AFRL-SR-BL-TR-98-

0185

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Real Time Control and Experimental Verification of Semiconductor Processing

5. FUNDING NUMBERS
F49620-93-1-0524

6. AUTHOR(S)
Kostas Tsakalis
Michael Kozicki

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Arizona State University
Box 87-1603
Tempe, Arizona 85287-1603

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
110 Duncan Avenue Room B115
Bolling AFB, DC 20332-8080

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This work demonstrated that wafer-scale parameters can be modeled and controlled. The modeling goal was representation of whole-wafer parameter trends using a minimal number of measurement sites. In such a way, useable wafer area for devices is maximized. The aim of the control was the attainment of a process parameter goal in a minimum number of process steps using an approximate model of the process.

A second-order polynomial model based upon a central composite design was created to map parameters on a wafer surface. It has been shown to adequately represent oxide thickness profiles created from a rapid thermal processor and measured by an ellipsometer. These models are then used with a run-to-run adaptive controller, applied to the oxidation process, to obtain a goal mean oxide thickness while simultaneously striving for optimum uniformity. Results have shown satisfactory control convergence, within an error margin of about 2%, in the mean oxide thickness parameter of the rapid thermal oxidation step after about four iterations.

19980223 122

14. SUBJECT TERMS
Feedforward and adaptive feedback control to semiconductor processing and device manufacturing

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

REAL TIME CONTROL AND EXPERIMENTAL VERIFICATION OF SEMICONDUCTOR PROCESSING

FINAL REPORT

This Final Report for the AFOSR-supported project "Real Time Control and Experimental Verification of Semiconductor Processing" is based on the following published works:

K. Stoddard, "Application of Feed-Forward and Adaptive Feedback Control to Semiconductor Device Processing," Master of Science Thesis, Department of Electrical Engineering, Arizona State University, 1994.

K. Stoddart, P. Crouch, M. Kozicki and K. Tsakalis, "Application of Feed-Forward and Adaptive Feedback Control to Semiconductor Device Manufacture", Proceedings of the American Control Conference, Baltimore, 892- 896 (1994).

D. L. Herald, "Control of Wafer-Scale Electrical Parameters," Master of Science Thesis, Department of Electrical Engineering, Arizona State University, 1997.

These publications are readily available and therefore are not reproduced as part of this report. However, the basic concepts developed in this program and results obtained are summarized below.

1. INTRODUCTION

For a typical production sequence consisting of numerous steps, the resulting silicon wafers contain a large number of devices. Regardless of the nature of the device, certain parameters will vary from one device to another across the wafer surface. These variations are largely due to imperfections in the nature of the fabrication process. For example, oxide growth may not be exactly uniform from one device to another due to uneven heating of the wafer surface during processing. Doping may vary as a function of position due to the nature of the ion implantation process. Variation in a process parameter is undesirable since it results in a deviation from a resulting device parameter goal, and too much deviation can cause a device to function improperly or fail.

Often, silicon-based devices, such as microprocessors, are mass-produced. Methods for determining product quality are statistical in nature [1]. A large quantity of product is produced and then studied to determine how many devices pass specifications. A small random sample of product determines statistically how much of the product will be suitable, and how much will fail. The process is adjusted such that the statistical failure number is suitably small. This approach is suitable for applications where a large quantity of product is desired. However, another approach, which focuses on each silicon wafer individually, is preferred for applications where only a small number of devices are required, or where each device has substantial cost, and using a bulk statistical methodology is not cost effective. Control techniques, such as multi-sensor real-time control of rapid thermal processing [2], and run-to-run adaptive control of the

same [3], are especially useful.

The control studied in this work focuses on each wafer individually. If one wafer, the statistical rarity, does start to drift out of specifications at one process step, the control will automatically make adjustments in order to pull the wafer back into specifications, in succeeding process steps. Control which feeds current wafer-status information forward to the next process step, in order for corrections to be made which pull the wafer back into specifications, is called feed-forward control. The results of the last wafer processed through a given process step, fed back to that same step for the current wafer, uses adaptive feedback control.

Parameters vary across the wafer surface, and control is needed to ensure that each wafer stays within specified limits. For some devices, the varying wafer parameters may determine whether or not a certain device on the wafer passes specifications. Therefore, a model which predicts the nature of the varying parameter is desired. The model is then inserted into control formulations which use this information to ensure that the maximum number of devices per wafer are within specified limits. Statistical methods and a least-squares approach are used with a measurement point distribution known as the central composite design [5], [6]. Similar work in this area can be found in [7], where many models are generated simultaneously using a methodology known as "Multiple Response Surfaces", and in [8], where statistical design and modeling is applied to the plasma etching process. Statistical methods used in engineering applications can be found in sources such as [6], [9], [10], and [11], and further examples of semiconductor research which utilizes statistical methods such as design of experiments (DOE) can be found in [12], [13], and [14].

2. WAFER MODEL DEVELOPMENT

Model Requirements

The first criterion for a useable model is that the model-building process should use a minimum number of sites on the wafer. In a real-world application of such a model, test sites have to be made on a wafer in order to extract data to satisfy the model. Typically, these test sites prohibit the inclusion of an electrical device at that site. Hence, wafer area that can be used for product devices is lost. There are two ways test sites can be implemented. One method is by use of a single test-wafer, where a very large number of test sites are used, and no devices are made on that wafer. This approach is more in keeping with the large-scale bulk manufacture of silicon-based devices, where the focus is not on each individual wafer but on the overall yield of wafers. This would not be desirable for this project. For this project, each wafer is itself a test-wafer. While this allows for information on a wafer-to-wafer basis, it also obviously uses up valuable wafer area. A model selected for use in this project must therefore provide enough information to be useful, while at the same time use the smallest set of measurement points on the wafer surface. This is therefore the first major model constraint in this project.

The second constraint for consideration of model choice is that of a measure of merit for the model. Obviously there needs to be a type of rating, or figure of merit, to determine whether a given model adequately represents the pattern of data on the wafer surface. In conjunction with the measure of merit, the user of the model needs to determine how much model error is acceptable. Too much model error, and the model is useless. A poor model with a large error in data fit will predict values of the needed parameter that are too far away from actual values, and

that will ultimately cause difficulty in control. A model that is too extensive could require more data points, produce unnecessary precision (for example, obtaining "exact" representations of the wafer surface, although the error in measurement exceeds the precise fluctuations of the mapping, therefore making those variations meaningless) or be unusually cumbersome or complex. So the upper limit of acceptable error is thus user-defined, while the lower limit is defined by error in the instruments used to gather the data for the model fit, repeatability of experiments, and other noise.

The third major constraint for model choice involves the data on the wafer surface that is being modeled. Experiments need to be run in order to determine the nature of variation of this data. Sharp variations, or data distributed which has more than one maximum, minimum, or several maximums and minimums may need to be modeled differently than data which has gradual variation, or just one maximum/minimum.

In plots of parameters such as oxide thickness and resistivity, it is important to note that there is one significant maximum or minimum. Thus there is one trend, with either increasing or decreasing values from that min or max. Also, no sudden, sharp changes in the surface are noted. This makes intuitive sense, because natural processes often involve gradual changes. Sharp changes are less frequently observed. To verify this information, observations were made on many wafers by the author. The raw data verify the gradual change assumption and the single trend assumption.

Central Composite Design

The model choice that has been selected for this research uses the least-squares method and a fitted polynomial equation [9], [10]. This method provides all of the required elements of a model for this project. Any number of data points may be used, depending upon the desired error margin for model fit and the nature of the experimental data. There is a measure of model fit that is widely used in statistical methods texts that is easily applied to this model. And a wide variety of data surfaces can be modeled with the use of a polynomial and the least-squares methodology. As a further bonus, the equation itself is easily determined by any number of mathematical software packages, such as Matlab, Mathcad, or statistical packages, and the statistical element is also quite common and easy to compute with available software. Finally, this method has been used successfully in industry for several decades [6].

Due to the nature of the observed data, it is clear that at least a second-order polynomial must be used to represent this data distribution. A second-order polynomial has the ability to represent a single peak or trough, as well as various rotations and placements of this surface. Equation 2.1 shows the equation of the second-order model chosen for use in this research.

$$Y = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \qquad (2.1)$$

When (2.1) is plotted in three dimensions, with x and y the dependant variables, it can be seen that the first coefficient shifts the resulting surface up or down. The second and third coefficients provide for linear relationships along the x and y directions. The fourth and fifth coefficients allow for curvature in both the x and y directions. Finally, the sixth term takes into

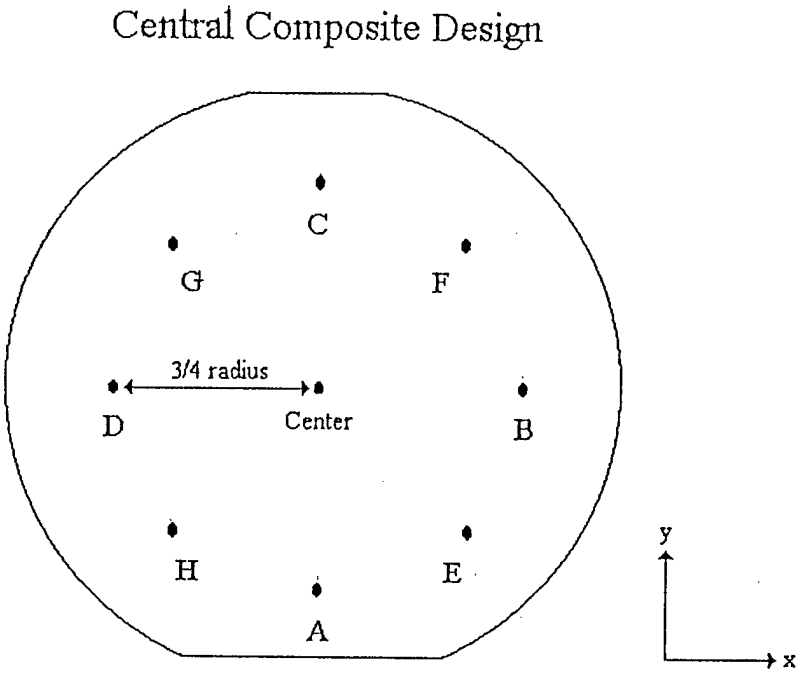


Fig. 2.1 Data gathering scheme for central composite design.

which will provide the ability for changes of curvature and more than one maximum or minimum. These qualities in the model equation are not needed as most real variations are best represented by (2.1) and the addition of these extra terms require more data to be gathered in order to adequately fit the equation, which is undesirable in our current scheme.

The basic data-gathering scheme for this design is shown in Fig. 2.1. The Central Composite Design accommodates a second-order equation, and provides enough data to fit the six regression coefficients, $\beta_0 - \beta_5$, of (2.1). The theoretical basis for this design, and the advantage of this allocation of data points over any other, can be seen in [5]. There are a total of nine data points. As shown, a coordinate system is established for this model. It is a rectangular coordinate system with the origin placed in the center of the wafer. This system was chosen for ease of use with process-line measurements. There is a data point placed at the center and four "axial" points (points located on the x and y axes). These axial points are located at a distance of three-quarters of a wafer radius from the center point of the model. This is a rule-of-thumb choice which allows for adequate model coverage. The remaining four points are located the same distance from the center as the axial points, but at an angle of 45 degrees from each neighboring axis. The center point measurement is repeated five times. This ensures that information about the resulting surface is estimated with a constant variance at fixed radial distances from the design center. This property makes the design "rotatable", and applies to any design where the plot of the variance function for the design results in concentric circles, spheres, or hyperspheres centered at the origin of the design. Such a design is preferred in situations where nothing is assumed in advance about the nature of the data to be described [5].

Use of Central Composite Design

The following section demonstrates the use of the Central Composite Design applied to oxide thickness and wafer resistivity data from processing. Statistical analysis is also briefly explained and used for determination of the goodness of model fit.

For most implementations of this design, the nine data points represent nine different combinations of varying levels of two parameters. For example, in the paper by Guo and Sachs [7], a nine-point system similar to the Central Composite, known as the 3^2 design, is used for variations in two gas flow rates in an LPCVD reactor, and the output measurement is uniformity of deposited polysilicon. However, for resistivity and oxide thickness profiles generated in this research, the two parameters are actually positions on a wafer surface, not controls to a process.

The least squares methodology is used to obtain the coefficients for (2.1). The following derivation is similar to that in [10]. Let z denote a vector of 13 data measurements (the nine sites, plus four additional center measurements), let **X** denote a matrix of the independent variables, let β denote a vector containing the six regression coefficients to be determined, and ε a vector of 13 random errors. Equation (2.2) results.

$$z=\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{13} \end{bmatrix}, \quad X=\begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & xy_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & xy_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{13} & y_{13} & x_{13}^2 & y_{13}^2 & xy_{13} \end{bmatrix}, \tag{2.2}$$

$$\beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{bmatrix}, \quad \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{13} \end{bmatrix}$$

Equation (2.1) can be rewritten in matrix notation, from (2.2), as

$$z=X\beta+\varepsilon \tag{2.3}$$

This equation contains the "actual" β's which provide the best possible fit to the assumed second-order model. This vector cannot be obtained exactly, but must be estimated from the supplied data. The errors represent the differences between the actual model and the measured z

values. Regression coefficients are sought which will minimize the ε's in (2.3). The methodology usually employed is to minimize the sum of squared ε's:

$$L = \sum_{i=1}^{13} \varepsilon_i^2 = \varepsilon'\varepsilon = (z - X\beta)'(z - X\beta) \qquad (2.4)$$

The minimization is achieved by taking the following derivatives:

$$\frac{\partial L}{\partial \beta} = 0 \qquad (2.5)$$

The equations resulting from (2.5) can be represented by the following matrices:

$$X'X\hat{\beta} = X'z \qquad (2.6)$$

Then (2.6) is solved for β:

$$\hat{\beta} = (X'X)^{-1}X'z \qquad (2.7)$$

The "hat" notation indicates that the equation is based upon estimated parameters, and is not the actual model for the process. Equation (2.7) is therefore the desired vector of least-squares estimated regression coefficients for the model equation (2.1). Equation (2.7) has been solved for directly as written using Matlab, and using the statistical analysis tool in Microsoft Excel.

What remains is to use statistical analysis methods on (2.7) to determine which terms in the suggested model of (2.1) should be retained, and which are candidates for withdrawal. The tests that have been used in this research are known as the t-test, the F-test, the p-statistic, and the adjusted R^2. While these tests are by no means perfect indicators of model goodness, they provide the user with educated guesses for how to proceed to a more optimum model.

There are underlying assumptions which the statistical methods are based upon, and which will be mentioned briefly. First, the ε's in (2.3) are assumed to be independent and normally distributed, with mean of zero and variance some σ^2. This means that run-to-run variation at any given point on the wafer surface follows a normal distribution centered upon the model's prediction. Also, the variance of ε across the region of interest (all values of wafer position of interest) is constant. These two conditions can be checked after obtaining a model, by analyzing the residuals to observe if they have an approximately normal distribution centered around the model prediction (normality check), and by noticing if the magnitude of the residuals become

larger for some values of the independent variables (non-constant variance).

If the residuals show a definite pattern, in other words not simply random variation about the model, then the most likely cause is improper model choice. In other words, there are terms that need to be added to the model to take into account those variations. Alternatively, the random error does not follow a normal distribution. However, even if this is the case, moderate departures from normality can still give meaningful statistical test results [9].

Models resulting from the above least-squares analysis can be seen in Figs. 2.2 and 2.3 (wafer oxidation using the Tamarack Model 180-M Rapid Thermal Processor). The Excel statistical analysis of the resulting models are in Table 2.1. Figs. 2.4 and 2.5 show plots of the residuals of a typical wafer oxidation model fit. As can be seen in the plots of residuals, there is no clear pattern that would suggest an incorrect model. Also, there is nothing to suggest that the variance in the error of the model changes significantly along any particular direction, since the magnitude of the residuals shows no pattern. So it seems that the assumptions stated above are valid.

The next step is to analyze the second-order model to determine the goodness of fit to the data. In Table 2.1, the fit for oxidation data for wafer #6 is analyzed by Microsoft Excel's regression tool. The R-square parameter suggests a poor fit to the data, and the adjusted R-squared parameter, a more accurate measure of model goodness, is even more pessimistic. In order to get an idea for why this parameter is so low, one needs to examine the contribution of each term of the polynomial to the model fit. The P value is very useful for this. The P-value essentially describes the probability for rejection of the parameter of interest. As can be seen in Table 2.1, the x and y parameters show P-values of 0.23 and 0.67, respectively. This indicates, in effect, that there is a high probability that these parameters don't contribute to the overall model. To refine this model further, these two terms should be removed, one by one, and the R^2 and P-values re-examined. The t-statistic provides similar information to the P, in that if the value is too small, it argues for rejection of the inclusion of a particular model parameter. In order to determine the critical t-value, below which the parameter should be rejected, statistical tables are consulted. In addition to the P and t-tests, the F-statistic provides at a glance an indication whether or not the overall model fits in any reasonable fashion. The "significance F" field in the statistical printouts indicates the probability that none of the parameters is significant, or in other words, the probability that regression is insignificant.

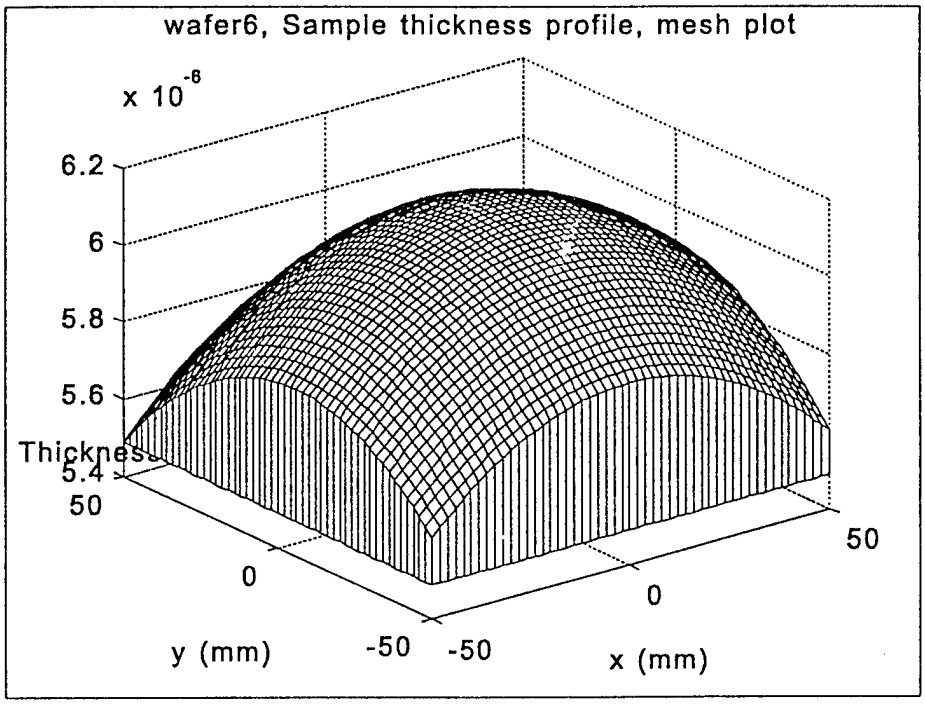


Fig. 2.2 Mesh plot of oxide thickness following rapid thermal oxidation on wafer #6 fitted to the second order polynomial of eqn. (2.1).

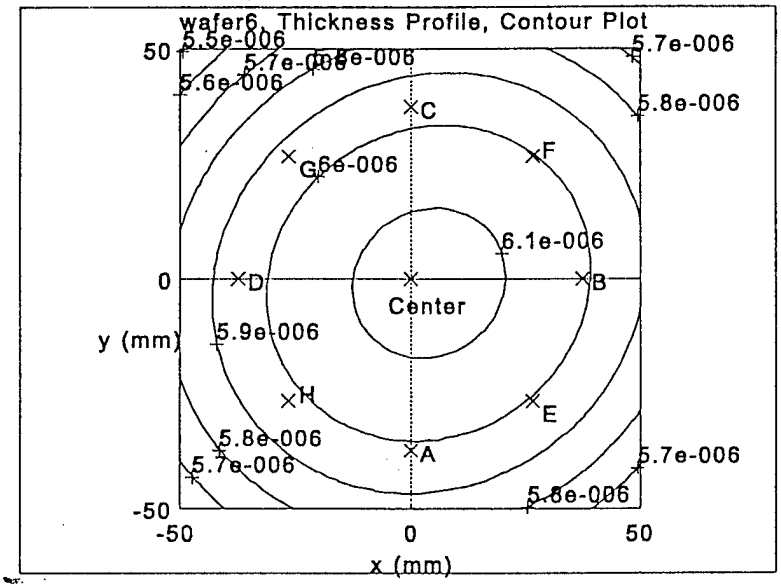
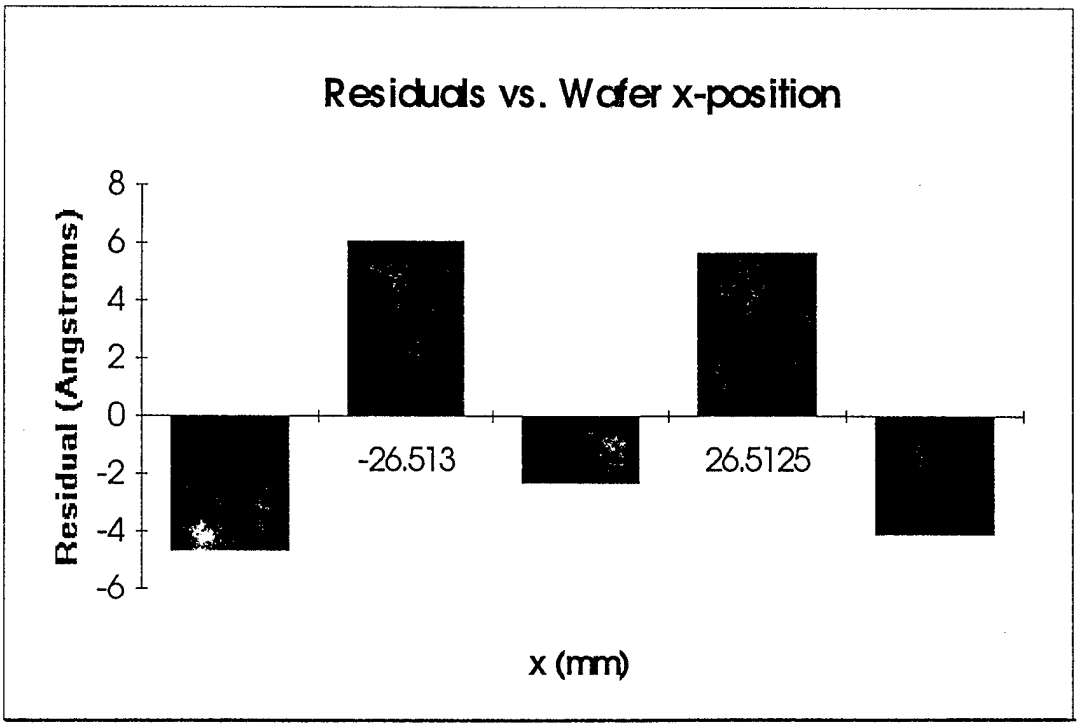


Fig. 2.3. Contour plot of oxide thickness following rapid thermal oxidation on wafer #6 fitted to the second order polynomial of eqn. (2.1). The letters represent actual measurement sites.

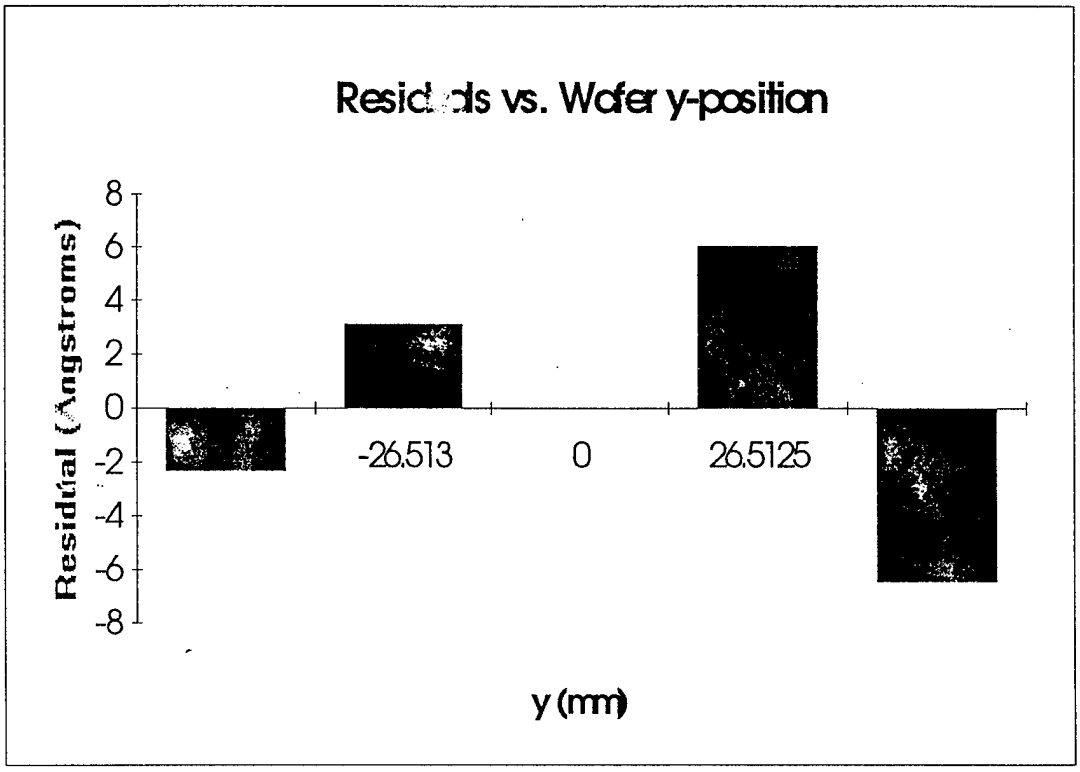
Regression Statistics	*(wafer 6 oxide)*					
Multiple R	0.902347013					
R Square	0.814230132					
Adjusted R Square	0.681537369					
Standard Error	4.948906677					
Observations	13					
Analysis of Variance						
	df	*Sum of Squares*	*Mean Square*	*F*	*Significance F*	
Regression	5	751.4298897	150.2859779	6.13620603	0.016978272	
Residual	7	171.4417411	24.4916773			
Total	12	922.8716308				
	Coefficients	*Standard Error*	*t Statistic*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	612.7150596	2.213218309	276.8434805	3.32054E-24	607.4816337	617.9484856
x	0.084198781	0.065990405	1.275924597	0.226126245	-0.071843619	0.240241181
y	-0.028826553	0.065990405	-0.436829459	0.669991971	-0.184868953	0.127215847
x-squared	-0.010418132	0.00266881	-3.903661549	0.002096144	-0.016728862	-0.004107403
y-squared	-0.011026132	0.00266881	-4.13147841	0.001391829	-0.017336862	-0.004715403
x*y	0.001860117	0.003520286	0.528399538	0.606856414	-0.00646403	0.010184264

Table 2.1. Statistical analysis of the model for oxide thickness.



In the examples shown here, regression is significant, with a confidence of at least 98%.

Although the value of the R-squared statistic in Table 2.1 is unfavorable, this is not the norm for the majority of the models that have been generated. Therefore, although in this specific case one could argue for exclusion of the x and y terms in the polynomial



based on the t statistic and P value, in general it is best to start with the full model of eqn. (2.1), and then eliminate terms if desired.

3. CONTROL METHODOLOGY

Control Concepts

Using the wafer model studied in the previous section, control equations are now developed. One of the goals of the control algorithm is to ensure that the largest area of a wafer surface contains devices with parameters within an acceptable range. Another desired result is insensitivity to noise (machine drifts, measurement errors, etc.). The control used in this research, in conjunction with the whole-wafer model developed in this work, strive to attain these goals.

An overall conceptual sketch of the process being undertaken is shown in Fig. 3.1. This is essentially the diagram originally shown in [15]. Two control concepts that are evident in this figure are feed-forward and adaptive feedback control. For a procedure which involves several different processes (such as thermal oxidation, etching, deposition, etc), each process could be represented by this diagram, and each process flows into the next process as shown. Measured results from each process are fed back into the controller as well as being fed into the next process.

Feed forward control involves taking knowledge from previous processes in order to predict what the output for the current process should be. For example, if in the previous process the result was below the desired goal for that process, the feed-forward controller can attempt to compensate by manipulating the desired result of the current process.

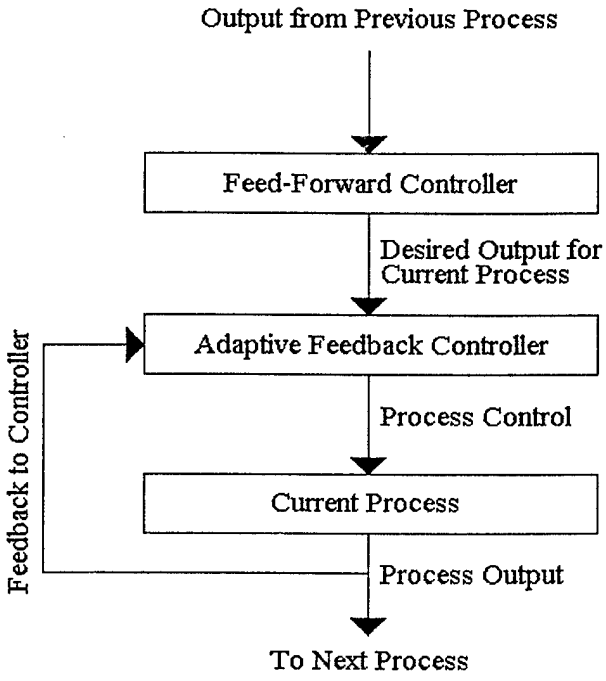


Fig. 3.1. Control flow diagram for a given step in a multi-step process.

The adaptive feedback controller attempts to tune the current process to the goal value that is given to it by the feed-forward controller. The feedback from the last output of the process is thus combined with the current input from the feed-forward controller to produce control signals that manipulate the output of the current process. In this research project, only one stage of Fig. 3.1 is considered, that being the rapid thermal oxidation of silicon wafers. However, the approach can easily be extended to multiple stages, as demonstrated in [15].

In this research, an optical pyrometer is used to give a temperature indication of the wafer surface, the heat being applied by 15 General Electric quartz halogen lamps arranged in a row over the top of the wafer. For control, time has been selected, temperature being held constant. The parameter selected for control is mean oxide capacitance, C, over the wafer surface, which is acquired through the achievement of a desired mean oxide thickness, μ. For the sake of control equation development, the desired parameters are denoted by a subscript asterisk (C_*, μ_*, etc.). Measured parameters are denoted by a subscripted index, k (C_1, C_2, ..., C_k).

Adaptive Control Equation Development

This development follows directly from [3]. The first assumption is that the process, in this case rapid thermal oxidation, can be expressed in the following form:

$$t = f(x) \cong \hat{f}(x, \theta_*) \qquad (3.1)$$

where t is oxidation time, and f(x) is the function which represents the actual oxidation process, x being the oxide thickness. This function cannot be known precisely, and thus is approximated by the function on the right, $\hat{f}(x, \theta_*)$. The control algorithm requires that this approximate function, for some choice of function parameters $\theta = \theta_*$, provides a good guess for oxidation time. The vector θ_* is not precisely known initially, but is close to an a priori estimate, θ_o.

The model methodology shown earlier is applied to this control situation to find the approximate function $\hat{f}(x, \theta_o)$. Oxidation time can be found as a function of mean oxide thickness, using the same second-order polynomial fit and central composite data point layout. Once this is found, the function can be solved for time.

Once this approximate function is found, what remains is to derive an update law which will insure that, after a small number of iterations (wafer runs), the desired average oxide thickness, and associated oxidation time, is obtained. In other words, $\mu_k \rightarrow \mu_*$ and $t_k \rightarrow t_*$. The equation for such a law, from [3], is:

$$t_k = \hat{f}(\mu_*, \theta_k); \; \theta_{k+1} = \theta_k + \gamma_k(\mu_* - \mu_k)$$
$$\gamma_k = \hat{f}_\theta^{-}(\mu_*, \theta_k)[\hat{f}_\mu(\mu_k, \theta_k) + \qquad (3.2)$$
$$\hat{f}_{\mu\theta}(\mu_k, \theta_k)\hat{f}_\theta^{-}(\mu_k, \theta_k)[t_k - \hat{f}(\mu_k, \theta_k)]]$$

where the subscripts μ and θ refer to derivatives taken with respect to these vectors, and $x^- = x^T[xx^T]^{-1}$, x being a row vector, and x^- the pseudo inverse of x.

In [15], (3.2) is applied to the Deal-Grove oxidation model [26]. Here, it is applied to a second-order polynomial, fitted with actual process data from the rapid thermal processor. The resulting equation for mean oxide thickness as a function of time and temperature is the following:

$$\mu = \hat{\beta}_o + \hat{\beta}_1 \text{Time} + \hat{\beta}_2 \text{Temp} + \hat{\beta}_3 \text{Temp}^2 + \hat{\beta}_4 \text{TimeTemp} \qquad (3.3)$$

where the coefficients from least-square analysis are

$$\hat{\beta}_o = .0001897;\ \hat{\beta}_1 = -2.03*10^{-8};\ \hat{\beta}_2 = -3.925*10^{-7};$$
$$\hat{\beta}_3 = 2.039*10^{-10};\ \hat{\beta}_4 = 2.284*10^{-11} \qquad (3.4)$$

Equation (3.3) can be re-written in the form of (3.1) to obtain

$$\text{Time} = \frac{\mu}{\hat{\beta}_1 + \hat{\beta}_4 \text{Temp}} - \frac{\hat{\beta}_o + \hat{\beta}_2 \text{Temp} + \hat{\beta}_3 \text{Temp}^2}{\hat{\beta}_1 + \hat{\beta}_4 \text{Temp}} \qquad (3.5)$$

In this work, temperature is to be held constant for each control experiment (a control experiment consisting of several wafer runs until the desired mean oxide thickness is obtained). From one control experiment to another, different temperatures may be used to obtain a better uniformity of thickness.

What remains is to develop the control equations around equation (3.5). Rewriting (3.5) once again, with respect to the θ vector in (3.1), and inserting μ_*, is

$$\hat{f}(\mu_*, \theta_o) = \text{Time} = \theta_o^T w_* \qquad (3.6)$$

where

$$\theta_o = \begin{bmatrix} \dfrac{1}{\hat{\beta}_1 + \hat{\beta}_4 \text{Temp}} \\ -\dfrac{\hat{\beta}_o + \hat{\beta}_2 \text{Temp} + \hat{\beta}_3 \text{Temp}^2}{\hat{\beta}_1 + \hat{\beta}_4 \text{Temp}} \end{bmatrix} \qquad (3.7)$$

and

$$w_* = \begin{bmatrix} \mu_* \\ 1 \end{bmatrix} \qquad (3.8)$$

The θ_o in equation (3.7) above is the "a priori" first estimate for the θ_* of equation (3.1). Using the time predicted by (3.6) for a given temperature, rapid thermal processing yields a wafer with a mean oxide thickness μ_k. Plugging this value into the gamma matrix of (3.2), a new theta vector, θ_k, is obtained, which in turn produces a new oxidation time (by replacing the θ_o by θ_k in (3.6)), as shown in (3.2). The new oxidation time yields a new mean, μ_{k+1}, and the process repeats itself until $\mu_k \to \mu_*$ (within the limits of the noise involved with this process).

All that remains to make the control useable is to solve the derivatives in the gamma matrix of (3.2). Starting from left to right, the pseudo inverse is

$$\hat{f}_\theta^-(\mu_*, \theta_k) = \frac{\partial \hat{f}^-}{\partial \theta}(\mu_*, \theta_k) = \begin{bmatrix} \frac{\partial \hat{f}}{\partial \theta(1)} \\ \\ \frac{\partial \hat{f}}{\partial \theta(2)} \end{bmatrix} \left(\frac{\partial \hat{f}^2}{\partial \theta(1)} + \frac{\partial \hat{f}^2}{\partial \theta(2)} \right)^{-1} \qquad (3.9)$$

Applying the derivatives in (3.9) to (3.6) yields

$$\hat{f}_\theta^-(\mu_*, \theta_k) = \begin{bmatrix} \mu_* \\ 1 \end{bmatrix} \left(\mu_*^2 + (1)^2 \right)^{-1} = \frac{1}{\mu_*^2 + 1} w_* \qquad (3.10)$$

A very similar derivation for the other pseudo inverse, which is in terms of measured mean oxide thickness instead of desired mean oxide thickness in (3.2), yields

$$\hat{f}_\theta^-(\mu_k, \theta_k) = \frac{1}{\mu_k^2 + 1} \begin{bmatrix} \mu_k \\ 1 \end{bmatrix} \qquad (3.11)$$

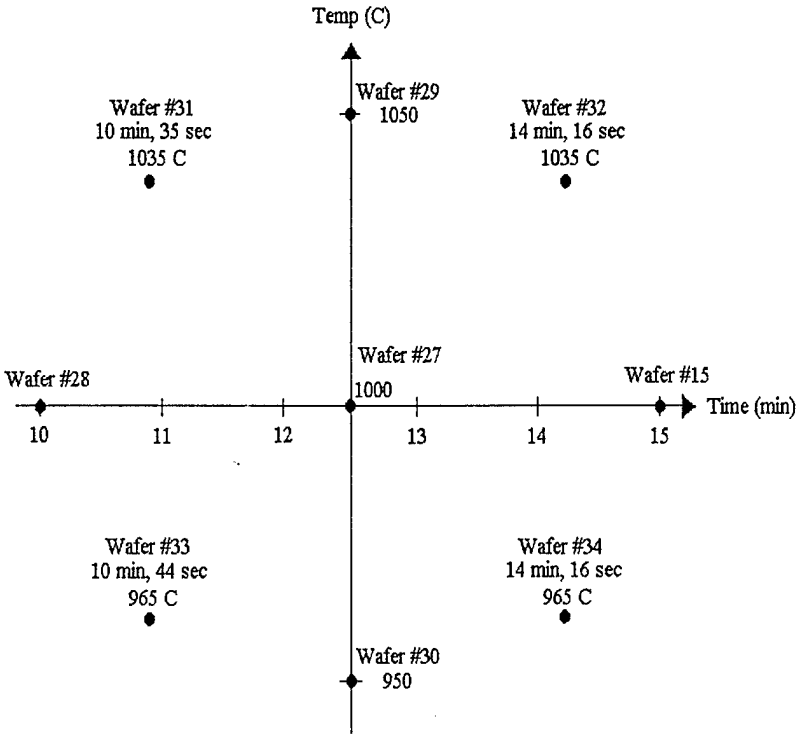
Next, the derivative with respect to mean oxide thickness:

$$\hat{f}_{\mu}(\mu_k,\theta_k)=\frac{\partial \hat{f}}{\partial \mu}(\mu_k,\theta_k)=\frac{1}{\hat{\beta}_1+\hat{\beta}_4\,\text{Temp}}=\theta_0(1) \qquad (3.12)$$

And finally, the derivative with respect to both mean oxide thickness and the theta vector:

$$\hat{f}_{\mu\theta}(\mu_k,\theta_k)=\frac{\partial^2 \hat{f}}{\partial\theta\partial\mu}=\frac{\partial}{\partial\theta}\left[\frac{\partial \hat{f}}{\partial \mu}(\mu_k,\theta_k)\right]=\frac{\partial}{\partial\theta}\left[\frac{1}{\hat{\beta}_1+\hat{\beta}_4\,\text{Temp}}\right]=[1\,0] \qquad (3.13)$$

Putting together (3.10) - (3.13), the gamma matrix of (3.2) can now be expressed in



equation (3.14). The w_k term is identical to the w_* vector, but with μ_k in place of the μ_*.

$$\gamma_k = \frac{1}{\mu_*^2+1} w_* \left[\theta_o(1) + [1\,0] \frac{1}{\mu_k^2+1} \begin{bmatrix} \mu_k \\ 1 \end{bmatrix} (t_k - \theta_k^T w_k) \right]$$

$$\gamma_k = \frac{1}{\mu_*^2+1} w_* \left[\theta_o(1) + \frac{\mu_k}{\mu_k^2+1} (t_k - \theta_k^T w_k) \right] \tag{3.14}$$

Oxide Models for Control

The model to which the control scheme of the previous section is applied is based upon the same central composite data point distribution as discussed earlier. However, in this case, the point distribution is over the time and temperature space, not coordinates of the wafer surface. Again, a second-order model is selected. In this case, some knowledge of the nature of the function is known already, since the oxidation process is described by the well-known Deal-Grove equation. Since the region of operation (~200–600 A) is in the parabolic realm of this equation, representing oxidation by a second-order model is justified as an initial guess for the relationship involved.

Fig. 3.2 shows the point layout used to fit the model. The operational limits of the Tamarack M180 rapid thermal oxidation system provide the constraints for the range of possible time and temperature combinations. For each data collection run, wafers are placed in precisely the same orientation on the wafer tray, and prepared the same way on the same day. Only the recipe varies. As seen in Fig. 3.2, the data distribution is exactly the same as Fig. 2.1. Namely, there are four axial points and four points taken at 45 degree rotations from the axes (at constant distance from the center), and a center point measurement. The center point was chosen at a time and temperature that was in the center of the operating region of interest. This region was determined both by the limitations of the Tamarack unit, and by the desire to grow the thickest oxides possible in order that noise in the process is a less significant component. In order that the design is rotatable, there needs to be 5 repeated center point measurements. However, this was not done here in order to save wafers. The control algorithm should compensate for any slight modeling errors resulting from this.

A total of nine wafers are required to gather the data for Fig. 3.2. Each wafer can then be analyzed using the methods discussed earlier to determine the model which represents the oxide thickness on the surface. Once this wafer model is obtained, a mean oxide thickness can be determined, either by taking the mean value of all nine of the measured data points, or by computing the predicted mean value from the fitted equation from least-squares analysis. In this research, the mean was computed from the response surface generated by the least-squares calculation. Therefore, for each point in Fig. 3.2,

there is an associated mean oxide thickness value. The collection of nine mean oxide thicknesses is then fitted to the time and temperature data to yield the equation for use in the control calculations. The same procedure can also be used to obtain models for thickness standard deviation and thickness uniformity.

The equation chosen for this curve fitting is, as mentioned, the full second-order model. However, upon statistical analysis, it is seen that this model is a poor fit to the data. Different combinations of variables are attempted to improve the situation. The best fit is with the $Time^2$ term removed from the model and all others kept. Figs. 3.3 and 3.4 show plots of this control equation. Another result from this analysis useful for control development is the elimination of the $Time^2$ term. With this term left in the equation, it becomes harder to solve the resulting equation for time, to be in the form of (3.1). Also, without doing any statistical analysis, one can observe that the relative size of the coefficient of the $Time^2$ term in the original full second-order fit is much smaller than any of the other terms' coefficients.

Other calculations of interest for control are standard deviation and uniformity of oxide thickness across the wafer surface. The standard deviation computed is the sample standard deviation,

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \qquad (3.15)$$

where y_i refers to oxide thickness calculations for a wafer, n is the number of wafer thickness calculations, and $\bar{y}$ is the mean of the thicknesses. As with the mean thickness calculation for each wafer, the standard deviation calculation in (3.15) uses the Section 2 model of oxide thickness and computes the statistic based on the thickness model

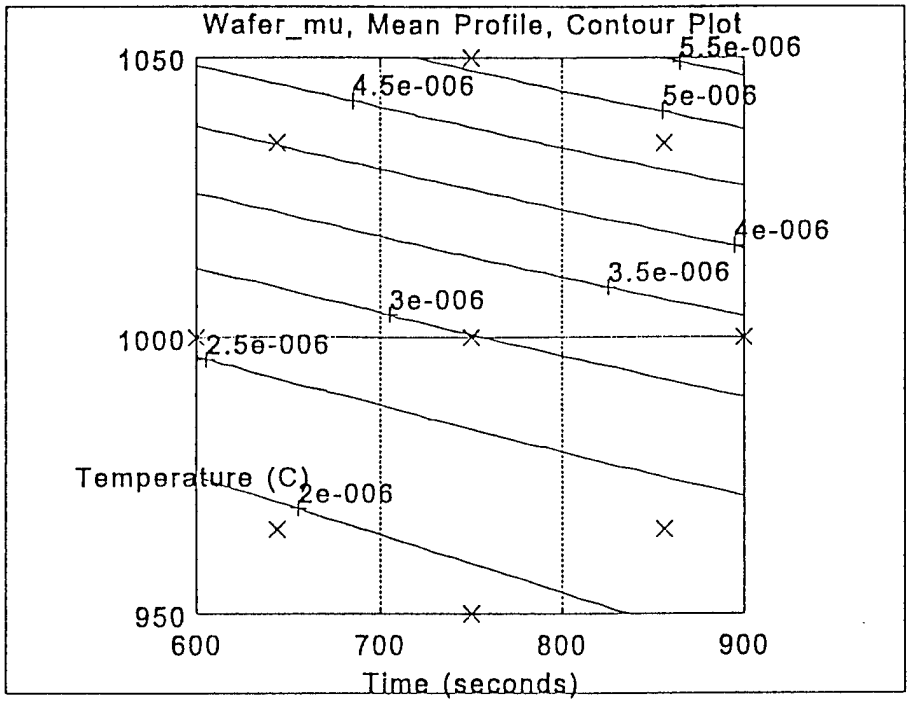


Fig. 3.3. Mean oxide thickness profile for control equation as a 3-D mesh plot.

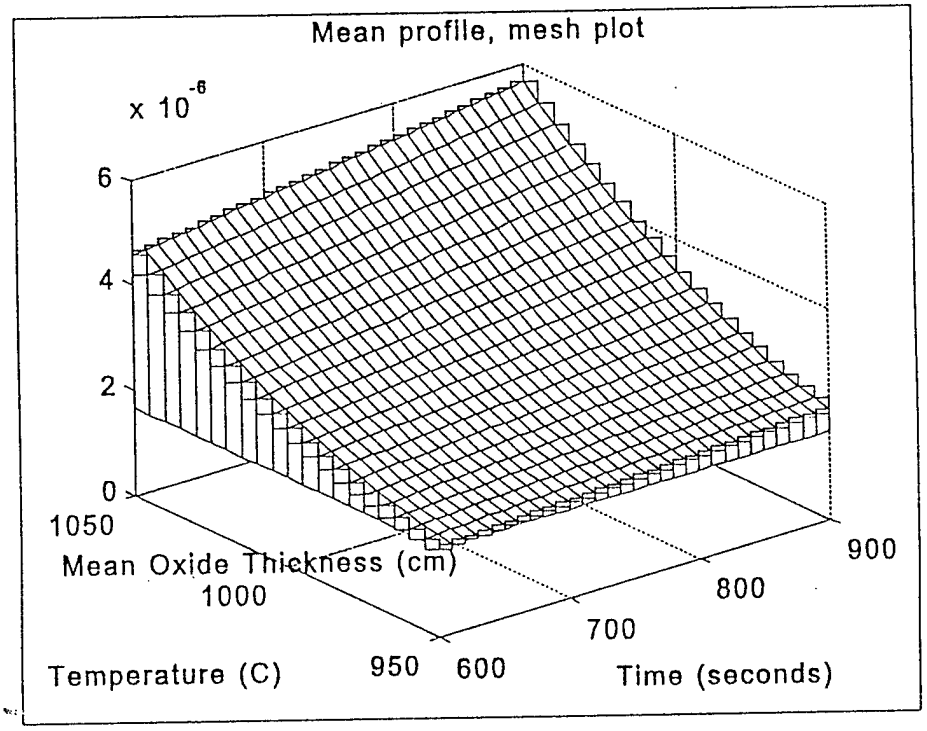


Fig. 3.4. Mean oxide thickness profile for control equation as a contour plot.

equation, not individual measurement points. The standard deviation gives a good indication what combination of time and temperature yields the flattest surface. Figs. 3.5 and 3.6 show the resulting plots of this data. The figures are from a full second-order least squares fit of the standard deviation data to the times and temperatures shown in Fig. 3.2. The statistical analysis is not done on this particular model, since uniformity is the

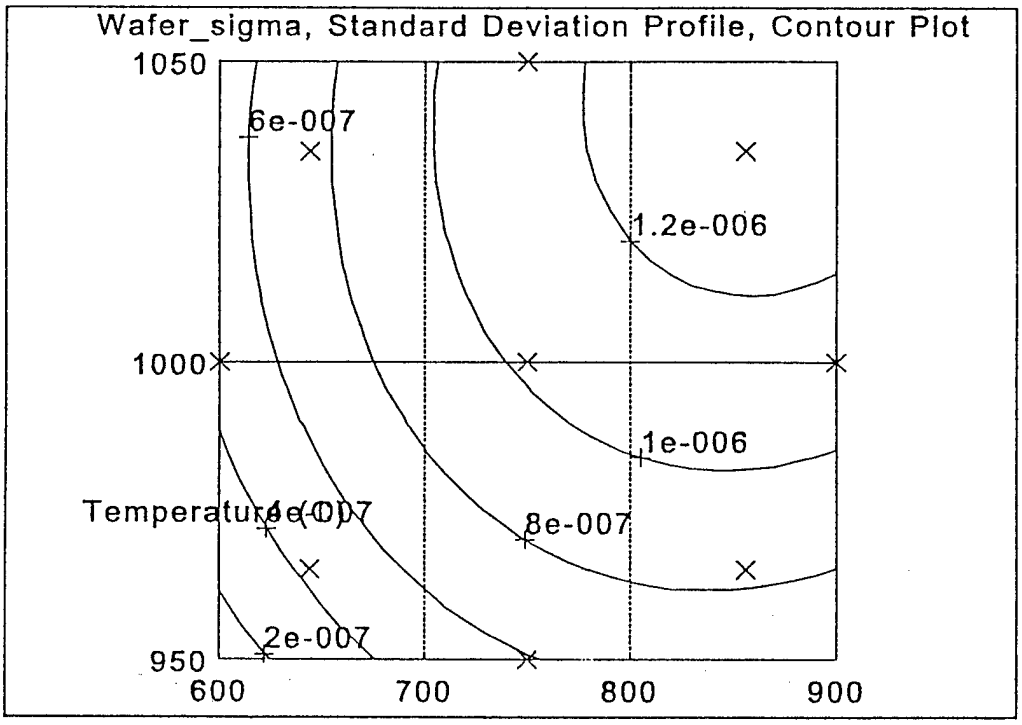
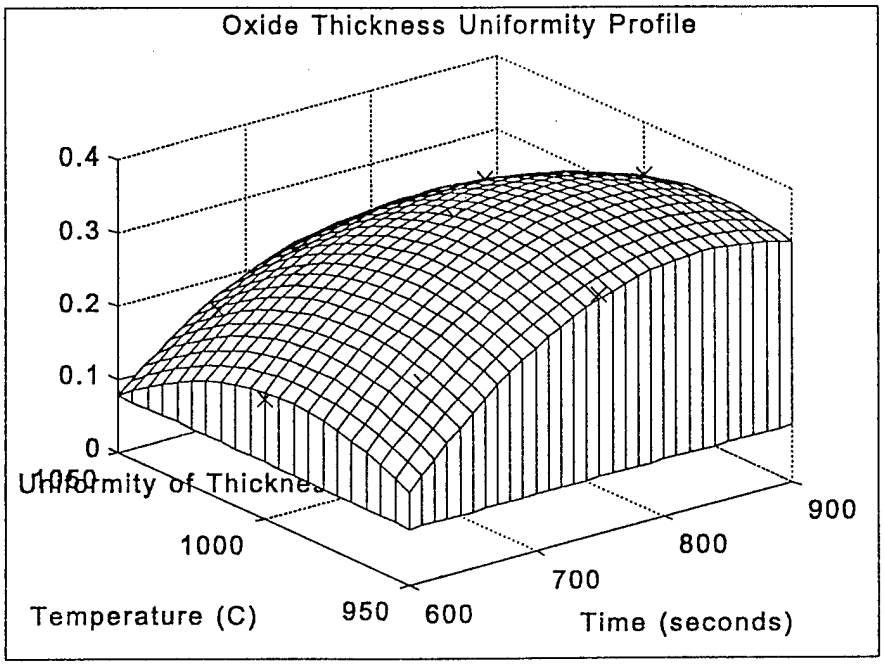
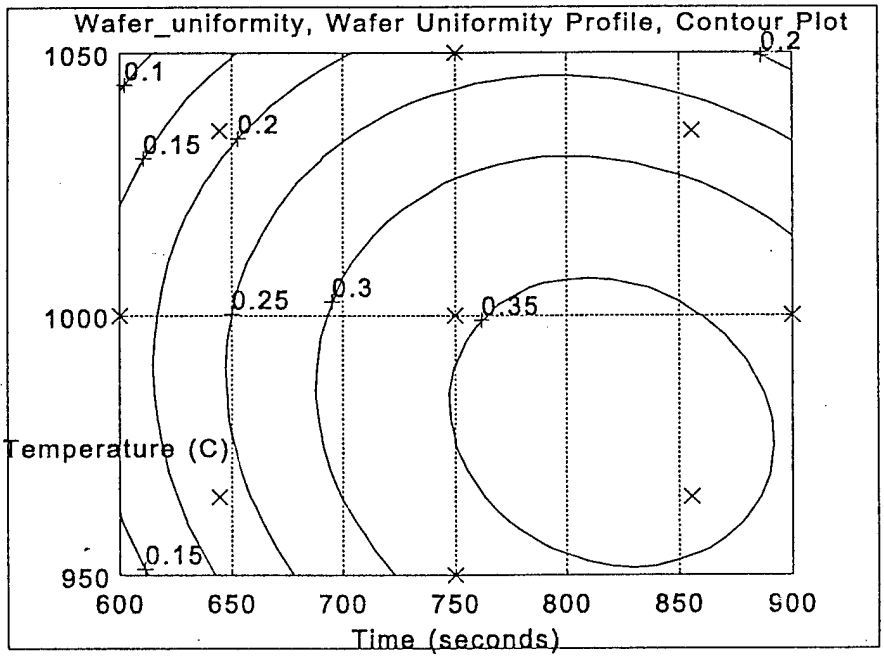


Fig. 3.6. Contour plot of the data of Fig. 3.5.


Oxide Thickness Uniformity Profile
0.4
0.3
0.2
0.1
0
1050
1000
950
Temperature (C)
600
700
800
900
Time (seconds)


Wafer_uniformity, Wafer Uniformity Profile, Contour Plot
1050
1000
950
Temperature (C)
0.1
0.15
0.2
0.25
0.3
0.35
600
650
700
750
800
850
900
Time (seconds)

As can be easily noted by equation (3.16), the best uniformity occurs when the value of this statistic is at a minimum. From Fig. 3.7, it can be seen that the optimum uniformities are obtainable for small oxidation times, and in particular small oxidation times and high temperatures. This differs from the standard deviation plot, which only indicated small times as the desired response.

For control use, the data presented in Fig 3.4 can be used to find out the combinations of time and temperature which yield the desired mean oxide thickness value. The line of constant thickness can be found and followed towards a temperature which yields the optimum uniformity, from Fig. 3.8. This temperature can now be used with the control formulae presented earlier and an optimum oxidation time converged upon.

4. SAMPLE RESULTS

Prior to an actual run using the processing equipment with actual wafers, the control developed in this work was tested with the help of Matlab. This was done using the Deal-Grove oxidation model to simulate the conditions of the Tamarack rapid thermal processor. In other words, the model from (3.5) is used to generate an initial oxidation time, and then the Deal-Grove equation is used to produce a "measured" oxide thickness for that time. Although obviously this thickness is not representative of the wafer-wide mean oxide thickness which is to be studied in actual runs, this gives an indication for the proper convergence of the control to a given system, using an approximate model.

$$t - t_o = \frac{(x^2 - x_i^2)}{B} + \frac{(x - x_i)}{(B/A)} \qquad (3.17)$$

The model shows convergence to within 1% of the desired oxide thickness after four iterations. Table 4.1 shows the various runs and the results. Equation (3.17) is the Deal-Grove equation and coefficients [26]. The test was done using Matlab software on a Hewlett-Packard workstation, but can easily be done on any programmable calculator.

In (3.17), B is termed the "parabolic" rate coefficient, and B/A the "linear" rate coefficient, x is the oxide thickness for an oxidation time t, and x_i is the pre-existing oxide

$$B = B_o e^{-E_A/kT}\ \mu m^2/s,\ B_o = 2.14 \times 10^7\ \mu m^2/s,\ E_A = 1.23\ eV$$
$$k = 8.62 \times 10^{-5}\ eV/K,\ T = \text{temperature (Kelvin)} \qquad (3.18)$$

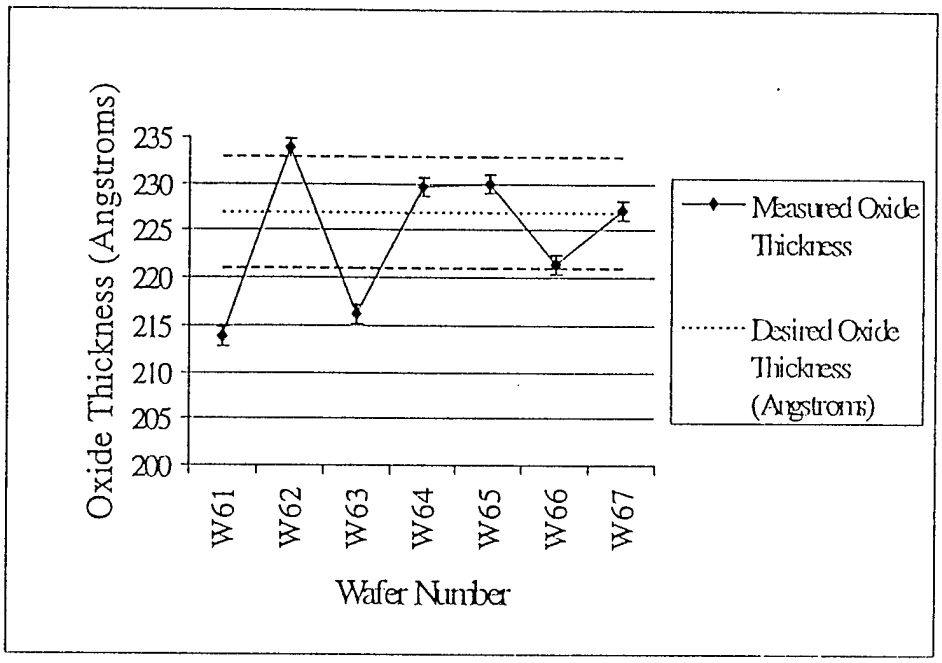
thickness resulting from an oxidation of time t_o. The coefficient B is:
The activation energy E_A in (3.18) is for dry oxidation, which is done with the Tamarack

$$x = \sqrt{(t - t_o)xB + x_i^2} \qquad (3.19)$$

unit. For the oxides studied here, thicknesses are large enough such that, as a good approximation, the second term of (3.17) can be ignored. Solving for oxide thickness, the equation which is used for testing is:
For further simplicity, t_o and x_i^2 are taken to be zero, since the thin thickness of native oxide is negligible in the calculation.

TABLE 4.1. Tests of Control Equations using Deal-Grove Model and Matlab Simulation
Temperature 1000 C, Desired Oxide Thickness 400 A

Iteration No.	Time from (3.5) (sec)	Oxide Thickness from (3.19) (A)
0	1141.7	575.4
-	Time from Control Equations (3.1) - (3.14)	-
1	451.2	361.7
2	601.9	417.8
3	531.9	392.7
4	560.5	403.2
5	548.1	398.7
6	553.3	400.6
7	551.1	399.8
8	552.0	400.1
9	551.6	400.0



First, the control is seen to converge to within the 6 angstrom zone, and the last wafer's mean oxide thickness is precisely on target.

Fig. 4.1. Example of oxide thickness convergence in oxidation control experiment.

5. SUMMARY/CONCLUSIONS

This work demonstrated that wafer-scale parameters can be modeled and controlled. The modeling goal was representation of whole-wafer parameter trends using a minimal number of measurement sites. In such a way, useable wafer area for devices is maximized. The aim of the control was the attainment of a process parameter goal in a minimum number of process steps using an approximate model of the process.

A second-order polynomial model based upon a central composite design was created to map parameters on a wafer surface. It has been shown to adequately represent oxide thickness profiles created from a rapid thermal processor and measured by an ellipsometer. These models are then used with a run-to-run adaptive controller, applied to the oxidation process, to obtain a goal mean oxide thickness while simultaneously striving for optimum uniformity. Results have shown satisfactory control convergence, within an error margin of about 2%, in the mean oxide thickness parameter of the rapid thermal oxidation step after about four iterations.

REFERENCES

[1] H. Wadsworth, K. Stephens, and A. Godfrey, *Modern Methods for Quality Control and Improvement.* New York: Wiley, 1986.

[2] S. Norman and S. Boyd, "Multivariable Feedback Control of Semiconductor Wafer

Temperature," *Proceedings of the American Control Conference*, San Francisco, CA, 1992.
[3] K. Tsakalis and P. Crouch, "A Simple Adaptive Controller for an Oxidation Process," *Proceedings of the American Control Conference*, San Francisco, CA, 1993.
[4] E. Sachs, R. Guo, S. Ha, and A. Hu, "Process Control System for VLSI Fabrication," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. 2, pp. 134–143, 1991.
[5] G. P. Box and J. N. Hunter, "Multi-factor Experimental Designs for Exploring Response Surfaces," *The annals of Mathematical Statistics*, vol. 28, pp. 195–241, 1957.
[6] D. Montgomery, *Design and Analysis of Experiments Third Edition.* New York: Wiley, 1991.
[7] Guo and Sachs, "Modeling, Optimization and Control of Spatial Uniformity in Manufacturing Processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 6, no. 1, pp. 41–57, 1993.
[8] G. May and C. Spanos, "Statistical Experimental Design in Plasma Etch Modeling," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. 2, pp. 83–97, 1991.
[9] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences.* San Francisco: Dellen Publishing Company, 1992.
[10] D. Montgomery and G. Runger, *Applied Statistics and Probability for Engineers.* New York: Wiley, 1994.
[11] G. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for experimenters: An Introduction to Design, Data Analysis and Model Building.* New York: Wiley, 1978.
[12] R. Jones and T. Mele, "Use of Screening and Response Surface Experimental Designs for the Development of a 0.5-μm CMOS Self-Aligned Titanium Silicide Process," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. 4, pp. 281–287, 1991.
[13] G. Gaston and A. Walton, "The Integration of Simulation and Response Surface Methodology for the Optimization of IC Processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 7, no. 1, pp. 22–33, 1994.
[14] D. Boning and P. K. Mozumder, "DOE/Opt: A System for Design of Experiments, Response Surface Modeling, and Optimization Using Process and Device Simulation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 7, no. 2, pp. 233–243, 1994.
[15] K. Stoddard, "Application of Feed-Forward and Adaptive Feedback Control to Semiconductor Device Processing," Master's thesis, Department of Electrical Engineering, Arizona State University, 1994.
[16] S. Norman and S. Boyd, "Multivariable Feedback Control of Semiconductor Wafer Temperature," *Proceedings of the American Control Conference*, San Francisco, CA, 1992.
[17] C. Shaper, "Real-Time Control of Rapid Thermal Processing Semiconductor Manufacturing Equipment," *Proceedings of the American Control Conference*, San Francisco, CA, 1993.
[18] R. Gyurcsik, T. Riley, and F. Sorrell, "A Model for Rapid Thermal Processing: Achieving Uniformity Through Lamp Control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. ?, pp. 9–13, 1991.

[19] P. Apte, S. Wood, K. Saraswat, and M. Moslehi, "Temperature Uniformity Optimization Using Three-Zone Lamp and Dynamic Control in a Rapid Thermal Multiprocessor," *Material Research Society Proceedings*, vol. 224, pp. 209–214, 1991.
[20] M. Elta, H. Etemad, J. Freudenberg, M. Giles, J. Grizzle, P. Kabamba, P. Khargonekar, S. Lafortune, S. Meerkov, J. Moyne, B. Rashap, D. Teneketzis, F. Terry, Jr., "Applications of Control to Semiconductor Manufacturing: Reactive Ion Etching," *Proceedings of the American Control Conference*, San Francisco, CA, 1993.
[21] C. Spanos, S. Leang, and S. Lee, "A Control and Diagnosis Scheme for Semiconductor Manufacturing," *Proceedings of the American Control Conference*, San Francisco, CA, 1993.
[22] E. Zafiriou, R. Adomaitis, and G. Gattu, "An approach to Run-to-Run Control for Rapid Thermal Processing," *Proceedings of the American Control Conference*, Seattle, WA, 1995.
[23] S. Butler and J. Stefani, "Supervisory Run-to-Run Control of Polysilicon Gate Etching using in situ Ellipsometry," *IEEE Transactions on Semiconductor Manufacturing*, vol. 7 no. 2, pp. 193–201, 1994.
[24] W. Kern and D. Puotinen, "Cleaning Solutions Based on Hydrogen Peroxide for use in Silicon Semiconductor Technology," *RCA Review*, 1970.
[25] W. Kern, "Purifying Si and SiO_2 Surfaces with Hydrogen Peroxide," *Semiconductor International*, pp. 94–99, April 1984.
[26] S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era Volume I: Process Technology*. Santa Ana, CA: Lattice Press, 1986.
[27] S. Wolf, *Silicon Processing for the VLSI Era Volume II: Process Integration*. Santa Ana, CA: Lattice Press, 1990.
[28] Dieter K. Schroder, *Semiconductor Material and Device Characterization*. New York: Wiley, 1990.
[29] S. M. Sze, *Physics of Semiconductor Devices 2nd Edition*. New York: Wiley, 1981.
[30] R. Pierret, *Field Effect Devices, Second Edition*. New York: Addison-Wesley, 1990.
[31] "MDC Application Note, Series Resistance Effects in MOS C-V Measurements," Materials Development Corporation, Chatsworth, CA, 1991.