Optimization of a Chemical Vapor Deposition Process Using Sequential Experimental Design

Paul J. Wissmann and Martha A. Grover*

Department of Chemical & Biomolecular Engineering, Atlanta, Georgia 30332

The exact mechanisms occurring during chemical vapor deposition are often not well understood or quantified, and CVD is therefore difficult to optimize efficiently. In this article, we implement a process optimization method based on sequential experimental design. This method simultaneously employs both mechanistic and empirical models, using confidence intervals to predict the region of potential optima. New experiments are constrained to this region, so that model predictions will be improved near the process optimum. After six initial experiments and two sequential experiments, a process temperature of 760 °C is computed, to achieve a desired film roughness of 7 ± 1.65 nm. The experimental data are obtained using a chemical vapor deposition system. The yttrium precursor material is deposited onto a 1 in. silicon wafer in a vacuum chamber. The samples are then analyzed using an atomic force microscope to determine the roughness of the samples. On the basis of these data, we create models that are purely empirical, and also models that are motivated by a more mechanistic understanding of the process. The rationale is that, rather than concentrating our experimental design on creating either empirical or mechanistic models, both types of models should be considered when optimizing a process. Furthermore, by developing both model types simultaneously, mechanistic understanding of the process can also be improved. Models are developed which predict the thin film roughness from the surface temperature and molar flow rate of the precursor. Five empirical models are considered, as well as two additional models which relate nucleation density to final film roughness using mechanistic modeling. The temperature is found to be most important for controlling the thin film roughness. An empirical model with two fitted parameters and a linear temperature dependence predicted performance best, while a hybrid model using nucleation density and two fitted parameters was second best. The confidence interval for the best model was reduced by 20% using our experimental methodology, using only two sequential experiments. The experiments indicate that the hybrid model does capture some of the trends of the experimental data but needs a stronger dependence on temperature to have more statistical significance.

1. Introduction

Chemical vapor deposition (CVD) is a common method for creating a thin coating of solid material, with applications including microelectronics,[1,2] thermal barrier coatings,[3,4] and fuel cells.[5−7] An advantage of CVD is the ease of depositing uniform films over nonuniform surfaces. Additionally, a high vacuum and line-of-site to the substrate are not required for CVD but are required for physical vapor deposition methods.[8] The underlying chemical processes in CVD are not always very well understood or quantified, and on the factory level, many wafers are wasted in calibration, to find a good recipe for the process.[9] If one had a better understanding of the underlying phenomena and sources of uncertainty, one could design better CVD equipment and recipes to deposit the desired film microstructure for a specific application. However, due to the rapidly changing nature of the microelectronics industry, always making devices smaller and smaller, a detailed study of these physical mechanisms is often impractical. Models must be developed rapidly if they are to be used in developing the process, and often this is accomplished using empirical models.

The three model types considered here are empirical, mechanistic, and hybrid models. Empirical models are built directly from experimental data to fit the trends in the experimental measurements. Data from designed experiments are often used to construct empirical models.[10] Such models work well when experimental data are readily available and the underlying detailed mechanisms are not well understood.[11] In contrast, a mechanistic model incorporates knowledge of the physical phenomena in the process and can be developed without reference to a particular experimental system and data.[12,13] However, when a mechanistic model is developed independently from a specific piece of hardware and the resulting experimental data, there may be unmodeled effects such that the model does not have the required accuracy needed for prediction of future experimental data. Additionally, a mechanistic model developed independently from data may not be directly applicable to an experiment, if certain key model outputs cannot be measured in the experiment, or if certain input factors in the experiment are not present in the mechanistic model. In CVD, mechanistic models have been generated using computational chemistry[14] and have been used in shaping the experimental procedure.[15] However, a purely mechanistic model used for experimental design or process optimization in CVD is rare.

A hybrid model is developed using a physical understanding of the process, but not all parameters for the system are known. Instead, a set of parameters is fit using the experimental data. Kennedy and co-workers[16] used hybrid models in the calibration of computer models to account for sources of uncertainty. Prasad and Vlachos[17] recently used experimental design concepts in a simulated catalysis process, for the purpose of improved parameter identification of chemical rate constants. There is a growing need for more descriptive (i.e., mechanistic or hybrid) process models in the microelectronics industry.[9,18] Unfortunately, the extremely tight manufacturing tolerances required in the microelectronics industry are typically more stringent than the prediction accuracy of mechanistic models. By combining empirical and mechanistic parts, a hybrid model may be able to span this divide between purely mechanistic models and purely empirical models.

* To whom correspondence should be addressed. E-mail: martha.grover@gatech.edu.

10.1021/ie901055e
Published on Web 05/19/2010

The common approach to CVD process optimization is factorial experimental design using empirical models. Some CVD experimenters use a nonsequential approach, with a single set of designed experiments,[19−23] whereas other experimenters use a more sequential approach, using results from the first set of screening experiments to design a second set.[24,25] Among these, most designs are factorial or fractional factorial experiments, some use orthogonal designs, and none use methods requiring a mechanistic model of the CVD system. CVD experiments are time-consuming and expensive, so a method for the most intelligent design of experiments is valuable.

When learning about a process, sequential experimentation is often used, as learning is a dynamic process—Box and Liu[26] observe that one set of experiments on the same process would probably yield different results given different experimenters. However, a sequence of experiments will likely lead to comparable conclusions. Research in sequential experimental design has focused on the limit when experiments are cheap and fast,[27] and asymptotic convergence properties have been proven in the limit of a large number of experiments.[28] When random error is small and experiments are fast, Daniel[29] advocated the use of one-factor-at-a-time plans. However, CVD experiments are costly, and often only a small number can be run, such that the first few experiments should be carefully designed to yield the most information. More recently, Frey and co-workers[30] compared sequential orthogonal and one-factor-at-a-time experimental designs. They noted that orthogonal designs are more affected by interactions between the factors but less affected by pure experimental error. They also suggest that one-factor-at-a-time experiments have their advantages when used in conjunction with physics-based mechanstic models, when only a small number of experiments will be performed.

Solving for the optimal experimental point is computationally challenging when a nonlinear mechanistic or hybrid model is used. Depending on optimality conditions, the optimal region will have an irregular shape. Using only a factorial or orthogonal design may miss potential optima or sample in regions that are not of interest. When searching for the optimum point in an expensive computer simulation, Jones and co-workers[31] used confidence bounds to design the next simulation point, which would have the highest expected improvement in the objective function. Henkenjohann et al.[32] later used a similar approach applied directly to experiments, to optimize a sheet metal spinning process. However, this approach requires a large number of sample points—Jones et al. recommend having 10 points per input dimension, before even beginning the sequential design portion. In this paper, we also use the confidence bounds on the optimal point to define regions where the next experiment can be run, to improve the prediction of the mean response near the optimal point of the process. In contrast, traditional sequential design of experiments "zoom in" to the point of interest based on previous experimental data or extrapolate to another region on the basis of a steepest ascent calculation.[10]

The article is organized as follows. Section 2 gives a brief description of the CVD system used to deposit thin films. In section 3, the sequential experimental design method recently proposed by Wissmann and Grover[33] is summarized. In section 4, we apply this method to our CVD experiments. In section 4.1, the initial data are described, while in section 4.2, the models used to fit the experimental data are introduced. The results of implementing the experimental design method on the CVD system are described in section 4.3, and concluding remarks are given in section 5.

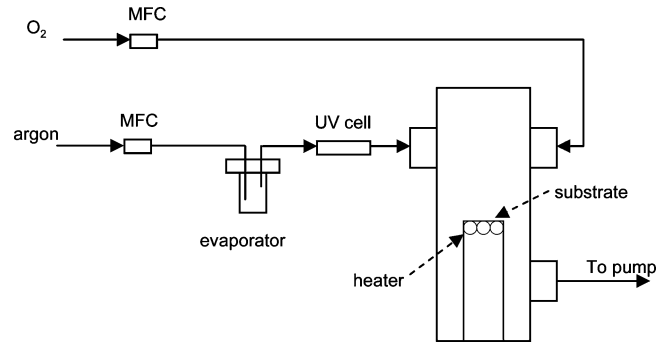

Figure 1. Schematic of the chemical vapor deposition system. The precursor is sublimated in the evaporator and is carried to the reaction chamber, where it reacts with the substrate to form a thin film.

Table 1. Experimental High and Low Settings for the Experimental Study

setting	high	low	units
T	775	600	°C
$\dot{n}$	200	100	μmol/min

2. The CVD Process

In a thermal chemical vapor deposition process, a wafer, or substrate, is heated to a high temperature and then is exposed to a reactive precursor gas, which contains the substance to be deposited.[8] At high temperatures, the precursor compound begins to decompose, and one or more of the reaction products deposits on the surface, creating a solid thin film. Metalorganic precursors are often used in CVD due to their low evaporation temperatures and high chemical reaction rates.[34] A schematic of the CVD system is presented in Figure 1. For more details, see Wissmann.[35] Using mass flow controllers for the inert argon carrier gas, the flow is directed through a heated evaporator to sublimate the $Y(tmhd)_3$ precursor [Strem Chemicals, CAS 15632-39-0], which is a powder at room temperature. The gas then flows through a sensor to monitor and control the mass flow rate of the precursor, using ultraviolet absorption. The precursor gas is then introduced into the cold-wall type CVD reaction chamber. Inside this reactor, the precursor gas is combined with oxygen gas and flows over a 1″ silicon substrate sitting atop a resistive heater [Heatwave Laboratories part: 101275−27]. The surface temperature and growth rate were monitored using an emissivity-correcting pyrometer [SVT Associates 4000]. The process settings varied in this optimization are surface temperature and precursor mass flow rate. The high and low values for the process settings are shown in Table 1.

After deposition, the substrate is removed from the CVD chamber for further characterization of the film, including atomic force microscopy (AFM) [Agilent PicoSPM II] to measure the surface height profile. The film roughness is calculated using the AFM[36,37] across a 2×2 μm area of the wafer. Three different locations on the wafer are measured, and one of the locations is measured twice to account for scan-to-scan variability. All AFM images were processed by flattening the image using a polynomial of degree three [Gwyddion], and the root-mean square roughness was then calculated from the image. A typical AFM image is shown in Figure 2.

3. Experimental Design Methodology

The basic steps in the methodology proposed by Wissmann and Grover[33] are presented here. The experimental settings to be varied may be determined via mechanistic models and domain knowledge, or with a screening experiment of the

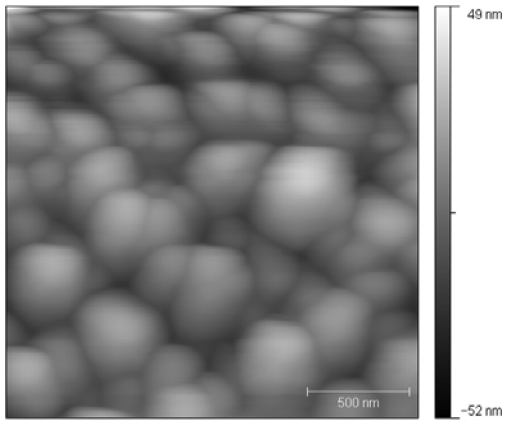

Figure 2. AFM image of experiment 1.

possible experimental settings. Here, we consider process settings of temperature and precursor flow rate. Each candidate model M_j, $j = 1, 2, ..., m$, may have unknown parameters θ_j. Initial estimates of the unknown parameters in each model are needed and may be available from the literature or acquired from a preliminary set of experiments.

The sequential design begins once initial estimates of θ_j are obtained. Any experimental design criterion can be used for the sequential portion. The three methods compared in this work are a random design of experiments, the D-optimal,[10] and the P-optimal design to minimize the variance on the mean response at the predicted optimal point.[33] In all cases, the most probable model—on the basis of all available data—is used to design the next experiment. We distinguish this P-optimal approach from another recently proposed approach with the same name,[38] which uses principal component analysis to extract the subset of parameters that can be robustly estimated from the measurements.

Once an experiment is completed, the parameters in each model M_j are estimated again, now using both the previous data and the new data point. In this work, the parameters are estimated using a least-squares minimization between the model and the data, although it would also be possible to estimate the parameters using other methods such as a Bayesian approach using *a priori* estimates for the parameters. Having obtained parameter estimates, θ_j, for each model M_j, predictions can be made. For example, the predicted optimal point from each model is

$$\hat{\mathbf{x}}_j = \arg\min_{\mathbf{x}} f(\hat{y}_j(\mathbf{x})) \tag{1}$$

where f is the objective function to be minimized. $\hat{\mathbf{x}}_j$ is the estimated optimal point from *model j* and is not necessarily $\bar{\mathbf{x}}$, the true optimal point of the *process*. $\hat{\mathbf{x}}_j$ will ideally converge toward $\bar{\mathbf{x}}$ as more data are acquired.

A unique feature of the method by Wissmann and Grover[33] is the simultaneous use of multiple models in the sequential experimental design. To compare the statistical significance of each model after each experiment is run, we compute the Akaike information criterion (AIC):[39]

$$\mathrm{AIC}_j = n(\ln(2\pi\mathrm{MSE}_j) + 1) + 2p_j \tag{2}$$

where n is the number of experiments run, p_j is the number of parameters in model j, and MSE_j is the mean squared error:

$$\mathrm{MSE}_j = \frac{\sum_{i=1}^{n} (y(\mathbf{x}_i) - \hat{y}_j(\mathbf{x}_i))^2}{n} \tag{3}$$

Here, $y(\mathbf{x}_i)$ is the experimental measurement, $i = 1, 2, ..., n$, and $\hat{y}_j(\mathbf{x}_i)$ is the prediction from model j at the i^{th} experimental setting. In eq 2, a model is penalized for having either more fitted parameters or a larger error than an alternative model. Note that eq 2 is used for mechanistic, empirical, and hybrid models. If *a priori* estimates $P(M_j)$ are available for each model, they can be used to compute a Bayesian model probability, but in this work, we use a frequentist approach which reflects our lack of knowledge on which model will be best.

The uncertainty on the mean value of y predicted by the model is quantified by the variance on the mean response $\sigma_j^2(\mathbf{x})$ of $\hat{y}_j(\mathbf{x})$:[40,41]

$$\sigma_j^2(\mathbf{x}) = \mathbf{a}^{(j)}(\mathbf{A}_j^T\mathbf{A}_j)^{-1}\mathbf{a}^{(j)T}\hat{\sigma}_j^2 \tag{4}$$

where $\mathbf{A}_j$ is the design matrix for model j (i.e., $\partial\hat{y}_j/\partial\theta_k(\mathbf{x}_i)$) on the basis of the n experiments already conducted, and $\mathbf{a}^{(j)}$ is the row vector $\partial\hat{y}_j/\partial\theta_k(\mathbf{x})$. The model variance $\hat{\sigma}_j^2$ used in eq 4 is calculated similarly to MSE_j in eq 3, except that the summation is divided by $n - p_j$, because p_j degrees of freedom are lost by estimating p_j parameters.

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{n} (y(\mathbf{x}_i) - \hat{y}_j(\mathbf{x}_i))^2}{n - p_j} \tag{5}$$

The variance on the mean response $\sigma_j^2(\mathbf{x})$ is used to calculate the confidence interval on $\hat{y}_{j^*}(\mathbf{x})$

$$\mathrm{CI}_j(\mathbf{x}) = \pm t_{\alpha/2,n-p_j}\sqrt{\sigma_j^2(\mathbf{x})} \tag{6}$$

where α is the level of confidence desired[10] and j^* is the most probable model based on the AIC.

Criteria must be specified for determining whether or not to conduct a new experiment. As with the experimental design, the criteria are calculated using the most probable model, M_{j^*}. The change in the MSE of model j^* is the first criterion and is satisfied if the change is less than ε_{tol}, where ε_{tol} is a prespecified constant value. The second criterion is satisfied if the confidence interval at $\hat{x}_{j^*}$ is below the desired level, ε_{CI}. If the first stopping criterion is true, then additional experiments may not improve the design or the confidence interval on the model prediction. If either criterion is true, then the experimenter needs to intervene to interpret the results. The experimenter should determine if the final optimal point and its confidence interval are satisfactory, or else if a new model—and possibly also new experiments—will be needed to further optimize the process. Note that at any time during the experiments, it is possible to add one or more new models and to continue to iterate.

Because all candidate models are assumed to have some error in their model structure, no model is expected to represent y perfectly for all $\mathbf{x}$. However, when restricted to a smaller region near the minimum of an objective function f, the models may better describe the local behavior in that region. Thus, each designed experiment is constrained to lie in the region of *potential* optimal points for the process. This region is defined using the predicted optimum point $\hat{\mathbf{x}}_{j^*}$, and using the confidence interval on the objective function $\mathrm{CI}(f(\hat{y}_{j^*}(\hat{\mathbf{x}}_{j^*})))$ at the predicted optimum. To minimize the computations required for the experimental design, the continuous variable $\mathbf{x}$ is discretized onto a grid. There are five steps to implement this approach:

1. Choose a set of grid points, $\mathbf{x}_g$, in the experimental domain, Z.

2. Evaluate a threshold value, $H(\hat{\mathbf{x}}_{j^*})$:

$$H(\hat{\mathbf{x}}_{j^*}) = f(\hat{y}_{j^*}(\hat{\mathbf{x}}_{j^*})) + \mathrm{CI}(f(\hat{y}_{j^*}(\hat{\mathbf{x}}_{j^*}))) \quad (7)$$

3. Determine which grid points have outputs that may be below the threshold value, $f(\hat{y}_{j^*}(\mathbf{x}_g)) - \mathrm{CI}(f(\hat{y}_{j^*}(\mathbf{x}_g))) < H(\hat{\mathbf{x}}_{j^*})$. (These are the potential optimal points, at the α confidence level.)

4. Evaluate the output of the experimental design function, $f_{\mathrm{ED}}(\mathbf{x})$, at these remaining grid points and pick the best point, $\mathbf{x}_{e,o}$, from this set.

5. Seek the optimal experimental point, $\mathbf{x}_e$, using $\mathbf{x}_{e,o}$ as a starting point, constraining the search within a box centered at $\mathbf{x}_{e,o}$, with sides equal to twice the grid spacing.

Equation 7 represents the case where the objective is to find the minimum value of some objective function $f(\mathbf{x})$. This is shown graphically in Figure 3, where all values of $\mathbf{x}$ having a lower bound less than the upper bound of the predicted optimum are potential minima (marked by open circles). The closed squares are not potential minima at a confidence level of α.

4. Results

The goal in this study is to find the temperature T and precursor molar flow rate $\dot{n}$ that minimize the objective function

$$f_{j^*}(\mathbf{x}) = (\mathrm{RMS}(\hat{y}_{j^*}(\mathbf{x})) - \mathrm{RMS}_{\mathrm{goal}})^2 \quad (8)$$

where $\mathbf{x} = [T\dot{n}]^T$. The film roughness RMS is the experimentally measured quantity

$$\mathrm{RMS} = \sigma_{\mathrm{h}} \quad (9)$$

where σ_{h} is the standard deviation of the height measurement from the AFM image. The target roughness of $\mathrm{RMS}_{\mathrm{goal}} = 7$ nm is our desired property for the deposited film.

Since not all models may depend on both T and $\dot{n}$, we also consider the secondary objectives of minimizing T and $\dot{n}$, which will minimize the processing cost. However, this is only used when there is no unique minimum to eq 8.

High and low settings were selected for each input factor, as shown in Table 1. The levels were selected to represent a typical range that is practical for the CVD experiment. Nominal settings for the process are shown in Table 2.

4.1. Initial Experiments. The initial experiments were performed at high and low settings as in a 2^2 factorial design, and the results were used to select candidate models and fit model parameters. Two additional replicates were performed at the low T/high $\dot{n}$ setting, so that the process variance could be estimated. Performing replicates only at a single setting is common in CVD, due to the high cost of each run.[21,25] The experiments were performed sequentially in numerical order. The results of the initial set of experiments are shown in Table 3, and experiments 4−6 are replicates. Table 3 shows both the nominal values of T and $\dot{n}$ used (in scaled variables) as well as the actual measured values of T and $\dot{n}$.

The AFM-measured roughness from the initial experiments is also shown in Table 3. For each experiment, the roughness is averaged over four AFM images at three different locations on the film. From Table 3, there are a few points to highlight. First, the deposition temperature seems to have a significant effect on the roughness of the film. A lower deposition temperature leads to rougher films while high deposition temperature films are smoother. In contrast, the molar flow rate does not appear to influence the roughness of the films. The flux of atoms to a surface is an important parameter in many mechanistic models for film growth[13,42] and is proportional to the molar flow rate of the precursor. However, the data in Table 3 have significant variability, and over this range of $\dot{n}$ values, no obvious trend exists.

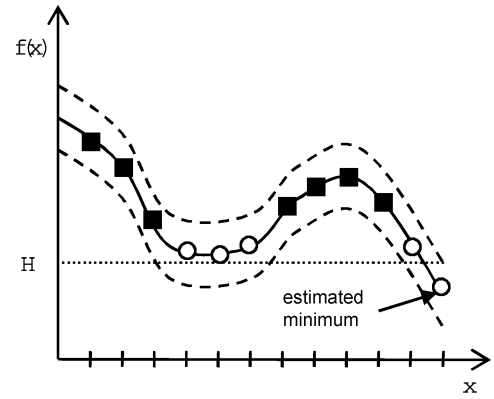

Figure 3. Graphical depiction of the region of potential optima. The upper and lower bounds on the prediction of the model are marked by dashed lines, and the dotted line is the threshold value, H. Grid points that are potential minima are marked by open circles, which are part of the reduced experimental design space, while squares are not potential minima.

Table 2. Typical Experimental Conditions for Thin Film Deposition

setting [units]	
precursor	$Y(tmhd)_3$
pressure [torr]	1.6
O_2 [sccm]	50
system T [°C]	155
evaporator T [°C]	140
gas flow through reactor [sccm]	200
film thickness [nm]	120

Table 3. Experimental Results from Initial Experiments

experiment	nominal scaled T	nominal scaled $\dot{n}$	measured T [°C]	measured $\dot{n}$ [μmol/min]	measured RMS [nm]
1	−1	−1	610	99	10.3 ± 0.6
2	1	1	770	200	5.6 ± 0.5
3	1	−1	782	101	6.3 ± 0.5
4	−1	1	607	205	8.8 ± 1.2
5	−1	1	605	201	12.2 ± 1.5
6	−1	1	610	202	11.6 ± 0.5

4.2. Models of the Experimental Data. The initial data are used to select candidate models of the process for use in the sequential experimental design. Our goal is to design a process to achieve the desired surface roughness of the thin film, which is important to film properties such as dieletric properties[43] and on the structure of multilayer films.[44]

Seven potential roughness models are included in the experimental design. The five empirical models are

$$\mathrm{RMS}_1 = \phi_{11} + \phi_{12}T \quad (10)$$

$$\mathrm{RMS}_2 = \phi_{21} + \phi_{22}T + \phi_{23}\dot{n} \quad (11)$$

$$\mathrm{RMS}_3 = \phi_{31} + \phi_{32}T + \phi_{33}T\dot{n} \quad (12)$$

$$\mathrm{RMS}_4 = \phi_{41} + \phi_{42}T\dot{n} \quad (13)$$

$$\mathrm{RMS}_5 = \phi_{51} \quad (14)$$

The empirical models all use coded variables for T and $\dot{n}$, based on the high and low settings in Table 1.

Table 4. Constants for the Film Growth Case Study

constant	value	units	description
κ	0.15		monolayers (Evans et al.)
ρ	2		number of adatoms per stable cluster
R	8.3145e-3	kJ/mol·K	gas constant
Ω	4.488e22	nm^3/mol	molar volume
h_{ML}	0.2651	nm/ML	thickness of a monolayer (Xu et al.)
$A_{substrate}$	5.07×10^{14}	nm^2	area of substrate
E_i	0	eV	binding energy
E_d	0.8	eV	diffusion activation energy
η	0.2		capture number

Table 5. Results for Roughness Models after Initial Experiments

model	equation	MSE	AIC	ϕ_1	ϕ_2	ϕ_3
1	$\phi_{11} + \phi_{12}T$	1.22	22.2	8.47	−2.48	
2	$\phi_{21} + \phi_{22}T + \phi_{23}\dot{n}$	1.21	24.2	8.49	−2.50	−0.08
3	$\phi_{31} + \phi_{32}T + \phi_{33}T\dot{n}$	1.19	24.1	8.37	−2.40	−0.34
4	$\phi_{41} + \phi_{42}T\dot{n}$	5.65	31.4	8.83	−0.85	
5	ϕ_{51}	6.24	30.0	9.13		
6	$\phi_{61} + \phi_{62}\sqrt{N_{isl}}$	1.73	24.3	19.5	−106	
7	$\phi_{71}\sqrt{N_{isl}}$	17.1	36.0	85.6		

The hybrid models are

$$\mathrm{RMS}_6 = \phi_{61} + \phi_{62}\sqrt{N_{\mathrm{isl,ss}}} \tag{15}$$

$$\mathrm{RMS}_7 = \phi_{71}\sqrt{N_{\mathrm{isl,ss}}} \tag{16}$$

which use the steady state density of islands, $N_{\mathrm{isl,ss}}$, to model the roughness of a thin film. N_{isl} is the time-varying density of atomic clusters, or "islands," on the substrate surface during the initial stages of deposition. It is calculated by integrating

$$\frac{dN_1}{dt} = F(1-\kappa) - (\rho+1)K_{\mathrm{nuc}}(\eta, T, E_i, E_d, N_1) - K_{\mathrm{agg}}(\eta, T, E_i, E_d) \tag{17}$$

$$\frac{dN_{\mathrm{isl}}}{dt} = K_{\mathrm{nuc}}(\eta, T, E_i, N_1) \tag{18}$$

This mechanistic model uses the constants listed in Table 4. These equations are described in detail in Evans et al.[42] The square root of N_{isl} represents the typical distance between islands and is used as the regressor in the hybrid models of eqs 17 and 18. In both models, the precursor flux, F, needed for eqs 17 and 18, is calculated from the molar flow rate as

$$F = \omega\dot{n} \tag{19}$$

where

$$\omega = \frac{1\ \mathrm{mol}}{10^6\ \mu\mathrm{mol}}\Omega\frac{1}{h_{\mathrm{ML}}}\frac{1}{A_{\mathrm{substrate}}} = 3341.07\frac{\mathrm{layers}}{\mu\mathrm{mol}}$$

In all seven models, RMS is linear in the unknown parameters. Consequently, all parameter fits were calculated exactly using linear least-squares regression. If a model is used that is nonlinear in the fitted parameters, a numerical optimization could be performed instead. The parameter fits for the roughness models and their respective mean squared errors and AIC values are given in Table 5. Model 1 is the most likely model after the initial experiments, as quantified by the AIC, due to the strong dependence on T described in section 4.1. Models 1 and 2 have similar errors as quantified by MSE, but model 1 has only two fitted parameters and is therefore the more likely of the two models. Model 6, a hybrid model, is the second most significant according to the AIC (comparable to models 2 and 3), but after the initial experiments, an empirical model (model 1) is selected for designing the next experiment. A smaller MSE for model 6 can be achieved by fitting some of the parameters shown in Table 4, although this would also result in a loss of degrees of freedom in the AIC calculation.

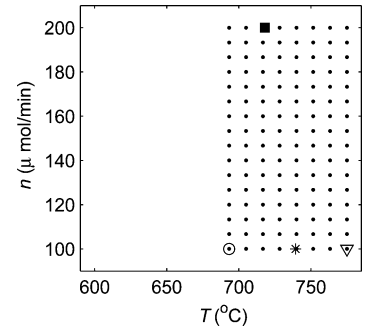

Figure 4. Results after the six initial experiments have been performed. The predicted optimal is shown with the ∗, and the next designed experiment (7) is marked with the filled square. For comparison, the D-optimal and P-optimal points are marked with the open triangle and circle, respectively. The grid points in the region of potential optima are marked with dots. Model 1 is used to compute the region and the D- and P-optimal experiments.

The results in Table 5 highlight an advantage of this experimental design approach. If one were performing a factorial design, after estimating the main effects, model 1 or 2 would be selected, and the rest of the experiments would be designed using one of these models. If a hybrid modeling approach was being taken, then model 6 would be chosen, and further experiments would be designed to improve the estimates of its parameters (if using D-optimal design). In the approach here, multiple models are specifically included because no model is perfect, and the most probable one may change as more data are collected. While it might be anticipated that the mechanistic model would provide a better regressor than the linear fit in model 1, in this case, the mechanistic model predicted a strong dependence on $\dot{n}$, while no clear dependence on $\dot{n}$ is evident in the experimental data. While this mechanistic model may be appropriate for a process that is limited by the transport of the precursor to the substrate, the data here suggest that the rate limiting step in the process may be surface reactions, which depend strongly on temperature.

Additionally, model 5—a constant—is considered as a guideline for the experimenter. If model 5 has a probability greater than or comparable to the probabilities of the other models being considered, then a new relationship between the settings should be considered, either for a modification to the current models or for a completely new model.

4.3. Sequential Experimental Design. Having selected the candidate models using the initial experimental data, the sequential experimental design was initiated. A confidence level of 95% was selected for the calculation of all confidence intervals, and all computations were performed using MATLAB. The previous simulation study by Wissmann and Grover[33] suggested that a uniform random selection of experiments from the region of potential optima was the best approach when significant modeling inaccuracies were present. Here, we also use a random selection from the region of potential optima. For comparison, we compute the D-optimal and P-optimal experiments at each iteration, based on the most probable model as determined by the AIC.

The designed experiment based on the initial experimental data is shown in Figure 4. Recall that model 1 is the most likely model after the initial six experiments, based on the AIC, and thus is used for designing the next experiment. The predicted

Table 6. Experimental Results after Initial Experiments

experiment	scaled T	scaled $\dot{n}$	roughness	T [°C]	$\dot{n}$ [μmol/min]
7	0.37	1.0	8.5 ± 0.5	720	200
8	0.6	0.46	8.8 ± 0.7	750	173

optimal point after the first iteration is shown by the $*$ symbol and is located at $T = 736$ °C. Any value of $\dot{n}$ could be used to match the desired roughness in model 1, so the minimum value of $\dot{n}$ is selected on the basis of the secondary criterion of minimizing process cost. The grid points in the region of potential optima are marked with dots and span approximately half of the design region. Because the region is calculated using a model with no dependence on $\dot{n}$, the region extends over all values of $\dot{n}$, but only the limited range of T surrounding the predicted optimum point. The next experiment (experiment 7), which is marked by the filled square in Figure 4, was randomly selected within the region of potential optima. The value of $\dot{n}$ is not fixed at the low level, so that further experimentation might uncover a trend in $\dot{n}$. For comparison, the D-optimal and P-optimal experimental points are also shown, but these experiments were not actually performed on the CVD system. Since they are also computed using model 1, any value of $\dot{n}$ could be used, but here they are shown at the minimum value of $\dot{n}$. Alternatively, they could be performed at random settings of $\dot{n}$, to further explore the dependence on $\dot{n}$. The most probable model at this step has a linear structure in T, which yields a D-optimal point (open triangle) at the highest value of T. Conversely, the P-optimal point (open circle) is at the minimum value in the region of potential optima. In our previous simulation study, we also found that the D-optimal and P-optimal points led to experiments concentrated at the boundaries of the domain.[33]

The settings for experiment 7 (the first sequentially designed experiment) and the measured roughness from experiment 7 are given in Table 6. After performing experiment 7, the parameter fits and model probabilities were recalculated and remained relatively unchanged, which indicates that the experiment confirmed the previously observed trends and did not provide much additional insight into the process. Remember, however, that the goal here is not to find the best or true model (model discrimination) nor to get the best parameter fits for the models (D-optimal) but rather to find the optimal point of the process. Looking at Figure 5, information has been added about the location of the optimal point. The predicted value is now $T =$ 746 °C, which has shifted to the right.

Figure 5 also shows the grid points that are potential optima, and this region is now smaller, compared to that in Figure 4,

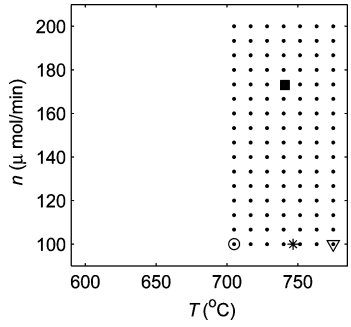

Figure 5. Results after the first sequential experiment has been performed. The predicted optimal is shown with the $*$, and the next designed experiment (8) is marked with the filled square. For comparison, the D-optimal and P-optimal points are marked with the open triangle and circle, respectively. The grid points in the region of potential optima are marked with dots. Model 1 is used to compute the region and the D- and P-optimal experiments.

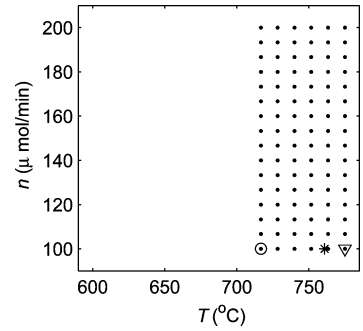

Figure 6. Results after the second sequential experiment has been performed. The predicted optimal is shown with the $*$. For comparison, the D-optimal and P-optimal points are marked with the open triangle and circle, respectively. The grid points in the region of potential optima are marked with dots. Model 1 is used to compute the region and the D- and P-optimal experiments.

Table 7. Confidence Intervals at the Predicted Optimal Point for Three Models in the Experimental Study As Well As the Confidence Interval on the Objective Function (CI_f) for Each Model

	10		14		15	
equation	CI_{RMS_1}	$CI_{f(x)}$	CI_{RMS_5}	CI_f	CI_{RMS_6}	CI_f
initial experiments	2.11	1.61	2.87	6.75	4.65	19.2
first iteration	1.72	1.15	2.32	5.18	3.96	16.0
second iteration	1.65	1.11	1.94	4.22	3.73	14.8

indicating that more confidence has been gained about the optimal process. The next experiment to be conducted (experiment 8) is plotted with the filled square, with the D-optimal and P-optimal experimental designs also shown for comparison. Note that both D- and P-optimal are both basically just replicates of the D-optimal and P-optimal experiments from the previous iteration. Only the P-optimal experiment has shifted slightly, due to the narrowing of the region of potential optima.

The second sequential experiment (experiment 8) was then performed, and the results are shown in Figure 6. The predicted optimal point is now located at $T = 760$ °C. The D-optimal and P-optimal points are shown in the figure and again are at the maximum and minimum temperatures, respectively. As with the previous experiment, the parameter estimates in the models do not change drastically. Improvement has come in the form of a tighter prediction on the optimal point, as can be seen in the shrinking region of grid points in Figure 6, as well as in Table 7 for models 1, 5, and 6. The confidence interval for each of the models' predictions is reduced as more experiments are performed, and one becomes more certain of where the true optimal point lies in the experimental space. The large confidence interval on the hybrid model (model 6) appears because we are calculating the confidence interval at the optimum temperature as predicted by model 1. When the confidence interval of model 6 is calculated instead at the predicted optimal of model 6, it is only slightly larger than the confidence intervals shown for model 1.

In this work, our primary goal was to evaluate the sequential experimental design method in an experimental system, and we did not directly apply the stopping criteria. However, we now compute the metrics previously suggested for the stopping criteria. After experiments 7 and 8, the sizes of the confidence intervals on RMS are 1.72 and 1.65 nm, respectively, which can be compared to a user specified value of ε_{CI} to determine if a desired confidence level on the film roughness has been achieved. For example, if $\varepsilon_{CI} = 1$ nm, then the experiments would continue. The change in MSE for each of the two iterations was 0.04 and 0.13, respectively, which can be

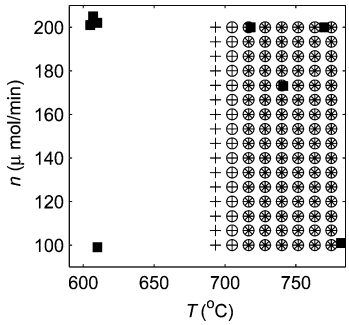

Figure 7. Points left in the grid after each experimental iteration. The grid points following experiment 6 are marked by +, after experiment 7 by ○, and after experiment 8 by ×. The eight experiments that were performed are marked by the filled squares.

compared against a stopping criterion of ε_{tol}. Note that the last experiment actually increased the value of the MSE for model 1.

After performing the eight designed experiments, significant understanding was gained about the process and its optimal settings. At all times, model 1 was the most significant, given the data, and it predicted that temperature can be used to achieve the desired film roughness of 7 nm. Figure 7 summarizes the results. The grid points at each of the three iterations are plotted, showing the gradual shrinking of the region of potential optima. The eight experiments performed are plotted with the filled squares. The final predicted value for the optimal value is somewhere between the high temperature setting, and the temperatures used for experiments 7 and 8. Thus, the knowledge gained by performing experiments in the vicinity of the predicted optimal is allowing us to refine our process temperature to achieve a desired film property, even in the presence of model inadequacy.

5. Conclusion

In this study, we implemented a recently proposed sequential experimental design method, by performing experiments on a chemical vapor deposition process. Using six initial experiments and two sequential experiments, we optimized the process settings to achieve a desired thin film surface roughness. Our results show that, when there is significant model inadequacy, the D-optimal and P-optimal experimental designs will not be desirable for locating the optimal process, since the experiments are only at the boundaries of the experimental domain. In our implementation, we use both empirical and hybrid models, since we do not know ahead of time which will be most useful for process optimization. In this study, the hybrid model was slightly less significant than one of the empirical models, so the hybrid model was not used for designing any of the experiments. However, by including the hybrid models, we were able to gain insight into the CVD process, concluding that the transport of the precursor to the substrate was not a rate limiting step in the evolution of the thin film roughness. This knowledge could be used to suggest a new mechanistic or hybrid model, which might ultimately be more valuable for additional experimental design. At the same time, additional empirical models could also be added, possibly including a higher-order dependence on temperature.

Another key feature in the experimental design method used here is the restriction of experiments to the region of potential optima. In the sequential experimentation, this region shrunk as additional experiments were performed. This was the intended effect, concentrating the experiments near the optimum, and further reducing the confidence interval at the predicted optimum. This idea could be further extended in the future to the parameter estimation portion of the method—the model parameters could also be fit using only the data that are in the current region of potential optima. Although the models would not have good global predictive capability, this is not required once the region of potential optima has been refined.

At the end of the eight experiments, the predicted optimal temperature was 760 °C. The predicted roughness at this temperature was 7 ± 1.65 nm, where this error is computed using a 95% confidence interval. Because the experimental data did not exhibit a clear trend in the mass flow rate of the precursor, this factor was not included in computing the process optimum, and the flow rate could be set to the low value to minimize process cost. In future application of this method, the effect of process uncertainty and noise could be included in the objective function, to identify major noise sources, such as the mass flow rate in this experimental data, and thus to create a robust process design.

Acknowledgment

This work was funded by the National Science Foundation. Discussions with Matthew Realff and J. C. Lu are gratefully acknowledged.

Literature Cited

(1) Murarka, S.; Gutmann, R.; Kaloyeros, A.; Lanford, W. *Thin Solid Films* **1993**, *236*, 257–266.

(2) Green, M.; Gusev, E.; Degraeve, R.; Garfunkel, E. *J. Appl. Phys.* **2001**, *90*, 2057–2121.

(3) Wahl, G.; Metz, C.; Samoilenkov, S. *J. Phys. IV* **2001**, *11*, 835–846.

(4) Haynes, J.; Pint, B.; More, K.; Zhang, Y.; Wright, I. *Comput. Chem. Eng.* **2002**, *58*, 513–544.

(5) Will, J.; Mitterdorfer, A.; Kleinlogel, C.; Perednis, D.; Gauckler, L. *Solid State Ionics* **2000**, *131*, 79–96.

(6) Liu, Y.; Zha, S.; Liu, M. *Adv. Mater.* **2004**, *16*, 256–260.

(7) Meng, G.; Song, H.; Dong, Q.; Peng, D. *Solid State Ionics* **2004**, *175*, 29–34.

(8) Seshan, K. *Handbook of thin-film deposition processes and techniques: principles, methods, equipment and applications*; Noyes Publications/William Andrew Pub.: Norwich, NY, 2002; Vol 2.

(9) Edgar, T. F.; Butler, S. W.; Campbell, W. J.; Pfeiffer, C.; Bode, C.; Hwang, S. B.; Balakrishnan, K. S.; Hahn, J. *Automatica* **2000**, *36*, 1567–1603.

(10) Montgomery, D. C. *Design and Analysis of Experiments*; John Wiley & Sons, Inc: New York, 2005; Vol 6.

(11) Henson, M. A. *Comput. Chem. Eng.* **1998**, *23*, 187–202.

(12) Liu, D.-J.; Weeks, J. D.; Johnson, M.; Williams, E. D. *Phys. Rev. B* **1997**, *55*, 7653–7659.

(13) Zhang, J.; Adams, J. *Comput. Mater. Sci.* **2004**, *31*, 317–328.

(14) Hukka, T. I.; Pakkanen, T. A.; D'Evelyn, M. P. *J. Phys. Chem.* **1995**, *99*, 4710–4719.

(15) Southwell, R.; Mendicino, M.; Seebauer, E. *J. Vacuum Sci. Technol. A* **1996**, *14*, 928–934.

(16) Kennedy, A.; Marc, C.; O'Hagan, A. *J. R. Stat. Soc. B* **2001**, *63*, 425–464.

(17) Prasad, V.; Vlachos, D. G. *Ind. Eng. Chem. Res.* **2008**, *47*, 6555–6567.

(18) *International Technology Roadmap for Semiconductors*; Technical Report, ITRS, 2007.

(19) Topol, A. W.; Dunn, K. A.; Barth, K. W.; Nuesca, G. M.; Taylor, B. K.; Dovidenko, K.; Kaloyeros, A. E.; Tuenge, R. T.; King, C. N. *J. Mater. Res.* **2004**, *19*, 697–706.

(20) Robbins, J. J.; Harvey, J.; Leaf, J.; Fry, C.; Wolden, C. A. *Thin Solid Films* **2005**, *473*, 35–40.

(21) Bao, T.; Morrison, P. W.; Woyczynski, W. A. *Thin Solid Films* **2005**, *485*, 27–41.

(22) Li, J. F.; Liao, H. L.; Ding, C.; Coddet, C. *J. Mater. Proc. Technol.* **2005**, *160*, 34–42.

(23) Coumes, C. C. D.; Courtois, S. *Chemom. Intell. Lab. Syst.* **2006**, *80*, 167–175.
(24) Nyutu, E. K.; Suib, S. L. *Surface Coatings Technol.* **2006**, *201*, 2741–2748.
(25) Nourbakhsh, A.; Ganjipour, B.; Zahedifar, M.; Arzi, E. *Nanotechnology* **2007**, *18*, 1–7.
(26) Box, G. E. P.; Liu, P. Y. *J. Quality Technol.* **1999**, *31*, 1–15.
(27) Chernoff, H. *Ann. Math. Stat.* **1959**, *30*, 755–770.
(28) Kiefer, J.; Sacks, J. *Ann. Math. Stat.* **1963**, *34*, 705–750.
(29) Daniel, C. *J. Am. Stat. Assoc.* **1973**, *68*, 454–360.
(30) Frey, D. D.; Engelhardt, F.; Greitzer, E. M. *Res. Eng. Des.* **2003**, *14*, 65–74.
(31) Jones, D. R.; Schonlau, M.; Welch, W. J. *J. Global Optimization* **1998**, *13*, 455–492.
(32) Henkenjohann, N.; Gobel, R.; Kleiner, M.; Kunert, J. *Qual. Reliability Eng. Int.* **2004**, *21*, 439–455.
(33) Wissmann, P. J.; Grover, M. A. *AIChE J.* **2009**, *55*, 342–353.
(34) Pulver, M.; Nemetz, W.; Wahl, G. *Surf. Coat. Technol.* **2000**, *125*, 400–406.
(35) Wissmann, P. J. Simultaneous Model Building and Process Design using Experimental Design: Application to Chemical Vapor Deposition, Thesis, Georgia Institute of Technology, Atlanta, Georgia, 2008.
(36) Pomfret, M. B.; Stoltz, C.; Varughese, B.; Walker, R. A. *Anal. Chem.* **2005**, *77*, 1791–1795.
(37) Pasko, S.; Hubert-Pfalzgraf, L. G.; Abrutis, A. *Mater. Lett.* **2005**, *59*, 1836–1840.
(38) Zhang, Y.; Edgar, T. F. *Ind. Eng. Chem. Res.* **2008**, *47*, 7772–7783.
(39) Burnham, K. P.; Anderson, D. R. *Model Selection and Inference: A Practical Information-Theoretic Approach*; Springer-Verlag: New York, 1998.
(40) Box, G. E. P.; Hill, W. J. *Technometrics* **1967**, *9*, 57.
(41) Montgomery, D. C.; Runger, G. C. *Applied Statistics and Probability for Engineers*, 4th ed.; John Wiley & Sons: Hoboken, NJ, 2007.
(42) Evans, J.; Pa, T.; Bartelt, M. *Surf. Sci. Rep.* **2006**, *61*, 1−128.
(43) Aspnes, D.; Theeten, J.; Hottier, F. *Phys. Rev. B* **1979**, *20*, 3292–3302.
(44) Chang, C.; Kryder, M. *J. Appl. Phys.* **1994**, *75*, 6864–6866.

Received for review July 10, 2009
Revised manuscript received February 19, 2010
Accepted April 27, 2010

IE901055E