# Efficient Overlay Optimization of Stepper Correctables

**Warren W. Flack, Susan Avlakeotes, David Chen, and Gary E. Flores**
**Ultratech Stepper, Inc.**
**San Jose, CA 95134**

A crucial aspect of overlay optimization is proper selection of stepper input corrections. Automated metrology systems provide the ability to rapidly amass extensive overlay data on lithography systems and processes. The data can then be used to provide feedback in the form of stepper input correction terms to improve overlay. A common approach is an analysis of the overlay data using conventional grid and lens models to determine apparent corrections to be applied to the stepper. However, the standard models do not necessarily account for all the variability in the measured data. Determination of optimal corrections is further complicated by cross-correlation of the stepper input correctable terms. In these cases, the simple application of the grid and lens modeled terms will not provide optimal results.

The use of efficient experimental design techniques can be used to reduce the large uncertainty involved in determining and applying these stepper input corrections. Using traditional experimental factorial and response surface design techniques, a descriptive model was developed for the six grid correction terms. The resulting empirical model was generated by using a six factor Box-Behnken experimental design. Multiple wafers were run at these conditions and overlay was measured using an automated metrology system. This empirical model was used to derive the optimal set of inputs to the stepper. This descriptive model is compared with input settings determined from a conventional grid model.

## 1.0 INTRODUCTION

Enhancing the manufacturability and product yield of an advanced microlithography line requires the continuous monitoring and control of overlay. Automated overlay metrology systems provide an opportunity to collect extensive data sets on microlithography tools and processes. The data can be used to provide feedback in the form of correction terms to improve stepper overlay. To maximize the value of this data, it is necessary to improve and evolve monitoring and diagnostic techniques. This can most effectively be implemented using statistical analysis and statistical process control (SPC) techniques.

SPC typically requires that a subset of the wafers at the lithography system are sampled using in-process measurement systems for monitoring overlay. Next, the measurement data is analyzed using SPC techniques with appropriate overlay models and diagnostic routines. From this analysis essential information on the stepper and lithography processes is derived. Typical information consists of possible process alarms and optimal stepper correction terms for overlay control. The output of the analysis can be used to control the stepper and processes using standard control techniques [1].

Overlay performance can be enhanced by monitoring and optimizing stepper grid and lens correction terms. This is a major challenge due to the inherent multivariable nature of these parameters. For a microlithography system, the overlay errors can be divided into intrafield (within one field) and interfield (across the wafer) systematic sources. The intrafield sources (dX and dY) are typically modeled as follows [2,3]:

$$dX(x,y) = T_{ix} + M_x x - \Theta_i y + \Psi_x xy + \Psi_y x^2 + D_3 x(x^2+y^2) + D_5 x(x^2+y^2)^2 + R_X \quad (1)$$

$$dY(x,y) = T_{iy} + M_y y + \Theta_i x + \Psi_y xy + \Psi_x y^2 + D_3 y(x^2+y^2) + D_5 y(x^2+y^2)^2 + R_y \quad (2)$$

where $x$ and $y$ are the coordinate location relative to the center of the field. For equations (1) and (2), the linear terms include die shift in $x$ ($T_{ix}$) and $y$ ($T_{iy}$), magnification in $x$ ($M_x$) and $y$ ($M_y$) and rotation ($\Theta_i$). The nonlinear terms include trapezoid in $x$ ($\Psi_x$) and $y$ ($\Psi_y$), third order ($D_3$), and fifth order ($D_5$) distortion. R represents the residual component that does not fit the specified model. These intrafield parameters are difficult to monitor in a production environment because of the large amount of data that must be collected inside each field. Typically, these parameters are set and only periodically adjusted as part of a maintenance program [4].

The interfield or grid sources ($E_x$ and $E_y$) are typically modeled as follows [3]:

$$E_X(X,Y) = T_{gx} + S_x X - \Theta_x Y + R_X \quad (3)$$

$$E_y(X,Y) = T_{gy} + S_y Y + \Theta_y X + R_y \quad (4)$$

where $X$ and $Y$ are the coordinate locations on the wafer. The grid parameters include offset error in $X$ ($T_{gx}$) and $Y$ ($T_{gy}$), wafer scaling magnification in $X$ ($S_x$) and $Y$ ($S_y$), and wafer rotation in $X$ ($\Theta_x$) and $Y$ ($\Theta_y$). Here $R$ represents the residual component that does not fit the grid model. These six stepper grid parameters can be easily determined and monitored in production environments using in-line metrology.

For optimal overlay control, it is desirable to simultaneously monitor and control these six grid parameters. However, it is possible the parameters are partially correlated to various degrees. For example, baseline data for an Ultratech 2244i stepper was collected over a four month period [5]. Table 1 shows the partial correlation matrix of the six grid parameters collected over the period. Each value in the table represents the partial correlation between various pairs of grid parameters.

The corresponding labels in the row and column identify the appropriate pair for each entry. The main diagonal of the matrix shows a value of “1” since each grid parameter is directly correlated to itself. A measure of the relationship or coupling between each pair of grid values is indicated by the size of the partial correlation values, where a “1” indicates a direct coupling and a 0 indicates no dependence of the variables. Since every value in the table is nonzero, there is a partial correlation among the entire set of grid values. Of particular significance is the large correlation value of 0.5580 for the $x$ and $y$ rotation. This means that 55.8 percent of the variability of $x$ rotation accounts for the variability of $y$ rotation. Under these conditions the simple application of the interfield model will not provide optimal results.

The six grid parameters are mathematically orthogonal under standard $X$, $Y$ axis folding symmetry and $90^o$ rotational symmetry conditions [2]. This means that the grid parameters should not interact for measurement plans with symmetry and spatial coverage [6]. Therefore, the observed partial correlations in Table 1 could be the result of higher order distortions incorporated into these parameters. Mathematically this implies the $R_x$ and $R_y$ terms in equations (3) and (4) are actually functions of $X$ and $Y$:

$$E_x(X,Y) = T_{gx} + S_x X - \Theta_x Y + R_x(X,Y) \quad (5)$$

$$E_y(X,Y) = T_{gy} + S_y Y + \Theta_y X + R_y(X,Y) \quad (6)$$

Thus, it is necessary to determine how $R_x$ and $R_y$ terms depend on $X$, $Y$ and the grid terms to adequately account for the cross correlations. The most efficient technique is to use experimental factorial and response surface techniques (RSM) which allows descriptive modeling of $R_x$ and $R_y$ [7]. One well established methodology is a three level sequential strategy referred to as a Robust Process Design (RPD) [8]. In the first level screening design, the goal is to identify factors which are dominant based on a 95 percent significance criterion. Only factors meeting this criterion are investigated in future experimental designs, while the remaining factors are maintained at constant levels that are practical. Following the screening design, the significant factors are further characterized using a two-level factorial design, which is commonly referred to as a limited response surface. At this stage a quantitative model results that includes estimates of linear and interaction terms providing prediction over the experimental region. Finally the third level experiment is used for modeling higher quality predictions that incorporate linear, interaction and quadratic effects. This RSM model enables location of an optimal process region simultaneously satisfying the experimental goals.

For the grid model it is already known that all six parameters are important and thus it is not necessary to perform a screening design. Therefore, a comprehensive strategy can be formulated in which the entire range of all six factors (variables) is explored in a single experiment and used for achieving a final optimized process. The resulting empirical model for the $R_x$ and $R_y$ terms can then be used to derive the optimal set of inputs to the stepper.

# 2.0 EXPERIMENTAL DESIGN

## 2.1 Experimental Design Techniques

Several basic design techniques are essential in experimental design schemes to enhance the information obtained from the statistical analysis. These include randomized experimental trials, replicate experimental trials, center point trials and higher order design trials [7]. By complete randomization of experimental trials, the effects of any process sequence dependencies such as unsuspected trends or patterns are eliminated. The use of replicated trials provides an estimate of the standard deviation of model coefficients and is also needed for significance testing. This can be used to distinguish model lack-of-fit error from pure experimental random error. The use of center point trials and higher order design trials provide additional benefits. These last two techniques create an efficient design that enables an estimate of curvature effects and lack of model fit. This minimizes the required number of experimental trials by eliminating the need to use a complete design with higher resolution to model curvature effects.

## 2.2 Response Surface Design

A higher order experimental design was selected to provide a response surface for the interfield errors $E_x(X,Y)$ and $E_y(X,Y)$ which incorporates linear and nonlinear factor effects as well as interactions. Potential design schemes included a three level factorial, Box-Behnken, central composite, and quadratic designs [7]. A full three level factorial design for six factors would require at least 729 trials plus additional trials for examining lack of fit and pure error. In comparison, a six factor Box-Behnken design employs a subset of this full factorial design and requires only 54 trials. The geometric character of the Box-Behnken design is a hypercube with experimental trials located at midpoint factor levels. Because of this approach the design suffers potentially large model prediction error at the edge of the design region. The selected design also includes 10 additional center point replicates to distinguish the types of error sources. Table 2 shows the six grid factors and their low, mid and high settings for this experiment. The factor settings were established based on long monitoring of the stepper to be studied in this project. The experimental design and statistical analysis was calculated using JMP® software from SAS Institute [9].

# 3.0 EXPERIMENTAL METHODS

An Ultratech Stepper model 4700 was used for the grid model optimization experiments. The Ultratech Stepper is based on the 1X Wynne-Dyson lens design employing broadband gh-line illumination from 390 – 450 nm [10]. The stepper utilizes a darkfield alignment detection system with Enhanced Global Alignment (EGA), where five fields are sampled per wafer as shown in Figure 1. The alignment targets for each EGA site are located along the horizontal scribe across

the top of the stepper field. The target design is a cross with a width of 2.0 microns and a darkfield (valley) polarity. The stepper generates the six grid parameters for each wafer based on equations (3) and (4) using the five sampled EGA sites.

Bare silicon wafers of 150 mm diameter were used for this study. A first level mask pattern was defined in a 1.1 micron thick photoresist film. The second level was aligned and exposed in the same photoresist with EGA corrections applied at the stepper. A six factor Box-Behnken design was used to set the grid correctables for second level alighment. Ten center point replicates (no corrections) were systematically distributed through the experimental matrix.

The photoresist-to-photoresist overlay pattern was then measured using a KLA 5107 Coherence Probe Microscope [11]. The overlay structures are a standard box-in-frame design. Characterization of the metrology system was performed to quantify the tool error. Multiple measurements were taken on the overlay targets during the tool setup to establish a three sigma tool variance of 7 nm. Additionally, the Tool Induced Shift (TIS) was characterized and found to be less than 10 nm for layers up to 15 microns thick.

The wafer layout consisted of 16 fields, as shown in Figure 1, with field dimensions of 44 mm in *x* and 22 mm in *y*. All the fields were sampled to reflect process variability over the entire wafer. Overlay was measured at five points in the field. One wafer was aligned for each process subgroup run (n). Thus, each process group or wafer included a total 65 measurements per subgroup run. This provides sufficient statistical degrees of freedom for extraction of the six wafer grid parameters using KLASS III® software [12].

## 4.0 RESULTS AND DISCUSSIONS

### 4.1 Data Reduction and Analysis Methodology

A single average overlay measurement was calculated for each field to minimize potential measurement errors and remove systematic intrafield contributions. After data averaging the remaining interfield or grid sources ($E_x$ and $E_y$) can then be modeled based on equations (3) and (4) using standard least squares techniques.

Six model schemes were applied to the data from the experimental design and then compared for model fit, lack of error, pure error and data dispersion. The simplest scheme is to examine the *raw data* or uncorrected results over the entire design space. This simple scheme can be extended by *removing the centerpoint trials* (a second scheme) from the data set and then examining the residual error. This is informative since it provides an indication of the data dispersion from the experimental design.

A third scheme is based on applying the *exact grid* model to the data set using equations (3) and (4). This exact grid model assumes that the conditions applied to the stepper in the DOE represent

the actual model response of the system. For example, a 1 ppm stage scale input to the stepper would result in a 1 ppm stage scale as measured on the wafer. Thus from equations (3) and (4) the exact DOE conditions are applied to the grid coefficient terms. This case is informative since it illustrates the error associated with the inadequacy of the stepper to metrology model. In this scheme the residual error after removing the various DOE conditions can then be applied to determine the inadequacy of the model.

The exact grid model can be extended to produce a *corrected grid* model (a fourth scheme). In this case a least squares regression is applied to the data to determine the apparent or effective grid terms in equations (3) and (4). The fifth scheme is to use the *DOE grid* model which incorporates the possible interaction and higher terms for the six grid parameters. This is essentially the main motivation for the project, since previous work has shown that the simple corrected grid model has difficulties in describing the behavior of the system over a large range of process conditions. The last scheme is to use a *full DOE grid* model, which adds the effects of nonlinear stage terms. These are the quadratic terms for the $X$ and $Y$ stage coordinates, which can be modeled from the raw measurement data since there are multiple field sites across a wafer.

### 4.2 Model Comparisons

Figures 2a through 2f are histograms of the raw and residual $x$ overlay error over the experimental design space for the six model schemes. Above each histogram are the average and standard deviation for the given scheme. Comparison of the raw data (Figure 2a) and the residual error after removing the centerpoint trials (Figure 2b) illustrates that the overlay error due to centerpoint trials is negligible. Using the exact grid model (Figure 2c) shows an average of 0.064 μm and standard deviation of 0.045 μm while the corrected grid model (Figure 2d) shows a standard deviation of 0.045 μm. Thus, there is no significant difference in the data dispersion for either the exact or corrected grid models.

Application of the DOE grid model (Figure 2e) which incorporates second order and interaction terms results in a standard deviation of 0.036 μm. This provides an additional 20 percent reduction in the data dispersion over the design space. The full DOE model (Figure 2f) shows identical behavior to the DOE grid model. This suggests that second order or high order staging effects are insignificant for this system. It is apparent that the main benefit for the grid model is the use of interaction terms. Figures 3a through 3f show the corresponding results for the $y$ overlay over the experiment design space. Behavior for the $y$ overlay error was analogous to that observed for the $x$ overlay error. These histogram comparisons support the necessity to apply the DOE grid model to most accurately describe the behavior of the system.

### 4.3 Normalized Grid Terms

A useful method to compare the various models is to examine the relative magnitudes of the grid terms. Figure 4 summarizes the grid terms normalized to the exact grid model. The grid terms are

the coefficients determined from least squares over the entire experimental design space which indicates an average effect. Next the determined grid terms are divided by the range of the corresponding input conditions applied to the stepper. For example, since the *x* offset range for the experimental design was from -0.05 to 0.05 µm, the modeled term is divided by 0.05. Thus the normalized magnitude of the grid terms should ideally be exactly equal to 1.0. For the *x* offset and *y* offset grid terms, all four model schemes illustrate a normalized magnitude near 1.0. However, the behavior is quite different for the scale terms.

Depending on the grid model the *x* scale term can be in error by as much as 35 percent. The *y* scale grid term shows a large departure from unity when using the corrected model. The normalized term has a value of approximately 0.55 for the corrected model, while the partial and full DOE models show a comparable value near 1.05. The *x* rotation term also shows a large departure from normal expectations with a value of approximately 0.5, however, both the partial and full DOE models show values near 1.20. For the *y* rotation terms, the behavior is more consistent across all the model types. This summary of the grid terms shows the apparent difficulty in attempting to optimize overlay using simple input of the grid terms to the stepper. If one assumes a perfect or exact system, serious deviations can occur from the expected results.

### 4.4 Model Fit Errors

A summary of the model errors from least squares analysis is an informative method to illustrate the deficiencies of the simple grid models. Tables 3 and 4 summarize four grid models based on the experimental design results for the *x* and *y* overlay. Included in the tables are the exact grid model, corrected grid model, partial DOE model and the full DOE model. Adjusted $R^2$ estimates the proportion of the variation in the measured overlay errors that can be attributed to the terms in each model versus the entire variation of error in the data. It is apparent that moving from the exact grid model to the corrected grid model the adjusted $R^2$ shows an improvement from 0.243 to 0.428 and 0.187 to 0.515 for the *x* and *y* overlay errors. This means that for the x overlay error, 42.8 percent of the data variability is accounted for using the corrected model, while only 24.3 percent is accounted for assuming the exact grid model.

Additional metrics in the tables are the lack of fit errors and the pure errors (in units of microns squared). Pure error is based on the replicate center point trials and does not vary with model type. The difference of the residual or total error from the model and the pure error is the lack of fit error. The F ratio of the lack of fit and pure error is also shown in the tables. This tests the hypothesis that the lack fit error is due to a random event. If there is a significant effect in the model which is not totally explained, there will be a low F ratio. The Prob > F is the probability of obtaining a greater F value by chance alone if the specified model fits no better than the overall response. Significant probabilities of less than 0.05 are often considered evidence that there is at least one significant factor in the model and an insignificant proportion of the error is explained by lack of fit. For the exact and corrected grid models, the Prob > F is much less than 0.05, indicating that the need for an improved model due to lack of fit. For both the DOE model and full DOE grid

models the lack of fit is no longer a significant issue. Thus the DOE grid model is more than sufficient to describe the system behavior as well as for determining optimal conditions for overlay.

## 5.0 CONCLUSIONS

A six factor Box-Behnken experimental design was used to descriptively model the overlay performance of a stepper. This design was applied to the six grid correction terms and the resulting overlay errors were measured. Modeling of the overlay error was compared using six separate empirical model schemes to determine appropriate model terms. The assumption that the overlay errors are described exactly using a typical grid model showed serious departures from expected behavior. For example, this assumption accounted for less than 24.3 percent of the data variation. A corrected grid model was next applied which provided a large improvement where 42.8 and 51.5 percent of the data variation was modeled. However, the model fit results still illustrated a statistically significant lack of model fit and inconsistencies in the measured grid responses. Inclusion of interaction terms in the grid model based on the DOE provided an improvement in the adjusted $R^2$ to nearly 70 percent. This methodology has merit in descriptively modeling the behavior of a stepper system. Using such an empirically-based model has the potential of achieving faster convergence on optimal overlay settings in comparison to a conventional grid based model.

## 6.0 REFERENCES

1. W. Levinson, “Statistical Process Control in Microelectronics Manufacturing,” *Semiconductor International*, November 1994.

2. J. Armitage, “Analysis of Overlay Distortion Patterns,” *Integrated Circuit Metrology, Inspection and Process Control II Proceedings*, SPIE **921** (1988).

3. M. van den Brink, C. de Mol and R. George, “Matching Performance for Multiple Wafer Steppers Using an Advanced Metrology Procedure,” *Integrated Circuit Metrology, Inspection and Process Control II Proceedings,* SPIE **921** (1988).

4. A. Yost, W. Wu, “Lens Matching and Distortion in a Multi-Stepper, Submicron Environment,” *Integrated Circuit Metrology, Inspection and Process Control III Proceedings*, SPIE **1087** (1989).

5. G. Flores et. al., “Process Control of Stepper Overlay Using Multivariate Techniques,” *OCG Microlithography Seminar Interface ‘95 Proceedings*, (1995).

6. I. Fink et. al., “Overlay Sample Plan Optimization for the Detection of Higher Order Contributions to Misalignment,” *Intregrated Circuit, Metrology, Inspection and Process Control VIII Proceedings*, SPIE **2196** (1994).

7. G. Box, W. Hunter, J. Hunter, “Statistics for Experimenters,” *John Wiley & Sons* (1978).

8. T. Donnelly, “Robust Product Design”, *Machine Design*, October 1987.

9. JMP® Reference Manual, SAS Institute Inc. (1994).

10. R. Hershel, “Characterization of the Ultratech Wafer Stepper,” *Optical Microlithography Proceedings*, SPIE **334** (1982).

11. M. Davidson et. al., “First Results of a Product Utilizing Coherence Probe Imaging for Wafer Inspection,” *Integrated Circuit Metrology, Inspection and Process Control II Proceedings*, SPIE **921** (1988).

12. KLASS III® Reference Manual, Appendix Models and Algorithms.

| | X Offset | Y Offset | X Scale | Y Scale | X Rot | Y Rot |
|---|---|---|---|---|---|---|
| X Offset | 1.0000 | 0.1723 | 0.3008 | -0.2263 | 0.2877 | 0.2686 |
| Y Offset | 0.1723 | 1.0000 | -0.2923 | -0.0320 | -0.1442 | -0.1831 |
| X Scale | 0.3008 | -0.2923 | 1.0000 | -0.1903 | -0.1164 | 0.0824 |
| Y Scale | -0.2263 | -0.0320 | -0.1903 | 1.0000 | -0.2410 | -0.2828 |
| X Rotation | 0.2877 | -0.1442 | -0.1164 | -0.2410 | 1.0000 | **0.5580** |
| Y Rotation | 0.2686 | -0.1831 | 0.0824 | -0.2828 | **0.5580** | 1.0000 |

Table 1: Partial Correlation Matrix for the Six Stepper Grid Parameters

| Factors | Low | Mid | High |
|---|---|---|---|
| X Offset (μm) | -0.05 | 0 | 0.05 |
| Y Offset (μm) | -0.05 | 0 | 0.05 |
| X Scale (ppm) | -1.00 | 0 | 1.00 |
| Y Scale (ppm) | -1.00 | 0 | 1.00 |
| X Rotation (ppm) | -1.00 | 0 | 1.00 |
| Y Rotation (ppm) | -1.00 | 0 | 1.00 |

Table 2: Experimental Design Low, Mid, and High Grid Factor Settings

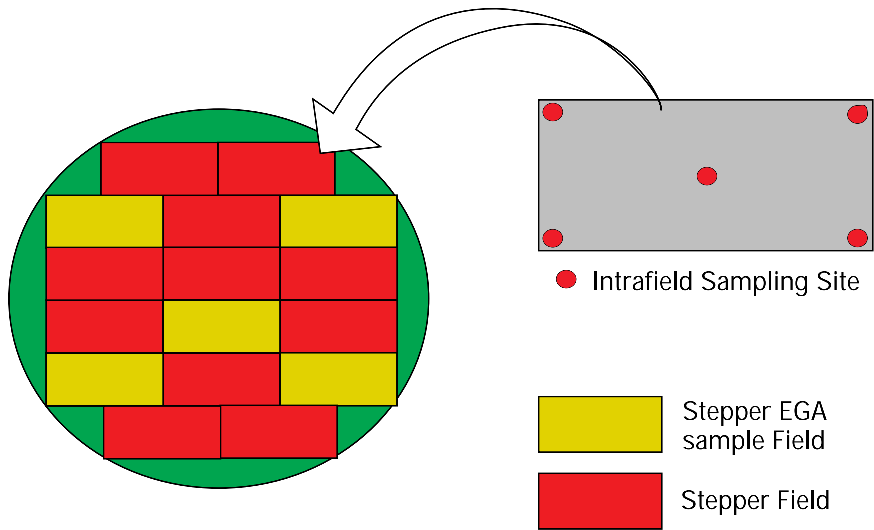



Figure 1: Wafer Layout and Sampling Schemes Used in Response Surface Design

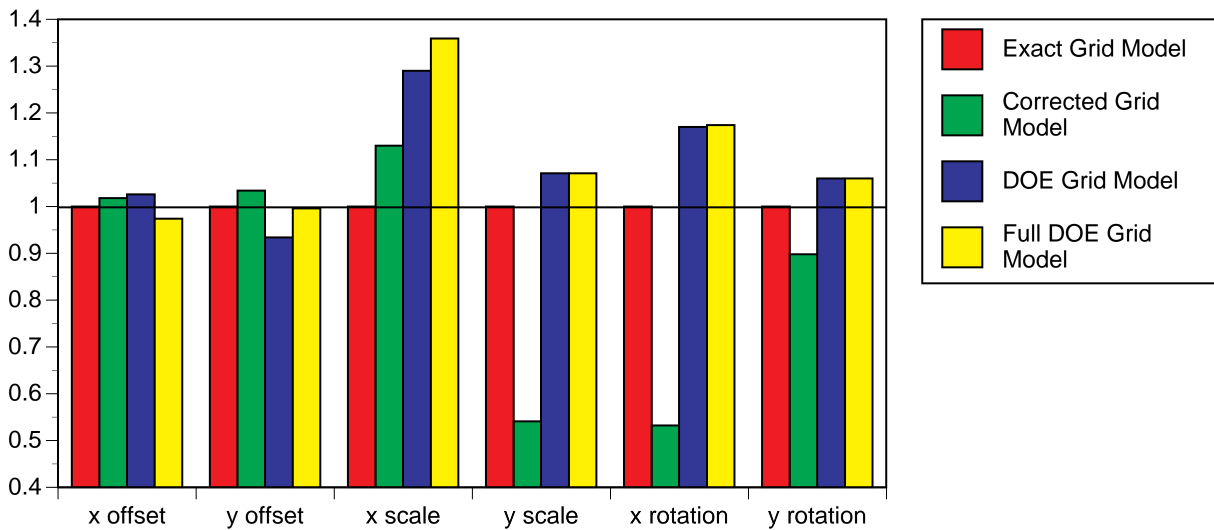
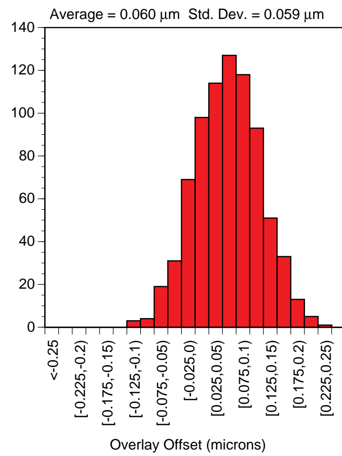
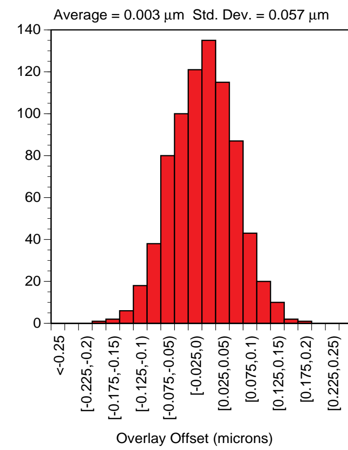



Figure 4: Normalized Magnitude of Grid Terms for Various Model Schemes

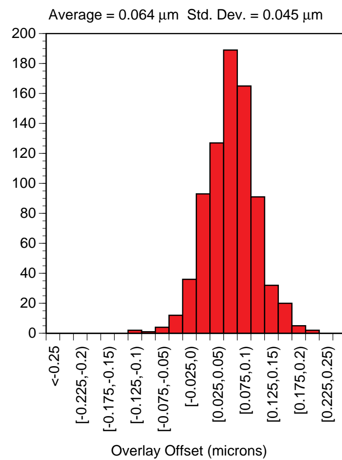
a) Raw data error



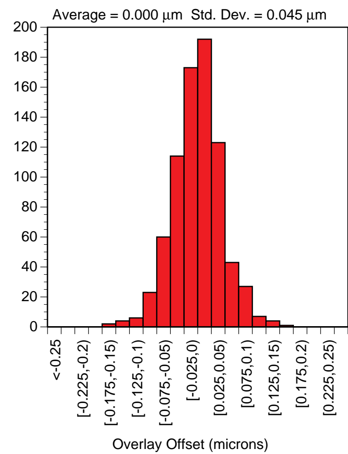

b) Residual error after removing center point trials



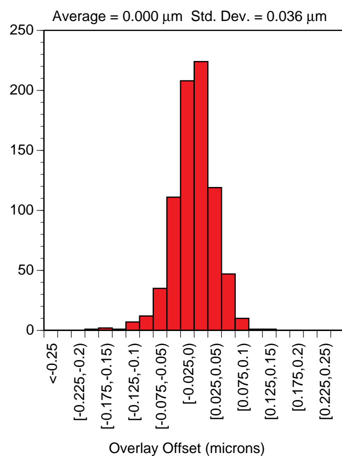

c) Residual error using exact grid model




d) Residual error using corrected grid model




e) Residual error using DOE grid model




f) Residual error using full DOE grid model

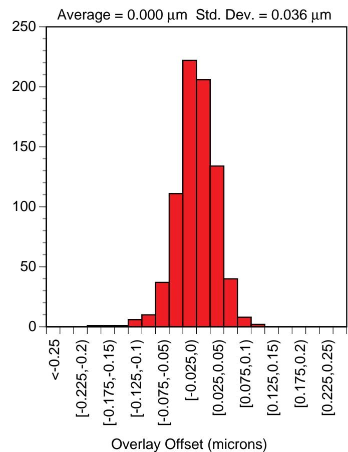
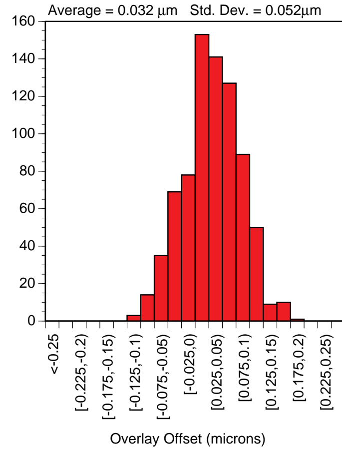
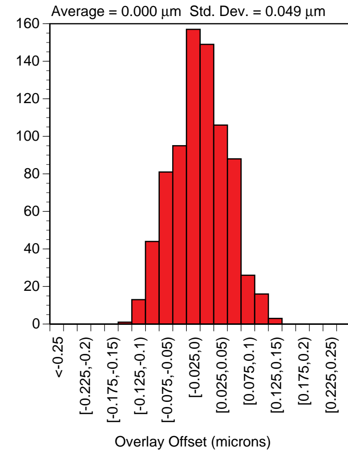



Figure 2: Histograms of Raw and Residual $x$ Overlay Error Over the Experimental Design Space

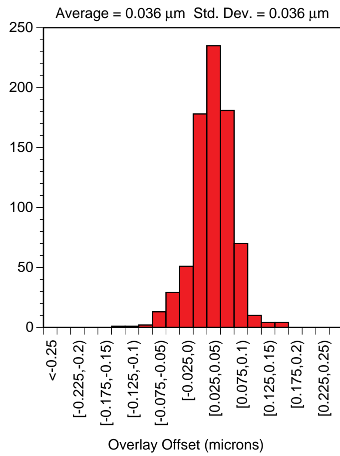
a) Raw data error



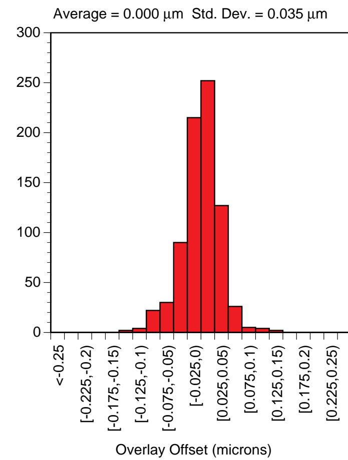

b) Residual error after removing center point trials



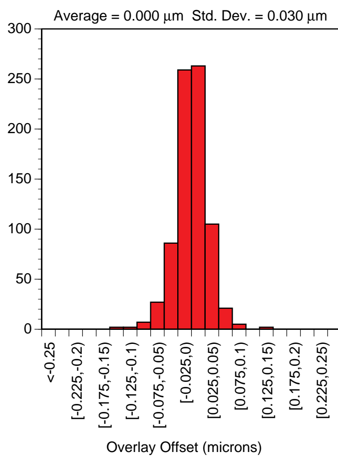

c) Residual error using exact grid model




d) Residual error using corrected grid model




e) Residual error using DOE grid model




f) Residual error using full DOE grid model

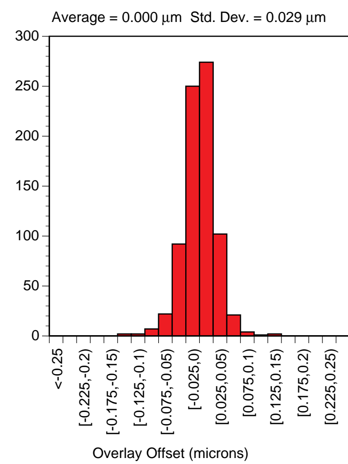



Figure 3: Histograms of Raw and Residual *y* Overlay Error Over the Experimental Design Space

| X Overlay | Exact Grid Model | Corrected Grid Model | DOE Model | Full DOE Model |
|---|---|---|---|---|
| Adjusted $R^2$ | 0.243 | 0.428 | 0.656 | 0.676 |
| Lack of Fit* | ** | 0.00529 | 0.0012 | 0.00121 |
| Pure Error* | ** | 0.00186 | 0.00187 | 0.0018 |
| F Ratio | ** | 2.852 | 0.653 | 0.642 |
| Prob > F | ** | 0.0001 | 1.0 | 1.0 |

*Mean Square Errors in Units of Microns **Not Calculated for Exact Grid Model

Table 3: Comparison of Model Fit of Various Grid Models Based on the Experimental Design Results for *x* Overlay Error

| Y Overlay | Exact Grid Model | Corrected Grid Model | DOE Model | Full DOE Model |
|---|---|---|---|---|
| Adjusted $R^2$ | 0.187 | 0.515 | 0.646 | 0.695 |
| Lack of Fit* | ** | 0.00282 | 0.00102 | 0.00099 |
| Pure Error* | ** | 0.00123 | 0.000784 | 0.000784 |
| F Ratio | ** | 2.29 | 1.23 | 1.27 |
| Prob > F | ** | 0.0002 | 1.0 | 1.0 |

*Mean Square Errors in Units of $\text{Microns}^2$ **Not Calculated for Exact Grid Model

Table 4: Comparison of Model Fit of Various Grid Models Based on the Experimental Design Results for *y* Overlay Error