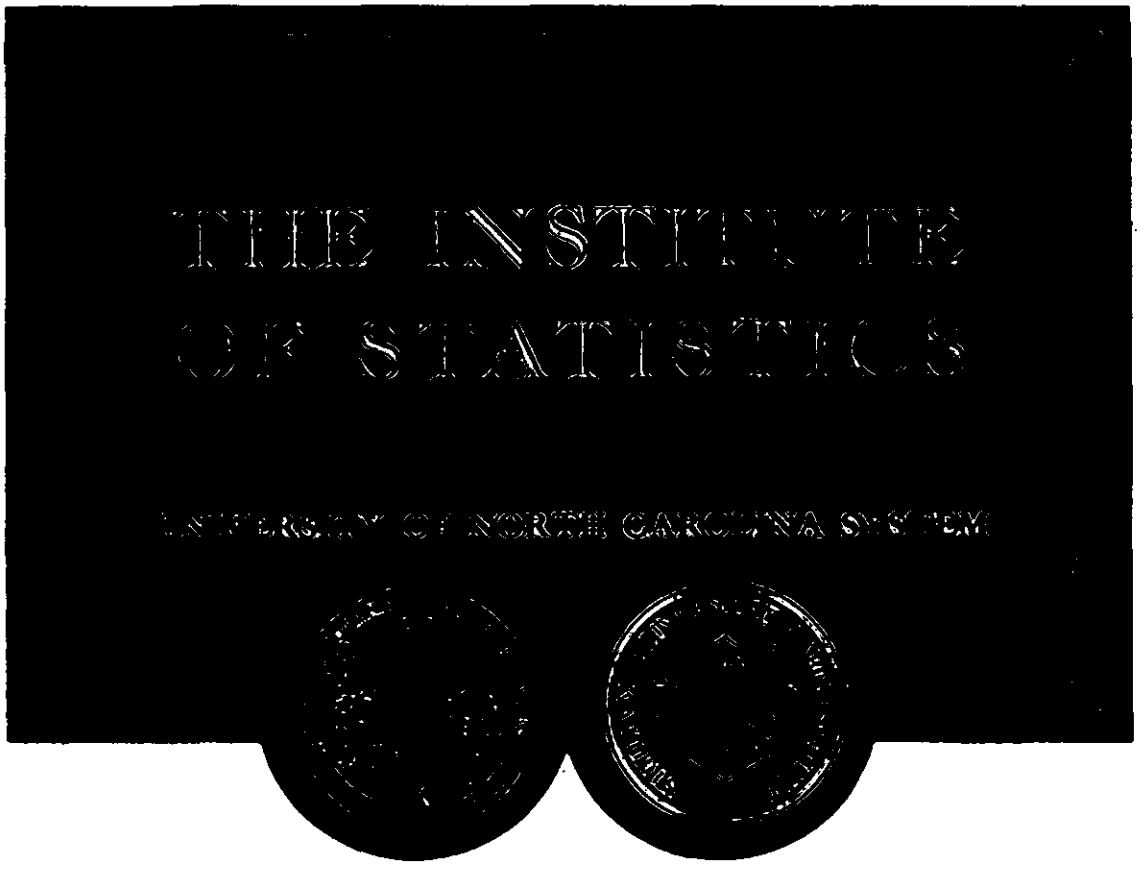


Achieving Uniformity in a Semiconductor Fabrication Process Using Spatial Modeling

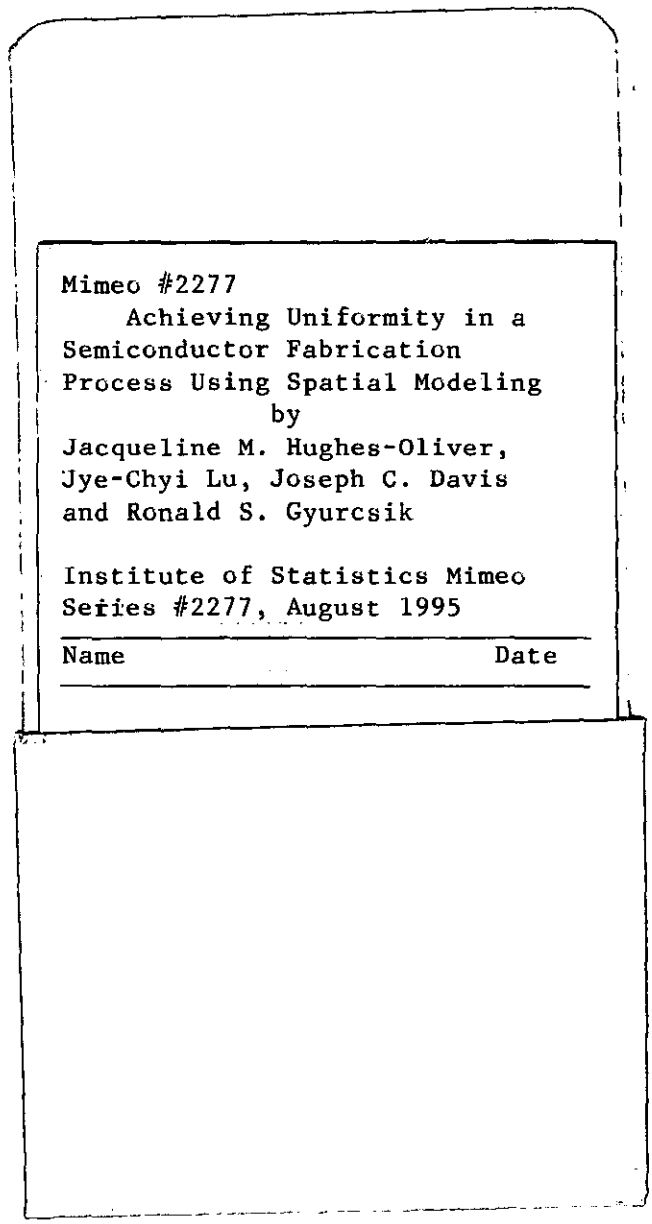
by

Jacqueline M. Hughes-Oliver, Jye-Chyi Lu, Joseph C. Davis
and Ronald S. Gyurcsik

Institute of Statistics Mimeo Series Number 2277

August, 1995

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina



Mimeo #2277
Achieving Uniformity in a Semiconductor Fabrication Process Using Spatial Modeling
by
Jacqueline M. Hughes-Oliver, Jye-Chyi Lu, Joseph C. Davis and Ronald S. Gyurcsik

Institute of Statistics Mimeo Series #2277, August 1995

Name Date

Achieving Uniformity in a Semiconductor Fabrication Process Using Spatial Modeling

Jacqueline M. Hughes-Oliver*, Jye-Chyi Lu*, Joseph C. Davis†, Ronald S. Gyurcsik†

An important diagnostic tool of semiconductor fabrication processes is within-wafer uniformity. Given some required standard (for example, deposition of 1000 angstroms of polysilicon on a blanket of silicon dioxide), the goal is to determine optimum process conditions that will achieve this requirement uniformly across the wafer, despite possible spatial anomalies in the fabrication process.

Processes are traditionally optimized with respect to the process operating conditions, without regard to spatial dependence. Nevertheless, this spatial dependence can be very important, and is thus accounted for in our modeling of the mean, variance, and correlation of measurements taken on a wafer. Moreover, the uniformity (or lack thereof) of the wafer surface is investigated by the use of different metrics—based on simple averages of absolute error, or on a thin plate spline of the error surface. These metrics are then employed in the optimization of the process. The techniques are illustrated through application to a Rapid Thermal Chemical Vapor Deposition process.

KEY WORDS: Optimality criteria; Spatial correlation; Thin plate spline; Variance modeling.

---

*Department of Statistics, North Carolina State University, Box 8203, Raleigh NC 27695-8203.
†Dept. of Electrical & Computer Engineering, North Carolina State University, Box 7911, Raleigh, NC 27695-7911.

We would like to thank Douglas Nychka for providing the routines for thin plate spline generation.

1. INTRODUCTION

There have recently been numerous articles in the area of semiconductor manufacturing that attempt to integrate statistical process control and computer-aided manufacturing systems in order to improve very large scaled integrated (VLSI) circuits fabrication processes. (For example, see Mozumder, Shyamsunder and Strojwas, 1988; Sachs, Guo, Ha and Hu, 1991.) Deposition of material onto the surface of a wafer in each of the several hundred steps required to process a semiconductor chip is intimately dependent on the chamber design, the processing conditions, the wafer condition, and the stochastic error, to name a few. It is thus very difficult to guarantee the uniformity of the quality of deposition at various sites across the wafer.

In order to determine process conditions for maximizing the yield (number of high-quality chips) of a wafer, as well as for building (spatial) signatures to be able to detect possible causes of malfunctions (Chang and Spanos, 1991), it is first necessary to develop a statistical model that accurately characterizes the spatial surface. Mozumder and Loewenstein (1992) first apply the least squares method to model the polysilicon thickness on a wafer at various locations. They then regress the least squares estimates with respect to the process conditions, and optimize the spatial uniformity by minimizing the slopes of the thickness models. Guo and Sachs (1993) use a multivariate regression model to optimize and to control the spatial uniformity. They compare two different formulations of Taguchi's (1987) signal-to-noise (SN) ratios for process optimization and conclude that modeling process responses and formulating the SN ratio for optimization is more effective than formulating the SN ratio and modeling the SN ratios for optimization.

Besides improving the within-wafer uniformity, there is also a target deposition thickness to be reached. Moreover, based on Taguchi's (1987) robust parameter design philosophy, one would like the quality of work on a wafer to be repeatable from wafer to wafer. To handle this multi-objective optimization problem, Mesenbrink, Lu, McKenzie and Taheri (1994), in a soldering process optimization case study, formulate summary statistics (e.g., mean response, uniformity, and their associated variances), fit models to these summary statistics, and then compare predictions to find the best process condition to achieve quality goals. The use of summary statistics (especially

the SN ratio), however, can result in loss of information and is not flexible when the quality goal is changed, say, to look for maximum yield in a certain region. We propose to model all the observations collected on the spatial surfaces from wafers processed in a design of experiment. We are then able to link this model to whatever optimization/quality goals are desired, to find the best operating process conditions. The objective of this article is to show one example of how to formulate variance and correlation models to characterize semiconductor deposition processes and how to utilize a thin-plate spline surface to optimize the process condition. Section 2 contains an example of a semiconductor wafer fabrication process. Section 3 contains models for the mean, variance, and correlation of the spatial responses. Section 4 contains two optimality criteria. Section 5 contains the estimated models and optimization for the wafer example. The discussion is contained in Section 6.

2. A SEMICONDUCTOR WAFER FABRICATION PROCESS

2.1 The Process

Rapid thermal processing (RTP) is quickly becoming the most novel method for semiconductor fabrication (Davis et. al., 1993). Instead of the conventional furnace where large numbers of wafers are simultaneously processed, thus leaving little room for flexibility, RTP processes a single wafer at a time. Rapid Thermal Chemical Vapor Deposition (RTCVD) is a particular RTP in which wafers are heated to facilitate the bonding of chemical vapors to the surface.

A single silicon wafer is covered with oxide on both sides then placed in a quartz chamber. Depending on the time setting for which the chamber is to operate, the wafer heats up to a temperature that compliments the set time. The wafer is heated by lamps placed vertically above the upper surface and horizontally beneath the lower surface. The heat of the wafer is monitored by a single pyrometer that records the heat at the center of the wafer. This temperature is fed into a power source which determines whether there should be more or less heat coming from the lamps. The chemical being deposited is polysilicon and the targeted thickness level is 1000 angstroms on the upper surface. (The lower surface also receives deposition, but at this point there is no interest in this thickness level.)

The design of the fabrication process is such that there are many possible causes for concern. The first is that the pyrometer "reads" temperature only at the center of the wafer, while it is known that the temperature is uneven across the wafer. This is due to conduction and loss of heat due to edge effects. In fact, it is known that temperature decreases with increasing distance from the center.

Even if several pyrometers were being utilized, the problems of uneven temperature would not be rectified. The pyrometer does not actually read the temperature but instead, calculates the temperature using a known functional relationship which depends on the emissivity of the surface. This surface emissivity changes as the polysilicon thickness increases, yet this is not reflected in the formula used by the pyrometer. The net effect is that as the polysilicon layer increases there is less control over the temperature of the wafer.

The response in this experiment is polysilicon thickness. As previously mentioned, this is targeted at 1000 angstroms and one would like to have uniform thickness across the wafer. In an attempt to monitor uniformity, the thickness is measured at 13 different spatial locations on the wafer. These measurement locations are displayed in Figure 1. In this diagram it can also be seen that the wafer is more or less circular, but with the bottom arc removed. The seventh measurement location, hitherto referred to as the center point, is the center of the wafer assuming the wafer forms a perfect circle.

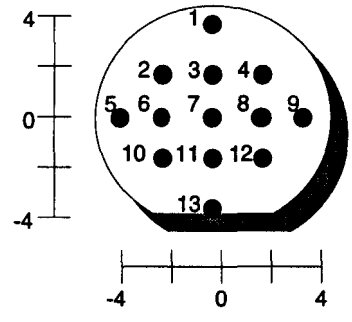


Figure 1: Spatial Measurement Locations.

Based on earlier discussions of the problem of maintaining the temperature across the wafer, it is expected that points closer to the edge will have smaller thickness than points close to the center.

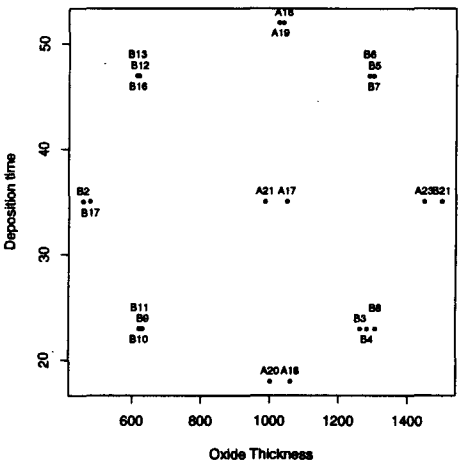


Figure 2: Experimental design, with wafer identifiers.

These edge points are also expected to display more variability. The 13^{th} measurement location is particularly prone to this because in addition to being far from the center point, it is closer to the edge than, say, point 9, because of the straight edge at the bottom of the wafer.

2.2 The Data

This experiment, being very exploratory in nature, has only two process or controllable variables: oxide thickness, the thickness of the oxide layer before deposition of the polysilicon, which ranges from 458 to 1498 angstroms; and deposition time, the time for which the wafer is subject to heating, which ranges from 18 to 52 seconds. The experimental design points are selected according to a quasi-replicated central composite design, as illustrated in Figure 2. For ease of reference, these quasi-replicated design points are numbered 1 (includes wafers B12, B13, B16), 2 (includes wafers A18, A19), 3 (wafers B5, B6, B7), 4 (wafers B2, B17), etc.

Figure 3 shows the relationship between polysilicon thickness and distance from the center point for three of the 22 wafers. Distance-from-center is clearly inversely related to polysilicon thickness. It is also clear that the variability of the thickness increases with increasing distance from the center. In Figure 4, polysilicon thickness is plotted as a function of spatial location for each wafer. As expected, location 13 consistently has the smallest thickness while location 7 (the center point) consistently has the largest. In addition, it is clear that uniformity of polysilicon thickness within a wafer varies for different design points, as does the thickness level.

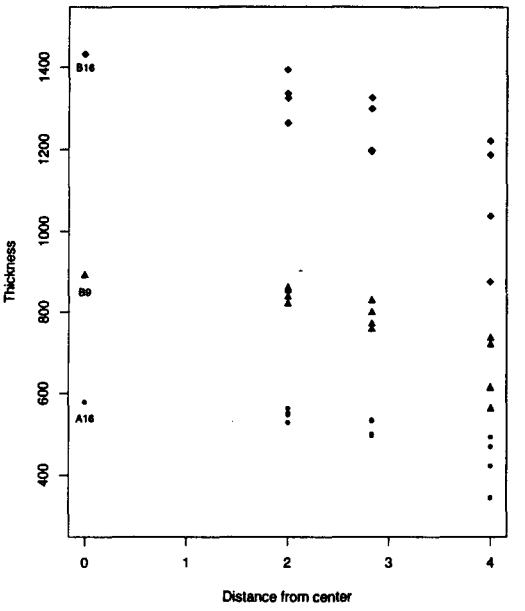


Figure 3: Polysilicon thickness versus distance from center, for 3 wafers.

3. MODELING MEAN, VARIANCE, AND CORRELATION

In order to account for the effects of the process conditions (e.g., oxide thickness and deposition time) as well as the correlation inducing factor (spatial location), models are proposed for not only

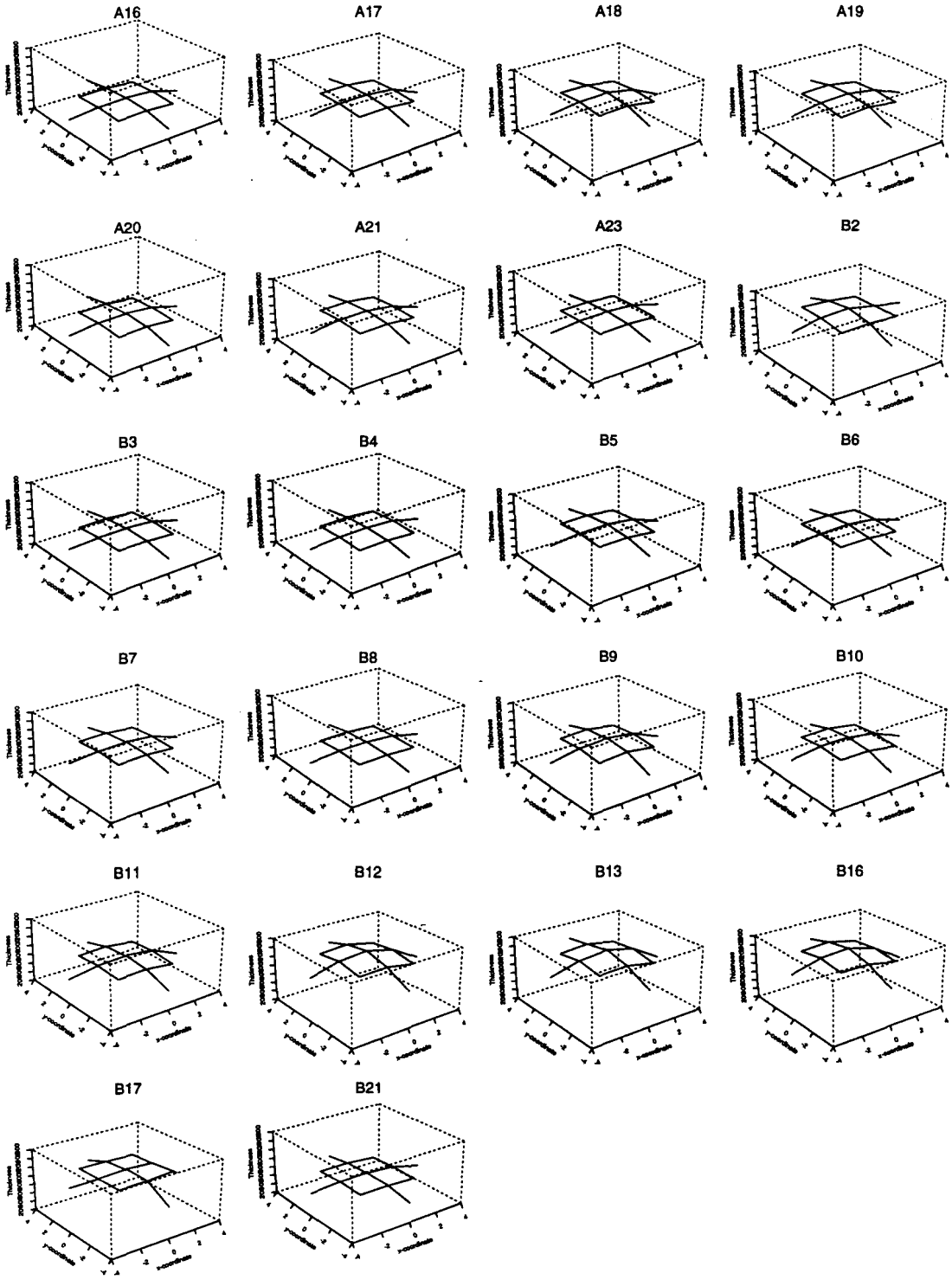


Figure 4: Polysilicon thickness for all 22 wafers at the 13 measurement sites.

the means and variances of the response, but also for the correlation existing between the various spatial measurements. The usefulness of the mean model will be in assessing the achievement of target, as well as uniformity. The variance model is useful for measuring repeatability of the process (Mesenbrink, Lu, McKenzie and Taheri (1994); Davidian and Carroll (1987)). The correlation model affects target, uniformity, and repeatability. It describes the behavioral relationship between the different spatial locations, as well as improving the estimates of the mean and variance models. Ignoring the correlation structure might result in less precise estimation of the parameters.

3.1 Notation

Let N represent the number of experimental runs (process operating conditions); S the number of spatial locations; C the number of process variables (possibly including interaction terms); $\boldsymbol{X}_1$ an $N \times C$ matrix of process variables; D the number of spatial variables; $\boldsymbol{X}_2$ an $S \times D$ matrix of spatial variables; y_{is} the response for the i^{th} run at location s; $\boldsymbol{y} = (y_{11}, \ldots, y_{1S}, y_{21}, \ldots, y_{2S}, \ldots, y_{N1}, \ldots, y_{NS})'$ the vector of responses.

3.2 The Mean Model

The most general model would allow the mean to depend on both process and spatial variables, as well as their interactions. Hence, the response is modeled as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\boldsymbol{X} = [\mathbf{1}_{NS} | \boldsymbol{X}_1 \otimes \mathbf{1}_S | \mathbf{1}_N \otimes \boldsymbol{X}_2 | \boldsymbol{X}_1 \otimes \boldsymbol{X}_2]$$

is an $NS \times (1 + C + D + CD)$ matrix with components of process variables, spatial variables, and their interactions; $\boldsymbol{\beta}$ is a vector of unknown parameters; and $\boldsymbol{\epsilon}$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{V}$.

3.3 The Variance Model

Unlike what used to be typical, it is assumed that the variance of the responses may also depend on the process conditions and the spatial location. Consequently, the variance-covariance matrix $\boldsymbol{V}$ of the responses is modeled as

$$\boldsymbol{V} = \boldsymbol{\Lambda}(\boldsymbol{I}_N \otimes \boldsymbol{R})\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is an $NS \times NS$ diagonal matrix of the standard deviations of the response $\boldsymbol{y}$; $\log(\mathrm{diag}(\boldsymbol{\Lambda})) = \boldsymbol{X\theta}$ ($\boldsymbol{X}$ is defined above) so that the log of standard deviations of the response is modeled as linear combinations of process variables, spatial variables and their interactions; $\boldsymbol{R}$ is an $S \times S$ correlation matrix of the spatial responses on a single wafer. This model assumes: 1) responses on separate wafers are independent; and 2) the spatial correlation within a wafer is the same for all experimental runs. If these assumptions are not viable, then $\boldsymbol{V}$ may be modeled as $\boldsymbol{\Lambda R \Lambda}$, where $\boldsymbol{R}$ is $NS \times NS$ and is the correlation matrix of the entire response vector. This latter case would contain far more parameters to be estimated.

3.4 The Correlation Model

The correlation between spatial locations s and t is modeled as a function of three distances: d_s, the distance between location s and the center point; d_t, the distance between location t and the center point; and h_{st}, the distance between locations s and t. The model is

$$R_{st} = \begin{cases} 1 & s = t \\ \delta_1 exp(-\delta_2 h_{st} exp(-\delta_3(d_s + d_t))) & s \neq t, \end{cases}$$

where R_{st} is the element in the s^{th} row and t^{th} column of R, and δ_1, δ_2, and δ_3 are parameters to be estimated.

This correlation model is constructed to satisfy several requirements based on empirical observations and the mechanics of the physical RTCVD process. There are many other possible correlation models that can take into account the effects of d_s and d_t. We use the correlation given above because the d_s and d_t effects act as adjustment factors for the h_{st} effect. To illustrate the features

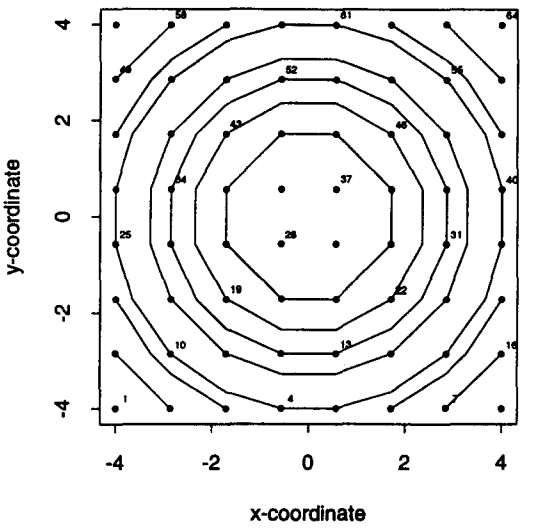


Figure 5: Grid of 64 spatial locations. Label represents the location number, and contours represent equal distance from center.

of this model, a grid of 64 spatial locations, as depicted in Figure 5, is used. Figure 5 also shows the distance contours of the locations from the center. Figures 6 and 7 show correlation contours obtained from this model for a variety of situations. The correlation between location 20 and all other locations is displayed in Figure 6. The correlation between location 55 and all other locations is displayed in Figure 7. Both figures 6 and 7 illustrate the correlation model for $\delta_1 = 1, \delta_2 = .25, .5,$ and $\delta_3 = -.1, 0, .1$.

First, it is expected that the correlation between two locations decreases exponentially as the distance between the locations increases. This is clearly seen in Figure 6, where, for example, $\rho_{20,28}$ is consistently larger than $\rho_{20,52}$. The effect of between-location distance on the correlation is controlled mainly by the parameter δ_2. Figures 6 and 7 show that as δ_2 increases, the correlation contours become closer, thereby dampening the effect of large between-location distances and exploding the effect of small distances.

The correlation between two locations is, however, believed to be more than just a function of the

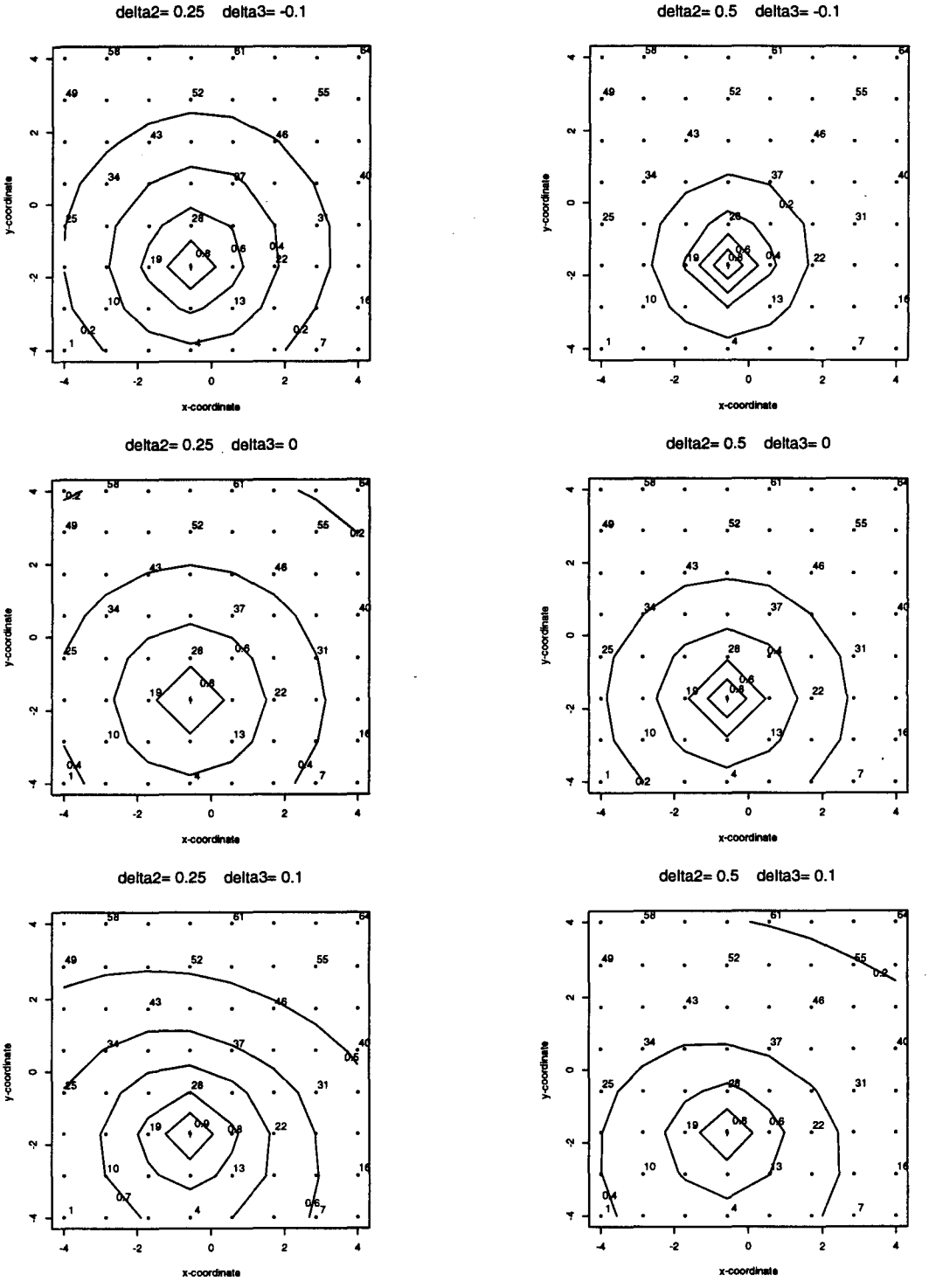


Figure 6: Correlation with location 20, for several choices of δ_2 and δ_3.

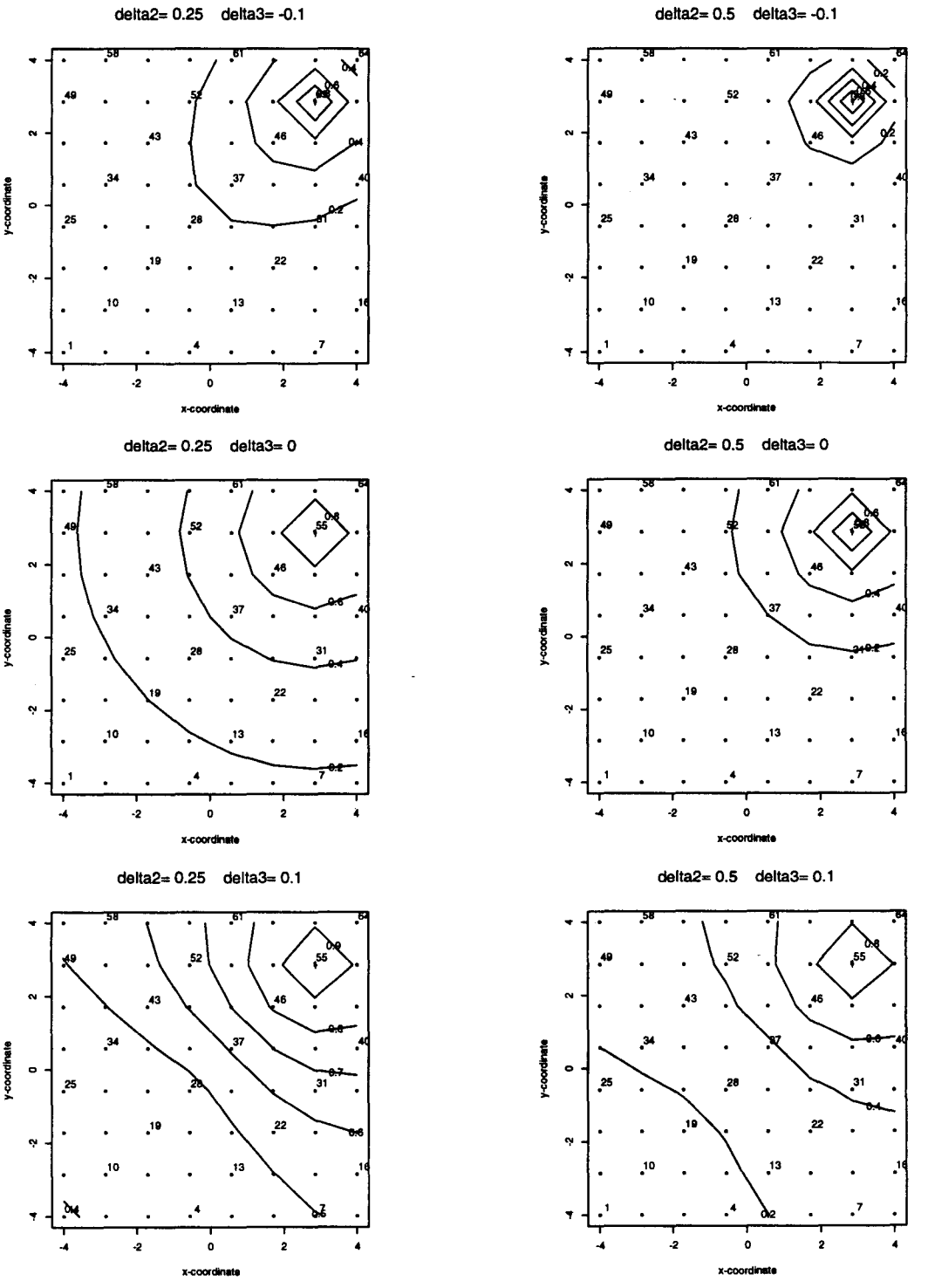


Figure 7: Correlation with location 55, for several choices of δ_2 and δ_3.

distance between them. Hence, all models assuming second order or intrinsic stationarity must be abandoned. In fact, given equal between-location distance, the correlation remains the same within radial bands from the center, but not across radial bands. For example, consider pairs of locations (19,43), (22,46), and (18,42). Although these pairs of locations share the same between-location distance (see Figure 5), since (19,43) and (22,46) fall on the same radial band, then $\rho_{19,43} = \rho_{22,46}$; on the other hand, since (18,42) is on a different radial band, $\rho_{18,42} \neq \rho_{19,43}$. In addition, for two locations in the same radial band, correlation decreases as distance between the locations increases, so that, for example, $\rho_{19,43} > \rho_{19,46}$.

The radial features of the model are controlled by δ_3. When $\delta_3 = 0$, there are no radial effects and the model becomes second order stationary. This is seen in Figures 6 and 7 as concentric-circle correlation contours. When $\delta_3 > 0$, pairs of locations further from the center are more highly correlated than locations closer to the center. For example, although pairs of locations (4,20) and (20,36) have the same between-locations distance, $\rho_{4,20} > \rho_{20,36}$ because the sum of the distances of 4 and 20 to the center $(d_4 + d_{20})$ is larger than the sum of the distances of 20 and 36 to the center $(d_{20} + d_{36})$. In other words, a positive value of δ_3 implies that "edge" points behave more similarly to each other than do "center" points. On the contrary, a negative value of δ_3 implies that "center" points behave more similarly than "edge" points.

Finally, in order to force the desirable positive correlations, as well as meet the requirements described above, the following constraints on the parameters are necessary:

$$0 < \delta_1 < \begin{cases} exp(\delta_2 d_{min} exp(-2\delta_3 d_{max})) & \text{if } \delta_3 > 0 \\ exp(\delta_2 d_{min} exp(-2\delta_3 d_{min})) & \text{if } \delta_3 < 0 \end{cases}$$

$$0 < \delta_2,$$

where d_{min} is the minimum possible distance between two distinct locations, and d_{max} is the maximum possible distance. [In the RTCVD example, $d_{min} = 2$ and $d_{max} = 8$.] A more restrictive, simpler constraint on δ_1 is $0 < \delta_1 < 1$, which still allows δ_3 to be unconstrained.

4. OPTIMALITY CRITERIA

Criteria for optimizing the process must be goal-specific. Hence, there must be penalties for: deviation from target response; lack of uniformity across the wafer; and large variability of response. The criteria described below address all these aspects with increasing complexity. The first criterion takes the "optimize on summary statistics" approach, while the second criterion takes the "optimize on the entire surface" approach. As a result of having closed form expressions for the criteria, they may be easily implemented using common minimization routines. Of course, the efficiencies of these routines will depend on your particular application. We hasten to add that the literature on optimality criteria is very large, and so no claim is made of this being an exhaustive presentation.

For ease of presentation, slightly differing notation than that used in Section 3 is introduced: s = # spatial measurement locations on a wafer; $\boldsymbol{y}$ = vector of responses on a single wafer; $\boldsymbol{m}$ = expected value of $\boldsymbol{y}$; $\boldsymbol{v}$ = vector of standard deviations of $\boldsymbol{y}$; $\boldsymbol{C}$ = correlation matrix of $\boldsymbol{y}$; T = target response.

The first criterion attacks the problem with brute force: minimize the average standard deviation, subject to the average absolute error being within tolerance. By focusing on minimization of the average standard deviation, much attention is given to the issue of repeatability. Low standard deviation indicates that repeating the experiment using the same process conditions will result in very similar responses. The constraint that absolute error is within a given tolerance addresses the issue of meeting target. The averaging, which is done across spatial measurement locations, attempts to address uniformity by looking at the average behavior rather than the individual or worst-case behavior. In symbols, the criterion is

Minimize

$$\frac{1}{s}\mathbf{1}'\boldsymbol{v}$$

subject to

$$\frac{1}{s}\mathbf{1}'\text{abs}(\boldsymbol{m} - T * \mathbf{1}) < \text{tolerance}.$$

There are, however, several disadvantages with this approach. One such disadvantage is that a value for tolerance needs to be provided. Another disadvantage is that there is no immediate measure of within-wafer uniformity. Also, the averaging done across measurement sites causes loss of uniformity information, as the mean may be unduly influenced by one or two stray values of the limited number of measurement locations available. Yet another disadvantage is the inability to assign unequal weights to the activities of making the process repeatable and achieving target. Finally, in this approach the correlation structure is ignored.

The second criterion is based on first obtaining the error surface of the wafer—the difference between the responses and the target. This surface will have both positive and negative values, so we choose to look at the squared error surface. The integral of this squared error surface can serve as a measure of within-wafer uniformity, with an ideal value of zero. This integral also measures the ability to meet target, since the difference is with respect to target. As a function of the responses, this integral metric is itself a random quantity with an associated distribution, expected value, and variance. The expected value of the integral metric is chosen to represent this distribution, and criterion 2 minimizes this value. The criterion is

Minimize

$$trace(\boldsymbol{A}\text{diag}(\boldsymbol{v})\boldsymbol{C}\text{diag}(\boldsymbol{v})\boldsymbol{A}') \; + \; (\boldsymbol{m} - T * \mathbf{1})'\boldsymbol{A}'\boldsymbol{A}(\boldsymbol{m} - T * \mathbf{1}), \qquad (1)$$

where the matrix $\boldsymbol{A}$ is described below. Repeatability issues are automatically considered because the variance and correlation of the responses are introduced in the first component of equation (1).

Unlike the first criterion, this one provides an immediate measure of uniformity. Changes in the correlation model and/or parameters is also allowed to influence the criterion and consequently the optimal process conditions chosen. This criterion also avoids the traps associated with using summary statistics such as the mean by instead integrating across the surface. The fact that repeatability issues are automatically accounted for is also nice.

The matrix $\boldsymbol{A}$ is determined by applying an interpolating natural thin plate spline to the s spatial responses in $\boldsymbol{y}$ in order to obtain the error surface on the wafer. The typical representation of a natural thin plate spline is

$$g(\boldsymbol{t}) = \sum_{i=1}^{n} \delta_i \eta(||\boldsymbol{t} - \boldsymbol{t}_i||) + \sum_{j=1}^{3} a_j \phi_j(\boldsymbol{t}),$$

where the spatial locations (x- and y-coordinates) are represented by $\boldsymbol{t}$; there are n data points at locations $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$; $\eta(\cdot)$ is a function of distance between locations; $\phi_1(\boldsymbol{t}) = 1$, $\phi_2(\boldsymbol{t}) =$ x-coordinate of $\boldsymbol{t}$, $\phi_3(\boldsymbol{t}) =$ y-coordinate of $\boldsymbol{t}$; $\boldsymbol{\delta}$ and $\boldsymbol{a}$ are solutions to a system of equations involving the data $\boldsymbol{z}$. It turns out, however, that this equation is a linear combination of the data, so that we may use the modified representation

$$g(\boldsymbol{t}) = \boldsymbol{u}'\boldsymbol{z}.$$

(For more detail, see Green and Silverman, 1994.) So to represent the spline at N selected spatial locations, it is possible to use the matrix representation $\boldsymbol{A}\boldsymbol{z}$, where $\boldsymbol{A}$ has n columns and N rows. Moreover, the sum of the columns of $\boldsymbol{A}$ is the vector of 1's.

The error surface for criterion 2 may then be written as $\boldsymbol{A}(\boldsymbol{y} - T * \mathbf{1})$, where the matrix $\boldsymbol{A}$ is obtained from a sufficiently fine grid of locations on the wafer. Numerical integration of the squared error surface is more tractable that exact integration and, provided a fine enough grid of locations is used to obtain $\boldsymbol{A}$, may be approximated as $k * (\boldsymbol{y} - T * \mathbf{1})'\boldsymbol{A}'\boldsymbol{A}(\boldsymbol{y} - T * \mathbf{1})$, where k is a constant representing the area of the grid points relative to the area of the wafer. Hence, ignoring the constant k, criterion 2 minimizes the expected value of the sum of the squared errors, as given above in equation (1).

5. APPLICATION TO RTCVD PROCESS

5.1 Model Fitting

For the wafer fabrication process described in Section 2, the matrix $\boldsymbol{X}$ is such that the mean and variance models are

$$\begin{array}{rcccccc} E(y_{is}) & = & \alpha_i & + & \tau_s & + & (\alpha\tau)_{is} \\ & & \uparrow & & \uparrow & & \uparrow \\ & & \text{design} & & \text{spatial} & & \text{interaction} \\ & & \text{effects} & & \text{effects} & & \text{effects} \\ & & \downarrow & & \downarrow & & \downarrow \\ \log\sqrt{Var(y_{is})} & = & \alpha_i^* & + & \tau_s^* & + & (\alpha\tau)_{is}^*. \end{array}$$

Maximum likelihood estimation is used to obtain the estimates of β, θ and δ. The estimated mean model is

$$984.5 - 260.2\text{ox} + 326.6\text{ti} + 58.8\text{ox}^2 - 42.7\text{ti}^2 - 97.6\text{ox*ti} + 0.96\text{d} - 11.6\text{d}^2 + 14.1\text{d*ox} - 8.6\text{d*ti},$$

where ox represents the coded oxide thickness [-1 represents 458 angstroms, 1 represents 1498 angstroms], ti represents the coded deposition time [-1 represents 18 seconds, 1 represents 52 seconds], and d represents the distance of the spatial location from the center point. The estimated log(standard deviation) model is

$$2.73 - 0.48\text{ox} + 0.08\text{d} + 0.07\text{d}^2 - 0.02\text{d*ox}.$$

The estimated correlation model is

$$exp(-0.61596 * h_{st} * exp(0.04077 * (d_s + d_t))),$$

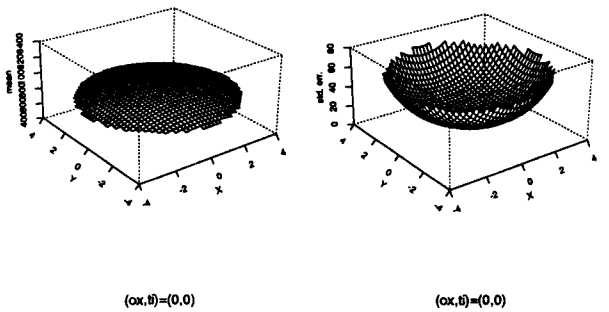


Figure 8: Estimated mean and standard deviation surfaces. Process conditions: ox=0, ti=0.

and likelihood ratio tests indicate the statistical significance of all three parameters $(\delta_1, \delta_2, \delta_3)$ of this model.

Figures 8–10 graph the estimated mean and standard deviation surfaces for selected settings of process conditions. Figure 8 shows that when oxide thickness and deposition time are at their center values (with respect to the experimental design) the mean surface is fairly uniform, but off target. At the same time the measurement locations closer to the edge show far more variability than those close to the center of the wafer. On the other hand, Figures 9 and 10 show both good achievement of target and repeatability.

To illustrate the correlation model, the finer grid of 64 spatial locations (not the 13 actual measurement locations) used in Section 3 is again used here (see Figure 5). Figure 11 contains the graph of the estimated correlation between these 64 locations. In order to demonstrate the effect of the "non-stationary" component of the correlation model, where distance to center point affects the correlation, Figure 12 shows the same correlation model *without* this piece.

5.2 Process Optimization

As previously stated, the goal of the experiment is to determine "optimum" levels of oxide thickness and deposition time, where "optimum" is defined as achieving the target thickness (at least within

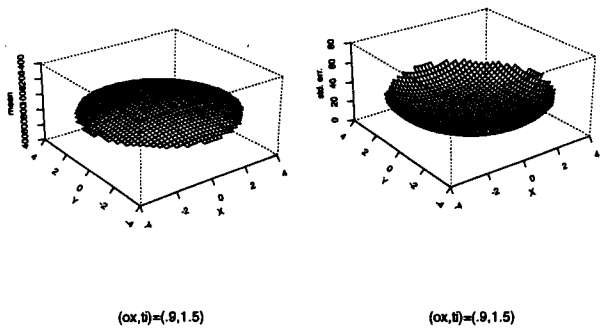


Figure 9: Estimated mean and standard deviation surfaces. Process conditions: ox=.9, ti=1.5.

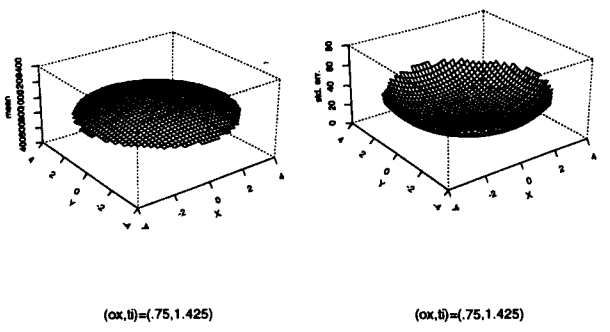


Figure 10: Estimated mean and standard deviation surfaces. Process conditions: ox=.75, ti=1.425.

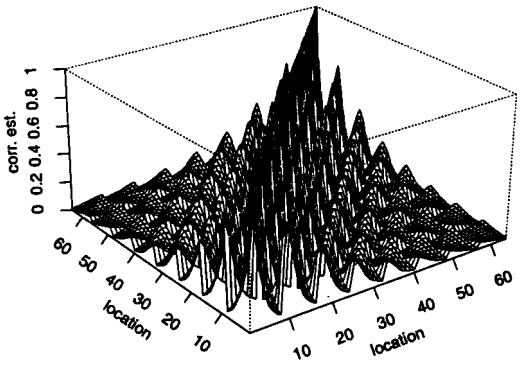


Figure 11: Estimated correlation surface.

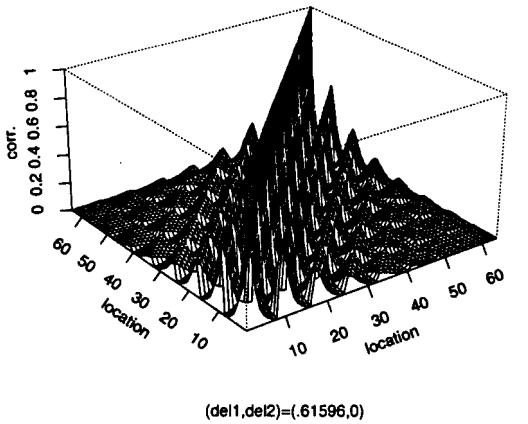


Figure 12: Modified correlation surface, without distance-to-center component.

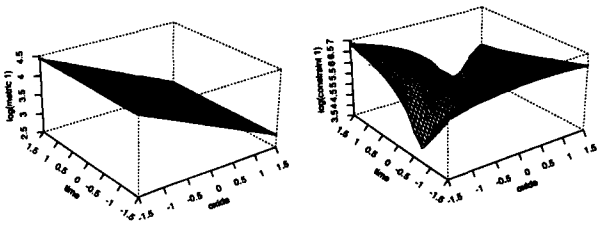


Figure 13: Optimality criterion 1

certain tolerance) while minimizing the variance of the thickness and achieving uniformity across the wafer. Based on the estimated models given in the previous section, criteria 1 and 2 are applied to this RTCVD process. A tolerance value of 60 angstroms (6% of the target) is used for criterion 1. Figures 13 and 14 show the logarithm of the metrics of the criteria over a grid of (ox, ti) settings. Figure 13 shows the average standard deviations, which is a plane since deposition time is not a part of the standard deviation model, and the average absolute error, which has a trough-like shape. Figure 14 for metric 2 has the same shape of the average absolute error surface, showing this as the dominant determinant of process conditions.

The optimal conditions, as obtained by the two criteria, are shown in Table 1. Criterion 1 has optimum process condition (.9, 1.5), and criterion 2 has the optimum of (0.75, 1.425). These optimal conditions equate to 1446 angstroms of oxide thickness, 60.5 seconds deposition time and 1368 angstroms oxide, 59.2 seconds deposition time, respectively. The mean and variance surfaces corresponding to these two sets of process conditions are displayed in Figures 9 and 10. Although these figures show good achievement of uniformity, target, and repeatability for both sets of operating conditions, the conditions determined by the first criterion yield slightly better repeatability at the expense of being more off target (see Table 1). This is not surprising, given the definition of the

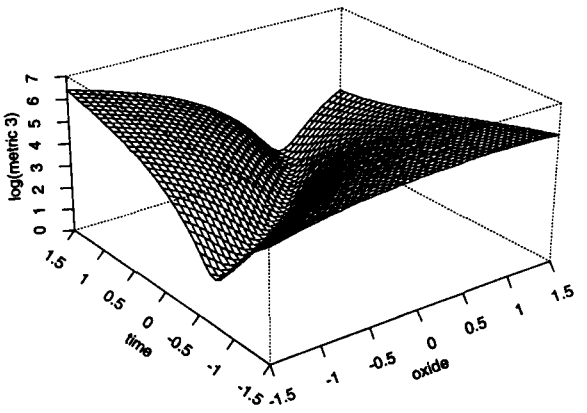


Figure 14: Optimality criterion 2

Table 1. Optimal process conditions for the two criteria.

	Optimum	Value of Metric for		
Criterion	(ox,ti)	Criterion 1	(Constraint 1)	Criterion 2
1	(0.9,1.350)	23.20393	58.60250	3578308
	(0.9,1.425)	23.20393	57.48159	3191891
	(0.9,1.500)	23.20393	56.47161	2899806
2	(0.75,1.425)	25.17286	51.31676	2321446

criteria.

6. DISCUSSION

This article contains recommendations for several key steps of robust process optimization. Models have been presented for not only the mean and variance, but also the correlation of spatial responses. It is very often the case that responses are correlated over space or time. The traditional approach in robust process optimization is to first "decorrelate" or make the responses independent prior to modeling the mean and variance surfaces. This article presents a technique flexible enough to model any correlation present while still allowing the possibility of independence. A particular spatial correlation structure has been suggested which goes further than the exponential decay model and, furthermore, does not assume second-order or covariance stationarity. It was designed for the particular wafer fabrication process presented, but is applicable to any situation for which it is

believed that the variability on an experimental unit is bowl-shaped.

Optimality criteria have also been presented and their use illustrated on a real problem. The metric involved in one of these criteria may also be used as an indicator of within-wafer uniformity, even in the absence of desires to optimize a process.

The issues considered here have long been present and are steadily being addressed in the semiconductor industry. For example, there are several proposals to combine the temporal and spatial information together for building a feedforward and feedbackward control system to continuously improve the process performance and for setting up process monitoring limits to guard against possible catastrophic faults. The search for optimal sampling and design of experiments in on-line and off-line improvement activities is important for facilitating spatial modeling and optimization studies. It is our hope that by providing a statistical view of these issues, albeit a very limited one due to the fact that this is the result of an initial study, we may generate more statistical activity in this area. We believe there is much potential for both theoretical and methodological advancements in this area, particularly with the continuation of current trends of productivity and quality improvement. Immediate concerns include the investigation of other correlation models, and diagnostic tools. Also of interest is the inferential properties of models obtained from a variety of estimation techniques, given the inclusion of the correlation model.

REFERENCES

Chang, N. H. and Spanos, C. J. (1991). Continuous equipment diagnosis using evidence integration: An LPCVD application. *IEEE Trans. on Semiconductor Manufacturing*, **4**(1), 43–51.

Davis, J. C., Gyurcsik, R. S., and Lu, J. C. (1993). Application of semi-empirical model building to the RTCVD of polysilicon. *Statistics in the semiconductor industry – Case studies of process/equipment characterization*, vol. 2, Chapter 4, 202–214. SEMATECH.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, **82**, 1079–1091.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman & Hall.

Guo, R. and Sachs, E. (1993). Modeling, optimization and control of spatial uniformity in manufacturing processes. *IEEE Trans. on Semiconductor Manufacturing*, **6**(1), 41–57.

Mesenbrink, P., Lu, J. C., McKenzie, R., and Taheri, J. (1994). Characterization and optimization of a wave soldering process. *Journal of the American Statistical Association*, **89**, 1209–1217.

Mozumder, P. K. and Loewenstein, L. M. (1992). Method for semiconductor process optimization using functional representation of spatial variations and selectivity. *IEEE Trans. on Components, Hybrids, and Manufacturing Technology*, **15**(3), 311–316.

Mozumder, P. K., Shyamsundar, C. R., and Strojwas, A. J. (1988). Statistical control of VLSI fabrication processes: A framework. *IEEE Trans. on Semiconductor Manufacturing*, **1**(2), 62–71.

Sachs, E., Guo, R., Ha, S., and Hu, A. (1991). Process control system for VLSI fabrication. *IEEE Trans. on Semiconductor Manufacturing*, **4**(2), 134–144.

Taguchi, G. (Technical editor for the English Edition: Don Clausing) (1987). *System of Experimental Design,* vols. 1 & 2, American Supplies Institute.