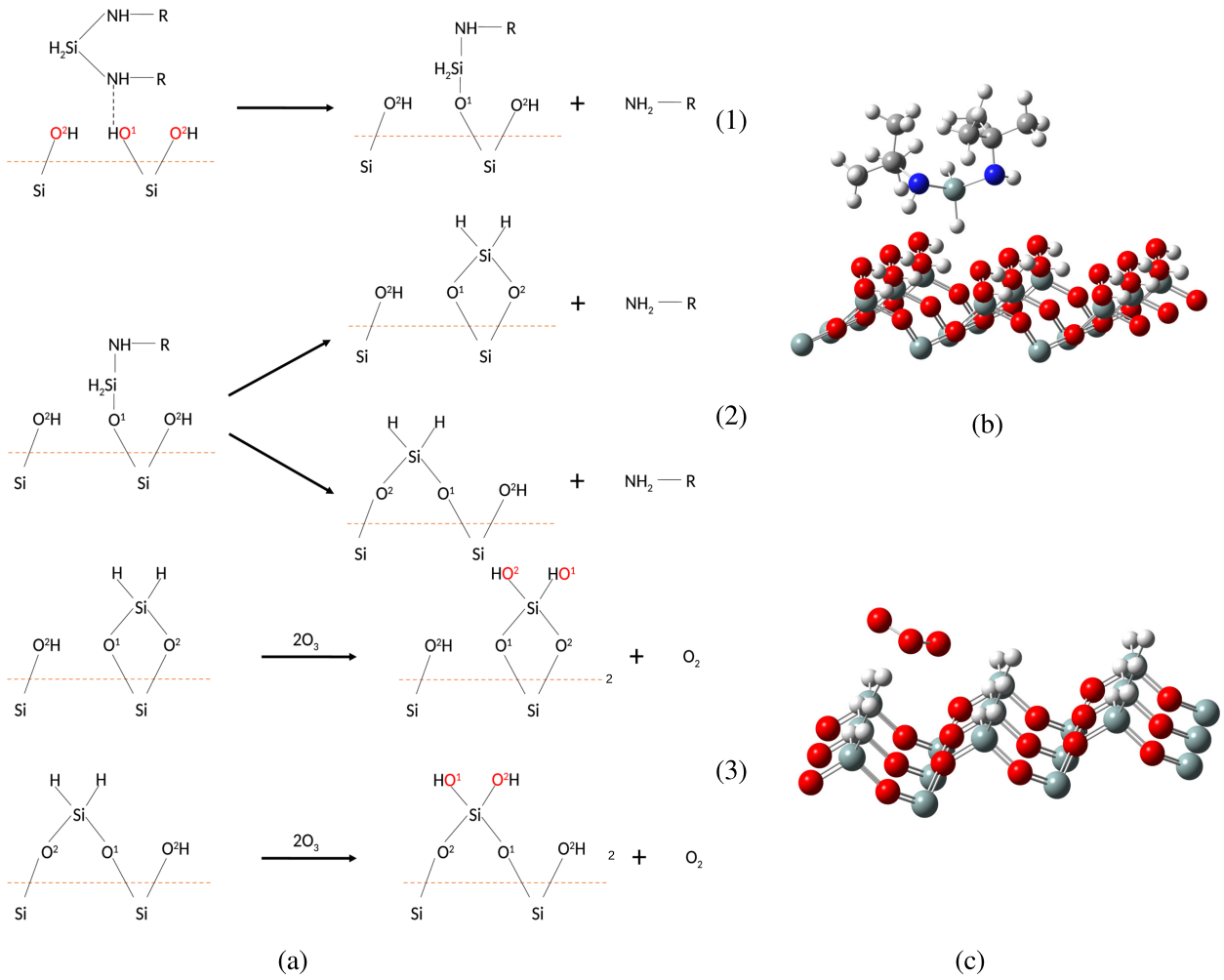


Figure 3.4: (a) Surface reaction mechanism: (1) First dissociative chemisorption step of BTBAS. (2) Second dissociative chemisorption step of BTBAS under self-binding and neighbour-binding mechanisms. (3) Oxidation of self-binding and neighbour-binding SiH_2 with ozone.(b) Physisorped BTBAS molecule demonstration. (c) Physisorped ozone molecule demonstration.

In reality, all reactions are simultaneously competing with each other and the kinetic rates for the above reactions are crucial for the kMC algorithm formulation. The physisorption rate can be described by the collision theory, which is applicable to gas-surface athermal barrierless processes, as follows:

$$r_{phs} = \frac{p_i}{RT}\sqrt{\frac{8k_bT}{\pi m_i}}s_{c,i}N_a\sigma \tag{3.7}$$

where r_{phs} is the physisorption reaction rate, p_i is the partial pressure of the species i, R is the ideal gas constant, T is the temperature, k_b is the Boltzmann constant, m_i is the mass of the molecule for species i, $s_{c,i}$ is the sticking coefficient of the species i at given temperature, N_a is the Avogadro number, and σ is the average area per surface site.

For thermodynamically activated reactions like chemisorption, abstraction, and oxidation, the transition state theory is adopted to characterize their rates [47]. The reaction rate can be described by an Arrhenius-type rate law of the equilibrated transition state complex as follows:

$$r_{rxn,i} = f_i^{\mathrm{TST}} \Big(\frac{k_b T}{h}\Big) \exp\Big(\frac{-E_{a,i}}{k_b T}\Big) \tag{3.8}$$

where $r_{rxn,i}$ is the reaction rate of the i^{th} thermodynamically activated reaction, $E_{a,i}$ is its activation energy for the transition state complex, k_b is the Boltzmann constant, h is the Plank's constant, T is the temperature, and f_i^{TST} is the ratio of the vibrational partition function between the transition state complex and the reactant.

Adopting the above rate equations, an n-fold hybrid kMC model can be utilized to determine event selection and time evolution. Specifically, the reaction rate of an individual reaction is normalized with respect to the total reaction rate, r_{total}, which is defined as:

$$r_{total} = \sum_{i=1}^{N} r_i \tag{3.9}$$

where r_i represents the respective rate of each event within an event set, which consists of total N events. The normalized indicator of the i^{th} event, $l_i \in (0,1]$, can be interpreted as the sum of the normalized probabilities of the first i events:

$$l_i = \frac{\sum_{j=1}^{i} r_j}{r_{total}}, i = 1, ..., N \tag{3.10}$$

Then, a uniformly distributed random number, $\gamma_1 \in (0,1]$, is generated. If γ_1 lies within the normalized interval of l_{i-1} to l_i, the corresponding i^{th} event will be chosen for execution in the

kMC algorithm.

In addition, given the chosen event, the kMC algorithm calculates the time evolution Δt of the physical process with the following equation:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{3.11}$$

where $\gamma_2 \in (0,1]$ is a second uniformly distributed random number.

Standard n-fold kMC can generally be adopted directly to the O-Cycle events. However, for the Si-Cycle, due to the competition between the kinetic and thermodynamic favorability of potential reaction pathways in the Si-Cycle, we need to take the forward and the reverse reaction of the sequential chemisorption into account. As a result, we distinguished physisorption and abstraction from the rest of the reactions in the Si-Cycle and adopted a discretized probability calculation process. Thus, a modified kMC model, with decoupled reaction sets, is formulated for the Si-Cycle and is explained in detail in [63].

3.2.3 DFT and Thermodynamic Calculations

As suggested by [75], experimental studies have been performed on a majority of the popular ALD precursors. However, as the experimental determination of the kinetic information is very difficult, most of the activation energies determined for the ALD reactions are the apparent equivalent activation energies, which may serve as a general rate estimator, but they are not sufficient for detailed structural simulations. Therefore, many thermodynamic properties of BTBAS that are important for transport phenomena calculation and surface reactions have yet to be determined. In order to derive the multiscale model from first-principles and accurately reproduce the experimental findings, DFT is adopted using the Gaussian09 software packages, which is one approach to approximate the solution of the complex Schrödinger equation for a many-body system

by solving the overall electron density [34, 76–78]:

$$\widehat{H}\Psi(r_i, R_I) = E\Psi(r_i, R_I) \tag{3.12}$$

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

For the CFD simulation, specific heat C_p, standard state entropy $S^o(298K)$, standard state enthalpy of formation $\Delta_f H^o(M, 298K)$, and viscosity μ are needed for the transport equations. Gaussian-4 (G4) theory, which is a highly accurate calculation procedure for the first three row elements, is adopted for calculating such parameters. In the G4 theory, a sequence of single-point energy calculations is performed using the appropriate spin-orbit, basis set, and the advanced higher level correction [79]. Specific heat C_p and the standard state entropy $S^o(298K)$ are directly obtained from this method and the determination of the standard state enthalpy of formation $\Delta_f H^o(M, 298K)$ is discussed in detail in [80]. Furthermore, the viscosity μ is approximated using the molecular theory of Newtonian gases [81].

For the surface deposition model, the vibrational partition coefficient is crucial to the determination of the abstraction reaction kinetics. The transition-state complexes of the abstracted BTBAS need to be calculated using the "optimization to transition state" (Opt=TS) method from an optimized ground-state $3 \times 3 \times 1$ SiO_2 surface lattice and a BTBAS molecule. Those two molecules are optimized to 6-31G+dp accuracy level using the Becke, 3-parameters, Lee-Yang-Parr (B3LYP) functional [52, 53]. Then, the G4 theory is again used to acquire an accurate vibrational frequency. The results, along with other important thermodynamic parameters are reported in [63].

3.2.4 Multiscale Data-Driven Model Construction

With the multiscale CFD model, it is possible to calculate the temporal deposition profile under different operating pressures, inlet precursor flow rates, and surface heating temperatures, which reveals valuable information of the thermal ALD process. However, a complete half-cycle

simulation of the multiscale CFD model takes 5 - 10 days to finish depending on the operating conditions, utilizing the parallel computation workflow with 36 cores on the UCLA Hoffman2 Cluster [72]. As a result, the prediction from the multiscale CFD model is deemed too time-consuming to be implemented for on-line operational optimization in industry. In order to speed up the computation while retaining the key process information, a multiscale data-driven model that accurately reproduces the results of both microscopic and macroscopic domains is developed and validated in this work.

3.2.4.1 Multiscale Data-based Model Construction

As mentioned above, each full multiscale CFD simulation is computationally intensive due to the communication overhead between the two simulation domains and the allocation of computational resources on the computation cluster. Therefore, it would be impossible to incorporate both domains into one single integrated data-driven model since it would be too time-consuming to generate a sufficiently large database to train a model that characterizes the actual process dynamics. As an alternative, a decoupling scheme is proposed to separately describe the macroscopic (gas phase) domain and the microscopic (film growth) domain. The decoupled multiscale data-driven model allows a quicker data collection while preserving the model fidelity to key process features.

For the microscopic deposition model, as introduced in the kMC section, its explicit inputs, which are generated by the macroscopic CFD model, include the surface temperature and the surface partial pressure of the half-cycle precursor. Additionally, to accurately predict the transient deposition rate, the surface configuration, which is predominantly represented by the surface coverage, also needs to be incorporated into the inputs for the microscopic data-driven model. Due to the non-linear and non-convex relationship between the microscopic domain input-output parameters, a traditional least square method or system identification method fails to provide satisfying regression result. Thus, feed-forward artificial neural network (ANN) is utilized to characterize this complex microscopic model. Among various training algorithms, it is observed

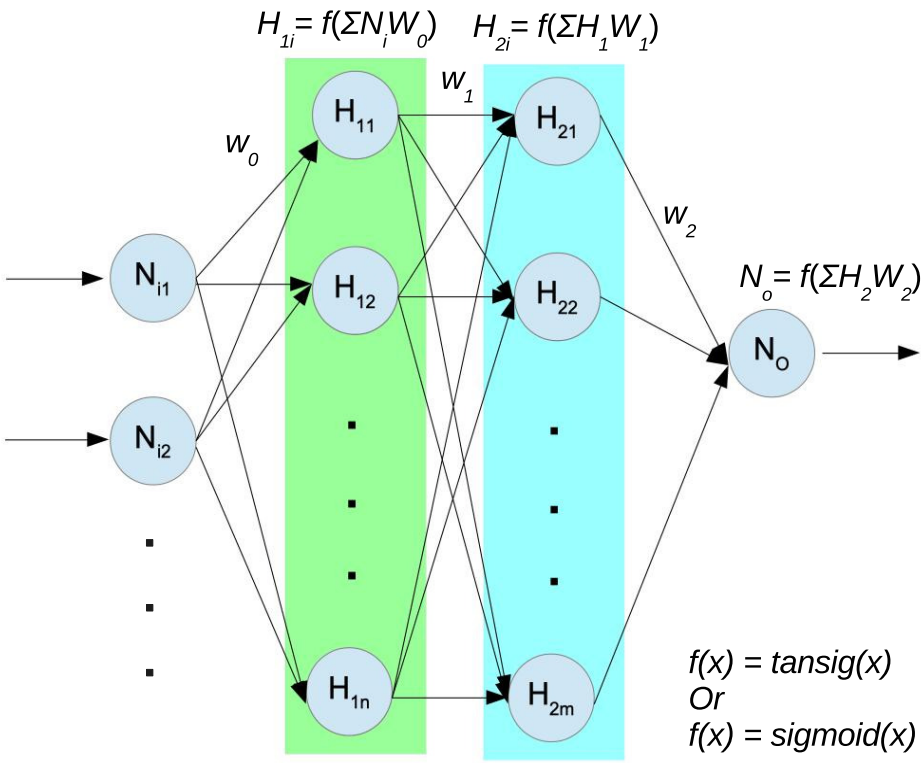


Figure 3.5: Structure of a feed-forward artificial neural network (ANN) with multiple inputs, two hidden layers, and one output. *W* is the weight for each ANN layer, *N*s and *H*s represent the neurons, and *f* represents the activation function, which can be tansig or sigmoid function based on the neural network structures being used.

that the traditional gradient descent algorithm and the more recent RMSprop algorithm both encounter difficulties in determining the appropriate amount of hidden layers and neurons. In addition, despite the popularity of the deep learning network, it suffers from the dilemma of balancing between the overall model accuracy and the over-fitting for solving a regression problem with high precision demand. Thus, a Bayesian regularized artificial neural network (BRANN) is adopted as an alternative structure by incorporating the probability theory into the algorithm, where the Bayes' theorem is used to calculate the probability of a current event based on its associated prior knowledge and conditions. This probability inference successfully makes BRANN more robust than the standard deep neural networks because the Bayesian regularization algorithm is formulated as a rigid regression model, which is a well-posed statistical problem for complicated non-convex and non-linear scenarios. BRANN also has the ability to systematically turn off

weight parameters that are less relevant in the training process, which is similar to the usage of drop-out regularization in a standard ANN. In addition, BRANN avoids the over-fitting problem by using the Occam's razor principle and incorporating evidence procedures under the Bayesian criterion [58]. In our model, the MATLAB machine learning toolbox is used for the BRANN construction, training and validation. The general structure of a multiple-input-single-output (MISO) feed-forward neural network with two hidden layers is shown in Figure 3.5, where H_{ij} are the hidden layer neurons, N_{in} are the input neurons, N_o is the output neuron, and W_i is the weight for the ith ANN layer. Specifically, in this neural network, the input layer contains three neurons, representing the surface temperature, the precursor partial pressure, and the current surface coverage, respectively. The output layer contains one neuron, representing the transient deposition rate. The neural network calculates the activation using feed-forward propagation:

$$A^l = \theta(W^l A^{l-1} + B^l) \tag{3.13}$$

where A^l is the activation matrix, W^l is the weight matrix, and B^l is the bias matrix for layer l, respectively. θ represents the activation function, which is the hyperbolic tangent sigmoid function (*tansig*) in the BRANN. It is a good candidate under Bayesian regularization because of its performance in the statistical inference [60]:

$$tansig(x) = \frac{2}{1+e^{-2x}} - 1 \tag{3.14}$$

Additionally, for probability update in BRANN backpropagation, the hyperparameters α and β, chosen from a uniform random distribution, are added to the standard mean square error (MSE) function, $S(w)$, which is typical for regression applications:

$$S(w) = \beta \sum_{i=1}^{N_D} [y_i - f(x_i, w)]^2 + \alpha \sum_{j=1}^{N_w} w_j^2 \tag{3.15}$$

where N_D is the number of data samples in the training dataset, w is the weight vector for all

hidden layers, which is assumed to have a Gaussian distribution, N_w is the number of weight parameters, w_j is the j^{th} entry in the weight vector, y_i is the desired output value, and $f(x_i, w)$ is the predicted value dependent on w and the input x_i. In order to compute the optimal w, a sequence of optimization problems are solved using the Levenberg-Marquardt algorithm [59]. Then, hyperparemeters α and β are updated using Bayesian inference calculation. The detailed explanation and construction of the optimization problems are discussed in detail by [60] and [58]. To further avoid saturation and high variance activation, batch normalization is applied between layers, thereby facilitating convergence speed and learning rate [82].

For the macroscopic model, at each transient time step, a convex relationship is observed between the input parameters (i.e., feed flow rate and operating pressure), and the output parameter (i.e., surface partial pressure). Also, it is demonstrated that the precursor usage feedback from the microscopic kMC model at different inlet flow rates does contribute to the gas phase development rate but does not significantly alter the overall shape of the gas phase response curves. Thus, a system identification method using a linear differential equation model with parameters dependent on operating pressure and inlet precursor flow rate (linear parameter varying model) is adopted to characterize the gas-phase profile. Under each set of operating conditions executed, a third-order linear parameter varying model is fitted to describe the partial pressure temporal development over time:

$$a_3(v,P_{op})\frac{d^3y}{dt^3} + a_2(v,P_{op})\frac{d^2y}{dt^2} + a_1(v,P_{op})\frac{dy}{dt} + a_0(v,P_{op})y = K(v,P_{op})v \tag{3.16}$$

$$y(0) = 0, \frac{dy}{dt}(0) = 0, \frac{d^2y}{dt^2}(0) = 0, \frac{d^3y}{dt^3}(0) = 0 \tag{3.17}$$

where y is the surface precursor partial pressure, and $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$, and $\frac{d^3y}{dt^3}$ are its first, second, and third-order time derivatives. K, a_0, a_1, a_2, a_3, are the gain and other model parameters, respectively, which are all functions of operating pressure P_{op}, and inlet precursor flow rate v, which is also the step input. Third-order linear parameter varying model is selected because it was found to be the best fit to our data, while lower order models failed to provide enough degrees of freedom to model the data. After we determined the parameters, which is dependent on v and P_{op}, from the

known datasets, the unknown conditions can be estimated with interpolation with respect to the input parameters.

3.2.4.2 Data Collection and Processing

According to our computational resource, on average one macroscopic model simulation or 3000 microscopic simulations can be completed within 200 hours. For the microscopic surface model, a large database under various fixed input conditions is constructed in order to capture the complex input-output relationship. The output of the model is the transient deposition rate, which depends on the surface temperature, precursor partial pressure and the transient film coverage. The surface temperature ranges from 475 K to 625 K with an interval of 1 K, partial pressure ranges from 30 Pa to 200 Pa with an interval of 1 Pa, and each set of inputs dynamically runs 400 time steps with a time step size of 0.05 s. Moreover, staring from zero, the transient film coverage is the accumulation of total precursor usage until the last time step. A raw dataset containing over 10 million data points is collected. Two steps are carried out to preprocess the dataset for the purpose of better fitting. First, the microscopic deposition steady-state identification is executed. With the term microscopic deposition steady-state, we refer to the case where an equilibrium is reached between the rates of material deposition into the field and the rate of abstraction (desorption) of material from the film to the gas phase. With high enough temperature and pressure, at the microscopic deposition steady-state, the deposited film achieves full coverage. However, under some other operating conditions, the deposition reactions will be slower than the abstraction reactions at certain coverages, thereby causing non-full coverages and non-uniformities of the films when the steady-state deposition rates are reached.

Locating the microscopic deposition steady-state not only reduces the amount of data needed for training of the ANN because the deposition rate after steady-state is reached always fluctuates around zero, and thus can be discarded, but also reduces the neural network fitting complexity by eliminating a non-differentiable turning point in the majority of the data. For each operating condition, the simulation data is fitted by a two-segment function described in Equation 3.18,

where the first segment represents the pre-steady-state process, fitted using the step-response of a first-order linear model, and the second segment represents the profile after achieving steady-state, fitted by a straight horizontal line:

$$y = \begin{cases} b_1\Big(1 - \exp\big(-\frac{b_2}{t}\big)\Big) & t \le b_3 \\ b_4 & t \ge b_3 \end{cases} \tag{3.18}$$

where b_1 and b_2 are the gain and the time constant of the first-order step-response function, respectively, b_3 represents the time in which the microscopic deposition steady-state occurs and b_4 represents the microscopic deposition steady-state surface coverage. It is noteworthy that the choice of a first-order step-response linear function as our function prototype is solely due to its flexible shape rather than proving an explicit relationship between deposition time and film coverage.

In order to accurately determine the values of the four parameters (b_1 to b_4), the deviation between the fitted function and the data is minimized through an optimization problem described as follows:

$$\begin{aligned} \min_{b_1,b_2,b_3,b_4} \quad & \sum_{i=1}^{N} (y_i - D_i)^2 \\ \text{s.t.} \quad & b_1\Big(1 - \exp\big(\frac{-b_2}{b_3}\big)\Big) \le b_4 \end{aligned} \tag{3.19}$$

where N is the total number of data points provided, y_i and D_i represent the corresponding fitted value and the provided data. This constraint ensures that there is no overshoot in the fitted function before the second segment, which makes physical sense because the coverage should in general be a monotonously increasing function. This optimization problem is solved in IPOPT using the interior point algorithm, obtaining the optimal parameter b_3 allows determining the time at which the steady-state deposition rate occurs [83].

After the microscopic deposition steady-states are identified for each operating condition, the first segments of all datasets are selected, which in total consist of about 2.5 million data points. Further noise reduction is performed by applying a moving average of every five consecutive data

points for each dataset. The smoothened datasets are then merged for the feed-forward ANN. The final dataset is divided into training, validation, and test data under the ratio of 7:1.5:1.5. The training dataset is only used to train the ANN model parameters, while the validation dataset is used to validate and improve the result in the training process. The test dataset is randomly chosen and separated from the data used for training in advance and is used to evaluate the final neural network performance.

For the macroscopic gas-phase model, because of the aforementioned convex relationship, a rather small dataset of the multiscale CFD model is needed to characterize the system. Four sets of CFD simulations at different precursor flow rates with and without microscopic kMC feedback workflow are performed to examine the gas-phase development and the influence of surface deposition on the macroscopic domain. After the macroscopic gas-phase model is identified, two more sets of multiscale CFD simulations are carried out for interpolation and validation purposes. Three operating pressures are tested to verify the model independence from operating pressure. Moreover, to determine the linear parameter varying model, a series of step inputs is generated based the CFD time step and the deposition half-cycle time under each precursor flowrate.

3.3 Multiscale Data-Driven Model Development and Validation

In this section, the resulting macroscopic linear parameter varying model and the microscopic ANN model are summarized, and the combined multiscale data-driven model is compared to the first-principles based multiscale CFD model. In addition, advantages and applications adopting this multiscale data-driven model are discussed.

3.3.1 Microscopic ANN Model Parameter Determination

In this section, the preprocessing of the training data for constructing the microscopic ANN and the training results are explained. As discussed in Section 5.2.1, the original dataset contains a

large number of data points that are collected after the microscopic deposition steady-states have occurred. Since those data points stay mostly constant regardless of the deposition time, they do not make positive contribution to the data-driven model prediction. Furthermore, a direct attempt to use the entire dataset is not computationally feasible because a complete grid search for the ANN structure with the available resources would take hundreds of hours to finish. By locating the microscopic steady-state using the method mentioned in 3.2.4.2, deposition steady-state times are successfully determined for all sets of temperature and pressure conditions. As shown in Figure 3.6 (a), two random sets of input conditions are plotted and the steady-state in each condition is successfully identified by the prediction model. In addition, there are fluctuating white noises, due to the inherent random nature of the kMC algorithm as introduced in Section 3.2. The white noise associated with the kMC predicted transient deposition rate and the accumulated coverage hinder the neural network fitting accuracy and precision. Figure 3.6 (b) shows that the moving mean algorithm successfully reduces the white noise in the transient deposition rate while retaining the correct mean of the original data. By performing noise reduction and data culling, this algorithm is proven to be highly accurate.

As introduced in Section 5.2.1, the ANN model takes three inputs: surface partial pressure, surface temperature, and current deposition coverage, and it generates one output: transient deposition rate. Two hidden layers, each containing 15 neurons, are chosen for the BRANN model based on a grid search. It is found that a single layer cannot generalize the complex input-output relationship, and that a two-layer network with too many neurons may result in over-fitting error. As shown in Figure 3.7 (a), a mean square error on the training dataset drops by four orders of magnitude and reaches a final performance of 7.2×10^{3} after 1000 epochs. This error corresponds to a normalized error that is less than 1.0×10^{-3} and less than 10% of the transient deposition rate. Figure 3.7 (b) shows that the testing error of the resulting neural network is normally distributed. In addition, Figure 3.7 (c) shows a good correlation (R^2 = 0.99) between the predicted and the simulated transient deposition rate, which indicates that the neural network achieves a good result in fitting the microscopic kMC simulation data.

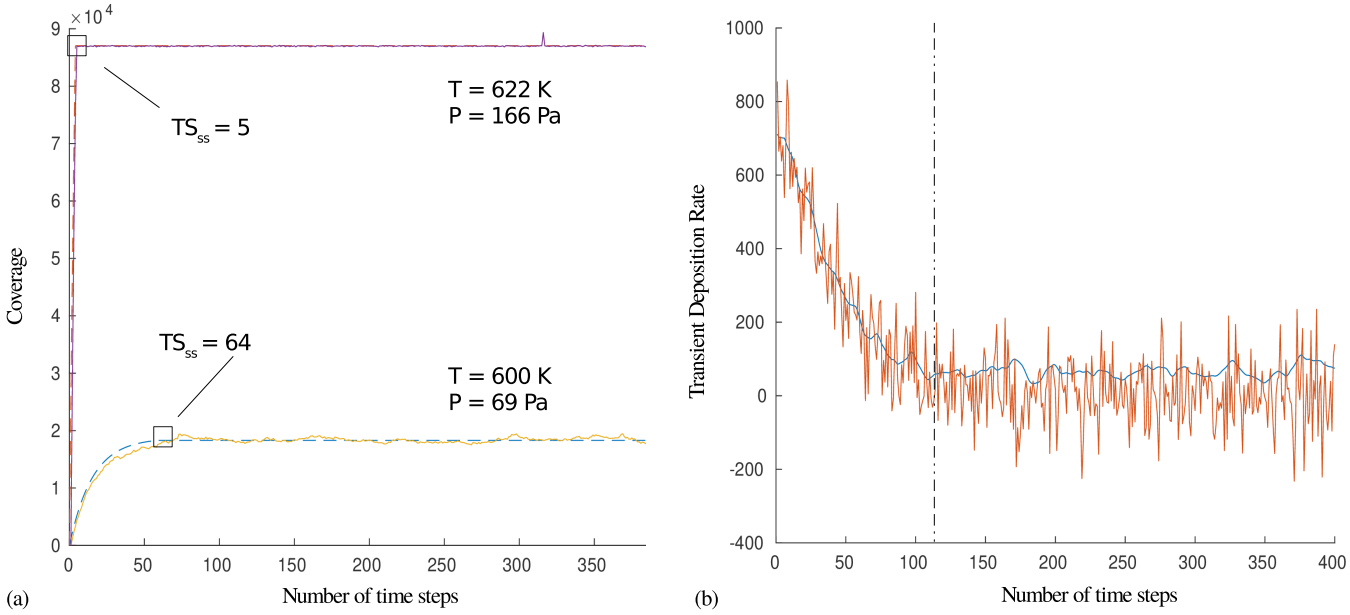


Figure 3.6: (a) Comparison between the kMC result and the steady-state fitting result at two randomly chosen operating conditions (T = 622 K, P = 166 Pa and T = 600 K, P = 69 Pa). The dashed lines are generated from the fitting and the solid lines are taken from the kMC simulation. The black boxes indicate the steady-states identified by the fitting algorithm and TS_{ss} indicates the number of time step when steady-state is reached. The *x*-axis is the number of time steps and the *y*-axis is the surface coverage, which has a maximum of 90000. (b) Comparison between the raw and the smoothened transient deposition rate at a randomly chosen operating condition (T = 570 K, P = 80 Pa). Dashed line indicates the location of the predicted steady-state. The *x*-axis is the number of time steps and *y*-axis is the transient deposition rate.

A visualization of the trained ANN model performance is shown in Figure 3.8. Because the system consists of three inputs and one output, the parameters' relationship cannot be directly plotted in a single figure. Since surface temperature will be kept constant in the multiscale model, we separate the temperature input for illustration purpose. A series of representative temperatures are selected and the results of the other three parameters are examined and plotted for each chosen temperature. The trained ANN is demonstrated to successfully predict the kMC simulation result in the majority of operating regions as most testing data points lie close to the fitted surface. However, there are some tendencies to overfit in the high temperature and high accumulation scenarios. The reason that over-fitting happened is that the reaction rate scales up rapidly with temperature which is indicated by the Arrhenius rate law. The surface reaction reaches steady-state very quickly

at high temperature, which results in a smaller amount of data points for training. Also, due to the stochastic nature of the kMC method, a fast deposition rate gives rise to a larger uncertainty at every time-step. The method of circumventing this over-fitting problem will be discussed in Section 3.2.1, where the multiscale data-driven model is validated.

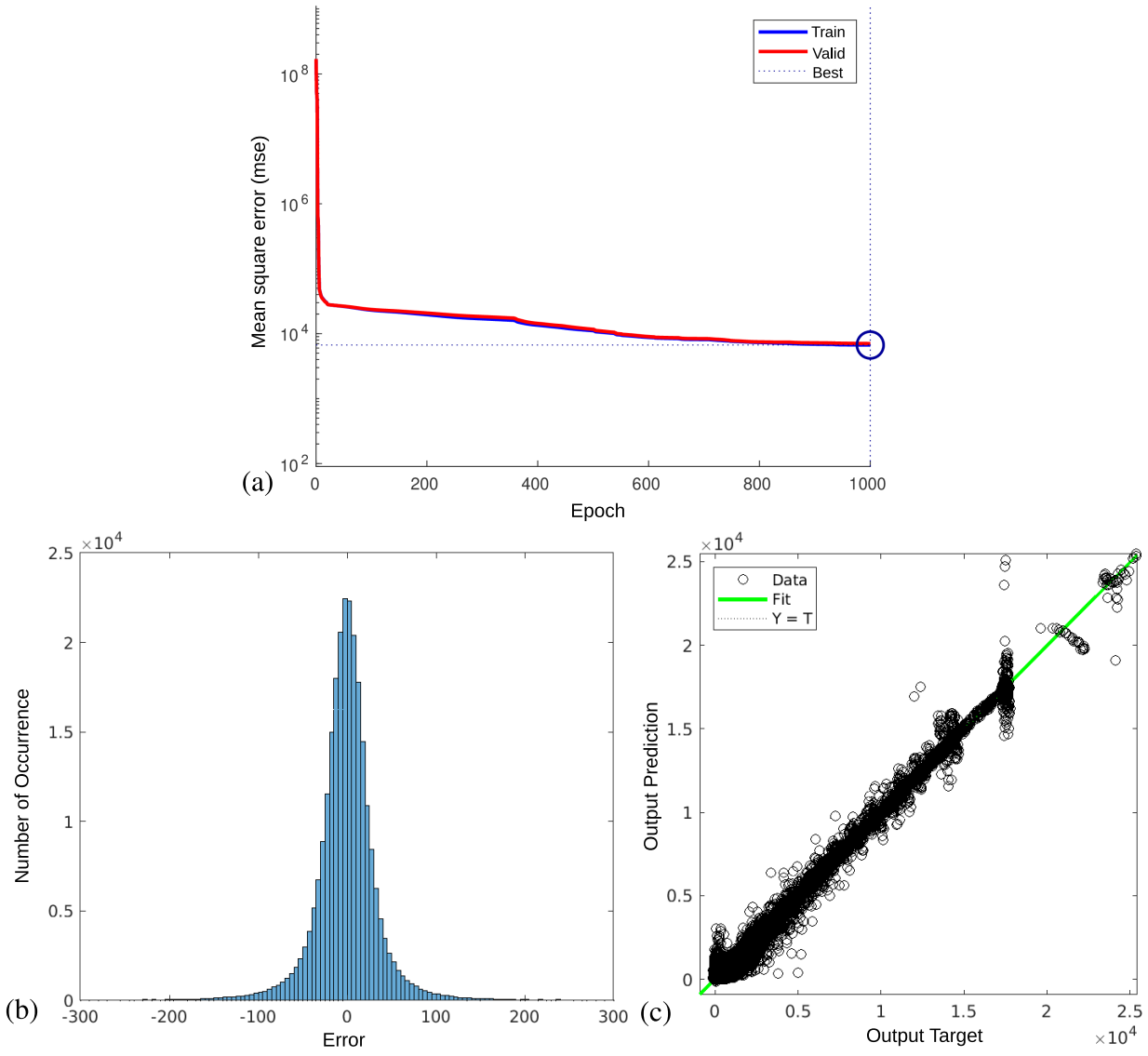


Figure 3.7: Training and performance of the constructed BRANN model: (a) Training and validation history. The x-axis is the training epoch and the y-axis is the MSE on output. (b) Prediction error distribution histogram of testing. The x-axis is the magnitude of error and the y-axis is the number of occurrences of the corresponding error. (c) Correlational accuracy of the predicted and the actual deposition rate of testing. The x-axis is the predicted transient deposition rate from the neural network and the y-axis is the actual rate from the kMC model.

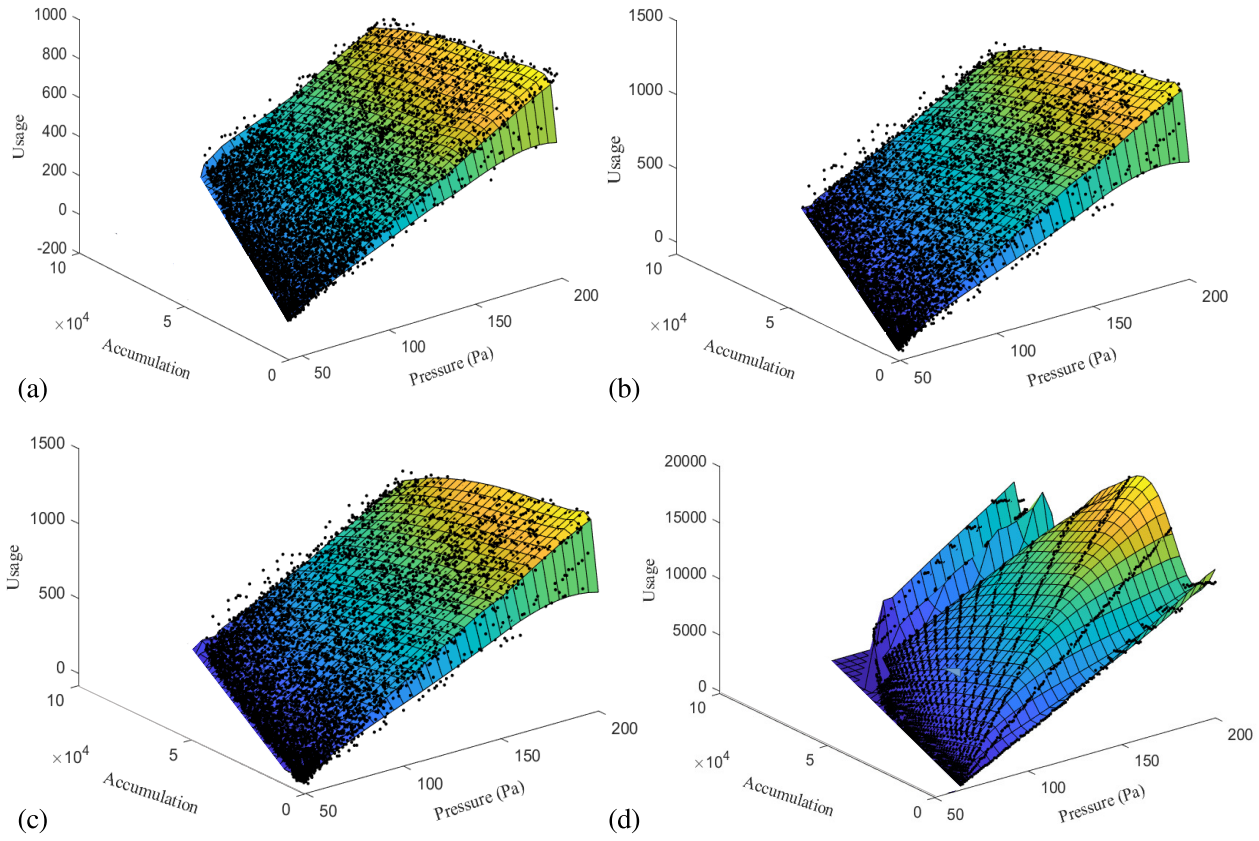


Figure 3.8: Fitting results of the ANN model. The black dots are the kMC simulation result and the surfaces are the neural network fitting result. The vertical z-axis is the output transient deposition rate as in precursor usage, and the horizontal x-axis and y-axis are the input accumulated coverage and surface partial pressure, respectively. This fitting is presented under various representative temperature inputs: (a) T = 475 K, (b) T = 500 K, (c) T = 550 K, and (d) T = 600 K.

3.3.2 Macroscopic Linear Parameter Varying Model Determination

In this section, the macroscopic gas-phase model linear parameter varying model validation and parameter determination are elaborated. In total, four pairs of multiscale CFD simulations are carried out under various operating conditions. Each simulation pair, under the same operating conditions, contains a standalone macroscopic CFD simulation and a multiscale CFD simulation, which couples the macroscopic CFD with the microscopic kMC model, taking the output of the microscopic kMC result as feedback to CFD at each time step. Under various operating conditions, the gas phase development profile response curves from all CFD simulations share

a similar overall shape. This shape can be characterized by a linear parameter varying model that is higher than first-order without zeros and varying parameters. In addition, we have endeavored to incorporate the radial information into a single gas-phase model. Nevertheless, it was found that the formulation of such a model is complicated and the error associated with this model is too significant to be representative. Thus, we have decided to build several linear parameter varying models in parallel for different locations. As a result, third-order linear parameter varying models are constructed as explained in Section 3.2 to characterize the gas-phase development and its result is compared with the simulation data.

The operating pressure effect shows an independent linear relationship with respect to the surface partial pressure output and therefore is considered separately from the linear parameter varying model. The decoupling of the operating pressure makes physical sense due to the fast transport of the incompressible gas mixture according to the physical dimensions of the ALD reactor [66]. The total pressure development is greatly affected by the momentum transfer of Argon left from the purging, which takes less than 0.1 s across the entire reactor at all operating conditions and is much faster than the precursor species transport rate. After the total pressure reaches steady-state, it is almost equal to the operating pressure throughout the reactor, with a relatively small decreasing differential pressure from the inlet to the outlet. It is further determined that a change in the operating pressure within an order of magnitude does not dramatically influence the species transport speed. Thus, the surface precursor partial pressure, which is the product of the precursor species concentration and the total pressure, can be considered to be directly proportional to the operating pressure setting after the initial period. As a result, the operating pressure can be independently considered in the gas-phase data-driven model.

The third-order linear parameter varying model with step response successfully reproduces the temporal surface partial pressure development for different inlet precursor flow rates as shown in Figure 3.9. The model accuracy, defined as the root mean squared error (RMSE), ranges from 1.147 to 3.778, and the fit to estimation ranges from 95.06% to 98.76% under various conditions.

Each of the third-order linear parameter varying model parameters shows a semi-log

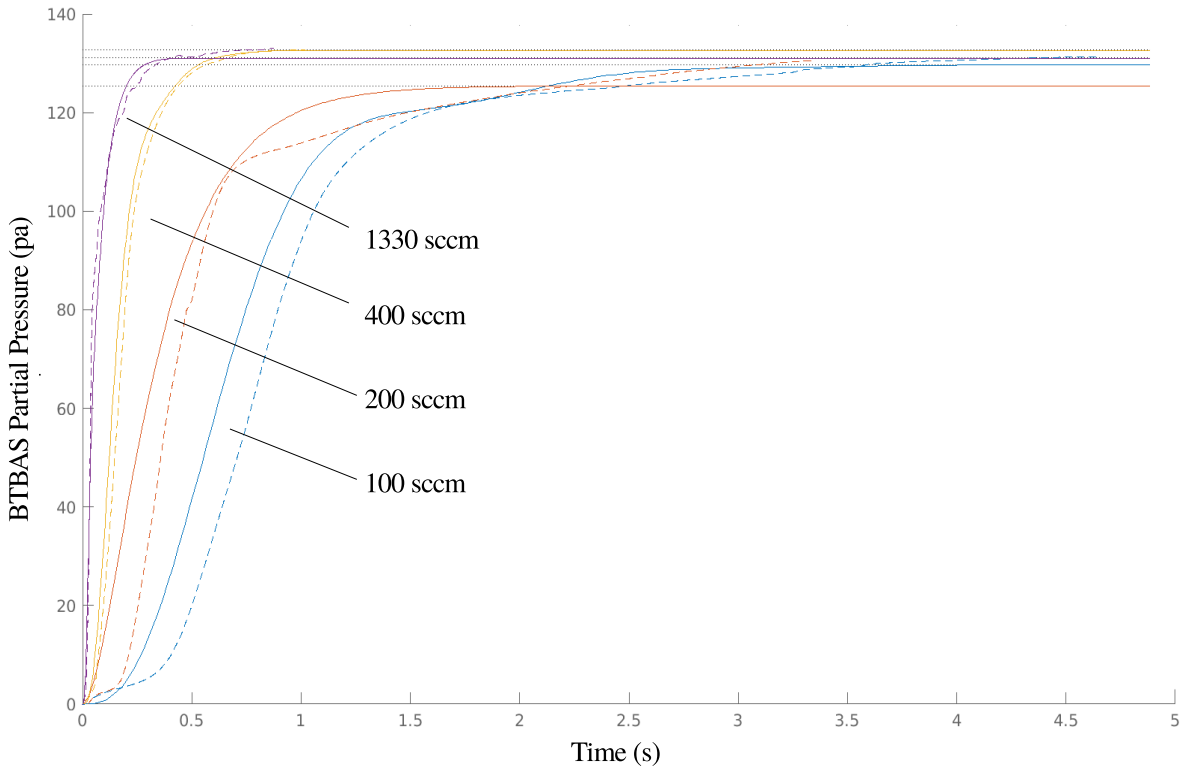


Figure 3.9: Comparison of the wafer-average linear parameter varying model fitting and the CFD gas-phase development. The x-axis is the ALD half-cycle physical time, and the y-axis is the surface BTBAS partial pressure. The dashed lines are the original CFD data and the solid lines are the fitting result from the linear parameter varying model. As indicated, the blue, red, yellow and purple lines are 100, 200, 400 and 1330 sccm inlet precursor flow rates, respectively.

relationship with inlet precursor flow rates. Thus, we find that an interpolation of the known parameters to generate process parameters of unknown interior operating conditions is promising. Since the linear parameter varying model is quite sensitive to its parameters, the fitting accuracy of the validation flow rate is better for higher flow rate where the linear parameter varying model provides a better fit. Nevertheless, all flow rates provide accurate estimations of half-cycle finishing time as demonstrated in the following section.

3.3.2.1 Multiscale Data-Driven Model Validation

After we have constructed the data-driven models for both microscopic and macroscopic domains, we combine the two models and incorporate the location information on the substrate surface for

a complete ALD process multiscale model. Validation runs are performed and compared with previous works to prove the model accuracy. One test scenario is demonstrated with conditions as follows: The precursor flow rate is 100 sccm, surface heating is 600K, and the operating pressure is 133 Pa, half-cycle process time is 3.5 seconds, and a time-step size is chosen to be the optimal ANN model step size to avoid the previously mentioned possible fitting problem. As discussed in the previous section, at each time step, the linear parameter varying model calculates the surface partial pressure under given input conditions and radial coordinates. Then, the microscopic ANN model draws this surface partial pressure input and the surface heating temperature input and calculates the transient deposition rate. This calculation process is executed for each of the 36 wafer grid locations [66].

As shown in Figure 3.10 (a), the two vertical lines specify the time for the fastest and slowest grid locations to achieve full coverage, which occur at t = 1.82 s and t = 3.11 s, respectively, from the data-driven model simulation. In addition, the central region shows a faster deposition rate than the peripheral region as predicted by the CFD model. By contrast, Figure 3.10 (b) shows the result from the multiscale CFD model for the same operating conditions reported in [66], and the differences of the fastest and slowest finishing time between the multiscale data-driven model and the multiscale CFD model are less than 10%. Therefore, it is demonstrated that the multiscale data-driven model correctly captures the general trend of the temporal coverage and the final completion times. Nevertheless, it is noted that the process curves produced by the data-driven model are not smooth because of its discrete nature. As the film growth gradually approaches its steady-state coverage, the slight model mismatch in microscopic deposition steady-state identification can cause the model to temporarily mis-identify a microscopic deposition steady-state value according to the current operating condition, thereby causing the prediction to plateau until the next steady-state is found after an updated operating condition is provided by the gas-phase.

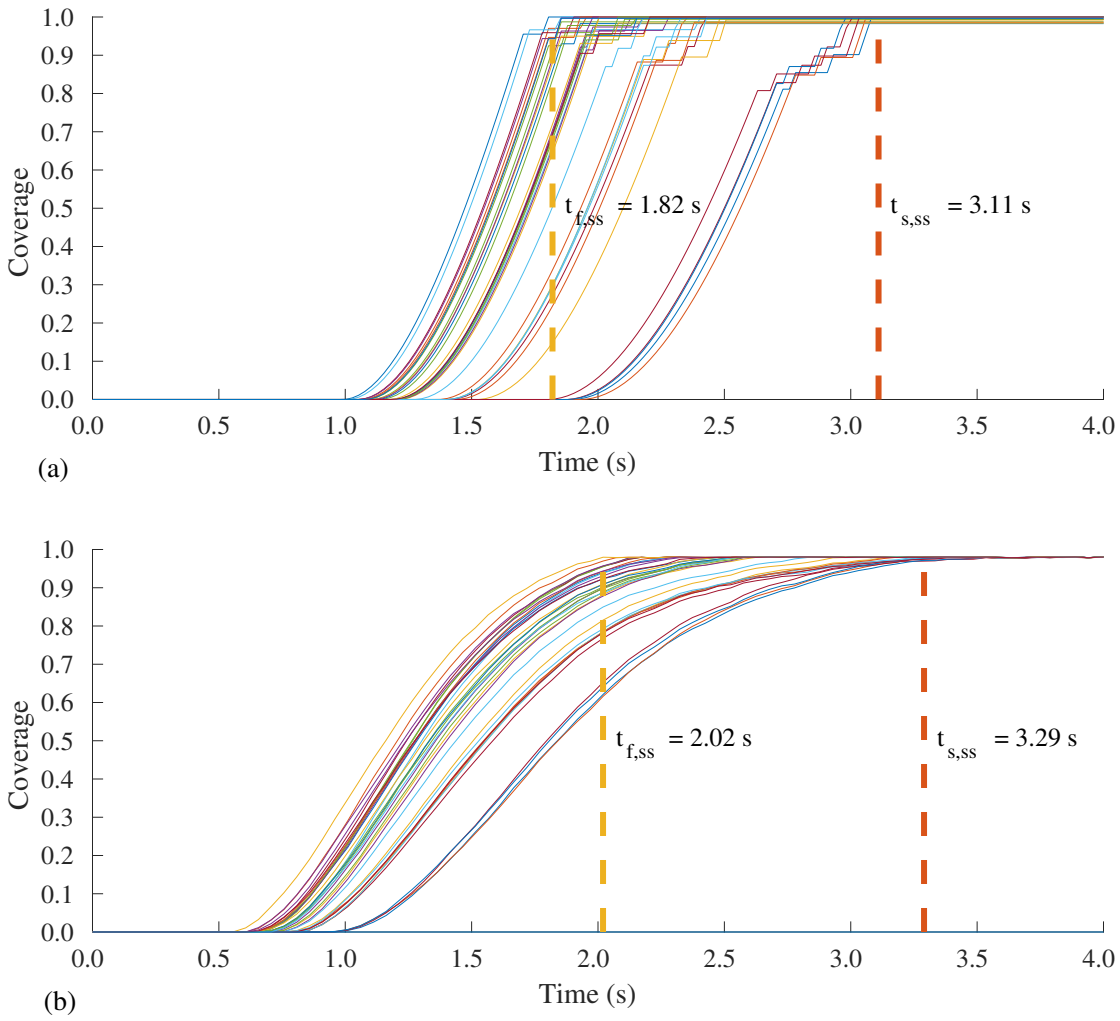


Figure 3.10: Comparison of one testing scenario between the data-driven model and the validation multiscale CFD data. The test scenario is under a substrate surface temperature of 600 K, an operating pressure of 133 Pa and an inlet feed flow rate of 100 sccm. The *x*-axis represents the process operation time, the *y*-axis represents the surface coverage and the dashed lines represent the occurrence of the fastest ($t_{f,ss}$) and the slowest ($t_{s,ss}$) deposition steady-states. Figure (a) presents the data-driven model results, where the completion of deposition on the wafer surface is achieved between 1.82 s and 3.11 s, and Figure (b) shows the multiscale CFD simulation results, where the completion of deposition is achieved between 2.02 s and 3.29 s. The deviation between the two calculations is less than 10%, respectively.

3.3.2.2 Multiscale Data-Driven Model Advantages and Applications

As discussed in previous sections, a full multiscale CFD simulation takes a long time (i.e., 5 - 10 days). Specifically, each time step needs, on average, from 0.6 to 1.2 hr to compute. By

contrast, when the multiscale data-driven model is adopted, an entire run now only takes around ten seconds, with each time step taking less than 0.1 s. Thus, it is demonstrated that the time scale of this model is comparable to the actual physical time progress. Due to the slight mismatch between the model-predicted and the actual shape of the process curve, real-time control within each half-cycle might not be achievable for a single-wafer reactor. However, this model is possible to be incorporated with on-line monitoring and half-cycle time prediction in real-time. Moreover, the accurate deposition completion time prediction allows for a fast cycle-to-cycle film quality control.

One application of this multiscale data-driven model is to provide an industrial operating guideline by making a fast and computationally feasible half-cycle time prediction under various input condition combinations, including the precursor flow rate, operating pressure and surface heating. An estimation can be made in advance of experiments so that the operating cost can be minimized. A variety of operating conditions that are in the industrial operating space is shown in Table 2 (a) and (b). Typically, in the industry, to ensure the achievement of good film coverage under all operating conditions, a half-cycle time of at least five seconds or even longer is adopted. However, with the prior-operation prediction using the multiscale data-driven model, the half-cycle time can be cut by at least 37.8%. In particular, under high flow rate operating conditions, the multiscale data-driven model can provide a huge economic advantage by cutting unnecessary process time and precursor usage in conventional practices, thereby increasing the throughput by accelerating the manufacturing process. In addition, this model can provide a more complete and more accurate prediction than the previously proposed estimation models. In our prior work, [63] has investigated into the estimation of half-cycle time using an ANN for the standalone kMC simulation with an independent consideration of the gas-phase development. As shown in Table 2 (a), we compare the predicted half-cycle time at these operating conditions using the standalone kMC model and the multiscale data-driven model. On top of the microscopic deposition time, the standalone kMC model predictions assume a fixed time of three seconds for gas phase profile development, which is a good estimation for an 100 sccm inlet flow rate according to [66]. On

the other hand, the multiscale data-driven model accounts for the coupled gas-phase and surface deposition profile and process time. Therefore, the multiscale data-driven model provides a more realistic half-cycle time prediction for a thermal ALD process by 5% to 15% depending on the operating temperature, thereby significantly reducing half-cycle time.

Table 3.2: (a) Comparison of the predicted deposition time to full coverage under various temperatures between the standalone kMC model and the mulstiscale data-driven model, and (b) additional temperature and inlet precursor flow rate predictions using the multiscale data-driven model.

Temperature (K)	Standalone kMC prediction (s)	Multiscale data-driven model prediction (s)
590	4.05	3.81
595	3.85	3.56
600	3.65	3.11
610	3.55	3.01

(a)

Temperature (K)	Flow rate (sccm)	Predicted time (s)
600	100	3.11
600	1330	1.45
550	1330	2.18
550	400	2.48

(b)

A machine-learning model opens up the possibility for run-to-run and real-time control. There are reactor environment changes, for example, materials deposited to the wall, that should be considered for run-to-run control purposes. These effects, nevertheless, do not dramatically alter the developed input-output relationships. Thus, the machine-learning model could be adapted online to accommodate those changes, using film coverage measurements after a complete deposition has occurred. This updated machine-learning model may then be used in a run-to-run controller to update the operating conditions. With respect to the operation of the control actuators, Proportional-Integral control schemes should be used to regulate the actuator outputs to the values calculated by the machine-learning model. Specifically, the machine-learning model calculates the substrate temperature, precursor feed flowrate, and operating pressure required to ensure full coverage within a specific deposition time. However, these three variables are regulated by control actuators. For example, the precursor feed flow rate is set by the valve position. These actuators

should be controlled by Proportional-Integral controllers with suitable tuning parameters to ensure that these three variables reach the values requested by the machine-learning model quickly given the short deposition duration in ALD.

3.4 Conclusion

In this chapter, we developed a multiscale data-driven model from a first-principles based multiscale CFD model of the thermal ALD SiO_2 thin-film deposition using BTBAS as precursor. Specifically, the resulting multiscale data-driven model consisted of an ANN-based model for the microscopic film growth domain and a linear parameter varying model for the macroscopic gas-phase domain. The final trained microscopic ANN model achieved a good prediction with a normalized error of 1.0×10^{-3} and a precise correlation of the predicted and kMC data with R^2 = 0.99. The macroscopic linear parameter varying model also captured the gas-phase partial pressure development under various flow rates, using the step response of a third-order linear parameter varying model as the form of fitting function. An average error of 3.07% was observed for the fitting under all flow rates. A validation test for the multiscale data-driven model was carried out at 600 K, 133 Pa and 100 sccm feed flow rate. The prediction results closely resembled those from the multiscale CFD model, with less than 10% deviation in the deposition completion time. The profile mismatch did not affect the prediction accuracy and still allowed for a fast cycle-to-cycle control. Moreover, the multiscale data-driven model enabled a low-cost on-line process operational space exploration, which may significantly reduce the amount of time and resource spent on experimental and industrial work. In addition, this versatile model could provide insights on different thermal ALD thin-film deposition processes under various working conditions and precursors.

Chapter 4

Microscopic Modeling and Optimal Operation of Plasma Enhanced Atomic Layer Deposition

4.1 Introduction

As mentioned in previous chapters, atomic layer deposition (ALD) has been widely utilized as a modern deposition technique in the semiconductor manufacturing industry. It has been developed to meet the demand of producing conformal thin-films in high aspect ratio structures for the metal-oxide-semiconductor field-effect transistors (MOSFETs) in the NAND-type flash memory devices [4]. To allow for the device size reduction to follow the Moore's Law and to enhance the performance of the microelectronic components, the thickness and width of the traditional oxide films (e.g., SiO_2, Si_3O_4 and Al_2O_3) have been constantly reduced to decrease the overall size of the transistor and to increase the gate capacitance effectiveness [2]. Nevertheless, as the thickness of the oxides scales below 20 Å, there exists a significant increase of leakage current through the oxide film due to quantum tunneling, where a substantial portion of the electrons passes through the closed transistor gate caused by the low potential barrier of the extremely thin oxide

films [84]. Thus, the spiking induced power consumption and the lack of device reliability limit further part miniaturization. As a result, replacing the traditional oxide film with high dielectric constant (high-κ) materials, where large enough potential barriers can be achieved with a similar thickness scale, has been extensively investigated [85]. Nevertheless, the manufacturing of those high-κ thin-films using thermal ALD has faced challenges of low-throughput and high operating cost [86]. The primary reason for these problems is the extremely slow reaction kinetics at low temperatures because of the high activation energy barrier of the traditional precursors, thus often requiring high temperature for energy input. To overcome this problem, plasma is introduced as an enhancement to the originally neutral precursors, in which free radicals and ions are created so that the deposition reactions can go through lower energy barrier pathways [87].

Therefore, in this chapter, a computationally efficient, yet sufficiently rigorous microscopic kMC model has been proposed for the PEALD process with TDMAH and O_2 plasma as precursors. Specifically, important structural details and reaction parameters are computed from the density functional theory (DFT) calculation. Depending on the calculated parameters, we adopt a 3D kMC algorithm to capture the surface reaction mechanism and the structure detail of the HfO_2 thin-film. From the developed simulation, a meaningful range of experimental conditions is explored to construct a deposition profile database. Based on this dataset, neural network regression is adopted to help to map the inlet parameter combinations to the film structures and growth profiles, and to figure out the optimal operating conditions according to throughput and economic concern. The workflow of this paper is shown in Figure 5.1.

4.2 PEALD Microscopic Modeling

This section addresses the microscopic film growth model of the PEALD. First, the HfO_2 structure detail and corresponding kMC lattice representation are discussed. Next, a selected set of reaction mechanisms that are critical to surface deposition, and their respective rate calculations are explained. After that, the kMC algorithm time progression and event selection are formulated in

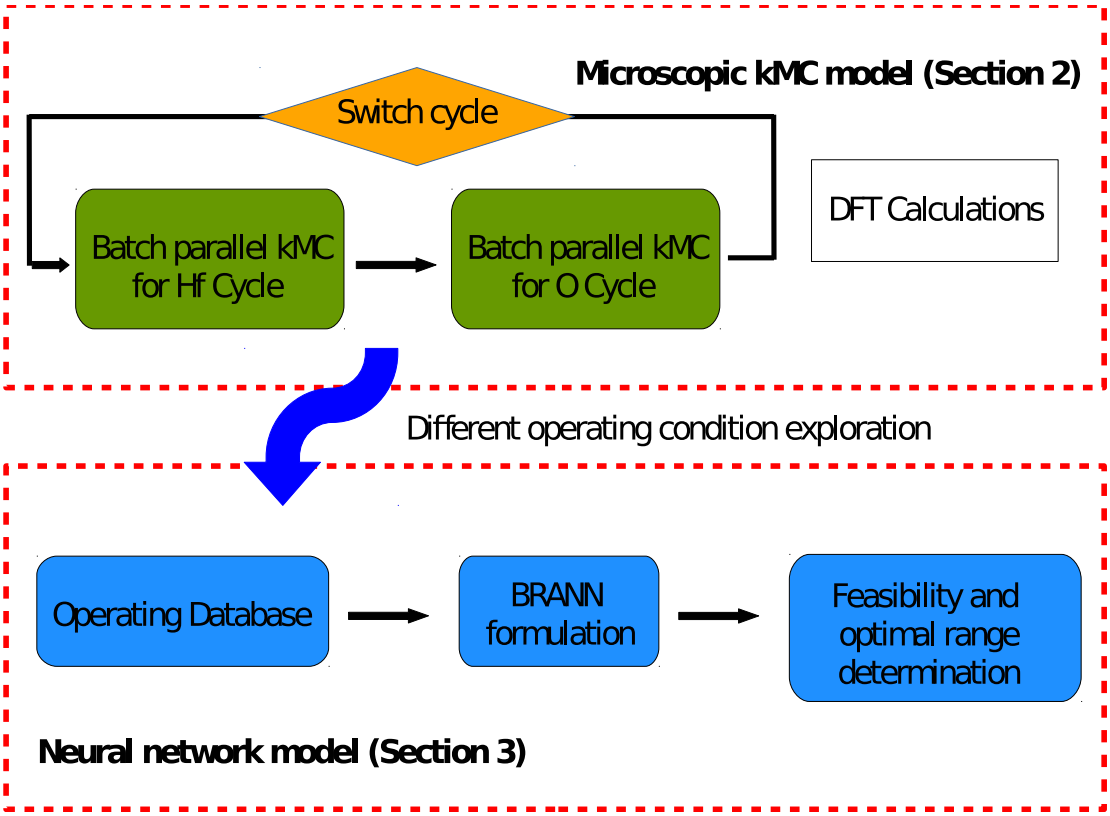


Figure 4.1: Workflow demonstrated in this paper and the respective section of each component.

detail. In addition, the DFT calculations are introduced for the determination of the important process and structural parameters. Finally, the automated workflow with cloud distributed computing is demonstrated.

4.2.1 Structural Characterization of the Deposited HfO_2

There are various crystalline structures that HfO_2 can take on, including tetragonal ($P4_2/nmc$), cubic ($Fm\bar{3}m$), monoclinic ($P2_1/c$), orthorhombic (Pbca and Pnma) and etc. [88]. Although it is known that the PEALD method often leads to an amorphous structure that is close to some crystalline phase, the exact phase is critically dependent on the deposition temperature. At very high temperatures, the HfO_2 often forms in the cubic phase, whereas the surface is often in the monoclinic phase at low temperatures [89]. In this paper, since we are investigating the PEALD near room temperature or at low temperature, we consider a surface structure that takes the

monoclinic-alike structure. The (111) surface, known as one of the most common surfaces of Hafnia, is chosen as the deposition surface, and the growth direction is shown as the z-direction in Figure 4.2 (a). Also, the hydroxyl-terminated surface, which is commonly prepared by the surface treatment with piranha solution and then in dilute HF solution to remove native oxide and organic compounds, is chosen as the initial substrate surface, which is shown in Figure 4.2 (b). Although it seems to be the most realistic if we adopt the exact monoclinic structure angle and bond length, the complex reaction pathways and structural detail in this real lattice make the kMC formulation computationally and conceptually challenging. As a result, we adopt a 3D triangular on-lattice model, developed in [63], as an approximation of the local tetrahedron alike crystal structure. In the approximated lattice structure, the atoms are considered on a crystal lattice, where the bonds connecting the Hf and the adjacent O atom all have 90 degree angles with respect to each other. This lattice structure assumption successfully preserves the key structural characteristics. The exemplary lattice miniature of our simulation is shown in Figure 5.4, where HfL2 is the deposited Hf with two unreacted methylamino ligands, and PsHf and CsHf are physisorbed and chemisorbed precursors, respectively. PsO1 or PsO2 are the physisorbed oxygen species which remain to be oxidized. This top view is shown to closely resembles the surface of the real PEALD lattice, as shown in Figure 4.2 (b). Also, the assumption leads to a similar cycle repetition as found in the realistic structure. The structural validation will be further demonstrated in the model validation section. The actual lattice used in the simulation has 1200×1200 sites per layer with a height that equals to the number of cycles simulated, which is proven to be large enough to be size-independent [35].

4.2.2 Reaction Mechanism

As introduced in the general PEALD process, a full cycle is comprised of two half-cycles with purging cycles in between. In each half-cycle and the beginning of its purging step, a specific or a set of specific precursors will contact and react with the substrate surface, which deposits the desired element onto the film. Tetrakis-dimethylamino-Hafnium (TDMAHf), at its realistic

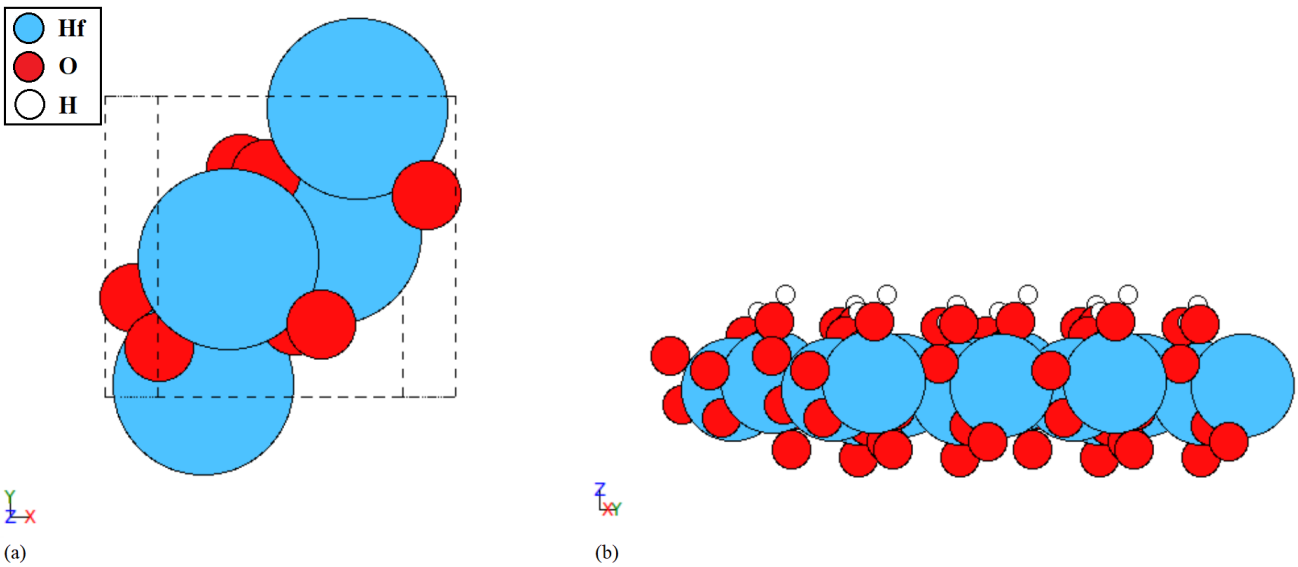


Figure 4.2: (a) Optimized monoclinic HfO_2 unit cell in the bulk, and (b) relaxed OH-terminated HfO_2 $2 \times 2 \times 1$ slab.

partial pressure with respect to the purging Argon, is considered to be the precursor for the Hafnium half cycle, and molecular oxygen ($^3O_2$), singlet molecular oxygen ($^1O_2$), and triplet atomic oxygen (^3O), generated from the plasma chamber, are considered to be the precursors of the oxygen half-cycle. The singlet atomic oxygen (^1O), according to the [90], recombines at a high rate in the gas transport in the main reactor, and will thus not be considered for the surface reaction. The detailed reaction mechanisms of both cycles were reported by [8] and [91]. However, the implementation of the full reaction set, which follows all the possible mechanisms of ligand rotation, proton diffusion, and other processes, will dramatically increase the computational demand, and thereby make the simulation of an industrial-sized wafer infeasible. Therefore, key reactions are selected with essential details on the proton transfer and structural influence considered, which would still constitute a full deposition reaction set and reproduce the experimentally reported growth rates and potential defects. The selected reaction mechanisms are explained below.

The first half-cycle, also referred to as the Hf-Cycle, contains physisorption, desorption, and a two-step dissociative chemisorption, which is shown in the top box of Figure 5.5. The precursor TDMAHf is exposed to the hydroxyl group terminated Hf surface, as shown in Figure 5.5(a).

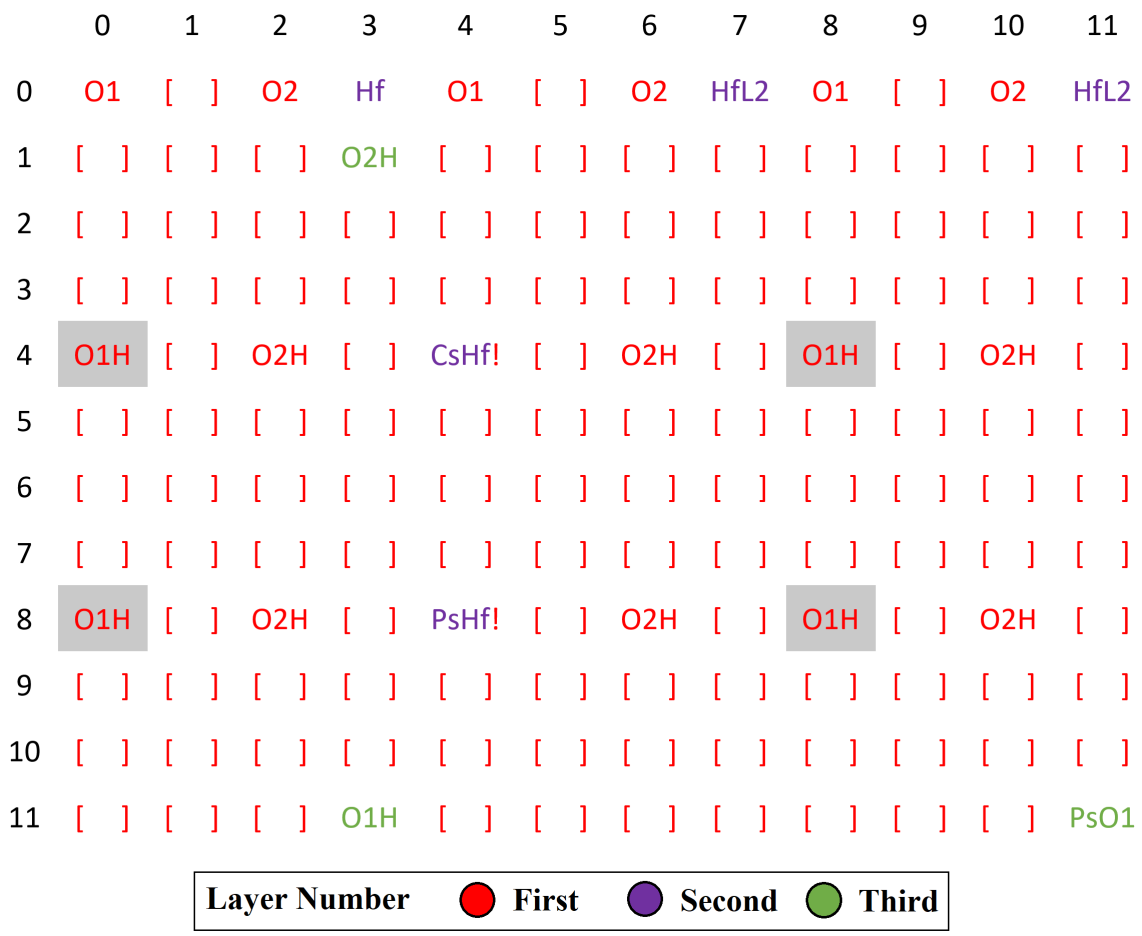


Figure 4.3: Top view of a 3-layer 12×12-site miniature demonstration of the full lattice used in the model. The three layers and the species on the lattice are shown using different colors and symbols, respectively. Grey background denotes the sterically hindered sites. The exclamation mark (!) denotes the OH group directly underneath.

According to [43], the two oxygen atoms in a cell of a surface oxide typically have different electronegativities, which is caused by the distinctive extent of exposure from the local bonding geometry. This conclusion is also implemented for HfO_2 in this work, in which the more electronegative oxygen atom is more likely to be attacked by the precursor. Therefore, as indicated by Figure 5.5(1), the nitrogen atom in the precursor particle enables strong physisorption onto the more electronegative hydroxyl group through an H-bond. After the physisorption, the precursor undergoes the first chemisorption step, as indicated by Figure 5.5(2), where the surface H from the attacked O will transfer across the ligands on the precursor, as shown in Figure 5.5(a).

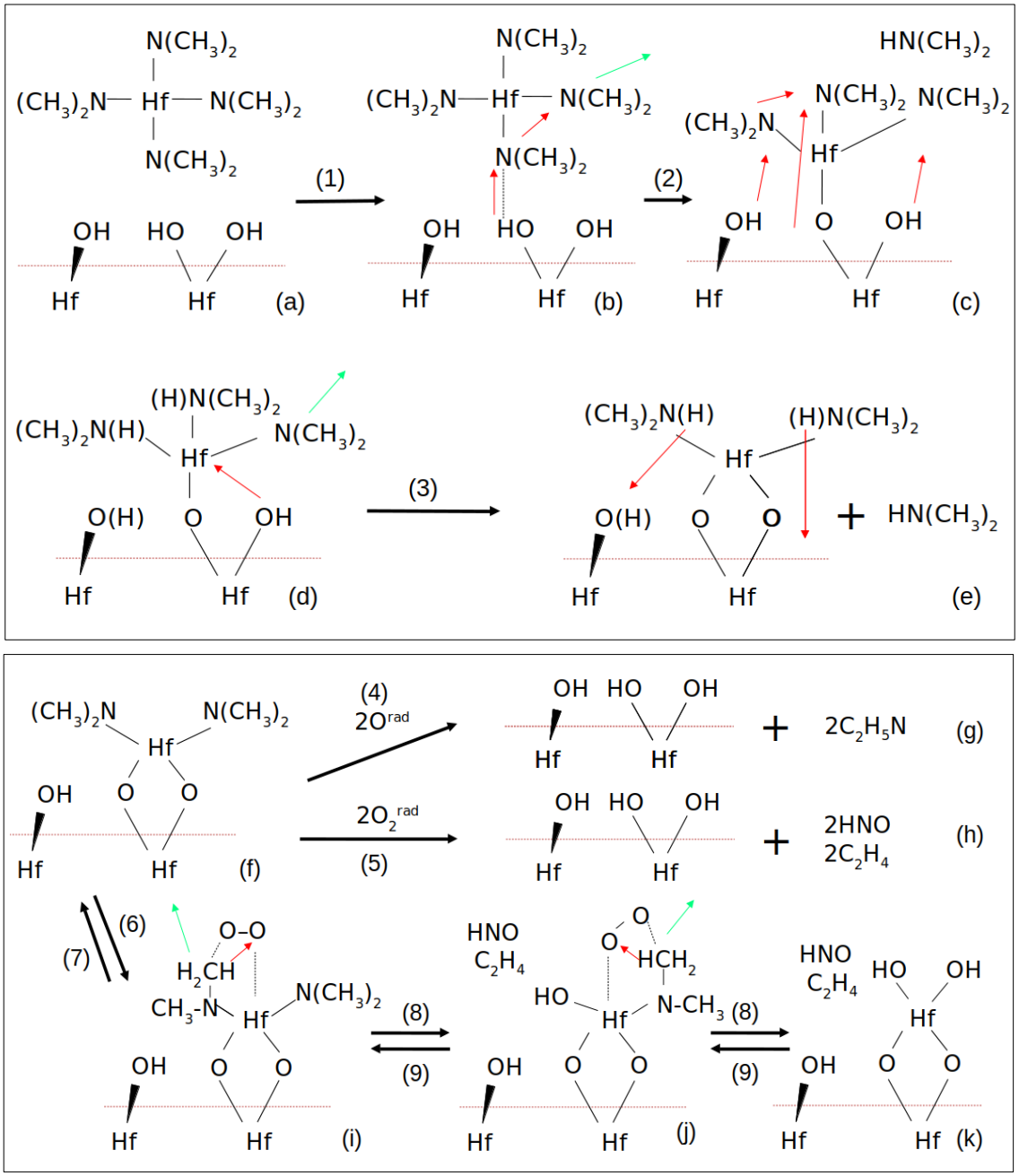


Figure 4.4: Illustration of the selected representative reaction set. The top part is the Hf-Cycle and the bottom part is the O-Cycle. The black arrows represent the overall reaction pathway, the red arrows represent the possible proton diffusion, and the *H* represent the potentially protonated positions. The number labels refer to the events and the alphabetic labels refer to the states. Please note that the bond lengths are not exact and drawn for descriptive purposes.

Eventually, the Hf is chemically bonded to the surface, releasing a dimethylamine (DMA) group, shown in Figure 5.5(c). After this first dissociative chemisorption step, more protons may be

transferred from the nearby substrate surface toward the precursor, making some of the ligands in the protonated state. Subsequently, one of the remaining DMA groups reacts with the adjacent hydroxyl group from the neighbour Hf atom, indicated by Figure 5.5(3), where again the surface H from the attacked O will transfer across the ligands on the precursor, shown in Figure 5.5(d), and release another DMA group, thus anchoring the Hf on the surface. This reaction pathway retains the original surface orientation and results in a thermodynamically favorable structure. After the reaction, the remaining two DMA groups become the new surface species, and some ligand protons will return to the nearby substrate surface, shown in Figure 5.5(e).

The second half-cycle is the O-Cycle, or the plasma cycle. Due to the high energy of radicals, i.e., plasma species, they often undergo and encounter different reaction routes and energy barriers than the neutral non-plasma species, thereby facilitating the deposition process. In particular, the O-cycle contains the following processes: physisorption, abstraction and surface oxidation, and the respective reverse reaction. While the majority of the works ignore the deposition effect of the neutral oxygen gas [92, 93], [94] stresses the importance of the oxygen gas contribution in the oxygen plasma. As a result, we take all of the potentially important species from the plasma reaction into consideration in our model: the molecular oxygen ($^3O_2$), the singlet molecular oxygen ($^1O_2$), and the triplet atomic oxygen (^3O). According to the kinetics reported by [91], the oxidation of the surface DMA group by the radicals, $^1O_2$ and ^3O, are extremely fast and kinetically barrierless. Therefore, as soon as those species are physisorbed onto the surface, they instantaneously oxidize and replace the surface DMA group with hydroxyl groups, as shown in Figure 5.5(4) and Figure 5.5(5), leaving the by-product gas of ethanimine compounds or the nitroxyl and ethane, respectively, indicated by Figure 5.5(g) and Figure 5.5(h). For the non-plasma species, $^3O_2$, however, it is essential to understand the competition of the surface chemical reactions because of the high transition state energy barrier. First, the $^3O_2$ physisorbs onto the DMA terminated surface, as indicated by the step Figure 5.5(6), to form a ring-shaped physisorbed complex, as shown in Figure 5.5(i). This physisorbed complex could be abstracted from the surface, as shown in Figure 5.5(7), because the addition of molecular oxygen may increase the

stability of the DMA groups attached to the surface [95]. Moving forward, the physisorbed complex may go through a series of oxidation steps, shown in Figure 5.5(8), producing nitroxyl and ethane in each step indicated by Figure 5.5(j) and Figure 5.5(k). The reverse reaction, indicated by Figure 5.5(9), is associated with high activation energy. To characterize the aforementioned O-Cycle reaction mechanisms, we combine our own DFT calculation with some reported reaction intermediate complex and transition state profile reported by [91].

It is important to note that the proton transfer back and forth from the ligand and the surface discussed in this section has a very minimal energy barrier and thus will happen extremely fast. Also, those events depend on the complex local coordination number and ligand rotation state, and thus, they are impossible to be deterministically captured. As a result, the possible protonated states will be randomly captured in the kMC, which will slightly influence the reaction energy. In addition, steric hindrance is an important concern that influences the surface site occupation. According to [8], due to the bulkiness of the DMA groups, once a TDMAHf is physisorbed on the surface, indicated by Figure 5.5(1), it will cause steric hindrance to the nearby sites, which means the closest hydroxyl groups will not be able to accept new precursor species, as roughly seen in Figure 5.5(b). This blocking effect will continue to affect the local deposition environment until the surface DMA groups are oxidized, as shown in Figure 5.5(g), (h), and (k). Those two specific considerations extract the important characteristic of the comprehensive model developed by [8] and the experimental findings, which will help our model more closely reproduce the realistic PEALD process conditions while maintaining a computationally acceptable algorithm.

4.2.3 Relative Rate Determination

The reaction kinetics of the pathways discussed in the previous section is crucial to the kMC event selection scheme, which influences the competitions between the aforementioned reactions and the overall deposition profile. There are majorly two kinds of reactions in the adsorption process, the physisorption and the surface chemical reactions, that can be characterized by different governing theories.

The chemisorptions, which in our case include the dissociative chemisorption, the desorption, and the forward and reverse oxidation reactions are thermodynamically activated reactions, which can be described by the general Arrenhius-type equation governed by the transition state theory (TST) [47]. The reaction rate is described by the thermodynamic properties of the transition state complexes at the quasi-equilibrium. As a result, the Arrenhnius-type reaction rate of the aforementioned reactions can be formulated as follows:

$$r_{rxn,i} = A_i \exp\left(\frac{-E_{a,i}}{k_b T}\right) \tag{4.1}$$

where i refers to the index each individual reaction, $E_{a,i}$ is the activation energy of the transition state complex, and A_i is the pre-exponential factor, which is governed by the following relationship:

$$A_i = f_i^{\mathrm{TST}} \left(\frac{k_b T}{h}\right) \tag{4.2}$$

where f_i^{TST} is the vibrational partition function ratio between the transition state complex and the reactants, k_b is the Boltzmann constant, and T is the temperature. The f_i^{TST} is around 1 for reactions happened on the slab surface [8].

In contrast, the physisorption r_{phs} can be described by the gas collision theory, which is reasonable for the gas-surface athermal barrierless processes:

$$r_{phs} = \frac{p_i}{RT}\sqrt{\frac{8RT}{\pi M_i}} s_{c,i} N_a \sigma \tag{4.3}$$

where for species i, p_i is the partial pressure of the precursor species, either TDMAHf or oxygen radicals, R is the gas constant, T is the absolute temperature, M_i is molar mass, $s_{c,i}$ is the sticking coefficient, σ is the unit cell surface area, and N_a is the Avogadro number. The sticking coefficient of TDMAHf changes exponentially with varying surface temperature, which is reported by [96].

4.2.4 Kinetic Monte Carlo Algorithm

In order to simulate the above reaction pathways and to realistically represent the reaction rates and structural evolution, first-principles based methods are adopted. Traditionally, Molecular Dynamics (MD) simulations are used, but a full MD simulation is too computationally expensive to be applied for the large-scale system investigated in this work [26]. As a result, we choose to utilize the n-fold kinetic Monte Carlo (kMC) algorithm in the framework proposed by earlier works [63]. kMC is a computational random sampling algorithm that, in this case, simulates the time progression of the surface deposition based on random variables and reaction kinetics. In particular, we focus on two aspects: event selection and time progression.

For the event selection, a random number, $\gamma_1 \in (0,1]$, is chosen from the uniform distribution. To utilize γ_1, we define a collective set of all possible events, which includes all the physical and chemical reactions. The total reaction rate r_{total} can be calculated as the sum of the unweighted rates calculated from the individual events:

$$r_{total} = \sum_{i=1}^{N} r^{i}_{unweighted} \tag{4.4}$$

where $r^{i}_{unweighted}$ represents the unweighted rate of the ith event within an N-event set, each calculated from the equations discussed in the previous section. Adopting the aforementioned equations, the event selection details of the two half-cycles can be computed as follows. For the Hf half-cycle, the total rate is:

$$r_{Hf,total} = r_{Hf,phs} + r_{Hf,des} + r_{Hf_a,che1} + r_{Hf_b,che2} \tag{4.5}$$

where $r_{Hf,phs}$ is the rate of TDMAHf physisorption, $r_{Hf,des}$ is the rate of TDMAHf desorption, and $r_{Hf_a,che1}$ together with $r_{Hf_b,che2}$ are the dissociative chemisorption rates of the TDMAHf first and second chemisorption, respectively. In contrast, for the O-Cycle, the total rate is:

$$r_{O,total} = r_{^3O,phs} + r_{^1O_2,phs} + r_{^3O_2,phs} + r_{^3O_2,des} + r_{^3O_2,oxidation} \tag{4.6}$$

where $r_{^3O,phs}$, $r_{^1O_2d,phs}$, and $r_{^3O_2d,phs}$ are the physisorption rates of atomic oxygen radicals, oxygen radicals and ground-state oxygen gas, respectively. While $r_{^3O_2,des}$ and $r_{^3O_2,oxidation}$ are the desorption and forward oxidation reaction, respectively. Chemisorption of the radical gas is not included in the event selection because the energies associated with atomic oxygen and oxygen radicals are extremely high. Thus, the barrier from the reaction complex to transition state can be easily overcome and the reaction can be considered instantaneous.

The first aspect of the kMC algorithm is the computation of the time progression to model the transient behavior of the surface deposition, utilizing a random number $\gamma_1 \in (0,1]$. The time of each selected event is governed by the total rate as follows:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{4.7}$$

Starting from $t = 0$, the simulation time clock is consecutively advanced by Δt of each event until it reaches the predefined half-cycle time.

The second aspect of the kMC is to execute the event selection scheme. In order to reproduce the realistic competitions between the surface physical and chemical reactions, we need to consider a variation from the standard kMC, where the surface species and vacant site concentrations are taken into consideration. Thus, the normalized event indicators under the competition of reaction pathways and directions are calculated by dividing the concentration-weighted reaction rates with respect to the total weighted rate as follows:

$$l_{Hf,i} = \frac{\sum_{j=1}^{i} r^j_{unweighted} R_j}{\sum_{k=1}^{N} r^k_{unweighted} R_k}, \;\; i = 1,...,N \tag{4.8}$$

where $l_{Hf,i} \in (0,1]$ represents the normalized indicator of the ith event in the surface reaction event set, $r^j_{unweighted}$ and $r^k_{unweighted}$ are the unweighted chemical reaction rate for the jth and kth event, respectively, and N is the total number of events in the hf-cycle or the o-cycle reaction event set. R is the number of available reactants for each associated reaction. Physisorption considers the acceptable empty sites, forward chemisorption and desorption consider the physisorbed species,

and reverse reactions consider the chemisorbed or oxidized species. A second random number γ_2 is chosen for event selection. If the number lies in the ith interval, between the normalized indicators l_{i-1} to l_i, the ith event is chosen for execution. It will be demonstrated in the model validation section that this tailored kMC algorithm takes in the first-principles-based characteristic as in the Langmuir adsorption model and the structural details to successfully reproduce the complex deposition process.

4.2.5 DFT and Thermodynamic Calculations

Parts of the first-principles analysis are reported for the interaction of the HfO_2 substrate surface with TDMAHf and O_2 plasma. However, to fully understand the exact surface configurations and deposition profile, we adopt the molecular visualization toolbox ASE and Density Functional Theory (DFT) calculation with the Gaussian09 software package to extract all necessary reaction mechanism, thermodynamic properties, and structural detail. Such information would help the microscopic kMC model to reproduce realistic film growth behavior. [34, 66, 78].

DFT is an approximation approach to the solution of the Schrödinger equation of the many-body system:

$$\widehat{H}\Psi = E\Psi \tag{4.9}$$

where Ψ is the state vector of the quantum system, and E is the energy eigenvalue associated with the quantum state Ψ. In addition, $\widehat{H}$ is the Hamiltonian operator, which consists of the nucleic kinetic energy $\widehat{T}_{nuc}(R)$, the coulombic potential between nuclei and electrons $\widehat{V}_{coulomb}$, and the electronic Hamiltonian $\widehat{H}_{elec}(r,R)$:

$$\widehat{H} = \widehat{T}_{nuc}(R) + \widehat{V}_{coulomb} + \widehat{H}_{elec}(r,R) \tag{4.10}$$

where r and R refer to the position of the electrons and nucleus respectively. It has been demonstrated in extensive research that a direct solution of the Schrödinger equation is computationally impossible for large scale systems. Thus, DFT instead looks at the electron

density of the many-body system, where a set of single-electron Schrödinger-like equations are solved, known as the Kohn–Sham equations [97, 98]. Because the ground state energy of the system can be identified as a unique functional of electron density, by assuming a valid energy functional and minimizing the overall system energy, the true electron density can be obtained. However, this energy functional contains the exchanges and correlations of the system, which is not known except for the free electron gas [99]. Thus, a highly-accurate energy functional would be important to the accuracy of the DFT calculation.

In this work, a clean hydroxylated HfO_2 (111) surface is chosen as the starting point of the deposition. In order to compute the correct kinetic parameters of the reactions happening on this surface, an accurate HfO_2 (111) surface slab must first be constructed. First, the monoclinic HfO_2 (space group $P2_1/c$) lattice structure is constructed based on the repetition of its unit cell. Then, Gaussian09 is used to optimize this unit cell with the built-in periodic boundary condition (PBC) calculation [34]. The resulting structure, which is shown in Figure 4.2(a), is then imported into the Python CatKit Package, and the (111) surface slab, which is shown in Figure 4.2(b) is calculated and generated according to the optimized lattice parameter using the CatGen module. The resulting slab is then terminated with hydroxyl groups, and the surface layer is relaxed while the atom positions in the bulk are fixed. Additionally, a single-$Hf(NCH_3)_2$-terminated HfO_2 slab is generated, as shown in Figure 4.5(a). A molecular oxygen gas, O_2, is placed onto the terminating -NCH_3 group, as shown in Figure 4.5(b), to simulate the O_2 desorption process and study the associated kinetics. The O_2-physisorbed slab is optimized to transitional state using the Hartree-Fock (HF) method with the Los Alamos National Laboratory 2-double-z (LanL2DZ) basis set to obtain an initial guess of the structure and the kinetics parameter for desorption [100–102]. Although it is ideal to compute the exact profile and the transitional state energy using the B3LYP functional, the large amount of atoms causes the computation very slow. Therefore, based on the initial guess, we perform trial-and-error to obtain the kinetics parameter for the O_2 desorption reaction that leads to reasonable simulation results as shown in the following section.

The important DFT-computed activation energies for Hf-cycle, which are derived from [8],

are listed as follows: $r_{Hf,des}$ = 23.29 kcal/mole, $r_{Hf_a,che1}$ = 20.52 kcal/mole, and $r_{Hf_b,che2}$ = 24.67 kcal/mole. The sticking coefficient of TDMAHf, which is critical to the physisorption rate, is discussed in [96]. In addition, the kinetics of the barrierless plasma species and the oxidation energy of neutral oxygen, $r_{^3O_2,oxidation}$ = 58.48 kcal/mole, are described in detail by [91]. The desorption energy, $r_{^3O_2,des}$ = 56.01 kcal/mole and the structural hindrance information are calculated and estimated in this paper.

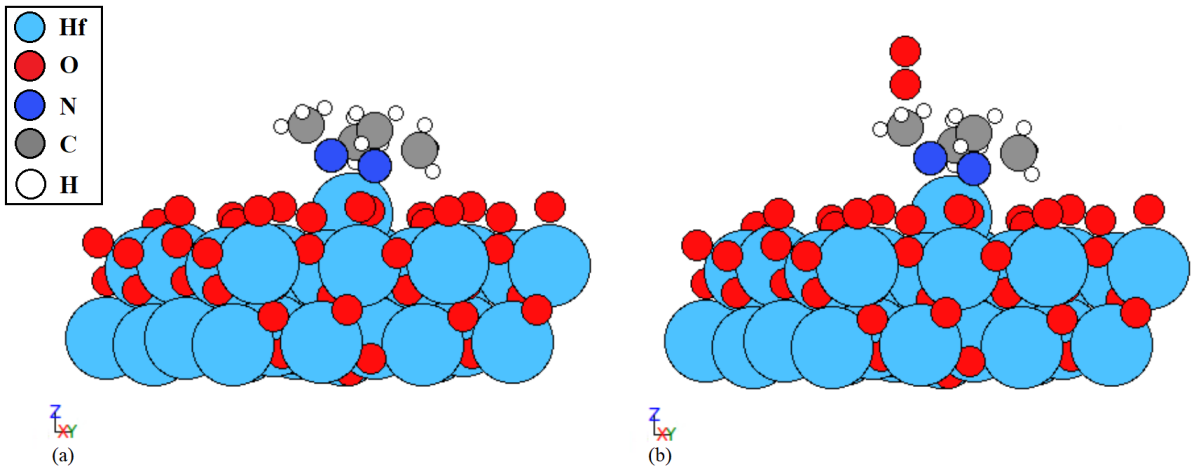


Figure 4.5: (a) single-Hf$(NCH_3)_2$-terminated HfO_2 $2 \times 2 \times 1$ slab, and (b) with O_2-physisorbed.

4.2.6 Simulation model result and validation

In order to validate the proposed kMC model, we compare the model output with experimental and simulation results reported in the literature [12, 86, 93, 94, 103–106]. Specifically, we will compare the exact growth per cycle (GPC) values under several operating conditions from the simulation model with the individual GPC reported from past works. In addition, the effect of individual operating parameter on the GPC is studied, where the major factors we consider are the substrate temperature and the precursor partial pressure of each half-cycle. It is noteworthy that the partial pressure development depends on the gas-phase momentum and species transports in the main reactor chamber, which are heavily influenced by the reactor's physical length scale and detailed design. In particular, there will transiently be a radial distribution of neutral and plasma species on the substrate surface. Nevertheless, a well-developed flow will lead to a rather uniform profile

Table 4.1: Comparison of GPC between literature values and simulation results under various operating conditions.

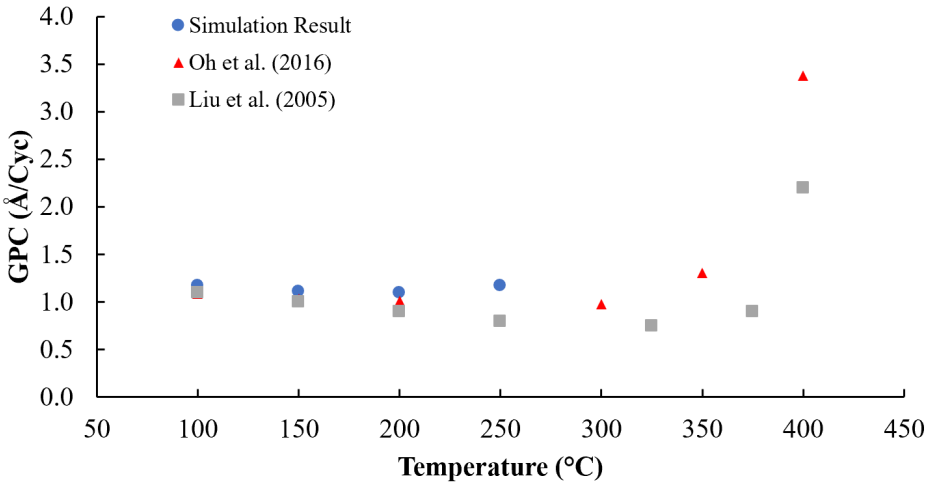
Condition	250 °C High dosage		200 °C Low dosage		Room temperature	
Method	[1] & [2]	Simulation	[3]	Simulation	[4]	Simulation
PEALD	1.06 - 1.3	$\sim$1.17	1.0 - 1.1	$<$1.17	0.8 - 1.1	1.1
ALD	1.0	$<$1.17	0.12 - 0.17	$<$0.12	N/A	low

based on the input operating condition. As a result, we will postulate a reasonable surface condition and investigate the exact influence of the gas-phase profile in future work. Moreover, although the proposed kMC model does not directly report the GPC as the experimental probing methods like ellipsometry and X-ray photoelectron spectroscopy (XPS), the lattice constants and the coverage information can be used to estimate the simulation growth rate. Base on the layer thickness of 2.7 Å, investigated from the slab optimization, the GPC is inferred with the accumulated coverage from the kMC simulation. Similar to the majority of experiments, the simulation model reports an overall linear growth with respect to increasing cycle numbers [12, 94].

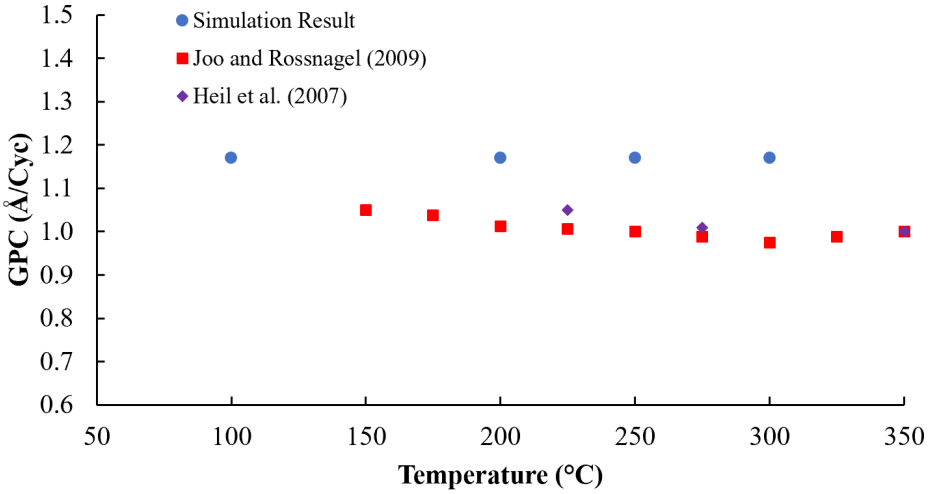
The GPC from the model is first compared with a variety of operating conditions for PEALD and ALD tested in literature, as shown in Table 4.1. The experimental condition is characterized by the substrate temperature and the precursor dosage, which is a function of precursor partial pressure and the half-cycle or pulse time. [103], as shown in Table 4.1[1], uses a substrate temperature of 250 °C and a long precursor pulse time of 7 s with a relatively high reactor pressure of 1 torr, and [105], as shown in Table 4.1[2], reports an even higher reactor pressure of 3 torr but with a slightly lower half-cycle time of 2 s. Under these conditions, preliminary calculations show that the supply is sufficient for the PEALD system and nearly sufficient for the corresponding ALD system. The above literature report PEALD GPC from 1.06 to 1.3 Å and ALD GPC of 1.0 Å, which are similar to the model output, where a steady-steady GPC of around 1.17 Å is reached for PEALD and a slower growth can be reached for ALD, depending on the exact boundary condition. In addition, [94], as shown in Table 4.1[3], investigates a relatively lower substrate temperature of 200 °C, 0.25 s of precursor pulse, 30 sccm flow of oxygen gas and a low flow rate of carrier gas, which can be characterized as a low overall dosage. It reports a slightly lower GPC for PEALD and

a substantially low GPC for ALD. This operating condition is tested with our simulation model, which shows that a limited dosage and reduced temperature have a huge influence on the deposition contributed by neutral O_2 species. Moreover, [104], as shown in Table 4.1[4], focuses on the room temperature PEALD on a small scale reactor with sufficient precursor dosage and reports a GPC of 0.8 to 1.1 Å under different probing methods. The room temperature PEALD is also tested in our simulation model, where we demonstrate that a similar GPC for PEALD can be reproduced and there is nearly no deposition from the neutral O_2 gas.

In addition to the GPC under literature-reported experimental conditions, we also validate our model with the effects of changing operating parameters on the deposition rate. As discussed in the previous paragraph, the substrate temperature is one of the most critical parameters in determining the GPC. However, the substrate temperature can be tricky to manipulate because it may have different effects on the two half-cycles due to different kinetic parameters. First, the influence of substrate temperature on the Hf-Cycle is shown in Figure 4.6(a), and is compared with [106] and [12], labeled as red upper triangles and grey squares, respectively. Both works focus on the Hf-Cycle from the thermal ALD with TDMAHf as the precursor and demonstrate that the temperature has a significant influence on Hf-Cycle. Similar to our model, labeled by the blue dot, the film deposition rate is slightly decreasing in the lower temperature range and has a rebounding trend in the higher temperature range, because the adsorption complex is stabilized under low temperature. The decreasing of GPC in the lower temperature range is more significant according to the data reported by [12] because the experimental film growth includes the impurity from carbon and hydrogen. Those impurities may have come from side reactions that are not considered in the kMC model reaction set. As the temperature increases, the impurity content gradually decreases and causes the apparent growth rate to decrease. Moreover, it can be seen that there is a dramatic increase in the film thickness, because of an undesirable rapid decomposition in the temperature range of 275 °C–300 °C as demonstrated in the FT-IR analysis. Our simulation does not identify this decomposition mechanism because it is a significant variation from the standard deposition reaction set and is known to yield undesirable products. On the other hand, the effect



(a)



(b)

Figure 4.6: Substrate temperature influence on GPC: (a) Influence on the Hf-cycle, and (b) influence on the O-Cycle.

on the plasma half-cycle (O-Cycle) is shown in Figure 4.6(b). [93] shows that the growth rate is fairly constant in the region of 230 °C-350 °C, as shown by the purple diamond. Similarly, [86] has developed a simulation model on the O_2 plasma and validate the plasma's influence on a small scale process. This simulation also shows a steady GPC, labeled by the red squares. A comparable trend is reflected in our simulation. We consider a scenario where a sufficiently supplied Hf-Cycle is followed by a limited plasma dosage (time $\times$ pressure), and the result shows that the temperature

does not have a drastic influence on the GPC, as shown by the blue dot. Additionally, the effect of excessive plasma pulse is investigated by [93] and [94], where no increasing GPC is observed from the extended pulsing time. Our simulation has demonstrated a similar result, in which the surface oscillates around steady states after enough dosage is given.

4.3 Machine learning

The half-cycle time, defined to be the time need to reach the self-limiting steady-state under each precursor pulse, is one of the most important considerations for the industrial thin-film production. According to the developed kMC model, the two half-cycles have a complicated interaction with each other on the structural detail like the existence of steric hindrance and the competition between available reaction cites. Therefore, the input-output relationship between half-cycle time and experimental conditions cannot be easily characterized by a small set of experiments. However, the cost of the research and development limits the exploration of a wide range of operating conditions for optimal throughput. Let alone the cost for energy and equipment maintenance, the precursor, TDMAHf, for a 100-cycle experiment of the industrial size wafer could cost more than 1,000 USD. On the other hand, the developed microscopic kMC model can provide cost-effective information but is still computationally expensive to be implemented in real-time and does not have a closed-form solution, as introduced in Section 1. As a result, it is crucial to collect a database and develop an efficient Neural Network (NN) model that estimates the half-cycle time needed, given a variety of input conditions. The objective of our model is to both investigate a wide range of operating conditions, which is required to predict the boundary of the feasible region and capture the precise input-output relationship, which is needed to determine the optimal operating condition. Thus, the desired model needs to pursue both good predicting precision and generality. Nevertheless, the traditional NNs with backpropagation often struggle with model accuracy and over-fitting. Thus, to successfully achieve both two aims, the Bayesian regularized artificial neural network (BRANN) is chosen as the solution. With the addition of the Bayesian inference to

the standard ANN, BRANN becomes a more robust network for precise regression because the Bayesian regularization algorithm reformulates the complex non-linear regression to a statistically well-posed problem. Also, by incorporating Occam's razor principle to turn off the non-relevant weight in the model training efficiently, the BRANN network penalizes the excessive complexity and resolve the over-training situation [58].

4.3.1 Model construction

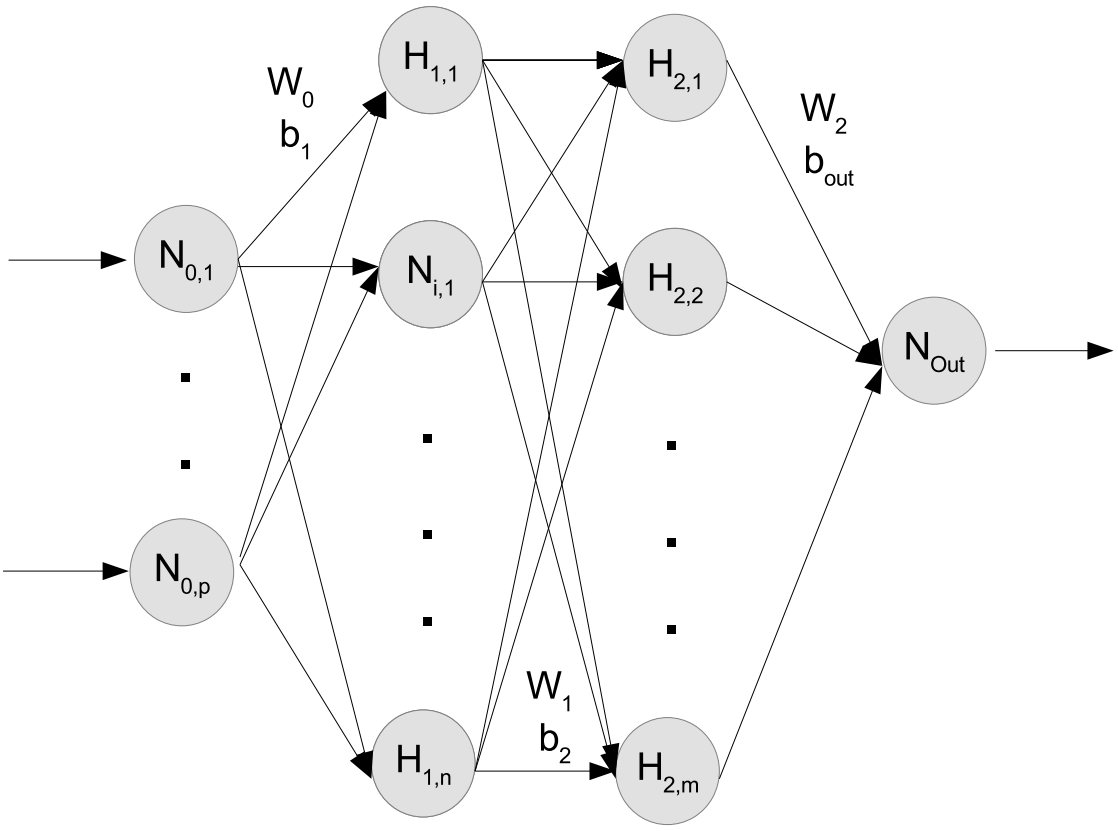


Figure 4.7: Illustration of the densely connected NN.

The construction of the BRANN network is implemented in the MATLAB machine learning toolbox. A dense connection is chosen, where all neurons in the neighbor layers are connected. A two-layer structure example is shown in Figure 4.7, and the mathematical formulation of the

neural network is shown as follows:

$$N_{1,j} = f_0(\sum_{i=1}^{p} W_{0,ij} N_{0,i} + B_{1,j}) \tag{4.11}$$

$$N_{2,j} = f_2(\sum_{i=1}^{n} W_{1,ij} N_{1,i} + B_{2,j}) \tag{4.12}$$

$$N_{out,j} = f_3(\sum_{i=1}^{m} W_{2,ij} N_{2,i} + B_{out,j}) \tag{4.13}$$

where each equation describes the mathematical relationship between the connected layers. N denotes the neuron, W denotes the weight matrix, and B denotes the bias matrix. The subscript notation 0 refers to the input layer, 1 and 2 refer to the first and second hidden layer, respectively, and *out* refers to the output layer. In each equation, i and j refer to the index of neurons in the previous and the subsequent layer, and ij refers to the matrix index in the weight matrix relating to i and j. p, n, and m describe the total number of neurons in the input, first and the second layer, respectively. In addition to the linear combination, activation function f is added to allow for the NN to describe the non-linearity. In the BRANN model, instead of the usual Rectified Linear Unit (ReLU) function, the hyperbolic tangent sigmoid function (tansig) is used as the transfer function:

$$tansig(x) = \frac{2}{1 + \exp(-2x)} - 1 \tag{4.14}$$

where x is the value passed in the network. This function has a range of [0,1], which makes it the right candidate to fit functions that map into probability spaces because the network can interpret the probability of an event in a more intuitive manner.

The goal of the training is to minimize the cost function, S(w), with respect to the parameters in the aforementioned structure. Unlike the traditional mean square error (MSE) function, which is typically used for regression application, the BRANN network adopts additional hyperparameters

to regularize the cost function as follows:

$$S(w) = \beta \sum_{i=1}^{N_D} (y_i - f(x_i, w))^2 + \alpha \sum_{j=1}^{N_w} w_j^2 \tag{4.15}$$

where i is the index for the data sample, N_D is the total number of training data, y_i is the desired output value of ith entry, and $f(x_i, w)$ is the predicted value dependent on w and the input x_i. w is the weight vector, which is initialized through Gaussian distribution, j is the index for the weight, and N_w is the total number of weight parameters. α and β are the hyperparameters for the weight regularization, initialized from a uniform random distribution. In order to arrive at the optimal weight vector, the maximum a posteriori (MAP) method is solved with the Levenberg–Marquardt algorithm (LMA) [59]. Similar to other numerical optimization algorithms, the LMA is an iterative procedure. It behaves like a gradient-descent method when the parameters are far from the optimal value, while it becomes more similar to the Gauss-Newton method when the parameters are close to optimal. Compared to the RMSProps and gradient descent method, LMA is demonstrated to have a more stable convergence progressing toward the global minimum [107]. The original backpropagation is modified with the addition of Bayes' theorem to approximate the exact Bayesian inference of the posterior probability, which leads the network to faster regularization and eliminates the need for dropout:

$$P(w|D) = \frac{P(w)P(D|w)}{P(D)} \tag{4.16}$$

where D refers to the dataset. In the meantime, the hyperparameters also need to be iteratively updated during each epoch according to Bayesian inference. This dual step optimization is repeated until the maximum epoch self-consistency is achieved. The detailed network construction and theory explanation are discussed in [58] and [60].

4.3.2 Dataset generation

In advance of the model training, a database should be generated, which contains the respective half-cycle times corresponding to a reasonable range of input operating conditions. Therefore, we have developed an automated workflow to run the kMC model for the two half-cycles and to test a variety of operating condition combinations. As shown in Figure 5.1, the top box specifies the kMC model workflow. The simulation starts at $t = 0$ where the operating temperature and partial pressure for the Hf-Cycle are given to the kMC model. The event selection scheme is executed and the time progression is calculated until the model reaches the termination criterion, where the surface is saturated under the restriction from steric hindrance. Then, the structural detail from the Hf-Cycle is passed to the O-Cycle, and the partial pressures for the O-Cycle precursors are given to the model as inputs. This workflow could be repeated until the desired cycle number is achieved. The overall time needed is stored as the final result. The automated workflow is utilized to explore a broad range of operating condition combinations. To boost the computational speed and efficiently utilize all available computational resources, we adopt a distributed computation using the UCLA Hoffman2 cluster. With the developed workflow, the kMC models are submitted as computation jobs to the Hoffman2 job queue, and the job scheduler will distribute the jobs onto different cores to allow parallel runs of the kMC models. Because of the competition of the reverse and forward reaction and the stochastic nature of kMC, the final lattice condition may vary under different input combinations and keep fluctuating. Nevertheless, the overall coverage will converge to steady-state, where the film structure does not differ dramatically. Thus, for the NN characterization, the half-cycle time needed for the kMC model to reach 99.7% of the final coverage is taken to minimize the noise associated with the fluctuation. The results are collected and compiled automatically. In this work, a single cycle simulation is investigated under a set of operating conditions where the substrate temperature ranges from 298 K to 503 K, and the overall pressure ranges from 1 Pa to 60 Pa, with an interval of 1 K and 1 Pa, respectively. The available computation resources allow us to run 500 jobs actively in parallel and queue over 4000 jobs automatically for execution. The full database can be generated under two weeks. In addition,

the database is further randomly divided into training, validation, and testing dataset with a ratio of 7:1.5:1.5. The training and validation datasets are applied in the model parameter updating and evaluation, while the testing dataset is used to assess the final model performance.

4.3.3 NN result and cycle time optimization

The numbers of neurons and layers are optimized via a grid search [55]. In particular, the input layer consists of two neurons, which represent the surface temperature and the total precursor pressure, respectively. The output layer consists of two neurons, which represent each of the half-cycle completion times to reach full coverage. Two hidden layers are included, which consist of 20 and 15 neurons, respectively. Good training and validation accuracy have been achieved as shown in Figure 4.8 (a), in which we can see that the training error quickly drops and the validation mean square converges to $2.84 \cdot 10^{-4}$ s^2. The fluctuation at the beginning of the training could be attributed to the stochastic nature of Bayesian inference used by BRANN. As the training progress moves towards convergence, the curve is smoother because the parameters are close to optimal values. The testing dataset is compared with the model prediction, where the NN achieves a mean absolute error of 0.017 s. Such a low testing error demonstrates a good model accuracy because it is already in the range of steady-state variation of the kMC runs. In Figure 4.8(b) and (c), the error distribution histograms demonstrate that, for both half-cycles, the deviation of the prediction from the actual half-cycle completion highly resemble the normal distribution, which demonstrates that the NN fitting is unbiased. Additionally, in Figure 4.8(d) and (e), the R-squared correlations between the kMC simulation results and the predicted time are 1.00 for both half-cycles, which demonstrates that the fitting result is in good agreement with the original kMC output. The grid search is completed on two computers with 64G RAM and 8-cores in one day. After the NN is trained, it takes less than one second to perform the prediction using the same computer.

With the trained BRANN model, we can explore the operating conditions to optimize the half-cycle time. An extended half-cycle time is not ideal for the PEALD process as it significantly reduces the industrial throughput and incurs unnecessary operating costs. Currently, appropriate

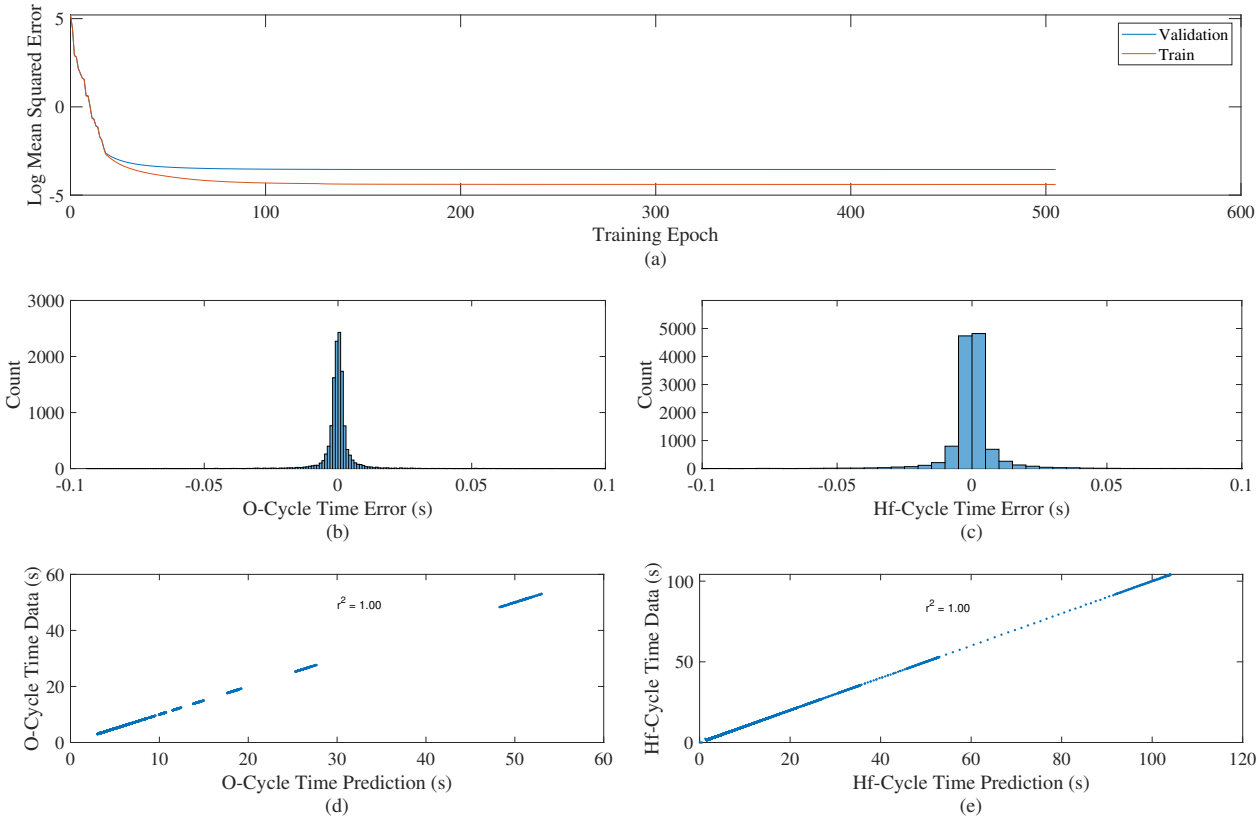
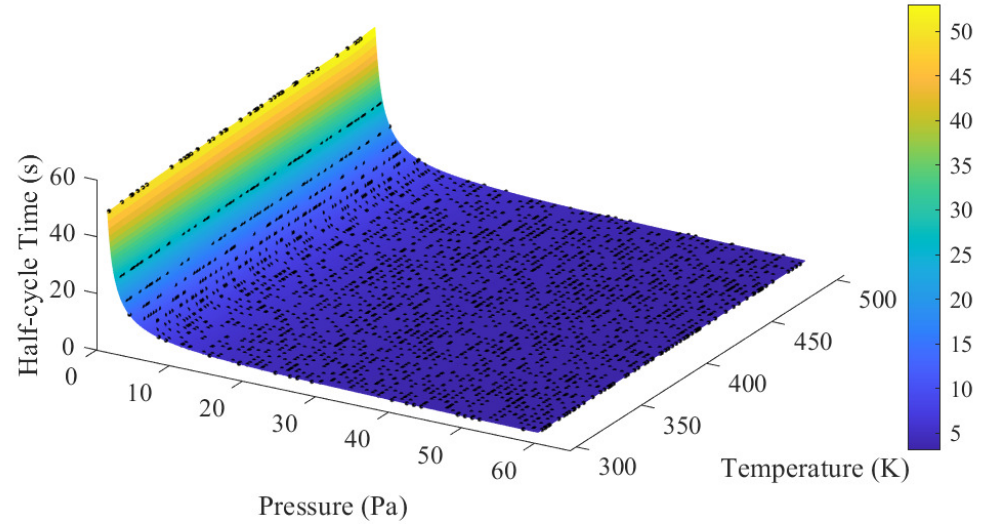
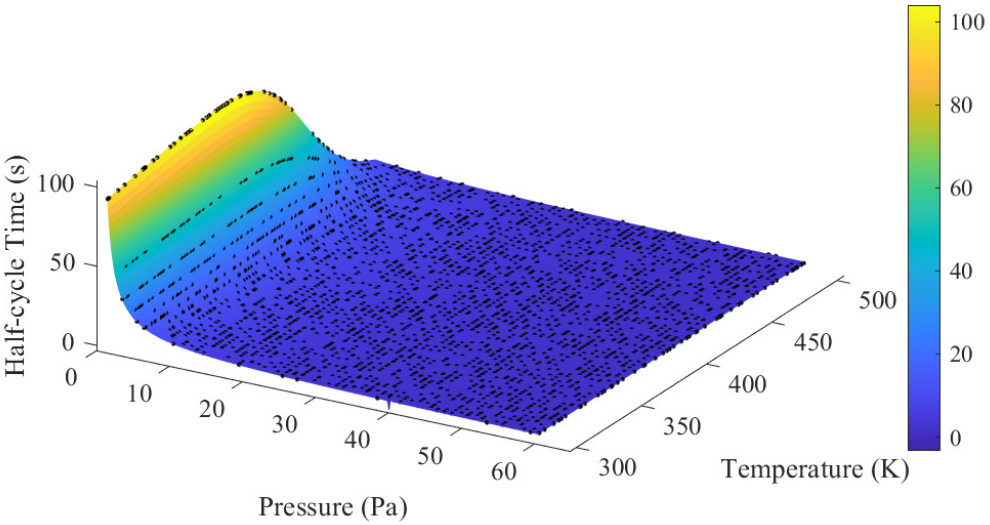


Figure 4.8: Performance of the NN: (a) Training and validation error over the training process. (b) Prediction error distribution histogram for O-Cycle. (c) Prediction error distribution histogram for Hf-Cycle. (d), (e) Correlation accuracy of the predicted time and the actual time for O-Cycle and Hf-Cycle, respectively, where the x-axis is the predicted time to reach steady-state from the neural network and the y-axis is the actual time to reach steady-state from the kMC model.

operating conditions are investigated through trial and error for each reactor design. For example, various works have investigated the O-Cycle pulse time ranging from 1 s to 60 s and the substrate temperature ranging from room temperature to 375 °C. Thus, a feasible operating range is valuable to be determined so that lab works can be completed with a reasonable experimental. A five-second time frame for each half cycle is chosen as the reference half-cycle time, which is the typical production time of a thin-film of the modern gate oxide. By using the developed BRANN model, as shown in Figure 4.10(a), we can identify the feasible range of operating conditions where both cycles can finish under 5 s. However, as discussed in Section 4.2.6, the time for gas-phase transport should also be accounted for. Our preliminary study has shown that a radial distributed gas-phase transport will lead to a typical development time ranges from 1 s to 1.5 s for a standard remote



(a)



(b)

Figure 4.9: Neural network fitting results and original training data. x-axis is pressure (Pa). y-axis is temperature (K). z-axis is the half-cycle time from prediction and training data. Black dots denote the training data, and the colormap denotes the magnitude of predicted time from the neural network. (a) NN fitting result for O-Cycle. (b) NN fitting result for Hf-Cycle.

PEALD reactor design. This rather long time range makes sure the prevention of potential damage onto the substrate surface through the high-energy plasma sheathing and high-speed perpendicular

flow. The BRANN model is also utilized to estimate the feasible operating region by subtracting the gas-phase transport time from the effective half-cycle time. As shown in Figure 4.10(b), when 1 s of gas-phase transport time is accounted for, there is a significant change and reduction of the feasible operating range. The curvature of boundary shown in Figure 10 is largely due to the nature of the kinetics. Since our kMC model modifies the standard kMC algorithm to include transient concentrations as weights, the direct relationship between the boundary and the rate of each reaction is not very straightforward. However, the nonlinearity in the boundary could provide us some insight into the overall trend of reaction rates at certain operating conditions. For example, at P = 18 Pa and T = 480 K, due to the increase in temperature, the desorption rate would increase to a point where it surpasses the increase of the forward deposition reaction rates. This overall trend shift would cause the feasibility region to shrink and lead to the kink at the boundary. Additionally, the narrow P-zone is due to the consideration of the operation of PEALD under fixed temperature and pressure in this work. In realistic operations, low P-zone would typically only exist a short amount of time, therefore it would not lead to the final incompletion.

Finally, we utilize the developed BRANN model to perform process optimization. As demonstrated in Table 4.2, a series of operations under fixed operating conditions are examined, and the predicted total time for both cycles are compared with the conventional fixed total time of 10 s. The precursor partial pressure is limited under 40 Pa to reduce the precursor cost and the substrate temperature is limited under 150 °C (423 K) to simulate a low-temperature deposition scenario which is critical to the temperature-sensitive materials. For experimental conditions 1 and 2 in Table 4.2, if the conventional fixed time is used, the product will be uncompleted because insufficient time is given for the cycle. For experimental conditions 3, 4, and 5 in Table 4.2, by using the accurate prediction from the NN model, the operation time could be reduced by up to 16%, which corresponds to a significant enhancement in the industrial throughput and operating cost.

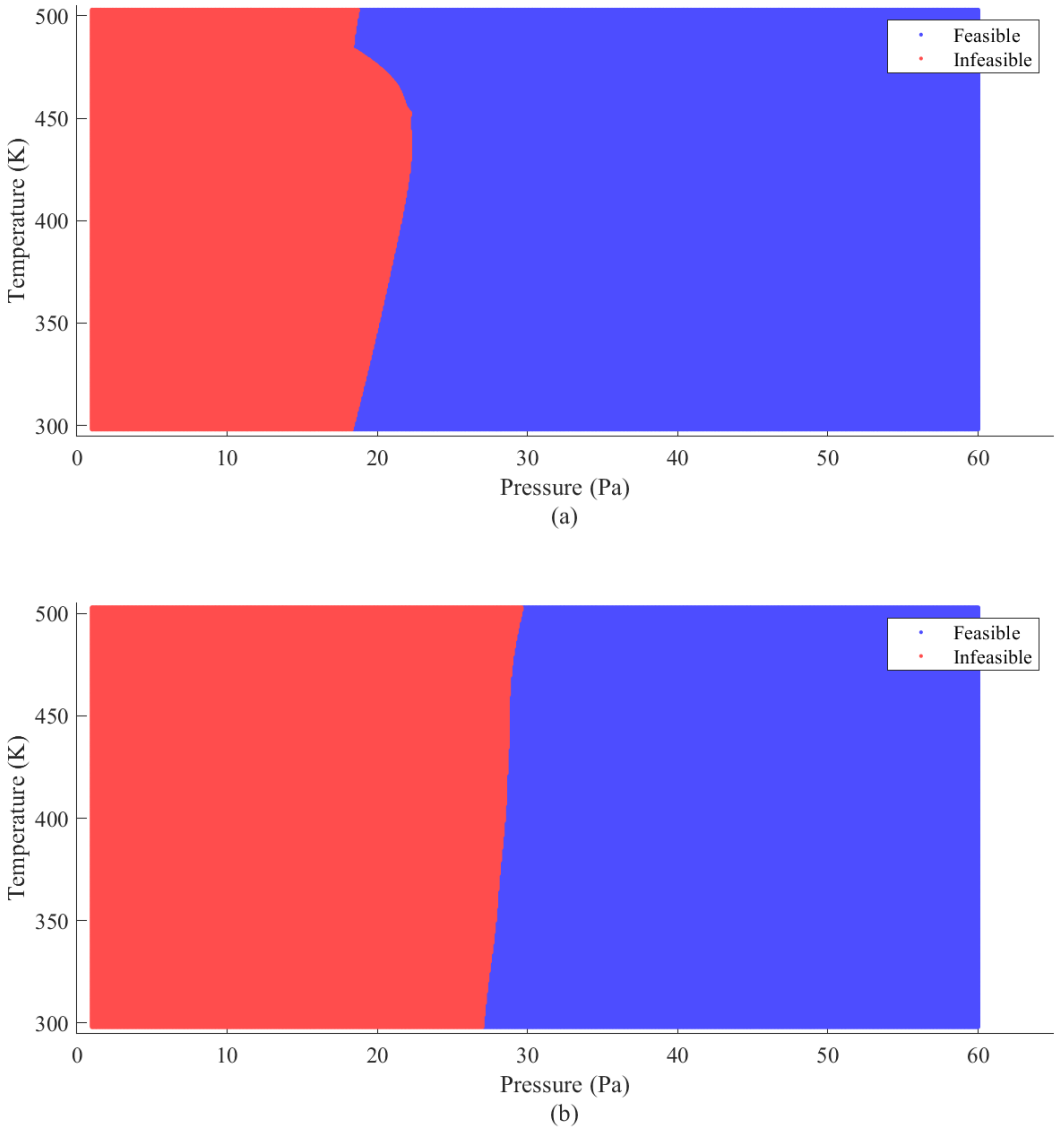


Figure 4.10: Operational feasibility of HfO_2 under fixed operating conditions. The blue region denotes the operating conditions where both half-cycles can be completed under desired half-cycle time, whereas the red region denotes otherwise. (a) Effective half-cycle time = 5 s, without considering gas-phase transport. (b) Effective half-cycle time = 4 s, accounting for 1 s of gas-phase transport.

4.4 Conclusion

In this work, we developed a microscopic model of the PEALD of HfO_2 using TDMAHf and oxygen plasma as precursors. A specially tailored kMC algorithm is utilized, which is dependent

Table 4.2: Temperature, pressure, predicted total time and conventional total time for process optimization. Predicted total time is also corrected with the one-second gas-phase transport time.

#	Temperature (K)	Pressure (Pa)	Predicted Total Time (s)	Original Total Time (s)
1	300	20	11.20	10.0
2	330	25	10.05	10.0
3	360	30	9.28	10.0
4	390	35	8.71	10.0
5	420	40	8.37	10.0

on the DFT calculation and is validated by experimental results. The GPC reported from the simulation model deviates from the experimental results up to 6.3 % under various operating conditions. Moreover, the model successfully characterized the influence of individual operating parameters on the GPC that are observed in the existing literature. After the simulation model has been validated, we constructed a BRANN model based on the database generated from the kMC model for a wide range of operating conditions. Using the BRANN model, we performed data analysis to determine the feasible operating region and implemented process optimization, where a reduction of half-cycle time up to 16% can be achieved. Instead of using the trial and error approach, the kMC model provides an operating guideline for both industrial manufacturing and experimental researches. By using the optimal time predicted by the model, the production throughput can be greatly improved, along with a significant reduction for labor and operating costs. It is notable that the developed kMC scheme is not limited to the deposition of HfO_2. It could be extended to the deposition of any XY_2-type compound with similar reaction routes. Therefore, in order to implement a generic microscopic PEALD model for XY_2-type compound, we generalized our kMC scheme and implemented the model in Python to take advantage of its modularity and accessibility. Thus, the developed model formulated a general methodology to explore the precursor applicability, boost the PEALD throughput, and reduce the operating cost.

Chapter 5

Machine Learning-Based Modeling and Operation of Plasma-Enhanced Atomic Layer Deposition of Hafnium Oxide Thin Films

5.1 Introduction

In this chapter, the construction of the data-driven model for PEALD is described. As mentioned in the previous chapter, simulation model is often a low-cost solution and can provide insights on the entire process operating domain. A variety of attempts have been made to capture the PEALD dynamics. To model the plasma generation and transport, [90] created a simulation model for O_2 plasma in a simplified remote inductively coupled plasma (ICP) reactor configuration, and [108] demonstrated a comparison between the experimental and simulation results of the similar plasma source. For the macroscopic gas-phase transport domain, computational fluid dynamics (CFD) softwares are widely utilized [109, 110]. For example, [111] constructed a CFD model to capture the chemical vapor deposition of amorphous silicon thin-films, and [66] designed

a CFD characterization of the SiO_2 ALD and optimized the showerhead reactor geometry. For the microscopic surface reaction domain, [8] developed a small-scale, high accuracy kinetic Monte Carlo (kMC) model for the ALD of HfO_2 accounting for the complete set of reactions. Adding the plasma half-cycle and adopting reduced reaction sets, [112] extended the kMC model to be used for PEALD. In addition, [113] recently formulated a brand-new multiscale CFD model that comprehensively captures all three individual domains for an industrial-scale PEALD process. Although the developed model is able to accurately describe the PEALD process, the time required to perform the simulation is much longer than the actual PEALD process time-scale and the computational resource needed is not locally obtainable, thus making it infeasible to be applied in the context of industrial on-line operational optimization.

In response to the aforementioned difficulty, data-driven modeling has been demonstrated as one of the most promising solutions by taking advantage of machine learning methodologies, especially neural networks. Attempts have been made to utilize feedforward neural network (FNN) to characterize the result of kMC simulation [31, 67]. The previous work by [28] developed a multiscale data-driven model using a decoupled linear parameter-varying model for the gas-phase domain and an FNN model for the microscopic domain. Although the FNN model was able to accurately characterize the steady-state input-output relationship, the dynamic system provides more information because of the transient change of the flow field and key thermodynamic and fluid mechanical properties. Several deep learning formulations like recurrent neural network (RNN) and convolutional neural network (CNN) are utilized to model sequential and temporal systems [114]. Among those deep neural network models, RNN is intuitively more suitable and has demonstrated its outstanding performance in text summarization and natural language processing [115]. Recently, it has also been applied in modeling many industrial systems [116]. For example, [117] investigated Phthalic Anhydride Synthesis based on CFD simulation and constructed a computationally efficient RNN characterization of the process. [118] utilized RNN to characterize a general nonlinear system under cyber-attack threats. Nevertheless, until this point, such a data-driven model has not yet been developed for thin film deposition processes

and especially for the industrial-scale PEALD system.

Motivated by the above considerations, in this chapter, we first construct a database using the previously developed multiscale CFD model [113]. Then, based on the collected data sets, a data-driven model is developed to incorporate the gas-phase transport profiles and the microscopic surface dynamics. Fully integrating both two domains into a unified model, the RNN-based data-driven model can accurately capture the interaction of both domains within each timestep. In this way, the data-driven model is able to fully preserve the input-output relationships between operating conditions and the deposition profile, while reducing the computational demand by taking advantage of existing data sets. As a result, the proposed integrated data-driven model allows both fast prediction of necessary cycle-time to achieve full coverage, and online estimation of film quality through the embedded information of transient dynamics. A significant amount of economic benefit can be potentially realized by the developed model. For example, the cost of precursor and operation to explore feasible operation conditions, estimated to be millions of dollars, could be substantially avoided. More importantly, such a method can serve as a general framework and is easily modifiable and extendable for other deposition processes.

5.2 Multiscale CFD Modeling of PEALD

While details can be found in [113], this section will briefly introduce each component of the developed HfO_2 thin-film PEALD multiscale CFD model. The simulations of ICP plasma generation in the remote plasma domain, gas-phase transport phenomena in the main ALD reactor chamber, and microscopic surface film deposition using kinetic Monte-Carlo (kMC) are reviewed.

5.2.1 Plasma Chamber CFD Model

The remote plasma system constructed in [113] is an inductively coupled plasma (ICP)-source, which utilizes alternative current (AC)-induced magnetic field to produce high purity and density plasma. The geometry of the simulated plasma generator is based on the Gaseous Electronics

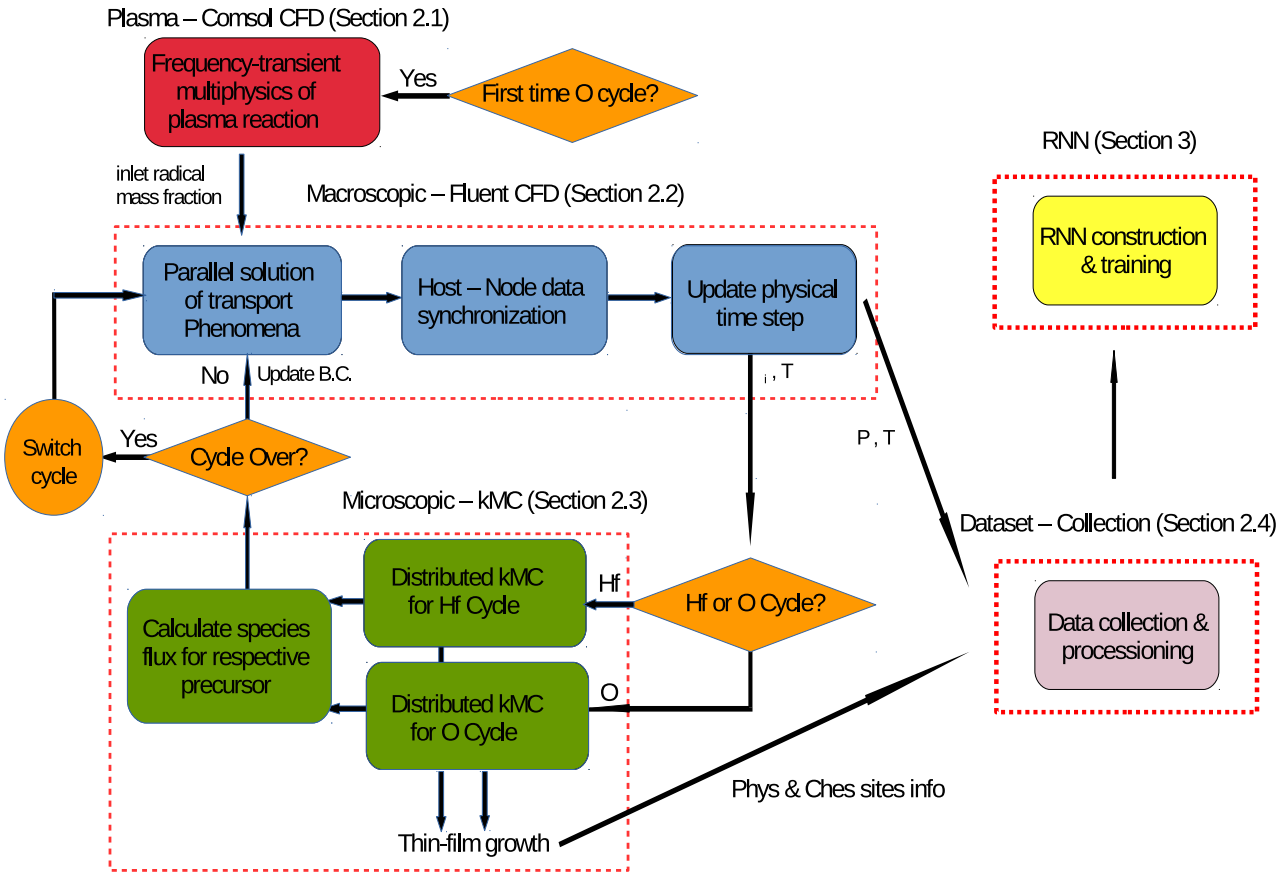


Figure 5.1: Integration of the multiscale computational fluid dynamics model, dataset generation and recurrent neural network development.

Conference (GEC) cell, a standard experimental and modeling prototype, proposed by the National Institute of Standards and Technology (NIST). For better connectivity to the main reactor, the outlet is modified as shown in Figure 5.2.

The AC generator operates at a power of 2000 W, and the current radio-frequency (RF) is 13.56 MHz. The changes in the electric field and the current density caused by the AC generate a magnetic field, which can be described by the Ampere-Maxwell equation as follows:

$$\vec{\nabla}(\varepsilon_{\mathrm{f}}\vec{B}) = \vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{5.1}$$

where ε_{f} is the electric permeability of materials, $\vec{B}$ is the magnetic field, $\vec{J}$ is the current density, ε_0 is the electric permeability of free space, and t is the time.

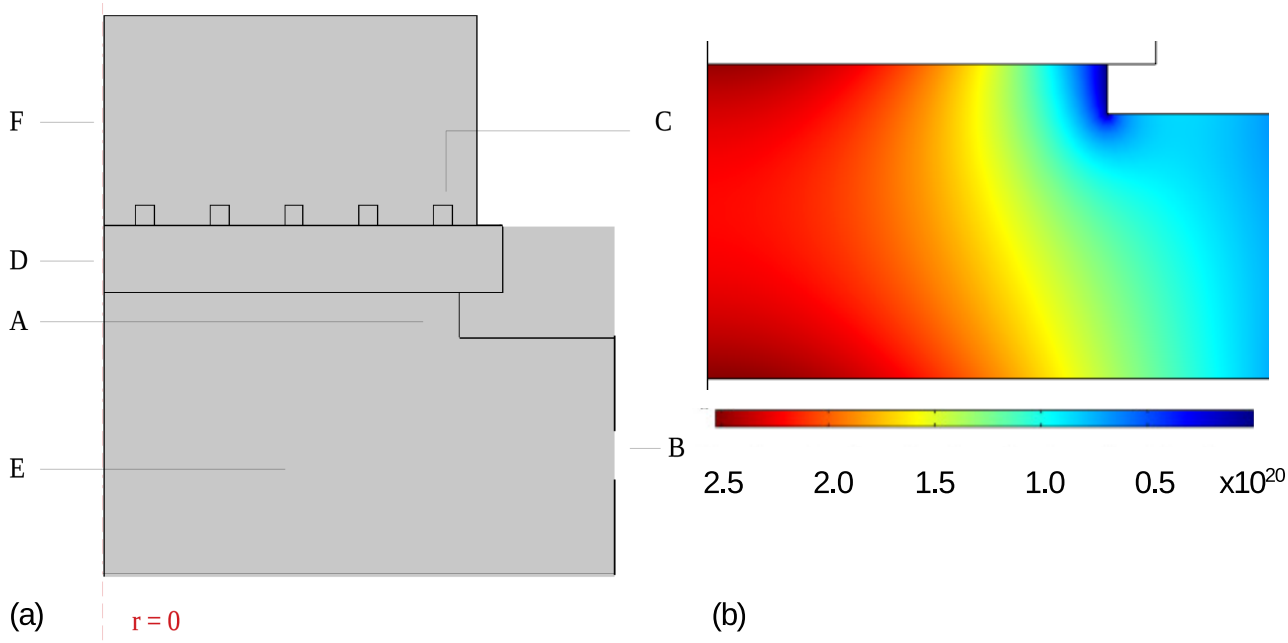


Figure 5.2: (a) Plasma generation chamber axis-symmetry. A - gas inlet, B - gas outlet, C - coil circuits, D - circuit wall, E - reactor chamber, and F - relevant surrounding. (b) Example of number density distribution of O radical, an important plasma species involved in the deposition reaction.

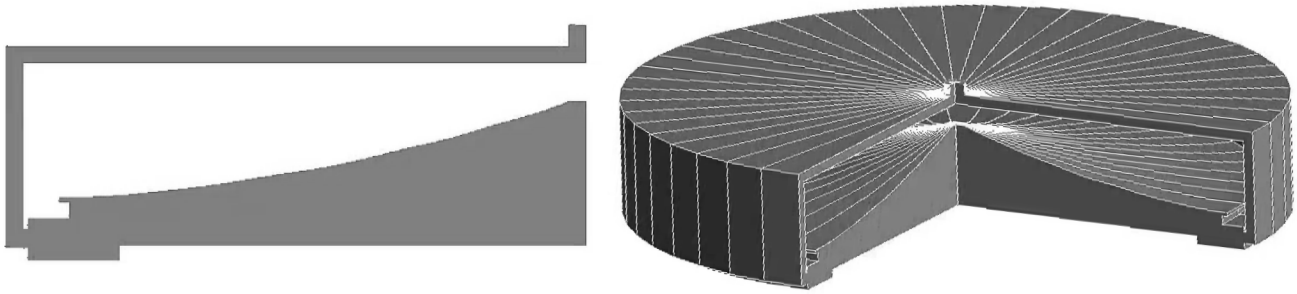
Inside the plasma chamber, a constant flow of argon is used to maintain the baseline pressure. Oxygen is injected through an inlet valve and the AC circuit is activated right before the initiation of the oxygen cycle. Three types of reactions are associated with the Ar/O_2 plasma generation: the electron impact reactions, the heavy particle reactions, and the surface reactions. The detailed reaction sets and rate characterization can be found in [113]. The developed plasma domain model is simulated in COMSOL Multiphysics, which integrates the AC/DC module, the laminar flow module, the heat-transfer-in-fluid module, and the plasma reaction module. A frequency-transient solution is obtained to investigate the plasma reaction and generation, and it provides the ion and radical profiles throughout the plasma chamber, which is then used as the inlet profile of the subsequent macroscopic gas-phase domain. Due to the difference in model time-scales, the plasma generation simulation can be decoupled from the gas transport in the main reactor chamber.

5.2.2 Macroscopic CFD Model

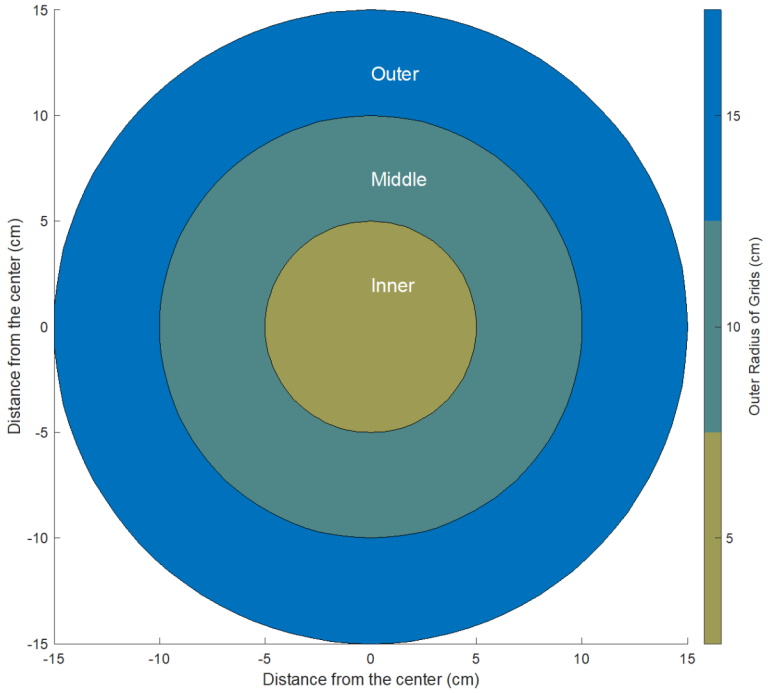
The macroscopic transport phenomena in the main reactor chamber directly affect the dynamics of surface reactions as well as the surface configuration of the deposited HfO_2 thin-film. The bulk reactor geometry used in this work is an optimized design based on the ASM International EmerALD XP reactor [19]. Specifically, as shown in Figure 5.3 [70], the reactor chamber design is modified to enhance the uniformity of the gas profile and is scaled up to allow the adoption of 300-mm diameter wafers. Precursors carried by inert argon gas are injected into the reactor from a 1-cm diameter inlet located on the top of the reactor. These gases first enter the upstream of the reactor, where the horn-shaped design facilitates the uniform flow distribution. Then, the gas species move downward and potentially encounter a showerhead panel, if used. After crossing the reactor downstream, precursors reach the substrate surface and deposition reactions occur. Finally, leftover precursors and carrier gas are pumped out of the reactor through the outlet channel. More details regarding meshing and reactor geometry are included in [113].

Twelve gas-phase species are characterized in the model: argon is the purging and carrier gas; TDMAHf is the Hf-cycle precursor; oxygen and its ionic and excited states are the O-cycle precursors. Volumetric and surface reaction sets involving all species are defined in the model, and the relevant reaction sets and parameters are reported by [113]. A 60 Pa base operating pressure is used, which is maintained through a constant flow of 300 sccm Ar and outlet pumping. The absolute pressure in each half-cycle may fluctuate due to precursor injection in each cycle. The baseline precursor feed flowrate at the inlet is equivalent to the feed from a gas-delivery bubbler at 348.15 K, which is the required temperature for TDMAHf vapor pressure to reach 1 Torr [119]. Using this setup, the governing transport equations are solved to obtain the profiles of velocity, pressure, and molar fraction of species in the gas-phase [72].

In addition, we use ANSYS Fluent for macroscopic domain simulation, which applies the finite volume method to solve the coupled governing equations, utilizing previously developed hybrid mesh [120]. Simulation accuracy and efficiency are greatly affected by the chosen calculation timestep [121]. The timestep size can be related to the number of cells that the fluid information



(a) (b)



(c)

Figure 5.3: (a) 2D axisymmetric PEALD reactor geometry. (b) 3D PEALD reactor geometry. (c) Indication of the inner, middle, and outer regions of the wafer.

travels in each timestep, known as the Courant number, C, the length of the spatial interval Δx, and the speed that information travels through the cells, u, according to the Courant-Friedrichs-Lewy (CFL) condition [74]:

$$\Delta t = \frac{C\Delta x}{u} \tag{5.2}$$

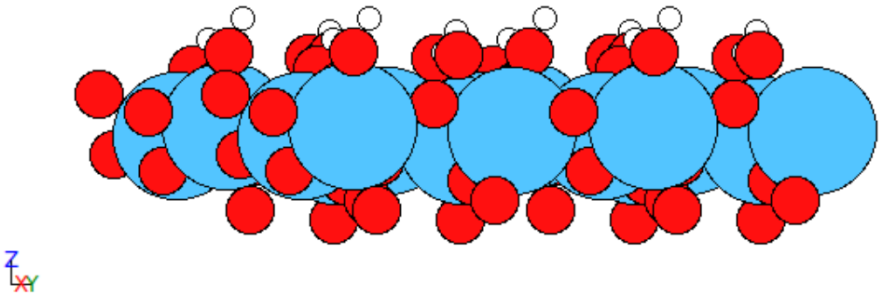
CFD computation time can be largely reduced when advanced CFD software is used [74]. Specifically, the pressure-based solver in this work uses a Courant number of 50.

5.2.3 Microscopic Surface Model

Besides reactor design and gas-phase transport, substrate surface reactions are necessary to be discussed at a microscopic level. The microscopic kMC model developed in [112] accurately characterizes the key surface reaction mechanisms and the developed HfO_2 thin-film structures. In the following subsections, HfO_2 structure, reaction kinetics and pathways, and the 3D kMC model are briefly presented, while more details can be found in [112].

A monoclinic-alike structure, reported in [89], is adopted for the simulation of low-temperature PEALD application. The hydroxyl-terminated (111) surface is assumed to be the starting surface, as shown in Figure 5.4. A 3D triangular on-lattice approximation developed in [63] is used to efficiently represent this surface as shown in Figure 5.4. The approximated lattice retains the connectivities between atoms and the cycle repetition pattern that are observed in the real lattice structure. The simulated lattice dimensions are set to be $1200\times1200\times N_{layer}$ to ensure size-independence [35].

Within each half-cycle, precursors associated with the cycle undergo surface reactions and deposit a layer of atoms of the desired element. Similar to the thermal ALD process, TDMAHf serves as the precursor for the Hf-Cycle. However, for the O-cycle, in addition to molecular oxygen, atomic oxygen and their excited and ionized derivatives from the macroscopic scale simulation are also involved in the deposition reactions. A simplified reaction set, selected from the complete mechanism reported in [8] and [91], is used to model the key surface dynamics while neglecting the proton diffusion, ligands rotation, and other detailed features. This simplification dramatically decreases the computational load, enabling industrial-sized wafer-scale simulation, and preserving key fidelity to the reaction mechanisms. The resulting mechanisms are demonstrated in Figure 5.5. Specifically, in the Hf-Cycle, TDMAHf goes through two steps of dissociative chemisorption, binding the hafnium atoms onto surface oxygen atoms with two terminating dimethylamine (DMA) groups and releasing two remaining DMA groups. In the O-Cycle, ground state O_2 particles and radicals go through their respective reaction pathways and eventually bind the O atoms with terminating hydroxyl groups, releasing different byproducts.



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

Figure 5.4: Surface slab and approximated lattice. Top: hydroxyl-terminated HfO_2 slab. Bottom: The approximated lattice with examples of adsorbed species. PsHf and CsHf represent the physisorbed and chemisorbed precursor, respectively, HfL2 represents the deposited Hf terminated with two dimethylamino ligands, and PsO1 and PsO2 represent two distinctively physisorbed oxygens, respectively.

The transition state theory (TST) [47] and the collision theory are used to characterize the rates of these reactions. The general Arrhenius-type equation from the TST can be used to describe thermodynamically activated reactions as follows:

$$r_{rxn} = A\exp\left(\frac{-E_a}{RT}\right) \tag{5.3}$$

where E_a is the transition state complex activation energy, A is the pre-exponential factor, R is

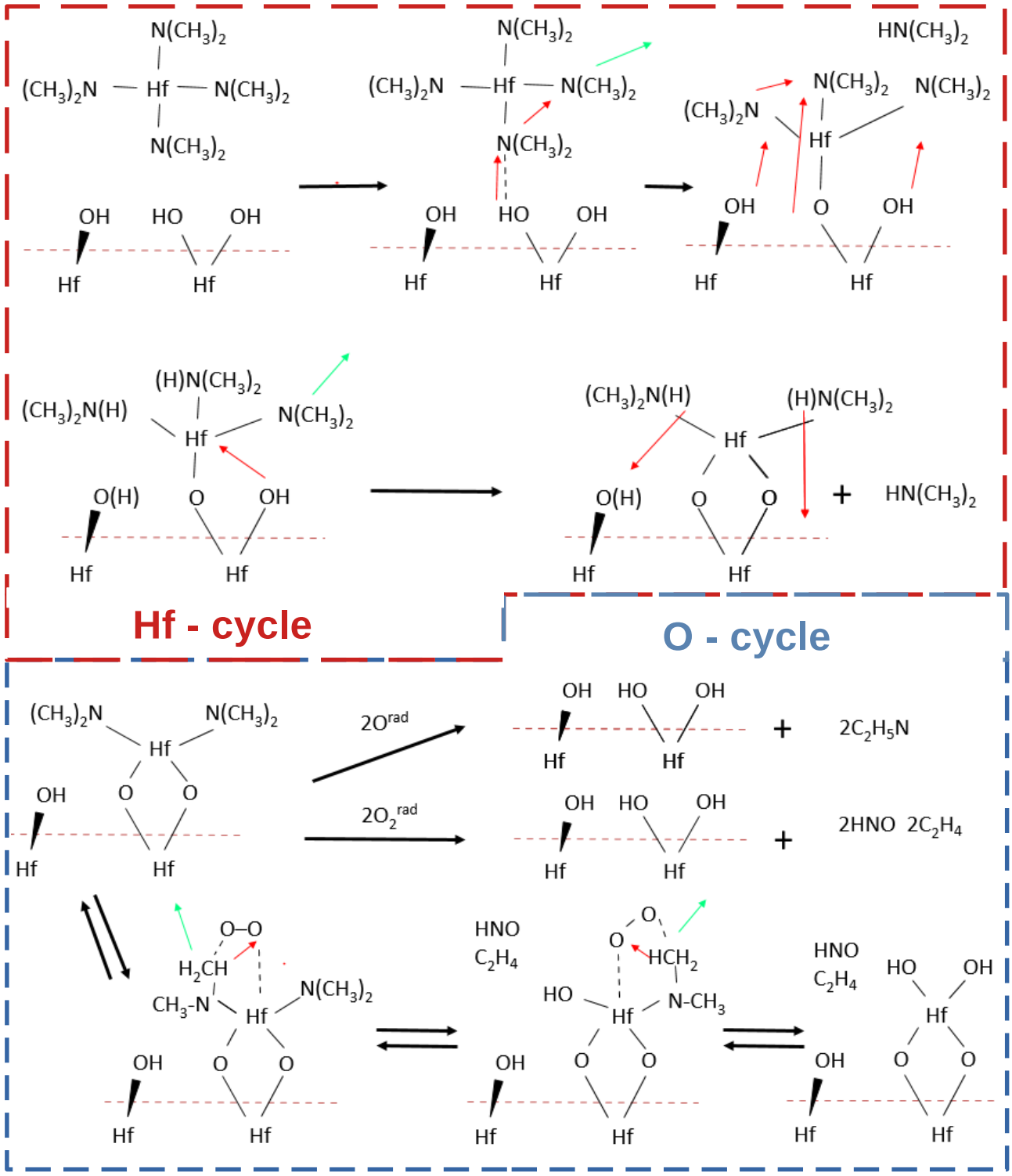


Figure 5.5: Illustration of reaction mechanisms for the precursor surface reactions. The black arrows denote the reaction pathways, and the red arrows denote potential proton diffusion.

the gas constant, and T is the substrate temperature. In contrast to the TST, the collision theory is used to describe the gas-surface athermal barrierless reactions such as physisorption. The rate formulation provided by the collision theory is shown as follows:

$$r_{phs} = \frac{p}{RT}\sqrt{\frac{8RT}{\pi m}} s_c N_a \sigma \tag{5.4}$$

where m is the molar weight of the precursor, σ is the unit cell surface area, N_a is the Avogadro number, and s_c is the sticking coefficient, as reported in [96].

The reaction selection and time evolution are then determined using the modified n-fold hybrid kMC algorithm by [63] to simulate the film growth and the deposition profile. The total reaction rate r_{total} can be calculated as the sum of each rate:

$$r_{total} = \sum_{i=1}^{N} r_i \tag{5.5}$$

where r_i represents the respective rate of N individual events. The reaction selection uses a random number $\gamma_1 \in (0,1]$ to locate the event according to its normalized indicator, $l_i \in (0,1]$, which is the sum of the normalized probabilities of the previous events:

$$l_i = \frac{\sum_{j=1}^{i} r^j_{unweighted} C_j}{\sum_{k=1}^{N} r^k_{unweighted} C_k}, \;\; i = 1,...,N \tag{5.6}$$

where the rate of reaction, $r_{unweighted}$, will be adjusted with respect to the availability of reactant, C, in each reaction, as explained in the previous work by [112]. In the case where the selected random number is between l_{i-1} to l_i, the i-th event will be executed. For the time evolution, a second random number $\gamma_2 \in (0,1]$ is used to calculate the elapsed time for each event:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{5.7}$$

5.2.4 Multiscale CFD Model Customization, Workflow and Data Generation

Figure 5.1 shows the automated workflow of the CFD simulation and RNN development and provides an overview of how information is relayed through each domain in this work. The simulation starts at COMSOL, if it is in the O-cycle, to solve for the plasma profile. Next, Fluent solves for macroscopic solution for one timestep. The surface partial pressure and temperature conditions are collected for each surface region and then transferred to the microscopic domain. Next, the kMC model will simulate the surface reaction and store the evolution of surface structure within the prescribed time interval. The synchronized kMC results will then be fed into Fluent to update the boundary condition for the next timestep. The simulation time clock continuously adds Δt to itself after each event until it reaches the designated half-cycle time. The subsequent half-cycle specification will then be updated for all domains. After each timestep, the macroscopic partial pressure of each important gas-phase species, as well as the surface site information (the amount of Hf physisorption sites and O physisorption sites), are collected in the database.

The aforementioned methodology serves as a general-purpose guideline for the multiscale model construction for the PEALD process, and it is subject to customization under various accuracy requirements and computational limitations. In this work, due to a large amount of potential operating conditions to explore, 2D axisymmetric CFD models are constructed for reactors with and without the showerhead. Moreover, surface kMC models are executed on the concentric rings at the wafer surface in the 2D multiscale CFD domain, instead of the grids as in the 3D multiscale CFD model, which is shown in Figure 5.6. Details regarding the 3D grid partition can be found in [112]. As we will demonstrate in the latter sections, the reduction from 3D to 2D will not affect the accuracy of the domain profile without the showerhead due to its axisymmetric nature. Moreover, for the reactor with showerhead, despite the existence of noticeable profile deviation, the results are consistent throughout the operating domain. Therefore, the 2D multiscale CFD model results for the showerhead reactor can provide general insights on the realistic 3D operation.

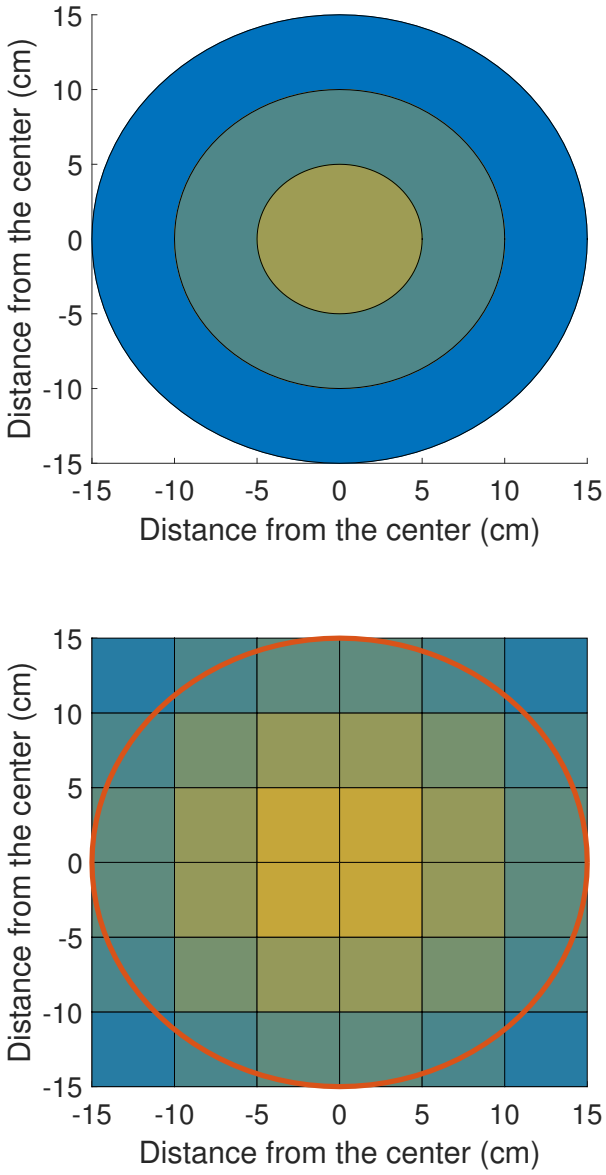


Figure 5.6: Comparison of 2D and 3D surface region partition. Color from yellow to blue represents the distance between the center of the region to the center of the wafer. (a) 2D surface concentric rings region representation. Each ring represents a separate kMC model. (b) 3D surface grid region representation. Each grid represents a separate kMC model.

5.3 Machine Learning-based Integrated Data-driven Model

Utilizing the data generated by the multiscale CFD model, an integrated data-driven model can be developed to describe and predict both the dynamic surface profile evolution as well as the

transient gas-phase profile development. Because of the non-linearity involved with the PEALD process, traditional machine learning methods fail to provide satisfying performance [28]. Thus, more advanced machine learning techniques need to be used to characterize the dynamic time series. The recurrent neural network model (RNN) and its variations are tailored for the prediction of time sequences and can be utilized to capture the complicated input-output relationship between operating conditions and deposition profiles. According to the universal approximation theorem, it can be proved that an RNN model with enough neurons can capture any given nonlinear dynamic system over finite time [122]. A simplistic view of the standard RNN structure is a stack of feedforward neural networks (FNN), where the output of the FNN is used repeatedly, along with additional real-time information, as the input to the network. The reused information is the hidden state, h, and the information fed in real-time is the input, x. Both values are used to predict the state of the system for the next timestep, of which the mathematical formulation is shown as follows:

$$h(t) = f(h(t-1), x(t-1), \theta) \tag{5.8}$$

where θ is the model parameter/weights, and the function f is the linear combination of hidden state, input, and weights, adjusted by the transfer functions, which provide the ability to capture non-linearity. The formulation of the RNN is very similar to a state-based model, with the state matrix replaced by a set of network neurons. The unrolling of the RNN shows that the cycled formulation forwards the hidden states, $h(t)$, and input information, $x(t)$, from the current timestep to the next timestep. This configuration makes it a perfect candidate to model the industrial time-series data, and thus it is selected in this work.

5.3.1 Long and Short Term Memory (LSTM) Method

Despite the intuitive compatibility, a typical challenge encountered in the training of RNN is the exploding or vanishing gradient, which happens because of the rolling vanilla structure. As we can see from the unrolling of the RNN, not only the forward information is cycled, the

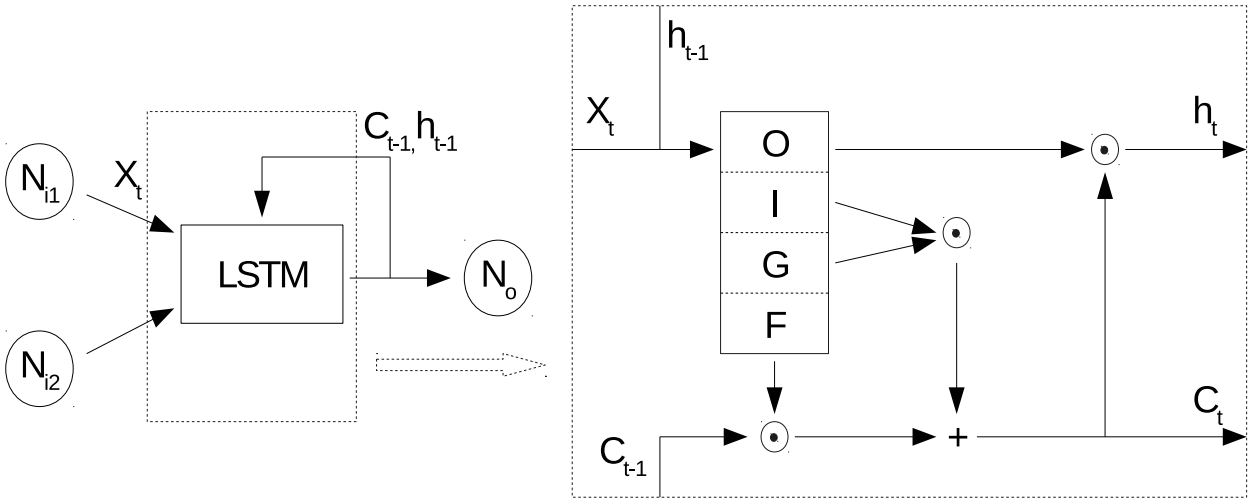


Figure 5.7: Long short-term memory (LSTM) recurrent neural network (RNN) structure. Left: general RNN formulation with LSTM cell. Right: detailed manifestation of the LSTM cell. N_i are input neurons, N_o is the output neuron, C_t and C_{t-1} are the cell state memory for training iteration t and $t-1$, and h_t and h_{t-1} are the hidden state for training iteration t and $t-1$. Merging is simple matrix stacking, $+$ denotes the element-wise addition, and circle denotes the element-wise product.

back-propagation is also repeated and passed through the activation functions multiple times. Therefore, this repetition leads to a significant accumulated input squeezing in the activation function, which further leads to the vanishing or explosion of the gradient and drastically decreases the validity of RNN. In addition, the traditional RNN treats all input information equally, while in reality, long term and short term information can provide quite different insights to the prediction of the new state.

To deal with these challenges, the long and short term memory (LSTM) is introduced. In the LSTM structure, the process characteristic parameter, θ, in the recurrent cell is replaced with four regulators: an input modulation gate (G), an input gate (I), an output gate (O), and a forget gate (F). These regulators selectively weight the long and short term memory (gradient) to avoid ill-conditioned propagation. Specifically, the LSTM weight matrix is comprised of four parts: the output gate controls the weights that reveal the cell state, the input gate and the input modulation gate control the weights that modify the cell state, and the forget gate controls the weights that erase the cell. In addition, the sigmoid activation is applied to the O, I, F gates, and the *tanh*

activation is applied to the G gate and the final output. It is demonstrated that the LSTM functions perfectly for processing and making predictions for time-series data. The exemplary configuration of an LSTM-based RNN is shown in Figure 5.7, and the model can be mathematically formulated as follows:

$$g = tanh(x_t U^g + h_{t-1} V^g) \tag{5.9}$$

$$i = \sigma(x_t U^i + h_{t-1} V^i) \tag{5.10}$$

$$f = \sigma(x_t U^f + h_{t-1} V^f) \tag{5.11}$$

$$s_t = s_{t-1} \circ f + g \circ i \tag{5.12}$$

$$o = \sigma(x_t U^o + h_{t-1} V^o) \tag{5.13}$$

$$h_t = tanh(s_t) \circ o \tag{5.14}$$

During the training step, all input and hidden state information will be provided as training data to the model, while the memory states are processed internally. Specifically, in order to capture the dynamic profile development in the reactor and on the substrate surface, three hidden states are selected from the simulation outputs: surface precursor partial pressure, the number of Hf physisorption sites and the number of O physisorption sites. Moreover, to correctly characterize the spatial difference of profiles in various wafer locations, each variable will be examined at the inner, middle, and outer regions of the wafer, which are shown in Figure 5.3. The precursor inlet flowrate, v, is used as the input to the model. Thus, in total ten states are included in the generated model.

To implement and train this LSTM-structured RNN, we adopt a high-level application programming interface (API), Keras, based on the Tensorflow backbend. The Adam optimizer is selected to adaptively minimize the loss function given a momentum estimation method by using

exponentially moving averages from the gradients evaluated on the current mini-batch:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 m_{t-1} + (1-\beta_2) g_t^2 \end{aligned} \tag{5.15}$$

where $t-1$ indicates the last mini-batch and t indicates the current mini-batch, v and m are the moving averages, g is the gradient, and the two betas are the hyperparameters of the model for learning rate, which are often selected to be 0.9 and 0.999, respectively. Then, these moving averages are used to scale the learning rates for all the parameters as follows:

$$w_t = w_{t-1} - \eta \frac{m_t}{\sqrt{v_t} + \varepsilon} \tag{5.16}$$

where w is the model weight, η is the step size of the learning and ε is the bias. In addition, to measure the deviation of the prediction from the provided CFD data, the least square error is chosen to be the loss function, which is the common choice in the regression application.

5.4 Integrated Data-driven Model Result and Validation

In the following sections, the training specifications and the performance of the developed RNN-based data-driven model are presented. The data-driven model is trained based on the data generated from the 2D multiscale CFD model for the reactor without the showerhead. The model accuracy is first validated by examining the alignment of model prediction with the original 2D multiscale CFD model result. Next, it is further compared with the 3D multiscale CFD model under the same operating parameters and reactor geometry.

5.4.1 2D Integrated Data-driven Model for PEALD Reactor

According to the formulation discussed in Section 5.3, an RNN model is trained based on the LSTM structure. The final configuration of the RNN is determined through a grid search, where

a two-layer structure consisting of 80 and 50 neurons, respectively, is chosen for the Hf-Cycle RNN model, and one layer of 100 neurons is chosen for the O-Cycle RNN model. The training dataset consists of the aforementioned transient deposition profiles for a range of inlet operating flowrates between 2.5×10^{-6} kg/s and 1.0×10^{-4} kg/s. In total, the training process takes around three to four hours on an Intel i7-8700 CPU with 64 GB of RAM, and the final normalized training performance, measured in terms of the mean standard error, reaches below 1.0×10^{-6} for both half-cycles. In addition, one full prediction for an entire time-series is demonstrated to take less than a minute to execute, which is a substantial reduction from the multiscale CFD model that takes about a day to run using 36 CPU cores on a powerful computational cluster. Also, the computational time involved is in line with the actual process operation time, thus making it feasible to be applied for on-line optimization and control purposes like model predictive control (MPC). Moreover, the prediction time can be even further shortened by taking advantage of cloud computing and other high-performance computational resources.

After the training is completed, the resulting models are validated with a set of test conditions. One of the most important indicators of the overall deposition progress is the availability of the physisorption sites. During the Hf-Cycle, Hf physisorption sites are the candidates for Hf precursor physisorption, while O physisorption sites are the product of the Hf-Cycle reactions and will be physisorbed in the following O-Cycle. Therefore, throughout the Hf-Cycle, the number of Hf physisorption sites will decrease while the number of O physisorption sites will increase, and vice versa in the O-cycle. The RNN accuracy for the prediction of the physisorption sites is demonstrated by comparing the RNN outputs with the multiscale CFD model results, as shown in Figure 5.8. A random flowrate, 2.3×10^{-6} kg/s, is chosen for testing and demonstration of the radially distributed thin-film growth profile of the Hf-Cycle. It can be seen from Figure 5.8 that, the model successfully predicts the surface deposition profile at all wafer regions. Specifically, for Hf precursor physisorption sites, all regions are accurately captured by the model. For the O precursor physisorption sites, the prediction of the inner and middle wafer regions closely corresponds to the CFD results. Although the prediction of the dynamic profile for the outer wafer region has

some deviations, the steady-state achieving time is accurately captured, as indicated by the top intersection of the green and red labels in Figure 5.8.

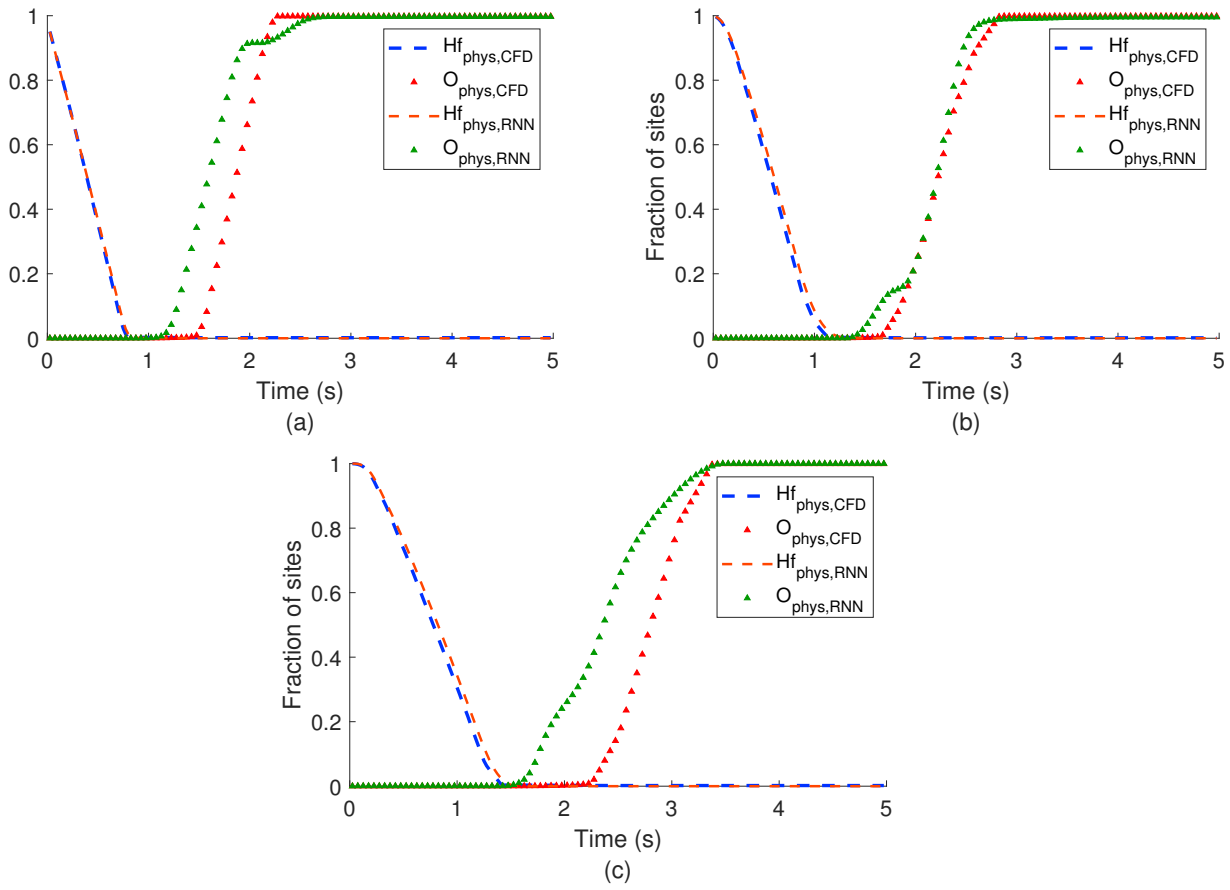


Figure 5.8: Comparison between RNN prediction and CFD simulation result without the showerhead for Hf-Cycle at inner, middle, and outer wafer regions, corresponding to (a), (b), and (c), respectively. The x-axis represents the time in a half-cycle, and the y-axis represents the fraction of sites. The blue and orange dashed lines represent the fraction of available sites for Hf precursor physisorption from CFD simulation and RNN prediction, respectively. The green and red triangles represent the fraction of available sites for O precursor physisorption from CFD simulation and RNN prediction, respectively.

Additionally, the prediction of the average partial pressure and physisorption sites are demonstrated in Figure 5.9, which corresponds to the profiles of Hf-Cycle at 6.8×10^{-5} kg/s and O-Cycle at 4.3×10^{-5} kg/s. As shown in Figure 5.9, the RNN predictions and CFD results closely match with each other, especially for O-Cycle. Due to the more complicated reaction routes in Hf-Cycle, the RNN model for Hf-Cycle has a slightly higher deviation. However, the model is

able to capture the overall trend and accurately determine the steady-state achieving times for both the Hf physisorption sites and the precursor partial pressure. In addition, it is worth noting that, compared to the linear parameter-varying model developed in [28], this RNN-based data-driven model fully integrates the microscopic domain and the macroscopic domain and is able to achieve higher accuracy.

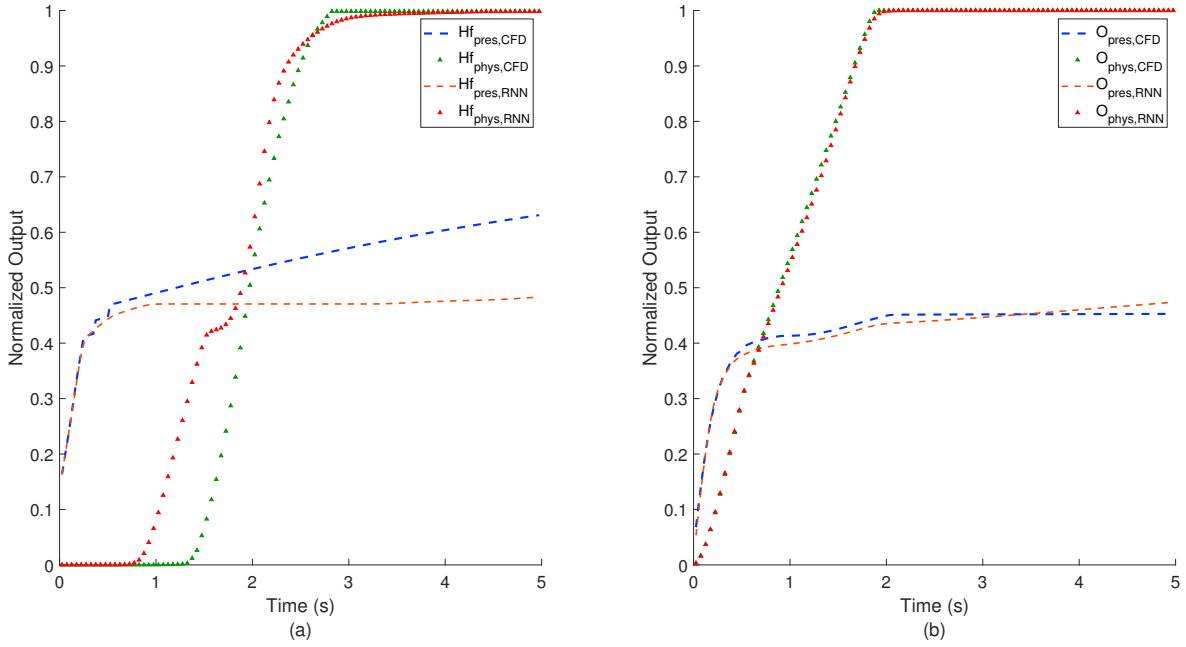


Figure 5.9: Comparison of RNN for two half-cycles in the reactor without the showerhead: (a) Hf-Cycle (b) O-Cycle. The x-axis represents the time within a half-cycle and the y-axis represents the normalized property value (partial pressure and physisorption sites availability). The blue and orange dashed lines represent the averaged precursor surface partial pressure from CFD simulation and RNN prediction, respectively. The green and red triangles represent the averaged surface physisorption sites availability from CFD simulation and RNN prediction, respectively.

5.4.2 Validation using 3D Multiscale Model

Although the RNN-based integrated data-driven model shows a good match with the 2D multiscale CFD model results, it is desirable to further validate the data-driven model with results computed from a full 3D multiscale CFD model. Thus, in this section, the data-driven model is compared with the computation result from the previously developed 3D multiscale CFD model by [113]. As

shown in Figure 5.3, the 3D multiscale model geometry can be viewed as the rotation of the 2D geometry around the central axis, and all the characteristic dimensions (wafer diameters, reactor size, outlet shape, etc.) are preserved.

For the O-Cycle, the average pressure and surface profiles are compared between the 3D multiscale CFD model and the RNN-based integrated data-driven model at a randomly chosen inlet flowrate of 1.25×10^{-5} kg/s. The starting point of both models is assumed to be the saturated surfaces after one complete Hf-Cycle. The deposition profiles of the inner, middle, and outer regions of the wafer over five seconds are presented in Figure 5.10. According to the CFD model result, film deposition steady-state is achieved around 4 s, which is also captured by the RNN model. Again, the RNN model prediction closely resembles the trend described by the 3D multiscale CFD model results. The normalized mean standard errors for the prediction of deposition profile at the inner, middle, and outer region and the prediction of the overall pressure are 2.54%, 1.19%, 2.85%, and 1.19%, respectively. Moreover, the error between the computed and predicted half-cycle time is 6.18%.

For the Hf-Cycle, the results are compared at another randomly chosen inlet flowrate at 4.0×10^{-5} kg/s. The starting surface profiles for both models are assumed to be the fully hydroxylated surface. Similar to the O-Cycle, the results over five seconds are demonstrated in Figure 5.11. RNN prediction accurately observes the trend computed by the multiscale CFD model. The normalized mean standard errors for the prediction of deposition profile in the inner, middle, outer region, and the prediction of the overall pressure are 1.85%, 4.60%, 1.50%, and 1.53%, respectively. Additionally, the error between the computed and predicted half-cycle time is 4.86%. Such performance is also quantitatively much better than the previously developed linear parameter-varying model, of which the errors range around 10%. Therefore, with validation by the 3D multiscale CFD model results, we can conclude that the developed RNN-based integrated data-driven model is capable of accurately characterizing the dynamic deposition profile and the transient gas-phase development in the PEALD reactors.

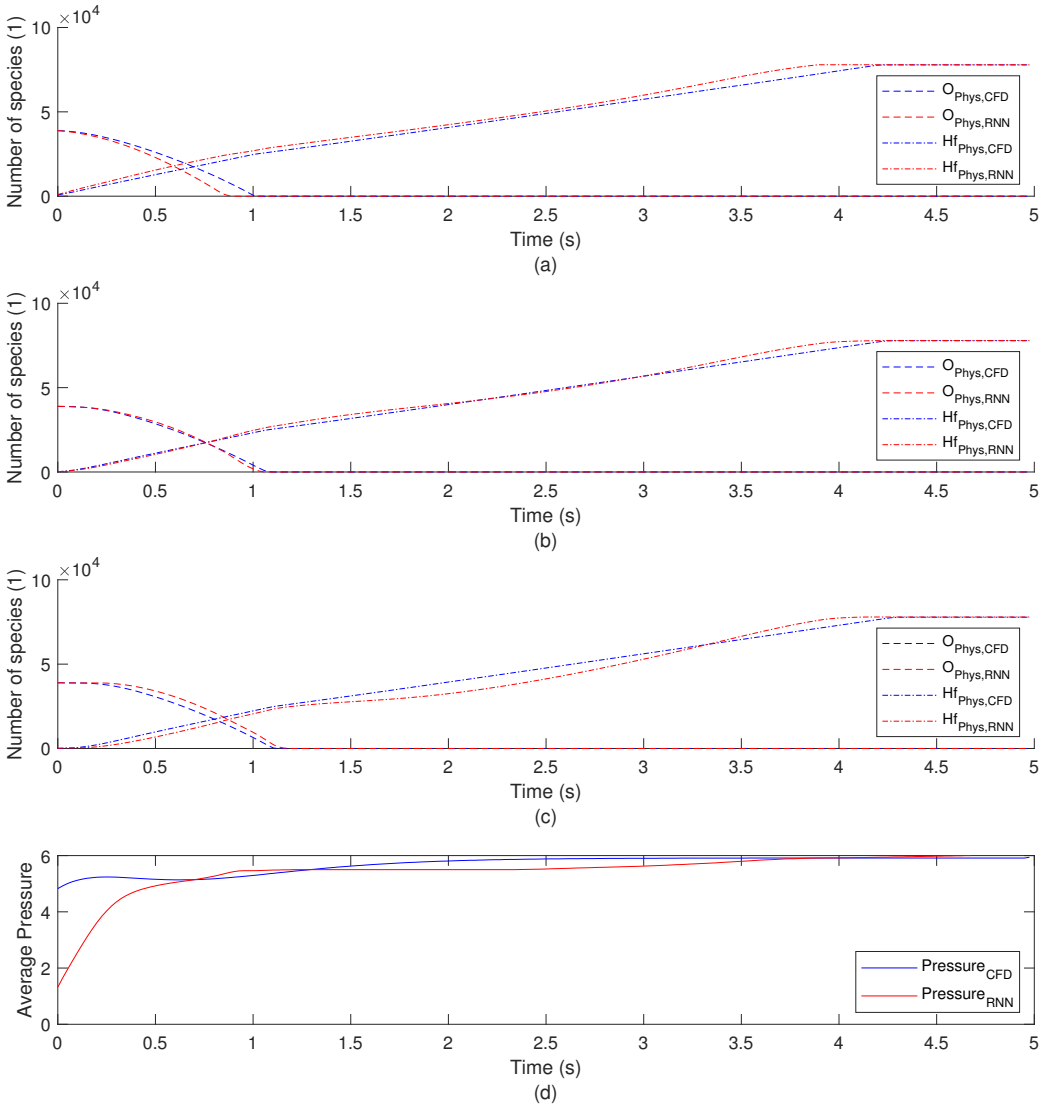


Figure 5.10: Comparison of 3D multiscale CFD model O-Cycle result with RNN-based data-driven model prediction under the inlet flowrate of 1.25×10^{-5} kg/s. Surface profiles over a duration of 5 s are demonstrated and surface steady-state is achieved towards the end of the cycle. x-axis represents the time, and y-axis represents the predicted profile. (a), (b), and (c) correspond to the profile development of physisorption sites at inner, middle, and outer regions of the wafer, respectively. (d) corresponds to the overall averaged pressure profile.

5.5 Operation Strategy Optimization

Now that the developed data-driven model has been validated, we can adopt the prediction capability of the data-driven model to identify the optimal operating strategy. In order to optimize

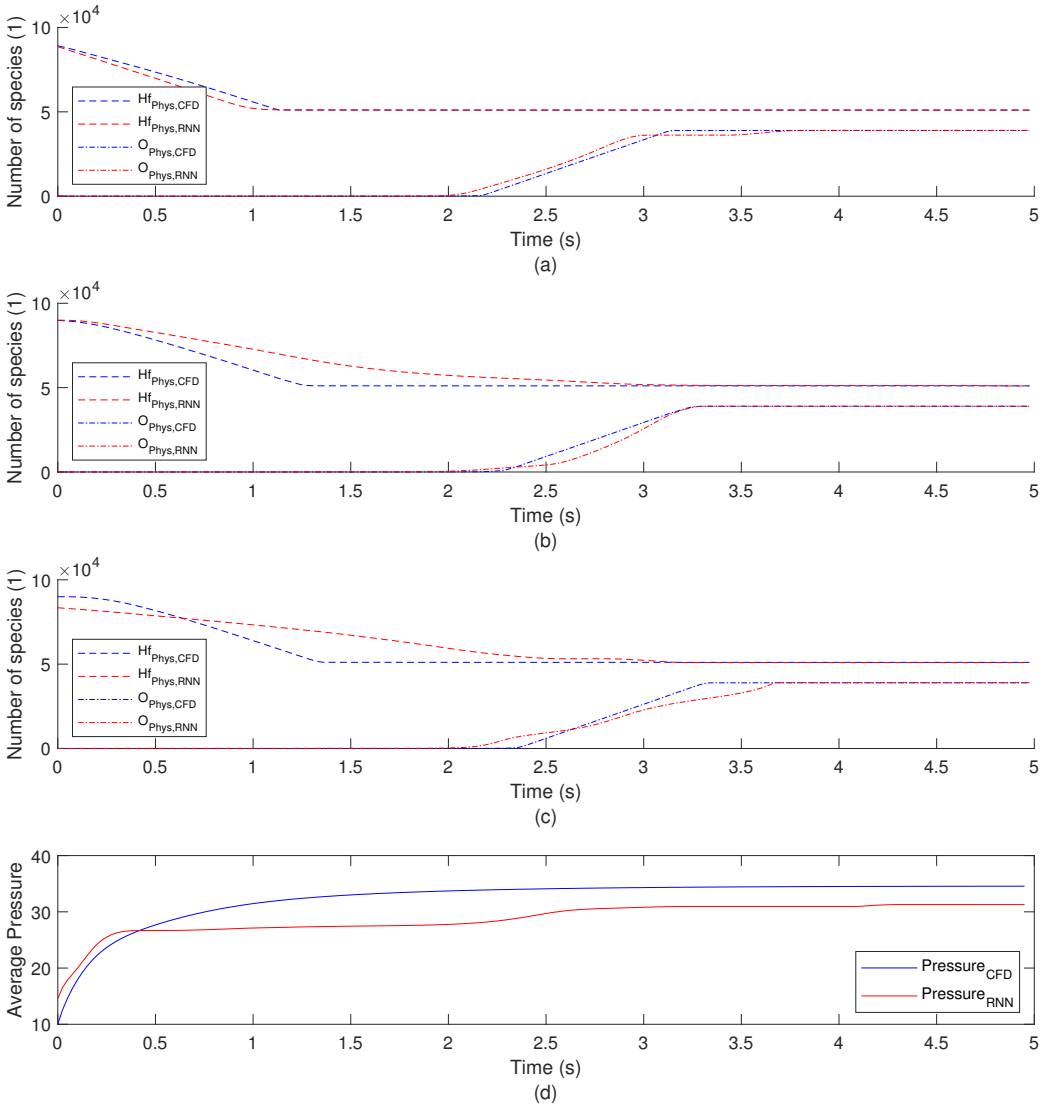


Figure 5.11: Comparison of 3D multiscale CFD model Hf-Cycle result with RNN-based data-driven model prediction under the inlet flowrate of 4.0×10^{-5} kg/s. Surface profiles over a duration of 5 s are demonstrated and surface steady-state is achieved towards the end of the cycle. x-axis represents the time, and y-axis represents the predicted profile. (a), (b), and (c) correspond to the profile development of physisorption sites at inner, middle, and outer region of the wafer, respectively. (d) corresponds to the overall averaged pressure profile.

the production throughput of PEALD processes, cycle-times need to be minimized, while the required film qualities need to be maintained at ideally the best achievable coverage. To satisfy

both requirements, the data-driven model is used to predict the system dynamics for a wide range of input flowrates from 2.5×10^{-6} kg/s to 9.75×10^{-4} kg/s, which is the common range of operating flowrates used in the experiments. By extracting the final film quality and the required cycle-time from the predicted film deposition profile, the range of feasible operating conditions can be summarized. Based on the feasible operating range, an elementary cost analysis is performed to determine the optimal operating condition.

For the O-Cycle, a database of deposition profiles for 200 flowrates uniformly distributed in the range of 2.5×10^{-6} kg/s to 9.75×10^{-4} kg/s is generated using the integrated data-driven model. The generated operating database is visualized in Figure 5.12 (a), (b), and (c) for the inner, middle, and outer wafer regions, respectively. Based on the predicted deposition profiles, the time to achieve steady-state and the final cycle progress are identified for each operating flowrate. The conventional half-cycle time for the PEALD process is taken to be five seconds. Therefore, only flowrates that allow the film to achieve full half-cycle coverage within five seconds are considered to be feasible, which are marked with blue dots at the steady-state achieving time in Figure 5.12 (a), (b), and (c). For O-Cycle, most flowrates in the given range are feasible except for the few lowest operating flowrates. After the feasible flowrates have been identified, an elementary cost analysis is performed to evaluate the optimal operating flowrate that minimizes the precursor usage. To evaluate precursor usage, the inlet flowrate are multiplied by their corresponding steady-state achieving times, which are shown in Figure 5.12 (d), (e), and (f) for the inner, middle, and outer wafer regions, respectively. Due to the nonlinear relationship between flowrate and deposition profile, the dependency of precursor usage on flowrate is also nonlinear, and the corresponding flowrates for minimal precursor usage at various wafer locations are marked with red diamonds in Figure 5.12 (d), (e), and (f), which is around 2.2×10^{-5} kg/s for all wafer regions. The reported trend of precursor usage over operating flowrates can also be applied in the more advanced cost analysis to customize the optimal operating condition, where additional concerns, such as overall expected throughput, specific wafer region quality, and exact precursor costs, are accounted for.

Similarly, for the Hf-Cycle, a database of deposition profiles for inlet operating flowrates within

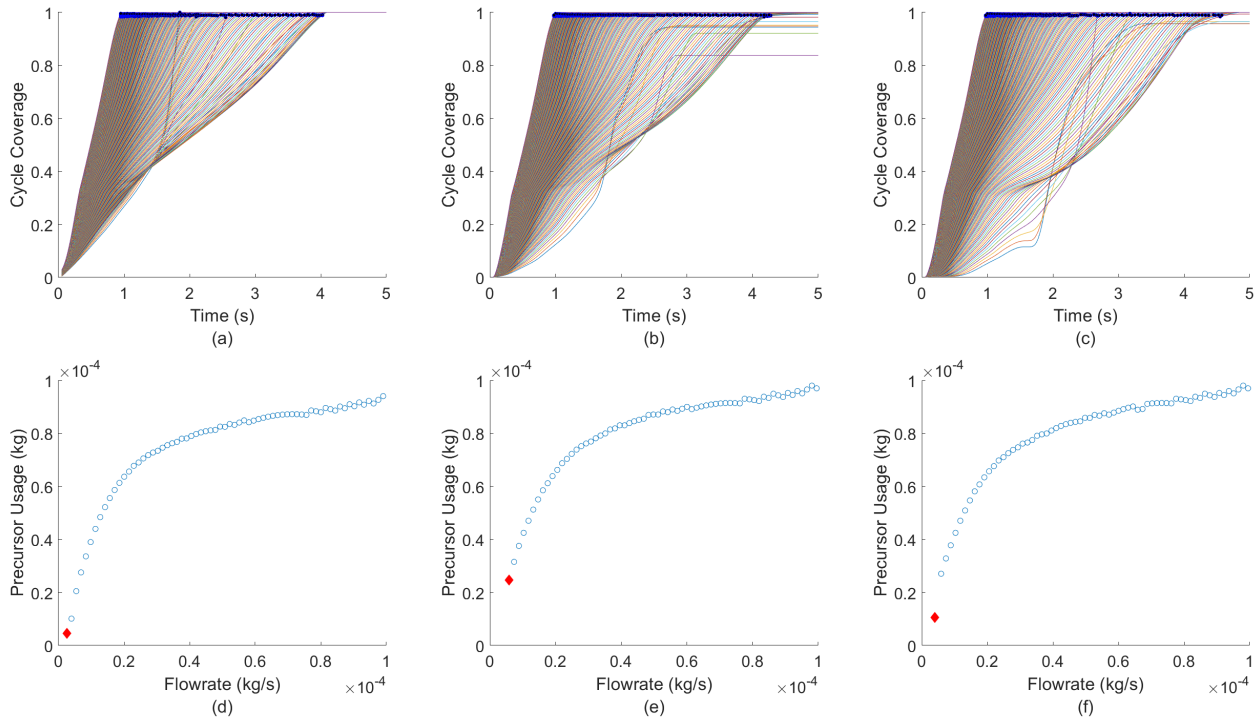


Figure 5.12: (a), (b), (c) Dynamic profiles for O-Cycle for flowrates between $\left[2.50 \times 10^{-6}, 9.75 \times 10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the O-Cycle deposition progress. The blue dots represent the identified steady-state achieving time for valid flowrates. (d), (e), (f) Precursor usage profiles for all valid flowrates. The x-axis is the flowrate and the y-axis is the amount of precursor required to finish the O-Cycle. The red diamond denotes the optimal operating flowrate to minimize the precursor usage.

the range of 2.5×10^{-6} kg/s to 9.75×10^{-4} kg/s is collected from the data-driven model prediction. The result is demonstrated in Figure 5.13 (a), (b), and (c) for the inner, middle, and outer wafer regions, respectively. The notations are similar as in the O-Cycle result and only feasible flowrates, marked with blue dots, are picked to identify the required half-cycle time and precursor usage. A flowrate of at least 2.5×10^{-5} kg/s is required for full half-cycle coverage. The precursor usage is then computed, and a non-linear relationship is observed between precursor usage and inlet flowrate. As shown in Figure 5.13 (d), (e), and (f), for all wafer regions, precursor usage increases with the flowrate until 5×10^{-5} kg/s, and then decreases as flowrate further increases. This nonlinear behavior can be attributed to the combination of gas-phase precursor transport and

reaction kinetics in the Hf-Cycle, which is explained in more detail in [112] and [113]. Therefore, the optimal operating condition to minimize precursor usage for Hf-Cycle occurs either at low or high flowrate, and it is subject to further cost and throughput analysis to determine the best operating flowrate in specific scenarios.

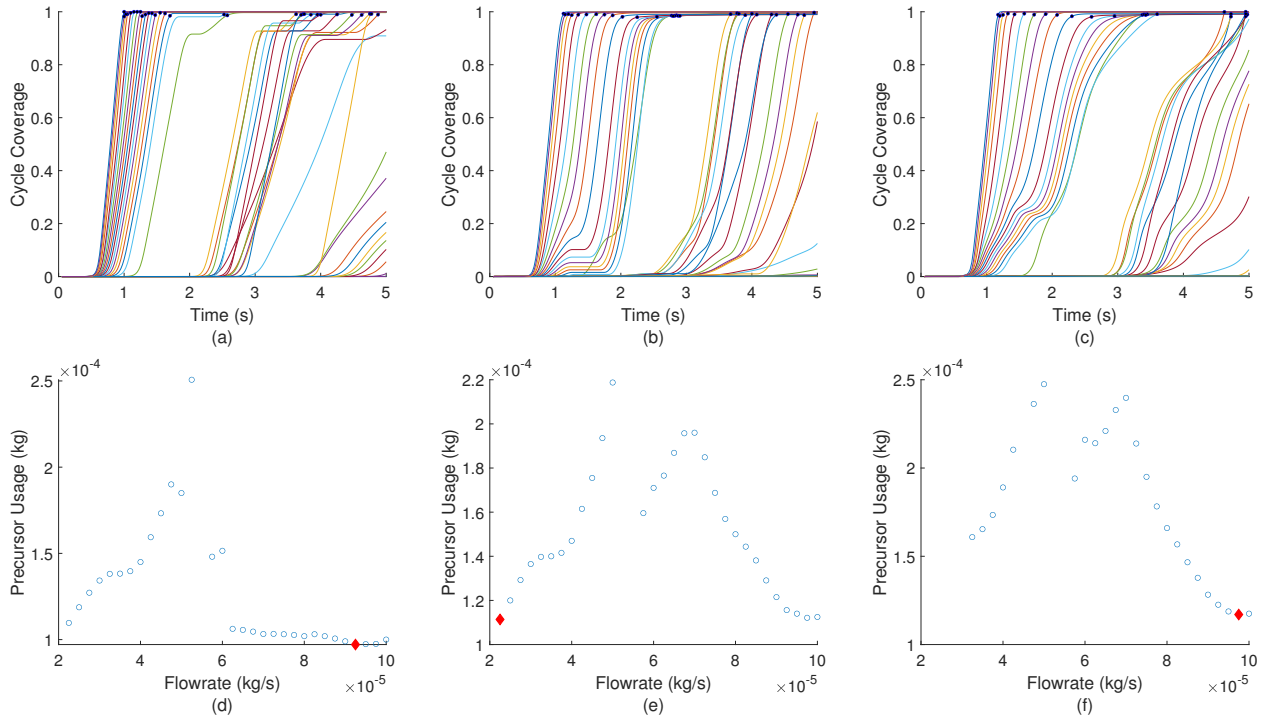


Figure 5.13: (a), (b), (c) Dynamic profiles for Hf-Cycle for flowrates between $[2.50 \times 10^{-6}, 9.75 \times 10^{-4}]$ kg/s for inner, middle, outer wafer regions, respectively. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the O-Cycle deposition progress. The blue dots represent the identified steady-state achieving time for valid flowrates. (d), (e), (f) Precursor usage profiles for all valid flowrates. The x-axis is the flowrate and the y-axis is the amount of precursor required to finish the O-Cycle. The red diamond denotes the optimal operating flowrate to minimize the precursor usage.

5.5.1 Integrated Data-driven model for PEALD Reactors with the Showerhead

Despite the intrinsic physical difference between 2D and 3D models for reactors with the showerhead, the 2D model result can still provide some insights for a general operating guideline.

Similar to the integrated data-driven model built for reactor without the showerhead, CFD simulations are performed to collect the operating profile for showerhead reactors, and the collected results are used to generate an RNN-based data-driven model, of which the training results are demonstrated in Figure 5.14 and Figure 5.15. The O physisorption site characterizations and the final steady-state achieving time for both half-cycles, as well as the partial pressure characterization for O-Cycle, are very accurate. Due to the complicated geometry of the showerhead reactor and the more complex reaction sets in the Hf-Cycle, the prediction of the Hf physisorption sites for Hf-Cycle demonstrates an offset from the CFD result. However, the RNN prediction captures the exact trend reported by CFD simulation and can be adjusted by further post-processing.

Using the built data-driven model, a range of deposition profiles under various operating flowrates are shown in Figure 5.16 and Figure 5.17. Similar to the reactor without the showerhead, the precursor usages under various flowrates are computed using the steady-state achieving time determined from the operating database. The trend of O-Cycle precursor usage versus operating flowrate for showerhead reactors is very similar to reactors without the showerhead, despite a nonlinear region for low operating flowrates. For the Hf-Cycle precursor usage, due to the resistance caused by the showerhead panel, the minimal flowrate required to achieve full coverage is higher than the reactor without the showerhead. As observed in Figure 5.17, the minimal precursor usage occurs around 9.0×10^{-5} kg/s for all wafer regions, and increases for both lower and higher flowrates. This nonlinearity could be potentially due to the complex competition between gas-phase precursor transport and the surface reaction kinetics. Moreover, despite the alignment of the result between the RNN model and the 2D multiscale CFD model, due to the limitation of the 2D showerhead reactor geometry, the showerhead reactor result does not perfectly reflect the full 3D transport and deposition behaviors. Therefore, the integrated data-driven model should only be used as a general guideline for the operation of PEALD showerhead reactors.

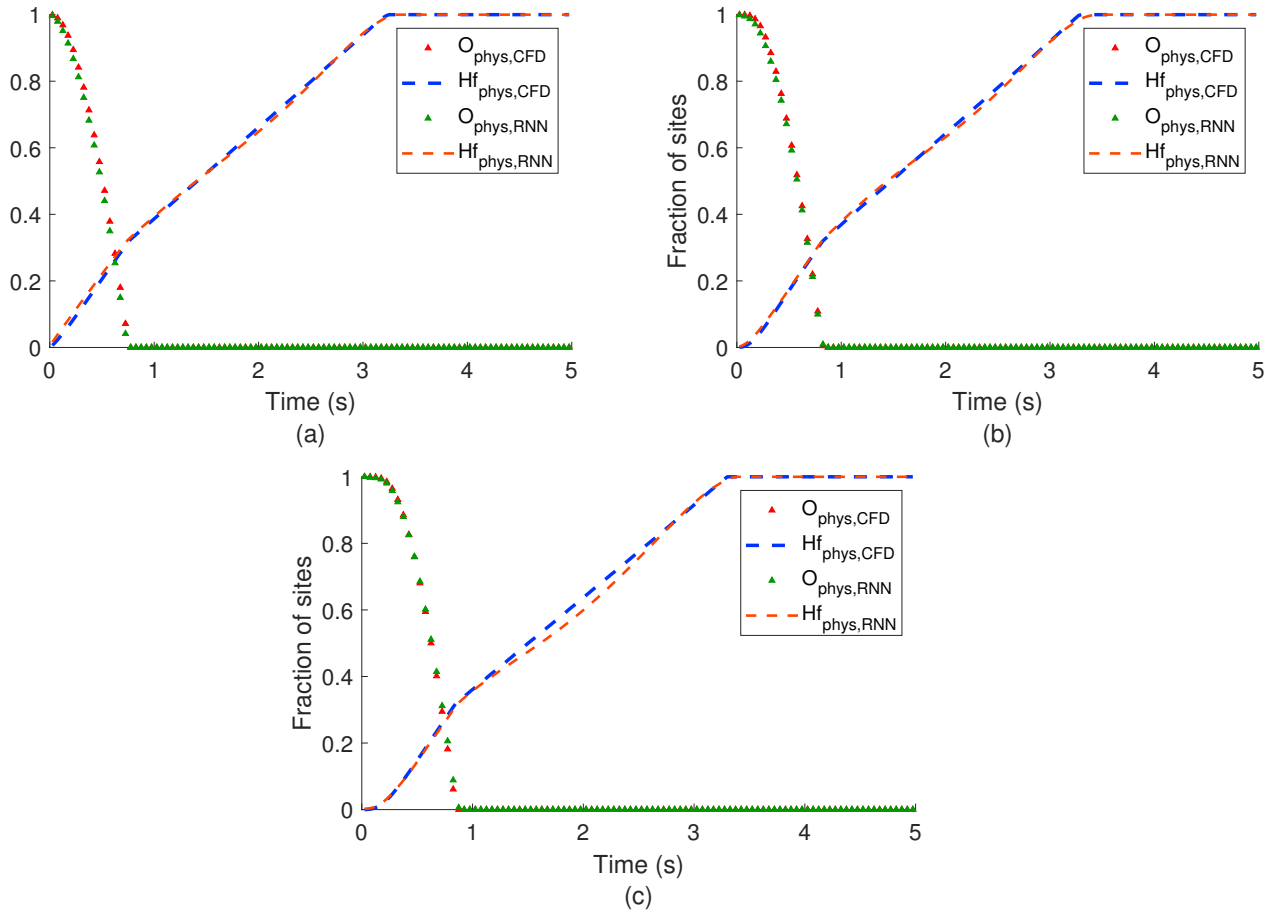


Figure 5.14: Comparison between RNN prediction and CFD simulation result with the showerhead for O-Cycle at inner, middle, and outer wafer regions, corresponding to (a), (b), and (c), respectively. The x-axis represents the time in a half-cycle, and the y-axis represents the fraction of sites. The blue and orange dashed lines represent the fraction of available sites for Hf precursor physisorption from CFD simulation and RNN prediction, respectively. The green and red triangles represent the fraction of available sites for O precursor physisorption from CFD simulation and RNN prediction, respectively.

5.6 Conclusion

In this chapter, an integrated data-driven model based on RNN is constructed using the previously developed multiscale CFD model for a PEALD process. Based on the datasets generated from CFD, the developed RNN is able to accurately predict the film deposition profile using the inlet feed flowrate. Although the built data-driven model is developed based on 2D axisymmetric CFD computation results, it is validated with the results from a full 3D CFD simulation. Specifically, for

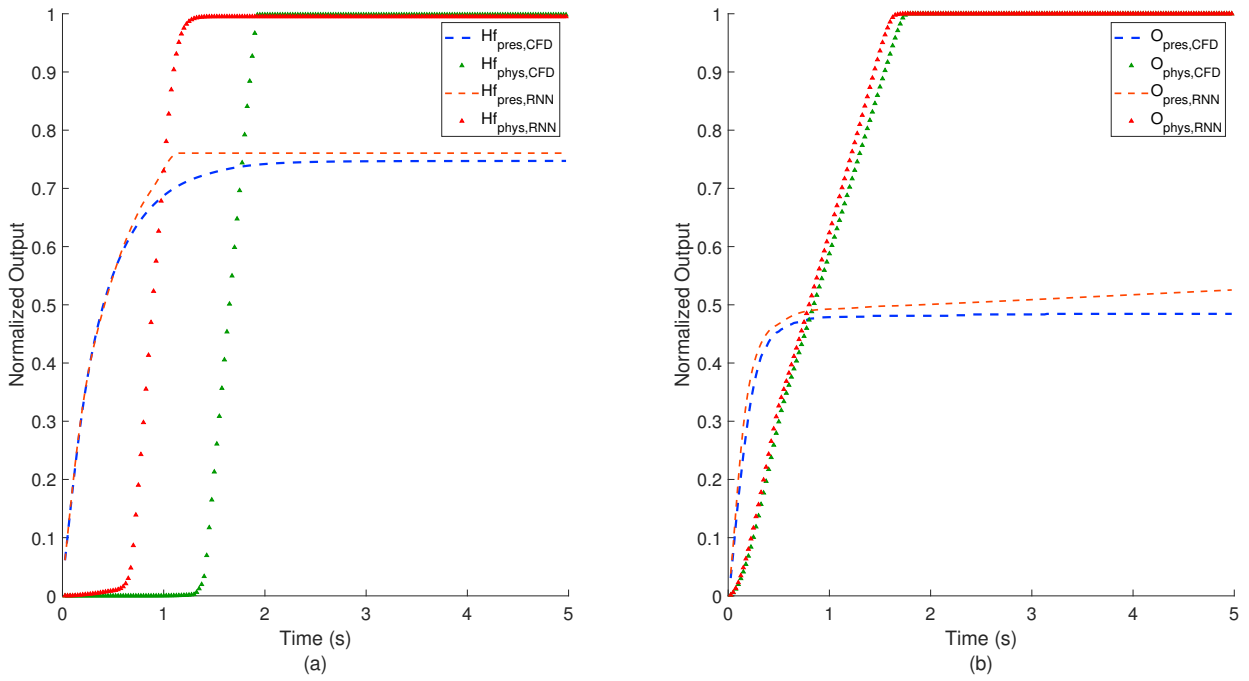


Figure 5.15: Comparison of RNN for two half-cycles in the reactor with the showerhead: (a) Hf-Cycle (b) O-Cycle. The x-axis represents the time within a half-cycle and the y-axis represents the normalized property value (partial pressure and physisorption sites availability). The blue and orange dashed lines represent the averaged precursor surface partial pressure from CFD simulation and RNN prediction, respectively. The green and red triangles represent the averaged surface physisorption sites availability from CFD simulation and RNN prediction, respectively.

each half-cycle, an RNN model has been tailored to capture the deposition behavior with less than or around 5% deviation from the CFD simulation results. Compared to the multiscale CFD model, which takes about a day to compute the profiles for one flowrate, the integrated data-driven model only takes a few seconds to predict the film growth, and thus can be incorporated in real-time process control and process monitoring. Moreover, an operation database is generated using the predictions from the integrated data-driven model. Using the operation database and based on industrial standard, a feasible operating region is determined in terms of the inlet flowrates. Furthermore, an optimal operating strategy is identified in the feasible operating domain for each half-cycle based on the precursor usage analysis. Additionally, for the showerhead PEALD reactor, a similar methodology has been adopted to generate its corresponding integrated data-driven model. Despite the fundamental geometric difference of showerhead between 2D and 3D, the

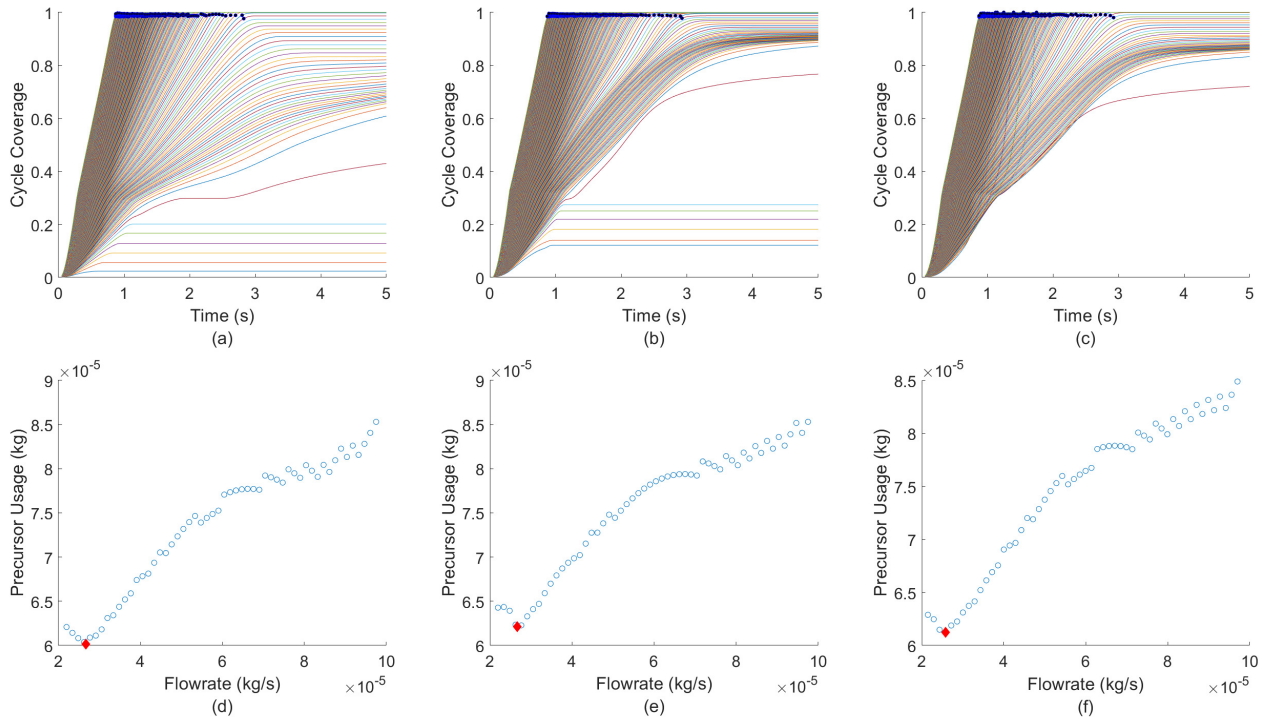


Figure 5.16: (a), (b), (c) Dynamic profiles for O-Cycle for flowrates between $\left[2.50\times10^{-6}, 9.75\times10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively, in a showerhead reactor. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the O-Cycle deposition progress. The blue dots represent the identified steady-state achieving time for valid flowrates. (d), (e), (f) Precursor usage profiles for all valid flowrates. The x-axis is the flowrate and the y-axis is the amount of precursor required to finish the O-Cycle. The red diamond denotes the optimal operating flowrate to minimize the precursor usage.

data-driven model can still provide a general guideline for the operation of showerhead reactors. In closing, it is worth mentioning that the LSTM based data-driven modeling strategy presented in this work is suitable for many other similar semiconductor fabrication processes. The challenges for the desired in-situ monitoring and the complexity of an exact on-line simulation representation can be avoided by extracting valuable information from combined experimental and simulated data at different conditions. In addition, with the developed data-driven model, an optimal operating strategy can be established based on the specific industrial need (e.g., optimal manufacturing throughput), and on-line control schemes can be explored, which can further enhance the process operational performance.

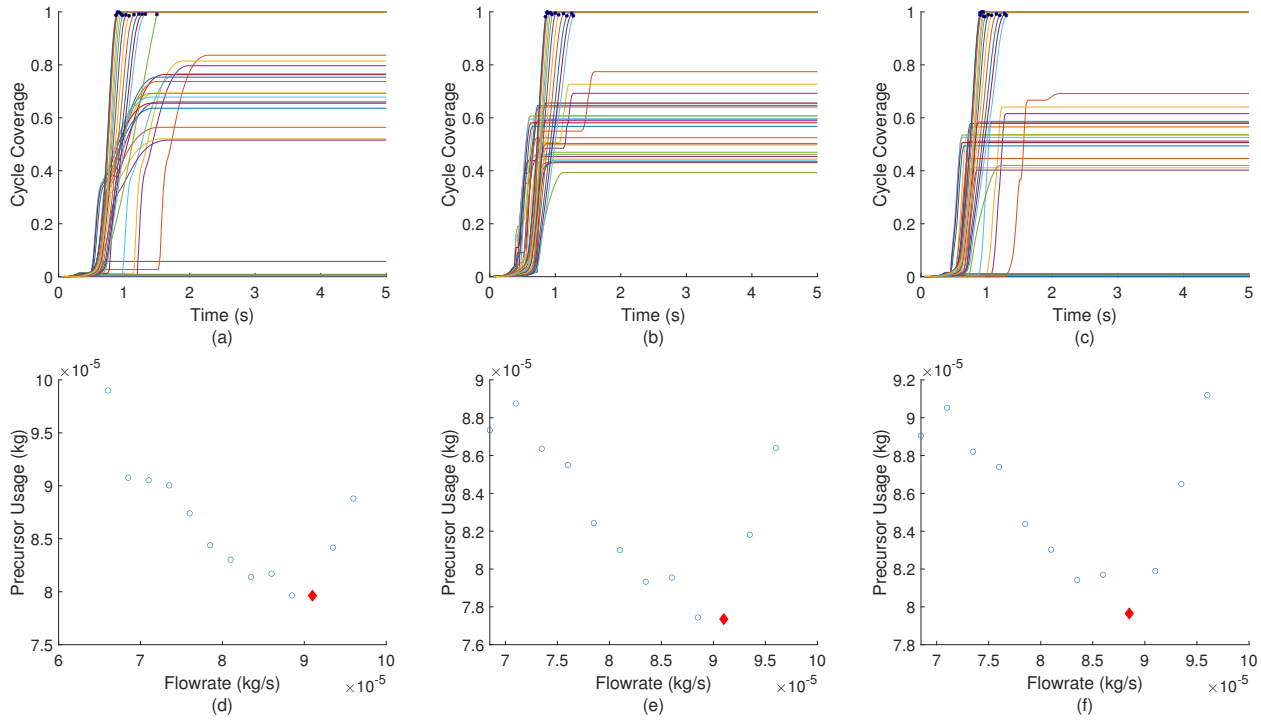


Figure 5.17: (a), (b), (c) Dynamic profiles for Hf-Cycle for flowrates between $\left[2.50\times10^{-6}, 9.75\times10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively, in a showerhead reactor. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the Hf-Cycle deposition progress. The blue dots represent the identified steady-state achieving time for valid flowrates. (d), (e), (f) Precursor usage profiles for all valid flowrates. The x-axis is the flowrate and the y-axis is the amount of precursor required to finish the Hf-Cycle. The red diamond denotes the optimal operating flowrate to minimize the precursor usage.

Chapter 6

Conclusion

This dissertation provides a general methodology for the development of microscopic surface models and data-driven models for atomic layer deposition (ALD) and plasma-enhanced atomic layer deposition (PEALD) applications. Microscopic models based on a customized hybrid kinetic Monte-Carlo algorithm were developed for ALD of SiO_2 and PEALD of HfO_2. Due to the lack of certain thermodynamics and kinetics parameters in the existing literature, density functional theory (DFT) calculations were performed to obtain these parameters for the successful implementation of the microscopic surface model. The microscopic model results are validated with past experimental reports and industrial standards. After the microscopic model has been successfully formulated, a database of the microscopic surface simulation results is constructed, which is used to train a feedforward artificial neural network (ANN). The neural network is able to expand the database by performing smart and efficient interpolation, which is then utilized for a preliminary exploration of the operating policy. The microscopic model development and ANN operating condition exploration for thermal ALD and PEALD are described in Chapter 2 and Chapter 4, respectively.

Chapters 3 and 5 describe the construction of the multiscale data-driven model and the operating policy exploration using the built data-driven model for ALD and PEALD, respectively. To provide a holistic view of the deposition profile inside the reactor, previously developed multiscale computational fluid dynamics (CFD) models are used. The multiscale CFD model

describes the entire ALD/PEALD reactor by coupling the kMC-based microscopic surface model) and the CFD-based macroscopic gas-phase model through an automated workflow. Despite the model accuracy, the multiscale CFD model is computationally expensive and is not suitable for real-time monitoring. Therefore, machine-learning based data-driven models are constructed to reduce the computation demand. Specifically, the data-driven model for SiO_2 ALD integrates a linear parameter varying model, which is used to characterize the macroscopic gas-phase domain, and an ANN-based data-driven surface model derived from the simulation results of the microscopic model, which is used to characterize the surface deposition. On the other hand, for the data-driven model for HfO_2, a more advanced approach is utilized. Instead of decoupling the microscopic and macroscopic domains for the derivation of the data-driven model, an integrated recurrent neural network (RNN) based model is used to characterize both the macroscopic and the microscopic domains together. Furthermore, by applying preliminary cost rubrics, an elementary cost analysis is performed using this integrated data-driven model for the exploration of optimal operating conditions. It is also noteworthy that, beyond the scope of this dissertation, the developed data-driven model can also be applied to construct a control scheme that enables the robust and stable operation of PEALD.

Bibliography

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

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

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

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

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

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

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

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

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

[10] N. P. Dasgupta, L. Li, X. Sun, Atomic layer deposition for energy and environmental applications, Advanced Materials Interfaces 3.

[11] M. C. Schwille, T. Schössler, J. Barth, M. Knaut, F. Schön, A. Höchst, M. Oettel, J. Bartha, Experimental and simulation approach for process optimization of atomic layer deposited

thin films in high aspect ratio 3D structures, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films 35 (2017) 01B118.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[26] A. Dkhissi, A. Estève, C. Mastail, S. Olivier, G. Mazaleyrat, L. Jeloaica, M. Djafari-Rouhani, Multiscale modeling of the atomic layer deposition of HfO_2 thin film grown on silicon: How to deal with a kinetic Monte Carlo procedure, Journal of Chemical Theory and Computation 4 (2008) 1915–1927.

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

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

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

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

[31] G. Kimaev, L. A. Ricardez-Sandoval, Nonlinear model predictive control of a multiscale thin film deposition process using Artificial Neural Networks, Chemical Engineering Science 207 (2019) 1230–1245.

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

[33] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

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

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

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

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

[38] S. Kamiyama, T. Mira, Y. Nara, Comparison between SiO_2 films deposited by atomic layer deposition with $SiH_2[N(CH_3)_2]_2$ and $SiH[N(CH_3)_2]_3$ precursors, Thin Solid Films 515 (2006) 1517–1521.

[39] Y. Kinoshita, F. Hirose, H. Miya, K. Hirahara, Y. Kimura, M. Niwano, Infrared study of tris(dimethylamino)silane adsorption and ozone irradiation on Si(100) surfaces for ALD of SiO_2, Electrochemical and Solid-State Letters 10 (2007) G80–G83.

[40] S. B. Baek, D. H. Kim, Y. C. Kim, Adsorption and surface reaction of bis-diethylaminosilane as a Si precursor on an OH-terminated Si(001) surface, Applied Surface Science 258 (2012) 6341–6344.

[41] G. Dingemans, C. A. A. V. Helvoirt, D. Pierreux, W. Keuning, W. M. M. Kessels, Plasma-assisted ALD for the conformal deposition of SiO_2: Process, material and electronic properties, Journal of the Electrochemical Society 159 (2012) H277–H285.

[42] L. Huang, B. Han, B. Han, A. Derecskei-Kovacs, M. Xiao, X. Lei, M. L. O′Neill, R. M. Pearlstein, H. Chandra, H. Cheng, First-principles study of a full cycle of atomic layer deposition of SiO_2 thin films with di(sec-butylamino)silane and ozone, The Journal of Physical Chemistry C 117 (2013) 19454–19463.

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

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

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

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

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

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

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

[50] K. Momma, F. Izumi, Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data, Journal of Applied Crystallography 44 (2011) 1272–1276.

[51] J. Li, J. We, C. Zhou, B. Han, E. J. Karwacki, M. Xiao, X. Lei, H. Cheng, On the dissociative chemisorption of tris(dimethyl-amino)silane on hydroxylated SiO_2(001) surface, Journal of Physical Chemistry C 113 (2009) 9731–9736.

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

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

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

[55] D. Svozil, V. Kvasnicka, J. Pospichal, Introduction to multi-layer feed-forward Neural Networks, Chemometrics and Intelligent Laboratory Systems 39 (1997) 43–62.

[56] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32^{nd} International Conference on Machine Learning, Vol. 37, 2015.

[57] S. Ruder, An overview of gradient descent optimization algorithms, CoRR abs/1609.04747.

[58] F. Burden, W. Dave, Bayesian regularization of Neural Networks, in: Artificial Neural Networks: Methods and Applications, Springer, 2008, pp. 23–42.

[59] J. J. Moré, The Levenberg-Marquardt algorithm: implementation and theory, in: Numerical Analysis, Springer, 1978, pp. 105–116.

[60] D. J. C. MacKay, Bayesian interpolation, Neural Computation 4 (1992) 415–447.

[61] B. El-Kareh, Fundamentals of semiconductor processing technology, Springer Science & Business Media, 2012.

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

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

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

[65] T. Weckman, M. Shirazi, S. D. Elliott, K. Laasonen, Kinetic Monte Carlo study of the atomic layer deposition of zinc oxide, The Journal of Physical Chemistry C 122 (2018) 27044–27058.

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

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

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

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

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

[71] M. Hintermüller, M. Ulbrich, A mesh-independence result for semismooth newton methods, Mathematical Programming 101 (2004) 151–184.

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

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

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

[75] G. Fang, L. Xu, J. Ma, A. Li, Theoretical understanding of the reaction mechanism of sio2 atomic layer deposition, Chemistry of Materials 28 (5) (2016) 1247–1255.

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

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

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

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

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

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

[82] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.

[83] A. Wächter, L. T. Biegler, On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming, Mathematical Programming 106 (1) (2006) 25–57.

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

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

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

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

[88] Q. Zeng, A. R. Oganov, A. O. Lyakhov, C. Xie, X. Zhang, J. Zhang, Q. Zhu, B. Wei, I. Grigorenko, L. Zhang, et al., Evolutionary search for new high-k dielectric materials: methodology and applications to hafnia-based oxides, Acta Crystallographica Section C: Structural Chemistry 70 (2014) 76–84.

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

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

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

[92] L. Melo, G. Burton, P. Kubik, P. Wild, Long period gratings coated with hafnium oxide by plasma-enhanced atomic layer deposition for refractive index measurements, Optics Express 24 (2016) 7654–7669.

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

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

[95] W. S. Epling, C. H. Peden, M. A. Henderson, U. Diebold, Evidence for oxygen adatoms on TiO_2 (110) resulting from O_2 dissociation at vacancy sites, Surface Science 412 (1998) 333–343.

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

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

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

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

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

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

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

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

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

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

[106] N. K. Oh, J. T. Kim, J. K. Ahn, G. Kang, S. Y. Kim, J. Y. Yun, The effects of thermal decomposition of tetrakis-ethylmethylaminohafnium (TEMAHf) precursors on HfO_2 film growth using atomic layer deposition, Applied Science and Convergence Technology 25 (2016) 56–60.

[107] T. Bazzi, R. Ismail, M. Zohdy, Comparative performance of several recent supervised learning algorithms, International Journal of Computer and Information Technology 7 (2018) 49–54.

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

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

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

[111] M. Crose, J. S. I. Kwon, M. Nayhouse, D. Ni, P. D. Christofides, Multiscale modeling and operation of PECVD of thin film solar cells, Chemical Engineering Science 136 (2015) 50–61.

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

[113] Y. Zhang, Y. Ding, P. D. Christofides, Multiscale computational fluid dynamics modeling and reactor design of plasma-enhanced atomic layer deposition, Computers & Chemical Engineering 147 (2020) 107066.

[114] Y. Wang, Y. Zhang, Z. Wu, H. Li, P. D. Christofides, Operational trend prediction and classification for chemical processes: A novel convolutional neural network method based on symbolic hierarchical clustering, Chemical Engineering Science 225 (2020) 115796.

[115] W. Yin, K. Kann, M. Yu, H. Schütze, Comparative study of CNN and RNN for natural language processing, arXiv preprint arXiv:1702.01923.

[116] H. Kim, M. Park, C. W. Kim, D. Shin, Source localization for hazardous material release in an outdoor chemical plant via a combination of LSTM-RNN and CFD simulation, Computers & Chemical Engineering 125 (2019) 476–489.

[117] Z. Wu, A. Tran, Y. M. Ren, C. S. Barnes, S. Chen, P. D. Christofides, Model predictive control of phthalic anhydride synthesis in a fixed-bed catalytic reactor via machine learning modeling, Chemical Engineering Research and Design 145 (2019) 173–183.

[118] S. Chen, Z. Wu, P. D. Christofides, Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control, Computers & Chemical Engineering 136 (2020) 106806.

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

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

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

[122] E. D. Sontag, Neural nets as systems models and controllers, in: Proceedings of the Seventh Yale Workshop on Adapative and Learning Systems, Yale University, 1992, pp. 73–79.