

A Bayesian Statistical Model for Hybrid Metrology to Improve Measurement Accuracy

R. M. Silver, N.F. Zhang, B.M. Barnes, J. Qin, H. Zhou, and R. Dixon
National Institute of Standards and Technology, Gaithersburg, MD 20899

ABSTRACT

We present a method to combine measurements from different techniques that reduces uncertainties and can improve measurement throughput. The approach directly integrates the measurement analysis of multiple techniques that can include different configurations or platforms. This approach has immediate application when performing model-based optical critical dimension (OCD) measurements. When modeling optical measurements, a library of curves is assembled through the simulation of a multi-dimensional parameter space. Parametric correlation and measurement noise lead to measurement uncertainty in the fitting process with fundamental limitations resulting from the parametric correlations. A strategy to decouple parametric correlation and reduce measurement uncertainties is described. We develop the rigorous underlying Bayesian statistical model and apply this methodology to OCD metrology. We then introduce an approach to damp the regression process to achieve more stable and rapid regression fitting. These methods that use *a priori* information are shown to reduce measurement uncertainty and improve throughput while also providing an improved foundation for comprehensive reference metrology.

1. INTRODUCTION

There has been significant interest in new methods to combine measurement techniques that reduce uncertainties and improve measurement throughput. Hybrid metrology can enable both improved throughput and accuracy or alternatively, improved measurement uncertainty in reference metrology [1]. In general, one can combine measurements from different systems on a weighted or equal basis, treating them as independent measurements, or one can integrate the measurements directly in the data fitting routines. In the latter implementation we can link the data analysis using a unified parametric fit across the methods or we can use a Bayesian statistical model that uses *a priori* information from the different measurements in a model-based regression. These methods combine measurements from different metrology tool configurations or platforms, potentially improving throughput and measurement accuracy.

Although methods such as scanning electron microscopy (SEM) are beginning to use model-based regression in improved edge detection, this hybrid approach has immediate utility when performing model-based optical critical dimension measurements [2, 3]. When modeling optical measurements, a library of curves is assembled through the simulation of a multi-dimensional parameter space. A least squares fitting routine is then used to choose the optimum set of parameters that yield the closest experiment-to-theory agreement. When using a model-based approach, it is essential that the model adequately describes the physical measurement conditions and that an acceptable goodness-of-fit is achieved with the best set of parameters.

In the data fitting process parametric correlation (uncertainty arising from the interaction between modeling parameters), measurement noise, and model inaccuracy all lead to error and measurement uncertainty. The cross-correlations among parameters can lead to very large uncertainties even when a measurement technique demonstrates good sensitivity to a single parameter [4]. The hybrid metrology approach can directly improve measurement uncertainty introduced through parametric correlation and has further utility in selecting the correct minima among multiple nearby local minima in the fitting space.

We develop a regression fitting routine and then apply Bayesian statistical methods to decouple parametric correlation and reduce measurement uncertainties. We then apply the rigorous statistical model to scatterometry and scatterfield measurement applications. Unlike a conventional chi-square minimization, the statistical regression allows us to both interpolate the calculated library grid and guide the library calculations through iteration. A damping term is also introduced to the regression procedure to achieve more stable regression fitting and a more robust iteration procedure. As a benefit to this approach we can also reduce the calculation volume of the parametric simulation space. To

demonstrate embedding information we use *a priori* information from AFM measurements in the optical fitting process. Both simulation results and experimental data demonstrating this methodology are presented.

Although in some circumstances a parallel regression of data from multiple platforms using a single uniform set of characterization parameters may provide the best fitting uncertainties, using embedded metrology allows us to take better advantage of best attributes of each technique. As an example we can use an OCD metrology tool in combination with an optical film thickness measurement and take into account the linewidth sensitivity of an OCD tool with the throughput of a film thickness tool to achieve an improved measurement uncertainty and throughput.

2. THE SCATTERFIELD HARDWARE PLATFORM AND SAMPLE PLAN

The scatterfield microscopy technique has been described in detail elsewhere [5, 6]. 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. Although for the data presented here we used the scatterfield platform in an angle-resolved mode, we have also used the platform in a spectroscopic mode which has been shown in scatterometry to result in lower parametric correlation [7].

The angle-resolved mode is similar to conventional angular scatterometry except that the high magnification optical train allows small targets or several targets to be measured simultaneously. We can perform massively parallel scatterometric 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.

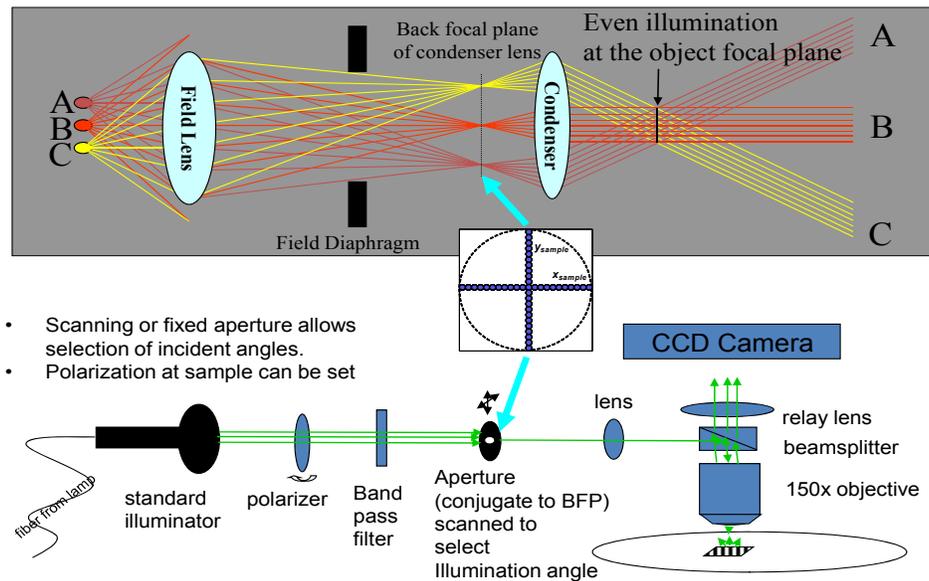


Figure 1. Schematic of the experimental apparatus

Two different types of samples are analyzed in this paper. We use an array of lines shown in Figure 2 and an array of pillars as described in Figure 3. The line array has nominal printed middle width values of 35 nm based on atomic force microscopy (AFM) measurements. The parameters used to characterize the line geometry are defined in the figure. The AFM is used to acquire line width profiles, pitch, and height which serve three purposes. First, the AFM values are used as the initial starting point for the simulation library. In this application the measurement uncertainty is not relevant since the AFM is merely used to define the approximate space of the simulation domain. Second, the AFM values are used as reference values to compare with the optical fitting data results based solely on optical critical dimension (OCD)

fitting routines. Third, the AFM measurement values and their uncertainties are used as *a priori* information to be embedded in the OCD fitting routines.

Figures 2 and 3 show the geometrical parameters used to approximate the shape. Also shown in the figures are an example of the die and their reference measurement values. In general, top width, middle width, bottom width, side wall angle, angle, pitch and height are all reported with measurement uncertainties.

A set of experimental data including parametric fitting results is shown in Figure 2 for each of the examples. Once the data are acquired and normalized, they can be analyzed using library-based fitting techniques. To compare the experimental signatures with electromagnetic scattering simulations, a comprehensive parametric analysis is used. When fitting the optical measurements, a library of curves is assembled through simulation of a multi-dimensional parameter space that includes the variation of n , k , height, pitch, sidewall, and CD. Simulations were completed here using a rigorous coupled waveguide analysis (RCWA) model or a three dimensional finite difference time domain model [8,9].

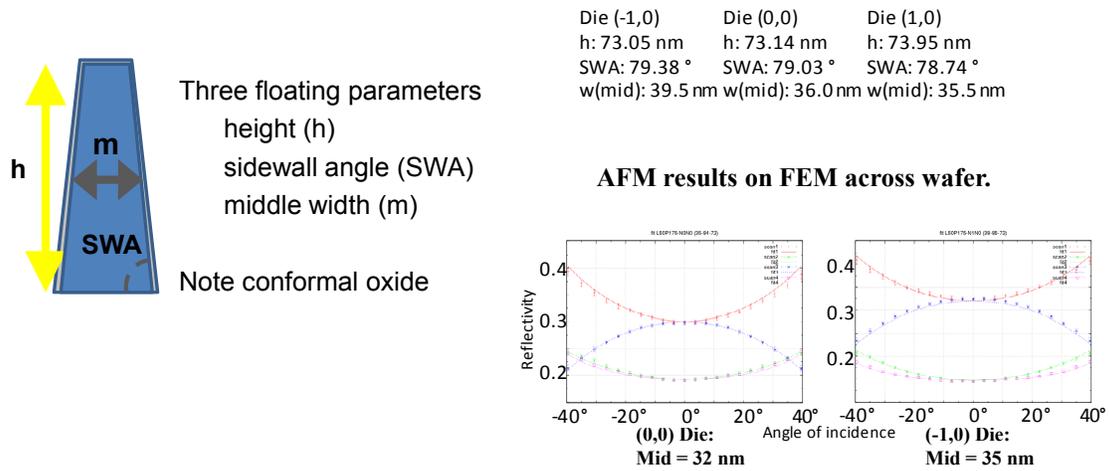


Figure 2. Geometrical parameters used to approximate the line shape. Also shown in the figure is an example of the die and their reference measurement values.

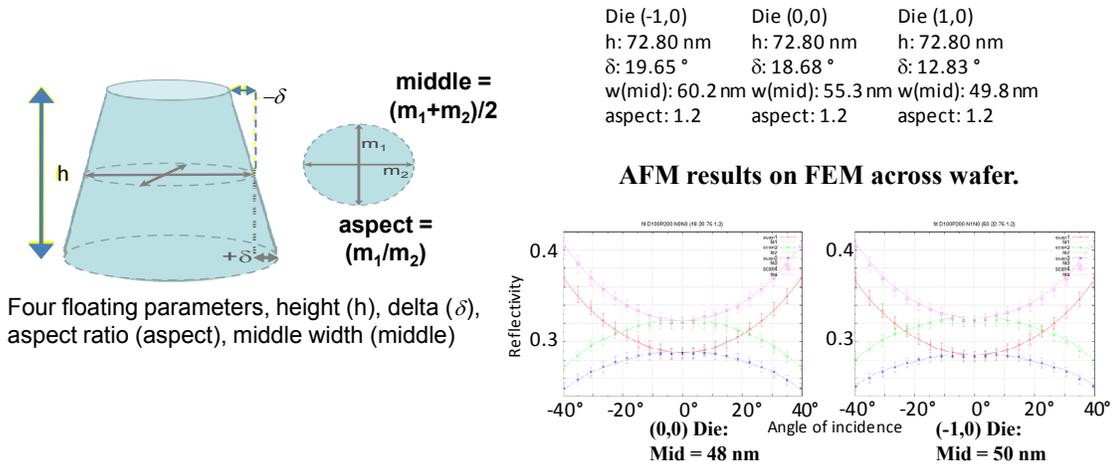


Figure 3. Geometrical parameters used to approximate the pillar shape. The figure also shows an example of two die from the FEM and their reference measurement values.

Historically, a least-square fitting routine has been used to choose the optimum set of parameters that yields the closest experiment-to-theory agreement and minimum uncertainties in the covariance matrix [4]. These data demonstrate good agreement between the simulated library of curves and the experimental data using standard chi-square analysis and the optical microscope normalization procedures described in Ref. [10]. The goodness-of-fit values and residuals are acceptable, although depending on the simulation set, the uncertainties may be unacceptably large. We next develop the regression theory and the Bayesian statistical approach to reduce uncertainties and improve measurement accuracy.

3. NONLINEAR REGRESSION DERIVATION

We now develop the nonlinear regression model and then expand the model using a Bayesian approach to include *a priori* information. In addition, a damping term is added to the regression for both embedded and non-embedded regression. The goal is to develop a rigorous method to embed reference metrology or other *a priori* information gained by knowledge of the manufacturing process. Only an overview of the derivation is given here, see Ref. 11 for a more detailed formalism.

In general a complete set of measurements consists of N data points acquired under a varying set of conditions for both simulations and experimental measurements. The K model parameters are expressed as a vector $\mathbf{a} = \{a_1, \dots, a_K\}$, and represent the model input parameters, for example, CD, sidewall, height. We have N measured values of Y denoted as $\{y_1, \dots, y_N\}$ and N simulated values $y(x_i; \mathbf{a})$ corresponding to the i^{th} data point x_i . We want to compare the measured $\{y_1, \dots, y_N\}$ with simulated $\{y(x_i; \mathbf{a}), i = 1, \dots, N\}$ and find an optimal estimator of the parameter vector $\mathbf{a} = \{a_1, \dots, a_K\}$. In general, $y(x_i; \mathbf{a})$ is a nonlinear function of \mathbf{a} . Treating $y(x_i; \mathbf{a})$ as a mean response of y_i , we have a nonlinear regression for y_i and $y(x_i, \mathbf{a})$ for $i = 1, \dots, N$ given by

$$y_i = y(x_i, \mathbf{a}) + \varepsilon_i \quad \text{for } i = 1, \dots, N, \quad (1)$$

where ε_i is the corresponding random error with zero mean. Using a first-order Taylor expansion, a linear approximation of the nonlinear regression is given by

$$y_i = y(x_i; \mathbf{a}(0)) + \sum_{k=1}^K \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)} (a_k - a_k(0)) + \varepsilon_i, \quad (2)$$

where $\mathbf{a}(0) = \{a_1(0), \dots, a_K(0)\}$ is an initial value or an optimal value of \mathbf{a} and ε_i is the corresponding random error with zero mean [11]. By re-parameterization, the model can be expressed as

$$y_i(0) = \sum_{k=1}^K D_{ik}(0) \beta_k(0) + \varepsilon_i, \quad (3)$$

with

$$D_{ik}(0) = \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \right]_{\mathbf{a}=\mathbf{a}(0)}, \quad (4)$$

and $\beta_k(0) = a_k - a_k(0)$ and $y_i(0) = y_i - y(x_i; \mathbf{a}(0))$, see Ref. 12. The covariance matrix of the experimental values $\{y_1, \dots, y_N\}$ is denoted by $\mathbf{V} = \text{diag}[\sigma_1^2, \dots, \sigma_N^2]$.

We can now write the re-parameterized model from Eqn. 2 in a matrix form.

$$\mathbf{Y}(\mathbf{0}) = \mathbf{D}(\mathbf{0})\boldsymbol{\beta}(\mathbf{0}) + \boldsymbol{\varepsilon} \quad (5)$$

with

$$D(\mathbf{0}) = \begin{bmatrix} D_{11}(\mathbf{0}) \dots D_{1K}(\mathbf{0}) \\ \dots \\ D_{N1}(\mathbf{0}) \dots D_{NK}(\mathbf{0}) \end{bmatrix}. \quad (6)$$

It can be shown that the generalized least squares estimator of $\boldsymbol{\beta}(\mathbf{0})$ is now given by

$$\hat{\boldsymbol{\beta}}(\mathbf{0}) = \left(\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0}) \right)^{-1} \mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{Y}(\mathbf{0}), \quad (7)$$

where $\hat{\boldsymbol{\beta}}(\mathbf{0}) = (\hat{\beta}_1(\mathbf{0}), \dots, \hat{\beta}_K(\mathbf{0}))$ are the best linear unbiased estimators of $\boldsymbol{\beta}(\mathbf{0})$. Namely, it is the linear unbiased estimator with the smallest variance as well as the smallest mean squared error. See Ref. 13.

It should be noted that the linear unbiased β estimators are based on the assumption of a linear space in the region where the $D(\mathbf{0})$ derivatives are calculated. Although the parameter space is smoothly varying, it is in general nonlinear due to the complexity of the electromagnetic scattering equations and parametric correlation. The space also potentially has multiple subtle minima depending on the geometry, materials stack, and correlation. To obtain a well behaved regression and limit overshoot or jumping into a nearby minimum, a damping factor is applied to the β values.

The generalized least square estimator of the original parameters and their covariance matrix are now given by $\hat{a}_k = \hat{\beta}_k(\mathbf{0}) + a_k(\mathbf{0}) \quad k = 1, \dots, K$, and $\text{Var}[\hat{\mathbf{a}}] = \left(D(\mathbf{0})^T \cdot V \cdot D(\mathbf{0}) \right)^{-1}$. Note that when $\{\varepsilon_i\}$ are Gaussian distributed, from (7) $\hat{\boldsymbol{\beta}}(\mathbf{0})$ and $\hat{\mathbf{a}}$ are also Gaussian distributed.

Following Ref. [14], p. 40-41 we can use the Gauss-Newton method to improve $\hat{\mathbf{a}}$ iteratively and keep improving the estimates until the desired minimal change is achieved. Thus, we can assume that $\mathbf{a}(\mathbf{0}) = \{a_1(\mathbf{0}), \dots, a_K(\mathbf{0})\}^T$ in (2) is an estimate vector from the Gauss-Newton method. This allows the use of an initially sparse grid, with detailed electromagnetic simulations for each iteration about the new $\mathbf{a}(\mathbf{0})$ and new set of $D(\mathbf{0})$ derivatives. In this approach it is important to check the residuals to determine whether the estimates are really improved.

When additional information on one or more parameters is available (e.g. the parameters and their uncertainties are obtained by AFM reference 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 a Bayesian statistical approach. In this implementation a parameter that was allowed to float freely is now influenced by the *a priori* information such as the mean value and its probability distribution from an external measurement of that floated parameter.

As an example, when the first parameter among the K parameters has a known distribution with a mean of a_1^{**} and variance of $\sigma_{a_1}^2$, we can obtain new parameter estimators \mathbf{a} and a new covariance matrix by embedding the prior information into the model. Referring to the regression model above in Eqn. (5), from [Ref 15], we can treat the prior information of β_1 as another “data point” in the regression. Thus, corresponding to (4), we have an expanded matrix model given by

$$\mathbf{Y}^*(\mathbf{0}) = \mathbf{D}^*(\mathbf{0})\boldsymbol{\beta}(\mathbf{0}) + \boldsymbol{\varepsilon}^*, \quad (8)$$

$$\mathbf{D}^*(\mathbf{0}) = \begin{pmatrix} \mathbf{D}(\mathbf{0}) \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} D_{11}(\mathbf{0}) \dots D_{1K}(\mathbf{0}) \\ \dots \\ D_{N1}(\mathbf{0}) \dots D_{NK}(\mathbf{0}) \\ 1, 0 \dots \dots \dots 0 \end{pmatrix}, \quad (9)$$

$$\hat{\boldsymbol{\beta}}^\#(\mathbf{0}) = \left(\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \right)^{-1} \mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{Y}^*(\mathbf{0}), \quad (10)$$

and for the parameter estimates $\hat{a}_k^\# = \hat{\beta}^\#(\mathbf{0}) + a_k(\mathbf{0})$ for $k=1, \dots, K$.

As an example we can evaluate $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ with $K=2$ floating model parameters and $p=1$ *a priori* measurements with a known mean value and its uncertainty. For this example $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is given by

$$\hat{\boldsymbol{\beta}}^\#(\mathbf{0}) = \frac{1}{|Q|} \left[|Q_1| \hat{\boldsymbol{\beta}}(\mathbf{0}) + |Q_1| \frac{\beta_1^*(\mathbf{0})}{\sigma_{a_1}^2} \begin{pmatrix} \text{Var}[\hat{\beta}_1(\mathbf{0})] \\ \text{Cov}[\hat{\beta}_1(\mathbf{0}), \hat{\beta}_2(\mathbf{0})] \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{g_2}{\sigma_{a_1}^2} \end{pmatrix} \right], \quad (11)$$

where $|Q|$ and $|Q_1|$ are the determinants of $\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0})$ in (10) and $\mathbf{D}(\mathbf{0})^T \mathbf{V}^{-1} \mathbf{D}(\mathbf{0})$ in (7).

In Eqn. 11 it can be seen that $\hat{\boldsymbol{\beta}}^\#(\mathbf{0})$ is a linear combination of terms including components from both the optical measurements and the AFM reference measurements. The posterior covariance matrix of the parameter estimators is given by

$$\text{Cov}[\hat{\mathbf{a}}^\#] = \text{Cov}[\hat{\boldsymbol{\beta}}^\#(\mathbf{0})] = \left(\mathbf{D}^*(\mathbf{0})^T \mathbf{V}^{*-1} \mathbf{D}^*(\mathbf{0}) \right)^{-1}. \quad (12)$$

It can also be shown that the parameter uncertainties are improved by including the additional information such that $\text{Var}[\hat{\beta}_k^\#(\mathbf{0})] \leq \text{Var}[\hat{\beta}_k(\mathbf{0})]$, $k = 1, \dots, K$ see Ref [11].

4. APPLICATIONS OF THE BAYESIAN APPROACH

We next demonstrate the Bayesian method using embedded *a priori* measurement values and their uncertainties from a CD-AFM reference metrology measurement to improve multiple aspects of the final measurement uncertainties. In the first example we use the line array described in Section 2 above. Figure 4 shows optical fitting results for three die from a focus exposure matrix (FEM) having design values of 45 nm linewidth and 157 nm pitch. The figure reports the middle width values, the side wall angles (SWA) and the line height values for each die based only on OCD fitting using the regression algorithm described above. There are no reference values used in these fitting data, although AFM reference values were used to define the starting point for the simulations [16]. A large enough parameter space is then simulated to cover all three die with ample parameter coverage to ensure the best estimate values are well within the space.

Each panel in the figure shows four scans, two polarization scans are shown for each scan axis. The tables show the floated parameter, the mean value reported, and the uncertainty for that parameter. The symbols are the experimental values and their error bars are shown for each data point as well. These data show good fits with small residuals and reduced chi-square values ranging from 1 to 3. The uncertainties reported in the figure are 1σ and are combined uncertainties. These combined uncertainties include all aspects of measurement uncertainty including repeatability and other systematic measurement errors such as wavelength and angle uncertainties [17].

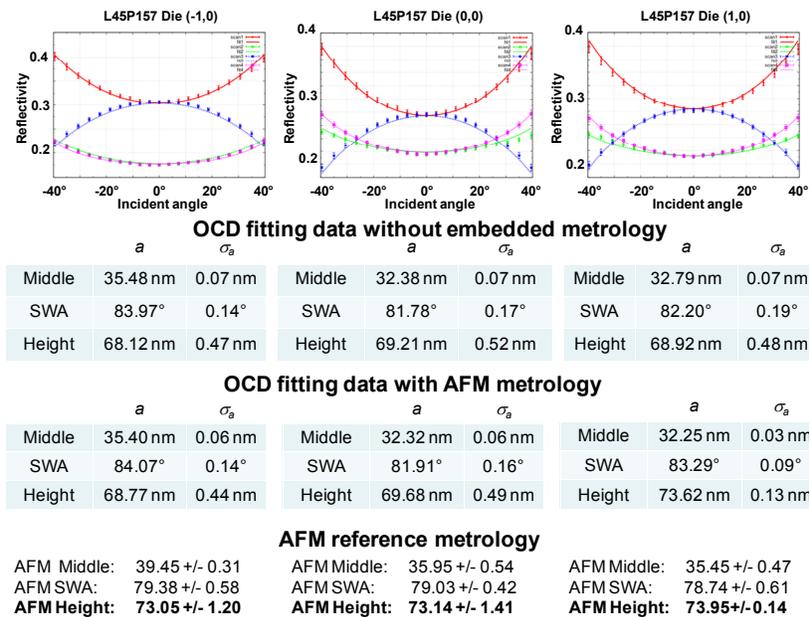


Figure 4. Parametric fitting of L45P157 linear arrays with and without embedded AFM.

Also shown in the figure are the results obtained by including AFM reference height data in the regression fits. The middle row of tables show the new OCD fit values with the AFM data embedded and the uncertainties. The lower set of tables show the AFM measured values for each of the embedded values and their associated uncertainties used in the simulations. In all cases the uncertainties are unchanged or reduced and the OCD best estimate values are seen to change relative to those values obtained in the without the embedded AFM reference measurements.

One notable item is the systematic offset between the AFM values and the OCD best fit values. This was explored in more depth. The samples are etched silicon and there is an expected conformal native oxide covering the lines. To test these assumptions, simulations were performed with a conformal oxide that showed a lack of sensitivity to the oxide at 450 nm measurement wavelength. On the other hand, the AFM is expected to measure this 1.5 nm to 2.0 nm oxide. The optical modeling results are shown below in Figure 5. These data are a likely explanation for the offset between the OCD fittings results and the AFM reference data.

4.1 Using appropriate types of uncertainties.

It is important to define the type of uncertainties used in the regression modeling. If a combined uncertainty is used for the reference measurements then we need to evaluate Type A and Type B uncertainties for the optical measurements as well. Although defined in detail in Ref [17], these correspond approximately to those elements determining measurement repeatability and systematic components such as wavelength and angle measurement errors. For both experimental data and simulation data, the appropriate uncertainties must be evaluated. If an AFM measurement with an uncertainty that includes both Type A and Type B components is used then the same must be used in the OCD regression fits. However, when combining SEM measurements and OCD measurements, for example, that do not have combined uncertainties for each method, one can still combine the methods by using equivalent uncertainties. What is essential is that comparable types of uncertainties be used to achieve the correct balance or influence in the regression analysis.

The situation is more constrained however when using combined uncertainties for reference metrology. Fundamentally the OCD electromagnetic scattering models require accurate parameter inputs since the models use SI units. If an unknown systematic error in a parameter, either fixed or floating, is introduced then the regression or chi square fitting

routines will go to an incorrect place in the parameter space to find the best fit. In this case introducing systematic errors in the parameters drives the minimization algorithms to an incorrect minimum.

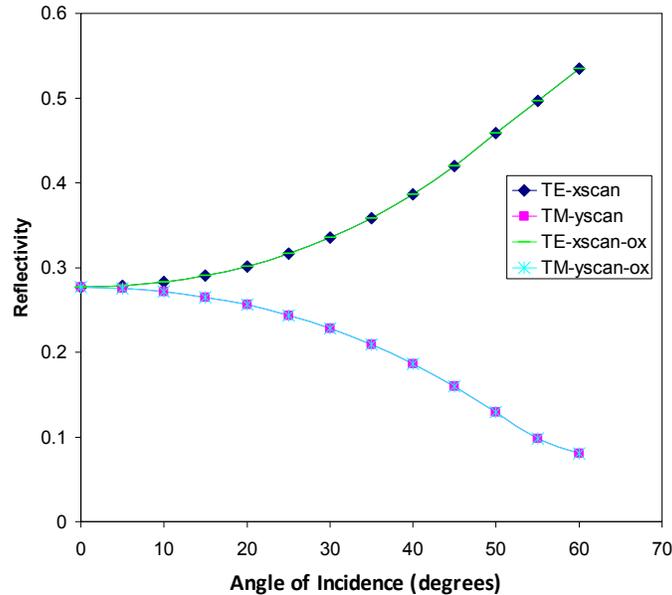


Figure 5. Simulations for an arrayed line width sample are shown with and without a conformal oxide.

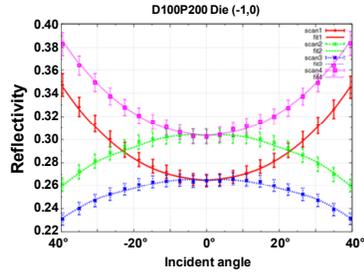
In the research presented here, repeat acquisitions are used to evaluate the experimental Type A uncertainties. Type B components are evaluated (*e.g.*, angle, spectral width, tool normalization) and also RMS added to the experimental repeatability data. We also need to evaluate systematic errors associated with the model and simulations. Examples of these errors are grid size, calculation convergence and also the set of errors due to any discrepancies between the model and the physical target (*e.g.*, rounding, footing).

4.2 Measuring a three-dimensional array of pillars.

The next example is the three dimensional pillar structure also described in Section 2. Figure 6 shows the optical data from Die (-1,0) and Die (0,0) with the scan data on the left and OCD regression and AFM reference values on the right, as labeled. The OCD-only data analyses have large uncertainties due to parametric correlation in the two dimensional structures. One noticeable attribute in the data is the significant gap in the two scan directions. This is due to asymmetry in the pillars in the x and y directions. There is nominally a 20 percent difference in the major and minor axes of the ellipse that defines the pillar cross section. Physically, this results from the lithography process.

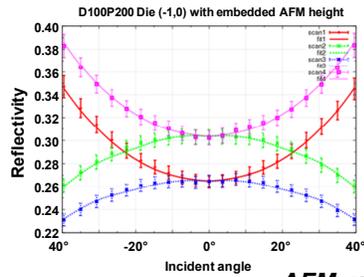
OCD regression fitting values are shown on the right in the second and fourth tables in Fig. 6 where damping is used and multiple iterations are shown. The damping coefficients are shown in the table and range from 0.32 to 1.0. For Die (0,0), the smallest residuals are seen in the first couple of iterations and the $D(0)$ matrix of derivatives send the $\mathbf{a}(0)$ estimators down a trough in parameter space to an unphysical height of 86 nm. This is the direct result of the subtle correlation effects between the various floating parameters and an effective lack of uniqueness seen in the regression minima. In this example, after six iterations, regression was stopped as the β values stayed within the grid size of the library. Though the last iterated χ^2 value for Die (0,0) is greater than that at the original $\mathbf{a}(0)$ starting point, the final iterated $\mathbf{a}(0)$ is adjacent to the point 44, 27, 86 with χ^2 of 5.09, which is lower than the starting value for that die.

To address the strong parametric correlation seen in the data as noted by the many nearby minima in the arrayed two dimensional structures we can introduce embedded AFM reference height values. With the embedded reference data the regression estimates quickly approach the best fit values requiring few iterations with little change seen in the β values with further iteration.



Die (-1,0), OCD regression

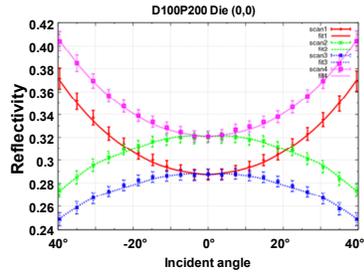
Iteration number	$a(0)$				β				dampen. ratio	χ^2
	mid (nm)	δ (nm)	h (nm)	aspect	mid (nm)	δ (nm)	h (nm)	aspect		
1	50	22	76	1.2	-0.26	1.54	0.07	<-0.01	1.00	6.60
2	50	24	76	1.2	-0.45	0.14	0.52	<0.01	1.00	14.71



Die (-1,0), OCD regression and AFM height embedded

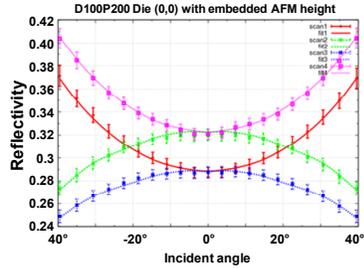
Iteration number	$a(0)$				β				dampen. ratio	χ^2
	mid (nm)	δ (nm)	h (nm)	aspect	mid (nm)	δ (nm)	h (nm)	aspect		
1	60	20	72	1.2	-9.99	5.05	1.39	-0.05	0.80	9853.30
2	52	24	74	1.125	-0.98	-3.11	-0.87	0.08	1.00	490.33
3	51	21	74	1.2	0.19	-0.61	-0.97	-0.01	1.00	7.41

AFM values: middle = 60.2 nm \pm 2.1 nm, δ = 19.6 nm \pm 3.8 nm, height = 72.8 \pm 2.0 nm



Die (0,0), OCD regression

Iteration number	$a(0)$				β				dampen. ratio	χ^2
	mid (nm)	δ (nm)	h (nm)	aspect	mid (nm)	δ (nm)	h (nm)	aspect		
1	48	20	76	1.2	-2.73	6.27	6.12	-0.01	0.32	6.71
2	47	22	78	1.2	-2.63	4.99	6.00	-0.02	0.33	5.49
3	46	24	80	1.2	-1.98	3.55	5.13	-0.01	0.39	6.24
4	45	25	82	1.2	-1.46	2.89	4.2	<-0.01	0.48	23.21
5	44	26	84	1.2	-0.17	1.54	1.66	<-0.01	1.00	67.99
6	44	28	86	1.2	-0.75	0.26	0.92	<0.01	1.00	38.66
7	43	28	86	1.2	0.42	0.02	0.44	<0.01	1.00	41.85



Die (0,0), OCD regression and AFM height embedded

Iteration number	$a(0)$				β				dampen. ratio	χ^2
	mid (nm)	δ (nm)	h (nm)	aspect	mid (nm)	δ (nm)	h (nm)	aspect		
1	48	20	76	1.2	1.01	-2.14	-2.27	<-0.01	0.88	9.36
2	49	18	74	1.2	0.13	-0.42	-0.40	-0.01	1.00	10.54

AFM values: middle = 55.3 nm \pm 2.4 nm, δ = 18.7 nm \pm 2.4 nm, height = 72.8 \pm 2.0 nm

Figure 6. Pillar arrays with and without embedded AFM metrology.

The discrepancy between the modeled and physical structures was further analyzed to better understand the offset between the AFM data and the OCD data. Since the AFM is known to have limited ability to measure footing and undercut at the bottom of line and pillar structures, simulations were run to evaluate the effects from corner rounding and footing. Although these parameters were not floated in the three dimensional modeling due to the extensive simulation times required, limited simulations were performed to evaluate the magnitude of the effects. Similar to the line array example, a native oxide is expected to create an offset that was not accounted for here.

The AFM had limited ability to identify correctly the elliptical cross section of the pillars as well as to provide an initial accurate target height due to the rounded pillar corners. Additional AFM algorithm development was necessary to adopt the height algorithms used on lines to measure the pillar heights. This is an example where the combination of techniques is invaluable for reference metrology. The two dimensional pillar structures highlighted subtleties were only

“seen” by using multiple measurement methods. In addition to reducing the combined uncertainties, we were able to flag likely causes of systematic offsets between the reference and measurement data results.

Although this hybrid metrology technique of using *a priori* information can be used to combine an individual reference measurement and single tool OCD regression fit, the method is also well suited to an iterative approach used between multiple model-based measurements. As an example, an SEM that uses edge profile fitting algorithms can be iteratively combined with an OCD fitting process. When using the Gauss-Newton method, the regression algorithms are well suited to iterate between a combination of tools that each use a regression algorithm to find the best-fit minima and parameter estimates, see Ref. [14]. This iterative implementation allows one to embed measurements from multiple tools that result in intricately linked hybrid measurements. Alternatively, a parallel regression amongst multiple methods can be carried out. This may be better suited to reference metrology where throughput is generally less of an issue. However, the parallel regression approach requires at least a partially overlapping set of floating characterization parameters.

5. CONCLUSION

An approach to hybrid embedded metrology was presented and demonstrated to improve reference metrology uncertainties. It has applications in both reference metrology and manufacturing process control metrology. This Bayesian approach enables the rigorous combination of diverse metrology solutions. A few related approaches were shown that can be applied to multiple measurement instruments to arrive at an optimum measurement result that combines the individual measurement results and their individual uncertainties. This approach is likely a requirement to improve future reference metrology uncertainties and enables a new architecture for, and approach to reference metrology.

This 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. Using the embedded metrology approach with a judicious choice in floating parameters can both reduce parametric correlation and reduce the required simulation volume of the measurement space, potentially saving significant time while improving accuracy.

A Gauss-Newton iteration procedure was introduced to guide the library calculations through iteration and generate a dense simulation space near the best estimator. A damping term was also introduced to the regression procedure to achieve more stable regression fitting and a more robust iteration procedure. An appropriate iteration implementation can effectively reduce the calculation volume of the parametric simulation space.

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 in-chip applications with reduced target size. This also enables parallel measurements of multiple targets having potential applications in both CD and overlay metrology. Using accurate background normalization and optical tool compensation, quantitative nanometer-scale measurements can be achieved.

ACKNOWLEDGEMENTS

The authors would like to thank SEMATECH for wafer fabrication and measurement support. The authors acknowledge Thomas Germer for providing the 2-D RCWA code and Francois Goasmat for assistance in computation. The authors thank John Villarrubia and James Potzick for useful discussions.

REFERENCES

- [1] R. M. Silver, N. F. Zhang, B. Barnes, H. Zhou, A. Heckert, R. Dixon, T. Germer and B. Bunday, "Improving optical measurement accuracy using multi-technique nested uncertainties," *Proc. SPIE* **7272**, 727202 (Mar, 2009).
- [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] R. 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* 65180U (2007).
- [5] 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).
- [6] 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).
- [7] B.M. Barnes, N.A. Heckert, R. Quintanilha, H. Zhou, and R.M. Silver, "Spectroscopic Scatterfield Microscopy," *AIP Conf. Proc.* **1173**, 261-265 (2009).
- [8] 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.
- [9] T. A. Germer, "Modeling the effect of line and trench profile variation on scatterometry measurements," *Proc. SPIE* 6518, 65180Z (2007).
- [10] R. M. Silver, B. Barnes, H. Zhou, N. F. Zhang, and R. Dixon, "Angle-resolved optical metrology using multi-technique nested uncertainties," *Proc. SPIE Vol. 7390* pp. 73900P-12 (2009).
- [11] N. F. Zhang, R. M. Silver, B. M. Barnes, and H. Zhou, "Use of Bayesian statistic to improve optical measurement uncertainty by combined multi-tool metrology", submitted to *Metrologia* Feb., 2011.
- [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] Bates, D. M. and Watts, D. G. (1998) *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
- [15] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., (2000) *Bayesian Data Analysis*, Chapman & Hall/CRC.
- [16] R. Dixon, J. Fu, N. Orji, R. Allen, and M. Cresswell, "CD-AFM Reference metrology at NIST and SEMATECH," *Proc. SPIE* 5752, 324 (2005)
- [17] U. S. Guide to the Expression of Uncertainty in Measurement, ANSI/NCSL Z540-2-1997.