

Improving Optical Measurement Accuracy using Multi-Technique Nested Uncertainties

R.M. Silver*^a, N.F. Zhang^a, B.M. Barnes^a, H. Zhou^a, A. Heckert^a, R. Dixon^a, T.A. Germer^a,
and B. Bunday^b

^aNational Institute of Standards and Technology, 100 Bureau Dr. MS 8212, Gaithersburg, MD, USA 20899-8212; ^bInternational Sematech, 2706 Montopolis Dr., Austin, TX USA 78741-6408

ABSTRACT

This paper compares and contrasts different combinations of scatterfield and scatterometry optical configurations as well as introduces a new approach to embedding atomic force microscopy (AFM) or other reference metrology results directly in the uncertainty analysis and library-fitting process to reduce parametric uncertainties. We present both simulation results and experimental data demonstrating this new method, which is based on the application of a Bayesian analysis to library-based regression fitting of optical critical dimension (OCD) data. We develop the statistical methods to implement this approach of nested uncertainty analysis and give several examples, which demonstrate reduced uncertainties in the final combined measurements. The approach is also demonstrated through a combined reference metrology application using several independent measurement methods.

Keywords: nested uncertainty, uncertainty analysis, Bayesian optical critical dimension, reference metrology

1. INTRODUCTION

Recently, there has been significant research investigating new optical technologies for critical dimension and overlay metrology for the 45 nm node and beyond. This work has been focused in two primary areas, scatterometry and more recently, scatterfield microscopy, a technique that combines well-defined angle-resolved illumination with image forming optics [1-4]. Experimental results have demonstrated nanometer-scale accuracy across a range of targets using angle-resolved scatterfield microscopy [5]. In this work the reflected intensity is measured using a well-characterized optical microscope operated in a scanned illumination mode. The experimental signatures are then compared with electromagnetic scattering simulations using a comprehensive parametric analysis. Although the introduction of an optical train significantly expands measurement system and target flexibility, the use of these hardware components introduces optical errors, which must be correctly normalized.

When modeling optical measurements, a library of curves is assembled through electromagnetic simulation over a multi-dimensional parameter space. A least square fitting routine is then used to choose the optimum set of parameters that yields the closest experiment-to-theory agreement. This approach assumes that the model is adequately describing the physical conditions and that an acceptable goodness of fit is achieved with the best set of parameters. However, parametric correlation, measurement noise, and model inaccuracy all lead to measurement error and measurement uncertainty in the fitting process [6]. Physical modeling inaccuracies and measurement noise can in principle be addressed directly through improved hardware, better sample parameterization, and improved optical characterization and normalization procedures. However, fundamental limitations exist as a result of parametric correlations.

Cross-correlations between parameters can lead to significant uncertainties even when a measurement technique demonstrates good sensitivity to a single parameter. As an example, an angle scan may have excellent sensitivity to a variation in the critical dimension (CD), but if that parametric response is coupled to a change in sidewall angle, then it reduces ones ability to identify the individual parameter that is changing. This ambiguity can lead to large measurement uncertainties in those parameters of interest. In this paper we outline a strategy to decouple parametric correlation and reduce measurement uncertainties. These techniques are applied to scatterometry and scatterfield measurements. The basis for this new approach is the systematic combination of multiple measurement methods designed specifically to reduce cross-correlation effects, resulting in reduced uncertainty in a given parameter of interest.
*silver@nist.gov; phone 1 301 975-5609; fax 1 301 869-0822; www.nist.gov

The goal is to improve measurements of those attributes that affect device performance or enable better process control. For example, when measuring CD, which is correctly broken into top CD, middle CD, and bottom CD, it may also be important to know line edge roughness (LER) and other geometrical properties of interest. This becomes complicated with optical measurements because we need to know the optical properties of the sample materials and there may be several different materials present and each of the materials has an estimated, but not perfectly known set of optical properties as well as height and geometry.

Recent studies have shown that OCD measurements are fundamentally limited by the correlation between measurement and fitting parameters [7]. The fundamental resolution and accuracy limits are due to the uncertainty and underlying correlation between sample characterization parameters and instrument parameters. If there is measurement noise or uncertainty in a measurement/tool parameter exists, the extent to which one can know the value of any given geometrical or materials parameter is limited. Ultimately the excellent sensitivity of optical metrology presents itself as a “double edge sword” with a price in parametric correlation and uncertainty.

The statistical approach that we have developed allows us to bring together different measurement techniques that may each have their own best attributes, such as excellent bottom width response versus sensitivity to sidewall changes or alternatively high throughput/good sensitivity production worthy metrology versus low throughput and excellent accuracy reference metrology. This approach allows us to combine the uncertainties of two individual measurements and arrive at a combined measurement uncertainty that is an improvement over the individual tools. Reference metrology performance can then be enhanced with input measurements from conventional reference instruments or *a priori* manufacturing knowledge used to restrict a fitting variable.

The goal of this work is to develop the statistical basis to use *a priori* information either gleaned from other measurement instruments or known based on physically relevant constraints. As an example, although n and k values can be independently measured, we should not use them in optical fitting routines as a fixed variable. They must be allowed to float. This approach allows us to effectively limit the range of float using known parameter estimations or uncertainty. In this application, the simulation volume of the n dimensional space can be minimized substantially, thereby reducing both simulation/processor time and library fitting time.

Although many of the examples discussed here will use optical techniques for illustration purposes, the same strategy is applicable to other measurement methods, such as scanning electron microscopy (SEM).

2. THE SCATTERFIELD HARDWARE PLATFORM AND DATA ACQUISITION

The technique of scatterfield microscopy has been described in detail elsewhere [8]. The basic instrument is based on a Köhler illuminated bright field microscope with a large accessible conjugate back focal plane. Data are acquired as a function of angle as shown in Figure 1. A charge coupled device (CCD) image is captured at each angle. A window (kernel) is placed in the image, the total intensity for that window is integrated and a normalized intensity per unit area is calculated. The intensity is then plotted as a function of angle. The intensity pattern may include only a zero-order specular reflection component or higher order diffraction components.

This mode of operation is similar to conventional scatterometry except that the high magnification optical train allows small targets or several targets to be measured simultaneously. We can perform massively parallel scatterometry measurements throughout the field of view by breaking the imaged field into an array of small targets or pixel groupings. Alternatively, we can perform scatterometry-type measurements on very small, embedded targets since the signals are spatially resolved.

Accurate modeling capabilities are a requirement when using optical methods that rely on library-based fitting. This, in turn, necessitates proper experimental normalization procedures to ensure accurate angle-resolved measurements. Although described in Ref. 8, a background is acquired for each image as a function of angle, either directly during image acquisition or at a nearby site. The angle-resolved background scan is then normalized using a known silicon reflectance. The image is then normalized to this background allowing the experimental data to be accurately compared to theory. This is essential due to the suite of tool-dependent effects that must be compensated.

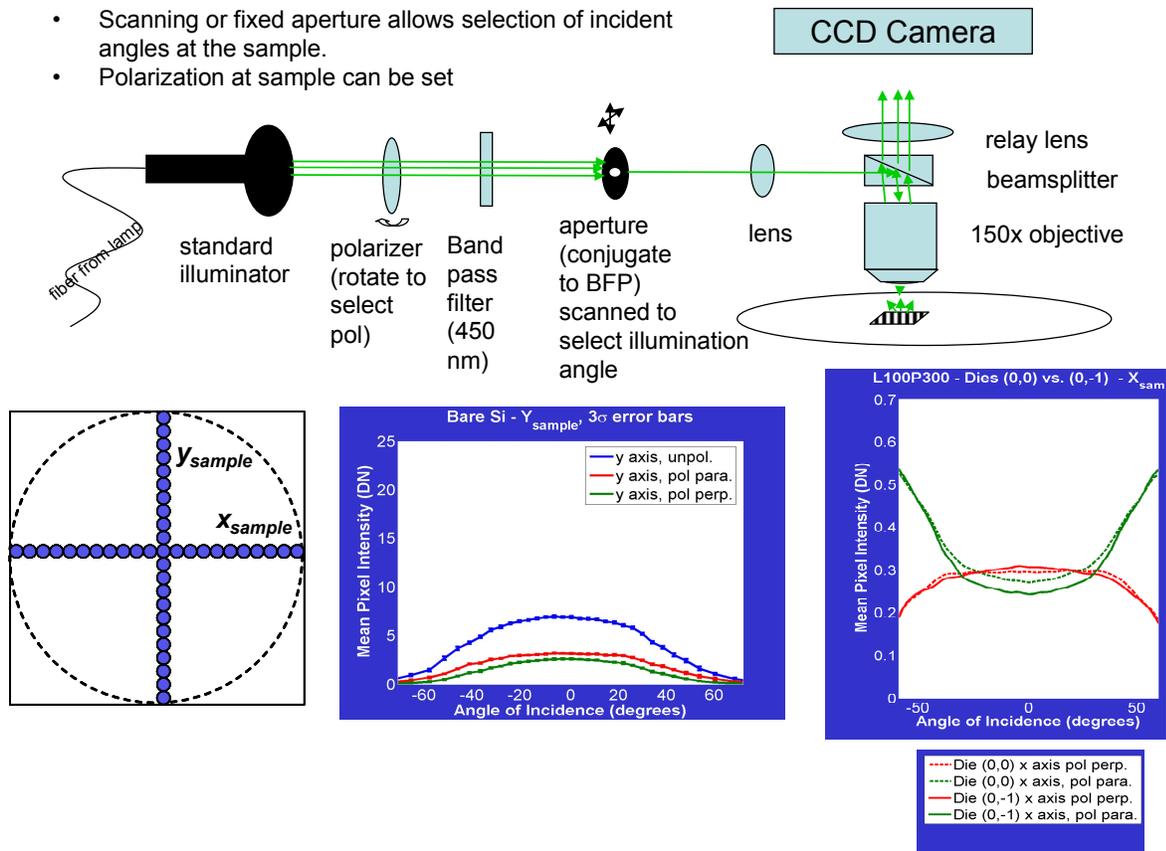


Fig. 1. Schematic of the experimental apparatus is shown in the upper figure. The lower portion shows the scan axes, a set of silicon normalization curves and the resulting data after raw angle-resolved reflectivity data are accurately normalized.

Once the data are acquired and normalized, they can be analyzed using library-based fitting techniques. The data are compared to simulation after building a comprehensive modeling library that typically involves a parametric analysis that includes varying n and k , height, pitch, sidewall, and CD. The starting point range for the geometrical parametric variations is typically based on atomic force microscopy (AFM) reference metrology that is provided with sidewall values and full uncertainty analysis or CD SEM reference values.

Comprehensive simulations were completed for the line arrays shown throughout this paper using an rigorous coupled waveguide analysis (RCWA) model while an finite difference time domain (FDTD) model was used for more complicated three-dimensional structures [9-10]. It can be seen in the analysis that the n and k parameters have a substantial effect on the intensity curves. Once an appropriate simulation library has been generated, a chi-squared analysis has been historically used to determine the best least square fit and evaluate the uncertainties in the fitting process [7].

3. CHI SQUARED UNCERTAINTY ANALYSIS

3.1 Basic formalism

The derivation of a general formula for multiple floating parameters is well-established [11]. We show the derivation of this formula for two parameters below to relay the importance of mathematical correlation effects and their physical meaning in the treatment of parametric uncertainty.

We start from the general linear least squares data fitting model and attempt to minimize the quantity

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - y(x_i; a)}{\sigma_i} \right]^2, \quad (1)$$

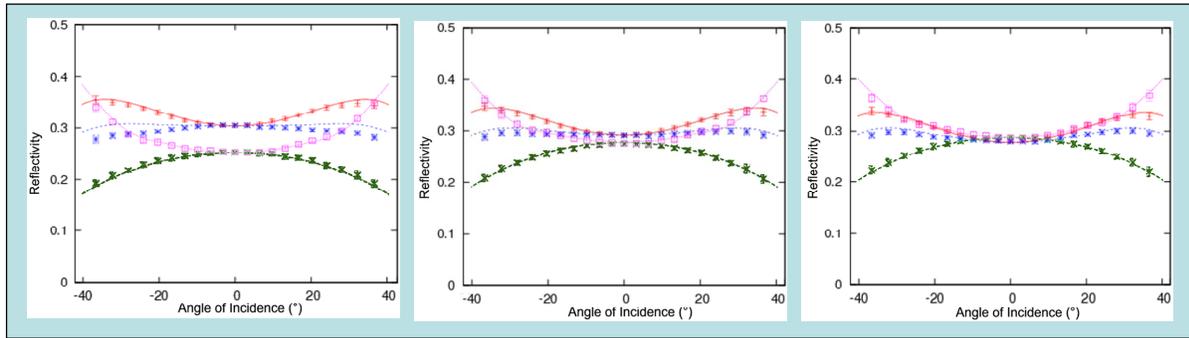
where a represents all floating parameters, σ_i is the uncertainty for each floating parameter, $a_1 \dots a_N$, y_i is the the measurand, and $y(x_i; a)$ is the simulated value for y_i at a given location within our parameter space a . It can be shown that the propagation of errors leads to a *sensitivity matrix* with elements

$$\alpha_{jk} = \sum_{i=1}^N \frac{1}{\sigma^2} \left[\frac{\partial y}{\partial a_j} \frac{\partial y}{\partial a_k} \right], \quad (2)$$

where a_i and a_j are separate floating parameters, e.g. critical dimension (CD) and line edge roughness (LER). The inverse of this sensitivity matrix is the covariance matrix, C_{jk} , where the estimated variance of a specific parameter is given by the diagonal elements

$$\sigma_{a_i}^2 = C_{ii}. \quad (3)$$

An example of a set of data is shown in Figure 2. These data demonstrate good agreement between the simulated library of curves and the experimental data using the experimental normalization procedure described above. The goodness of fit values and residuals are acceptable; however, the uncertainty in the fitting process is greater than desired, 1 nm to 2 nm, 1σ . This is an example in which correlation effects due to cross-correlation between the several fitting parameters and measurement noise are introducing uncertainty in the final results and ultimately undermining the confidence limits in the measurements.



Top = 115 nm
Middle = 110 nm
Bottom = 136 nm

Top = 118 nm
Middle = 107 nm
Bottom = 142 nm

Top = 120 nm
Middle = 112 nm
Bottom = 143 nm

Fig. 2. An example of experimental data and library data fits for three die from the overlay metrology advisory group (OMAG) 3 wafer. Good agreement with the reference values was obtained for the top and middle CDs.

To better understand the cause of correlation effects, we can look at the sensitivity matrix and covariance matrix and examine the effects of parametric correlation. To illustrate the role of the sensitivity matrix in the uncertainty analysis, we will focus on a simple two-parameter case. Suppose we have a given line width sample with an unknown CD and unknown LER. If we note CD as a_1 and LER as a_2 , we can display the entire 2x2 sensitivity matrix as

$$\begin{bmatrix} A & B \\ B & D \end{bmatrix}, \quad (4)$$

where we define A, B, and D as

$$A = \sum_i \frac{1}{\sigma^2} \left[\frac{\partial y}{\partial a_1} \right]^2, \quad B = \sum_i \frac{1}{\sigma^2} \left[\frac{\partial y}{\partial a_1} \frac{\partial y}{\partial a_2} \right], \quad \text{and} \quad D = \sum_i \frac{1}{\sigma^2} \left[\frac{\partial y}{\partial a_2} \right]^2. \quad (5)$$

Taking the inverse of the sensitivity matrix, we arrive at the covariance matrix, which is

$$\begin{bmatrix} \frac{D}{AD-B^2} & -\frac{B}{AD-B^2} \\ -\frac{B}{AD-B^2} & \frac{A}{AD-B^2} \end{bmatrix}. \quad (6)$$

Focusing on the uncertainty of one parameter, *e. g.* the critical dimension,

$$\sigma_{CD}^2 = \frac{D}{AD-B^2} = \frac{1}{A} \left(\frac{AD}{AD-B^2} \right). \quad (7)$$

When $B \neq 0$, the uncertainty of CD is increased by a cross-term factor

$$F = \left(\frac{AD}{AD-B^2} \right). \quad (8)$$

Since $A > 0$, $D > 0$, and $AD - B^2 > 0$, it follows that $F > 1$ as well. Stated differently, the larger the cross-term B gets, the larger this multiplicative factor F becomes, increasing the uncertainty of our floating parameter.

This mathematical picture becomes much more complex when more than two parameters are involved, as the contributions of cross-terms in the sensitivity matrix are much more convoluted than presented above. However, by calculating cross-term factors between each pair of parameters, we still can identify the major contributing factors to the uncertainties.

3.2 Contact holes example

To examine these correlation effects in more detail we next look at a more complicated three-dimensional contact hole process stack having several layers, thicknesses, and optical constants. For this example we use a simulation-based analysis comparing spectroscopic and angle-resolved scatterometry. The method for performing these simulation-based scatterometry sensitivity studies including modeled measurement and sample noise was discussed in Ref. 6. Simulations of the reflected and scattered fields from patterned contact holes were calculated using a commercially-available FDTD model. A schematic of the contact hole geometry is shown in Figure 3. The dimensions of these contact hole stacks studied correspond to the 22 nm and 32 nm nodes. For both cases, simulations have been performed both for scans of wavelength at certain fixed angles and for scans of incident angle at certain fixed wavelengths. These are commonly known as “angle scans” and “lambda scans.” For all the calculations performed, the incident light is assumed to be linearly polarized either in the transverse electric (TE) or transverse magnetic (TM) modes.

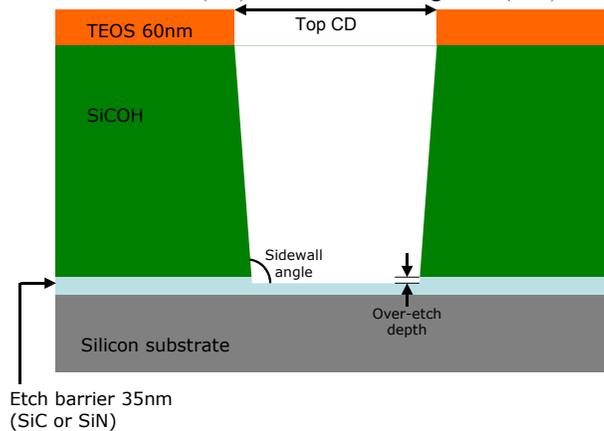


Fig. 3. Contact hole material stack and basic geometrical layout simulated for both 22 nm and 32 nm nodes.

The dimensions varied in the parametric analysis include sidewall angle, critical dimension for the diameter, and over-etch depth. The actual dimensions for the complete analysis are based on the 32 nm node and the 22 nm node. For the 32 nm node, the top width is 48 nm and the bottom width is 26 nm; for the 22 nm node, the top width is 33 nm and the bottom width is 18 nm. An example is shown in Table 1 for wavelength scan and angle scan uncertainty data for spectroscopic scatterometry and angle-resolved scatterometry for the wavelengths and incident angles indicated. These data are for 22 nm node contact holes. The data in the figure show the uncertainties for each of the parameters varied as well as the sensitivity matrix and covariance matrix as determined for various CDs, sidewall angles, and over-etch values based on a simulated fitting library used in the statistical analysis.

Examination of the data shows regions of high and low sensitivity in the sensitivity matrices for both the wavelength and angle scans. The diagonal sensitivity matrix elements are a direct measure of sensitivity for a given parameter; the off-axis elements are a measure of the cross-correlation. Uncertainties are shown for each of the three floating parameters; all other parameters are treated as fixed. CD uncertainties are typically compared with 2% of the top CD width to determine whether a measurement has performed at an acceptable level. Wavelength scans typically exhibit lower uncertainties than the angle scans when comparing the longer wavelength angle-resolved scans, while comparable uncertainties are achieved at the shorter wavelengths for specific polarization states, as seen in the figure. Similar uncertainties were observed at each of the angles simulated for the spectroscopic scans. However, the TM polarization data are interesting in that the sensitivity values are comparable to some of the wavelength scan values with the difference being the large correlation values in the covariance matrices. These correlation effects are important and will be further examined.

Table 1. These data show simulation sensitivity results, covariance matrices and the resulting uncertainties for different angle-resolved and wavelength-resolved scans. Sensitivity matrix values are in units of per varied unit squared and the covariance matrix is in varied units squared for the diagonal elements. Uncertainty values are 1σ.

Wavelength Scan					Angle Scan							
	σ (CD)	σ (SideWall Angle)	σ (overetch depth)	Sensitivity Matrix	Covariance Matrix		σ (CD)	σ (SideWall Angle)	σ (overetch depth)	Sensitivity Matrix	Covariance Matrix	
Wavelength Scan TE θ = 10°	0.21nm	0.12 °	0.26 nm	67.4 -62.8 -20.7 -62.8 170 -13.5 -20.7 -13.5 31.1	0.045 0.020 0.039 0.020 0.015 0.020 0.039 0.020 0.066		0.57nm	0.58 °	1.41 nm	55 -65.7 -5.96 -65.7 99.2 15 -5.96 15 4.15	0.326 0.321 -0.69 0.321 0.337 -0.759 -0.69 -0.759 1.99	
Wavelength Scan TE θ = 25°	0.36nm	0.17 °	0.48 nm	17.5 -29.9 2.18 -29.9 92.3 -8.7 2.18 -8.7 5.24	0.13 0.044 0.019 0.044 0.028 0.028 0.019 0.028 0.23		Angle Scan TE λ = 200nm	0.62nm	0.50 °	0.54 nm	37.4 -54.6 19 -54.6 86.4 -31.8 19 -31.8 15.5	0.38 0.28 0.107 0.28 0.253 0.175 0.107 0.175 0.292
Wavelength Scan TE θ = 45°	0.27nm	0.12 °	0.26 nm	28.2 -45.5 2.8 -45.5 146 -12.7 2.8 -12.7 4.09	0.074 0.026 0.029 0.026 0.018 0.039 0.029 0.039 0.347		Angle Scan TM λ = 200nm	1.82nm	1.47 °	2.36 nm	10.8 -19.8 -4.26 -19.8 43.1 12 -4.26 12 4.4	3.32 2.61 -3.91 2.61 2.15 -3.34 -3.91 -3.34 5.55
Wavelength Scan TM θ = 10°	0.21nm	0.09 °	0.20 nm	67.4 -62.8 -20.7 -62.8 170 -13.5 -20.7 -13.5 31.1	0.045 0.020 0.039 0.020 0.015 0.020 0.039 0.020 0.066		Angle Scan TM λ = 310nm	3.02nm	3.23 °	2.82 nm	18.2 -30.9 16.4 -30.9 55.8 -31.6 16.4 -31.6 19	9.14 9.64 8.07 9.64 10.4 8.99 8.07 8.99 7.98
Wavelength Scan TM θ = 25°	0.35nm	0.17 °	0.45 nm	17.3 -26.4 1.22 -26.4 81.7 -9.09 1.22 -9.09 6.2	0.122 0.044 0.041 0.044 0.031 0.036 0.041 0.036 0.206							
Wavelength Scan TM θ = 45°	0.32nm	0.22 °	0.73 nm	10.3 -5.33 1.01 -5.33 36.6 -6.62 1.01 -6.62 3.06	0.105 0.015 -0.002 0.015 0.047 0.097 -0.002 0.097 0.536							

Figures 4 shows reflectivity data calculated for spectroscopic scans in the top and angle-resolved scatterometry results in the lower part of the figure. These data are also from the 22 nm contact hole example. The uncertainties and covariance matrix are shown on the right side of the figure. Large regions of excellent sensitivity can be seen in the angle scans, yet the uncertainties are better for the wavelength scans. The angle scans are more sensitive than wavelength scans in this example. It is the larger cross-terms in the covariance matrix that lead to larger uncertainty values for CD, sidewall, and over-etch. It is important that the analysis with respect to parametric correlation highlights the double-edged sword of sensitivity as it can manifest itself in a potentially large cross-correlation between the floating parameters.

This raises the question of whether measurement uncertainty can be improved by restricting the range of floating parameters with reference metrology. Can we integrate SEM and optical measurements with appropriate sampling strategies or develop a best combination of angle scans and spectroscopic scans to optimize uncertainties by reducing correlation effects? The schematic on the left in Figure 5 illustrates the effect of small changes in two parameters and the resulting shift in the optical intensity profiles. If the two parameters are correlated, then the upshift in the reflectivity

curve could be due to a change in parameter a_1 or a_2 . If the effective floating range were reduced or known for one parameter, then the remaining movement in the curve could be attributed to the other parameter.

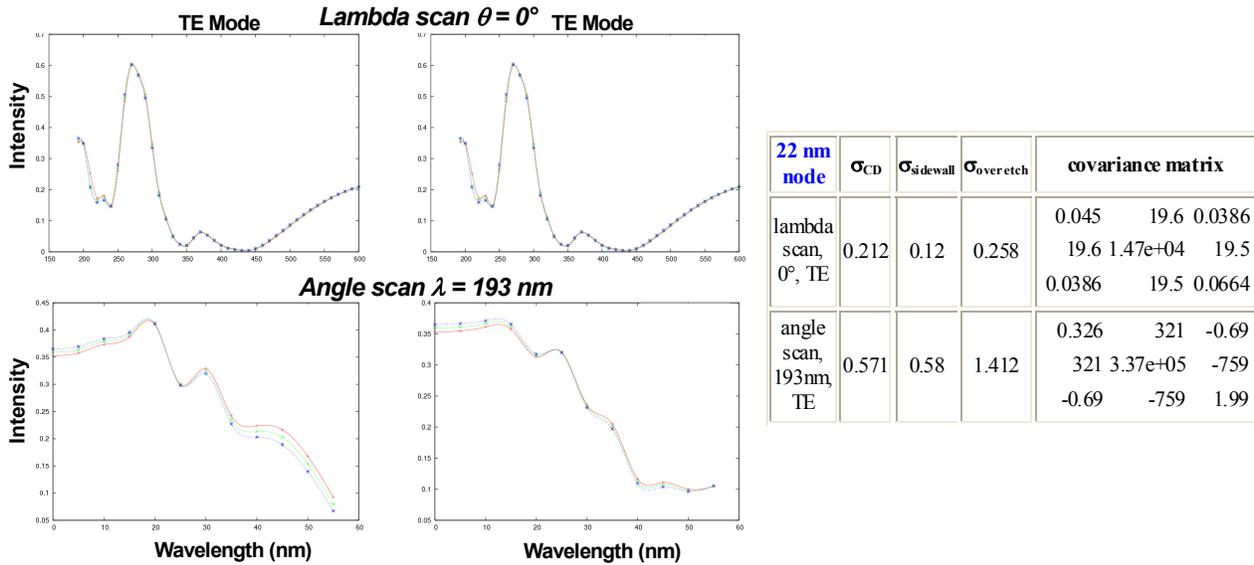


Fig. 4. TE and TM scans for different sidewall angles varied by 0.5° are shown. Angle scans are more sensitive than wavelength scans in this example. The parametric correlation analysis highlights the large correlation between floating parameters. The cross-terms in the covariance matrix lead to larger uncertainty values.

Without reference data or metrology, the confidence intervals of the fitting parameters have to enclose the elongated ellipse area seen on the right side of the figure. By employing the reference metrology in the parameters, we can enhance our confidence interval in the other correlated parameters.

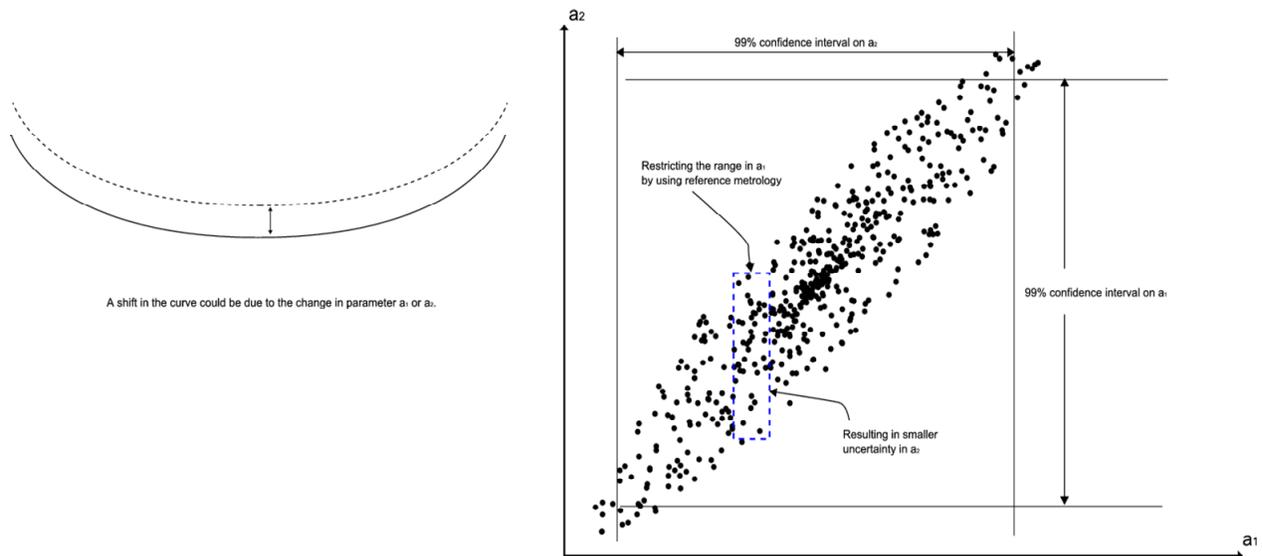


Fig. 5. An illustration of the effects of movement in an optical intensity profile (left) from an unknown change in the optical parameters and the resulting confidence limits (right).

4. NON-LINEAR REGRESSION

4.1 Derivation

In this section, we first develop the non-linear regression model and then expand this model using a Bayesian approach to include the *a priori* information. The goal is to develop a rigorous method to embed reference metrology or other *a priori* information gained by knowledge of manufacturing variability or alternative physical constraints.

We assume that in a regression setting, Y is the dependent variable, *e.g.*, intensity, and X is an independent variable, *e.g.*, λ or θ . In general a complete set of measurements consists of N data points obtained under a varying set of conditions for both simulations and experimental measurements. The i^{th} data point represents the condition under which $i = 1, \dots, N$ measurements were made. The K model parameters are expressed as a vector $\vec{a} = \{a_1, \dots, a_K\}$, and represent the model input parameters, for example, CD, sidewall, height, *etc.* We have N measured values of Y denoted as $\{y_1, \dots, y_N\}$. For the i^{th} data point the physically measured value corresponding to x_i is y_i . We also have the N simulated values with the i^{th} data point x_i corresponding to the simulated values $y(x_i; \vec{a})$ for $i = 1, \dots, N$. We want to compare the measured $\{y_1, \dots, y_N\}$ with simulated $\{y(x_i; \vec{a})\}, i = 1, \dots, N\}$ and find the optimal estimator of the parameter vector $\vec{a} = \{a_1, \dots, a_K\}$.

In general, $y(x_i; \vec{a})$ is a nonlinear function of \vec{a} , at a specific point (*e.g.*, an initial value or an optimal value) of $\vec{a}(0) = \{a_1(0), \dots, a_K(0)\}$. Using a first-order Taylor expansion, a linear approximation of the nonlinear regression is given by

$$y_i = y(x_i; \vec{a}(0)) + \sum_{k=1}^K \left[\frac{\partial y(x_i; \vec{a})}{\partial a_k} \right]_{\vec{a}=\vec{a}(0)} (a_k - a_k(0)) + \varepsilon_i(0), \quad (9)$$

where $\varepsilon_i(0)$ is an approximation error that we want to minimize [12]. By re-parameterization, the model can be expressed as

$$y_i(0) = \sum_{k=1}^K D_{ik}(0) \beta_k(0) + \varepsilon_i(0), \quad (10)$$

with

$$\beta_k(0) = a_k - a_k(0), \quad (11)$$

$$D_{ik}(0) = \left[\frac{\partial y(x_i; \vec{a})}{\partial a_k} \right]_{\vec{a}=\vec{a}(0)}, \quad (12)$$

and

$$y_i(0) = y_i - y(x_i; \vec{a}(0)). \quad (13)$$

The variance–covariance matrix of the experimental values $\{y_1, \dots, y_N\}$ is denoted by $V = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2]$. The generalized least square estimator of $b(0)$ is then given by

$$\hat{\beta}(0) = (D(0)^T V^{-1} D(0))^{-1} D(0)^T V^{-1} Y(0), \quad (14)$$

with $\hat{\beta}(0) = \{\hat{\beta}_1(0), \dots, \hat{\beta}_K(0)\}$ and

$$D(0) = \begin{bmatrix} D_{11}(0) \dots D_{1K}(0) \\ \dots \\ D_{N1}(0) \dots D_{NK}(0) \end{bmatrix}. \quad (15)$$

See Ref. 13. The generalized least square estimator of the parameters and variance – covariance matrix is now given by

$$\hat{a}_k = \hat{\beta}_k(0) + a_k(0) \quad k = 1, \dots, K, \quad (16)$$

and

$$V \text{ ar}[\hat{a}] = (D(0)^T \cdot V \cdot D(0))^{-1}. \quad (17)$$

When additional information from one or more parameters is available (e.g. the parameters and their uncertainties are obtained by AFM metrology [or other sources]), we can treat these as prior information and embed these in the model to obtain new parameter estimates and their corresponding uncertainties using the Bayesian statistical approach. In the application of this approach few parameters are treated as fixed and instead either a parameter is allowed to float freely or *a priori* information is used to estimate a mean value and its probability distribution, which has the effect of a weighting to limit the floating range. In this treatment, these parameters become unknowns that have their own probability distributions.

As an example, when the first parameter among the K parameters has a known distribution with a mean of a_1^* and variance of σ , we obtain new parameter vector estimates of \hat{a}^* and a new variance matrix of $V \text{ ar}[\hat{a}^*]$ by embedding these into the model.

In this scenario, if additional information about one parameter is available, we can treat this as prior information to get better estimates of the parameters and their uncertainties. To use this prior information, a Bayesian approach is applied. Referring to the regression model above in Eqn. (10), from [14, 15], we can treat the prior information of β_1 as another “data point” in the regression. Thus, corresponding to (10), we have an expanded model given by

$$Y^*(0) = D^*(0) \cdot \beta(0) + \varepsilon^*(0), \quad (18)$$

$$D^*(0) = \begin{pmatrix} D(0) \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} D_{11}(0) \dots D_{1K}(0) \\ \dots \\ D_{N1}(0) \dots D_{NK}(0) \\ 1, 0, \dots, \dots, 0 \end{pmatrix}, \quad (19)$$

$$\hat{\beta}^\#(0) = (D^*(0)^T V^{*-1} D^*(0))^{-1} D^*(0)^T V^{*-1} Y^*(0), \quad (20)$$

and for the parameter estimates

$$\hat{a}_k^\# = \hat{\beta}_k^\#(0) + a_k(0) \quad \text{for } k=1, \dots, K. \quad (21)$$

The posterior variance matrix of the parameter estimators is given by

$$\text{Var}[\hat{a}^\#] = \text{Var}[\hat{\beta}^\#(0)] = \left(D^*(0)^T \cdot V^{*-1} \cdot D^*(0) \right)^{-1}. \quad (22)$$

The two-parameter example from Eqn. (6) becomes

$$Q^{-1} = \begin{pmatrix} \frac{D}{AD - B^2 + D/\sigma_{\beta_1}^2}, & \frac{-B}{AD - B^2 + D/\sigma_{\beta_1}^2} \\ \frac{-B}{AD - B^2 + D/\sigma_{\beta_1}^2}, & \frac{A + 1/\sigma_{\beta_1}^2}{AD - B^2 + D/\sigma_{\beta_1}^2} \end{pmatrix}, \quad (23)$$

and

$$\text{Var}[\hat{\beta}_1^\#(0)] = \frac{D}{AD - B^2 + D/\sigma_{\beta_1}^2}. \quad (24)$$

This last equation shows the effect on the sensitivity matrix and uncertainty matrix for the previous example when including the *a priori* information. It can be shown that the uncertainties of all parameters can be improved such that

$$\text{Var}[\hat{a}^*] \leq \text{Var}[\hat{a}] \quad k = 1, \dots, K.$$

4.2 Applications of the Bayesian Approach

We next apply embedded uncertainties to develop reference metrology applications and to improve measurement uncertainties for measurement hardware compatible with high volume manufacturing. First, we apply the statistical techniques to the contact hole example from Figure 3 and Table 1. We recalculate the uncertainties with assumed *a priori* knowledge of the top CD and sidewall angles. The first section of Table 2 shows a series of uncertainty calculations for various wavelength and angle scans with all three parameters floating, as labeled in the table. The second and third sections show the uncertainties with an *a priori* assumption of 2 nm uncertainty for the top CD and a 1° known sidewall uncertainty distribution.

Table 2. Improving uncertainties for the 22 nm node contact hole example. Comparing the uncertainties from the lower two cases, the uncertainties have been reduced for all values by embedding the *a priori* reference data.

Uncertainties without <i>a priori</i> information							$\sigma_{\text{noise}} = 1\%$
	λ scan TE $\theta = 25^\circ$	λ scan TM $\theta = 25^\circ$	λ scan TE $\theta = 45^\circ$	λ scan TM $\theta = 45^\circ$	θ scan TE $\lambda = 310$ nm	θ scan TM $\lambda = 310$ nm	
σ_{CD}	2.98	2.92	3.83	2.66	17.10	1.83	
σ_{SWA}	0.70	0.70	0.90	0.68	8.21	0.61	
σ_{Overetch}	1.47	1.36	1.86	1.23	6.70	0.67	

Uncertainties with 2 nm Top CD reference						
	λ scan TE $\theta = 25^\circ$	λ scan TM $\theta = 25^\circ$	λ scan TE $\theta = 45^\circ$	λ scan TM $\theta = 45^\circ$	θ scan TE $\lambda = 310$ nm	θ scan TM $\lambda = 310$ nm
σ_{CD}	1.66	1.65	1.77	1.60	1.99	1.35
σ_{SWA}	0.53	0.54	0.52	0.44	2.08	0.49
σ_{Overetch}	1.41	1.30	1.64	1.20	1.38	0.67

Uncertainties with 1.0° Sidewall Angle reference						
	λ scan TE $\theta = 25^\circ$	λ scan TM $\theta = 25^\circ$	λ scan TE $\theta = 45^\circ$	λ scan TM $\theta = 45^\circ$	θ scan TE $\lambda = 310$ nm	θ scan TM $\lambda = 310$ nm
σ_{CD}	0.59	0.58	0.76	0.53	1.90	0.36
σ_{SWA}	0.14	0.14	0.18	0.13	0.85	0.12
σ_{Overetch}	0.29	0.27	0.37	0.25	0.72	0.14

The next application of the Bayesian embedded metrology approach is to the L100P300 etched silicon target, Table 3. In this experimental example, best fit values and uncertainties are shown for the regression analysis with and without embedded AFM reference metrology [16]. The data show a change in the mean values as well as an improvement in the uncertainties. The effects of the embedded reference measurement data are to pull the OCD scatterfield results toward the AFM measured values. The resulting uncertainties from the combined measurements are lower than each of the individual measurement results.

Table 3. A chart showing OCD measurements with and without embedded AFM reference metrology. The data show reduced uncertainties in the combined analysis.

	OCD Fitting	AFM	OCD w/ AFM		OCD Fitting	AFM	OCD w/ AFM
Top	120	119.2	121	σ_{Top}	1.05	0.75	0.35
Middle	112	117.3	115	σ_{Middle}	1.58	0.75	0.60
Bottom	143	132.8	141	σ_{Bottom}	0.78	0.75	0.42

5. REFERENCE MEASUREMENT ARCHITECTURE

In this section, we demonstrate the combination of four different measurement techniques into a monolithic optical measurement. Although the three reference measurements are embedded in the optical fitting process, the influence of each measurement result is weighted by its associated measurement uncertainty, thereby resulting in an integrated final measurement. AFM, SEM, CD X-ray scattering (SAXS), and OCD were all used to measure line patterns etched into a nitride film on a polysilicon substrate. A standard focus exposure matrix (FEM) was used to create etched line arrays. An example of the various reference measurement data is shown in the lower part of Figure 6. The two graphs in the top of the figure show the parametric fitting results using angle-resolved scatterfield microscopy intensity profiles. For these optical data the top, middle and bottom CDs were allowed to float as well as the nitride layer thickness and the optical properties for both the nitride and the substrate. Allowing all of these parameters to float was essential to obtaining good agreement. The data show typical results for the various die across the FEM field. Although most measurements across the wafer agree within a couple of nanometers on the mean CD, the methods yielded discrepant sidewall angle values. This comparison was carried out in four different laboratory locations, and the results were not normalized or offset with respect to one another to obtain better agreement. No special care was taken to match the results, but rather they were used as independent evaluations of the measurand.

The residuals and goodness of fit are good for this optical parametric fitting example. The uncertainties, however, in the optical fits are not adequate due to a combination of low sensitivity and correlation between parameters. This regression fit will benefit from embedded reference metrology.

Another attribute or complication of the fitting process is observed when two local minima are near one another. In this situation, the residuals from two sets of parameters are close in value and the least square fitting routine or regression fit may return an incorrect value. This is essentially a uniqueness problem that arises randomly, depending on how well behaved the n dimensional parameter space is and where the best fit resides in this space. The embedded uncertainty distributions represent a solution to this problem as they help the regression algorithm “snap” to the correct minimum. Figure 7 is an example showing two minima near to one another. The best fit has a slight undercut and is, in fact, the incorrect solution based on reference metrology. When AFM data are embedded, the regression algorithm selects the

second minimum as the best fit. Although the residuals and apparent goodness of fit values are slightly worse, the correct values are selected.

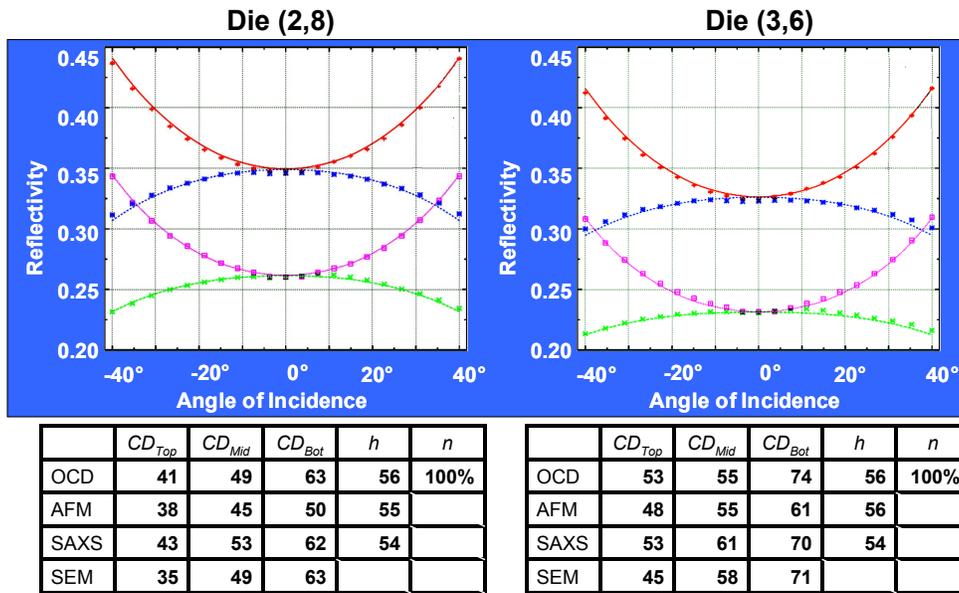


Fig. 6. The top graphs show four scatterfield OCD scans representing two polarizations and two scan axes. The agreement between simulation and experiment is good for both die based on an absolute measurement using the standard normalization procedure. The lower tables show the various reference measurements for each die.

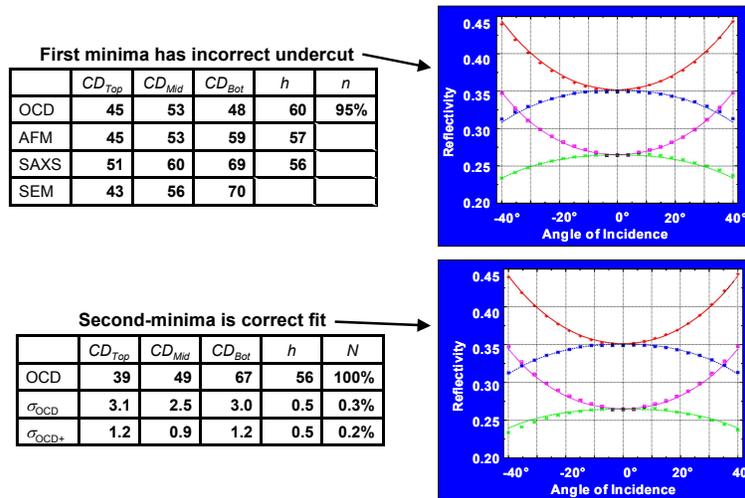


Fig. 7. This experimental example demonstrates the effect of local minima with similar residual values. Often the parameter set will “snap” between two sets of values. The graphs on the right show the subtleties in curves for nearby minima.

This example of local minima demonstrates both a reduction in uncertainty and the selection of the correct measurement value. We have observed this multiple minima effect multiple times throughout several more exhaustive OCD studies. The lower table in Figure 7 shows the uncertainty improvements resulting from the embedded reference measurement. All five floating parameters in this case are “snapping” between two local minima. In our standard algorithm output, we report the first ten minima and this type of effect is easily uncovered using this analysis algorithm.

The final example combines all four reference measurement techniques as shown in Table 4. These data show only scatterfield OCD parametric fitting results, OCD combined with bottom width CD SAXS, and a combination of all four measurement methods. The CD SAXS and CD SEM measurements were not reported with complete expanded uncertainties. As a result, 2 nm 1σ expanded uncertainties were assumed solely for demonstration purposes. These data again show a substantial reduction in the measurement uncertainties.

Table 4. This example demonstrates the combination of multiple measurement methods. First, mean values and uncertainties for OCD are reported, followed by embedded CD SAXS bottom width data. The combination of all four measurement results with the assumptions described in the text is shown in the last column

$CD_{Top} = 53 \pm 4.5 \text{ nm}$	$CD_{Top} = 53 \pm 4.4 \text{ nm}$	$CD_{Top} = 53 \pm 2.6 \text{ nm}$
$CD_{Mid} = 55 \pm 2.9 \text{ nm}$	$CD_{Mid} = 55 \pm 2.3 \text{ nm}$	$CD_{Mid} = 55 \pm 1.1 \text{ nm}$
$CD_{Bot} = 74 \pm 3.5 \text{ nm}$	$CD_{Bot} = 74 \pm 1.7 \text{ nm}$	$CD_{Bot} = 74 \pm 1.2 \text{ nm}$
$h = 56 \pm 0.6 \text{ nm}$	$h = 56 \pm 0.5 \text{ nm}$	$h = 56 \pm 0.4 \text{ nm}$
$n = (n_{stated}) \pm 0.3 \%$	$n = (n_{stated}) \pm 0.2 \%$	$n = (n_{stated}) \pm 0.1 \%$
OCD only parametric fit	$CD_{Bot, SAXS}$ embedded	AFM, CD SAXS, and CD SEM embedded

One important aspect to highlight in this work is that the method of reporting uncertainties has an immediate impact on the weight given to any embedded component. It is important to draw a distinction between repeatability measurement distributions and expanded uncertainty measurement distributions [17]. Although discussed in depth in Ref. 18, the major difference is the incorporation of simply measurement repeatability versus systematic components, such as errors in the geometry modeled or uncertainties in the measurement wavelength and angle. The optical modeling and library-fitting technique, depending on the parameters floated, necessarily includes some of the systematic uncertainty components such as sidewall variation, optical constants, and layer thickness to the extent that these parameters were floated.

6. CONCLUSIONS

Scatterfield microscopy was used to demonstrate quantitative measurements of dense line arrays with dimensions that result in only specular reflected light. Although the technique was applied to larger scatterometry arrays that fill the optical field of view, this technique is capable of scatterometry-type measurements on very small targets, enabling chip applications with reduced target size. This also allows for parallel measurements of multiple targets having potential applications in both CD and overlay metrology. Using accurate background normalization and optical tool compensation, quantitative modeling was demonstrated. Nanometer-scale measurements can be achieved using angle-resolved microscopy. Next generation tool development includes a 193 nm scatterfield metrology tool as well as a new spectroscopic scatterfield instrument. Both tools are now operational and are expected to yield lower measurement uncertainties.

A new approach to embedded metrology was presented and demonstrated to improve reference metrology uncertainties. It has applications in both reference metrology and process control metrology. This Bayesian approach enables the rigorous combination of diverse metrology solutions, and it can be applied to multiple measurement instruments to arrive at a monolithic measurement result that combines the individual components based on or weighted by their individual uncertainties. This provides an immediate method to improve reference metrology uncertainties and enables a new architecture or systems approach to reference metrology.

This new method has important implications in devising measurement strategies that take advantage of the best measurement attributes of each individual technique. This includes both sensitivities to geometrical aspects or materials attributes as well as consideration of measurement throughput. Experimental examples were used to demonstrate the embedded uncertainty approach including an example that combined optical measurements of n and k with thickness

measurements from independent tools. These data demonstrate substantial effects on the resulting OCD measurements and uncertainties. This was given as an example implementation to describe applications in a production environment.

7. ACKNOWLEDGEMENTS

The authors would like to thank Pete Lipscomb, Jimmy Price, and Michael Bishop of ISMI for wafer fabrication and measurement support. The authors acknowledge Wen Li Wu of NIST for CD SAXS measurement data and Ravikiran Attota for optical property measurements. The authors are grateful for financial support from the ISMI, Office of Microelectronics Programs at NIST and the Scatterfield Optical Competence also from NIST.

REFERENCES

1. R. M. Silver, B. Barnes, R. Attota, J. Jun, M. Stocker, E. Marx, and H. Patrick, "Scatterfield Microscopy to Extend the Limits of Image-based Optical Metrology," *Applied Optics*, Vol. 46, 20, pp. 4248-4257 (2007).
2. H Patrick, R. Attota, B. Barnes, T. Germer, R.G. Dixon, M. Stocker, R. M. Silver, and M. Bishop, "Optical Critical Dimension Measurement of Silicon Grating Targets using Back Focal Plane Scatterfield Microscopy," *J Micor/Nanolith., MEMS and MOEMS* 7(1), 0137011, (2007).
3. C. Raymond M. Littau, A. Chuprin, and S. Ward, "Comparison of solutions to the scatterometry inverse problem," *Proc. SPIE* Vol. 5375, p. 564, (2004).
4. P. Leray, and S. Cheng, "Optimization of scatterometry parameters for the gate level of the 90 nm node," *Proc. SPIE* Vol. 5752, p. 1402, (2004).
5. R. M. Silver, B. Barnes, A. Heckert, R. Attota, R. Dixon, and J. Jun, "Angle Resolved Optical Metrology", *Proc. SPIE* 6922, 69221M (2008).
6. R. M. Silver, T. Germer, R. Attota, B. Barnes, B. Bunday, J. Allgair, E. Marx, and J. Jun, "Fundamental Limits Optical Critical Dimension Metrology: A Simulation Study," *SPIE Proc. Feb. Vol.* 6518 (2007).
7. R. M. Silver, T. Germer, R. Attota, E. Marx, M. Davidson, B. Barnes, J. Jun, and R. Larrabee, "Fundamental Optical Critical Dimension (OCD) Limits: A Simulation-based Study," SEMATECH Tech. Transfer #06044749A-TR, May 2006.
8. B.M. Barnes, L.P. Howard, J. Jun, P.Lipscomb, and R.M. Silver, "Zero-order imaging of device-sized overlay targets using scatterfield microscopy," *Proc. SPIE* **6518**, 65180F (2007).
9. T.A. Germer, "Modeling the effect of line and trench profile variation on scatterometry measurements," *Proc. SPIE* 6815, (2007).
10. M.G. Moharam, D.A. Pommet, E.B. Grann, and T.K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," *J. Opt. Soc. Am. A* 12, 1077-1086 (1995).
11. W.H. Press, S.A. Teukolsky, W.T. *Recipes in C, The Art of Scientific Computing*, 2nd Edition (Cambridge University Press, 1992), Section 15.6.
12. Neter, J., Wasserman, W., and Kutner, M. (1983), *Applied Linear regression Models*, Richard D. Irwin, Inc, Homewood, Illinois.
13. Ravishanker, N, and Dey, D. K. (2002), *A First Course in Linear Model Theory*, Chapman & Hall/CRC.
14. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., (2000) *Bayesian Data Analysis*, Chapman & Hall/CRC.
15. Golub, G. H. and Van Loan, C. F. (1989) *Matrix of Computations*, 2nd edition, The John Hopkins Press, Baltimore.
16. R. Dixon, J. Fu, N. Orji, R. Allen, and M. Cresswell, "CD-AFM Reference metrology at NIST and SEMATECH," *Proc. SPIE* Vol. 5752-32, (2005).
17. U. S. Guide to the Expression of Uncertainty in Measurement, ANSI/NCSS Z540-2-1997.
18. T. Germer, H. Patrick, R. M. Silver, and B. Bunday, "Developing an Uncertainty Analysis for Optical Scatterometry," *Proc. SPIE San Jose CA*, (2009).

**Some equipment or software is identified in this publication solely for the purpose of fully explaining the technical content. NIST does not endorse the use of any of those materials or instruments identified.

SEMATECH, Inc. SEMATECH, and the SEMATECH logo are registered servicemarks of SEMATECH, Inc. International SEMATECH Manufacturing Initiative, ISMI, Advanced Materials Research Center and AMRC are servicemarks of SEMATECH, Inc. All other servicemarks and trademarks are the property of their respective owners.