Run-to-Run Control of PECVD Systems: Application to a Multiscale Three-Dimensional CFD Model of Silicon Thin Film Deposition

Marquis Crose, Weiqi Zhang and Anh Tran
Dept. of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095

Panagiotis D. Christofides*
Dept. of Chemical and Biomolecular Engineering and Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095

DOI 10.1002/aic.16400
Published online October 4, 2018 in Wiley Online Library (wileyonlinelibrary.com)

Deposition of amorphous silicon thin films via plasma-enhanced chemical vapor deposition (PECVD) and batch-to-batch operation under run-to-run control of the associated chambered reactor are presented in this work using a recently developed multiscale, three-dimensional in space, computational fluid dynamics model. Macroscopic reactor scale behaviors are linked to the microscopic growth of amorphous silicon thin films using a dynamic boundary which is updated at each time step of the transient in-batch simulations. This novel workflow is distributed across 64 parallel computation nodes in order to reduce the significant computational demands of batch-to-batch operation and to allow for the application and evaluation in both radial and azimuthal directions across the wafer of a benchmark, run-to-run based control strategy. Using 10 successive batch deposition cycles, the exponentially weighted moving average algorithm, an industrial standard, is demonstrated to drive all wafer regions to within 1% of the desired thickness set-point in both radial and azimuthal directions across the wafer surface. This is the first demonstration of run-to-run control in reducing azimuthal film nonuniformity. Additionally, thin film uniformity is shown to be improved for poorly optimized PECVD geometries by manipulating the substrate temperature alone, without the need for re-tooling of the equipment. © 2018 American Institute of Chemical Engineers *AIChE J*, 65: e16400, 2019

Keywords: multiscale modeling, computational fluid dynamics, thin film growth, parallel computing, run-to-run control, thin film silicon solar cells

Introduction

Continual efforts are being made to increase the uniformity of thin film semiconductor devices and quality of solar cell products while maintaining high production quality. Due to the high cost associated with collecting experimental data and re-tooling production machinery (e.g., deposition chambers), the push for accurate modeling of chemical vapor deposition (CVD) is greater than ever (e.g., Refs. 1-6). Recently Crose et al.[7] developed a three-dimensional (3D) multiscale computational fluid dynamics (CFD) model for plasma-enhanced chemical vapor deposition (PECVD), with specific focus on capturing the deposition of amorphous silicon thin films. The model proved successful in capturing both the macroscopic, reactor scale behavior, as well as the microscopic surface interactions associated with thin film growth. While Crose et al.[7] further demonstrated that models of this type may be used in the design of improved reactor geometries (i.e., geometries that reduce thickness nonuniformities in thin film products), there is room for further improvement. Specifically, in altering the PECVD reactor geometry the product thickness nonuniformity was reduced from 8% to <4%; however, demand in the semiconductor and solar cell industries for microelectronic devices of high quality requires products with even tighter film thickness uniformity. To that end, in this work we propose the addition of operational control within successive batch deposition cycles using an exponentially weighted moving average (EWMA) algorithm. The goal of the run-to-run controller is to drive the thickness of the amorphous silicon film in both radial and azimuthal directions across the wafer to the desired set-point of 300 nm by adjusting the temperature of the wafer substrate in four distinct concentric zones between deposition cycles. The 3D multiscale CFD model is used to demonstrate the effectiveness of the run-to-run control algorithm in driving the product to within 1% of the set-point and maintaining it there for all batches thereafter.

The structure of this manuscript is as follows: first, a brief discussion of the process description and of each modeling domain are provided, including the dynamic boundary condition which connects the two simulation regimes. Next, a parallel programming approach is explored which is shown to reduce the computational time of multiscale CFD model solution to within reasonable limits while maintaining a model with high fidelity to the physical system. Steady-state results then yield evidence of significant nonuniformity in the gas-phase species concentrations right above the wafer surface, as

Correspondence concerning this article should be addressed to P. D. Christofides at pdc@seas.ucla.edu

© 2018 American Institute of Chemical Engineers

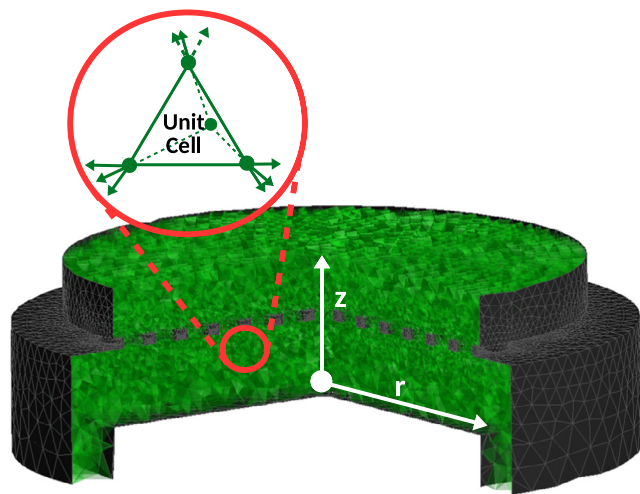

Figure 1. Schematic of the PECVD reactor with 1.5 million tetrahedral cells which define the unstructured, three-dimensional mesh.
[Color figure can be viewed at wileyonlinelibrary.com]

well as the resulting thickness of the *a*-Si:H thin film. Finally, the EWMA algorithm is presented and its impact in reducing film thickness nonuniformity in both radial and azimuthal directions across the wafer surface is demonstrated for the first time in the literature. Specifically, 10 transient batch simulations are conducted for each of two common PECVD geometries and manipulation of the substrate temperature shows success in driving the film thickness everywhere on the surface of the wafer to the 300 nm thickness set-point.

Process Description and Modeling

In the recent publications of Crose et al.,[7,8] the usefulness of 3D reactor CFD modeling in capturing the complex geometry of PECVD systems was demonstrated. The 3D geometry used was able to account for not only the showerhead layout which plays a key role in the distribution of process gases into the reactor, but also the spatial dependency of thin film uniformity on the wafer substrate. As such, the framework developed in Crose et al.[7] will provide the foundation for the study in this work. We again use a cylindrical PECVD reactor geometry with showerhead holes arranged in a rectangular array (see Figure 1). Although PECVD reactor designs vary widely between manufacturers, the spacing between the parallel plates, the diameter of the wafer substrate, and the outlet port regions represent dimensions common to industry and can be thought of as a representative base case. The polar showerhead hole geometry explored in Crose et al.[7] is also applied in this study for comparison, further details of which can be found in the run-to-run control section below.

An unstructured mesh containing ~1.5 million tetrahedral cells is used to discretize the PECVD reactor geometry to allow for numerical solutions to the partial differential equations necessary in capturing the gas-phase mass, momentum, and energy balances. While previous efforts in modeling PECVD systems have relied on structured meshing (including Crose et al.[8]), the curvature of the cylindrical reactor shell and the showerhead holes favor the use of unstructured mesh compositions. More specifically, the tetrahedral cells which define the mesh geometry are distributed with nonuniform cell density. As presented in Figure 1, regions which are expected to experience significant gradients in temperature, flow velocity, species concentration, and so on (e.g., near walls and curved surfaces) have been given significantly higher cell density compared to the bulk fluid regions. Furthermore, the boundary layers which form at the interfaces between the fluid and the reactor surfaces must be considered; cell density is tuned such that the boundary layer is captured within one “layer” of the mesh. Given the relatively low flow rate of the process gas (75 SCCM) and low chamber pressure (P = 1 Torr), flow through even the narrow passages is expected to be laminar (preliminary results and past experiments suggest a maximum Reynold’s number of Re = 4.56×10^{-1}). In order to obtain industrially relevant plasma distributions and thin film growth rates, accurate flow modeling of the process gas throughout the chamber is of great importance, strongly supporting the use of 3D CFD modeling.

It is further essential to note that while ANSYS software[9] is applied in this work, specifically to the creation of the geometric mesh and as a solver for the partial differential equations (PDEs) describing transport phenomena and chemical reactions in the gas phase, the software alone cannot yield the multiscale, 3D CFD model that will allow evaluating the run-to-run control strategy. In particular, three user-defined functions (UDFs) have been developed to tailor the ANSYS solver to the particular problem of PECVD of *a*-Si:H. Specifically, the connection between the UDFs, CFD domain, the parallel programming structure and the incorporation of run-to-run control through the manipulation of substrate temperature in four concentric zones is of particular novelty to this work. The 34 dominant gas-phase reactions which define silane plasmas are accounted for via the volumetric reaction scheme discussed in Crose et al.[7] in a UDF. A second UDF is necessary to calculate accurate electron densities throughout the cylindrical reactor space (i.e., to provide data necessary to carry out the chemical reactions). Finally, the third UDF facilitates the multiscale modeling on which this work is based; specifically, a microscopic simulation domain is defined which runs tandem to the transient CFD calculations. The details concerning the development of the three UDFs can be found in Ref. 7.

While the ultimate goal of this work is to control the growth of amorphous silicon layers within PECVD systems, these results are tied intimately to the physio-chemical phenomena that govern the macroscopic gas phase. Traditional analytic solutions to the gas-phase flow field are viable only for simplified geometries (e.g., the earlier work of Crose et al.[10]) and often fail to provide the resolution necessary for meaningful operation of industrially used systems. Thus, by implementing the mesh structure, the governing equations can be solved with high fidelity using finite volume methods available through the ANSYS software. Again, extended functionality of the software is achieved using the aforementioned UDFs on a cell-by-cell basis.

Rigorous derivation of the continuity, momentum, and energy equations will not be provided here due to the standard formulation used; nonetheless, a generalized vector form is given by the following system:

$$\frac{\partial}{\partial t}(\rho\vec{v}) + \nabla(\rho\vec{v}\vec{v}) = -\nabla p + \nabla\bar{\bar{\tau}} + \rho\vec{g} + \vec{F} \tag{1}$$

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

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

$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla \cdot (\rho \vec{v} Y_i) = -\nabla \cdot \vec{J_i} + R_i + S_i \quad (4)$$

$$\vec{J_i} = -\rho D_{i,m} \nabla Y_i - D_{T,i} \frac{\nabla T}{T} \quad (5)$$

where ρ is the density of the gas, $\vec{v}$ is the physical velocity vector, p is the static pressure, $\bar{\bar{\tau}}$ and I are the stress and unit tensors, J is the diffusive flux, Y_i is the mass fraction of species i, D_i is the diffusion coefficient of species i, and S_h, R_i, and S_i are terms specific to the UDFs.[7] Details concerning the formulation and implementation of the above PDEs may be found in the Fluent user manual.[9]

The R_i term appearing in Eq. 4 is a product of the volumetric reaction set which defines the first UDF. As discussed previously, a set of plasma-phase reactions is used to tailor the functionality of the Fluent solver to the deposition of amorphous silicon. Although extensive reaction networks have been proposed in literature which detail all possible intermediate and aggregate species (e.g., Ref. 11), here we limit the scope to the 12 most dominant species and their associated 34 gas-phase reactions; the reader may refer to Crose et al.[7] for a complete listing of the reactions, mechanisms, and rate constants. Thus, the R_i terms in the mass balance are updated by the first UDF during each time step of the transient solver. Additionally, the right-hand side of Eqs. 3 and 4 contain source/sink terms for energy and mass, respectively. During the deposition of amorphous silicon, energy is exchanged through the formation and breaking of chemical bonds along the surface of the substrate (S_h has units of J s^{-1} m^{-3}). Likewise, growth of the thin film requires mass to be drawn from the gas phase in the form of SiH_3 and H radicals, while mass is similarly reintroduced to the macroscopic domain through the process of hydrogen abstraction (units of kg s^{-1} m^{-3}). Thus, S_h and S_i act as a dynamic boundary between the two simulation domains and their values are updated based on the results from each time step. Finally, with respect to the microscopic thin film growth domain model and the computational implementation of the thin film growth process via lattice-based kinetic Monte Carlo, we note that 60 distinct kMC thin film growth simulators are used that uniformly distributed across the wafer surface and the reader may refer to Ref. 7 for all the details.

Multiscale Simulation Execution

The means by which the macro- and microscopic simulation domains are connected is discussed below. At the start of each batch simulation, $t = 0$, every cell of the mesh must first solve the governing equations (i.e., Eqs. 1–5) with respect to their reduced spatial coordinates via finite volume methods, then the boundaries along adjacent cells are resolved iteratively. Forward integration in time is accomplished using an Implicit Euler scheme. Additionally, the volumetric chemical reaction and electron density UDFs are executed and the results of which are fed into Fluent's PDEs. Once the end of a time step is reached, $t = t_1$, the upper layer of the simulation workflow completes and the temperature, pressure, and species concentrations along the boundary of the wafer substrate are transferred to the microscopic domain. This information is then used to initialize the discrete lattice-based kMC simulations allowing for the growth of a-Si:H thin film layers to begin. Time within the microscopic domain is advanced until all 60 wafer regions have reached t_1. The dynamic boundary condition which links the two domains is then updated for each cell bordering the wafer surface based on the exchange of mass and energy within the associated region. As before, the macroscopic PDEs are solved such that the time progresses to t_2 and the cycle continues. In this way, the multiscale CFD simulation advances until the end time of a batch deposition process, $t = t_{batch}$.

In the interest of clarity, two remarks concerning the interconnection of the micro- and macroscopic domains must be made: first, spatial distribution of kMC simulations has been discussed at length in Ref. 7; however, the spatial distribution introduces a need to interpolate between known data points when updating the boundary conditions. Specifically, each set of three nearest data points forms a triangular surface which provides a means to interpolate boundary condition values for cells which lie between the predefined kMC locations. This methodology smooths out the transition between nearby mesh cells and avoids sharp jumps between adjacent wafer regions. Furthermore, due to the relatively brief start-up time within the macroscopic domain, an obvious computational advantage could be exploited by switching off the transient CFD calculations after the initial gas-phase transients die out. In other words, while the batch growth of the microscopic thin films are inherently transient, it is feasible to operate the macroscopic gas-phase model in a quasi-steady-state mode. This approach significantly improves the computational speed of the model in some cases at the cost of generality to other systems with slower gas-phase dynamics, not to mention a loss in accuracy at the boundary between the macroscopic and microscopic domains.

Parallel Computation of the Multiscale CFD Model

The use of 3D in space reactor models allows for detailed analysis and understanding of PECVD systems, but comes at the cost of significant computational burden. Recent publications from Crose et al.[7,8] detail a sharp jump in computational requirements in moving from 2D to 3D CFD models, with current 3D complete batch simulations requiring days to weeks of continuous processing per batch, making running a large number of batches a challenging task. In this work, the computational demands are no less significant; not only does the macroscopic, CFD domain contain ~1.5 million tetrahedral cells, but every batch requires 60 discrete kMC simulations in order to span the substrate surface. Additionally, the run-to-run operations described in the following sections require 10 consecutive batch deposition cycles per geometry, with redundant batches executed for each to improve accuracy. As a result, serial computation of the 3D multiscale CFD model on a single processor or standalone workstation represents an infeasible task. To that end, a parallel computation strategy is used based on the commonly used message passing interface structure which has been well documented in literature.[12,13]

The process by which a parallel program is created can be easily understood through three basic steps: (1) the original serial task, or tasks, are decomposed into smaller computational elements; (2) the smaller tasks are distributed across multiple processors (nodes); and (3) a host node orchestrates the communication between the multiple, distinct processors at the completion of each time step (we note that for processes operating at steady-state the orchestration occurs between each iteration). The number of available nodes is dependent on the architecture of the workstation or computational cluster used; nodes in modern computational clusters typically contain

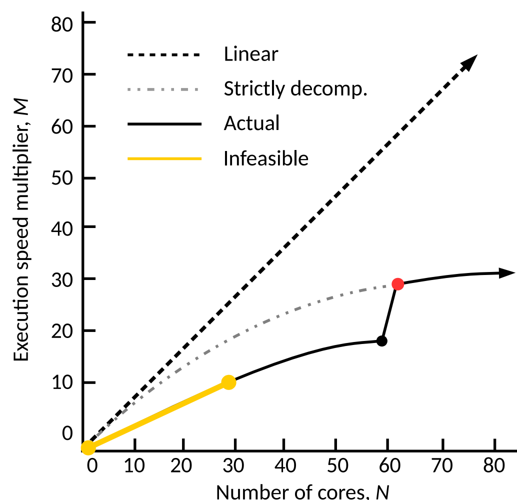

Figure 2. Ideal and actual speedup due to parallelization of the calculations of the multiscale 3D CFD model across *N* nodes.

[Color figure can be viewed at wileyonlinelibrary.com]

multiple computing cores and this is the case in the cluster used in our calculations. In order to generalize the discussion and avoid dependence on a specific process system architecture, nodes and cores will be used interchangeably throughout this section, representing a hypothetical cluster with one core per node. Following the described parallel programming approach, the maximum theoretical speedup that can be achieved with parallel processor computations can be calculated by the following formula:

$$M(N) = \frac{1}{(1-P) + \frac{P}{N}}, \tag{6}$$

where M is the maximum achievable speedup, P is the fraction of the simulation which is available for parallelization (i.e., the fraction of the program tasks which may be discretized), and N is the number of processors used.[14] In practice, the maximum expected speedup deviates from the one calculated by this formula for two key reasons: first, for any parallel architecture, the overhead time required for communication between node and host scales with the number of nodes used. Second, only the serial computations which define the macroscopic (CFD) domain may be strictly decomposed into smaller tasks. In other words, the PECVD reactor mesh can be distributed among any number of computational cores within practical limits; however, the microscopic kMC simulations are unable to be further decomposed. As discussed previously, the wafer substrate requires 60 discrete kMC simulations (each one run at a specific spatial location) in order to span the surface efficiently, but decomposing these 60 simulations further would require significant computational overhead in order to resolve the boundaries of the triangular lattice. The physical speedup which results from the combination of these factors can be seen in Figure 2. Again, a linear 1:1 speedup is out of reach due to the communication overhead, but it is included in the plot as it provides a useful benchmark for comparison purposes. The dashed curve represents the theoretical multiplier if the overall simulation was strictly decomposable. Lastly, the actual multiplier curve (i.e., the solid curve in Figure 2) exhibits a sharp jump between 59 and 60 cores. Synchronized parallelization is used in this work (i.e., faster nodes must wait for slower nodes to complete a task before execution continues) in order to enforce that all boundary conditions and required parameters are available to the host process. Hence, if even one node is required to run two kMC simulations, the remaining $N-1$ nodes must sit idle till completion (a constraint that leads to deviation from the ideal speedup). With 60 kMC locations used to span the substrate surface, the number of nodes is recommended to be >60. The highlighted region between $N = 0$ and 30 is considered infeasible for this work as one or more nodes would be required to run three kMC simulations. No data was collected for this case due to excessively long computational times. Finally, we note that both 64 and 128 nodes were tested during the development of the multiscale, 3D CFD model and associated run-to-run controller implementation presented in this work. Given negligible differences in computational time between 64 and 128 nodes, 64 nodes were used in the collection of all the results presented in the following sections, with each deposition batch requiring roughly 20 h to complete.

Open-Loop Simulation Results

While the simulations discussed in this work are transient in nature, the startup period of the PECVD reactor is relatively brief and therefore will not be discussed in the present work as it is not critical for the evaluation of the performance of run-to-run control. Instead, we present the long-time behavior of the reactor at standard operating conditions of $T = 475$ K, $P = 1$ Torr and a precursor gas flow rate of 75 SCCM with a 10:1 ratio of hydrogen to silane with emphasis on spatial uniformity. To be clear, the reactor never reaches a true steady-state due to the continually developing *a*-Si:H thin film layer; however, given a maximum film thickness of 300 nm and a chamber height of 3 cm, the change in film dimension is trivial compared to the macroscopic geometry and dynamics. Reported thickness measurements have been drawn at the completion of batch simulations, $t_{batch} = 640$ s (i.e., the time required to deposit a 300 nm thick film at the nominal operating conditions). Additionally, each thickness data point represents an averaged value across 10 redundant batch simulations in an effort to minimize stochastic effects inherent in Monte Carlo models (see, Ref. 15 for more discussion on this issue).

Nonuniform film thickness

As reported in the recent publication of Crose et al.,[7] the effect of the showerhead holes on the flow characteristics within the reactor cannot be ignored. Specifically, at the edges of the bottom plate, a relatively high flow rate is experienced compared to the significant dead zone surrounding the center of the reaction zone. For this geometry, which will be referred to as the nominal geometry in this work, the showerhead holes have uniform diameter and spacing across the showerhead region. Nonetheless, the concentration of the key deposition species, SiH_3, within the reaction zone shows significant spatial nonuniformity in both the azimuthal and radial directions (i.e., see Figure 14 in Crose et al.[7]).

Given that wafer substrates lie at the base of the reaction zone, deposited thin films are expected to display nonuniform thickness across their radii, as well as in the azimuthal direction. This phenomena was explored by Crose et al.,[7] and it was concluded that the hot-spots and radial nonuniformity were due, in large part, to the design of the showerhead. Moreover, the thickness nonuniformity was quantified using transient batch simulations at the operation conditions listed

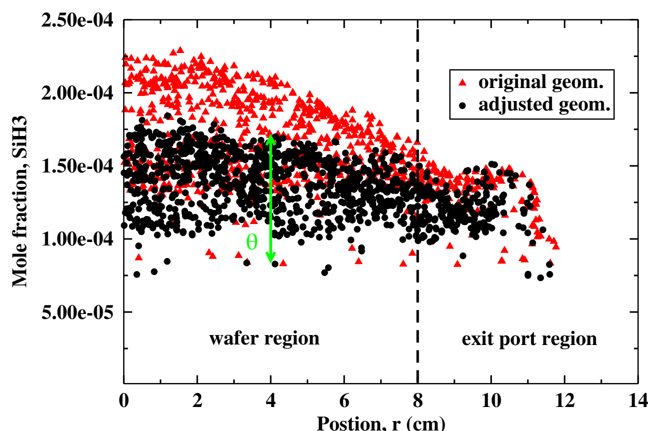

Figure 3. SiH_3 mole fraction as a function of radial position, *r*, and azimuthal position, *θ*.
[Color figure can be viewed at wileyonlinelibrary.com]

above, revealing an 8% difference between the product thickness at the center of the wafer and the edge.

Adjusted reactor geometry

In an effort to improve product quality, an adjusted showerhead geometry was suggested by Crose et al.[7] Specifically, the diameter of the showerhead holes near the center were halved and the spatial arrangement adjusted such that the holes were laid out in a polar array, as opposed to the rectangular array used by the nominal geometry. In doing so the hot-spots were significantly reduced, as were variations along the θ direction, as evidenced by Figure 3. The improvement in the spatial uniformity of x_{SiH_3} resulted in an *a*-Si:H thin film product with only a 4% difference in thickness across its radius.

As mentioned in the motivations for this work, product quality in the microelectronics and solar cell industries is of the highest priority. While a reduction in thickness nonuniformity from 8 to 4% represents a nontrivial improvement to the *a*-Si:H product, in this work we aim to surpass that margin through the use of batch-to-batch control. The following sections provide details for the development and performance of the run-to-run control algorithm which is the key topic of this manuscript.

Run-to-Run Control

Given that the spatial distribution and concentration of the primary deposition species (e.g., SiH_3 and H) cannot be directly controlled, an alternative manipulated variable must be chosen in order to further improve on the remaining 4% offset from the thickness set-point of 300 nm. Fortunately, the growth rates of *a*-Si:H thin film layers have recently been demonstrated to depend roughly linearly on the temperature of the wafer substrate in the neighborhood of the nominal deposition condition of $T = 475$ K, for fixed species concentrations.[16] Hence, a run-to-run control strategy is applied in this work to the multiscale, 3D CFD model which requires only post-batch measurements of the product thickness as an input in order to update the wafer temperature for the successive batch.

Specifically, four concentric zones are defined along the surface of the wafer (see Figure 4), and within each zone an EWMA algorithm is applied in order to iteratively update the substrate temperature such that the *a*-Si:H thin film product may be driven to the desired thickness of 300 nm. We note that in each zone the substrate temperature is uniform in both the radial and azimuthal directions, yet the control actions are applied to the 3D multiscale CFD model to evaluate the effect of the controller in reducing both radial and azimuthal thin film thickness nonuniformity across the wafer. It is also important to note here that the choice to use four concentric zones is not arbitrary; while finer control would be made possible by the use of more zones, practical restrictions on the design and manufacture of PECVD reactors limits this number.

The proposed EWMA algorithm is of the form:

$$\varepsilon_{k+1}^{i} = (1-\lambda)\varepsilon_{k}^{i} + \lambda\left(\tau_{\text{s.p.}} - \tau_{\text{meas.}}^{i}\right) \tag{7}$$

where ε_{k+1}^{i} is the parameter correction for the $k + 1$ batch in zone i, $\tau_{\text{s. p.}}$ is the thickness set-point and $\tau_{\text{meas.}}^{i}$ is the thickness measurement for zone i in the current batch (evaluated via film thickness averaging in both radial and azimuthal directions within zone i). The term λ is known as a learning factor and can be thought of as the weight given to the currently measured error (i.e., $\tau_{\text{s.p.}} - \tau_{\text{meas.}}^{i}$) vs. the historic, or recursively calculated, error ε_{k}^{i}.

Application of the generalized EWMA algorithm provided above requires a change of units; specifically, the error term, ε_{k+1}^{i}, has units of nanometers while the zone temperatures which must be updated have units of Kelvin. Fortunately, due to the linear relationship between thin film growth rate and substrate temperature discussed previously, a growth rate function and thickness approximation can be readily defined:

$$G\left(T^{i}\right) = \alpha^{i} T^{i} + \beta^{i} \tag{8}$$

$$\tau^{i} = G\left(T^{i}\right) \cdot t_{\text{batch}} \tag{9}$$

where $G(T^i)$ is the growth rate as a function of temperature in zone i and the parameters α^i and β^i are a product of the linearization of the growth rate around the nominal deposition temperature of $T = 475$ K. Given that the startup period of the reactor is relatively short lived (i.e., typically <5% of the deposition period, t_{batch}), the thickness for a given zone may be closely approximated by τ^i. Thus, we can approximate the effect of the parameter correction, ε_{k+1}^{i}, as the difference between the thickness of the $k + 1$ and k batches,

$$\varepsilon_{k+1}^{i} = \tau_{k+1}^{i} - \tau_{k}^{i} = \alpha^{i} T_{k+1}^{i} t_{\text{batch}} - \alpha^{i} T_{k}^{i} t_{\text{batch}} \tag{10}$$

Rearranging Eq. 10 yields the desired equation for manipulating the substrate temperature:

$$T_{k+1}^{i} = T_{k}^{i} + \frac{\varepsilon_{k+1}^{i}}{\alpha^{i} \cdot t_{\text{batch}}} \tag{11}$$

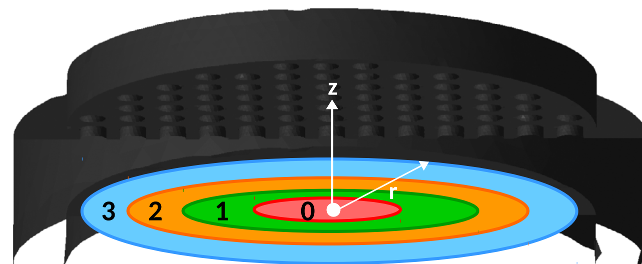

Figure 4. Four concentric wafer zones as defined by the EWMA algorithm.
[Color figure can be viewed at wileyonlinelibrary.com]

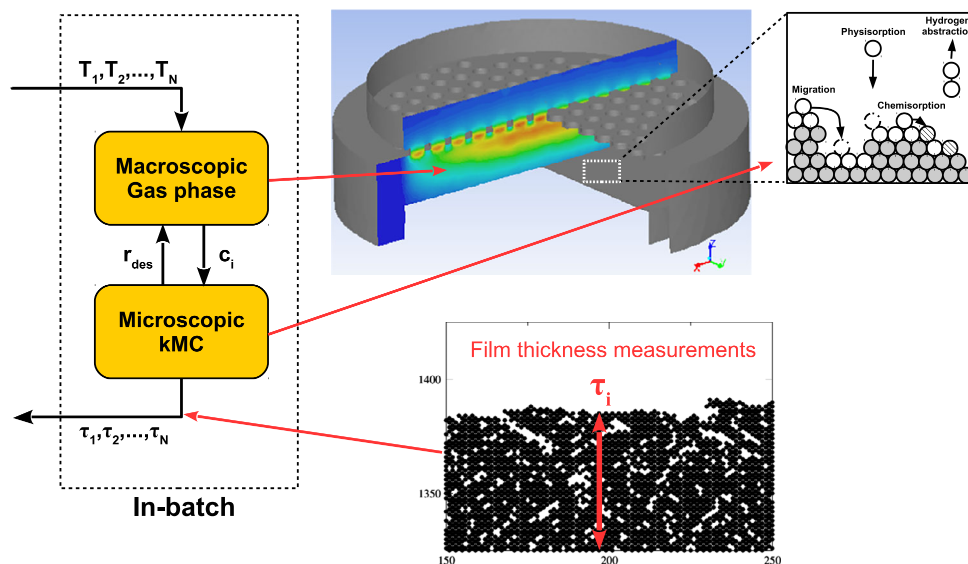

Figure 5. Run-to-run control scheme application to the multiscale CFD PECVD reactor model.
[Color figure can be viewed at wileyonlinelibrary.com]

where T_{k+1}^{i} is the updated substrate temperature in zone i for the k + 1 batch and T_{k}^{i} is the substrate temperature applied to the previous batch.

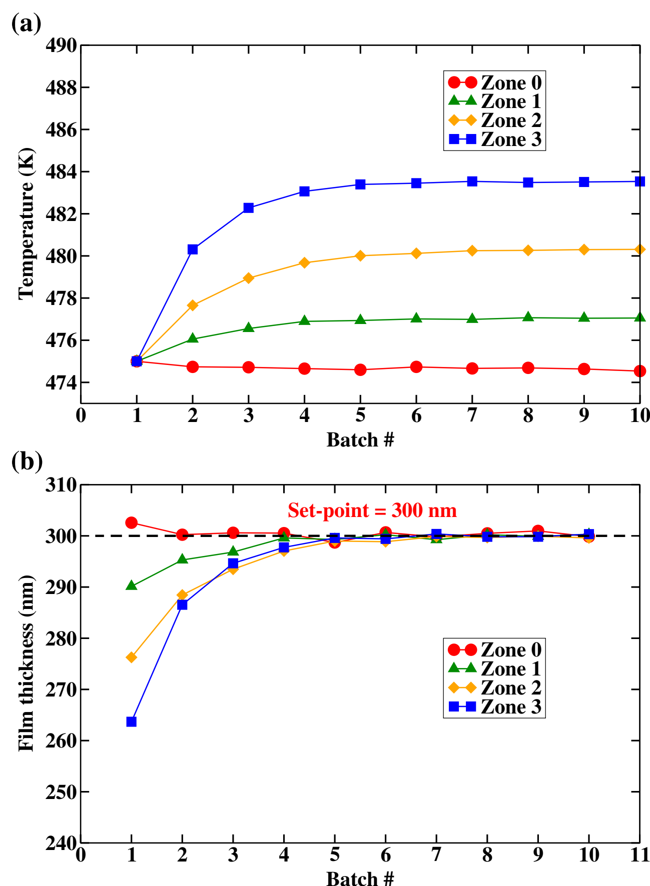

Figure 6. (a) Control action resulting from EWMA algorithm. Temperatures shown correspond to PECVD reactor with polar showerhead geometry. (b) Batch-to-batch thin film thickness within polar PECVD geometry.
[Color figure can be viewed at wileyonlinelibrary.com]

As an important note, while Eqs. 8 and 10 are approximations, they do not affect the fidelity of the overall multiscale model to the physical PECVD system, only the speed of convergence to the set-point. In other words, higher order growth rate functions and more complex thickness relations may be used, but the only tangible benefit would be the speed at which the EWMA algorithm predicts correct wafer temperatures (an off-line calculation). As shown in the following section, specifically Figure 6a, the simplified growth rate and thickness relations provided above quickly drive the system to the set-point and are thus sufficient for this work. The methodology for the simulation and collection of growth rate data will not be provided here. Interested readers are encouraged to consult Crose et al.[16]; we note that in that work, the run-to-run control scheme was applied to a multiscale model that used a simplified model for the gas phase and not a 3D CFD modeling description. Nonetheless, the resulting linear growth relations necessary to the function of the EWMA algorithm are given below:

$$G(T^0) = 0.0350T^0 - 14.46\,(\text{nm}\cdot\text{s}^{-1}) \tag{12}$$

$$G(T^1) = 0.0352T^1 - 14.64\,(\text{nm}\cdot\text{s}^{-1}) \tag{13}$$

$$G(T^2) = 0.0322T^2 - 13.35\,(\text{nm}\cdot\text{s}^{-1}) \tag{14}$$

$$G(T^3) = 0.0287T^3 - 11.88\,(\text{nm}\cdot\text{s}^{-1}) \tag{15}$$

As a final note, the sensitivity of the system to changes in the learning factor, λ, has been shown to be low[16]; nonetheless, a higher value for λ will place more weight on the current thickness offset (i.e., $\tau_{\text{s.p.}} - \tau_{\text{meas.}}^{i}$) and is expected to drive the product to the set-point using fewer batches. The results presented in the following section were collected with λ = 0.75. Figure 5 displays the application of the run-to-run control scheme to the multiscale CFD PECVD reactor model.

Batch-to-Batch Operation

Ten data sets are shown in Figures 6a, b; each of which contains four data points which correspond to the concentric wafer zones. In the first data set (i.e., batch 1), all zones were maintained at the nominal deposition temperature, T = 475 K.

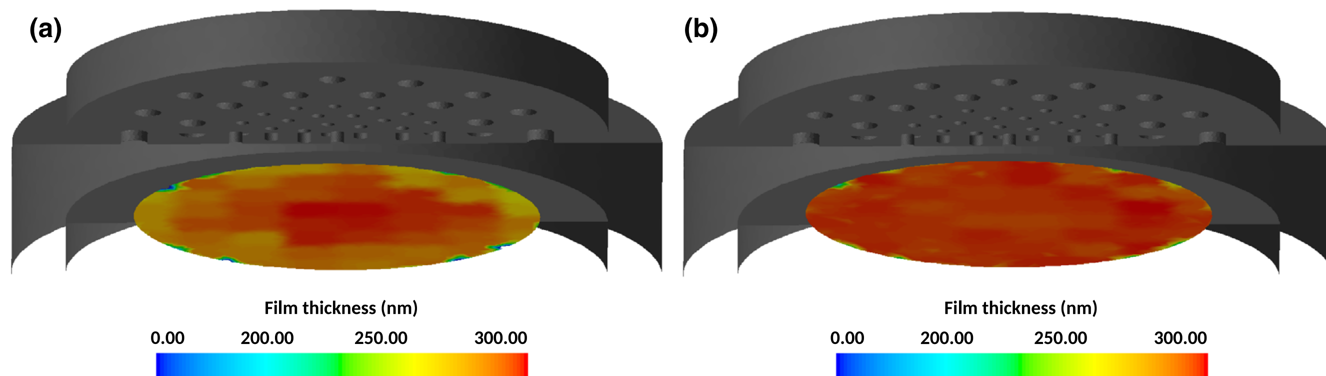

Figure 7. (a) Batch 1 film thickness. (b) Batch 10 film thickness.
[Color figure can be viewed at wileyonlinelibrary.com]

The resulting thickness of the *a*-Si:H thin film demonstrates significant nonuniformity, most apparent in zones 2 and 3 which lie in the outer-half of the wafer substrate. This result is expected due to the loss in SiH_3 density between r = 4 and 8 cm as discussed in the open-loop results.

Readers may note that the thickness nonuniformity at the nominal operating conditions (i.e., the first batch sets shown in Figures 6b and 8) is greater than that predicted in the recent publication of Crose et al.[7] Despite using identical deposition conditions and multiscale models with great similarity, the thickness difference is expected for two key reasons: first, the number of zones has been reduced to four in this work due to the practical constraints discussed previously. Second, convergence of the model for each time step has been refined since last published. Specifically, the species residuals which define convergence for each cell of the mesh, in particular for the cells which lie along the wafer boundary, have been tightened from 10^{-4} to 10^{-6}. As such, the increased nonuniformity in batch 1 is more representative of physical PECVD systems and further motivates the simulation results which follow.

As the transient simulation progresses and additional batch deposition cycles complete, the EWMA algorithm begins to adjust the substrate temperature profile; see batches 2–10 in Figure 6a. Zone 0 maintains a near constant 475 K as its position is in the center of the reactor and receives the highest concentration of SiH_3. Conversely, the temperature of zone 3 is set to nearly 484 K by the R2R controller (i.e., through the use of parameter correction via the EWMA algorithm) in an effort to drive the thin film product thickness to the set-point of 300 nm. The evolution of the thin film thickness in each concentric zone can be seen in Figure 6. Initially zone 0 slightly overshoots the product target of 300 nm thickness, while zones 1–3 significantly undershoot. Nonetheless, by batch 5 all zones are within 1% of the product set-point in both radial and azimuthal directions across the wafer and are maintained inside this margin for all batches thereafter.

While the data points discussed in Figure 6b represent the average thickness within each control zone, in the interest of clarity it is worth discussing the thickness at specific spatial locations. To that end, Figure 7 provides a comparison of the thin film thickness achieved in batches 1 and 10, respectively, in both the radial and azimuthal directions across the entire wafer; to the best of our knowledge, this is the first time the impact of run-to-run control on the azimuthal direction is evaluated as all previous run-to-run control applications were made on models obtained under the assumption of axisymmetric conditions. The vast majority of the defined wafer regions in batch 10 lie within 1% of the thickness set-point; unfortunately, along the edge of the wafer substrate there exist locations with relatively little change in thickness between batches 1 and 10. In industry these segments are typically known as edge die or edge chips and are often discarded.

It is important to note here that while the results presented thus far correspond to the PECVD reactor with improved geometry (i.e., the polar showerhead arrangement determined by Crose et al.[7]), the run-to-run control scheme described in this work may be equally applied to less-optimized reactor geometries. Specifically, the rectangular showerhead arrangement described previously is applied here and the resulting film thickness profiles are shown in Figure 8. Significant thickness nonuniformity can be seen in batch 1 with offset in zone 3 >16% from the 300 nm set-point. The EWMA algorithm is able to recognize the offset and drive zones 1–3 to the set-point within six batches of operation, only requiring one additional batch to reach the set-point compared to the optimized reactor geometry.

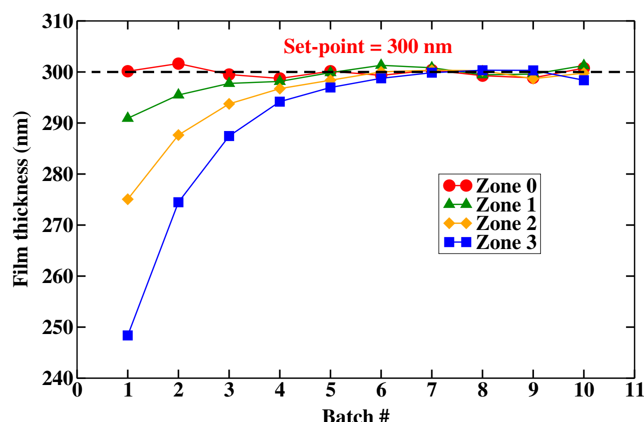

Figure 8. Batch-to-batch thin film thickness within nominal (i.e., rectangular) PECVD showerhead geometry.
[Color figure can be viewed at wileyonlinelibrary.com]

Conclusions

Recent improvements in reactor design have shown promise for reducing thin film thickness nonuniformity. Nevertheless, demand in the microelectronics and solar cell industries for products of increasing quality necessitates novel means of deposition control. A run-to-run control algorithm, which uses substrate temperature correction as manipulated variables via an EWMA formulation, has been applied to 10 consecutive

batch deposition cycles within a multiscale, 3D in space, CFD modeling framework of thin film growth via PECVD. The transient, 3D multiscale simulation which defines each batch suggests the *a*-Si:H product can be driven to the 300 nm thickness set-point within five batches of operation in both radial and azimuthal directions across the wafer surface. Additionally, the run-to-run controller has been demonstrated to be efficient for poorly optimized PECVD geometries, reaching the set-point in six batches and maintaining the product within 1% of the set-point for all batches thereafter.

Acknowledgment

Financial support from the National Science Foundation (NSF), CBET-1262812, is gratefully acknowledged.

Literature Cited

1. Collins DJ, Strojwas AJ, White DD. A CFD model for the PECVD of silicon nitride. *IEEE Trans Semiconduct Manuf*. 1994;7:176-183.
2. da Silva AN, Morimoto NI. Gas flow simulation in a PECVD reactor. In: Proceedings of the 2002 International Conference on Computational Nanoscience and Nanotechnology, San Juan, Puerto Rico, 2002:22–25.
3. Gerogiorgis D, Ydstie B. Multiphysics CFD modelling for design and simulation of a multiphase chemical reactor. *Chem Eng Res Des*. 2005;83:603-610.
4. Rasoulian S, Ricardez-Sandoval LA. Uncertainty analysis and robust optimization of multiscale process systems with application to epitaxial thin film growth. *Chem Eng Sci*. 2014;116:590-600.
5. Rasoulian S, Ricardez-Sandoval LA. A robust nonlinear model predictive controller for a multiscale thin film deposition process. *Chem Eng Sci*. 2015;136:38-49.
6. Gakis GP, Vergnes H, Scheid E, Vahlas C, Caussat B, Boudouvis AG. Computational fluid dynamics simulation of the ALD of alumina from TMA and H2O in a commercial reactor. *Chem Eng Res Des*. 2018; 132:795-811.
7. Crose M, Zhang W, Tran A, Christofides PD. Multiscale three-dimensional CFD modeling for PECVD of amorphous silicon thin films. *Comput Chem Eng*. 2018;113:184-195.
8. Crose M, Tran A, Christofides PD. Multiscale computational fluid dynamics: methodology and application to PECVD of thin film solar cells. *Coatings*. 2017;7:23.
9. ANSYS Fluent Theory Guide 15.0 (November). ANSYS Inc.; 2013.
10. Crose M, Kwon JSI, Nayhouse M, Ni D, Christofides PD. Multiscale modeling and operation of PECVD of thin film solar cells. *Chem Eng Sci*. 2015;136:50-61.
11. Kushner MJ. A model for the discharge kinetics and plasma chemistry during plasma enhanced chemical vapor deposition of amorphous silicon. *J Appl Phys*. 1988;63:2532-2551.
12. Ingle NK, Mountziaris T. A multifrontal algorithm for the solution of large systems of equations using network-based parallel computing. *Comput Chem Eng*. 1995;19:671-681.
13. Kwon JSI, Nayhouse M, Christofides PD. Multiscale, multidomain modeling and parallel computation: application to crystal shape evolution in crystallization. *Ind Eng Chem Res*. 2015;54:11903-11914.
14. Culler DE, Singh JP, Gupta A. *Parallel Computer Architecture: A Hardware/Software Approach*. Gulf Professional Publishing; 1999.
15. Lou Y, Christofides PD. Feedback control of growth rate and surface roughness in thin film growth. *AIChE J*. 2003;49:2099-2113.
16. Crose M, Kwon JSI, Tran A, Christofides PD. Multiscale modeling and run-to-run control of PECVD of thin film solar cells. *Renew Energy*. 2017;100:129-140.

Manuscript received Aug. 2, 2018, and revision received Aug. 22, 2018.