Equipment Fault Detection Using Spatial Signatures

Martha M. Gardner, Jye-Chyi Lu, Ronald S. Gyurcsik, *Member, IEEE,* Jimmie J. Wortman, *Senior Member, IEEE,* Brian E. Hornung, *Member, IEEE,* Holger H. Heinisch, *Student Member, IEEE,* Eric A. Rying, *Student Member, IEEE,* Suraj Rao, Joseph C. Davis, *Member, IEEE,* and Purnendu K. Mozumder, *Senior Member, IEEE*

***Abstract*—This paper describes a new methodology for equipment fault detection. The key features of this methodology are that it allows for the incorporation of spatial information and that it can be used to detect and diagnose equipment faults simultaneously. This methodology consists of constructing a virtual wafer surface from spatial data and using physically based spatial signature metrics to compare the virtual wafer surface to an established baseline process surface in order to detect equipment faults. Statistical distributional studies of the spatial signature metrics provide the justification of determining the significance of the spatial signature. Data collected from a rapid thermal chemical vapor deposition (RTCVD) process and from a plasma enhanced chemical vapor deposition (PECVD) process are used to illustrate the procedures. This method detected equipment faults for all 11 wafers that were subjected to induced equipment faults in the RTCVD process, and even diagnosed the type of equipment fault for 10 of these wafers. This method also detected 42 of 44 induced equipment faults in the PECVD process.**

***Index Terms*— Equipment fault diagnosis, process improvement, simulation, statistical metrology.**

I. Introduction

EQUIPMENT faults are often the cause of major variations in semiconductor manufacturing processes. Considering the expense of processing, these variations can cause dramatic yield losses [1]. Traditionally, the mean or signal-to-noise ratio of the wafer surface data is modeled, and the resulting model is used to detect equipment faults according to statistical process control (SPC) techniques; however, as wafer sizes increase and film thicknesses are reduced, the use of integrated spatial information will have a greater impact on detecting equipment faults.

Manuscript received March 24, 1997; revised September 29, 1997. This work was supported in part by Texas Instruments, Inc. and the NSF Engineering Research Centers Program through the Center for Advanced Electronics Materials Processing, Grant CDR 8721505, the Semiconductor Research Corporation, SRC Contract 94-MP-132, and the SRC SCOE program at NCSU, SRC Contract 94-MC-509.

M. M. Gardner was with the Department of Statistics, North Carolina State University, Raleigh, NC 27695 USA and is now with General Electric, Nishayuna, NY 12309 USA

J.-C. Lu are with the Department of Statistics, North Carolina State University, Raleigh, NC 27695 USA.

R. S. Gyurcsik is with the Semiconductor Research Corporation, Research Triangle Park, NC 27709 USA.

J. J. Wortman, H. H. Heinisch, and E. A. Rying are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA.

B. E. Hornung is with Motorola, Austin, TX 78721 USA.

S. Rao, J. C. Davis, and P. K. Mozumder are with Texas Instruments, Inc., Dallas, TX 75265 USA.

Publisher Item Identifier S 1083-4400(97)09163-8.

The use of site-specific models has been shown to have better sensitivity, with respect to spatially dependent process variations, than mean-based models [1]. However, detection of equipment faults identified from models based on data from different sites can have inconsistent results; i.e., some site models may detect a certain type of equipment fault, while other site models do not [1]. Saxena, *et al.* [2] have used a monitor wafer controller (MWC) to fix this to some degree. It has also been shown that the use of a virtual wafer surface, rather than specific sites on a wafer, captures even more information about the spatial signatures generated from different equipment conditions [3]. Kibarian and Strojwas [4] have also developed models which account for spatial dependencies and shown how the models can be used to separate spatial dependencies from other causes.

The detection and diagnosis of equipment faults in semiconductor processes is usually a two step procedure. Detection refers to the identification of the occurrence of an equipment fault, whereas diagnosis refers to the classification of equipment faults. Faults are detected using one method. Then faults are classified using another method. Current research in the literature has concentrated on equipment fault diagnosis, rather than the detection of the existence of equipment faults. For example, pattern recognition techniques including statistical discriminant analysis techniques [1], fuzzy logic techniques [5], and neural networks [6] have been used for diagnosis purposes. Hu *et al.* [7], Butler and Stefani [8], and Bombay and Spanos [9] have applied empirical (or semi-empirical) polynomial modeling techniques to relate process outputs to process settings, and May and Spanos [10] have used evidential reasoning to integrate in-line, off-line, and maintenance data for fault diagnosis. However, methods such as statistical discriminant analysis do not make use of the spatial information and the physical knowledge of equipment faults.

The equipment fault detection methodology described in this paper is unique not only in that it incorporates the use of integrated spatial information in a virtual wafer surface, but also in that it can be used to *detect* and *classify* equipment faults at the same time. The main focus of this paper is using the spatial signatures of the differences between observed and expected virtual wafer surfaces to construct physically based metrics which can be used to detect and diagnose various types of equipment faults. When establishing an equipment fault signature library, it would be ideal to have experiments conducted to model wafer spatial measurements at process conditions without faults and with certain known faults; however, historical data on existing faults may also be used. Using the

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

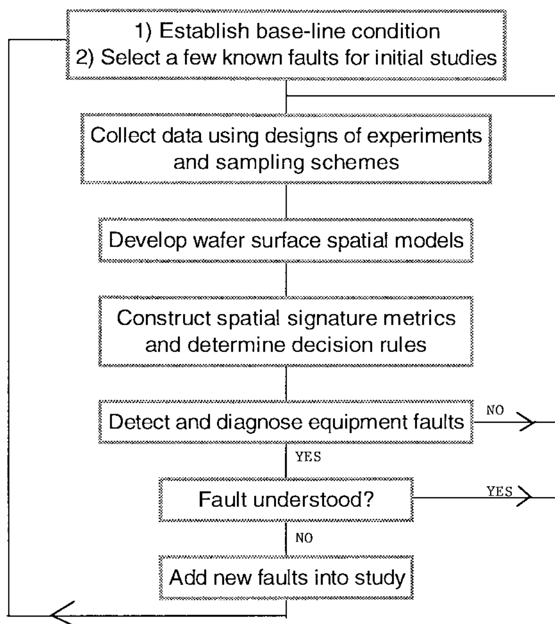

Fig. 1. Equipment fault detection and diagnosis chart.

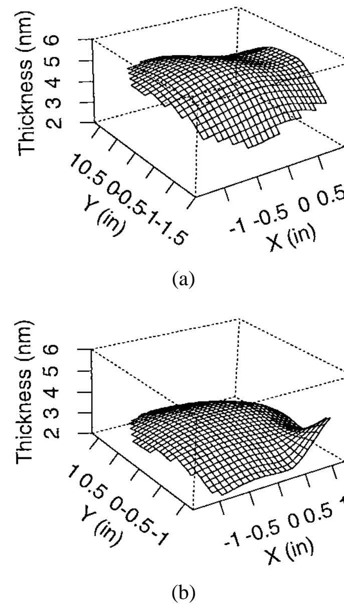

Fig. 2. Fitted wafer surfaces from wafers processed (a) with no equipment faults and (b) with known equipment faults.

experimental data, one can construct physically based signature metrics to detect and identify equipment faults. When the basic faults are understood, new faults can be added into the study. Fig. 1 shows a flow chart of all the steps in the process.

If a certain type of fault is known to have a specific shape, then classification of faults can also be verified by comparing a newly fitted surface to the known fault surface. In this case, by treating the fault surface as the "target," the methodology of spatial signature metrics can be used to statistically compare the newly fitted surface to this "target" to determine if the newly fitted surface belongs in this fault class.

Section II describes the equipment fault detection methodology using spatial signatures in detail. Sections III and IV provide illustrating examples from experiments conducted at North Carolina State University (NCSU) and Texas Instruments, Inc. (TI), respectively. Section V draws conclusions from this study and points to potential future work.

II. Fault Detection Methodology

Fig. 2(a) and (b) show how equipment faults can be manifest in the spatial response of the process. Fig. 2(a) shows the gate oxide thickness surface of a wafer that was processed under fault-free conditions. Fig. 2(b) shows the gate oxide thickness surface of a wafer processed under known equipment faults. X and Y represent the x and y distances from the center of the wafer. Not only is there an apparent decrease in thickness between the two surfaces, but also a change in spatial pattern.

The next five subsections present the new methodology of using spatial signatures to detect equipment faults:

1) modeling wafer surface data using thin-plate splines;
2) estimation of the baseline or "fault-free" surface;
3) construction of physically based signature metrics for comparing wafer surfaces;
4) estimation of the statistical distribution of metrics;

A. Modeling Wafer Surface Data Using Thin-Plate Splines

While recognizing that other modeling methods are available, this study uses thin-plate splines to model the virtual wafer surface. A virtual wafer surface model of spatial process behavior is less sensitive to the position of measurement sites, measurement error, and angular orientation than techniques focusing on individual data points [11]. The thin-plate spline can be viewed as a multi-dimensional extension of the cubic smoothing spline. Although splines, in general, are constrained to pass through the knots of the function [e.g., gate oxide thickness measurements at (x, y) distances from the center of a wafer], the thin-plate spline attempts to produce the smoothest curve possible between the knots and, therefore, does not have the requirement that the surface actually pass through the knots. The estimator of the thin-plate spline g is the minimizer of the following penalized sums of squares [12]:

$$\sum_{i=1}^{n}[y_i - g(x_i)]^2/n + \lambda J_m(g) \quad (1)$$

where the first term represents the lack of fit, $J_m(g)$ is the roughness penalty function, and λ is the spline smoothing parameter. For this study, thin-plate spline fittings were formed by using a collection of routines called FUNFITS written for use in the S-plus statistical software [13]. A thin-plate spline can then be used to predict the response at any location on the wafer and thus can be used to predict the entire wafer surface. In this study, λ is set to be very small (0.001) which gives more of an interpolating surface as recommended by Davis *et al.* [3].

B. Estimation of the Target Surface

A target surface needs to be specified for evaluating equipment performance. In many cases, the target surface may be known; e.g., a non-uniformity study where the target thickness

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

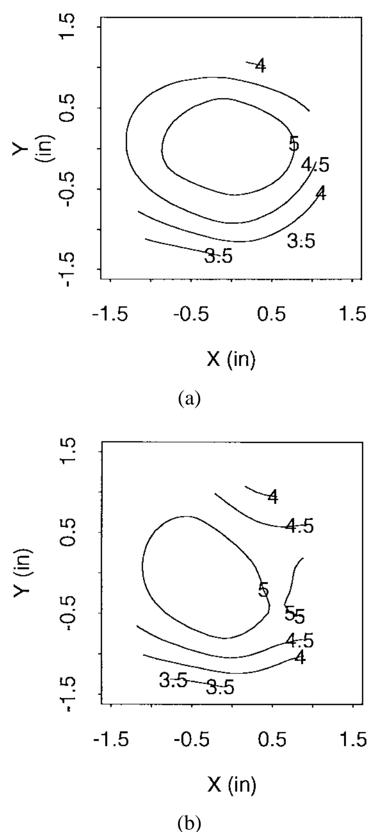

Fig. 3. Replicate wafer surfaces from wafers processed under fault-free conditions in a RTCVD experiment.

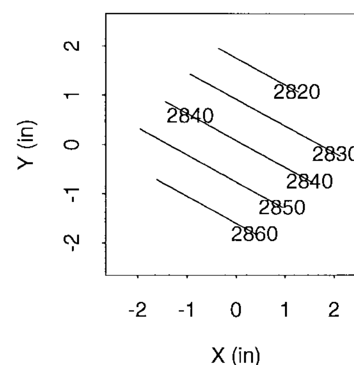

Fig. 4. Individual wafer surface from wafer processed under fault-free conditions in a PECVD experiment.

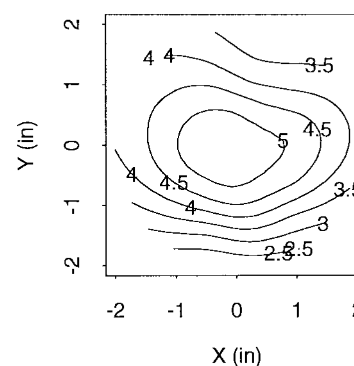

Fig. 5. Two replicates from fault-free conditions averaged to form target surface.

a result of equipment design, the wafer surface may not be flat even though there is no equipment fault. Assuming a flat target surface in this situation may lead to incorrect equipment fault detection. Thus, if the experimental wafer surfaces under equipment fault-free conditions are not flat, then these surfaces should be used as the target surface rather than a constant. For example, the two surfaces shown in Fig. 3(a) and (b) are the surfaces from two replicates at the fault-free condition in a RTCVD experiment conducted at NCSU and have a nonlinear pattern. Fig. 4 shows a surface from the fault-free condition in a PECVD silicon nitride experiment conducted at Texas Instruments, and this surface has a linear pattern. However, none of the surfaces shown in Fig. 3(a) and (b) or Fig. 4 reflect a constant baseline process surface.

In addition, the target surface should be validated after preventive maintenance or any other procedure which alters the tool. The proposed methodology can be used to determine if any significant changes in the tool have occurred. If no significant changes have occurred, then the new data can be used in conjunction with the historical data to update the target surface. The following method allows for a target surface to be estimated from data collected from fault-free runs.

Data is collected from wafers under the equipment fault-free condition to obtain a good estimate of the target surface. If there is slow drift, and this slow drift is considered to be typical phenomenon, then the wafers are still considered fault-free. Statistical outlier diagnosis can be used to screen the data. Af-

the target surface is obtained by averaging location specific parameter estimates from the individual spline equations. As an example, by averaging the two surfaces in Fig. 3(a) and (b) from the RTCVD experiment at NCSU, we obtain the target surface as shown in Fig. 5. A randomization procedure for use with wafer surfaces processed under the fault-free condition is currently being studied to better incorporate wafer-to-wafer variation in the proposed methodology, but this randomization procedure is beyond the scope of this paper.

A typical method for deriving the target surface is to first average the data collected from specific sites on "replicated" wafers at the fault-free condition and then fit a spline surface to the averaged data to create the target surface. However, this approach averages the data at n sites, where n is the number of data collected on a wafer, and requires that data be collected at the same sites on all wafers, as well as does not take into account any wafer-to-wafer variation. The approach described in the previous paragraph averages the spline estimates, which, in the intuitive sense, averages the spline surface at all possible sites. In the spatial signature metrics developed in next section, a grid of close to 700 sites is used for prediction.

If the purpose of the statistical test is to compare an average of wafer surfaces to a target, then averaging the wafer data first would be appropriate. However, our concentration in this work is to compare the spatial surface of a *single* wafer to the target, subject to random variation, and the variance is underestimated if the data from the target wafers are averaged before the spline is fit. For example, if the spline fits from three

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

averaging the variance after the splines are fit yields a variance of $\overline{V} = (V_1+V_2+V_3)/3$. Now if we let $\overline{X}$ represent the vector of averaged data, then since the variance of a mean $\overline{X}$ is σ^2/m where σ^2 is the variance of X, by averaging the data first, an extra factor of 1/3 will be introduced into the variation before the spline is ever fit. Thus, the variance of the spline fit when averaging the wafer data first will be $(1/m)$ times smaller than the variance of the spline fit to the individual wafer data, where m is the number of replicated wafers.

Another alternative is to treat data from all wafers processed under the fault-free condition as coming from a single wafer, then construct a spline to estimate the target surface. Although this method can capture variation at a particular site without deflating it, this method also loses individual wafer characteristics.

C. Construction of Physically Based Spatial Signature Metrics

Different equipment faults may produce distinct spatial signatures. For instance, an equipment fault may affect only a specific region of the wafer surface rather than the entire wafer surface. In this case, a certain performance evaluation metric may better detect this particular type of equipment fault. It is also possible that several metrics may have to be used simultaneously to detect certain types of faults. Understanding the physical processes that create faults, and their resulting signatures, also greatly aids in constructing and deciding what types of evaluation metrics to use. Four metrics are presented below as examples of how different metrics may be needed for detecting certain fault signatures. The metrics are extensions of the uniformity metrics presented by Davis *et al.* [3] with an expected surface used as the target surface. All metrics discussed here are based on loss functions. For all these metrics, g denotes a newly fitted thin-plate spline surface, T denotes the target surface, and $\mathbb{R}$ denotes the wafer surface region.

The quadratic and absolute loss functions are commonly used in many fields to quantify the penalty from departing from the target. The first two metrics used in this work are a squared deviate from target metric and an absolute value deviate from target metric. Both statistics are general metrics used to quantify the surface difference $(g-T)$ and are nonlinear functions of the error volume between two spline surfaces. The metrics are calculated as

$$\mathrm{M_{SQ}} = \int_{\mathbb{R}} (g-T)^2 \, d\mathbb{R}, \quad \mathrm{M_{ABS}} = \int_{\mathbb{R}} |g-T| \, d\mathbb{R}. \quad (2)$$

The squared metric, M_{SQ}, penalizes much more than the absolute metric, M_{ABS}, with respect to larger departures from the target. Both metrics cover the entire wafer surface and place equal weights on information at all wafer sites. In addition, both metrics yield the same equipment fault detection results in this study.

The following metric is an example of a metric that can be used to detect an equipment fault that leads to a thicker wafer surface. This metric is calculated as

$$\mathrm{M_{WT}} = \int_{\mathbb{R}} h_1(g-T) \, d\mathbb{R} \quad \text{if } g > T;$$

where h_i $(i = 1, 2)$ could be any functions. For example, h_1 could be the squared error loss function and h_2 set to 0. This example would only penalize surfaces thicker than the target. Another example is to place different weights on the penalties for $g > T$ than for $g \leq T$. Again, understanding the reasons for getting equipment faults and their resulting spatial signatures plays an important role in the selection of the h_i functions.

Another type of metric that should be considered is one which allows for different regions of the wafer surface to be weighted differently. For instance, error in the center of the wafer may be of more importance than error toward the edge of the wafer. Also, certain equipment faults may cause defects, such as a thinner surface, in specific regions of the wafer surface rather than the entire wafer surface. An example of a metric that weights wafer surface regions differently is calculated as

$$\mathrm{M_{REG}} = \sum_{i=1}^{k} \int_{\mathbb{R}_i} w_i h_i(g-T) \, d\mathbb{R}_i \quad (4)$$

where k denotes the number of nonoverlapping regions, and w_i and h_i denote the weight and penalty functions for the ith region, respectively. This metric has the potential to be very useful, particularly in the stage of equipment fault diagnosis, since it is more general than the other suggested metrics. In fact, the previous metrics may be considered as special cases of this metric.

D. Estimation of Signature Metrics

The calculation of the metrics discussed in Section II-C involves integration of the difference of two spline surfaces over certain regions. This integration can be done by using a numerical integration technique where the fitted surface g is evaluated on an $N \times N$ grid (with points outside the radius of the wafer removed) and the evaluated results, e.g., $(g-T)^2$, are then summed for each of the $N \times N$ grid points and multiplied by the area of one of the grid elements. The metrics may be approximated as [3]

$$\text{metrics} \approx h(\underline{T} - \mathbf{W}_g \underline{x})^T * \underline{1}_{N^2 \times 1} * A \quad (5)$$

where h is the loss function incorporated into the metric, $\underline{T}$ is an $N^2 \times 1$ vector of the target, $\mathbf{W}_g$ is an $N^2 \times n$ matrix of thin-plate spline coefficients for the measurements, $\underline{x}$ is an $n \times 1$ vector of the measurements, n is the number of measurements taken on the wafer, and A is the area of one of the grid elements. This approximation was shown to have good results for $N \geq 30$ [3]. Thus, $N = 30$ is used in the experimental examples presented in this paper.

E. Use of the Signature Metrics in Equipment Fault Detection

If the metrics indicate that the surface of a newly processed wafer is statistically significantly different from the target wafer surface, then the conclusion is that an equipment fault has occurred. Therefore, the null distributions of the metrics should be studied in order to set up the "cut-off value" for

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

TABLE I
WAFER LABELS FOR WAFERS PROCESSED UNDER EACH COMBINATION OF EXPERIMENTAL CONDITIONS

	Low Flow Rate	Medium Flow Rate	High Flow Rate
All Lamps Working	7	1, 11, 12	3, 8
Bottom Lamp Out	4	10, 14	2, 15
Side Lamp Out	5	6, 13	9

fitted surface and the target wafer surface. In other words, the distributions of the metrics under the "fault-free" condition are needed to determine the "cut-off values" in the tail(s) of the distributions for a specified level of significance.

An analytical approach based on standard statistical asymptotic normal approximation theory was first considered. Approximation theory cannot be applied in this study since traditional distributional spline results are for the independent identically distributed case; however, as n (the number of spatial measurements) goes to a very large number, many devices are being sampled on a fixed size wafer, and the data become more dependent because of spatial correlations. This resulting increasing dependence is called infill-asymptotics [14]. An alternate Bayesian (simulation) approach can be taken to determine the null distribution of the metrics using the following steps.

1) According to a procedure given in Green and Silverman [12], assuming a Gaussian prior distribution, the posterior distribution of the spline surface g has the following multivariate normal distribution [12]:

$$g \sim \mathrm{MVN}[\underline{\hat{g}}, \hat{\sigma}^2 \mathbf{A}(\hat{\lambda})] \tag{6}$$

where $\underline{\hat{g}}$ is the vector of fitted values, $\hat{\sigma}^2$ is calculated as: (the residual sums of squares about the fitted curve)/equivalent error degrees of freedom, and $\mathbf{A}(\hat{\lambda})$ is the projection matrix which maps the vector of observed values to their predicted values. Since there were multiple wafers processed independently at the baseline conditions in this study, the averages of the $\underline{\hat{g}}$'s and $\hat{\sigma}^2 \mathbf{A}(\hat{\lambda})$'s from all baseline wafers were used in (6).

2) The following parametric bootstrapping approach can be used to simulate independent observations from the null posterior distribution of the metric. First, 5000 sets of n observations are simulated from the multivariate normal model (6). A spline surface is fitted to each set of observations, and the spatial signature metrics are calculated using (2)–(5). As a result, 5000 independent observations are obtained from the null posterior distribution of each signature metric.

3) With a pre-specified significance level, $\alpha = 0.01$ (or 0.05), a "cut-off" value is defined as the 99th (or 95th) percentile from the null distribution of a specific metric. If the calculated metrics (2), (3), or (4) for a newly fitted surface are larger than their respective "cut-off values," it is statistically significant at level α that the wafer was not processed under fault-free conditions. The decision rule is still applicable even if the sampling scheme changes since the target surface remains the same. The decision

and the Type II error (false negative) as follows: Fix $\alpha =$ Prob(Type I error) below a chosen level, e.g., 0.01, and select the test that minimizes $\beta =$ Prob(Type II error). The choice of α is up to the discretion of the practitioner. In this study, more conservative "cut-off" values were desired, so $\alpha = 0.01$ was selected.

III. EXPERIMENTAL VERIFICATION WITH NCSU LABORATORY EQUIPMENT

A. Design of Experiment and Data Collection

To test the proposed methodology, a laboratory experiment was conducted at NCSU where the following two types of equipment faults were induced in a prototype RTP (rapid thermal processing) single wafer system: lamps burning out and a miscalibrated SiH_4/Ar mass flow controller. The response measured was SiO_2 thickness. There were 15 wafers available for the experiment. Preliminary experiments, in which one lamp at a time was removed, were performed to decide on how many lamps should be disengaged during an experiment. The removal of a single lamp caused a marked decrease in oxide thickness. For example, when a side lamp was disengaged, a 13.5 Å decrease in average oxide thickness was observed. When a bottom center lamp was disengaged, the average oxide thickness decreased by 10.22 Å.

The final experimental design used to induce equipment faults was an unbalanced 3^2 design. While the N_2O flow was held constant at 500 sccm, three 10% SiH_4/Ar flow rates of 20, 25, and 30 sccm were used for low, medium, and high flow rates, respectively. Table I shows all possible experimental conditions along with the wafer labels of each wafer processed under each combination of conditions. The baseline "fault-free" state was the condition with all lamps working and with medium flow rate. Three replicates were allocated for estimating the target surface and for constructing the posterior distributions of the spatial signature metrics. For each wafer, the oxide thickness was estimated at 17 points. The sampling scheme was designed to cover the wafer regions evenly as shown in Fig. 6.

The oxide thickness was estimated by measuring the C–V characteristics of capacitors located at the measurement points using a Keithley 595 Quasistatic Meter and a Keithley 590 C–V Analyzer. The gate oxide thickness was then extracted from the C–V data while accounting for polysilicon depletion and quantum effects using a program written by Hauser [15]. Unfortunately, data could not be collected from wafer 5 (low flow rate and a side lamp out) since the wafer was damaged during processing. However, since the goal of this study is not an experimental design analysis, this was not a major concern.

B. Application of the Proposed Methodology to the Experimental Data

The laboratory experiment had three runs processed under the baseline condition. One of the wafers appeared to be consistently thicker than the other two wafers at all sampled sites. The data from this wafer was excluded from the esti-

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.