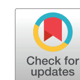
Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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



Microscopic modeling and optimal operation of thermal atomic layer deposition

Yangyao Ding [a], *Yichi Zhang* [a], *Keegan Kim* [a], *Anh Tran* [a], *Zhe Wu* [a], *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*

ARTICLE INFO

Article history:
Received 17 February 2019
Received in revised form 1 March 2019
Accepted 4 March 2019
Available online 20 March 2019

Keywords:
Atomic layer deposition
Microscopic modeling
Neural networks
Optimal operation
ALD cycle time optimization

ABSTRACT

This work develops a comprehensive framework for first-principles-based microscopic modeling, data-driven input/output modeling and optimal operation of thermal atomic layer deposition (ALD) of SiO_2 thin-films using bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. Specifically, we initially utilize Density Functional Theory (DFT)-based calculations for the computation of the key thermodynamic and kinetic parameters, which are then used in the microscopic modeling of the ALD process. Subsequently, a detailed microscopic model is constructed, accounting for the microscopic lattice structure and atomic interactions, as well as multiple microscopic film growth processes including physisorption, abstraction and competing chemical reaction pathways. Kinetic Monte-Carlo (kMC) algorithms are utilized to obtain computationally efficient microscopic model solutions while preserving model fidelity. The obtained kMC simulation results are used to train Artificial Neural Network (ANN)-based data-driven models that capture the relationship between operating process parameters and time to ALD cycle completion. Specifically, a dense two-hidden-layer feed-forward ANN is constructed to find a feasible range of ALD operating conditions accounting for industrial considerations, and a Bayesian Regularized ANN is constructed to implement the cycle-to-cycle optimization of ALD cycle time. Extensive simulation results demonstrate the effectiveness of the proposed approaches. The kMC models successfully achieves a growth per cycle (GPC) of 1.8 Å per cycle, which is in the range of reported experimental values. The ANN models accurately predict deposition time to steady-state from the given operating condition input, and the cycle time optimization using BRANN model reduces the conventional BTBAS cycle time by 60%.

© 2019 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

1. Introduction

Thin-film deposition is one of the most important building blocks in the semiconductor industry. Various deposition techniques, such as epitaxy, chemical vapor deposition (CVD), and physical vapor deposition (PVD), have been developed to deposit high quality thin-films of various materials, e.g., Al_2O_3, Hf_2O_3, RuO_2, SiO_2, etc. (Nalwa, 2002). However, the requirements in the production of advanced memory devices have become more and more demanding. For example, the dimensions of new high-k gate dielectrics are under transition to sub-10-nm scale and the associated film thickness is required to be under 30 Å (George et al., 1996; Schuegraf et al., 2013). Also, new transistor designs often involve complex three-dimensional structures rather than two-dimensional planar surfaces, along with the demand of conformal films

* *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.2019.03.004
0263-8762/© 2019 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

with a stringent criterion on uniformity and defects. Thus, the atomic layer deposition (ALD) process has been widely adopted by industry to meet the requirements of major design breakthroughs (Kääriäinen et al., 2013). ALD is a thin-film deposition method originally derived from CVD. In an ALD process, a substrate surface is exposed to alternating gas-phase precursor streams such that only one type of reactant is in contact with the substrate surface at each half-cycle. Once in contact, the precursor undergoes self-limiting surface reactions that allow a nearly complete and conformal surface coverage given sufficient exposure time and appropriate reactor conditions. In between the alternating precursor cycles, the reactor is purged with an inert gas, ensuring all previously-entered precursors are removed from the chamber prior to the exposure of the film to the next precursor, avoiding undesirable reactions and a decrease in film purity (George, 2009). The ALD method enables a layer-by-layer film growth with film uniformity at atomic level, which is more precise and controllable than the traditional CVD approach (Tanner et al., 2007; Foong et al., 2010; Shirazi and Elliott, 2014; Ishikawa et al., 2017). Therefore, in the field of microelectronics 3D integration, where ultra-thin and highly-conformal films are needed, ALD has gained significant popularity.

Currently, there is a wealth of ALD research on both laboratory and industrial scales (Kääriäinen et al., 2013). This significant research activity on ALD has led to the discovery of novel precursors and mechanisms which make high throughput film processing possible while allowing various substrate lay-outs (Dasgupta et al., 2016). However, experimental and industrial works on ALD remain expensive and time-consuming due to the cost of precursors and ALD-specific equipment, as well as due to the limited throughput (Shirazi and Elliott, 2014). Additionally, the real-time in-situ monitoring of film growth is not possible because molecular structure can only be understood through methods like scanning electron microscopy (SEM) and scanning tunneling microscope (STM), which are accurate but destructive to the deposited film (Schwille et al., 2017a). Thus, a model for ALD that provides insights on the details of real-time film profile and the overall growth rate can be beneficial to both industrial and research work.

Various deposition models have been developed over the past two decades in different fields of microscopic simulation, for example, crystal growth, CVD, and plasma-enhanced CVD (PECVD) (Nayhouse et al., 2013; Ikegawa and Kobayashi, 1989; Crose et al., 2018). In particular, Molecular Dynamics (MD), and more recently, kinetic Monte-Carlo (kMC) are among the most popular simulation methods (Elliott and Greer, 2004; Rasoulian and Ricardez-Sandoval, 2014; Crose et al., 2018). An ab initio MD model keeps track of all the particle movements and requires an overwhelming amount of computational resources, making it impossible to perform a simulation on an industrial scale process (Battaile and Srolovitz, 2002). However, the kMC method has a crucial advantage in computational efficiency as it tracks a single event at a time in a predefined lattice space. Despite this simplification, the kMC method has been used in deposition models to successfully reproduce realistic profiles (Rey et al., 1991; Dkhissi et al., 2008; Crose et al., 2018). Recently, Crose et al. (2018) proposed a novel multiscale computational fluid dynamics (CFD) simulation that used a surface microscopic n-fold hybrid kMC model and demonstrated its validity with a PECVD system. Moreover, many groups have showed the validity of using kMC in ALD simulation. For instance, Knoops et al. (2010) used raw probabilities of reaction and recombination to construct a kMC model for general plasma-enhanced ALD. Shirazi and Elliott (2014) modeled a small scale Al_2O_3 ALD deposition using kMC based on first-principles analysis.

However, previous works have not investigated the simulation of thermal ALD of SiO_2 thin-film, which is an important material for gate oxides in MOSFET and MEMS devices, sacrificial layers, and conformal dielectric films in the front-end-of-line (FEOL) semiconductor wafer processing (Murray et al., 2014). These ultra thin-films contain only around ten atomic layers (around 20 Å) and are only viable to prepare with ALD method (Schuegraf et al., 2013). Current research has yet to propose a model that is scalable to the industrial size, which have proven to be useful in traditional thin-film deposition systems (Crose et al., 2018). Moreover, computationally efficient data-driven approaches, such as system identification, have not yet been applied to characterize the proposed ALD system (Lee and Lee, 2005; Djurabekova et al., 2007; Nicolas and Lorenzo, 2010; Rasoulian and Ricardez-Sandoval, 2015b; Mhaskar et al., 2018; Kimaev and Ricardez-Sandoval, 2017; Singh Sidhu et al., 2018). Such data-driven models may make on-line optimization, run to run (R2R) control and real-time control possible for ALD processes (Wang et al., 2009; Kwon et al., 2015a,b; Rasoulian and Ricardez-Sandoval, 2015a, 2016; Oh and Lee, 2016; Chaffart and Ricardez-Sandoval, 2017, 2018).

In this paper, we propose a kMC model to simulate the atomic layer deposition of SiO_2 thin-film on a 3D lattice. Bis(tertiary-butylamino)silane (BTBAS) and ozone are chosen as the precursors, and density functional theory (DFT) is used to obtain thermodynamic and kinetic parameters of the precursors that were not previously reported. Those DFT-calculated parameters such as the intermediate complex activation energies and pre-exponential coefficients crucially affect the microscopic kMC model event selection to reproduce realistic growth rates and structure. This model is also an extension of the previous 2D-lattice kMC models proposed, for example in Crose et al. (2018). Although there are many advantages in using a 2D simplification, including easy setup and computational efficiency, a 3D lattice is required to simulate ALD due to the importance of spatial influence between species. Moreover, the modeling of deposition onto high-aspect ratio (AR) features requires the analysis of edges and corners which is not possible with a 2D geometry (Schwille et al., 2017a). Therefore, adopting a 3D microscopic lattice structure also enables simulating ALD with high AR design in the future. To further improve model performance, we develop a 3D triangular lattice approximation of real crystal lattice while maintaining important structural characteristics. After building the kMC simulation and validating its performance with experimental results, we derive a data-driven model via machine learning techniques to predict the steady-state film growth behavior for cycle-to-cycle optimization. Although kMC simulation can provide information about film growth in real-time, it is computationally expensive to be implemented in a control scheme for a large-scale system such as an entire wafer. Therefore, it is useful to derive a data-driven model that can provide a closed-form solution and can capture key film growth characteristics. Due to the stochastic nature of kMC and the non-linearity involved in the reaction mechanisms, non-linear regression models are applied to capture the input–output relationship. Traditional algebraic input–output models such as the least-squares method are subjected to prediction error and over-fitting error. Therefore, Artificial Neural Networks (ANNs), a more robust and more systematic way

of parameter determination for non-linear problem, can be readily tailored to perform such tasks (Goodfellow et al., 2016). Specifically, a dense two-hidden-layer feed-forward ANN and a Bayesian Regularized ANN are implemented to find the feasible range of ALD operating conditions and to optimize ALD throughput cycle-to-cycle, respectively. The proposed kMC-model achieves a growth per cycle (GPC) rate of 1.8 Å, which lies in the experimentally reported range of 1.4–2.1 Å per cycle. Extensive simulation results demonstrate the validity of the proposed ANN approach in calculating optimal deposition times with respect to the operating parameters. The resulting model is demonstrated to reduce the industrial conventional cycle time by 60%. Furthermore, the modeling approach developed in this work can serve as a general guideline and be extended to the ALD of other thin-film materials using different precursors and operating conditions.

2. ALD process description and modeling

This work focuses on developing a microscopic model that describes the deposition of SiO_2 thin-film via thermal ALD, which captures the structural details, the reaction mechanisms and the growth rate of SiO_2 thin-films. In this section, the approximation of the 3D SiO_2 lattice is introduced and validated. Then, precursor selection is discussed based on experimental results and thermodynamic data. Subsequently, the reaction mechanism and associated kinetics are discussed in detail, including the DFT calculations of kinetic rate parameters and activation energies by Gaussian09/Gaussview software package (Frisch et al., 2000). Next, a hybrid *n*-fold model-specific kMC algorithm is developed to simulate the ALD process. Finally, we present the machine learning algorithms involved with the data-driven analysis on the relationship between operating conditions and cycle completion time.

2.1. Structural assumptions of deposited SiO_2

Our microscopic model aims to simulate the deposition of α-Quartz SiO_2, which crystallizes in the trigonal crystal system of space group $P3_121$ and has a local SiO_4 structure similar to tetrahedron. Although it may be tempting to use a true α-Quartz lattice structure, such a lattice structure would not be suitable for kMC implementation. The chemical nature of SiO_2 ALD process requires the consideration of multiple reaction pathways, structural geometry and defect generations. A true α-Quartz lattice kMC model would be conceptually complex and computationally challenging (Shirazi and Elliott, 2014). Thus, instead of a fully realistic 3D lattice, a 2D triangular model (i.e., each monolayer is off-shifted from the monolayer below it) adopted in our previous work by Crose et al. (2018) is extended to 3D as an approximation of the actual α-Quartz crystal structure. In our model, a bond angle of 90° is assumed for the connectivity between Si and O atom instead of 109.5°. As shown in Figs. 1 and 2, the top view of the simulated lattice closely resembles that of the real lattice, with some angle distortion. This assumption leads to a lattice repetition every four cycles instead of three cycles as appears in the α-Quartz SiO_2. For any silicon atom in the approximated lattice, another silicon atom appears directly above it every four cycles, whereas in a realistic lattice, such pattern repeats every three cycles. However, this simplification does not influence the connectivity of the individual lattice cell or the validity of the model. We will also later demonstrate that our simulation captures accurately the growth rate and defect generation pattern of α-Quartz SiO_2 deposition reported by experimental results. Thus, this 3D triangular model is a valid simplification of the true structure. A lattice size of 1200×1200 sites per layer is used, which is large enough for the simulation to be size-independent but still computationally efficient as demonstrated by Huang et al. (2010a,b), and the height depends on the number of cycles simulated.

2.2. Precursor selection

Surface reactions in the ALD process govern the growth rate and the structural pattern of SiO_2 films. Therefore, the selection of oxygen and silicon precursors is an important topic. In the past, many silicon precursors have been selected and studied to improve the uniformity and the growth rate of SiO_2 deposition. In recent years, aminosilane-based precursors have gained significant popularity because of the low activation energy of the sequential dissociative chemisorption mechanism caused by the H–N hydrogen bonds formed during the adsorption stage. These characteristics lead to high reaction rate and greatly improve the efficiency of SiO_2 deposition (O'Neill et al., 2011). Among those aminosilane precursors, the most popular ones are: bis(tertiary-butylamino)silane (BTBAS), bis(diethylamino)silane (BDEAS), bis(dimethylamino)-silane (BDMAS), tris(dimethylamino)silane (TDMAS) and di(sec-butylamino)silane (DSBAS) (Kamiyama et al., 2006; Kinoshita et al., 2007; Baek et al., 2012; Dingemans et al., 2012; Huang et al., 2013; Putkonen et al., 2014). In order to pick the most favorable precursor for our simulation, we account for the following factors: the existence of experimental data (e.g., growth per cycle (GPC) and precursor exposure time), the existence of theoretical data (e.g., reaction mechanism and associated kinetic parameters), and the availability of additional information such as film quality, sticking coefficient and steric hindrance studies. Based on the above considerations, BTBAS is chosen as the Si precursor due to its fast reported growth rate (1.4–2.1 Å per cycle), adequate experimental and theoretical data, and a detailed mechanism available to model the process (O'Neill et al., 2011; Han et al., 2011). With respect to the oxygen precursor, ozone (O_3) is chosen among the common candidates for thermal ALD of oxide films, because ozone is chemically reactive and does not introduce hydrogen-involved side products in the thermal ALD process. Furthermore, ozone is extensively used in the industry and is widely studied in experiments, which makes its major chemical properties and reaction mechanisms readily accessible (Nishiguchi et al., 2002; Prechtl et al., 2003; Han et al., 2011).

2.3. Reaction mechanism

A full deposition cycle in the ALD process consists of two half-cycles, each using a specific precursor species to introduce the desired element onto the film. As mentioned above, we choose BTBAS and ozone as the precursors for SiO_2 deposition simulation. The reaction mechanism using these two precursors was reported by Han et al. (2011) and is explained in detail below.

The first half-cycle is referred to as the Si-Cycle, which contains physisorption, abstraction and a two-step dissociative chemisorption. In our model, we picked a fully hydroxylated $SiO_2(001)$ surface as our starting point, shown in Fig. 1. The sil-

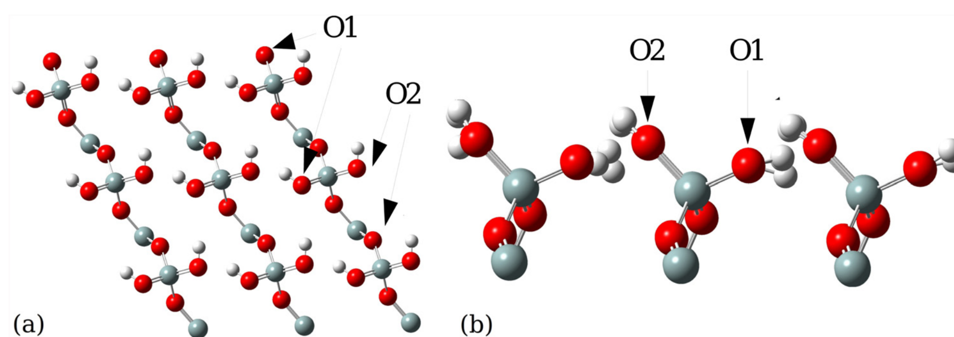


Fig. 1 – (a) Top view of the hydroxylated $SiO_2(001)$ surface. (b) Side view of the hydroxylated $SiO_2(001)$ surface, where O1 is the more electronegative oxygen. The double bonds are due to Gaussian display format, which does not influence the validity of the structure.

	0	1	2	3	4	5	6	7	8	9	10	11
0	Si	O2H	[ ]	O1	Si!	O2	[ ]	CsP!	Si	O2H	[ ]	PsP!
1	[ ]	[ ]	[ ]	[ ]	H2	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
2	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
3	[ ]	[ ]	H1	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	H1	[ ]
4	Si	O2	Si	O1	Si	O2H	[ ]	O1H	Si	O2	Si	O1
5	[ ]	[ ]	PO2!	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	H2	[ ]
6	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
7	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	O1H	[ ]	[ ]	[ ]	[ ]	[ ]
8	Si	O2H	[ ]	[ ]	Si	O2	Si	O1	Si	O2H	[ ]	O1H
9	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	H2	[ ]	[ ]	[ ]	[ ]	[ ]
10	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
11	[ ]	[ ]	[ ]	[ ]	H1	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Fig. 2 – Top view of a 5-layer 12 × 12-site miniature demonstration of the full kinetic Monte-Carlo simulation lattice. The five layers and the species on the lattice are shown using different colors and symbols, respectively. The first (bottom) layer, labeled red, contains the base Si atoms. The second layer, labeled black, contains oxygen atoms or hydrogenated oxygens. The third layer, labeled yellow, contains the species from the first silicon half-cycle: Si is the neighbor-binding silicon, Si! is the self-binding silicon, and PsP and CsP are the physisorbed and chemisorbed precursors, respectively. The fourth layer, labeled green, contains the species from the first oxygen half-cycle: O and OH are the oxygen atoms and hydrogenated oxygens. The fifth (top) layer, labeled blue, contains physisorbed ozones (PO1 and PO2), which remain to be oxidized. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

icon precursor, BTBAS, is first physisorbed onto the substrate surface under specific temperature and pressure. According to Han et al. (2011), the two oxygen atoms in a SiO_2 cell have different electronegativities. The more electronegative oxygen atom, denoted as O^1, is more reactive and is therefore more likely to be electrophilicly attacked by precursor particles than the less electronegative oxygen atom, denoted as O^2. Therefore, as shown in Fig. 3(a), the precursor particle is first physisorbed onto the O^1-type hydroxyl group through a strong H-bond to form the reactant. Then, the physisorbed precursor goes through the first dissociative chemisorption step, forming a monoamine intermediate and releasing one of the two aminoethyl groups. Next, the remaining aminoethyl group electrophilicly attacks an adjacent O^2-type hydroxyl group, which can be either from the neighbor Si atom, i.e., neighbor-binding route, or from the same substrate Si atom, i.e., self-binding route, as shown in Fig. 3(b). The former reaction pathway retains the original surface orientation, resulting in a thermodynamically favorable structure, whereas the latter, which is more kinetically favorable as shown in Table 1, causes a deviation from the (001) surface orientation and leads to defect formation. After the electrophilic attack, the other aminoethyl group is released from the surface structure and another O–Si bond is formed. The remaining two H atoms from the Si atom then become the new substrate surface. The competition of kinetic and thermodynamic favorability is crucial in explaining the structural non-uniformity of SiO_2. Therefore, both reaction pathways and their reverse reactions are incorporated in our kMC model, and the reaction kinetics will be explained in more details in the next section.

The second half-cycle is referred to as the O-Cycle, which contains the ozone physisorption, abstraction and surface

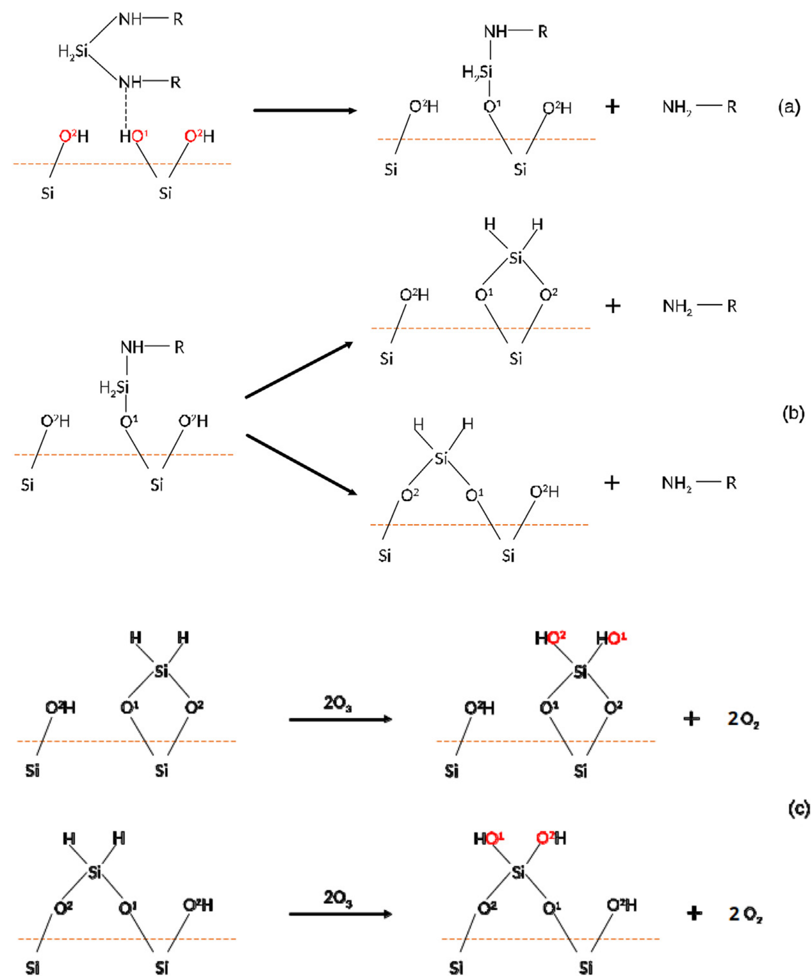


Fig. 3 – (a) First dissociative chemisorption step of BTBAS. (b) Second dissociative chemisorption step of BTBAS under self-binding and neighbor-binding mechanisms. (c) Oxidation of self-binding and neighbor-binding SiH_2 with ozone.

Table 1 – Activation energies and partition function ratios of reactions.

Reaction	Activation energy (kcal/mole)	Vibrational partition function ratio
$r_{si,chem}$	8.9	1
$r_{si,neigh,f}$	20.1	1
$r_{si,neigh,r}$	33.6	1
$r_{si,self,f}$	16.1	1
$r_{si,self,r}$	14.4	1
$r_{si,des}$	17.5	9.56e−8
$r_{oa,f}$	17.7	1
$r_{o_b,f}$	15.4	1
$r_{o,des}$	9.224	1e−4

oxidation. The oxidation steps of self-binding and neighbor-binding H–Si groups are shown in Fig. 3(c). Once the surface is partially/fully chemisorbed, both terminating H atoms will be oxidized by ozone to hydroxyl group (–OH), which are utilized in the next Si-Cycle.

2.4. Relative rate determination

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

$$r_{phs} = \frac{p_i}{RT}\sqrt{\frac{8k_bT}{\pi m_i}}s_{c,i}N_a\sigma \tag{1}$$

where r_{phs} is the physisorption reaction rate, p_i is the partial pressure of the species i, R is the ideal gas constant, T is the temperature, k_b is the Boltzmann constant, m_i is the molecular weight of species i, $s_{c,i}$ is the sticking coefficient of the species i at given temperature, N_a is the Avogadro number, and σ is the average area per surface site. Although the sticking coefficient of BTBAS is not reported in previous works, we obtain its value through an analogy with the sticking coefficient of BDEAS because of structural and electronic similarity (Schwille et al., 2017b).

On the contrary, chemisorption, abstraction and oxidation are thermodynamically activated kinetic reactions, which are generally described by the Transition State Theory (TST) (Cortright and Dumesic, 2001). Assuming quasi-equilibrium is achieved between the complex and the reactant, the reaction rate can be estimated using the thermodynamic properties of the transition state complexes, which are computed using DFT. Thus, the reaction rate equation can be formulated with standard Arrhenius-type rate law as follows:

$$r_{rxn,i} = A_i \exp\left(\frac{-E_{a,i}}{k_bT}\right) \tag{2}$$

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

$$A_i = f_i^{TST}\left(\frac{k_b T}{h}\right) \quad (3)$$

where h is the Planck's constant, k_b is the Boltzmann constant, T is the temperature, and f_i^{TST} is the ratio of the vibrational partition function between the transition state complex and the reactant, calculated with DFT. In this work, all DFT calculations are performed using the Gaussian09 software, which will be illustrated in more details below. The resulting parameters are summarized in Table 1 and the associated nomenclature is explained in Section 2.5. Since the desorption reactions lead to gas-phase products, the DFT-calculated vibrational partition function ratios of those reactions are small than one, which match the results reported in literature. On the contrary, the other reactions are entirely surface reactions. Therefore, their vibrational partition function ratios all equal to one (Cortright and Dumesic, 2001).

2.5. *Kinetic Monte-Carlo algorithm*

As mentioned in Section 1, a first-principles Molecular Dynamic simulation is too computationally demanding to be feasible for the scale of system discussed in this work (Battaile and Srolovitz, 2002; Rey et al., 1991; Dkhissi et al., 2008). Thus, we adopt an n-fold hybrid kMC algorithm in the framework proposed by earlier works (Lou and Christofides, 2004; Christofides et al., 2008; Crose et al., 2018). kMC is a stochastic algorithm that uses the kinetic rate information and uniformly distributed random numbers to determine event execution and system time evolution. Specifically, we define an event set as a collection of all events that have comparable rates. A total rate, r_{total}, is defined as:

$$r_{total} = \sum_{i=1}^{N} r_i \quad (4)$$

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

$$l_i = \frac{\sum_{j=1}^{i} r_j}{r_{total}}, \quad i = 1, \ldots, N \quad (5)$$

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

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

$$\Delta t = \frac{-\ln \gamma_2}{r_{total}} \quad (6)$$

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

$$r_{o,total} = r_{o,phs} + r_{o,des} + r_{o_a,f} + r_{o_b,f} \quad (7)$$

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

$$r_{si,total} = r_{btbas,phs} + r_{btbas,des} + r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} \quad (8)$$

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

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

$$r_{si,rxn} = r_{neigh,f} + r_{neigh,r} + r_{self,f} + r_{self,r} + r_{btbas,des} \quad (9)$$

$$r_{si,ads} = r_{btbas,phs} \quad (10)$$

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

$$J_{si,ads} = \frac{r_{si,ads}}{r_{si,total}} = 1 - J_{si,rxn} \quad (11)$$

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

pathways and directions is calculated by the concentration-weighted reaction rates as follows:

$$l_{si,i} = \frac{\sum_{j=1}^{i} r_{rxn,j} R_j}{\sum_{k=1}^{N} r_{rxn,k} R_k}, \quad i = 1, \ldots, N \tag{12}$$

where $l_{si,i} \in (0, 1]$ represents the normalized indicator of the ith event in the surface reaction event set, $r_{rxn,j}$ is the un-weighted chemical reaction rate for the jth event calculated from Eq. (2), R_j is the number of reactants for each surface reaction, and N is the total number of events in the Si-Cycle surface reaction event set. The normalized indicators are then used to execute the event selection following the same approach performed in the standard kMC algorithm. In Section 3, it is demonstrated that this decoupling scheme achieves desired accuracy.

2.6. DFT and thermodynamic calculations

Although the reaction activation energies and mechanisms have already been analyzed for BTBAS by Han et al. (2011) as discussed in Section 2.3, many fundamental thermodynamic and kinetic properties of BTBAS have yet to be investigated, including its entropy, enthalpy, vibrational partition and others. Since the above properties are difficult to measure experimentally yet essential to the accurate microscopic simulation of ALD behavior, in this work, we utilize Density Functional Theory (DFT) with Gaussian09 package to compute them (Frisch et al., 2000).

In the Si-Cycle, to calculate properties of BTBAS and its reaction kinetics with SiO_2 lattice, we first need to investigate the configuration of the hydroxylated surface lattice and the structure of the physisorbed BTBAS transition state complex. Specifically, to construct an optimal surface lattice, a generic bulk α-quartz SiO_2 unit cell is modified to generate a desired hydroxylated surface layer. The bulk unit cell is first imported into VESTA 3, which is a 3D visualization program widely adopted to construct crystalline structures (Momma and Izumi, 2011). The uppermost layer of Si atoms is removed, leaving two single bonded oxygen atoms per unit cell. Each oxygen atom is terminated with one hydrogen atom, and the new O–H bond is assumed to have the typical bond angle and bond length of 0.98 Å (Li et al., 2009). Then, the hydroxylated unit cell is imported to the Gaussview molecule builder tool. A $3 \times 3 \times 1$ SiO_2 lattice is constructed using the hydroxylated SiO_2 unit cell with Gaussian Periodic Boundary Condition (PBC) cell symmetry replication (Frisch et al., 2000; Mankad and Jhu, 2016). A series of optimization steps is carried out, with all atoms other than the surface hydrogen and oxygen atoms fixed during structure optimization (Han et al., 2011). The lattice structure is first optimized using Hartree–Fock (HF) method with basis set 3-21G to obtain an initial guess of the structure. Next, the B3LYP method, a hybrid Becke's three-parameter exchange functional (B3) with Lee–Yang–Parr gradient correction functional (LYP) and triple valence plus polarization, is applied to optimize the structure to an acceptable energy minimum with basis set 6-31G+dp accuracy level (Lee et al., 1988; Becke, 1993). Subsequently, an initial guess of the TS complex is obtained by structuring together an optimized BTBAS molecule and a $3 \times 3 \times 1$ SiO_2 surface lattice. Then, the optimization to transition state keyword (opt=TS) in Gaussian09 is used to calculate an optimized TS complex structure that is most energetically favorable. The calculation is carried out with the same level basis set accuracy as surface lattice structure optimization, plus modified coordinate definition and force constant calculation of every atomic position. The resulting TS structure has one negative vibrational frequency as expected (Frisch et al., 2000). Similarly, for the O-Cycle reaction, the thermodynamic and kinetic properties of gas-phase ozone molecule, H-Si surface lattice and physisorbed ozone are investigated. The same calculation procedures as Si-Cycle are carried out for the O-Cycle, respectively.

Finally, in order to perform a precise vibrational frequency calculation, Gaussian-4 (G4) theory is adopted. G4 theory is a complex computation method for accurate calculation of molecular energy based on ab initio molecular-orbital theory. It provides thermodynamic results for compounds containing second row (Li-F) and third row (Na-Cl) elements, which is applicable to our reacting molecules. Parallel computational work with Linda worker from Gaussian09 package is carried out to find out all the vibrational frequencies to calculate the ratio f^{TST} (Frisch et al., 2000). The final vibrational partition function along with other important thermodynamic properties are summarized in Table 1.

2.7. Artificial neural network model and non-linear regression

The half-cycle time plays an important role in both industrial production and experimental studies of ALD, yet the actual time needed for each half-cycle for various experimental conditions of temperature and pressure remains unknown. Specifically, according to kMC simulation results and experimental analysis in SiO_2 thin-film ALD, Si-Cycles require longer time than O-Cycles at high temperature around 600 K, which impacts film coverage and quality (Han et al., 2011). Therefore, it is important to develop a model that can estimate the required Si-Cycle time, given operating temperature and pressure. Although the kMC model can be used to simulate the transient behavior of ALD and provide a reference to cycle time, it is computationally demanding to be applied for multiple-cycle film production in real-time. Moreover, kMC model is not a closed-form model, and thus, cannot be directly utilized for optimization and control. Therefore, instead of using the kMC model to perform real-time optimization, we take advantage of the data-driven modeling approach and build Artificial Neural Network (ANN) models that correlate the Si-Cycle completion time with operating inlet temperature and pressure, using kMC-generated databases. Compared with traditional input–output models such as the least-squares method, the ANN approach is chosen for its advantages to lowering prediction error and its robustness against over-fitting error. Since it takes a long time to generate a database from the kMC model solutions, especially at low temperature and pressure, two ANN-based models are developed to serve different levels of precision. The first database covers a wider but sparse range of operating conditions, aiming to predict the suitable boundary for operating conditions and to provide a general reference of cycle completion time with acceptable accuracy. However, a higher accuracy is necessary for the real-time control and cycle-time optimization. Therefore, the second database is developed, which focuses on a smaller range with higher resolution. Due to the difference in data ranges, we adopt two levels of regularization to accurately train our neural networks: (1) The standard un-regularzied feed-forward neural network for the feasible range, and (2)

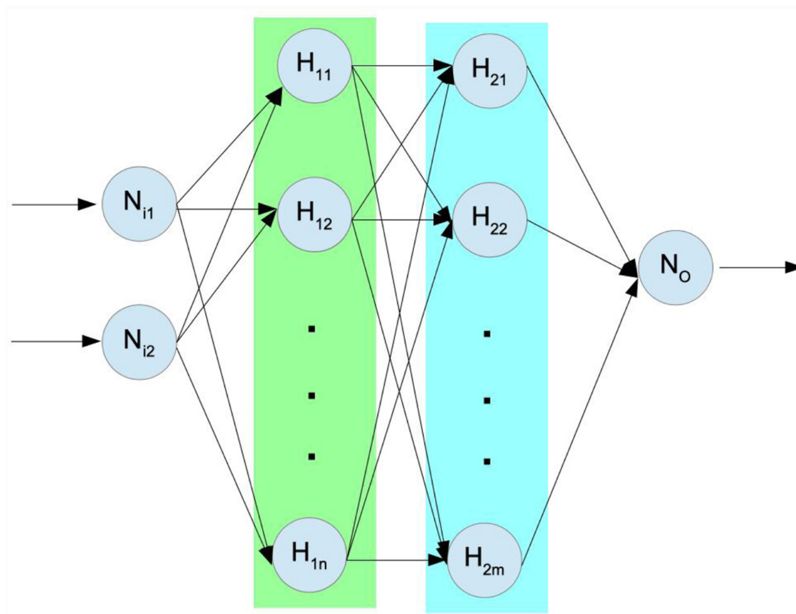


Fig. 4 – Feed-forward Artificial Neural Network with two inputs, two hidden layers, and one output.

The Bayesian Regularized Artificial Neural Network (BRANN) for the optimal range.

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

$$ReLu(x) = x^{+} = \max(0, x) \tag{13}$$

Additionally, all layers are densely connected, and the structure of the ANN is optimized via a grid search (Svozil et al., 1997). The general structure of a feed-forward two-input-single-output neural network with two hidden layers is given in Fig. 4, where N_{ij} are the input neurons, H_{ij} are the hidden layer neurons, and N_o is the output neuron. The above ANN structure is constructed with Tensorflow's keras module, a high-level application programming interface (API) designed to build and train a deep-learning ANN model which is widely used in deep learning applications.

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

$$S(w) = \frac{1}{N_D}\sum_{i=1}^{N_D}[y_i - (f(x_i, w))]^2 \tag{14}$$

where N_D is the number of data samples in the training dataset, y_i is the desired output value, w is the weight vector for all hidden layers, and $f(x_i, w)$ is the predicted value dependent on input x_i and weight w. The proper weight vector is obtained by solving an optimization problem to minimize the cost function $S(w)$ using standard back-propagation. Batch normalization is applied after each hidden layer to avoid saturation and high variance activation, thereby facilitating convergence speed and learning rate (Ioffe and Szegedy, 2015). Dropout regularization layers with a rate coefficient of 0.5, which is the typical value for hidden layers, is used to perform model averaging with bagging method, which enhances the generality of the network and reduces over-fitting. The RMSProp optimizer is adopted for model training, which utilizes normalized gradient from recent iterations by keeping a moving average of the squared gradient for each weight using the following equation:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma)g_t^2 \tag{15}$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}}g_t \tag{16}$$

where w_t and g_t are the weight parameter and its gradient at iteration t respectively, $E[g^2]_t$ denotes the running average of g^2 at iteration t, $E[g^2]_{t-1}$ denotes the running average of g^2 at iteration $t-1$, and ϵ is a smoothing term that avoids division by zero. A learning rate $\eta = 0.01$ and a momentum factor $\gamma = 0.9$ are used, which are the recommended values for RMSprop method (Ruder, 2016).

For the second database, our objective is to identify an ANN model that captures the precise input–output relationship over a smaller operating range for real-time cycle-to-cycle optimization. However, the traditional RMSprop mechanism faces a dilemma between model accuracy and over-fitting. Thus BRANN is introduced as an alternative solution, by adding Bayesian regularization to the standard ANN. BRANN is more robust than standard neural networks for precise regression because the Bayesian regularization algorithm converts complex non-linear regression into a rigid regression, which is a well-posed statistical problem. By efficiently turning off the weights that are not relevant in the training process and incorporating Occam's razor principle, BRANN avoids the over-fitting and over-training problems by optimally penalizing excessive complexities in models (Burden and Dave, 2008). In our model, BRANN for the second database is constructed, trained, and implemented using MATLAB machine learning

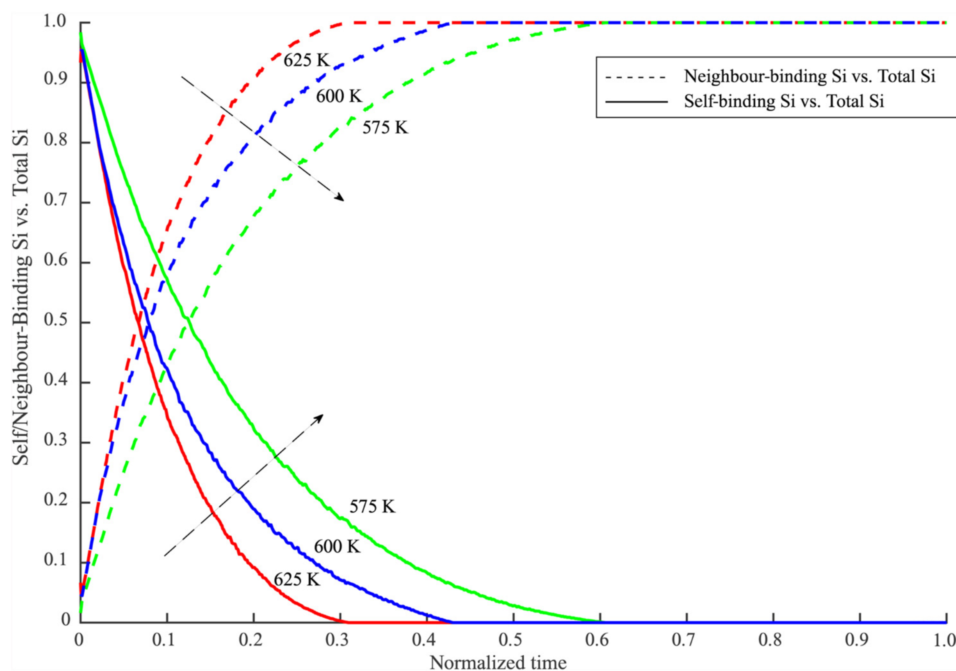


Fig. 5 – Competition between self-binding and neighbor-binding silicons with respect to the normalized cycle time. The dashed lines represent the ratio of neighbor-binding silicons versus total built silicons, and the solid lines represent the ratio of self-binding silicons versus total built silicons. The dashed arrows represent the direction of decreasing temperature.

package. Specifically, the input layer and the output layer of the BRANN are constructed using the same approach as the standard ANN above. For the inner structure of the BRANN, one hidden layer with 25 neurons is constructed. The hyperbolic tangent sigmoid function (*tansig*) is used as the activation function for the BRANN.

$$tansig(x) = \frac{2}{1+e^{-2x}} - 1 \qquad (17)$$

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

$$S(w) = \beta \sum_{i=1}^{N_D} [y_i - (f(x_i, w))]^2 + \alpha \sum_{j=1}^{N_w} w_j^2 \qquad (18)$$

where N_D is the number of data samples in the training dataset, w is the weight vector for all hidden layers, which is assumed to have a Gaussian distribution, N_w is the number of weight parameters, w_j is the jth entry in the weight vector, y_i is the desired output value, and $f(x_i, w)$ is the predicted value dependent on w and the input x_i. To compute the optimal weight vector w and the continuously updated hyperparameters α and β, a sequence of optimization problems are solved using the Levenberg–Marquardt algorithm (Moré, 1978). The Bayesian inference calculation and the construction of the optimization problems are discussed in detail by MacKay (1992) and Burden and Dave (2008).

Finally, both databases are generated based on the kMC model using the UCLA Hoffman2 Distributed Cluster. The two ranges of operating conditions are: (1) T = 550–700 K (ΔT = 5 K) with P = 80–60 Pa (ΔP = 2 Pa), and (2) T = 590–610 K (ΔT = 0.5 K) with P = 120–150 Pa (ΔP = 1 Pa). The first set of simulation results is divided into training, validation and test data under the ratio of 8:1:1, and the second set of simulation results is divided into training, validation and test data under the ratio of 7:1.5:1.5. The training dataset is used to determine the model parameters. The validation dataset is used in the training process to validate and improve the training. The testing dataset is randomly chosen from the entire dataset in advance to evaluate the final result and is not used in the training process.

3. Simulation results

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

3.1. *Validation of microscopic kMC model with experimental data*

The kMC model is validated by observing its behavior under varying temperature for a total of ten ALD cycles. Specifically, the precursor partial pressure is kept constant at 133 Pa, and the operating temperature is varied from 555 K to 625 K with an increment of 20 K or 30 K for each simulation run. Each half-cycle is assigned 2 s, which is sufficient for the cycle to reach steady-state as observed. The same amount of cycle

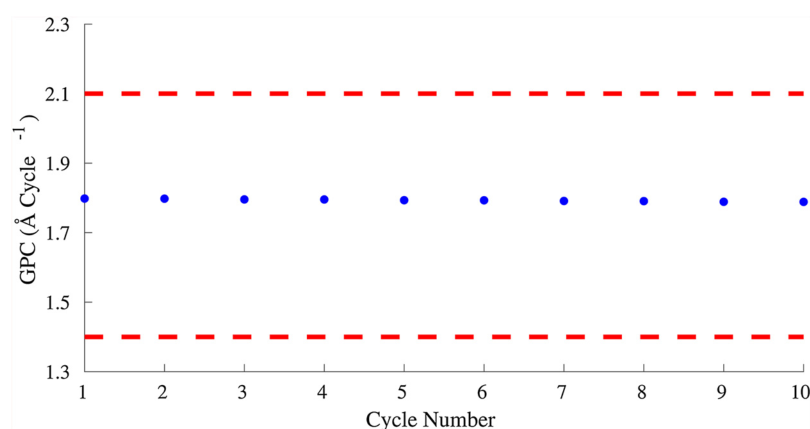


Fig. 6 – Steady-state GPC for the first ten cycles at 600 K and 133 Pa, where the dashed lines represent the upper and lower GPC limits reported in the literature (Putkonen et al., 2014).

time is reported in experimental work by O'Neill et al. (2011), although the specific value may vary since it also accounted for the time of gas-phase development. The precursor partial pressure on the substrate surface is important in determining physiosorption reaction rate. Therefore, the coupled effect of the gas-phase mass and momentum transfer in the reactor has a significant contribution to the cycle time and film quality, which we will discuss in our future work on the investigation of multi-scale computational fluid dynamics modeling of thermal ALD. Additionally, as mentioned in Section 2.7, since the O-Cycle time is much faster than Si-Cycle, we will focus on the Si-Cycle results only.

During the Si-Cycle, the competition between neighbor-binding and self-binding pathways is observed. As discussed in Section 2.3, self-binding events are kinetically favorable, while neighbor-binding events are thermodynamically favorable. The designed kinetics in our model is able to capture this behavior, as shown in Fig. 5. At the beginning of each cycle, self-binding events dominate. As the simulation time goes on, self-binding silicons undergo reverse reaction and decrease, as shown by the solid lines, and neighbor-binding silicons start to form, as shown by the dashed lines. If a cycle is given enough time to develop, the fraction of neighbor-binding silicons will approach unity, while the fraction of self-binding silicons approaches zero. Since the neighbor-binding dominance is thermodynamically driven, as temperature decreases, more self-binding silicons appear in the initial deposition stages and a longer time is required for the lattice to develop into the desirable neighbor-binding-dominant profile, shown by the direction of the dashed arrow in Fig. 5. This observation is consistent with the calculation and experimental analysis reported in Han et al. (2011).

ALD growth rate is also successfully simulated by the kMC model, which is characterized by the average growth of film thickness per cycle (GPC). Although the kMC model does not report the GPC directly as in experimental approaches, we could use the final coverage information to compute the GPC based on the literature values of SiO_2 unit cell lattice constants: $a = 5.5407$ Å, $b = c = 4.918$ Å (El-Kareh, 2012). An atom-to-atom measurement is performed in Gaussview to find the relative distance between Si and O atom within a SiO_2 cell, and the ideal layer thickness is calculated using the lattice constants and the atomic radii of oxygen and silicon atoms, which are 0.65 Å and 1.18 Å, respectively (El-Kareh, 2012). Then, the GPC is inferred through the simulated surface coverage and the ideal layer thickness. The simulated

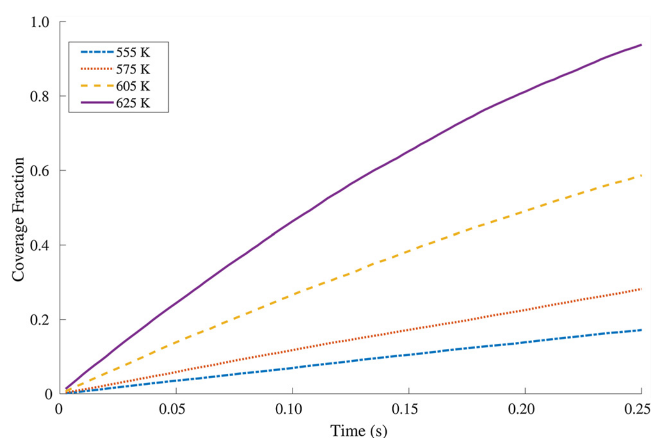


Fig. 7 – Comparison of initial transient deposition profiles for different temperatures at 133 Pa.

GPC at standard industrial operating condition of 600 K and 133 Pa is shown in Fig. 6. The GPC varies little with increasing cycle numbers but does show a slightly decreasing pattern as reported in experimental works (George et al., 1996). The average growth rate over 10 cycles is 1.8 Å per cycle, which is in the range of the SiO_2 GPC of 1.4-2.1 Å per cycle reported by O'Neill et al. (2011).

Although the temperature does not impact GPC when sufficient time is given for each cycle to reach steady-state and achieve full coverage, it is a crucial factor for the transient deposition rate within each cycle. Fig. 7 demonstrates that the transient deposition rate increases as temperature increases, and the rate increment is approximately proportional to the temperature increment. This effect is observed in experimental results from the work by Putkonen et al. (2014), where the deposition rate is demonstrated to increase with temperature for unsaturated surfaces. Moreover, it is noteworthy that not all temperatures allow the surface to reach full coverage at steady-state. The selection of an appropriate temperature region will be introduced in the next section.

3.2. *ANN results for Si-Cycle*

3.2.1. *Feasible operating regime*

In this section, a set of kMC simulations covering a wide range of operating conditions is carried out under fixed temperature and pressure throughout each simulation. The simulation is terminated either when steady-state is achieved under the

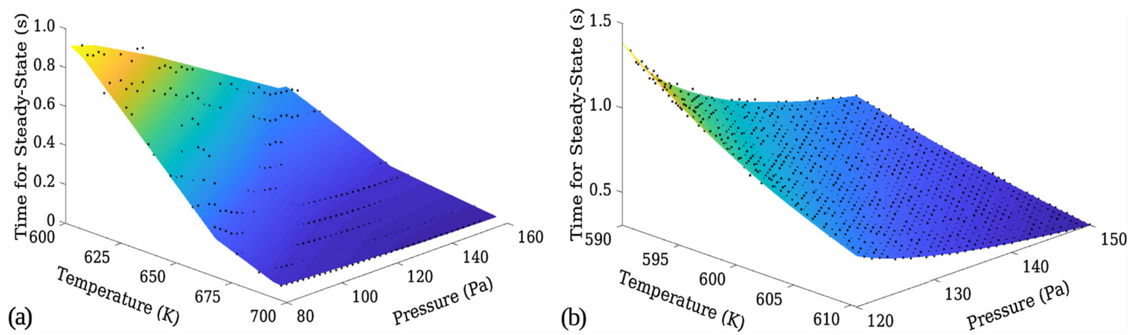


Fig. 8 – Dependence of the time to reach steady-state on the operating conditions, where black markers (dots) represent the training data and the surface represents the fitting result: (a) Large-range operating condition fitting. (b) Small-range operating condition fitting.

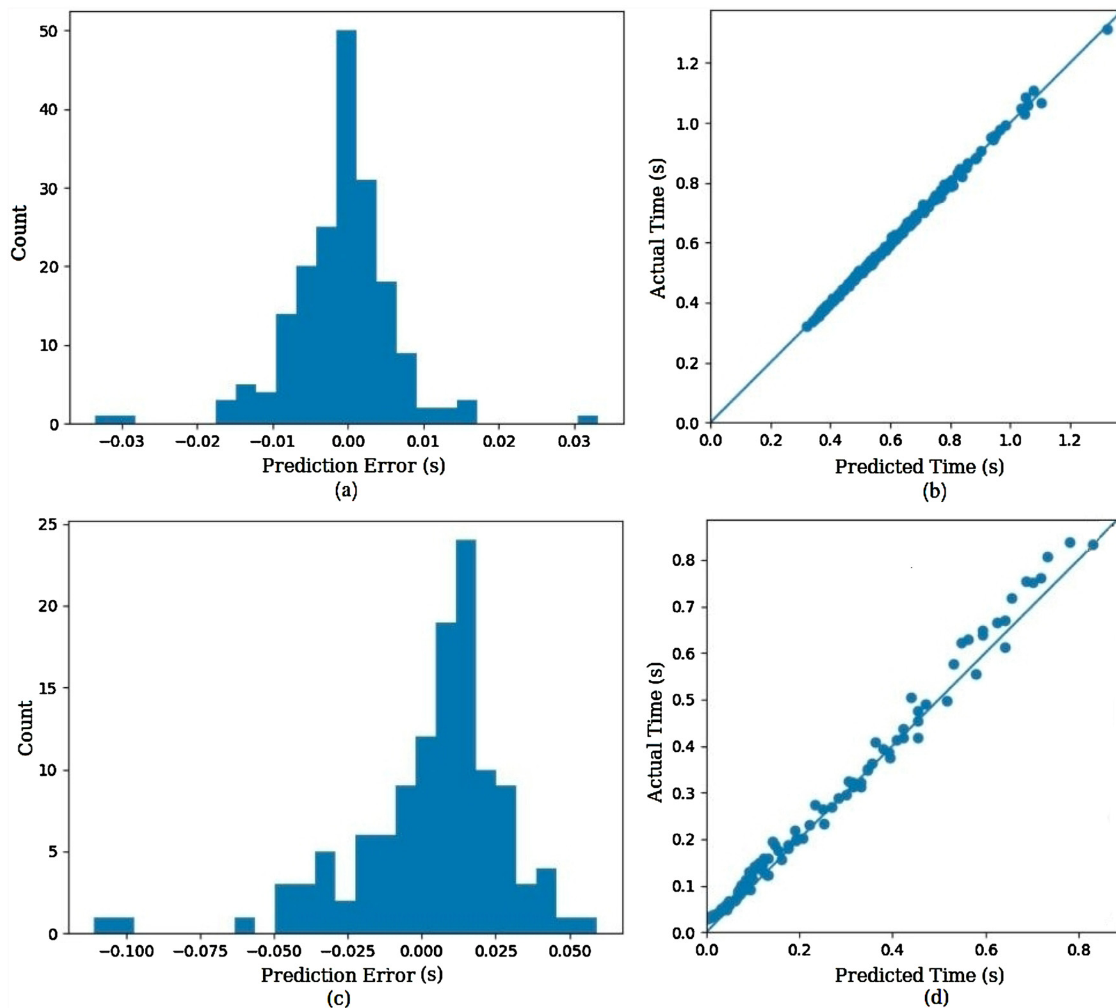


Fig. 9 – Performance of the Artificial Neural Network models: (a) Prediction error distribution histogram for the large-range operating conditions. (b) Correlation accuracy of the predicted time and the actual time for the large-range operating conditions, where the x-axis is the predicted time to reach steady-state from the neural network and the y-axis is the actual time to reach steady-state from the kMC model. (c) Prediction error distribution histogram for the small-range operating conditions. (d) Correlation accuracy of the predicted time and the actual time for the small-range operating conditions, where the x-axis is the predicted time to reach steady-state from the neural network and the y-axis is the actual time to reach steady-state from the kMC model.

given conditions, or when the simulation time exceeds 5 s, which is too long to be considered industrially relevant (Acton, 2012). Due to the stochastic nature of the kMC algorithm, the lattice surface configuration keeps changing at steady-state. However, the overall coverage at steady-state, which is one of the most crucial attributes of ALD processes, will be maintained at a certain value, with fluctuation under 0.5%. Then, the ANN model is developed to capture the relationship between the time to achieve 98% of the final coverage and the operating temperature and pressure. It is noted that, in the ANN model, the time for the system to reach 98% of the final coverage is used instead of the time to reach the final coverage to reduce the noise and inaccuracy involved in the steady-state fluctuations.

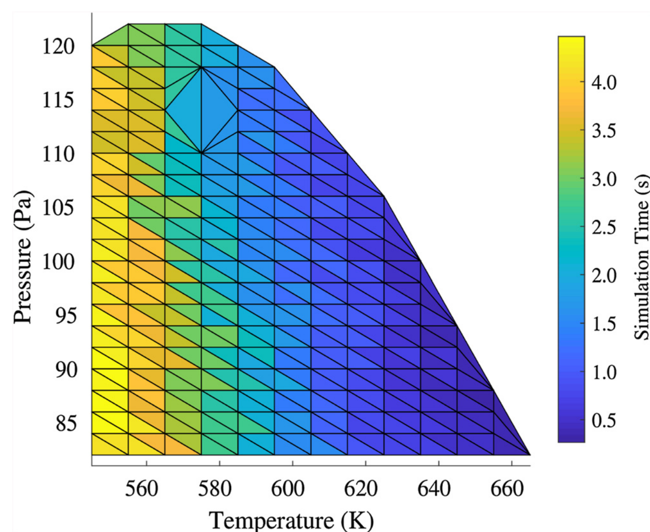


Fig. 10 – Range of operating conditions where the deposited films are not able to reach full coverage at steady-state.

Two hidden layers are used for the ANN model and the numbers of neurons are determined to be 35 and 30 for the first and second hidden layer respectively, based on a grid search. A single-hidden-layer feed-forward ANN does not yield a good solution, and the performance cannot be improved by simply adding more neurons since over-fitting error is observed. Therefore, a two-hidden-layer structure is chosen, which is conventionally adopted to capture the exponential-like behavior in our model. A mean absolute error on the test dataset is reported to be 8.00×10^{-3} s. As shown in Fig. 8(a) and (b), the ANN model achieves desired performance on the test dataset. In Fig. 9(a), the error distribution histogram shows a nearly normal distribution for the error between predicted and actual completion time, with a mean close to zero. Additionally, in Fig. 9(b), the R-squared correlation between predicted steady-state time and the simulated time is 0.979, demonstrating that the neural network result has a good resemblance with the actual simulation result.

Moreover, in Fig. 10, a region of unsuitable operating conditions is identified where either the steady-state is not reached within 4 s or the steady-state coverage cannot achieve full coverage. For example, at $T = 655$ K and $P = 86$ Pa, although the steady-state is reached at 0.33 s, only 95% of the surface is covered. Additionally, at $T = 545$ K and $P = 112$ Pa, the surface does not reach the steady-state in 4 s. The maximum allowable time to steady-state is set to be 3.5 s considering industrial usefulness. Typically, a SiO_2 ALD half-cycle will take up to 5 s including the precursor flushing stage, which allows for the substrate surface to be fully saturated with precursor particles. Although the duration of precursor flushing may vary according to reactor geometry, 1 to 1.5 s is the usual lower limit allowed in a typical ALD reactor design to prevent damage to reactor caused by high precursor flow rate and pressure (Acton, 2012). This result demonstrates that in order to achieve a good coverage, the contribution of both temperature and pressure should be accounted for. Since the optimal operating conditions might not be easily achieved or needed, the ANN model mentioned above can be used as an initial guideline to determine the feasible pressure and temperature conditions in experiments and industrial productions to operate the ALD process.

Table 2 – Temperature, pressure and predicted time for multi-cycle simulation.

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

3.2.2. *Optimal operating conditions*

In this section, we focus on a narrower range of operating conditions with higher data resolution. For this dataset, one hidden layer with 25 neurons is chosen for the BRANN and the neuron number is determined via a grid search. A single layer BRANN is able to accurately capture the complex non-linearity based on its regularization algorithm. The trained neural network achieves a desired performance with a mean absolute error on the test data of 5.01×10^{-3} s. In Fig. 9(d), the statistical analysis shows that the neural network accurately predicts the simulation data. The R-squared correction between the predicted steady-state time and the simulated time is 0.99, characterized by an almost linear correlation curve. Moreover, as shown in Fig. 9(c), although limited by the number of data points in the test data, the error distribution is close to a normal distribution, which further validates the model fitness.

According to the results in Section 3.2.1, the operating conditions chosen in this section guarantee full coverage of the newly deposited surface at steady-state within 3.5 s. Under these operating conditions, the time necessary to achieve steady-state ranges from 0.31 s to 1.33 s, and a higher temperature and pressure will reduce the time to reach steady-state. The trained BRANN accurately predicts the time to reach steady-state and full coverage from given pressure and temperature inputs, which will be used for cycle time optimization in the next section.

3.3. *Cycle time optimization for multi-cycle SiO_2 ALD*

Although the neural network trained for the dataset of the small range of operating conditions in Section 3.2.2 is based on single-cycle simulations, it is demonstrated to be applicable for predicting cycle completion time for multi-cycle simulation when all layers are close to full coverage. In a typical industrial setting, cycle time for the entire deposition process is fixed at given operating conditions. However, cycle time can be reduced based on the knowledge of cycle completion time according to temperature and pressure input from BRANN results. Such optimization is demonstrated by performing two sets of five-cycle simulations. Specifically, five different pairs of temperature and pressure are chosen for each cycle, as shown in Table 2. For each cycle in the first set, a fixed duration of 3.5 s is given, as discussed in Section 3.2.2, to be industrially practical. In contrast, for the second database the steady-state times predicted by the ANN are given based on the respective operating conditions. Within each cycle, temperature and pressure are kept constant. Both sets result in similar deposited five-layer SiO_2 thin-film with almost full coverage (>99.9%). Therefore, by optimizing operating time from cycle to cycle, we can reduce the surface deposition time by 60% using ANN model result, ignoring the gas-phase development time, which will be analyzed in our future work on the multi-scale CFD modeling. In addition, the cycle-time calcu-

lation from ANN is almost instantaneous and is much faster than that from the kMC model, which needs approximately 1 h to finish a cycle. Moreover, as the film develops thicker, the kMC model will become even slower. This reduction in calculation time is essential for the purpose of cycle-to-cycle optimization and real-time control.

4. Conclusions

In this work, we developed an integrated framework for first-principles-based microscopic modeling, data-driven modeling and optimal operation of thermal atomic layer deposition (ALD) of SiO_2 thin-films using bis(tertiary-butylamino)silane (BTBAS) and ozone as precursors. The performance of the 3D kMC model for the SiO_2 ALD process using fundamental chemical properties from DFT calculations was corroborated by experimentally reported data. The thermodynamic-kinetic competing reaction mechanism was reproduced and a GPC of 1.8 Å per cycle under according operating conditions lied in the range of experimental growth rates. Then, two ANN-based data-driven models for different ranges of operating conditions were constructed. The derived ANN models enabled us to predict the time to achieve steady-state for the film growth at given temperatures and precursor partial pressures. Specifically, the ANN model based on the dataset of large operating range predicted a feasible boundary of operating conditions, and the ANN model based on the dataset of the small range can be utilized in the cycle time optimization to reduce throughput time of the Si-Cycle by 60%, neglecting the gas-phase development time. As a result, the overall approach holds promise for developing an accurate general model for SiO_2 ALD with only precursor thermodynamic and kinetic properties, seeking proper operating conditions while increasing industrial throughput, and significantly saving time and resources that would otherwise have been spent on the testing and manufacturing of physical reaction chambers. Moreover, the generality of the modeling approach in this work makes it possible to extend the current model to the thermal ALD of other thin-film materials with different precursors.

Acknowledgment

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.

Baek, S.B., Kim, D.H., Kim, Y.C., 2012. Adsorption and surface reaction of bis-diethylaminosilane as a Si precursor on an OH-terminated Si(001) surface. Appl. Surf. Sci. 258, 6341–6344.

Battaile, C.C., Srolovitz, D.J., 2002. Kinetic Monte Carlo simulation of chemical vapor deposition. Annu. Rev. Mater. Res. 32, 297–319.

Becke, A.D., 1993. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98, 5648–5652.

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., 2017. Robust dynamic optimization in heterogeneous multiscale catalytic flow reactors using polynomial chaos expansion. J. Process Control 60, 128–140.

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.

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. Advances in Catalysis, vol. 46.

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.

Dasgupta, N.P., Li, L., Sun, X., 2016. Atomic layer deposition for energy and environmental applications. Adv. Mater. Interfaces 3.

Dingemans, G., Helvoirt, C.A.A.V., Pierreux, D., Keuning, W., Kessels, W.M.M., 2012. Plasma-assisted ALD for the conformal deposition of SiO_2: process, material and electronic properties. J. Electrochem. Soc. 159, H277–H285.

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. At. 255, 8–12.

Dkhissi, A., Estève, A., Mastail, C., Olivier, S., Mazaleyrat, G., Jeloaica, L., Djafari-Rouhani, M., 2008. Multiscale modeling of the atomic layer deposition of HfO_2 thin film grown on silicon: how to deal with a kinetic Monte Carlo procedure. J. Chem. Theory Comput. 4, 1915–1927.

El-Kareh, B., 2012. Fundamentals of Semiconductor Processing Technology. Springer Science & Business Media.

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

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.

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

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(001) surface. J. Phys. Chem. C 116, 947–952.

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.

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

Ikegawa, M., Kobayashi, J., 1989. Deposition profile simulation using the direct simulation Monte Carlo method. J. Electrochem. Soc. 136, 2982–2986.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning.

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.
Kamiyama, S., Mira, T., Nara, Y., 2006. Comparison between SiO_2 films deposited by atomic layer deposition with $SiH_2[N(CH_3)_2]_2$ and $SiH[N(CH_3)_2]_3$ precursors. Thin Solid Films 515, 1517–1521.
Kimaev, G., Ricardez-Sandoval, L.A., 2017. A comparison of efficient uncertainty quantification techniques for stochastic multiscale systems. AIChE J. 63, 3361–3373.
Kinoshita, Y., Hirose, F., Miya, H., Hirahara, K., Kimura, Y., Niwano, M., 2007. Infrared study of tris(dimethylamino)silane adsorption and ozone irradiation on Si(100) surfaces for ALD of SiO_2. Electrochem. Solid-State Lett. 10, G80–G83.
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.
Kwon, J.S.I., Nayhouse, M., Orkoulas, G., Ni, D., Christofides, P.D., 2015a. A method for handling batch-to-batch parametric drift using moving horizon estimation: application to run-to-run MPC of batch crystallization. Chem. Eng. Sci. 127, 210–219.
Kwon, J.S.I., Nayhouse, M., Orkoulas, G., Ni, D., Christofides, P.D., 2015b. Run-to-run-based model predictive control of protein crystal shape in batch crystallization. Ind. Eng. Chem. Res. 54, 4293–4302.
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.
Lee, J.M., Lee, J.H., 2005. Approximate dynamic programming-based approaches for input–output data-driven control of nonlinear processes. Automatica 41, 1281–1288.
Li, J., We, J., Zhou, C., Han, B., Karwacki, E.J., Xiao, M., Lei, X., Cheng, H., 2009. On the dissociative chemisorption of tris(dimethyl-amino)silane on hydroxylated SiO_2(001) surface. J. Phys. Chem. C 113, 9731–9736.
Lou, Y., Christofides, P.D., 2004. Feedback control of surface roughness of GaAs(001) thin films using kinetic Monte Carlo models. Comput. Chem. Eng. 29, 225–241.
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 α-SiO_2(110) surface. AIP Adv. 6, 085001.
Mhaskar, P., Garg, A., Corbett, B., 2018. Modeling and Control of Batch Processes: Theory and Application. Springer Science & Business Media.
Momma, K., Izumi, F., 2011. Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data. J. Appl. Crystallogr. 44, 1272–1276.
Moré, J.J., 1978. The Levenberg-Marquardt algorithm: implementation and theory. In: Numerical Analysis. Springer, pp. 105–116.
Murray, C.A., Elliott, S.D., Hausmann, D., Henri, J., LaVoie, A., 2014. Effect of reaction mechanism on precursor exposure time in atomic layer deposition of silicon oxide and silicon nitride. ACS Appl. Mater. Interfaces 6, 10534–10541.
Nalwa, H. (Ed.), 2002. Handbook of Thin Films, vol. 1. Academic Press, Burlington.
Nayhouse, M., Kwon, J.S.I., Christofides, P.D., Orkoulas, G., 2013. Crystal shape modeling and control in protein crystal growth. Chem. Eng. Sci. 87, 216–223.
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.
Nishiguchi, T., Nonaka, H., Ichimura, S., Morikawa, Y., Kekura, M., Miyamoto, M., 2002. High-quality SiO_2 film formation by highly concentrated ozone gas at below 600 °C. Appl. Phys. Lett. 81, 2190–2192.
Oh, S.K., Lee, J.M., 2016. Iterative learning model predictive control for constrained multivariable control of batch processes. Comput. Chem. Eng. 93, 284–292.
O'Neill, M.L., Bowen, H.R., Derecskei-Kovacs, A., Cuthill, K.S., Han, B., Xiao, M., 2011. Impact of aminosilane precursor structure on silicon oxides by atomic layer deposition. Electrochem. Soc. Interface 20, 33–37.
Prechtl, G., Kersch, A., Icking-Konert, G.S., Jacobs, W., Hecht, T., Boubekeur, H., Schröder, U., 2003. A model for Al_2O_3 ALD conformity and deposition rate from oxygen precursor reactivity. In: Proceedings of the International Electron Devices Meeting, Washington, DC, USA.
Putkonen, M., Bosund, M., Ylivaara, O.M., Puurunen, R.L., Kilpi, L., Ronkainen, H., Sintonen, S., Ali, S., Lipsanen, H., Liu, X., 2014. Thermal and plasma enhanced atomic layer deposition of SiO_2 using commercial silicon precursors. Thin Solid Films 558, 93–98.
Rasoulian, S., Ricardez-Sandoval, L.A., 2014. Uncertainty analysis and robust optimization of multiscale process systems with application to epitaxial thin film growth. Chem. Eng. Sci. 116, 590–600.
Rasoulian, S., Ricardez-Sandoval, L.A., 2015a. Robust multivariable estimation and control in an epitaxial thin film growth process under uncertainty. J. Process Control 34, 70–81.
Rasoulian, S., Ricardez-Sandoval, L.A., 2015b. A robust nonlinear model predictive controller for a multiscale thin film deposition process. Chem. Eng. Sci. 136, 38–49.
Rasoulian, S., Ricardez-Sandoval, L.A., 2016. Stochastic nonlinear model predictive control applied to a thin film deposition process under uncertainty. Chem. Eng. Sci. 140, 90–103.
Rey, J.C., Cheng, L., McVittie, J.P., Saraswat, K.C., 1991. Monte Carlo low pressure deposition profile simulations. J. Vacuum Sci. Technol. A 9, 1083–1087.
Ruder, S., 2016. An Overview of Gradient Descent Optimization Algorithms, CoRR abs/1609.04747.
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.
Schwille, M.C., Schössler, T., Barth, J., Knaut, M., Schön, F., Höchst, A., Oettel, M., Bartha, J., 2017a. Experimental and simulation approach for process optimization of atomic layer deposited thin films in high aspect ratio 3D structures. J. Vacuum Sci. Technol. A: Vacuum Surf. Films 35, 01B118.
Schwille, M.C., Schössler, T., Florian, S., Oettel, M., Bartha, J.W., 2017b. Temperature dependence of the sticking coefficients of bis-diethyl aminosilane and trimethylaluminum in atomic layer deposition. J. Vacuum Sci. Technol. A: Vacuum Surf. Films 35, 01B119.
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.
Singh Sidhu, H., Siddhamshetty, P., Kwon, J., 2018. Approximate dynamic programming based control of proppant concentration in hydraulic fracturing. Mathematics 6, 132.
Svozil, D., Kvasnicka, V., Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. Chemom. Intell. Lab. Syst. 39, 43–62.
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.
Wang, Y., Gao, F., Doyle, F.J., 2009. Survey on iterative learning control, repetitive control, and run-to-run control. J. Process Control 19, 1589–1600.