THE INSTITUTE FOR SYSTEMS RESEARCH

ISR TECHNICAL REPORT 2008-12

Full Wafer Mapping and Response Surface Modeling Techniques for thin Film Deposition Processes

Leon, Maria del Pilar and Adomaitis, Raymond

Full wafer mapping and response surface modeling techniques for thin film deposition processes

María del Pilar León and Raymond A. Adomaitis[1]
Department of Chemical and Biomolecular Engineering and Institute for Systems Research, University of Maryland, College Park, MD, 20742, USA

Abstract

Computational techniques for representing and analyzing full wafer metrology data are developed for chemical vapor deposition and other thin-film processing applications. Spatially resolved measurement data are used to produce "virtual wafers" that are subsequently used to create response surface models for predicting the full-wafer thickness, composition, or any other property profile as a function of processing parameters. Statistical analysis tools are developed to assess model prediction accuracy and to compare the relative accuracies of different models created from the same wafer data set. Examples illustrating the use of these techniques for film property uniformity optimization and for creating intentional film-property spatial gradients for combinatorial CVD applications are presented.

1 Introduction

Response surface models are used extensively in chemometrics [3] and in semiconductor processing applications [8]. The primary motivation for creating a response surface model is to be able to predict a system's output corresponding to a given set of conditions or to determine the necessary conditions that would yield a desired response. There are three steps involved in obtaining a response surface model. First, a series of experiments must be designed and performed in such a way that adequate and reliable measurements of the response of interest are gathered. Then, a mathematical model is fitted to the data through regression analysis (usually the least squares method). Finally, the empirical model is subjected to error analysis and statistical testing to verify its validity.

In this paper, we present a numerical procedure for generating response surface models for thin-film deposition processes where the model prediction corresponds to the full spatial profile over the substrate. This work is motivated by the increasing availability of full-wafer metrology methods and the potential of wafer-profile modeling software to make full use of such measurements. We demonstrate the utility of our full wafer modeling approach for both uniformity optimization and for chemical vapor deposition single-wafer combinatorial processing.

1.1 Combinatorial CVD

In semiconductor device manufacturing processes such as chemical vapor deposition (CVD), one typically strives for uniform film properties (thickness, microstructure, composition, etc.) across

[1]Corresponding author, adomaiti@umd.edu

the substrate (wafer). In some cases, however, it is desirable to control deposition rate or a film property profile to a specific, spatially non-uniform distribution [1]. Such situations arise in single-wafer combinatorial processing where the objective is to deposit a film with properties that vary as a function of location on the wafer.

Combinatorial CVD is an emerging technology that enables efficient development of new materials and the corresponding process recipes for this crucial step in semiconductor manufacturing. The objectives of combinatorial CVD are to (reproducibly) deposit intentionally non-uniform films across the substrate and to have the ability to model the deposition system to sufficient degree that gas phase conditions above any point of the wafer can be determined to correlate processing conditions to desired film qualities [9].

There are relatively few examples of CVD reactor systems with combinatorial capabilities. Those that do exist, however, all demonstrate the capability to produce films with graded properties over a portion of the substrate surface. For example, the CVD reactor design of Gladfelter [11, 16] features three feed tubes in a triangular arrangement across the substrate; a different single-source precursor is fed through each tube, generating compositional spreads of three metal dioxides over the substrate. In Wang [13, 14, 15], thickness graded films of hydrogenated silicon were deposited in a hot-wire CVD system featuring a mask and motorized shutter; control of the shutter speed was used to create strips of graded films over the substrate. Hyett and Parkin [7] describe a cross-flow reactor configuration where separate precursor inlet nozzles are used to produce films with composition gradients of tungsten and titanium oxides and to likewise generate films with varying ratios of TiO_2 rutile/anatase phases. In Taylor and Semancik [12], microhotplate devices were used to control the temperature in an array of micro-scale substrate samples; it was found that temperature gradients in the microhotplate supports resulted in a microstructurally graded film on the support legs. Finally, the Programmable Reactor system [5, 6, 9, 10] features a segmented shower head design where each segment is fed individually with reactant gases and exhaust gas is pumped back up through each segment. This concept was tested in a three-zone prototype tungsten deposition system to evaluate the system's ability to manipulate gas phase composition across the wafer surface [5, 6] and to demonstrate its true programmable nature using a model-based approach to controlling spatial deposition patterns across the wafer surface [9, 10].

Given the promise of the emerging field of combinatorial CVD and the continuing importance of uniformity control in thin film processing, the goal of this work is to develop the numerical techniques and associated software for interpreting, modeling, and optimizing the film properties across a variety of substrate geometries to make full use of the capabilities of next generation CVD reactor design.

2 Virtual wafers

Full wafer metrology data, such as film thickness, composition, dopant concentration, phase composition, etc., generally are measured using a uniform grid or as a set of uniformly spaced points on a line across the wafer surface. A representative sample of tungsten film thickness as determined using four-point probe measurements [6] is shown in Fig. 1.

Because our goal is to model and analyze full wafer data in a computationally accurate manner, using the data in its original, uniformly discretized form is inappropriate. Consider, for example, the problem of accurately determining the rotationally averaged wafer profile – for this problem, ex-

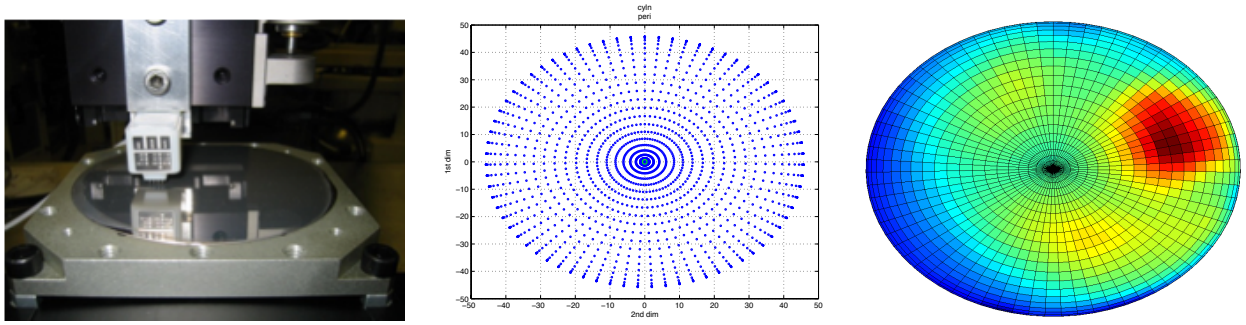


Figure 1: *Film thickness measured using four-point probe measurements (left); quadrature grid onto which the data are to be interpolated (center); "virtual wafer" map (right) as defined on the quadrature grid.*

pressing the wafer profile in polar coordinates *and* at Gaussian quadrature points makes it possible to perform calculations to within the roundoff error of the computer used. Furthermore, because the quadrature weights w_k used for integration operations need only be calculated once, the resulting computational procedure is very efficient. Likewise, subsequent interpolation operations necessary to extract wafer spatial data (points, lines, and subregions) and the statistical analysis of wafer spatial features also can be done in a very precise manner.

The wafer data are interpolated to the quadrature grid by first defining the quadrature points (r_k, θ_k) as roots of a polynomial taken from a sequence of orthogonal polynomials in the radial (r) direction and equidistant points in θ (see [2] and Fig. 1). For each (r_k, θ_k), the nearest experimental measurement points are located and the measured values at those points are used to interpolate the corresponding value Y_k at the quadrature point. It is important to point out that values of our measured film property Y are now defined in an optimal manner by interpolating polynomials between points (r_k, θ_k), and so the virtual wafer map actually defines a function $Y(r, \theta)$ (Fig. 1), resulting in the accurate calculations described previously.

The virtual wafer generation operations are all carried out using the object-oriented features of MATLAB[2]. In our computational approach, we define a MATLAB class called `scalarfield` [2] in which computational objects containing the Y_k values, quadrature weights w_k, and quadrature point locations (r_k, θ_k) are defined; discrete differential and other operators likewise are set up. Wafer maps in this form then are combined with additional wafer information, such as the wafer ID and wafer processing conditions $x_1, x_2, \ldots, x_m$ inside objects representing the parameterized data (`paramdata`), a class that allows for efficient sorting of large sets of wafer data as well as the development of response surface models. Ultimately, the user needs only to choose the wafer map object class to implement all of these numerical procedures. In this paper, we discuss two: `wafer`, which describes one- and -two-dimensional wafer maps on circular substrates, and `rectwafer` which also represents one- and two-dimensional data but on rectangular wafers. Other wafer map classes, such as those appropriate for planetary reactor systems, are under development and will be discussed in subsequent publications. All of the methods described are collected in the graphical user interface `waferview.m` application shown in Fig. 2.

[2] www.mathworks.com

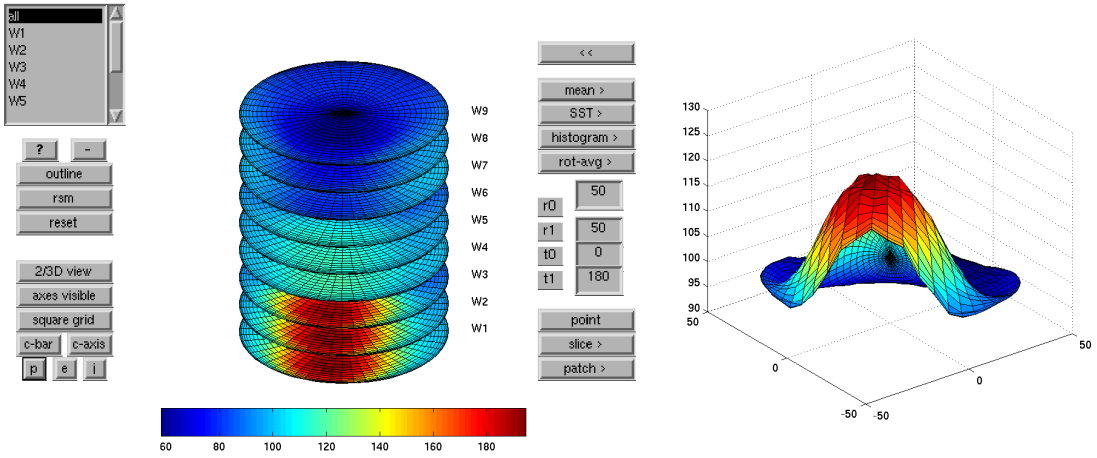


Figure 2: *Graphical user interface to the wafer visualization and characterization methods illustrating a stack of 9 wafers (left) and the mean thickness profile of the set of wafers (right).*

3 Building response surface models

Given a vector of wafer map objects $Y = [Y_1, Y_2, \ldots, Y_n]$ and corresponding operating conditions $x = [x_1, x_2, \ldots x_p]$, we turn to the problem of creating a model to predict the full wafer spatial features given a user-defined set of operating conditions. To facilitate the model derivation description, the variable and parameter definitions are summarized in Table 1.

x_1 to x_m	process inputs
$Y_1(r,\theta)$ to $Y_n(r,\theta)$	experimental data represented by virtual wafers
$Y_{i,k}$	wafer measurement at point k on virtual wafer i
$\beta_1(r,\theta)$ to $\beta_p(r,\theta)$	model parameter true values
$b_1(r,\theta)$ to $b_p(r,\theta)$	model parameter values estimated by regression
$y(r,\theta)$	model prediction
(r_1,θ_1) to (r_q,θ_q)	quadrature point locations on the wafer surface

Table 1: *Response surface model notation.*

The linear least squares method is used to compute estimates $b_j(r,\theta)$ of the parameters $\beta_j(r,\theta)$ of a fitted model when these *spatially dependent* parameters are coefficients of functions of only the independent variables (process inputs $x_1, \ldots, x_m$). In other words, the observed response $Y_i(r,\theta)$ is linear in $\beta_j(r,\theta)$ but not necessarily in the independent variables (i.e., operating conditions). A single observation may be represented as

$$Y_i(r,\theta) = \sum_{j=1}^{p} \beta_j(r,\theta) f_j(x_{1,i}, x_{2,i}, \ldots, x_{m,i}) + \epsilon_i(r,\theta) \quad i = 1, \ldots n$$

where f_j's are the functions of the independent variables related to each parameter $\beta_j(r,\theta)$ and $\epsilon_i(r,\theta)$ is the random error of the measurement. The totality of observations is expressed in matrix

form as

$$Y(r,\theta) = X\beta(r,\theta) + \epsilon(r,\theta).$$

The result of applying the linear least squares technique to this system is a set of equations that are expressed in matrix form as

$$X'Xb(r,\theta) = X'Y(r,\theta)$$

where b is the vector of the β estimates. The values for the b vector are easily obtained from this expression as

$$b(r,\theta) = (X'X)^{-1}X'Y(r,\theta).$$

While standard in its appearance, we emphasize that the implementation of the least squares technique is new in that the least squares method is performed in every point (r_k, θ_k), $i = 1, \ldots, q$ of the quadrature grid of the virtual wafers. As a result, the estimated parameters b_j are functions of r and θ instead of scalar values and a full wafer map is obtained from the response-surface model. A representative response-surface model object is displayed using the `wafermodelview.m` interface graphical user interface shown in Fig. 3.

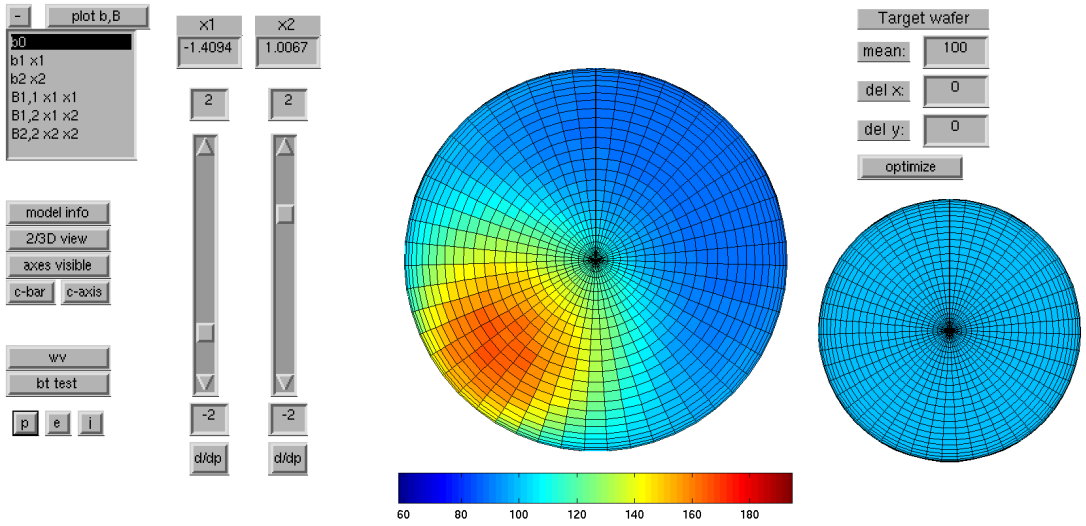


Figure 3: *Graphical user interface to the wafer response surface model illustrating a sample of the model predition (center) corresponding to process conditions set by the slider bars (left), and the target wafer to be used for optimization (right).*

4 Error analysis and statistical testing

Model validity is assessed through an analysis of variance that divides the total variability of a set of data into meaningful components. For instance, the total variation, known as the total sum of squares (SST) can be partitioned into two contributions: the sum of squares of variability explained by the fitted model, known as the sum of squares due to regression (SSR), and the sum of squares

unaccounted for by the fitted model, known as the sum of squares of residuals (SSE):

$$SST = SSR + SSE.$$

For the full-wafer model analysis, the total sum of squares is computed as the weighted average of the sum of the squares of the deviations of the observed responses $Y_{i,k}$ about their average $\overline{Y_k}$ for each point (denoted with the subscript k) of the quadrature grid

$$SST = \frac{\sum_{k=1}^{q}\sum_{i=1}^{n} w_k (Y_{i,k} - \overline{Y_k})^2}{\sum_{k=1}^{q} w_k} = \frac{\sum_{k=1}^{q} w_k sst_k}{\sum_{k=1}^{q} w_k} \tag{1}$$

where the w_k are the quadrature weights associated with the $k'th$ point of the quadrature grid, q is the total number of points in the grid, and sst_k is the total sum of squares of the $k'th$ individual point. By the nature of the discrete-ordinate collocation formulation, the *sst* are actually true functions of r and θ over the substrate physical domain. The number of degrees of freedom associated to SST is $n-1$, where n is the total number of observations (i.e., total number of wafers used to generate the model).

The sum of squares due to regression is calculated by the weighted sum of the squares of the difference between the model prediction $y_{i,k}$ and the average of the observed values for each point of the grid:

$$SSR = \frac{\sum_{k=1}^{q}\sum_{i=1}^{n} w_k (y_{i,k} - \overline{Y_k})^2}{\sum_{k=1}^{q} w_k} = \frac{\sum_{k=1}^{q} w_k ssr_k}{\sum_{k=1}^{q} w_k}$$

There are $p-1$ degrees of freedom associated to SSR, where p is the number of parameters in the fitted model.

Lastly, the sum of squares of residuals is the weighted average of the sum of the squares of the difference between the observed values and the predicted values.

$$SSE = \frac{\sum_{k=1}^{q}\sum_{i=1}^{n} w_k (Y_{i,k} - y_{i,k})^2}{\sum_{k=1}^{q} w_k} = \frac{\sum_{k=1}^{q} w_k sse_k}{\sum_{k=1}^{q} w_k}$$

The degrees of freedom associated to SSE are $n-p$. Again, we see that *ssr* and *sse* are true functions of r and θ.

A measure of the adequacy of a model is the ratio between the sum of squares due to regression and the sum of squares of residuals divided by their respective degrees of freedom:

$$f_0 = \frac{SSR/(p-1)}{SSE/(n-p)}$$

The value of f_0 is compared to a tabulated value $F_{\alpha,p-1,n-p}$ that represents the upper $100\alpha\%$, with $\alpha \in [0,1]$, of the F-distribution. If f_0 is greater than $F_{\alpha,p-1,n-p}$ the variation accounted for by the model is significantly greater than the unexplained variation, thus the model is considered adequate. However, the possibility that another model is a better fit to the data is not rejected. A "better" model may include other variables or the deletion of one or more of the variables considered in the model.

4.1 Test of the individual parameters

The least squares method assumes that the estimated parameters $b_j(r,\theta)$ are normally distributed with the mean being the true value of the parameter $\beta_j(r,\theta)$. The variances of the estimated parameters $\sigma^2_{b_j}(r,\theta)$ are defined by the diagonal elements of the inverse of the $X'X$ matrix (used to calculate the b_j's) and the model variance $\sigma^2(r,\theta)$. The model variance is estimated by the ratio of the sum of squares of residuals and its respective degrees of freedom

$$\sigma^2(r,\theta) \approx \frac{sse(r,\theta)}{(n-p)}$$

which is computed from the values of *sse* at the collocation points.

Having knowledge of the distribution of the parameters makes it possible to test hypotheses about parameter values. For example, it can be tested whether or not β_j equals a set value β_{o_j} using a t-test

$$t_o = \frac{b_j - \beta_{o_j}}{s_{b_j}}$$

where s_{b_j} is the estimate of the standard deviation σ_{b_j} for parameter b_j and is calculated by taking the square root of the variance of b_j

$$\sigma_{b_j} \approx s_{b_j} = \sqrt{[X'X]^{-1}_{j,j}\sigma^2}.$$

The null hypothesis ($H_o : \beta_j = \beta_{o_j}$) is accepted if $t_o \in [-t_{\alpha/2,n-p}, t_{\alpha/2,n-p}]$ for more than $\phi \in [0,1]$ weighted percentage of the quadrature grid points[3]. The value $t_{\alpha/2,n-p}$ is obtained from tables (or calculated using appropriate software) and depends on α that is defined as the level of significance of the test (an α level of significance corresponds to a $100(1-\alpha)$% level of confidence).

A special case is when is when $\beta_{o_j} = 0$ in which case, if the null hypothesis is accepted, the parameter β_j must be removed from the model and a new model calculated. These numerical procedures are implemented in our MATLAB function `bttest.m` and will be described in the applications that follow this derivation.

4.2 Model comparison

More than one model can be fitted to the same set of data and it is important to compare these models to determine the one that represents the data most accurately. One way to make this decision is based on the F-distribution.

Both the SSE of a model ($n-p$ degrees of freedom) and the SSE of the difference of two models ($p_1 - p_2$ degrees of freedom) follow a χ^2 distribution. By definition, the ratio of two χ^2 distributions divided by their respective degrees of freedom follows the F-distribution. Therefore, the following ratio should follow the F distribution with $(p_1 - p_2)$ and $(n - p_1)$ degrees of freedom in the numerator and denominator, respectively

$$\frac{(sse_{p_2} - sse_{p_1})/(p_1 - p_2)}{sse_{p_1}/(n-p_1)} = F(\alpha, p_1 - p_2, n - p_1)$$

[3]The value of ϕ is selected by the user; typically, values greater that 0.5 are chosen, with a default value of 0.75 (75%).

where subscript 1 refers to the model with a larger number of parameters ($p_1 - p_2$ must be positive). Rearranging to obtain a ratio between sse_{p_2} and sse_{p_1}

$$\begin{aligned} \text{critical ratio} &= (p_1 - p_2)\frac{F(\alpha, p_1 - p_2, n - p_1)}{n - p_1} + 1 \\ \text{actual ratio} &= \frac{sse_{p_2}}{sse_{p_1}} \end{aligned}$$

If the critical ratio is larger than the actual ratio it can be concluded that adding the extra $p_1 - p_2$ terms to the model with larger number of parameters does not represent a significant improvement of the fit.

In determining which model is "best" for accurate prediction of the full wafer map, two approaches may be taken: 1) comparing the spatially weighted average of the actual ratio sse_{p_2}/sse_{p_1} to the critical ratio value, and 2) comparing the critical ratio value to the ratio of the overall sum of squares SSE_{p_2}/SSE_{p_1}. As presented in the second case study that follows, the first approach will be used for model comparison in this paper.

5 Case studies

To demonstrate the use of the wafer mapping, model building, and statistical analysis tools, we present three test cases that illustrate elements of each.

5.1 Uniformity optimization

In the first example, we examine the problem of building a response surface model that predicts the full-wafer film thickness profile as a function of the deposition reactor inputs. For this example, we consider a chemical vapor deposition process and a goal of manufacturing a spatially uniform film with a mean thickness of 100mn. To test the wafer modeling and optimization software, we use wafer thickness map data that is created artificially using the following equation

$$Y(r,\theta) = \beta_0 + \beta_1(r,\theta)(x_1 - 1) + \beta_2(r,\theta)(x_2 + 1) + \beta_{1,2}(r,\theta)(x_1 - 1)(x_2 + 1) + \epsilon(r,\theta). \quad (2)$$

In this system there are two process inputs x_1 and x_2, which represent operating parameters such as heater power input and reactant gas flow. We note that because β_0 is a constant, the process at $x_1 = 1$, $x_2 = -1$ produces perfectly uniform films except for the noise term $\epsilon(r,\theta)$.

Using the process inputs x_1 and x_2, we perform a full-factorial design of experiments with $x_1 \in \{-2, 0, 2\}$ and $x_2 \in \{-2, 0, 2\}$, thus creating a total of 9 individual wafer profile data sets. These data are used as input to the `wafer.m` constructor method, creating the set of 9 wafer map (virtual wafer) objects plotted in Fig. 4.

We observe in the collection of film thickness profiles, that there is significant spatial variation among the wafers; in fact, the wafers produced by this design of experiments feature film thicknesses ranging from 85nm to 110nm. Computing $sst(r,\theta)$ (equation 1), we plot the variation of the data in Fig. 5 to illustrate the magnitude of the uniformity optimization problem.

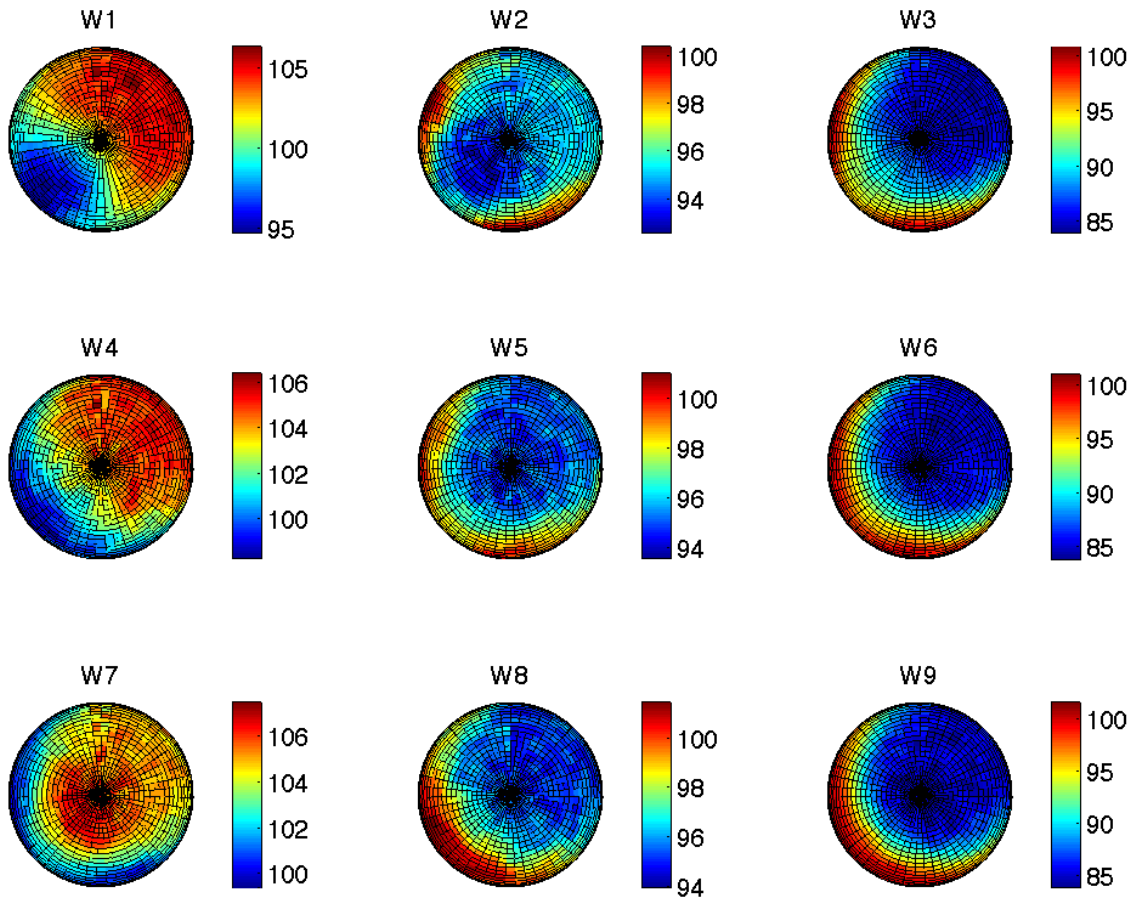


Figure 4: *Plots of the 9 data wafers for the uniformity optimization demonstration.*

In the first step of the model development and subsequent optimization, we use the nine wafer maps as input to our response surface model constructor method `rsmodel.m`, which by default will attempt to identify a second-order model of the form

$$y(r,\theta) = b_0(r,\theta) + b_1(r,\theta)x_1 + b_2(r,\theta)x_2 + b_{1,2}(r,\theta)x_1x_2 + b_{1,1}(r,\theta)x_1^2 + b_{2,2}(r,\theta)x_2^2$$

The model constructor method successfully identifies an acceptable model for this case; we then proceed to apply to null hypothesis to each of the identified parameters using `bttest.m`, which correctly identifies that parameters $b_{1,1}$ and $b_{2,2}$ should be eliminated from the model (c.f., eqn 2).

We now define an optimization setpoint corresponding to a uniform film thickness $y_{set} = 100nm$ and perform an optimization with respect to the two input parameters x_1 and x_2 to determine the set producing a wafer that satisfies the following minimization problem

$$\min_{x_1,x_2} ||y_{set}(r,\theta) - y(r,\theta)||^2 = \min_{x_1,x_2} \int_0^{2\pi} \int_0^{R_w} (100nm - y\,(r,\theta))^2\, r dr\, d\theta$$

subject to $x_1 \in [-2,2]$ and $x_2 \in [-2,2]$, and where R_w is the wafer radius.

This constitutes a constrained optimization problem that is efficiently solved using the `fmincon.m` function of the MATLAB optimization toolbox. Results are shown in Fig. 5 using the optimized

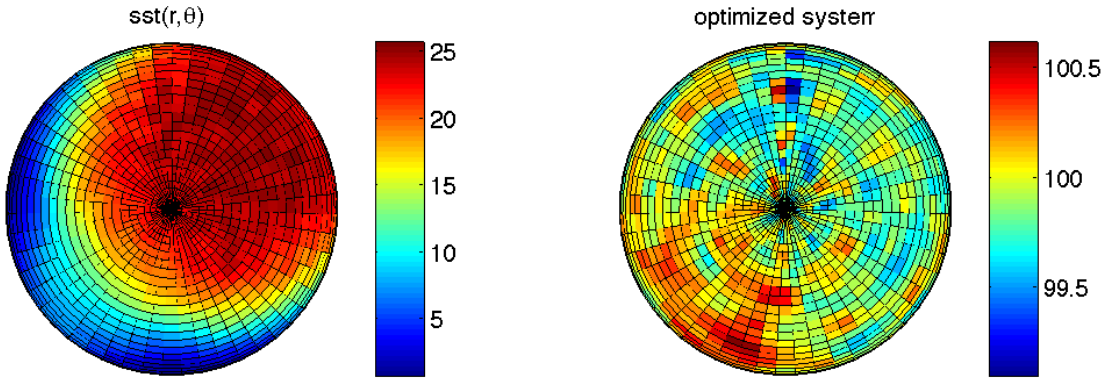


Figure 5: *A plot of sse*(r, θ) *for the demonstration data (left) and results produced by the thickness optimization procedure (right).*

values for $x_1 = 1.0062$ and $x_2 = -1.0017$. Note that the only variations in the optimized wafer correspond to the measurement noise of the original data.

5.2 Model comparison for the spatially programable CVD reactor

In the second example we consider data produced by the Programmable CVD reactor system [5, 6, 9, 10], a reactor designed to test spatially controlled CVD concepts. The main feature of this system is its segmented showerhead design that allows independent control of feed gas composition to each segment. Furthermore, rather than drawing residual gas over the wafer surface, exhaust gas is pumped back up through each segment, resulting in periodic gas velocity fields over the wafer surface. This design enables placement of the showerhead arbitrarily close to the substrate, and results in diffusion being the dominant across-wafer gas species transport mechanism.

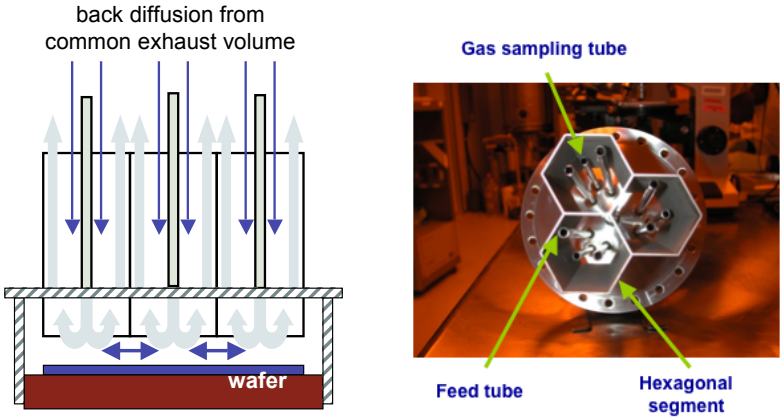


Figure 6: *Illustration of segmented design (right) and three-segment implementation (right).*

Experimental data were obtained from a 3-segment prototype in which blanket tungsten (W) de-

position was performed using the hydrogen reduction process:

$$WF_6(g) + 3H_2(g) \rightarrow W(s) + 6HF(g).$$

Film thickness maps were obtained using a 4-point probe (see Fig. 1); a sequence of deposition runs was performed in which the segment recipes were varied as well as the showerhead/substrate gap size. The specific experiments performed and the resulting wafer maps are shown in Fig. 13.

The motivation for developing a response surface model based on the data collected is to determine the segment recipes and gap size need to achieve a certain deposition pattern over the wafer surface. In [10], an empirical model was developed using physically motivated arguments based on inter- and intra-segment gas phase transport of gas phase reactants. The model then was used to demonstrate the ability to program the reactor system to produce spatially graded films over a subregion of the wafer surface for combinatorial CVD applications. The model developed in the cited study had the form

$$M_1: \quad y(r,\theta) = b_1(r,\theta)x_1 + b_2(r,\theta)x_2 + b_3(r,\theta)x_3 + b_{1,4}(r,\theta)x_1 g + b_{2,4}(r,\theta)x_2 g + b_{3,4}(r,\theta)x_3 g$$

where the subscripts $i = 1,2,3$ denote the segment number, and the double subscripts $(i,4)$ denote the segment feed/gap interaction terms (i.e., 4 represents the fourth model input which is gap size). In this paper we seek to demonstrate that this model is indeed a better fit to the data than the simpler linear model

$$M_2: \quad y(r,\theta) = b_0 + b_1(r,\theta)x_1 + b_2(r,\theta)x_2 + b_3(r,\theta)x_3$$

by statistically comparing them.

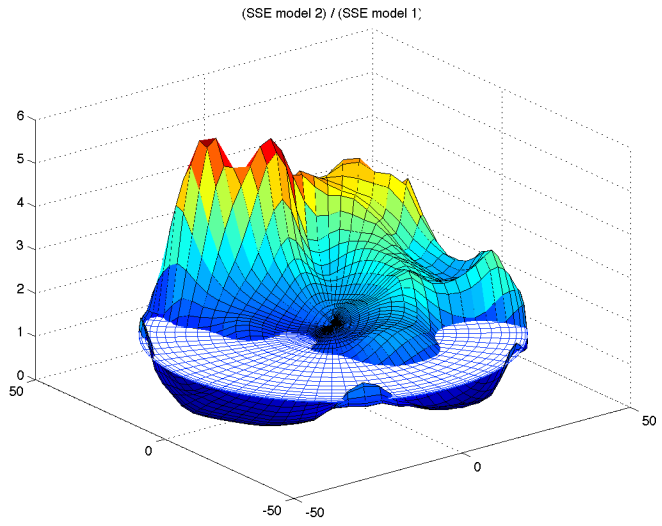


Figure 7: *The ratio of the linear model* $sse(r,\theta)$ *(*M_2 *or model 2) and that of* M_1 *(model 1) plotted as a solid surface, as compared to the critical ratio value (plotted as the flat mesh surface).*

To make this comparison, we generate both models using the wafer data set listed in Fig. 13 (note that we do not use the validation data in producing the models). Comparing the $sse(r,\theta)$ of the models by computing the ratio of each (linear model M_2 over the full model M_1), we see in Fig. 7 that while approximately half of the ratio per unit of substrate area fall to each side of the critical

ratio, the portion above the critical ratio corresponding to the linear model M_2 has much greater magnitude relative to that of the nonlinear model M_1. Integrating over the substrate we find the weighted average actual ratio to be 1.74 as compared to the critical ratio of 1.37 for this case. From this analysis we conclude that M_1 is more likely to provide a more accurate prediction of the film thickness over the entire wafer. This conclusion is supported by a comparison of both model predictions to the validation data set in Fig. 8, where we clearly observe the superior performance of the nonlinear model (M_1).

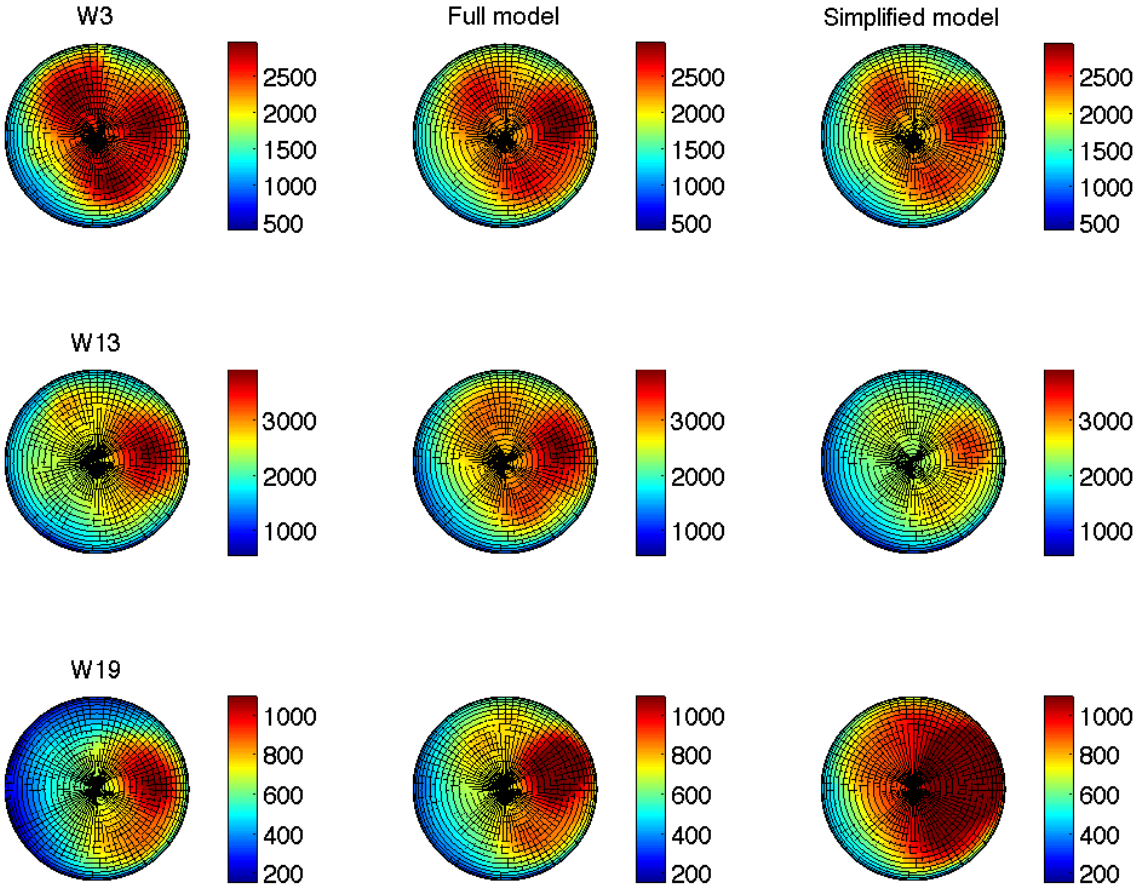


Figure 8: *A plots of validation wafer data (left column) versus predictions of model with segment recipe setting/gap interaction terms M_1 (center) and the simplified linear model M_2 (right).*

5.3 Gradient control for combinatorial CVD

In the third example, we consider the problem of determining operating conditions best suited for the combinatorial investigation of thin films deposited in a CVD reactor. For this example, we deposit a compound semiconductor $A_xB_{1-x}C$ in a rectangular, duct-flow reactor system where the inlet is split in half transverse to the total flow direction in the reactor, where the species A precursor is injected through one of the half-width ports while the species B precursor is fed through the other (Fig. 9). This reactor configuration is patterned after the design of [7]. As is typically the case with these hot-wall MOVPE systems, deposition takes place after the thermal decomposition of the

precursors and occurs on all surfaces of the reactor as well as the substrate itself.

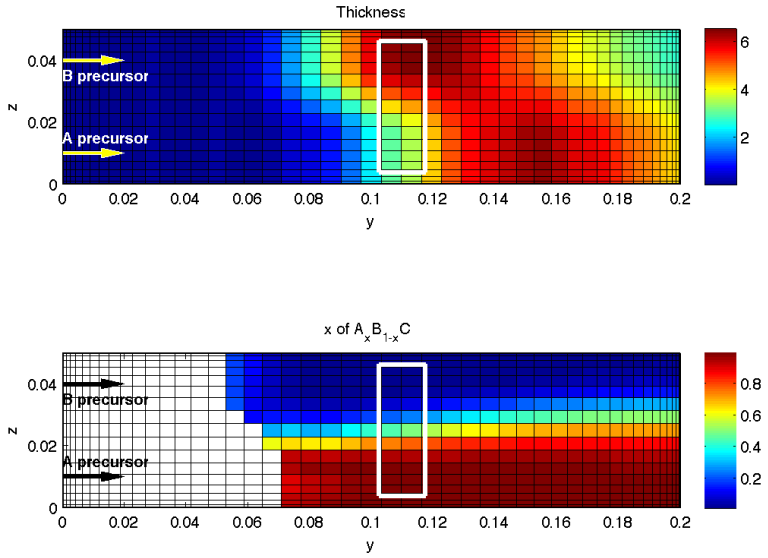


Figure 9: *Plots of total film thickness (top) and the fraction x of species A in the $A_xB_{1-x}C$ film (bottom); we note that film composition is not displayed in the inlet region corresponding to extremely thin or nonexistent films. The test substrate strip area is outlined in white.*

Our goal is to examine films of varying composition in terms of the A/B ratio, represented by the subscript $x \in [0,1]$ over a narrow strip of the substrate with local spatial coordinates (y_s, z_s); this strip is positioned with its major axis perpendicular to the direction of gas flow. To achieve this combinatorial goal, we wish to deposit a graded film of the form

$$x = (z_{max} - z_s)/z_{max} \quad \text{with } z_s \in [0, z_{max}]$$

where z_s is the substrate strip coordinate transverse to the flow path and z_{max} is the width of the strip we consider. Together with the composition set point profile above, a second objective is to maintain as uniform a thickness profile as possible in the test-strip region. As can be seen in Fig. 9, under a nominal set of operating conditions, both the film composition and thickness can vary substantially as a function of 2-dimensional position over the reactor system.

To carry out this study, we consider a full factorial design over the reactor operating conditions, setting the dimensionless carrier gas flow rate $V_o \in \{0.8, 1, 1.2\}$ and dimensionless substrate temperature $T_s \in \{875, 900, 925\}/900$ (dimensional temperature is in K). Resulting film profiles over the test strip are shown in Fig. 10 demonstrating the considerable variability in composition and thickness profiles that are possible, where the latter are all adjusted to have zero mean. Because of the multi-objective nature of the optimization criterion (a specified gradient in composition *x and* thickness uniformity), our approach is to examine the response surface curves as a function of the two operating parameters generated from the data given in Fig. 10. Using both the thickness and composition profile data sets, we generate a 2nd-order response surface model for each; checking the individual b_j using `bttest.m` we find a suitable response surface model for predicting the test strip thickness profile as

$$W_t(y_s, z_s) = b_0(y_s, z_s) + b_2(y_s, z_s)T_s + b_{1,2}(y_s, z_s)V_oT_s + b_{2,2}(y_s, z_s)T_s^2$$

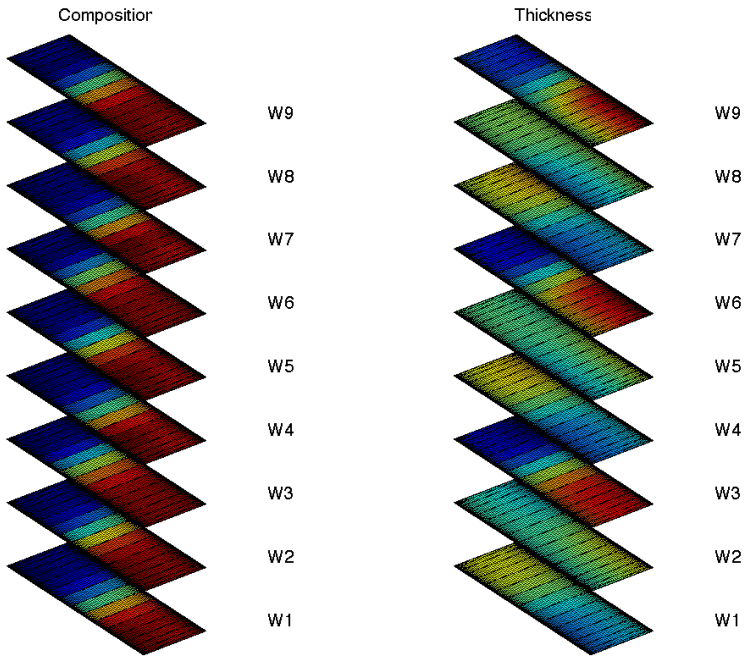


Figure 10: *Composition (left) and thickness (right) spatial patterns generated from the design of experiments for the combinatorial CVD example.*

and the composition profile as

$$W_x(y_s, z_s) = b_0(y_s, z_s) + b_1(y_s, z_s)V_o + b_2(y_s, z_s)T_s + b_{1,2}(y_s, z_s)V_oT_s + b_{2,2}(y_s, z_s)T_s^2.$$

The two objective functions to be minimized are

$$G_t = ||W_t||^2 \qquad G_x = ||W_x - (z_{max} - z_s)//z_{max}||^2.$$

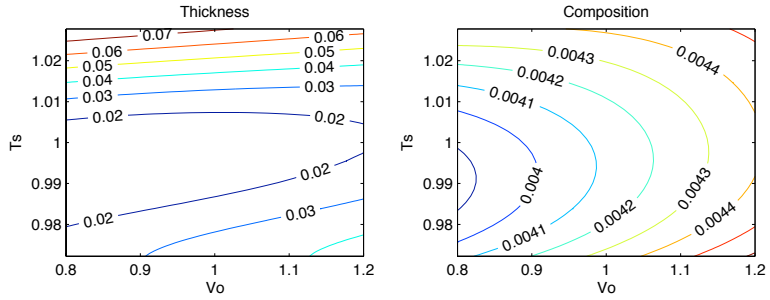


Figure 11: *Response surfaces illustrating optimization criteria G_t (left) and G_x (right) as a function of V_o and T_s.*

Contour plots of the two objective function values as a function of the operating parameter values are shown in Fig. 11 and are generated using our `modelsurf.m` method. From these plots, we observe that optimal composition gradient is found at $V_o = 0.8$ and $T_s \approx 0.99$ and the most uniform thickness profile (minimum value of G_t) is found nearly at the same value. Using these parameter

values in the *simulator used to generate the original data*, the resulting composition and thickness profiles of each are shown in Fig. 12. Note that the thickness and composition range scales of Fig. 12 are comparable to those of Fig. 9, indicating the optimized operating conditions do indeed result in deposition conditions favorable for combinatorial studies.

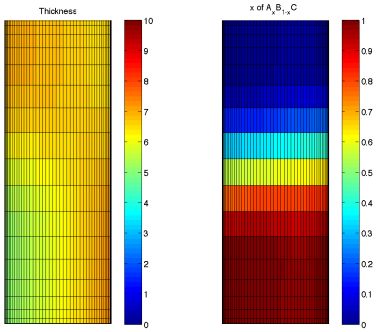


Figure 12: *Thickness (left) and composition (right) gradients optimized for combinatorial operations; conditions correspond to* $V_o = 0.8$ *and* $T_s = 0.99$.

6 Conclusions

In this paper, we presented the development and implementation of a library of computational tools for representing, analyzing, and modeling thin-film properties as a function of spatial position across the substrate surface. While presented in the context of CVD examples, we note that the numerical techniques can be readily applied to a wide range of deposition and etching operations.

Whether the ultimate goal of a thin-film process is to deposit or etch films in a spatially uniform or intentionally non-uniform manner (e.g., for combinatorial operations), we see the need for computational methods to 1) identify the desired spatial profile, and 2) adjust the processing conditions to achieve that goal. A framework for coupling process models to numerical optimization methods is necessary for this paradigm of thin-film processing, and the software developed in this paper was shown to be effective in addressing this need.

Acknowledgments

The authors acknowledges the support of the National Science Foundation through grant CTS-0554045.

References

[1] Adomaitis, R. A. “Intentionally Patterned and Spatially Non-Uniform Film Profiles in Chemical Vapor Deposition Processes,” Surface and Coatings Technology **201** (2007) 9025-9029.

[2] Adomaitis, R. A. “Objects for MWR,” *Computers & Chem. Engng* **26** (2002) 981-998.

[3] Box, G. E. P. and N. R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley Series in Probability and Mathematical Statistics, J. Wiley, 1987.

[4] Cai, Y., L. Henn-Lecordier, G. W. Rubloff, R. Sreenivasan, J.-O. Choo, and R. A. Adomaitis, "Multiplexed Mass Spectrometric Sensing in a Spatially Programmable Chemical Vapor Deposition System," *J. Vac. Sci. Tech. B* **25** (2007) 1288-1297.

[5] Choo, J. -O., R. A. Adomaitis, G. W. Rubloff, L. Henn-Lecordier, and Y. Liu "Simulation-Based Design and Experimental Evaluation of a Spatially Controllable Chemical Vapor Deposition Reactor," *AIChE Journal* **51** (2005) 572-584.

[6] Choo, J.-O., R. A. Adomaitis, L. Henn-Lecordier, Y. Cai, and G. W. Rubloff, "Development of a Spatially Controllable Chemical Vapor Deposition Reactor with Combinatorial Processing Capabilities," *Rev. Sci. Instr.* **76** (2005) 062217.

[7] Hyett, G. and I. P. Parkin, "A combinatorial approach to phase synthesis and characterization in atmospheric pressure chemical vapour deposition," *Surf. & Coatings Tech.* **201** (2007) 8966-8970.

[8] May, G. S. and C. S. Spanos *Fundamentals of Semiconductor Manufacturing and Process Control* IEEE, Wiley-Interscience (2006).

[9] Sreenivasan, R. R. A. Adomaitis, and G. W. Rubloff "A Demonstration of Spatially Programmable Chemical Vapor Deposition: Model-Based Uniformity/Non-uniformity Control." *J. Vac. Sci. Tech. B* **24** (2006) 2706-2715.

[10] Sreenivasan, R., R. A. Adomaitis, and G. W. Rubloff "A Comparative Study of Reactor Designs for the Production of Graded Films with Applications to Combinatorial CVD," *J. Crystal Growth* **310** (2008) 270-283.

[11] Smith, R. C., N. Hoilien, J. Roberts, S. A. Campbell, and W. L. Gladfelter, Combinatorial chemical vapor deposition of metal dioxides using anhydrous metal nitrates, *Chem. Mater.* **14** (2002) 474-476.

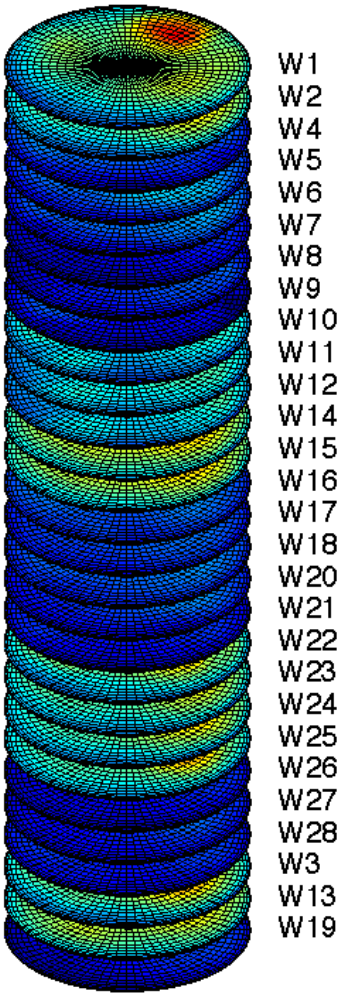
[12] Taylor, C. J. and S. Semancik, Use of microhotplate arrays as microdeposition substrates for materials exploration, *Chem. Mater.* **14** (2002) 1671-1677.

[13] Wang, Q., J. Perkins, H. M. Branz, J. Alleman, C. Duncan, and D. Ginley, Combinatorial synthesis of solid state electronic materials for renewable energy applications, *Appl. Surf. Sci.* **189** (2002) 271-276.

[14] Wang, Q., Combinatorial hot-wire CVD approach to exploring thin-film Si materals and devices, *Thin Solid Films* **430** (2003) 78-82.

[15] Wang, Q., F. Liu, and D. Han, High-throughput chemical vapor deposition system and thin-film silicon library, *Macromol. Rapid Commun.* **25** (2004) 326-329.

[16] Xia, B., R. C. Smith, T. L. Moersch, and W. L. Gladfelter, Balancing reactor fluid dynamics and deposition kinetics to achieve compositional variation in combinatorial chemical vapor depositions, *Appl. Surf. Sci.* **223** (2004) 14-19.



wafer ID	H2s1 (sccm)	H2s2 (sccm)	H2s3 (sccm)	gap (mm)
W1	16	32	48	3
W2	48	16	32	3
W4	32	0	0	3
W5	0	32	0	3
W6	0	0	32	3
W7	0	0	32	1
W8	0	32	0	1
W9	32	0	0	1
W10	48	16	32	1
W11	32	48	16	1
W12	16	32	48	1
W14	32	48	16	5
W15	32	32	32	5
W16	32	0	0	5
W17	0	32	0	5
W18	0	0	32	5
W20	0	32	0	3
W21	32	0	0	3
W22	32	32	32	3
W23	48	16	32	3
W24	32	48	16	3
W25	16	32	48	3
W26	0	0	32	2
W27	0	32	0	2
W28	32	0	0	2
W3	32	32	32	3
W13	16	32	48	5
W19	0	0	32	3

Figure 13: *Operational conditions for each wafer generated using the Programmable Reactor system; upper (larger) set corresponds to training data and lower (W3, W13, and W19) set corresponds to validation data.*