Contents lists available at ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

# Machine learning-based modeling and operation for ALD of $SiO_2$ thin-films using data from a multiscale CFD simulation

*Yangyao Ding*[a], *Yichi Zhang*[a], *Yi Ming Ren*[a], *Gerassimos Orkoulas*[c], *Panagiotis D. Christofides*[a,b,*]

[a] *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA*
[b] *Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA*
[c] *Department of Chemical Engineering, Widener University, Chester, PA 19013, USA*

ARTICLE INFO

*Article history:*
Received 14 August 2019
Received in revised form 28 August 2019
Accepted 2 September 2019
Available online 11 September 2019

*Keywords:*
Atomic layer deposition
Multiscale modeling
Computational fluid dynamics modeling
Kinetic Monte Carlo modeling
Machine learning modeling
ALD operation

ABSTRACT

Atomic layer deposition (ALD) is a widely utilized deposition technology in the semiconductor industry due to its superior ability to generate highly conformal films and to deposit materials into high aspect-ratio geometric structures. However, ALD 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. In this work, a multiscale data-driven model is proposed and developed to capture the macroscopic process domain dynamics with a linear parameter varying model, and to characterize 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 to 1.2 h 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.



## 1. Introduction

Deposition, along with photo-lithography and etching, is one of the most important building blocks in thin-film material development. In the semiconductor industry, the stringent demands on the scale and complexity of the electronic devices have continuously pushed the deposition techniques to be more precise and controllable. For instance, the state-of-the-art flash memory devices, such as the 3D NAND (not-and) memory and the dynamic random-access memory (DRAM),

* *Corresponding author at*: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA.
E-mail address: pdc@seas.ucla.edu (P.D. Christofides).
https://doi.org/10.1016/j.cherd.2019.09.005

require sophisticated non-planar 3D design and ultra-thin gates (<10 nm) with demanding conformity and defect formation criteria (George et al., 1996; Schuegraf et al., 2013). Originated as a derivative of the chemical vapor deposition (CVD), atomic layer deposition (ALD) has been developed to build the thin-films that meet the demands for more uniform films and higher aspect-ratio micro-structures (Kääriäinen et al., 2013). In a typical ALD process, two precursor gases are introduced in pulses into the reactor sequentially, with inert purge gas pulse injected between the two precursor pulses. Each precursor pulse is called a half-cycle, and the inert gas pulses are called purging. In each half-cycle, ideally, only one precursor will be in contact and react with the substrate surface. Given the appropriate operating conditions and cycle-time, an ideal ALD half-cycle will be self-limiting, resulting in a highly conformal and fully covered substrate surface terminated with the desired element introduced by the precursor species (George, 2009). Following this alternating precursor scheme, ALD can precisely deposit materials layer-by-layer and effectively allow the uniform coverage on complex geometries (Tanner et al., 2007; Foong et al., 2010; Shirazi and Elliott, 2014; Ishikawa et al., 2017). As a result, ALD has been widely utilized in the field of nanoelectromechanical systems (NEMS), especially for the manufacturing of semiconductor memory devices.

The past decade has witnessed an increasing research interest on ALD and a growing market in the industrial manufacturing of thin-film materials (Raaijmakers, 2011; Kääriäinen et al., 2013). The ALD industry also holds a promising future with its global market expected to reach US$3.2 billion by 2026. Novel precursors and reaction mechanisms were discovered, making the film processing of high aspect-ratio substrate layouts more efficient and feasible. Yet, the limited throughput and the high manufacturing cost of ALD production still trouble the manufacturers, especially amid the fluctuating global semiconductor market. Despite the urgent need for further enhancement of the ALD processes, the overall cost of precursors and equipment thwarts the progression of more extensive research efforts (Shirazi and Elliott, 2014). More specifically, one of the major roadblocks to a deeper understanding of the ALD process is the coupled effect between the reactor gas-phase development and the microscopic thin-film deposition process. Beside the apparent absence of a clear and direct theoretical explanation, it is also extremely difficult to obtain an empirical database of film growth during the experiments due to the limitations in real-time in-situ sensing. Scanning electron microscopy (SEM) and scanning tunneling microscopy (STM) provide detailed film profiles but are often expensive and destructive to the deposited film, and therefore, they cannot be used to perform on-line monitoring (Chen, 1993; Goldstein et al., 2017). Quartz crystal microbalance (QCM) does a good job of measuring the overall deposition rate by examining the total mass of the substrate surface in real-time, but fails to unveil local changes within the substrate (Elam et al., 2002). Another option is the in-situ spectroscopic ellipsometry, which is capable of providing local structural information of the film. However, the operation of spectroscopic ellipsometry equipment is very complicated and thus is not prevalent in industry (Pittal et al., 1993; Dalton et al., 1994). With all aforementioned shortcomings, the existing hardware monitoring methods, despite their fidelity and capability to report the actual film profiles, are not yet applicable for a complete measurement of local film coverage profiles during the industrial ALD process. Therefore, an accurate and all-inclusive simulation model for the ALD processes could be helpful in estimating the transient film coverage and in understanding the real-time development on both microscopic and macroscopic scales, which could be of significant value to both industrial and experimental works.

Many efforts have been done in the past to develop an appropriate model for thermal ALD processes. 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., Ding et al., 2019). 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 (Christofides et al., 2008). 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 (Pan et al., 2014; Crose et al., 2018). 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 (Knoops et al., 2010; Elliott and Greer, 2004; Weckman et al., 2018; Crose et al., 2018). 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 (Ding et al., 2019). Specifically, using a standalone 3D surface deposition kMC model, we successfully reported a growth rate that lies within the experimental range (1.4 Å ~ 2.1 Å per cycle); and in Zhang et al. (2019), 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. Djurabekova et al. (2007) and Nicolas and Lorenzo (2010) have employed an artificial neural network (ANN) to characterize the result of kMC simulation for lattice microscopic structures. Chaffart and Ricardez-Sandoval (2018) and Kimaev and Ricardez-Sandoval (2019) 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 (Zhang et al., 2019). 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.

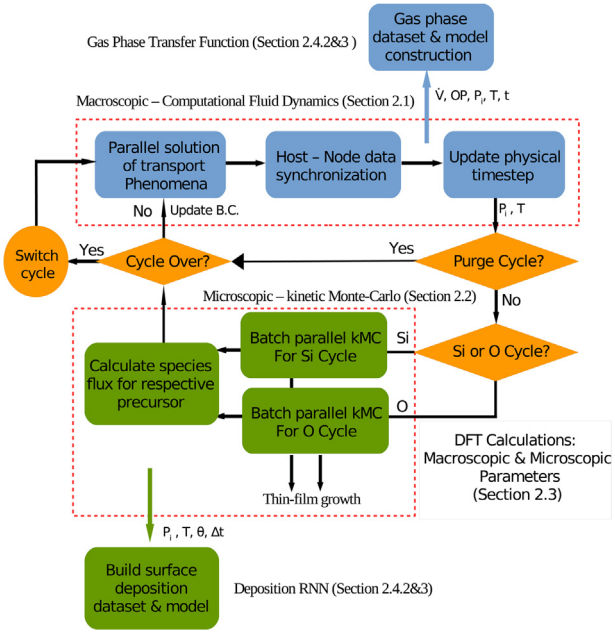


**Fig. 1 – Multiscale workflow, parallel processing and information exchange illustration.**

## 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.

Fig. 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 Zhang et al. (2019) as well as the data collection in both microscopic and macroscopic domains.

### 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, Zhang et al. (2019) 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 Fig. 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 (King, 2011). As shown in Fig. 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 (Zhang et al., 2019). 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 Zhang et al. (2019), 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 (Hintermüller and Ulbrich, 2004). 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 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. (Fluent, 2013).

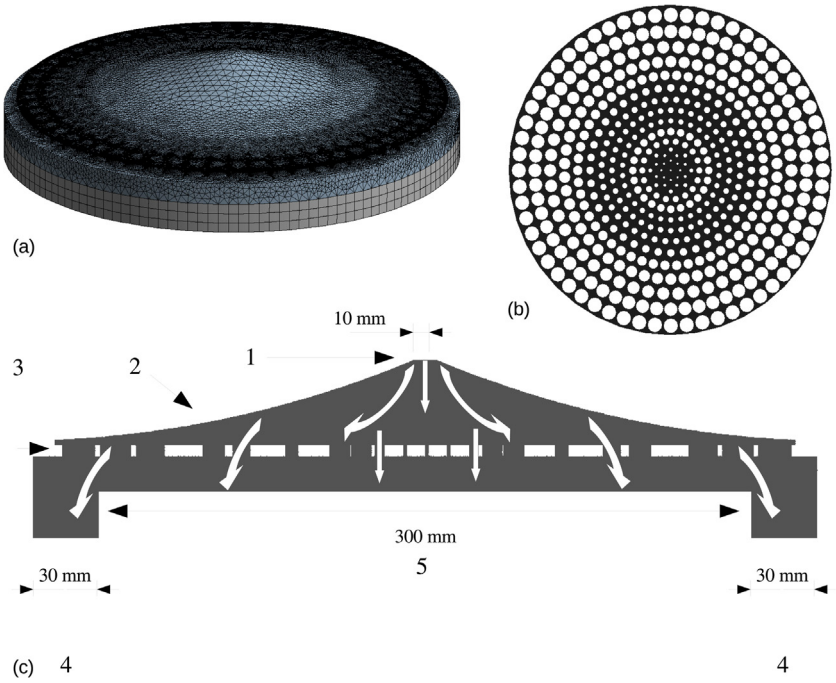


**Fig. 2 – (a) Hybrid reactor mesh. (b) Optimized radially distributed showerhead configuration. (c) x-cross-section geometry of the optimized ALD reactor. The gas inlet is labeled 1; the upstream region is labeled 2, where the gas-profile gradually develops; the showerhead holes is labeled 3; the gas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

**Table 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 |

The transient gas-phase CFD calculation is realized using the ANSYS Fluent pressure-based solver (Fluent, 2013). 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 2.3. 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 (Young et al., 2010). 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 Zhang et al. (2019). 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 (Acton, 2012; Fluent, 2013):

$$\bar{\bar{\tau}} = \mu\left[(\nabla\vec{v} + \nabla\vec{v}^{T}) - \frac{2}{3}\nabla\cdot\vec{v}I\right] \tag{1}$$

$$\frac{\partial(\rho\vec{v})}{\partial t} + \nabla\cdot(\rho\vec{v}\vec{v}) = -\nabla P + \nabla\cdot(\bar{\bar{\tau}}) + \rho\vec{g} + \vec{F} \tag{2}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla(\vec{v}(\rho E + p)) = \nabla(k\nabla T - \Sigma h\vec{J} + (\bar{\bar{\tau}}\vec{v})) + S_h \tag{3}$$

$$\vec{J}_i = -\rho D_{m,i}\nabla Y_i - D_{T,i}\frac{\nabla T}{T} \tag{4}$$

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla\cdot(\rho\vec{v}Y_i) = -\nabla\cdot\vec{J}_i + R_i + S_i \tag{5}$$

where $\bar{\bar{\tau}}$ is the stress tensor, $\mu$ is the fluid viscosity, $\vec{v}$ is the velocity of gas-phase species mixture, $I$ is the unit tensor, $\rho$ is the density of the gas-phase species mixture, $P$ is the static pressure, $\vec{g}$ is the acceleration of gravity, $\vec{F}$ 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, $\vec{J}_i$ is the diffusive flux of species i, and $D_{m,i}$ and $D_{T,i}$ are the mass and heat diffusivities of species i, 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 (Moura and Kubrusly, 2012):

$$\Delta t = \frac{C\Delta x}{u}, \tag{6}$$

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, $\Delta x$ 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 (Fluent, 2013).

### 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 $1200 \times 1200$ 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 (Huang et al., 2010a,b). A 3D triangular lattice is adopted to approximate the true α-Quartz $SiO_2$ lattice as shown in Fig. 3(a) and (b), which is demonstrated to preserve the fidelity of the key features in the microscopic structure (Ding et al., 2019).

As mentioned in Section 1, 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 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 Fig. 4(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 Han et al. (2011).

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:

$$r_{phs} = \frac{p_i}{RT}\sqrt{\frac{8k_bT}{\pi m_i}}s_{c,i}N_a\sigma \tag{7}$$

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 $\sigma$ 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 (Cortright and Dumesic, 2001). The reaction rate can be described by an Arrhenius-type rate law of the equilibrated transition state complex as follows:

$$r_{rxn,i} = f_i^{TST}\left(\frac{k_bT}{h}\right)\exp\left(\frac{-E_{a,i}}{k_bT}\right) \tag{8}$$

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:

$$r_{total} = \sum_{i=1}^{N} r_i, \tag{9}$$

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:

$$l_i = \frac{\sum_{j=1}^{i} r_j}{r_{total}}, \quad i = 1, \ldots, N \tag{10}$$

Then, a uniformly distributed random number, $\gamma_1 \in (0, 1]$, is generated. If $\gamma_1$ 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 $\Delta t$ of the physical process with the following equation:

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}}, \tag{11}$$

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 Ding et al. (2019).

### 2.3. DFT and thermodynamic calculations

As suggested by Fang et al. (2016), 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

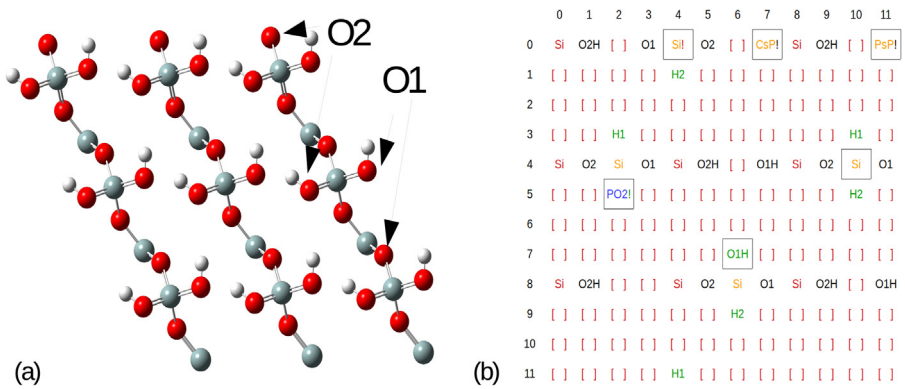


**Fig. 3 – (a) Top view of the hydroxylated $SiO_2$ (0 0 1) surface, where O1 is the more electronegative oxygen. (b) Top view of a 5-layer 12 × 12-site miniature demonstration of the full kMC simulation lattice. The five layers and the species on the lattice are shown using different colors and symbols, 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)**

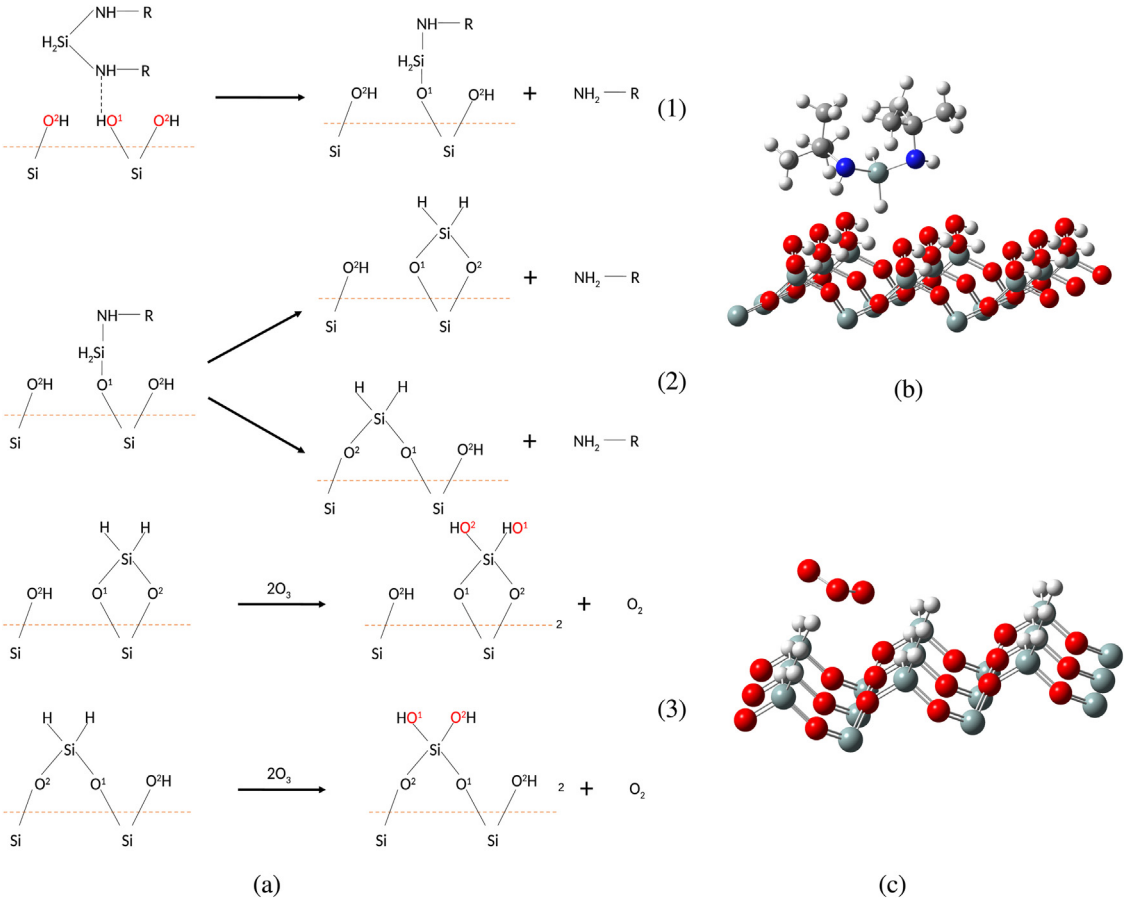


**Fig. 4 – (a) Surface reaction mechanism: (1) First dissociative chemisorption step of BTBAS. (2) Second dissociative chemisorption step of BTBAS under self-binding and neighbour-binding mechanisms. (3) Oxidation of self-binding and neighbour-binding $SiH_2$ with ozone.(b) Physisorbed BTBAS molecule demonstration. (c) Physisorbed ozone molecule demonstration.**

Schrödinger equation for a many-body system by solving the overall electron density (Frisch et al., 2000; Council et al., 2003; Fermeglia and Pricl, 2009; Liu and Liu, 2018):

$$\hat{H}\Psi(r_i, R_I) = E\Psi(r_i, R_I), \tag{12}$$

where $\Psi$ 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 $\hat{H}$ 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 $\mu$ 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 (Curtiss et al., 2007). 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 Ochterski (2000). Furthermore, the viscosity $\mu$ is approximated using the molecular theory of Newtonian gases (Bird et al., 2007).

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 (Lee et al., 1988; Mankad and Jhu, 2016). Then, the G4 theory is again used to acquire an accurate vibrational frequency. The results, along with other important thermodynamic parameters are reported in Ding et al. (2019).

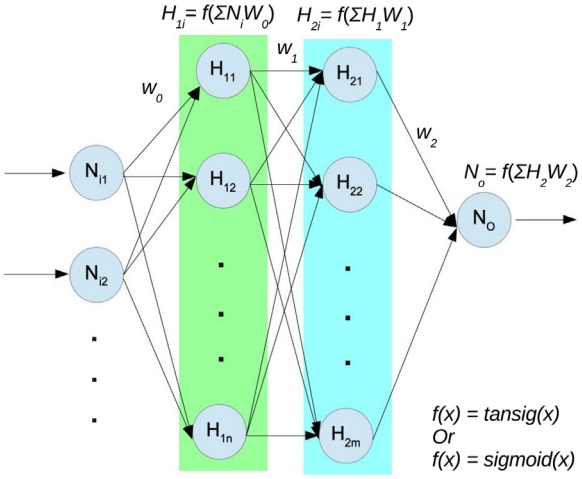


**Fig. 5 – Structure of a feed-forward artificial neural network (ANN) with multiple inputs, two hidden layers, and one output. *W* is the weight for each ANN layer, *N*s and *H*s represent the neurons, and *f* represents the activation function, which can be tansig or sigmoid function based on the neural network structures being used.**

### 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 (Fluent, 2013). 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.

#### 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 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 (Burden and Dave, 2008). 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 Fig. 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 $i$-th 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:

$$A^l = \theta(W^l A^{l-1} + B^l), \tag{13}$$

where $A^l$ is the activation matrix, $W^l$ is the weight matrix, and $B^l$ is the bias matrix for layer $L$, respectively. $\theta$ 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 (MacKay, 1992):

$$tansig(x) = \frac{2}{1+e^{-2x}} - 1 \tag{14}$$

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

$$S(w) = \beta \sum_{i=1}^{N_D} [y_i - f(x_i, w)]^2 + \alpha \sum_{j=1}^{N_w} w_j^2, \tag{15}$$

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 (Moré, 1978). Then, hyperparameters $\alpha$ and $\beta$ are updated using Bayesian inference calculation. The detailed explanation and construction of the optimization problems are discussed in detail by MacKay (1992) and Burden and Dave (2008). To further avoid saturation and high variance activation, batch normalization is applied between layers, thereby facilitating convergence speed and learning rate (Ioffe and Szegedy, 2015).

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:

$$a_3(v, P_{op})\frac{d^3y}{dt^3} + a_2(v, P_{op})\frac{d^2y}{dt^2} + a_1(v, P_{op})\frac{dy}{dt} + a_0(v, P_{op})y = K(v, P_{op})v \tag{16}$$

$$y(0) = 0, \quad \frac{dy}{dt}(0) = 0, \quad \frac{d^2y}{dt^2}(0) = 0, \quad \frac{d^3y}{dt^3}(0) = 0 \tag{17}$$

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.

### 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 h. 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 Eq. (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:

$$y = \begin{cases} b_1\left(1 - \exp\left(-\frac{b_2}{t}\right)\right) & t \le b_3 \\ b_4 & t \ge b_3 \end{cases} \tag{18}$$

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:

$$\begin{aligned} \min_{b_1,b_2,b_3,b_4} \quad & \sum_{i=1}^{N}(y_i - D_i)^2 \\ \text{s.t.} \quad & b_1\left(1-\exp\left(\frac{-b_2}{b_3}\right)\right) \le b_4 \end{aligned}, \tag{19}$$

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 (Wächter and Biegler, 2006).

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. 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.

### 3.1. 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 2.4, 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 Section 2.4.2, deposition steady-state times are successfully determined for all sets of temperature and pressure conditions. As shown in Fig. 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 2. The white noise associated with the kMC predicted transient deposition rate and the accumulated coverage hinder the neural network fitting accuracy and precision. Fig. 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 2.4, 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 Fig. 7(a), a mean square error on the training dataset drops by four orders of magnitude and reaches a final performance of $7.2 \times 10^3$ after 1000 epochs. This error corresponds to a normalized error that is less than $1.0 \times 10^{-3}$ and less than 10% of the transient deposition rate. Fig. 7(b) shows that the testing error of the resulting neural network is normally distributed. In addition, Fig. 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.

A visualization of the trained ANN model performance is shown in Fig. 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.

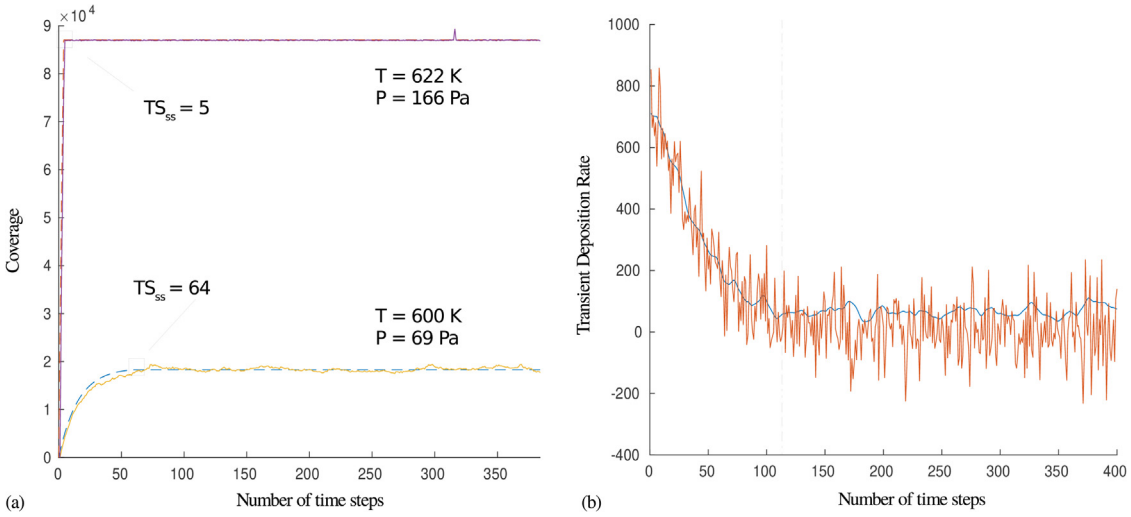


**Fig. 6 – (a) Comparison between the kMC result and the steady-state fitting result at two randomly chosen operating conditions ($T$ = 622 K, $P$ = 166 Pa and $T$ = 600 K, $P$ = 69 Pa). The dashed lines are generated from the fitting and the solid lines are taken from the kMC simulation. The black boxes 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 90,000. (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.**

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.

### 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 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 (Zhang et al., 2019). 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 Fig. 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 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 accu-

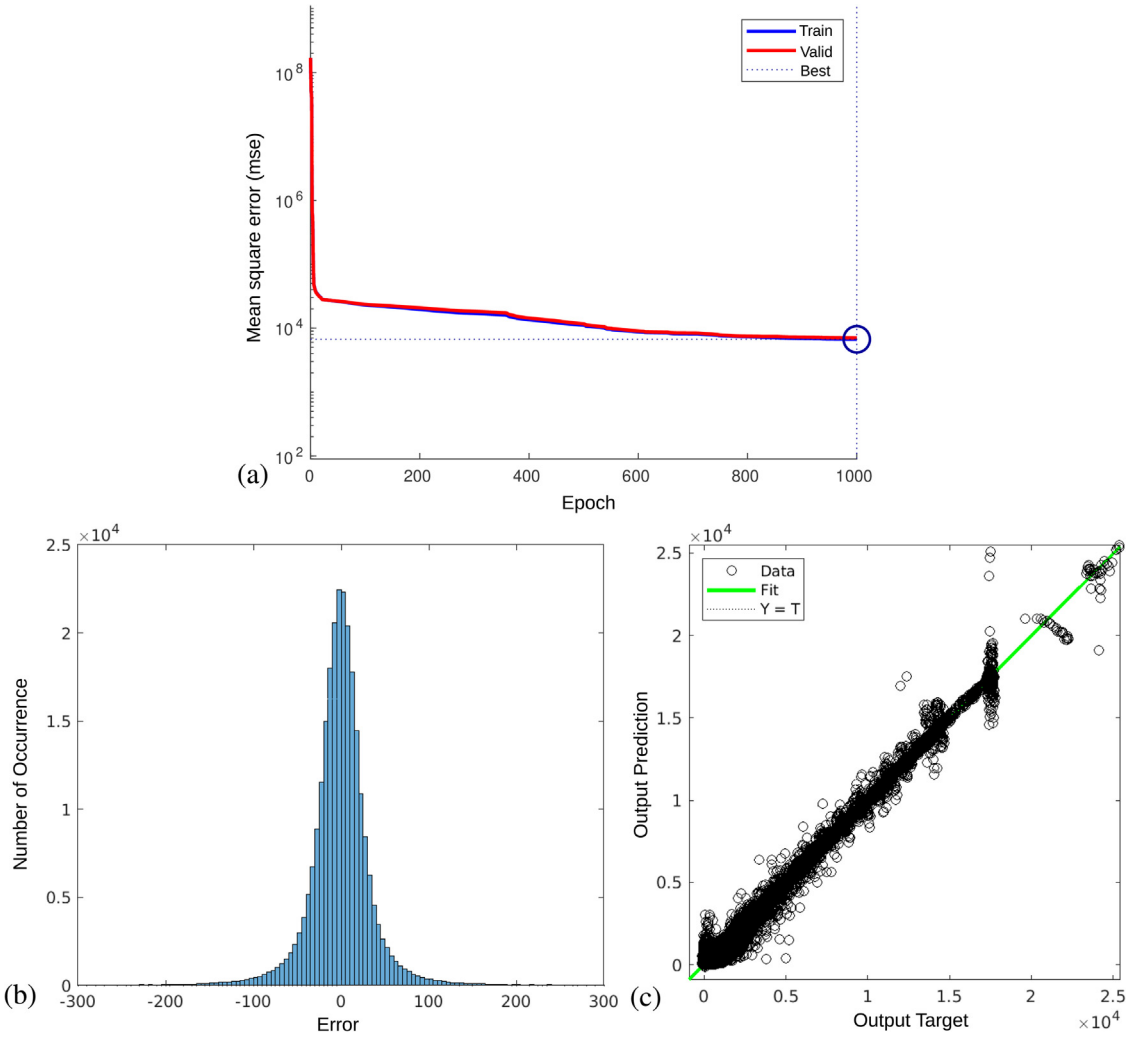


**Fig. 7 – Training and performance of the constructed BRANN model: (a) Training and validation history. The *x*-axis is the training epoch and the *y*-axis is the MSE on output. (b) Prediction error distribution histogram of testing. The *x*-axis is the magnitude of error and the *y*-axis is the number 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.**

rate estimations of half-cycle finishing time as demonstrated in the following section.

### 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 600 K, the operating pressure is 133 Pa, half-cycle process time is 3.5 s, 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 (Zhang et al., 2019).

As shown in Fig. 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, Fig. 10(b) shows the result from the multiscale CFD model for the same operating conditions reported in Zhang et al. (2019), 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.

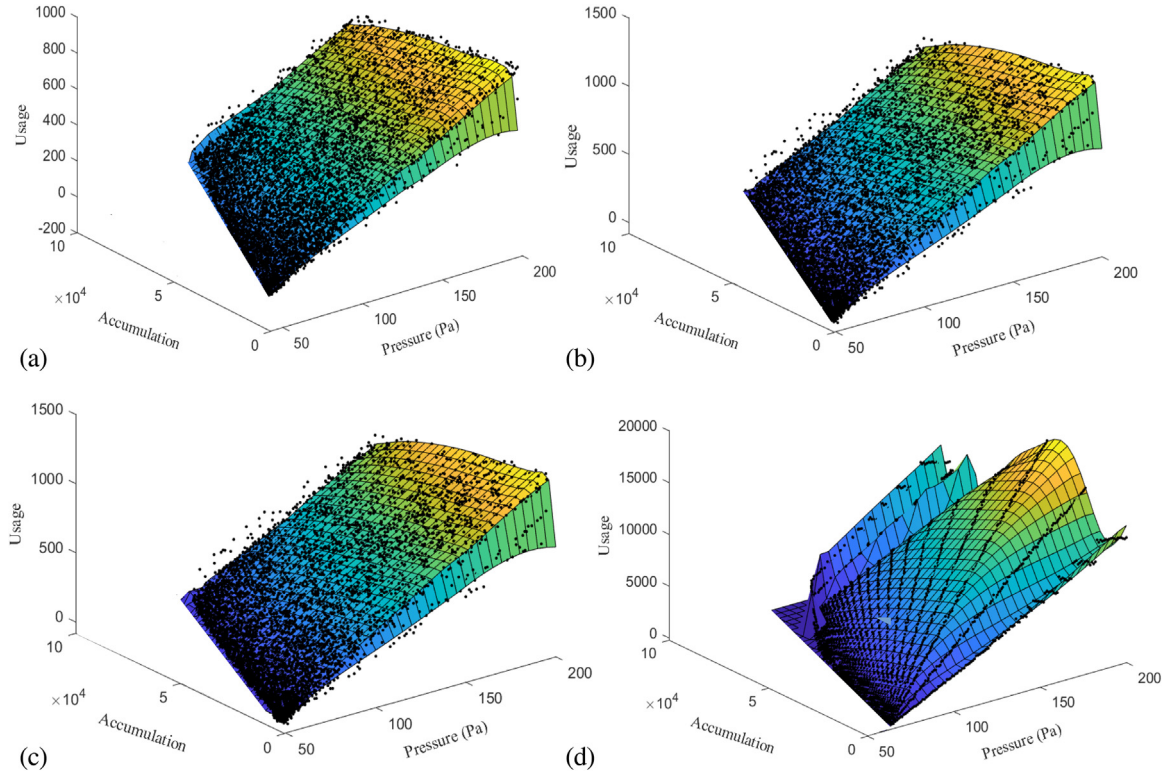


**Fig. 8 – Fitting results of the ANN model. The black dots are the kMC simulation result and the surfaces are the neural network fitting result. The vertical *z*-axis is the output transient deposition rate as in precursor usage, and the horizontal *x*-axis and *y*-axis are the input accumulated coverage 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.**

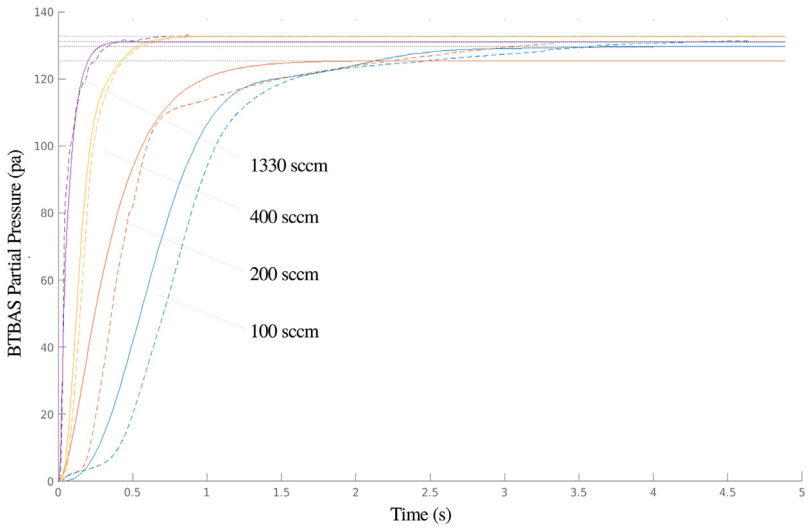


**Fig. 9 – Comparison of the wafer-average linear parameter varying model fitting and the CFD gas-phase development. The *x*-axis is the ALD half-cycle physical time, and the *y*-axis is the surface BTBAS partial pressure. The dashed lines are the original CFD data and the solid lines are the fitting 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)**

### 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 h 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.

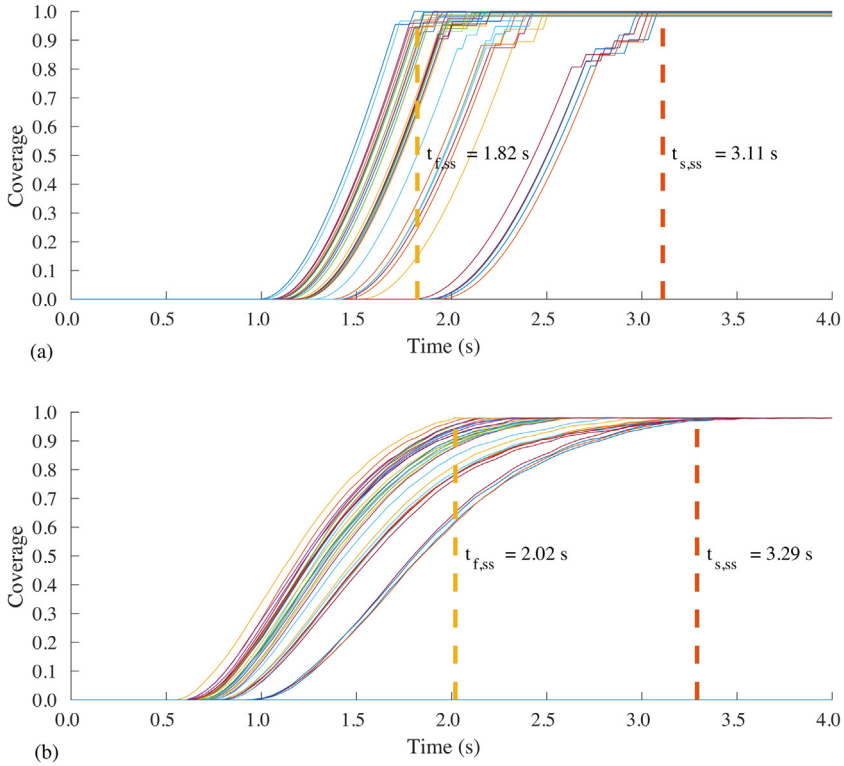


**Fig. 10 – Comparison of one testing scenario between the data-driven model and the validation multiscale CFD data. The test scenario is under a substrate surface temperature of 600 K, an operating pressure of 133 Pa and an inlet feed flow rate of 100 sccm. The x-axis represents the process operation 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. (a) 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 (b) 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.**

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, Ding et al. (2019) 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 3 s for gas phase profile development, which is a good estimation for an 100 sccm inlet flow rate according to Zhang et al. (2019). 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

**Table 2 – (a) Comparison of the predicted deposition time to full coverage under various temperatures between the standalone kMC model and the multiscale data-driven model, and (b) additional temperature and inlet precursor flow rate predictions using the multiscale data-driven model.**

(a)

| 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 |

(b)

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

prediction for a thermal ALD process by 5% to 15% depending on the operating temperature, thereby significantly reducing half-cycle time.

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.

## 4. Conclusion

In this work, 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 $1.0 \times 10^{-3}$ 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.

## Acknowledgement

Financial support from the National Science Foundation is gratefully acknowledged.

## References

Acton, Q.A., 2012. Chemical Processes—Advances in Research and Application: 2012 Edition: ScholarlyBrief. ScholarlyEditions.

Bird, R.B., Stewart, W.E., Lightfoot, E.N., 2007. Transport Phenomena. John Wiley & Sons.

Burden, F., Dave, W., 2008. Bayesian regularization of neural networks. In: Artificial Neural Networks: Methods and Applications. Springer, pp. 23–42.

Chaffart, D., Ricardez-Sandoval, L.A., 2018. Optimization and control of a thin film growth process: a hybrid first principles/artificial neural network based multiscale modelling approach. Comput. Chem. Eng. 119, 465–479.

Chen, C.J., 1993. Introduction to Scanning Tunneling Microscopy, vol. 4. Oxford University Press on Demand.

Christofides, P.D., Armaou, A., Lou, Y., Varshney, A., 2008. Control and Optimization of Multiscale Process Systems. Springer Science & Business Media.

Cortright, R.D., Dumesic, J.A., 2001. Kinetics of heterogeneous catalytic reactions: analysis of reaction schemes. Adv. Catal. 46, 161–264.

Council, N.R., et al., 2003. Beyond the Molecular Frontier: Challenges for Chemistry and Chemical Engineering. National Academies Press.

Crose, M., Zhang, W., Tran, A., Christofides, P.D., 2018. Multiscale three-dimensional CFD modeling for PECVD of amorphous silicon thin films. Comput. Chem. Eng. 113, 184–195.

Curtiss, L.A., Redfern, P.C., Raghavachari, K., 2007. Gaussian-4 theory. J. Chem. Phys. 126, 084108.

Dalton, T.J., Conner, W.T., Sawin, H.H., 1994. Interferometric real-time measurement of uniformity for plasma etching. J. Electrochem. Soc. 141, 1893–1900.

Ding, Y., Zhang, Y., Kim, K., Tran, A., Wu, Z., Christofides, P.D., 2019. Microscopic modeling and optimal operation of thermal atomic layer deposition. Chem. Eng. Res. Des. 145, 159–172.

Djurabekova, F.G., Domingos, R., Cerchiara, G., Castin, N., Vincent, E., Malerba, L., 2007. Artificial intelligence applied to atomistic kinetic Monte Carlo simulations in Fe-Cu alloys. Nucl. Instrum. Methods Phys. Res., Sect. B: Beam Interact. Mater. Atoms 255, 8–12.

Elam, J., Groner, M., George, S., 2002. Viscous flow reactor with quartz crystal microbalance for thin film growth by atomic layer deposition. Rev. Sci. Instrum. 73, 2981–2987.

Elliott, S.D., Greer, J.C., 2004. Simulating the atomic layer deposition of alumina from first principles. J. Mater. Chem. 14, 3246–3250.

Fang, G., Xu, L., Ma, J., Li, A., 2016. Theoretical understanding of the reaction mechanism of $SiO_2$ atomic layer deposition. Chem. Mater. 28, 1247–1255.

Fermeglia, M., Pricl, S., 2009. Multiscale molecular modeling in nanostructured material design and process system engineering. Comput. Chem. Eng. 33, 1701–1710.

Fluent, A., 2013. ANSYS Fluent Theory Guide 15.0. ANSYS, Canonsburg, PA.

Foong, T.R.B., Shen, Y., Hu, X., Sellinger, A., 2010. Template-directed liquid ALD growth of $TiO_2$ nanotube arrays: properties and potential in photovoltaic devices. Adv. Funct. Mater. 20, 1390–1396.

Frisch, A., Nielsen, A.B., Holder, A.J., 2000. GaussView User Manual. Gaussian Inc., Pittsburgh, PA, pp. 556.

George, S.M., 2009. Atomic layer deposition: an overview. Chem. Rev. 110, 111–131.

George, S.M., Ott, A.W., Klaus, J.W., 1996. Surface chemistry for atomic layer growth. J. Phys. Chem. 100, 13121–13131.

Goldstein, J.I., Newbury, D.E., Michael, J.R., Ritchie, N.W., Scott, J.H.J., Joy, D.C., 2017. Scanning Electron Microscopy and X-ray Microanalysis. Springer.

Han, B., Zhang, Q., Wu, J., Han, B., Karwacki, E.J., Derecskei, A., Xiao, M., Lei, X., O'Neill, M.L., Cheng, H., 2011. On the mechanisms of $SiO_2$ thin-film growth by the full atomic layer deposition process using bis(t-butylamino)silane on the hydroxylated $SiO_2$ (0 0 1) surface. J. Phys. Chem. C 116, 947–952.

Hintermüller, M., Ulbrich, M., 2004. A mesh-independence result for semismooth Newton methods. Math. Program. 101, 151–184.
Huang, J., Hu, G., Orkoulas, G., Christofides, P.D., 2010a. Dependence of film surface roughness and slope on surface migration and lattice size in thin film deposition processes. Chem. Eng. Sci. 65, 6101–6111.
Huang, J., Hu, G., Orkoulas, G., Christofides, P.D., 2010b. Dynamics and lattice-size dependence of surface mean slope in thin-film deposition. Ind. Eng. Chem. Res. 50, 1219–1230.
Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv preprint 1502.03167.
Ishikawa, K., Karahashi, K., Ichiki, T., Chang, J.P., George, S.M., Kessels, W., Lee, H.J., Tinck, S., Um, J.H., Kinoshita, K., 2017. Progress and prospects in nanoscale dry processes: how can we control atomic layer reactions? Jpn. J. Appl. Phys. 56, 06HA02.
Kääriäinen, T., Cameron, D., Kääriäinen, M.L., Sherman, A., 2013. Atomic Layer Deposition: Principles, Characteristics, and Nanotechnology Applications. John Wiley & Sons.
Kimaev, G., Ricardez-Sandoval, L.A., 2019. Nonlinear model predictive control of a multiscale thin film deposition process using artificial neural networks. Chem. Eng. Sci. 207, 1230–1245.
King, S.W., 2011. Plasma enhanced atomic layer deposition of $SiN_x$:H and $SiO_2$. J. Vac. Sci. Technol. A: Vac. Surf. Films 29, 041501.
Knoops, H.C.M., Langereis, E., Van De Sanden, M.C.M., Kessels, W.M.M., 2010. Conformality of plasma-assisted ALD: physical processes and modeling. J. Electrochem. Soc. 157, G241–G249.
Lee, C., Yang, W., Parr, R.G., 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B 37, 785.
Liu, Y., Liu, H., 2018. Development of 3D polymer DFT and its application to molecular transport through a surfactant-covered interface. AIChE J. 64, 238–249.
MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4, 415–447.
Mankad, V., Jhu, P.K., 2016. First-principles study of water adsorption on $\alpha$-$SiO_2$ (1 1 0) surface. AIP Adv. 6, 085001.
Moré, J.J., 1978. The Levenberg–Marquardt algorithm: implementation and theory. In: Numerical Analysis. Springer, pp. 105–116.
Moura, C.A.d., Kubrusly, C.S., 2012. The Courant-Friedrichs-Lewy (CFL) Condition: 80 Years After its Discovery. Birkh”auser Basel.
Nicolas, C., Lorenzo, M., 2010. 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. J. Chem. Phys. 132, 074507.
Ochterski, J.W., 2000. Thermochemistry in Gaussian. Gaussian Inc., pp. 1–19.
Pan, D., Li, T., Chien Jen, T., Yuan, C., 2014. Numerical modeling of carrier gas flow in atomic layer deposition vacuum reactor: a comparative study of lattice Boltzmann models. J. Vac. Sci. Technol. A: Vac. Surf. Films 32, 01A110.
Pittal, S., Snyder, P.G., Ianno, N.J., 1993. Ellipsometry study of non-uniform lateral growth of ZnO thin films. Thin Solid Films 233, 286–288.
Raaijmakers, I.J., 2011. Current and future applications of ALD in micro-electronics. ECS Trans. 41, 3–17.
Schuegraf, K., Abraham, M.C., Brand, A., Naik, M., Thakur, R., 2013. Semiconductor logic technology innovation to achieve sub-10 nm manufacturing. IEEE J. Electron Devices Soc. 1, 66–75.
Shirazi, M., Elliott, S.D., 2014. Atomistic kinetic Monte Carlo study of atomic layer deposition derived from density functional theory. J. Comput. Chem. 35, 244–259.
Tanner, C.M., Perng, Y.C., Frewin, C., Saddow, S.E., Chang, J.P., 2007. Electrical performance of $Al_2O_3$ gate dielectric films deposited by atomic layer deposition on 4*H*-SiC. Appl. Phys. Lett. 91, 203510.
Wächter, A., Biegler, L.T., 2006. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Math. Program. 106, 25–57.
Weckman, T., Shirazi, M., Elliott, S.D., Laasonen, K., 2018. Kinetic Monte Carlo study of the atomic layer deposition of zinc oxide. J. Phys. Chem. C 122, 27044–27058.
Young, D.F., Munson, B.R., Okiishi, T.H., Huebsch, W.W., 2010. A Brief Introduction to Fluid Mechanics. John Wiley & Sons.
Zhang, Y., Ding, Y., Christofides, P.D., 2019. Multiscale computational fluid dynamics modeling of thermal atomic layer deposition with application to chamber design. Chem. Eng. Res. Des. 147, 529–544.