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

Author
Chung, Ho Yeon

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

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

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science
in Chemical Engineering

by

Ho Yeon Chung

2020

ABSTRACT OF THE THESIS

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

by

Ho Yeon Chung

Master of Science in Chemical Engineering

University of California, Los Angeles, 2020

Professor Panagiotis D. Christofides, Chair

Plasma-enhanced atomic layer deposition (PEALD) has demonstrated its superiority at coating ultra-conformal high dielectric thin-films, which are essential to the fin field-effect transistors (FinFETs) as well as the advanced 3D V-NAND (vertical Not-AND) flash memory cells. Despite the growing research interest, the exploration of the optimal operation policies for PEALD remains a complicated and expensive task. Our previous work has constructed a comprehensive 3D multiscale computational fluid dynamics (CFD) model for the PEALD process and demonstrated its potential to enhance the understanding of the process. Nevertheless, the limitation of computational resources and the relatively long computation time restrict the efficient exploration of the operating space and the optimal operating strategy. Thus, in this work, we apply a 2D axisymmetric reduction of the previous 3D model of PEALD reactors with and without the showerhead design. Furthermore, a data-driven model is derived 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 thesis of Ho Yeon Chung is approved.

Dante A. Simonetti

Carlos Morales-Guio

Panagiotis D. Christofides, Committee Chair

University of California, Los Angeles

2020

Contents

List of Figures

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Panagiotis D. Christofides for his guidance and support throughout the course of the thesis.

I would like to thank Professor Dante Simonetti and Professor Carlos Morales-Guio for participating in my Master's thesis committee.

This work was submitted with the same title for publication in *Computers and Chemical Engineering*, and is co-authored by Yangyao Ding, Yichi Zhang, and Professor Panagiotis D. Christofides. I would like to acknowledge their contributions to this thesis and express great thanks for their help.

Chapter 1

Introduction

The rapid growth of the semiconductor industry is backed by the increasing demand for highly compact microelectronic devices. In order to meet the design challenges of constant device miniaturization, high dielectric constant (high-κ) materials are introduced to resolve the quantum tunneling effects associated with extremely thin gate oxides [1], like TiO_2 [2], HfO_2 [3], and ZrO_2 [4]. In particular, one of the most adopted high-κ thin-film materials is HfO_2 due to its high thermal stability, extraordinary gate capacitance, good charge mobility, and very large dielectric constant, which is four to six times higher than that of SiO_2. Nevertheless, the reactions associated with the deposition process often have high energy barriers, which reduces the production throughput, increases the energy footprint, and introduces design challenges for the traditional thermal atomic layer deposition (ALD) operation policies, especially for temperature-sensitive materials [5]. Therefore, plasma-enhanced atomic layer deposition (PEALD) has been invented to overcome the aforementioned problems by taking advantage of the active plasma species [6].

Derived from the traditional thermal ALD, PEALD follows the scheme of sequential precursor pulsing and inert gas purging. Conventionally, each precursor pulsing stage is called a half-cycle, during which the precursor deposits the corresponding element in a self-limiting manner under appropriate operating conditions. Purge steps follow the pulsing half-cycles, where an inert species is introduced into the reactor to clean the unreacted precursor species and

by-products [7]. With the aid of high energy plasma species, ultra-uniform high-κ dielectric thin-films can be produced under relatively low operating temperatures in a layer-by-layer manner with high controllability [8]. Therefore, many novel precursors for the deposition of HfO_2 using PEALD, together with their reaction mechanisms and the associated PEALD reactor design, have been investigated. For the hafnium (Hf) half-cycle precursors, amino-based metal-organic compounds have reaction mechanisms facilitated by H-N bonds. Materials including tetrakis(dimethylamino)hafnium (TDMAHf) [9], tetrakis(ethylmethylamino)hafnium (TEMAHf) [2], tetrakis(diethylamino)hafnium (TDEAHf) [10] have been extensively studied. For the oxygen (O) half-cycle, which involves plasma species, typical candidates include O_3 plasma [11], H_2O plasma [12], and O_2 [11] plasma. In addition to precursor selection, the reactor design also critically affects the PEALD throughput and effectiveness. Remote plasma reactors have been designed as a balanced solution for the PEALD process, in which plasma species are generated from a remote chamber at an appropriate distance from the substrate surface [13]. As high energy species travel across the reactor, their energy is reduced to the desired level where sputtering can be avoided, while still being high enough to overcome the reaction energy barrier [14]. Thus, in this work, the discussion will focus on the operation of PEALD of HfO_2 thin-film using TDMAHf and oxygen plasma in a remote plasma reactor.

Due to the great potential of PEALD, various experimental efforts have been carried out to explore, understand, and optimize the operating policies of the PEALD process to maximize its economical benefits. However, the PEALD process is often associated with high operating costs, complicated process components, and the absence of efficient monitoring techniques. For example, a PEALD reactor requires ultra-precise gas flow controllers, complex radio-frequency (RF) power sources, and sophisticated pumping systems [15, 16]. Also, the deposition speed and the properties of the deposited thin-film are highly dependent on both gas-phase transport phenomena and microscopic surface dynamics. Such relationships are reactor design-specific and highly complicated to be determined experimentally [15, 17]. Moreover, despite the existence of in-situ analysis techniques such as quartz crystal microbalance (QCM) and spectroscopic

ellipsometry, the amount of data that can be obtained in real-time is limited [18]. Also, ex-situ microstructure analysis methods such as scanning tunneling microscopy (STM) and scanning electron microscope (SEM) are often destructive to the deposited film [19]. These difficulties restrict the effective exploration of PEALD operating conditions.

As an alternative to the experimental methods, 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, [20] created a simulation model for O_2 plasma in a simplified remote inductively coupled plasma (ICP) reactor configuration, and [21] 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 [22, 23]. For example, [24] constructed a CFD model to capture the chemical vapor deposition of amorphous silicon thin-films, and [25] designed a CFD characterization of the SiO_2 ALD and optimized the showerhead reactor geometry. For the microscopic surface reaction domain, [26] 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, [27] extended the kMC model to be used for PEALD. In addition, [28] 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 [29, 30]. The previous work by [31] 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 [32]. Among those deep neural network models, RNN is intuitively more suitable and has demonstrated its outstanding performance in text summarization and natural language processing [33]. Recently, it has also been applied in modeling many industrial systems [34]. For example, [35] investigated Phthalic Anhydride Synthesis based on CFD simulation and constructed a computationally efficient RNN characterization of the process. [36] 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 work, we first construct a database using the previously developed multiscale CFD model [28]. 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.

Chapter 2

Multiscale CFD Modeling of PEALD

While details can be found in [28], 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.

2.1 Plasma Chamber CFD Model

The remote plasma system constructed in [28] 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 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 2.2.

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:

$$\vec{\nabla}(\varepsilon_{\mathrm{f}}\vec{B}) = \vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \tag{2.1}$$

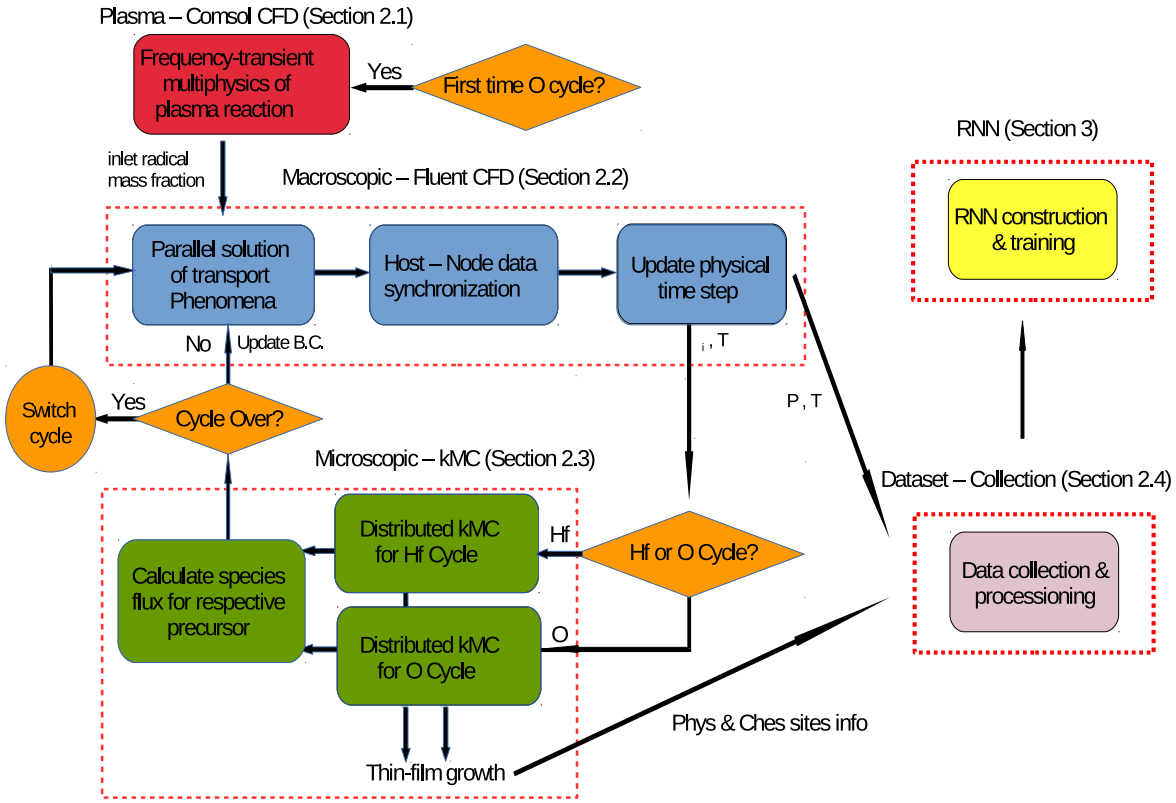


Figure 2.1: Integration of the multiscale computational fluid dynamics model, dataset generation and recurrent neural network development.

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.

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 [28]. 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

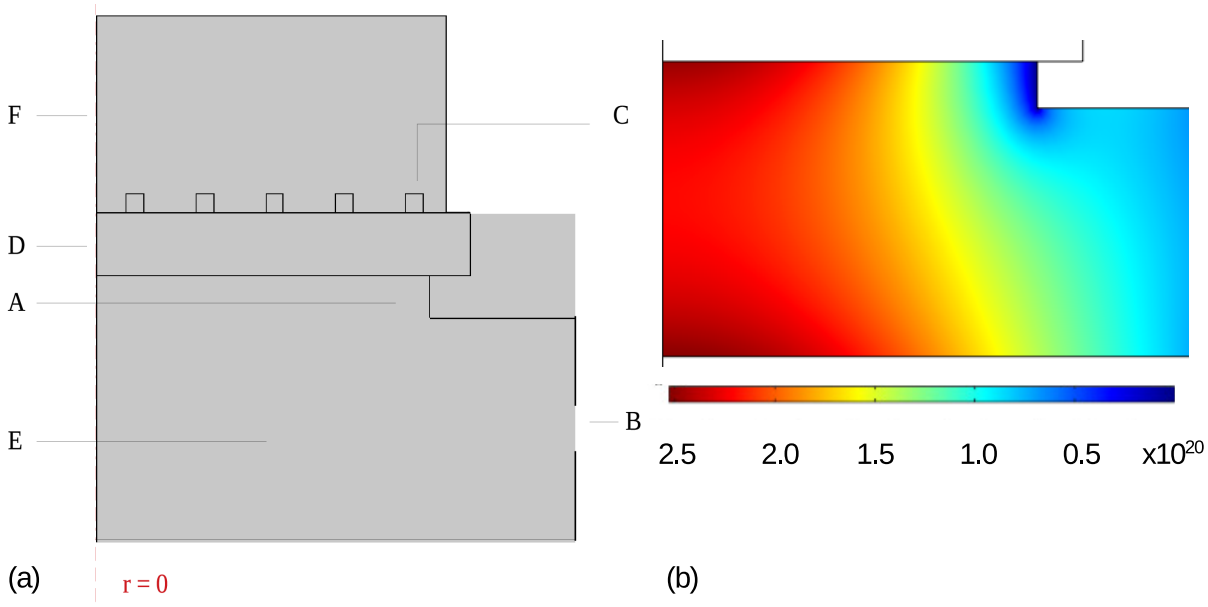


Figure 2.2: (a) Plasma generation chamber axis-symmetry. A - gas inlet, B - gas outlet, C - coil circuits, D - circuit wall, E - reactor chamber, and F - relevant surrounding. (b) Example of number density distribution of O radical, an important plasma species involved in the deposition reaction.

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.

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 [15]. Specifically, as shown in Figure 2.3 [37], 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 [28].

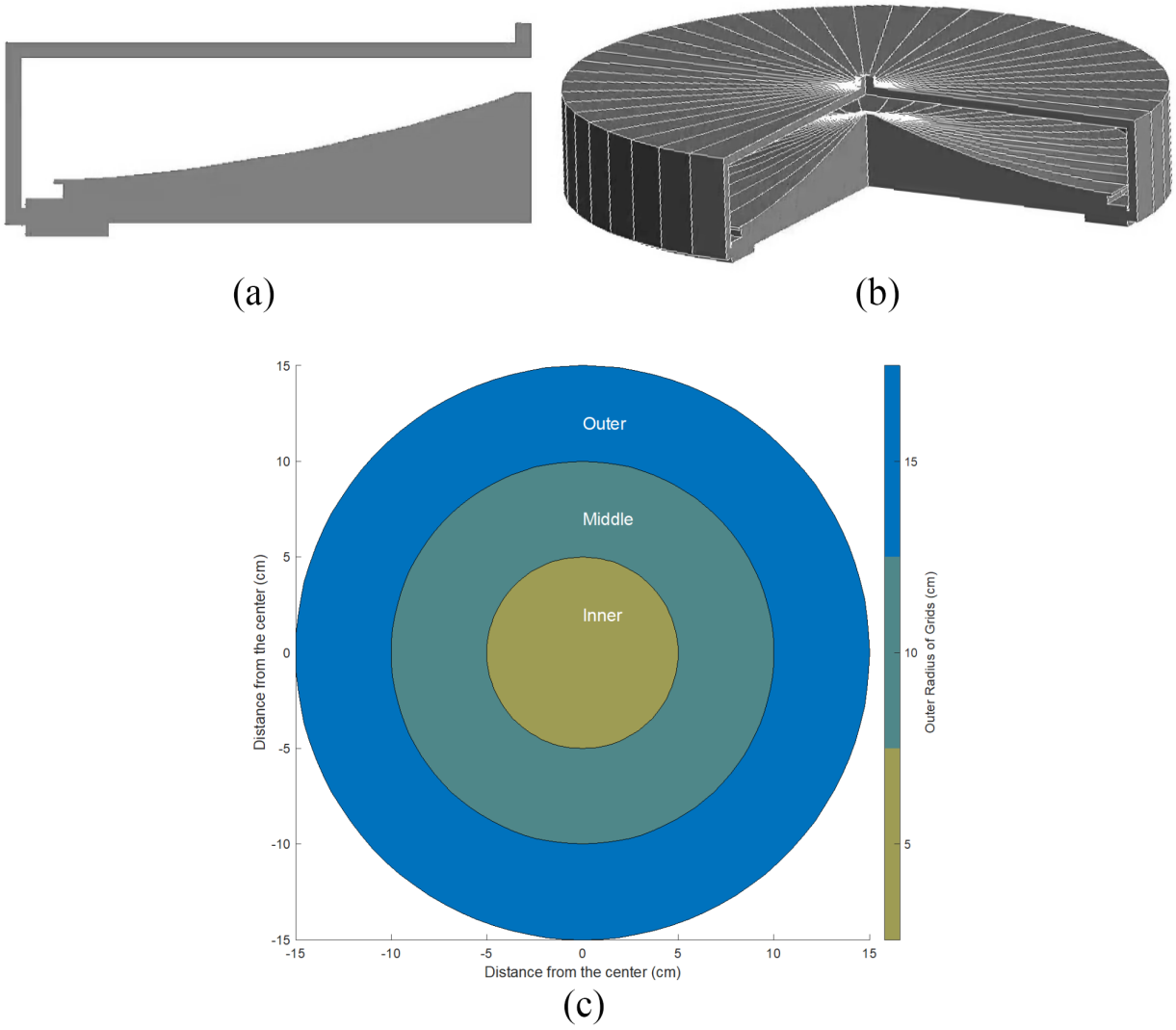


Figure 2.3: (a) 2D axisymmetric PEALD reactor geometry. (b) 3D PEALD reactor geometry. (c) Indication of the inner, middle, and outer regions of the wafer.

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 [28]. 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 [38]. 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 [39].

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 [40]. Simulation accuracy and efficiency are greatly affected by the chosen calculation timestep [41]. The timestep size can be related to the number of cells that the fluid information 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 [42]:

$$\Delta t = \frac{C\Delta x}{u} \tag{2.2}$$

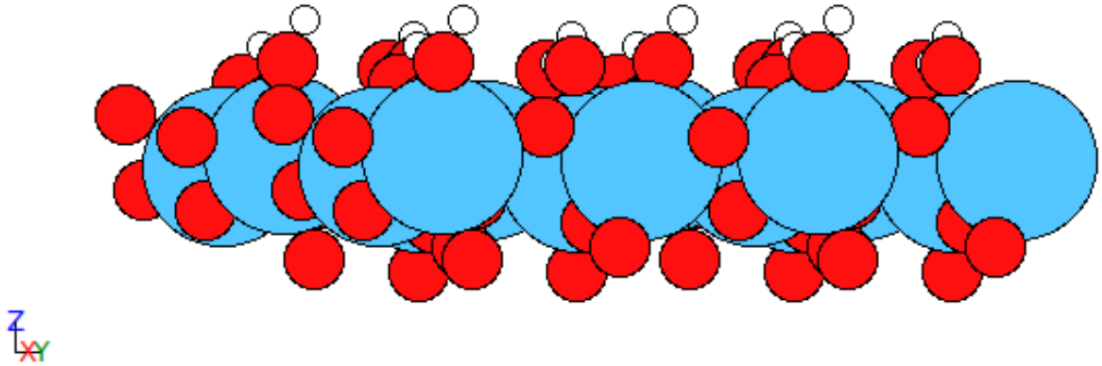
CFD computation time can be largely reduced when advanced CFD software is used [42]. Specifically, the pressure-based solver in this work uses a Courant number of 50.

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 [27] 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 [27].

A monoclinic-alike structure, reported in [43], 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 2.4. A 3D triangular on-lattice approximation developed in [44] is used to efficiently represent this surface as shown in Figure 2.4. 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 [45].



	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 2.4: 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.

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 [26] and [46], 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 2.5. 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.

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:

$$r_{rxn} = A\exp\left(\frac{-E_a}{RT}\right) \tag{2.3}$$

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

$$r_{phs} = \frac{p}{RT}\sqrt{\frac{8RT}{\pi m}} s_c N_a \sigma \tag{2.4}$$

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 [48].

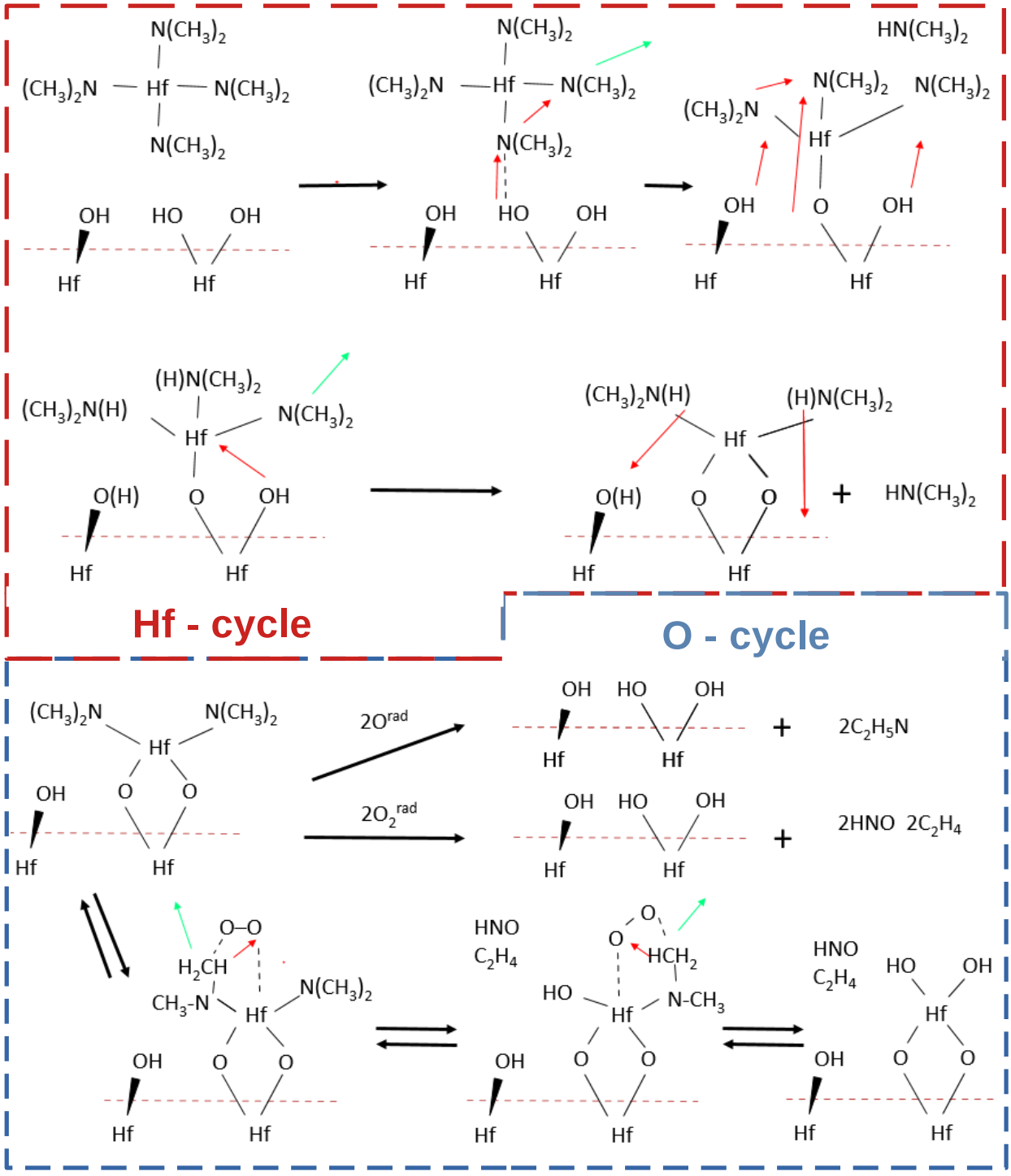


Figure 2.5: Illustration of reaction mechanisms for the precursor surface reactions. The black arrows denote the reaction pathways, and the red arrows denote potential proton diffusion.

The reaction selection and time evolution are then determined using the modified n-fold hybrid kMC algorithm by [44] to simulate the film growth and the deposition profile. The total reaction

rate r_{total} can be calculated as the sum of each rate:

$$r_{total} = \sum_{i=1}^{N} r_i \tag{2.5}$$

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:

$$l_i = \frac{\sum_{j=1}^{i} r_{unweighted}^{j} C_j}{\sum_{k=1}^{N} r_{unweighted}^{k} C_k}, \quad i = 1,...,N \tag{2.6}$$

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 [27]. 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:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \tag{2.7}$$

2.4 Multiscale CFD Model Customization, Workflow and Data Generation

Figure 2.1 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 2.6. Details regarding the 3D grid partition can be found in [27]. 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.

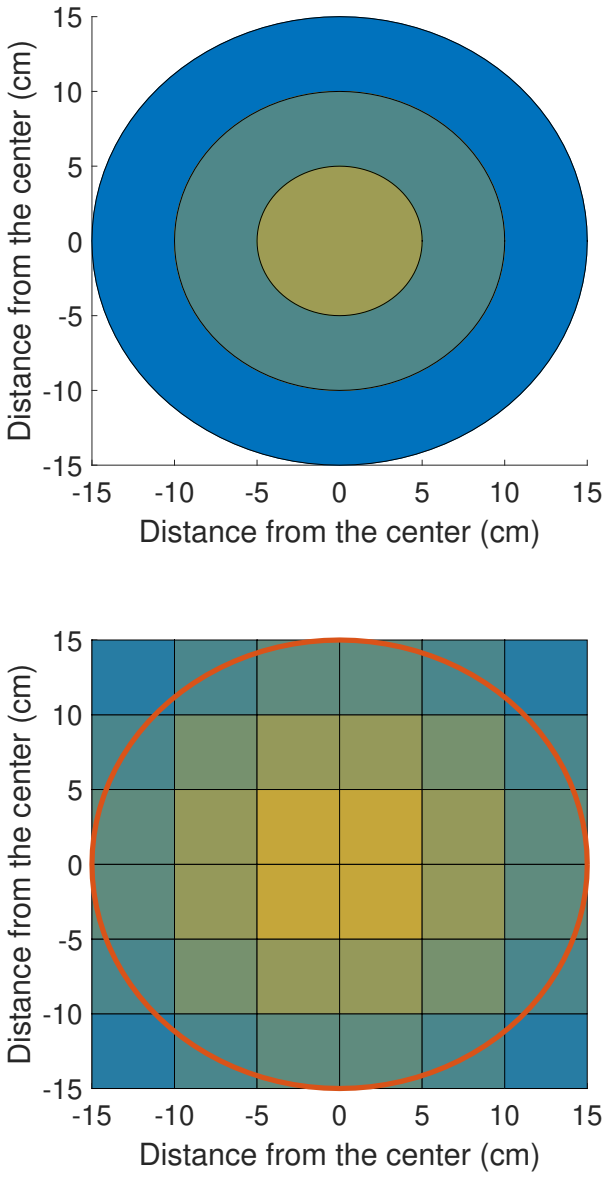


Figure 2.6: Comparison of 2D and 3D surface region partition. Color from yellow to blue represents the distance between the center of the region to the center of the wafer. (a) 2D surface concentric rings region representation. Each ring represents a separate kMC model. (b) 3D surface grid region representation. Each grid represents a separate kMC model.

Chapter 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 [31]. 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 [49]. 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:

$$h(t) = f(h(t-1), x(t-1), \theta) \tag{3.1}$$

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.

3.1 Long and Short Term Memory (LSTM) Method

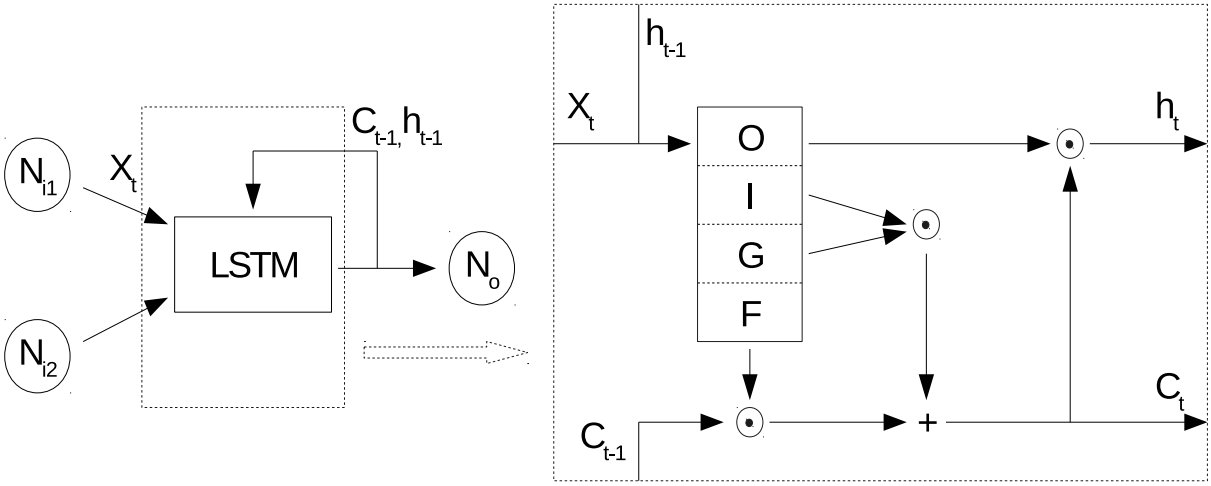


Figure 3.1: Long short-term memory (LSTM) recurrent neural network (RNN) structure. Left: general RNN formulation with LSTM cell. Right: detailed manifestation of the LSTM cell. N_i are input neurons, N_o is the output neuron, C_t and C_{t-1} are the cell state memory for training iteration t and $t-1$, 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.

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 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 3.1, and the model can be mathematically formulated as follows:

$$g = tanh(x_t U^g + h_{t-1} V^g) \tag{3.2}$$

$$i = \sigma(x_t U^i + h_{t-1} V^i) \tag{3.3}$$

$$f = \sigma(x_t U^f + h_{t-1} V^f) \tag{3.4}$$

$$s_t = s_{t-1} \circ f + g \circ i \tag{3.5}$$

$$o = \sigma(x_t U^o + h_{t-1} V^o) \tag{3.6}$$

$$h_t = tanh(s_t) \circ o \tag{3.7}$$

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 2.3. 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:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 m_{t-1} + (1-\beta_2) g_t^2 \end{aligned} \tag{3.8}$$

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:

$$w_t = w_{t-1} - \eta \frac{m_t}{\sqrt{v_t} + \varepsilon} \tag{3.9}$$

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.

Chapter 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.

4.1 2D Integrated Data-driven Model for PEALD Reactor

According to the formulation discussed in Section 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 4.1. 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 4.1 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 4.1.

Additionally, the prediction of the average partial pressure and physisorption sites are demonstrated in Figure 4.2, 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 4.2, the RNN predictions and CFD results closely

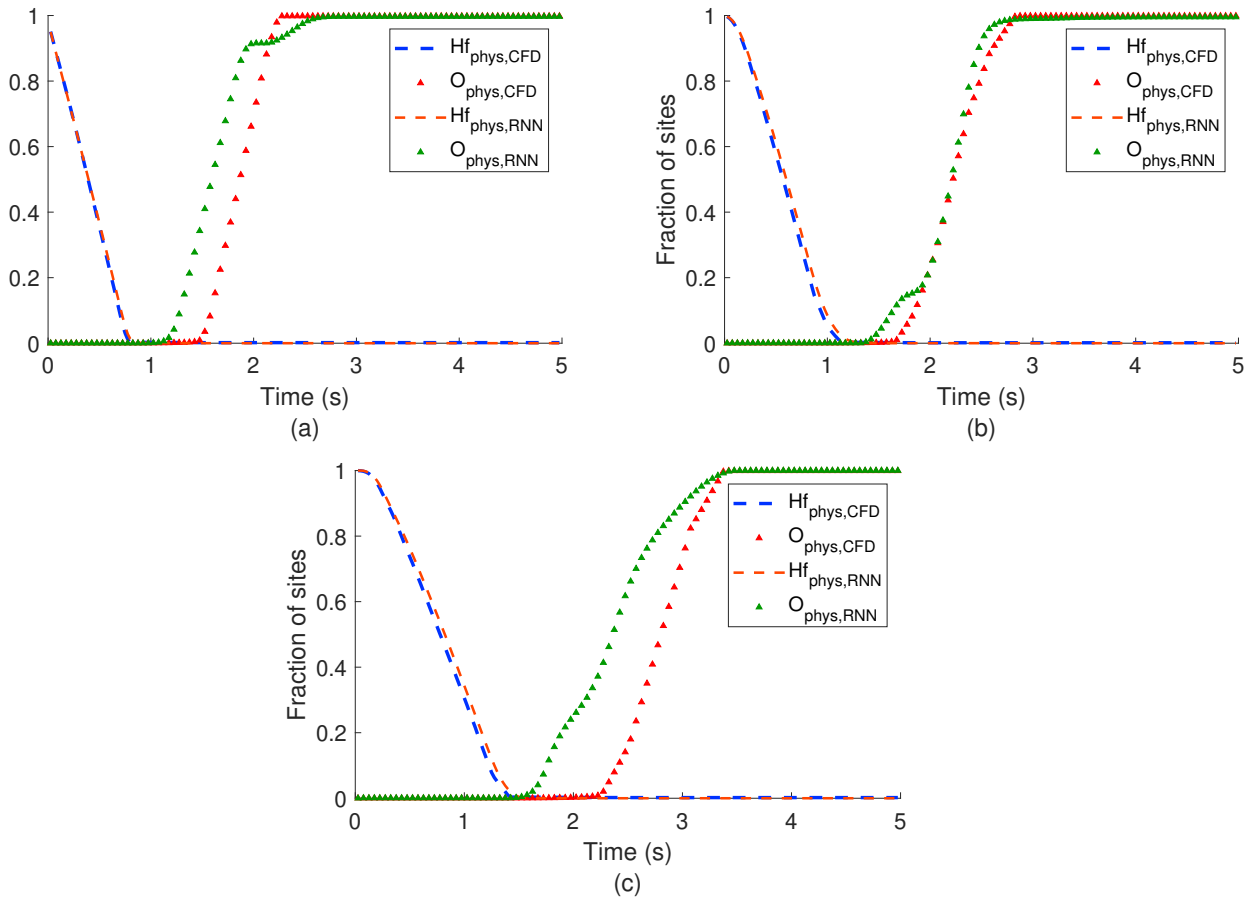


Figure 4.1: Comparison between RNN prediction and CFD simulation result without the showerhead for Hf-Cycle at inner, middle, and outer wafer regions, corresponding to (a), (b), and (c), respectively. The x-axis represents the time in a half-cycle, and the y-axis represents the fraction of sites. The 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.

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 [31], this RNN-based data-driven model fully integrates the microscopic domain and the macroscopic domain and is able to achieve higher accuracy.

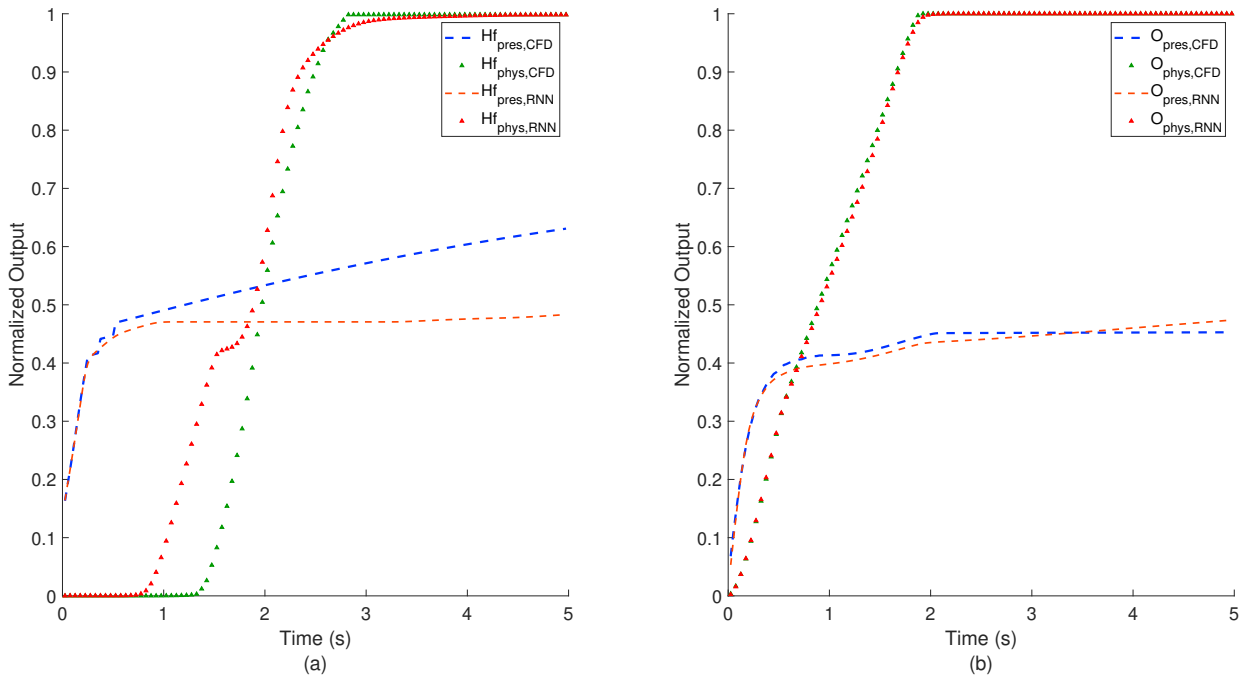


Figure 4.2: Comparison of RNN for two half-cycles in the reactor without the showerhead: (a) Hf-Cycle (b) O-Cycle. The x-axis represents the time within a half-cycle and the y-axis represents the normalized property value (partial pressure and physisorption sites availability). The blue and orange dashed 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.

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 [28]. As shown in Figure 2.3, 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 4.3. 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 4.4. 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.

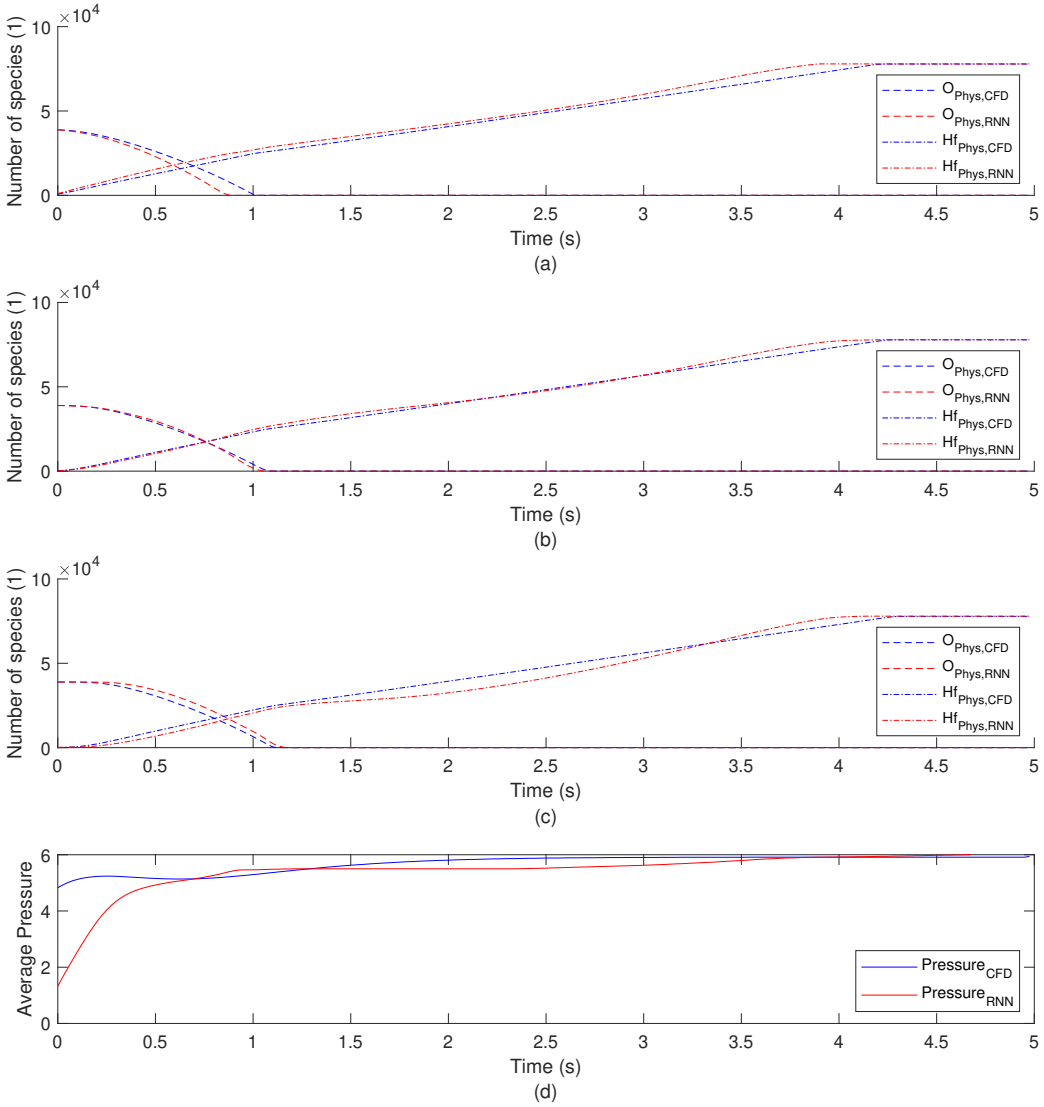


Figure 4.3: Comparison of 3D multiscale CFD model O-Cycle result with RNN-based data-driven model prediction under the inlet flowrate of 1.25×10^{-5} kg/s. Surface profiles over a duration of 5 s are demonstrated and surface steady-state is achieved towards the end of the cycle. x-axis represents the 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.

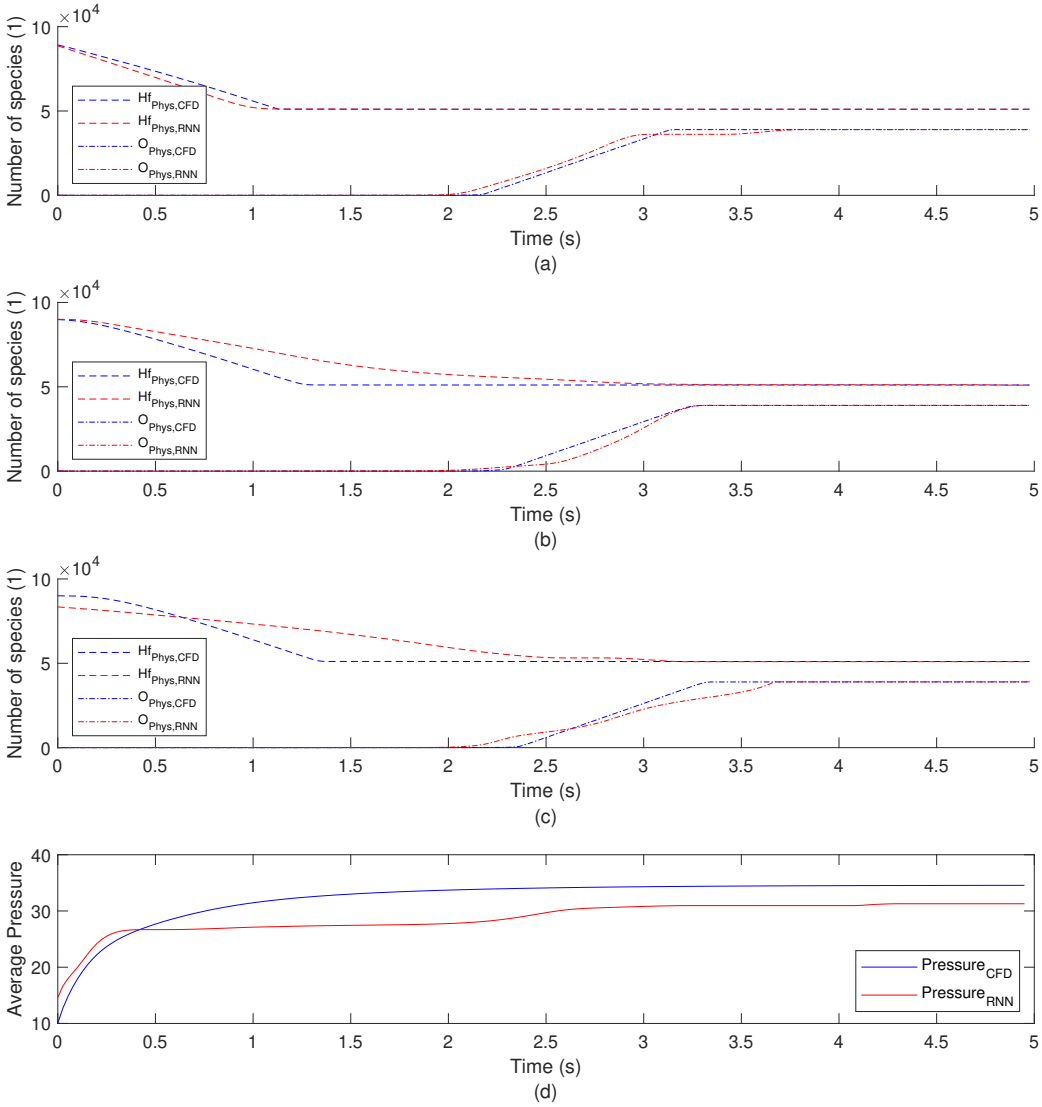


Figure 4.4: Comparison of 3D multiscale CFD model Hf-Cycle result with RNN-based data-driven model prediction under the inlet flowrate of 4.0×10^{-5} kg/s. Surface profiles over a duration of 5 s are demonstrated and surface steady-state is achieved towards the end of the cycle. x-axis represents the 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.

Chapter 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 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 5.1 (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 5.1 (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 5.1 (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 5.1 (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 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 5.2 (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 5.2 (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 [27] and [28]. 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.

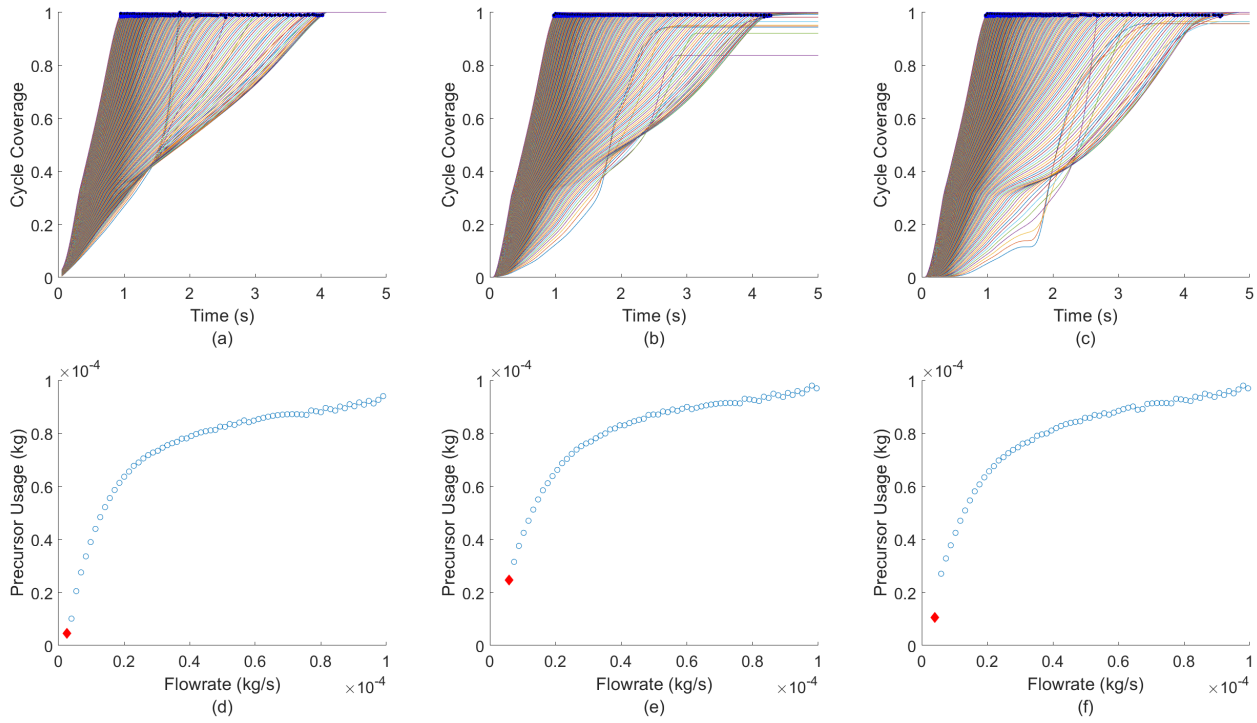


Figure 5.1: (a), (b), (c) Dynamic profiles for O-Cycle for flowrates between $\left[2.50 \times 10^{-6}, 9.75 \times 10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the O-Cycle 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.

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 5.3 and Figure 5.4. The O physisorption site characterizations and the final

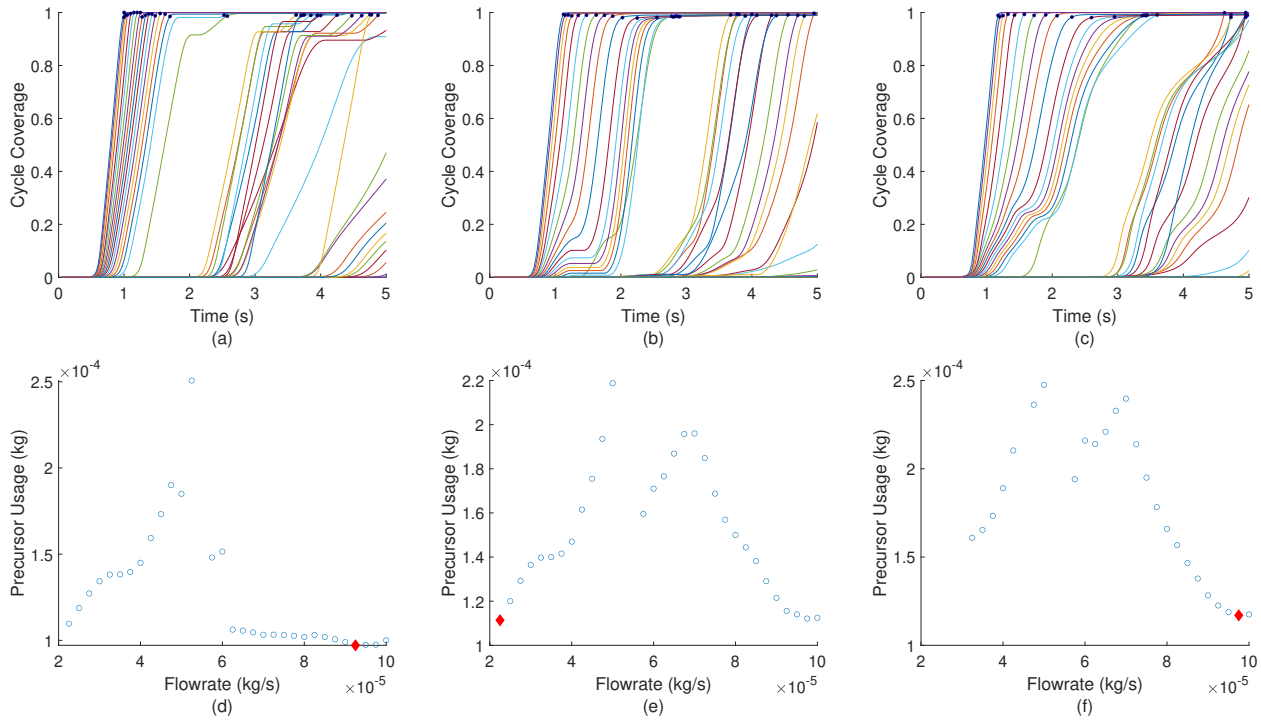


Figure 5.2: (a), (b), (c) Dynamic profiles for Hf-Cycle for flowrates between $\left[2.50\times10^{-6}, 9.75\times10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and the y-axis is the O-Cycle 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.

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 5.5 and Figure 5.6. 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

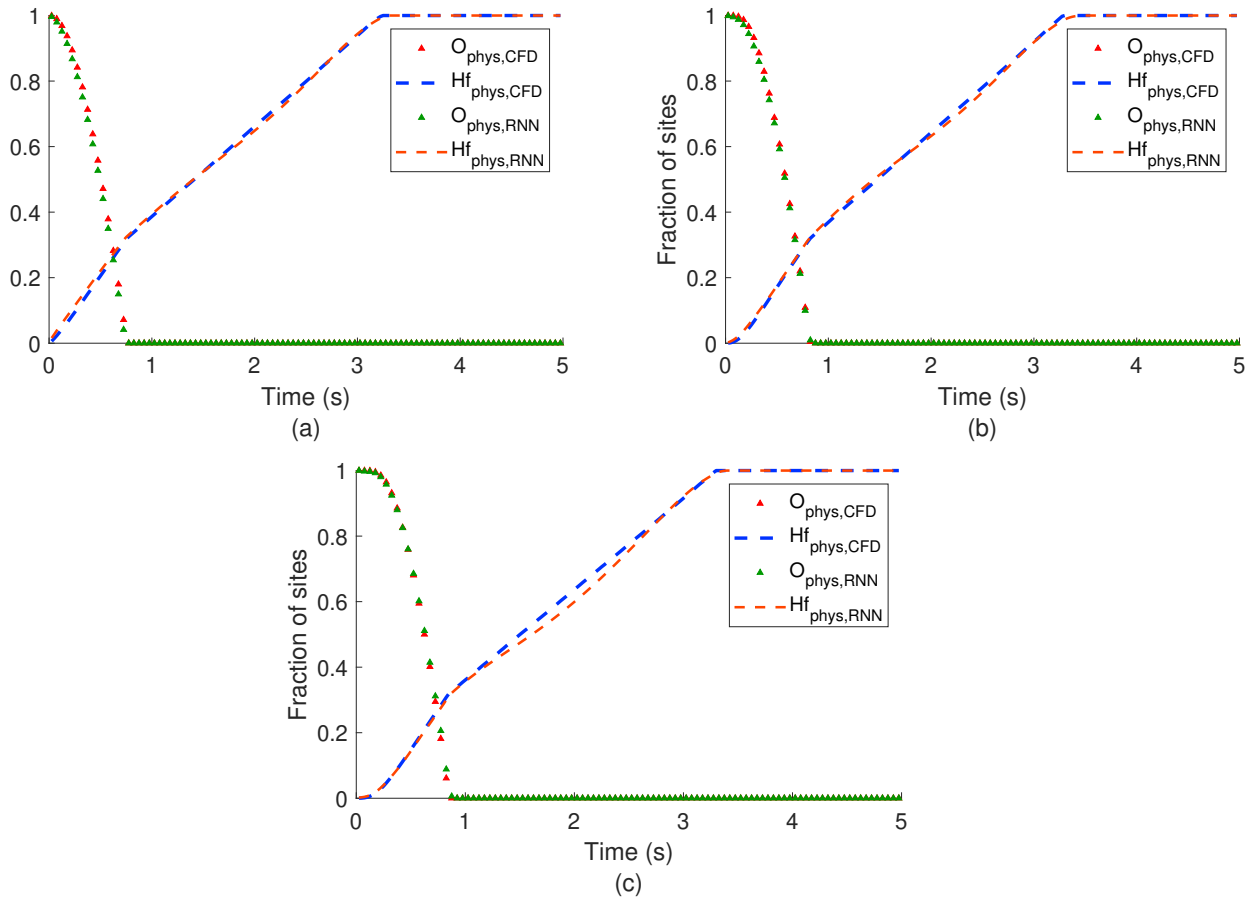


Figure 5.3: Comparison between RNN prediction and CFD simulation result with the showerhead for O-Cycle at inner, middle, and outer wafer regions, corresponding to (a), (b), and (c), respectively. The x-axis represents the time in a half-cycle, and the y-axis represents the fraction of sites. The blue 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.

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 5.6, 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

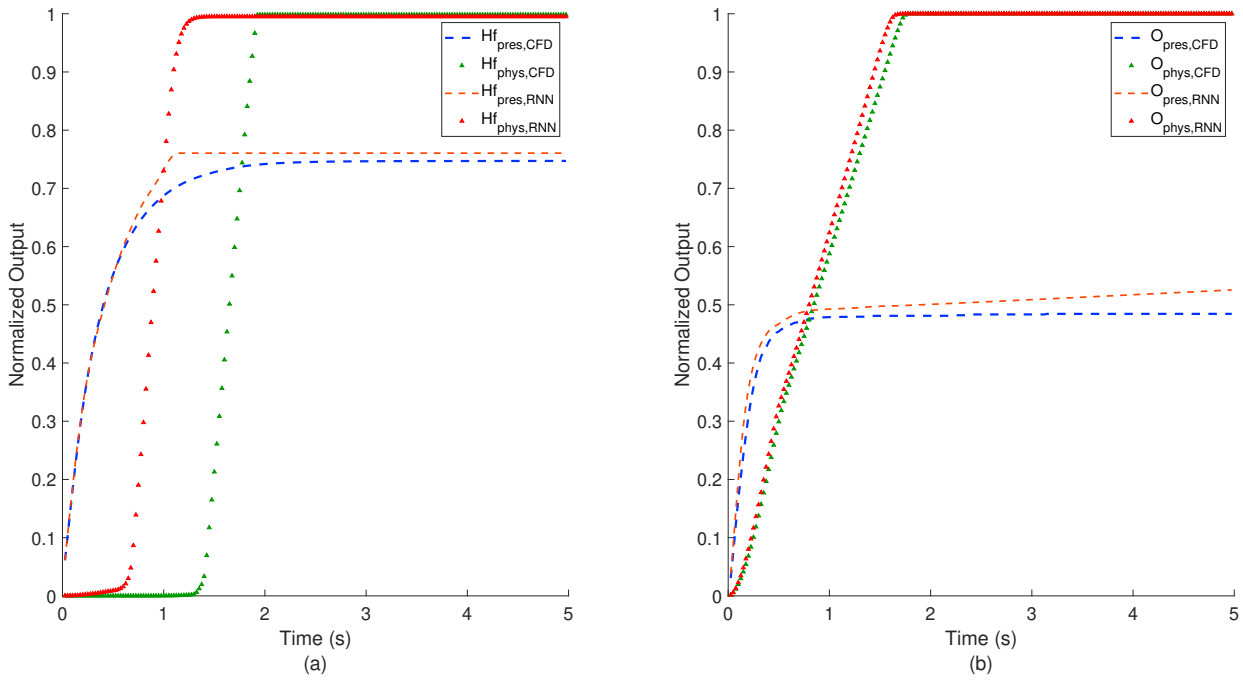


Figure 5.4: Comparison of RNN for two half-cycles in the reactor with the showerhead: (a) Hf-Cycle (b) O-Cycle. The x-axis represents the time within a half-cycle and the y-axis represents the normalized property value (partial pressure and physisorption sites availability). The blue and orange dashed 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.

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.

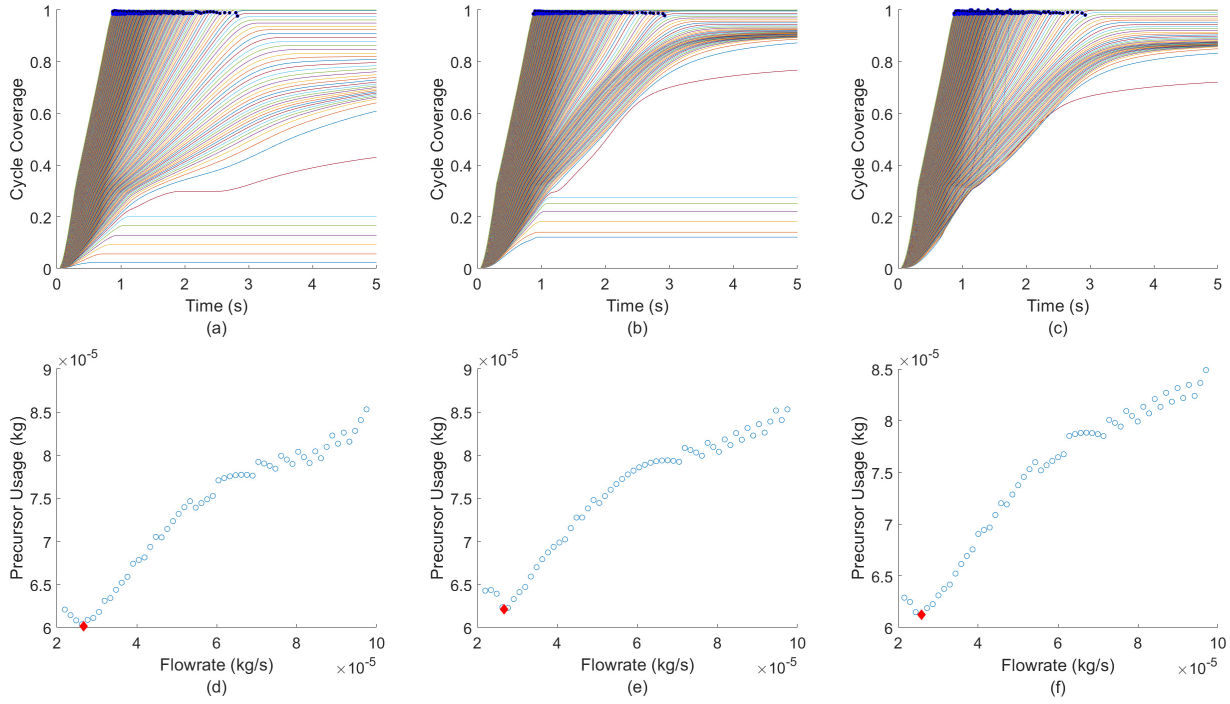


Figure 5.5: (a), (b), (c) Dynamic profiles for O-Cycle for flowrates between $\left[2.50\times10^{-6}, 9.75\times10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively, in a showerhead reactor. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and 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.

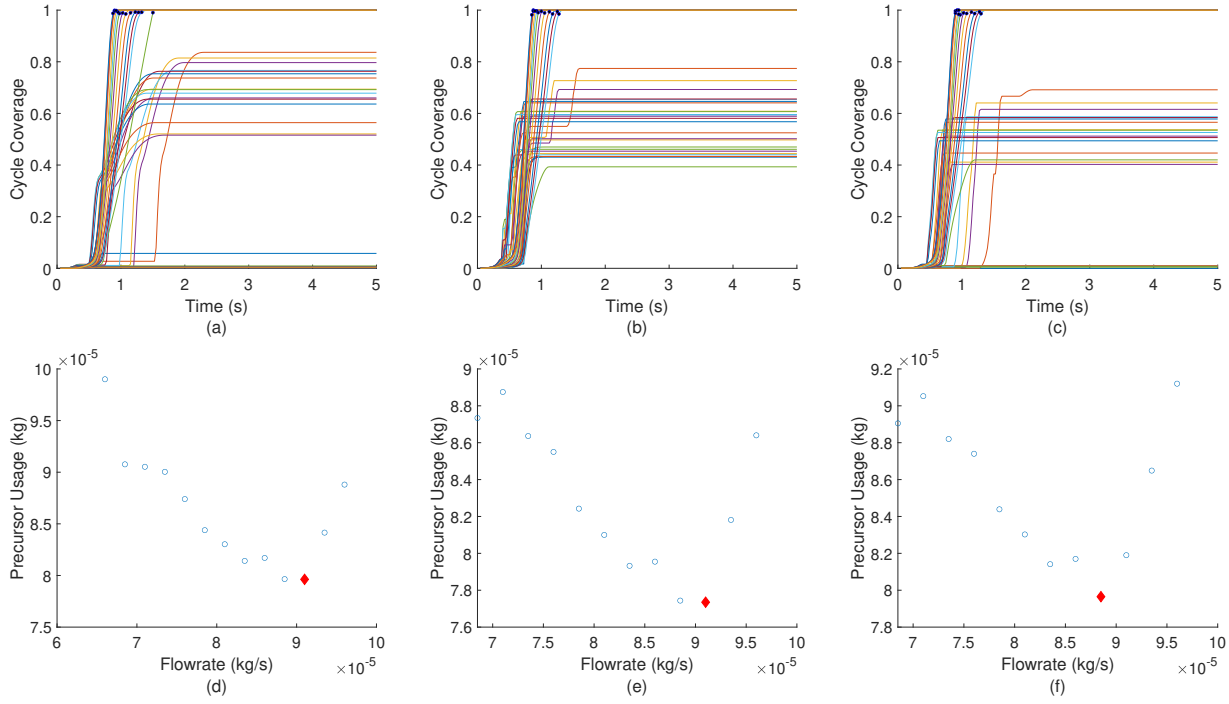


Figure 5.6: (a), (b), (c) Dynamic profiles for Hf-Cycle for flowrates between $\left[2.50\times10^{-6}, 9.75\times10^{-4}\right]$ kg/s for inner, middle, outer wafer regions, respectively, in a showerhead reactor. Each line corresponds to the profile of a specific flowrate. The x-axis is the time and 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

In this work, 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 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 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.

Bibliography

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

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

[15] 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).

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

[17] 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.

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

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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

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

[41] 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.

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

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

[49] 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.