

Journal of
**Micro/Nanolithography,
MEMS, and MOEMS**

Nanolithography.SPIEDigitalLibrary.org

Optimizing hybrid metrology: rigorous implementation of Bayesian and combined regression

Mark-Alexander Henn
Richard M. Silver
John S. Villarrubia
Nien Fan Zhang
Hui Zhou
Bryan M. Barnes
Bin Ming
András E. Vladár

Optimizing hybrid metrology: rigorous implementation of Bayesian and combined regression

Mark-Alexander Henn,^{a,*} Richard M. Silver,^a John S. Villarrubia,^a Nien Fan Zhang,^b Hui Zhou,^a Bryan M. Barnes,^a Bin Ming,^a and András E. Vladár^a

^aNational Institute of Standards and Technology, Engineering Physics Division, 100 Bureau Drive MS 8212, Gaithersburg, Maryland 20899-8212, United States

^bNational Institute of Standards and Technology, Statistical Engineering Division, 100 Bureau Drive MS 8212, Gaithersburg, Maryland 20899-8212, United States

Abstract. Hybrid metrology, e.g., the combination of several measurement techniques to determine critical dimensions, is an increasingly important approach to meet the needs of the semiconductor industry. A proper use of hybrid metrology may yield not only more reliable estimates for the quantitative characterization of three-dimensional (3-D) structures but also a more realistic estimation of the corresponding uncertainties. Recent developments at the National Institute of Standards and Technology feature the combination of optical critical dimension measurements and scanning electron microscope results. The hybrid methodology offers the potential to make measurements of essential 3-D attributes that may not be feasible otherwise. However, combining techniques gives rise to essential challenges in error analysis and comparing results from different instrument models, especially the effect of systematic and highly correlated errors in the measurement on the χ^2 function that is minimized. Both hypothetical examples and measurement data are used to illustrate solutions to these challenges. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMM.14.4.044001](https://doi.org/10.1117/1.JMM.14.4.044001)]

Keywords: hybrid metrology; electromagnetic simulation; sensitivity and uncertainty evaluation; Bayesian data analysis.

Paper 15086P received May 20, 2015; accepted for publication Oct. 8, 2015; published online Nov. 12, 2015.

1 Introduction

Hybrid metrology, e.g., the combination of distinct measurement techniques to determine critical dimensions (CDs), is an increasingly important approach to meet the needs of the semiconductor industry. A proper use of hybrid metrology may yield not only more reliable estimates for the quantitative characterization of three-dimensional (3-D) structures but also a more realistic estimation of the corresponding uncertainties. Ideally it helps to reduce the overall uncertainties by combining the individual strengths of each of the measurement techniques, making subnanometer uncertainties a realistic goal as CDs approach 10 nm.

Recent developments^{1–3} in hybrid metrology at the National Institute of Standards and Technology (NIST) feature the combination of optical critical dimension (OCD) and scanning electron microscope (SEM) measurements. The challenges and possible solutions have been outlined by some of these authors in a previous proceedings paper.³ Various methods have been presented to combine measurement results from different tool platforms, revealing two related but distinct challenges. There must be an overlapping parameter set for combined regression,³ such that each individual parametric geometry must share at least one parameter in common (e.g., height). Additionally *a priori*, each individual method must also yield parametric values that together with their uncertainties are statistically consistent, usually quantified by a Z-test; see Ref. 2 for more details.

This paper can be seen as a continuation of that work and Ref. 4, with an emphasis on the proper treatment of the measurement errors, including highly correlated and systematic errors and their influence on hybrid metrology. Tool-induced errors for OCD and scale errors for SEM are investigated, and it is shown how the parametric uncertainties can be decreased if those issues are addressed accurately in the hybridization.

Since the term hybrid metrology has gained increased significance in the dimensional metrology community outside NIST,^{5–7} we will start this work with a short overview of two of the most common techniques in Sec. 2, namely the Bayesian approach and combined regression. The measured targets and the generalized parameter sets that describe them are discussed in Sec. 3 before we give a detailed description of the performed error analysis for both OCD and SEM in Sec. 4 and their impact on the hybridization in Sec. 5. We will close with the conclusions in Sec. 6.

2 Hybrid Metrology

Hybrid metrology has gained significant recognition in recent times as an approach to considerably reduce parametric uncertainties by combining different measurements of the same measurand. We want to use this section to identify the main differences and similarities of two of the most common hybrid approaches, namely the use of *a priori* information in a Bayesian sense and combined regression. We start with the Bayesian approach and continue with combined regression. In order to keep the notation simple, throughout this section we will assume that only two measurement techniques are combined with each individual method, yielding a statistically consistent set of parameters. Note that even if

*Address all correspondence to: Mark-Alexander Henn, E-mail: mark.henn@nist.gov

some of the notations are different, the presented approaches are equivalent to those given in Ref. 2. Additional information for those who are not familiar with all of the terminology of Bayesian data analysis can be found in Refs. 8–10.

2.1 Bayesian Approach

The Bayesian approach treats information provided by each of the measurement tools quite differently. The first tool provides m values of measurement data that can be described as a vector in an m -dimensional real vector space

$$\mathbf{y} = (y_1, \dots, y_m)^T \in \mathbb{R}^m \quad (1)$$

that contains only indirect information about the quantity of interest. We therefore need to analyze the data in terms of an inverse problem.¹¹ A common approach to solve an inverse problem is to set it up as a regression problem. Initially, we need to provide a model function

$$\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \mathbf{f}(\mathbf{p}) = [f_1(\mathbf{p}), \dots, f_m(\mathbf{p})]^T \quad (2)$$

that maps the parameters of interest (e.g., the height, the width, etc.), that are identified with a vector in an n -dimensional real vector space $\mathbf{p} = (p_1, \dots, p_n)^T$, to the simulated quantities. These simulated data are generated in the same m -dimensional space defined by the measurement. The regression problem then amounts to minimizing the difference between the modeled and the measured data. We therefore solve for the parameter vector $\hat{\mathbf{p}}$ that minimizes the so-called χ^2 function, measuring the goodness of fit of the simulation to the measurement data given by the weighted norm

$$\chi^2(\mathbf{p}) = \|\mathbf{f}(\mathbf{p}) - \mathbf{y}\|^2 = [\mathbf{f}(\mathbf{p}) - \mathbf{y}]^T \mathbf{V}^{-1} [\mathbf{f}(\mathbf{p}) - \mathbf{y}]. \quad (3)$$

In this formulation, we implicitly assume that the measurement data \mathbf{y} are a noisy realization of the model, and that we have an additive error model where

$$\mathbf{y} = \mathbf{f}(\mathbf{p}) + \epsilon. \quad (4)$$

Note that the errors that are added to each of the model values can also be arranged as an m -dimensional vector. We assume the errors to be normally distributed with zero mean and $m \times m$ covariance matrix Σ_ϵ . The matrix $\mathbf{V} \in \mathbb{R}^{m \times m}$ in Eq. (3) is usually chosen to be an estimate of Σ_ϵ . We will discuss the importance of a good choice of \mathbf{V} in more detail later. If we ignore normalization constants, we see that the function in Eq. (3) is proportional to the negative log-likelihood function for the chosen error model.¹⁰ We furthermore assume that we have information about the parameter vector of interest \mathbf{p} , such as an estimate $\boldsymbol{\mu}_p$ and an uncertainty or, more specifically, a covariance matrix Σ_p . This can be based on expert knowledge or an already completed analysis of measurement data from a second measurement tool. If we assume the parameters of interest to be normally distributed, we have all the information we need to express this prior information in terms of a probability density function (PDF). This PDF is usually called the prior distribution π_{pri} . In the case of a normally distributed random vector \mathbf{p} , it is given by

$$\begin{aligned} \pi_{\text{pri}}: \mathbb{R}^n &\rightarrow \mathbb{R}, \\ \pi_{\text{pri}}(\mathbf{p}) &= \left[(2\pi)^n \det \left(\sum_p \right) \right]^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{p} - \boldsymbol{\mu}_p)^T \Sigma_p^{-1} (\mathbf{p} - \boldsymbol{\mu}_p) \right]. \end{aligned} \quad (5)$$

By subtracting the prior distribution or, more precisely, its logarithm, from the function in Eq. (3), the negative log-likelihood, we get a function that is proportional to the negative logarithm of the posterior probability distribution. If we again ignore normalization constants we get the function that serves as the modified χ^2 function

$$\tilde{\chi}^2(\mathbf{p}) = [\mathbf{f}(\mathbf{p}) - \mathbf{y}]^T \mathbf{V}^{-1} [\mathbf{f}(\mathbf{p}) - \mathbf{y}] + (\mathbf{p} - \boldsymbol{\mu}_p)^T \Sigma_p^{-1} (\mathbf{p} - \boldsymbol{\mu}_p). \quad (6)$$

Note that the second term that reflects the prior information acts as a penalty or regularization term, penalizing possible solutions to the inverse problem for measurement tool 1 that are not consistent with the prior information. The function in Eq. (6) is finally minimized to find the parameter vector estimate $\hat{\mathbf{p}}$.

2.2 Combined Regression

In combined regression we start with two distinct sets of measurement data, \mathbf{y}_A and \mathbf{y}_B , that come from two different measurement techniques. Their model functions, $\mathbf{f}_A(\mathbf{p}_A)$ and $\mathbf{f}_B(\mathbf{p}_B)$, depend on parameter vectors \mathbf{p}_A and \mathbf{p}_B , respectively. The models must have at least one common model parameter in order to perform combined regression. Determining what the common parameters of the two models are can be a challenging task; see Ref. 3 for further details. In combined regression we define the combined χ^2 function to be the sum of the individual χ^2 functions for each of the tools. Note that this is only possible if we assume that the two measurements are independent of each other

$$\chi^2(\mathbf{p}_{A \cup B}) = \chi_A^2(\mathbf{p}_A) + \chi_B^2(\mathbf{p}_B). \quad (7)$$

Here $\mathbf{p}_{A \cup B}$ is the vector that consists of the union of the elements in \mathbf{p}_A and \mathbf{p}_B . The solution $\hat{\mathbf{p}}_{A \cup B}$ to the inverse problem in combined regression is then found by minimizing the above combined χ^2 function.

3 Measured Targets and Generalized Parameter Sets

The investigated targets and the geometrical parameterizations used have already been described in detail in Ref. 3, so we will only give a brief overview. We investigate finite 30-line arrays of Si on Si with a thin native conformal oxide (see Fig. 1). The nominal widths are 14, 16, and 18 nm. A schematic representation of the geometrical parameterizations can be found in Fig. 2. For the OCD analysis, the geometry is fully characterized by the height, width, Δ_{top} , and Δ_{bot} of a single line. The physics by which this geometry interacts with incident light to produce a signal was approximated by the rigorous coupled-wave analysis (RCWA) model that is based on a semianalytical treatment of Maxwell's equations.^{12,13} The OCD data used in this study have been generated using this RCWA model for 30 lines and a measurement setup similar to the actual experiment at NIST.^{14,15}

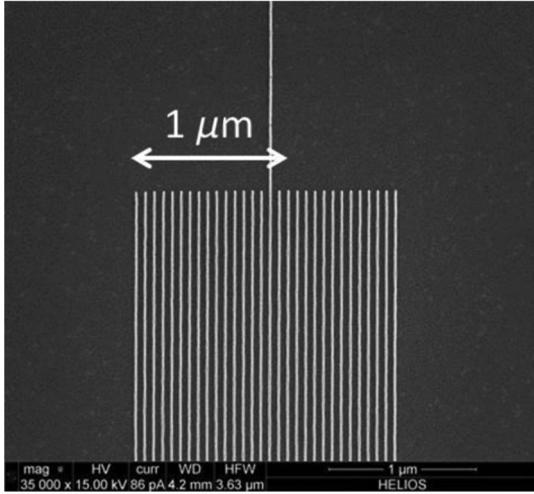


Fig. 1 Scanning electron microscope (SEM) image showing the 30 line arrays. Horizontal field of view is 3.63 μm .

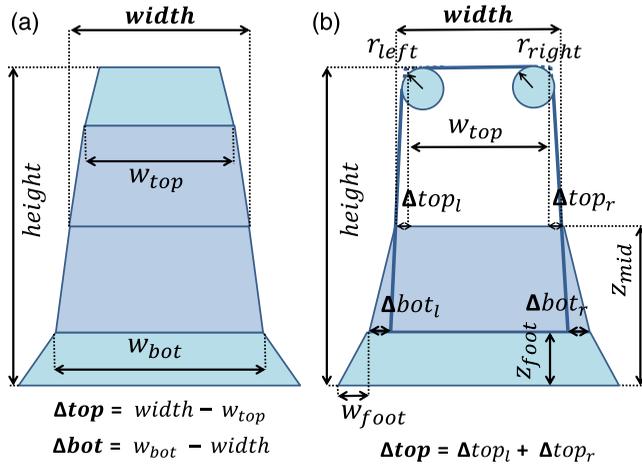


Fig. 2 Initial parameters for (a) optical critical dimension (OCD) and (b) SEM modeling. The parameters that are in bold font are part of the reduced subset of these parameters that has been used in this work.

The nominal values are height = 36.0 nm, width = 17.1 nm, $\Delta\text{top} = 2.9$ nm, and $\Delta\text{bot} = 0.5$ nm. The noise that has been added to this simulated data has been generated using the same correlated errors as described in Sec. 4.1.

The initial geometrical parameters for the SEM fit of each line included the line position, its width at half height, and its left and right sidewall edge slopes (represented by the edge widths). Although the left and right sidewall slopes of an individual line often differed, this was a manifestation of roughness. On average, left and right edge widths were indistinguishable. Consequently, for hybridization with OCD, which averages over a comparatively large area, it was sufficient to treat sidewall width as a single parameter. Similarly, line position was not important for OCD. Hybridization was limited, therefore, to two parameters: the width at the middle and Δtop . The geometry is related to the measured signal by JMONSEL, a single-scattering Monte Carlo simulator with a choice among a number of physical models of the scattering phenomena that affect electron transport. We used tabulated Mott cross sections to model scattering of electrons from nuclei and a dielectric function theory

approach to secondary electron generation. Electron–phonon interactions and electron refraction (or total internal reflection) at the vacuum/Si interface were included. Details of JMONSELs treatment of these models are available elsewhere.¹⁶

4 Error Analysis

Equipped with the proper set of overlapping parameters, we will now focus on a more realistic modeling of the present measurement data. We will investigate the influence of tool-induced measurement errors for OCD and the influence of scaling errors for SEM.

4.1 Optical Critical Dimension

The specific way in which the OCD measurements are performed, by taking images on the xy plane going through focus, requires increased attention to possible systematic and correlated errors. For example, one can imagine that being slightly off axis in the illumination will affect the symmetry of the entire set of collected images. Determining such correlations from measurement data alone is often not possible, and other methods need to be developed in order to give a quantitative description of those effects. We use the Monte Carlo method,¹⁷ which is based on the following reasoning. If we denote by ν the vector of k fixed parameters of the measurement setup, we can make use of the following more general model function:

$$\mathbf{f}: \mathbb{R}^{n+k} \rightarrow \mathbb{R}^m, \quad \mathbf{f}(\mathbf{p}, \nu) = [f_1(\mathbf{p}, \nu), \dots, f_m(\mathbf{p}, \nu)]^T. \quad (8)$$

The effect that a slight deviation of the parameters ν from the nominal values ν_0 has upon the simulated image can then be estimated from the following five steps:

1. Assume a distribution for ν , based on tool specifications, expert knowledge, etc.
2. Draw a sample $\{\nu_i\}_{i=1, \dots, N}$ from the distribution.
3. Calculate $\{\mathbf{f}(\mathbf{p}, \nu_i) = [f_1(\mathbf{p}, \nu_i), \dots, f_m(\mathbf{p}, \nu_i)]^T\}_{i=1, \dots, N}$.
4. Define $\mathbf{r}_i = \mathbf{f}(\mathbf{p}, \nu_0) - \mathbf{f}(\mathbf{p}, \nu_i)$.
5. Calculate the sample covariance matrix via

$$\tilde{\mathbf{V}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})^T. \quad (9)$$

In this paper, the ν vector included components for the collection numerical aperture (CNA), illumination numerical aperture (INA), focus heights, and phase. This Monte Carlo procedure, therefore, estimates the effect of errors propagated from variations in those instrument parameters, e.g., we choose the INA to be normally distributed with a mean of 0.13 and a standard deviation of 0.01. Figures 3 and 4 show a graphical representation of the first 752 entries of the sample covariance matrix that correspond to the first four focus heights in X polarization for phase, focus, illumination, and CNA variations, along with the respective concatenated images.

One can clearly see both the positive and negative correlation between errors as colored areas off the diagonal. In order to have a reliable model for the measurement errors, it

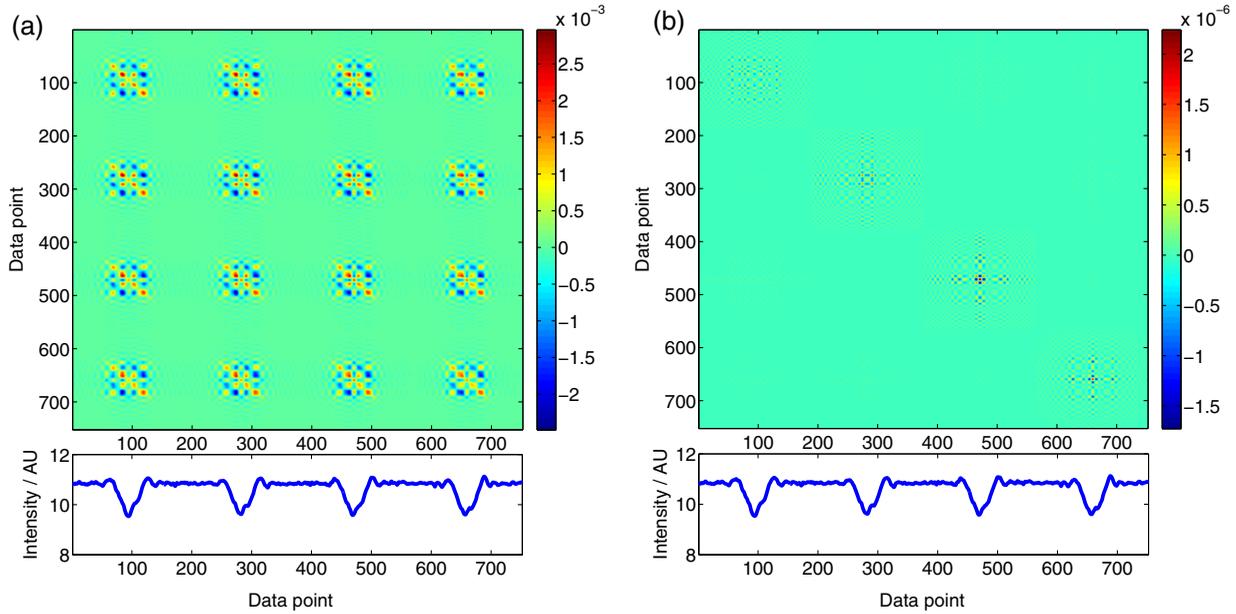


Fig. 3 Estimated covariance matrices for variations in the (a) phase and (b) focus. Random phase perturbations normally distributed with a mean of 0 rad and a standard deviation of $\pi/5$ rad, and focus offset normally distributed with a mean of 0 nm and a standard deviation of 2 nm. The variance induced by phase variations is 3 orders of magnitude greater than those from focus variation. Four concatenated OCD intensity profiles are duplicated below the matrices in Figs. 3 and 4 to illustrate the positional dependence of the covariance. In general changes correlate to the position of the finite target.

is therefore very important to account for this effect in the covariance matrix that is being used in the χ^2 function [see Eq. (3)]. This has an influence not only on the best fit values, but also on the estimation of the parametric uncertainties or, more precisely, the covariance matrix Σ , diagonal elements of which are the uncertainties in the estimated parameters, given by²

$$\sum = (\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J})^{-1}. \quad (10)$$

Here \mathbf{J} denotes the Jacobian matrix, i.e., the matrix of all first-order partial derivatives of the vector-valued model function \mathbf{f} , at the best fit vector $\hat{\mathbf{p}}$. A comparison of the parametric uncertainties based on using a diagonal \mathbf{V} and the full \mathbf{V} in Eq. (10) can be found in Table 1.

Note that there is a notable difference between the estimated parametric uncertainties, with the most significant change in the uncertainty of the width. It is also very important to note that a given parametric uncertainty might increase or decrease if correlations in the measurement data are taken into account. Thus, a general statement about the effect correlated errors have on the estimated parametric uncertainties cannot be given, and it must be investigated separately for each new problem.

4.2 Scaling Errors in Scanning Electron Microscope

The biggest contribution to the SEM's measurement error in this experiment is due to pixel calibration. Errors in this calibration directly influence the obtained values for the CDs. The usual approach is to simply add the uncertainty due to the calibration in quadrature with the estimated parametric uncertainty after the reconstruction. The estimation of the parameters of interest, i.e., the vector \mathbf{p} and their uncertainties in combined regression, are based on the combination of

the OCD and the SEM data, while the error induced by pixel calibration only affects SEM data; hence, it is not possible to simply include the uncertainty due to the calibration afterward. We will therefore model the effect a variation in the scale has on the measurement in a simple way, multiplying the model parameter vector $\mathbf{p} = (\text{width}, \Delta\text{top})^T$ by a scaling parameter κ such that the modified model is given by

$$\tilde{\mathbf{f}}: \mathbb{R}^3 \rightarrow \mathbb{R}^m, \quad \tilde{\mathbf{f}}(\kappa, \mathbf{p}) = \mathbf{f}(\kappa \cdot \mathbf{p}) \quad (11)$$

with \mathbf{f} being the SEM's model function. In this description, $\kappa = 1$ corresponds to no scale error, $\kappa = 1.01$ to 1% scale error, etc. It is clear that this approach leads to a high parametric correlation in the parameters of the model. We will explain this effect using a simple model with an added scaling factor in the following:

Let

$$\mathbf{f}: \mathbb{R} \rightarrow \mathbb{R}^m, \quad \mathbf{f}(x) = [f_1(x), \dots, f_m(x)]^T, \quad \text{and} \\ \mathbf{J} = (J_{i,1})_{i=1, \dots, m}, \quad J_{i,1} = Df_i \quad (12)$$

be a model function depending on only one parameter x and denote by Df_i the derivative of f_i with respect to this parameter. If we assume the measurement errors to be independent and identically distributed (i.i.d.) with unit variance [hence $\mathbf{V} = (\delta_{i,j})_{i,j=1, \dots, m}$] we have for the estimated covariance matrix

$$\sum = (\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J})^{-1}, \quad \text{with} \quad \mathbf{J}^T \mathbf{V}^{-1} \mathbf{J} = \sum_{i=1}^m [Df_i]^2, \quad (13)$$

which is well defined if $\det(\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J}) = \sum_{i=1}^m [Df_i]^2 \neq 0$.

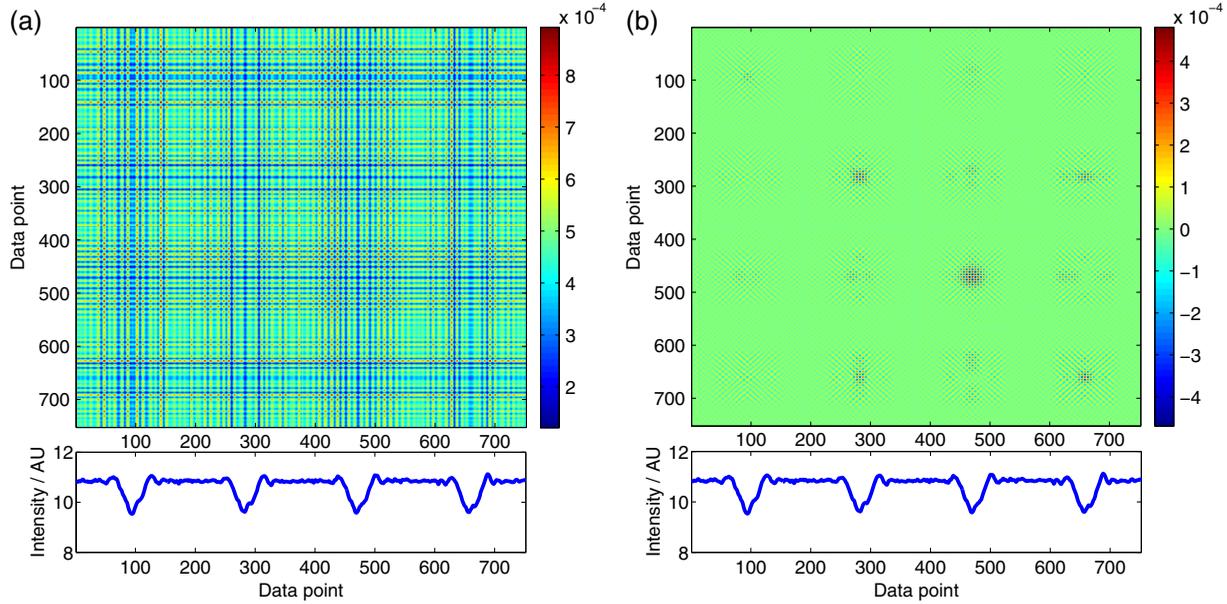


Fig. 4 Estimated covariance matrices for variations in the (a) illumination numerical aperture (INA) and (b) collection numerical aperture (CNA). INA normally distributed with a mean of 0.13 and a standard deviation of 0.01, CNA normally distributed with a mean of 0.95 and a standard deviation of 0.1. The variances induced by INA and CNA variations are one magnitude less than those from phase errors.

Now assume that we add a scale parameter to the above model by defining a slightly modified function

$$\begin{aligned} \mathbf{F}: \mathbb{R}^2 &\rightarrow \mathbb{R}^m, \quad \mathbf{F}(\kappa, x) = \mathbf{f}(\kappa \cdot x), \quad \text{and} \\ \mathbf{J} &= (J_{i,j})_{i=1, \dots, m; j=1,2}, \quad J_{i,1} = x \cdot Df_i, \\ J_{i,2} &= \kappa \cdot Df_i. \end{aligned} \quad (14)$$

The estimated covariance matrix for this two-parameter model is then given by

$$\begin{aligned} \Sigma &= (\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J})^{-1}, \quad \text{with} \\ \mathbf{J}^T \mathbf{V}^{-1} \mathbf{J} &= \begin{bmatrix} \sum_{i=1}^m x^2 [Df_i]^2 & \sum_{i=1}^m \kappa x [Df_i]^2 \\ \sum_{i=1}^m \kappa x [Df_i]^2 & \sum_{i=1}^m \kappa^2 [Df_i]^2 \end{bmatrix}. \end{aligned} \quad (15)$$

Table 1 Estimates of the parametric uncertainties for simulated OCD data using diagonal \mathbf{V} , i.e., only accounting for uncorrelated random noise, and full \mathbf{V} , i.e., taking correlations into account.

	Width (nm)	Δ_{top} (nm)	Δ_{bot} (nm)	Height (nm)
$\hat{\mathbf{p}}$	17.09	2.04	0.10	35.37
σ diagonal \mathbf{V}	0.05	1.37	0.53	0.97
σ full \mathbf{V}	0.14	1.29	0.63	0.92

However,

$$\begin{aligned} \det(\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J}) &= \sum_{i=1}^m x^2 [Df_i]^2 \cdot \sum_{i=1}^m \kappa^2 [Df_i]^2 \\ &\quad - \left\{ \sum_{i=1}^m \kappa x [Df_i]^2 \right\}^2 = 0. \end{aligned} \quad (16)$$

Since the above term is always equal to zero, the estimated covariance matrix Σ is not defined and we cannot assign a parametric uncertainty. Since we have prior information about the scale, the actual error lies between 1% and 2%; we can use the Bayesian approach as described in Refs. 2 and 9 under the premise that the prior information can be expressed in terms of normal distributions. The prior information on the parameter κ is treated as an additional data point the model function has to account for, such that we have a function that still depends on κ and x but now maps into an $(m+1)$ -dimensional space, with the $(m+1)$ 'th value simply being κ . This also adds additional terms to the Jacobian, $J_{m+1,1} = (\partial/\partial\kappa)\kappa = 1$ and $J_{m+1,2} = (\partial/\partial x)\kappa = 0$, such that

$$\begin{aligned} \tilde{\mathbf{F}}: \mathbb{R}^2 &\rightarrow \mathbb{R}^{m+1}, \quad \tilde{\mathbf{F}}(\kappa, x) = [\mathbf{F}(\kappa, x), \kappa]^T, \quad \text{and} \\ J_{m+1,1} &= 1, \quad J_{m+1,2} = 0. \end{aligned} \quad (17)$$

Since we know that $\kappa \sim \mathcal{N}(1, \sigma_\kappa^2)$, we add an additional entry to \mathbf{V} , namely $V_{m+1,m+1} = \sigma_\kappa^2$ with $\sigma_\kappa = 0.01 - 0.02$, and obtain the estimated covariance matrix, again using Eqs. (10) and (13),

$$\begin{aligned} \Sigma &= (\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J})^{-1}, \quad \text{with} \\ \mathbf{J}^T \mathbf{V}^{-1} \mathbf{J} &= \begin{bmatrix} \sum_{i=1}^m x^2 [Df_i]^2 + \frac{1}{\sigma_\kappa^2} & \sum_{i=1}^m \kappa x [Df_i]^2 \\ \sum_{i=1}^m \kappa x [Df_i]^2 & \sum_{i=1}^m \kappa^2 [Df_i]^2 \end{bmatrix} \end{aligned} \quad (18)$$

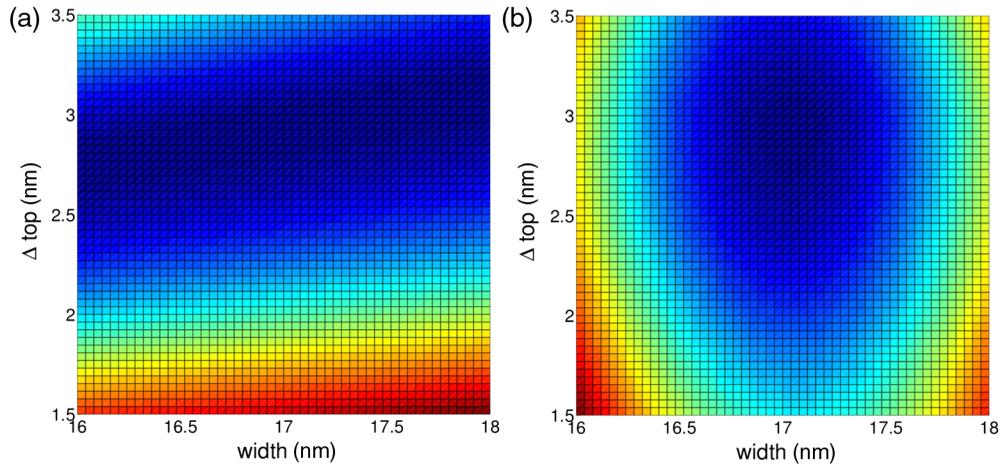


Fig. 5 χ^2 surface for SEM data (a) without and (b) with Bayesian input for the scale κ .

and

$$\det(\mathbf{J}^T \mathbf{V}^{-1} \mathbf{J}) = \left\{ \sum_{i=1}^m x^2 [Df_i]^2 + \frac{1}{\sigma_\kappa^2} \right\} \cdot \sum_{i=1}^m \kappa^2 [Df_i]^2 - \left\{ \sum_{i=1}^m \kappa x [Df_i]^2 \right\}^2. \quad (19)$$

Equation (19) implies that as long as the prior knowledge about κ is not too vague, i.e., σ_κ is not too large, the above term is in general not equal to zero and the estimated covariance matrix is well defined. A graphical representation of the above-described phenomenon for the measured SEM data is shown in Fig. 5. Note that the χ^2 surface with the added prior information about κ has a distinct minimum at $\hat{\mathbf{p}} = (1, 17.01 \text{ nm}, 2.86 \text{ nm})$, while it is hard to determine where the minimum is for the χ^2 surface without prior information. In fact, there is not a single distinct minimum but an infinite set of possible minima. Obviously, $\hat{\mathbf{p}} = (1, 17.01 \text{ nm}, 2.86 \text{ nm})$ is also the minimum for the χ^2 without prior information, but so is any vector $\hat{\mathbf{p}} = [\kappa, (17.01/\kappa) \text{ nm}, (2.86/\kappa) \text{ nm}]$ with $\kappa \neq 0$. Defining a

parametric uncertainty is therefore not possible. In contrast, the error estimation for the model with prior information for the scale κ yields a 2% parametric uncertainty if we assume an error of 2% in the scale as expected; here, this strict linearity only holds since the random errors in the SEM data are much smaller than those attributed to the scale error.

5 Results

We now combine the results that we found in the previous section with the hybridization of OCD and SEM data by combined regression. As pointed out in Sec. 2, this is done by minimizing the sum of the respective χ^2 functions. Note that we use prior information about the scale κ , so that the χ^2 function for the SEM data is modified as shown in Eq. (6). The individual χ^2 surfaces in dependence on the width and Δ_{top} are shown in Figs. 6 and 7. For these plots, the height and Δ_{bot} have been fixed for OCD. The plots also show the individual minima and the assigned parametric uncertainties. The χ^2 surface from the combined regression is shown in Fig. 8, and the results from the combined regression are presented in Table 2. The combined minimum is close to the SEM's minimum and the parametric

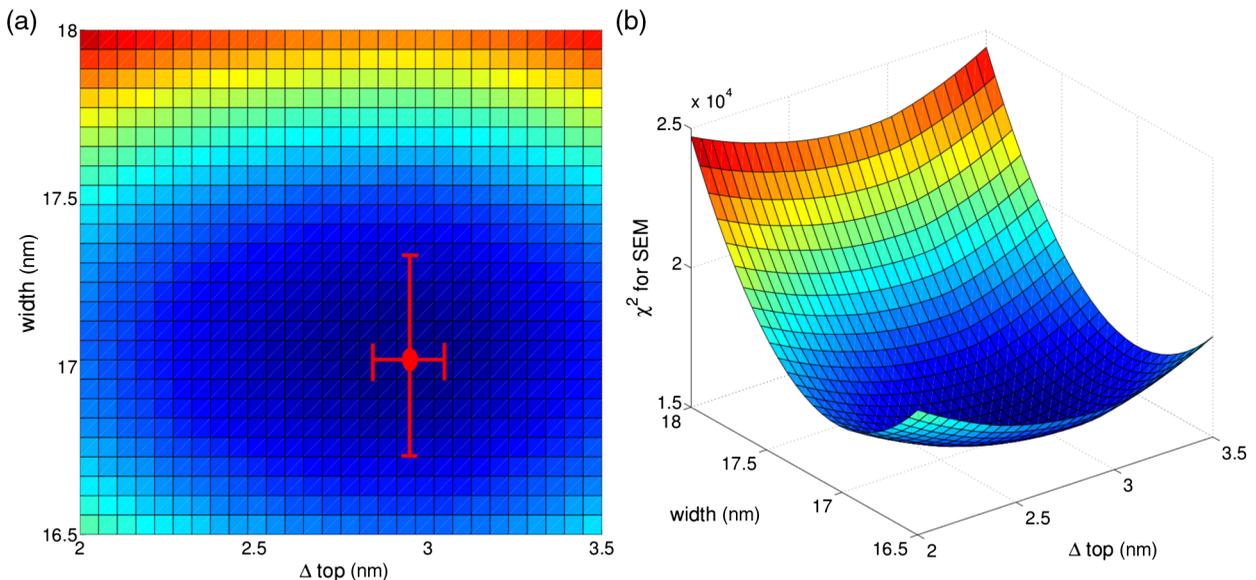
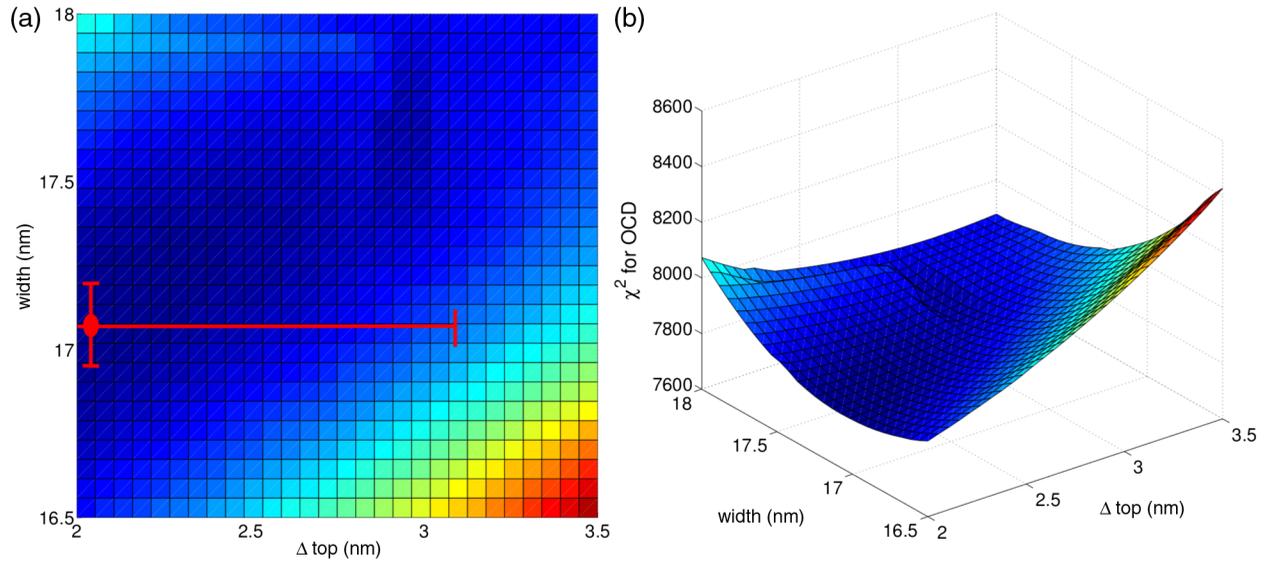
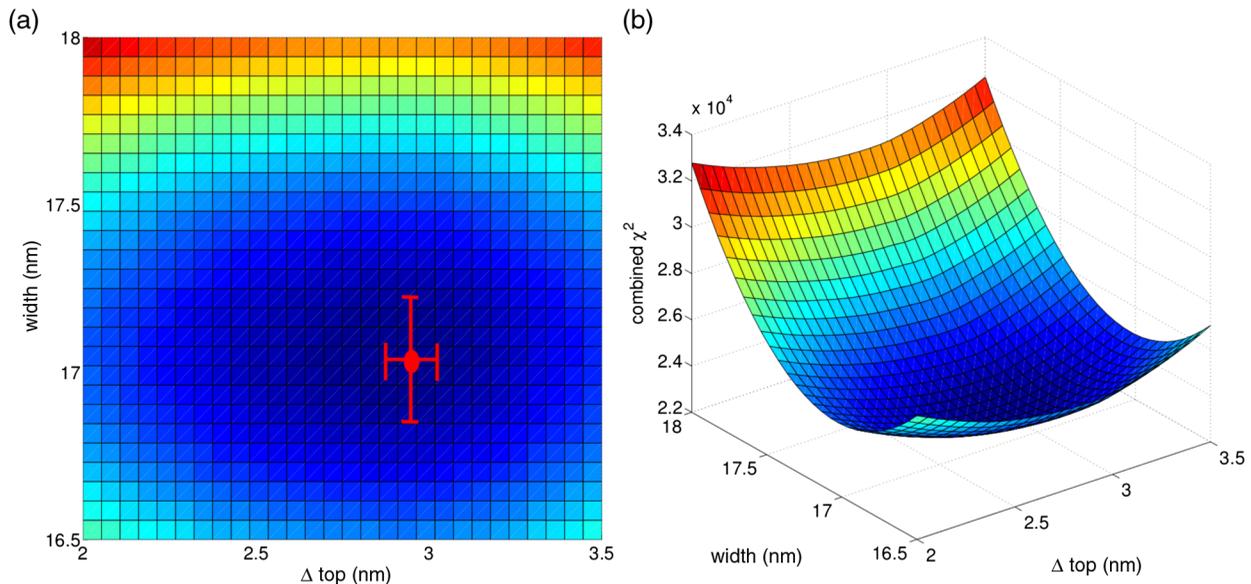


Fig. 6 χ^2 surfaces for the SEM data.


Fig. 7 χ^2 surfaces for the simulated OCD data.

Fig. 8 Combined χ^2 surface for the hybridization.

uncertainties for combined regression are lower than the individual ones, even for the parameters that are only present in the OCD model. This is due to the strong parametric correlations in the models.

Table 2 Parameter estimates and parametric uncertainties obtained from combined regression.

	Width (nm)	Δ_{top} (nm)	Δ_{bot} (nm)	Height (nm)
$\hat{\mathbf{p}}_{\text{hybrid}}$	17.01	2.94	0.59	36.11
σ_{hybrid}	0.23	0.04	0.58	0.21
σ_{SEM}	0.34	0.06	—	—
σ_{OCD}	0.14	1.29	0.63	0.92

6 Discussion and Conclusion

Following the approach outlined in Ref. 3, we studied the challenges in hybrid metrology due to measurement errors. Those included highly correlated tool-induced errors for the OCD data and systematic errors due to scaling errors in SEM. We have demonstrated how slight variations in the measurement setup for OCD, e.g., in the focus heights, the phase, INA, and CNA lead to highly correlated errors in the measurement data that manifest themselves as nonzero elements in the sample covariance matrix \mathbf{V} . Including those off-diagonal elements in the estimation of a parametric uncertainty can lead to either an increased or a decreased parametric uncertainty compared to the case where only the diagonal of the \mathbf{V} matrix is used, depending on the individual nature of that particular full \mathbf{V} matrix. Furthermore, we demonstrated the influence of scaling errors on the analysis of SEM data. Attempting to account for such scale errors by including the scale as a fully free parameter would lead to

unreasonable results due to strong, or even perfect correlations. This problem has been solved using prior information about the scale in a Bayesian approach. Finally, we demonstrated how the more sophisticated error analyses could be used in the hybridization of OCD and SEM data. With the proper treatment of those errors, we could achieve a sub-nanometer parametric uncertainty. It is important to note that the presented framework can be extended to include additional measurement techniques, such as atomic force microscopy or CD small angle X-ray scattering. In addition, it may also be applied across a homogeneous multiple-tool platform. However, for every added measurement technique, it is crucial to perform a careful error analysis in order to use its full capabilities.

References

1. R. M. Silver et al., "Improving optical measurement accuracy using multi-technique nested uncertainties," *Proc. SPIE* **7272**, 727202 (2009).
 2. N. F. Zhang et al., "Improving optical measurement uncertainty with combined multitool metrology using a Bayesian approach," *Appl. Opt.* **51**, 6196–6206 (2012).
 3. R. M. Silver et al., "Optimizing hybrid metrology through a consistent multi-tool parameter set and uncertainty model," *Proc. SPIE* **9050**, 905004 (2014).
 4. M.-A. Henn et al., "Optimizing hybrid metrology: rigorous implementation of Bayesian and combined regression," *Proc. SPIE* **9424**, 94241J (2015).
 5. A. Vaid et al., "A holistic metrology approach: hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM," *Proc. SPIE* **7971**, 797103 (2011).
 6. A. Vaid et al., "Implementation of hybrid metrology at HVM fab for 20 nm and beyond," *Proc. SPIE* **8681**, 868103 (2013).
 7. J. Foucher et al., "Hybrid CD metrology concept compatible with high-volume manufacturing," *Proc. SPIE* **7971**, 79710S (2011).
 8. D. S. Sivia, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Oxford (1996).
 9. A. Gelman et al., *Bayesian Data Analysis*, Chapman&Hall/CRC, London (2014).
 10. J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer Science & Business Media, Berlin (2006).
 11. A. Tarantola and B. Valette, "Generalized nonlinear inverse problems solved using the least squares criterion," *Rev. Geophys.* **20**, 219–232 (1982).
 12. M. Moharam et al., "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A* **12**, 1068–1076 (1995).
 13. M. Moharam et al., "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).
 14. R. M. Silver et al., "Scatterfield microscopy for extending the limits of image-based optical metrology," *Appl. Opt.* **46**, 4248–4257 (2007).
 15. B. M. Barnes et al., "Zero-order and super-resolved imaging of arrayed nanoscale lines using scatterfield microscopy," *AIP Conf. Proc.* **931**, 397–401 (2007).
 16. J. S. Villarrubia et al., "Scanning electron microscope measurement of width and shape of 10 nm patterned lines using a JMONSEL-modeled library," *Ultramicroscopy* **154**, 15–28 (2015).
 17. D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo Methods*, John Wiley & Sons, Hoboken, NJ (2013).
- Mark-Alexander Henn** is a guest researcher at the National Institute of Standards and Technology. He received his diploma in mathematics from the Technical University of Berlin, Germany, in 2008, and his PhD degree in theoretical physics from the Technical University of Berlin in 2013. Prior to that he worked for the Physikalisch-Technische Bundesanstalt in Berlin. His current research interests include Bayesian data analysis, electromagnetic modeling, and Fourier optics.
- Richard Silver** received his BA in physics from the University of California at Berkeley and his PhD in physics from the University of Texas at Austin. He is the Surface and Nanostructure Metrology Group Leader, a fellow of SPIE, co-chair of the European Optical Modeling Conference and on the Program Committee of the Metrology, Inspection, and Process Control Conference. He has over 100 archived publications and was a recipient of the 2013 R&D 100 Award for Quantitative Hybrid Metrology and the 2013 Intel Outstanding Researcher Award.
- John S. Villarrubia** received his PhD in physics from Cornell University in 1987. He was a visiting scientist for two years at IBM, where he did scanning tunneling microscopy of silicon surfaces. At the National Institute of Standards and Technology since 1989, he has been working on aspects of nanometer-scale dimensional metrology, particularly modeling the instrument function for Atomic Force and Scanning Electron Microscopes and using these models of the instrument function to correct images for measurement artifacts.
- Nien Fan Zhang** is presently a Mathematical Statistician of the Statistical Engineering Division at the US National Institute of Standards and Technology (NIST). He received MS and PhD in statistics in 1983 and 1985 respectively, at Virginia Polytechnic Institute and State University.
- Hui Zhou** is a research scientist and a consultant at Dakota Consulting, Inc. His main research area is computational electrodynamics, and his interest spreads to all areas between physics and computing. When he is not at work, you may find him reading a random non-fiction book, playing a game of go, or dreaming of computation and physics in the future.
- Bryan M. Barnes** is a physicist at the National Institute of Standards and Technology. He received his BA degree in mathematics and physics from Vanderbilt University in 1995, and his MS and PhD degrees in physics from the University of Wisconsin-Madison in 1997 and 2004, respectively. He is the author of more than 40 proceeding and journal papers and holds one patent. His current research interests include optical defect inspection, hybrid metrology, and critical dimension metrology.
- Andras E. Vladar**, PhD is the leader of the Three-Dimensional Nanometer Metrology Project at the National Institute of Standards and Technology (NIST) USA. He is an expert in scanning electron microscopy and dimensional metrology, one of the best-known research scientists and a technical leader of this field. His research interest is in SEM-based sub-10 nm three-dimensional measurements for semiconductor and nanotechnology applications.
- Bin Ming**: Biography is not available.