

Machine learning driven measurement of high-aspect-ratio nanostructures using Mueller matrix spectroscopic ellipsometry



Cite as: J. Vac. Sci. Technol. B 43, 012801 (2025); doi: 10.1116/6.0004058

Submitted: 7 September 2024 · Accepted: 3 December 2024 ·

Published Online: 7 January 2025



Shiva Mudide,^{1,a)} Nick Keller,² G. Andrew Antonelli,² Geraldina Cruz,³ Julia Hart,³ Alexander R. Bruccoleri,³ Ralf K. Heilmann,⁴ and Mark L. Schattenburg⁴

AFFILIATIONS

¹Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

²Onto Innovation, Wilmington, Massachusetts 01887

³Izentis LLC, Cambridge, Massachusetts 02139

⁴Space Nanotechnology Laboratory, MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Note: This paper is part of the Special Topic Collection: Papers from the 67th International Conference on Electron, Ion and Photon Beam Technology and Nanofabrication (EIPBN 2024).

^{a)}Author to whom correspondence should be addressed: shiva271@mit.edu

ABSTRACT

Accurate fabrication of high-aspect ratio (HAR) structures in applications from semiconductor devices to x-ray observatories is essential for their optimal performance because their performance directly depends on their structure. High-efficiency critical-angle transmission (CAT) gratings enable high-resolution x-ray spectroscopy in astrophysics, but their performance is only ideal when certain performance-critical parameters, like the bar tilts introduced during deep reactive-ion etching, are tuned to precise values. Traditional measurement methods like small-angle x-ray scattering (SAXS) are accurate, but limit the development of robust control algorithms to nudge performance-critical parameters toward favorable values because they are slow and often destructive. We present a fast, accurate, nondestructive measurement method using Mueller matrix spectroscopic ellipsometry and machine learning. Given a HAR structure, we train on rigorous coupled-wave analysis simulation data to predict Mueller matrix spectra from input performance-critical parameter values. We then invert this forward problem by freezing our network weights, measuring experimental Mueller matrix spectra, and vanilla gradient descending on performance-critical parameters to values that correspond to the input Mueller matrix spectra. Introducing machine learning to invert the forward problem reduces computation time, and experimental results demonstrate close agreement between our method's determined tilt and SAXS measurements. Our accurate, fast measurement method paves the way for the development of robust control algorithms that adjust fabrication parameters in response to measurement, ensuring optimal performance in not only CAT gratings but also HAR structures embedded in applications from semiconductor to microelectromechanical systems fabrication.

© 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1116/6.0004058>

I. INTRODUCTION

High-aspect ratio (HAR) microelectronic and photonic devices are critical components in advanced technological applications, including microelectromechanical systems (MEMSs), semiconductor devices, solar cells, and x-ray observatories. Accurate fabrication is essential for HAR device performance because

optimal performance is by virtue of the precise structure itself. For example, smooth and straight sidewalls in HAR device structures minimize defects and roughness, reducing electrical resistivity and improving performance in MEMS.¹

Accurate fabrication is only possible through measurement. Critical structure parameters (like a parameter that formally defines

smoothness) that yield optimal device performance must be measured, ideally in real-time. If measurements deviate from desired values, control algorithms can correct fabrication parameters to nudge performance-critical structure parameters to these desired values. For example, in the critical-angle transmission (CAT) diffraction grating bars for x-ray observatories discussed in this paper, near-zero-degree tilt bars are crucial for optimal performance. When fabricating diffraction gratings, accurate measurement of the grating bar tilt can reveal deviations from zero-degree tilt bars. Control algorithms can then perturb etch parameters like substrate bias to ensure that the tilt remains near zero.²

A current limitation to the existence of such robust control algorithms for all HAR applications from MEMS to semiconductors, is the speed of measurement. Traditional techniques like scanning electron microscopy (SEM) and atomic force microscopy (AFM) are accurate but often slow and limited in real-time measurements.³ In this paper, we present a method for fast and accurate measurement of performance-critical HAR structure parameters. We demonstrate the efficacy of our method by using it to measure bar tilts in CAT gratings, but the method can be generalized to measure performance-critical HAR structure parameters in many other micro electro/photonic devices, especially those that are periodic. This paves the way for robust control algorithms that perturb fabrication parameters in real-time, ultimately accelerating optimal performance microelectronic and photonic device development.

CAT gratings, fabricated from silicon-on-insulator (SOI) wafers, are ultra-HAR structures used in high-resolution x-ray spectroscopy.⁴⁻⁷ They are blazed transmission gratings, reflecting x rays off their sidewalls at grazing angles below the critical angle for total external reflection, thus maximizing diffraction efficiency in higher orders and enabling high spectral resolving power (Fig. 1). They combine the high diffraction efficiency and resolving power of blazed reflection gratings with the practical advantages of transmission gratings.

One critical challenge in fabricating CAT gratings is the introduction of undesired grating bar tilts during the deep reactive-ion etching⁸ (DRIE) step (Fig. 2). These tilts significantly affect the incident x-ray angle, impacting the blazing behavior.^{9,10} Accurate measurement and characterization of these bar tilts are essential for fine-tuning of the DRIE step and thus optimizing grating performance.² For CAT gratings and other HAR structures, traditional methods like small-angle x-ray scattering (SAXS) can be accurate but require destructive sample thinning if used early in the fabrication process and are time consuming.⁹

The basic CAT grating fabrication steps are shown in Fig. 3. Initial steps (1 and part of 2) are performed on specially designed 200 mm-diameter SOI wafers patterned at MIT Lincoln Lab. Subsequent steps are performed on campus labs including MIT.nano. Mueller matrix spectroscopic ellipsometry (MMSE) can be applied immediately after the critical DRIE of the device layer (Step 3), providing prompt feedback without requiring thinning the back side layer as required for SAXS.

We propose using MMSE for the nondestructive characterization of bar tilts in CAT gratings. MMSE measures changes in polarization as light reflects off a sample, providing detailed information about the sample's optical properties and structure. By capturing

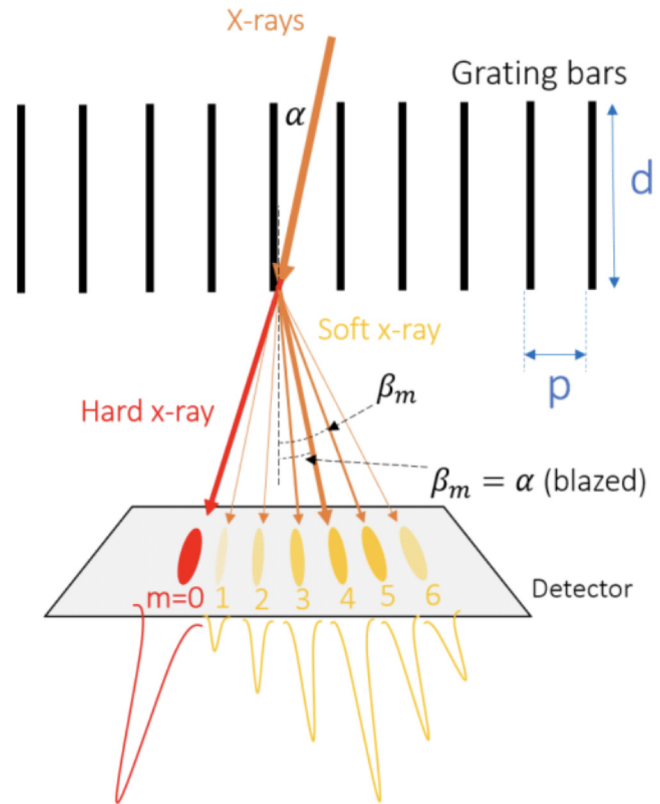


FIG. 1. Principle of CAT diffraction gratings. X rays efficiently reflect from nano-polished sidewalls of grating bars. Diffraction efficiency is enhanced (blazed) when the angle of incidence onto the side walls, α , is comparable to the angle, β_m , of a particular m th diffraction order.

07 January 2025 16:12:47

experimental MMSE spectra from the sample grating, we can build a model of the grating using a rigorous coupled-wave analysis (RCWA)-based electromagnetic solver.¹⁴

The experimental setup captures the MMSE spectra, which are then compared to the modeled spectra. The optimization process involves calculating the gradient of the square deviation between the experimental and modeled spectra with respect to the free parameters. This allows us to iteratively adjust the free parameters until the model spectra converge with the experimental data. Traditional approaches perform gradient calculations through finite differences with RCWA simulations at each step, which is computationally intensive and time-consuming. The novelty in our method is in that we replace RCWA with a neural network, and gradient descend on the input space to find a solution that best approximates experimental spectra. To our knowledge, this technique inspired by generative artificial intelligence has not been applied in HAR metrology before. We do this by first generating training data with RCWA-simulated spectra across free parameters. We train the neural net to solve the forward problem of predicting an MMSE spectra, given a point in the free parameter space. Once trained, the neural network approximates RCWA in an analytical

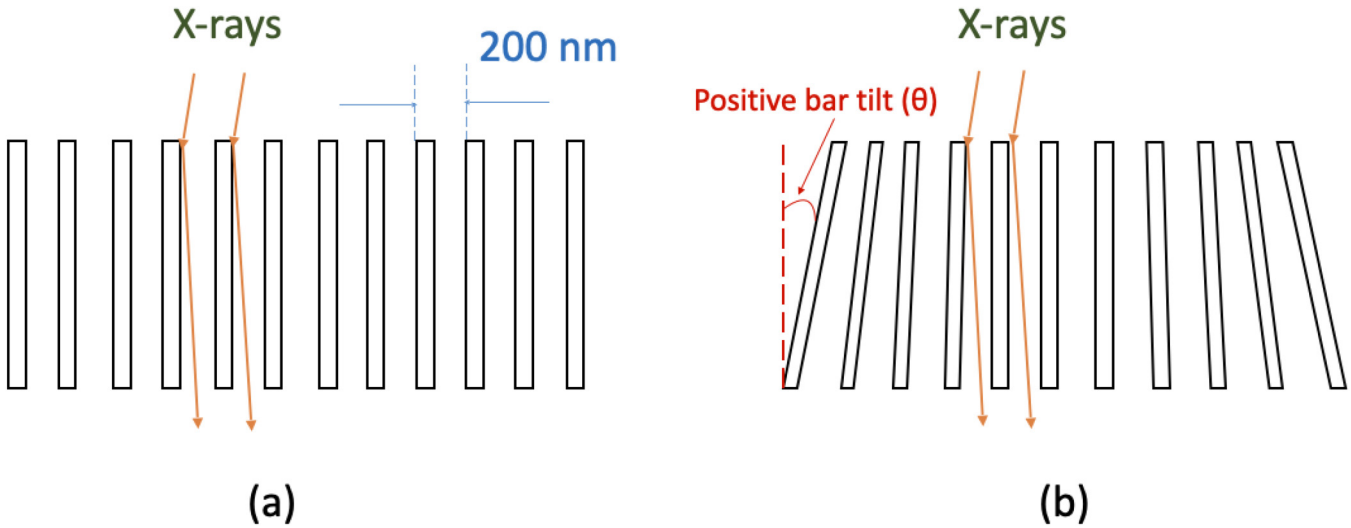


FIG. 2. Etch chamber electrostatics can cause undesirable bar tilts across wafers. (a) Ideal grating with no bar tilt. (b) Grating with bar tilt increasing with wafer radius.

form. The gradient is then calculated analytically through the network with the chain rule by freezing the network weights.

The robustness of our method is evident in its speed, consistency, and accuracy compared to calculating the gradient with finite differences using RCWA. Our neural network-based approach not only accelerates the optimization process but also maintains high accuracy in parameter estimation. Our method provides detailed results of the bar tilt across the entire wafer. By measuring multiple points on the wafer, we map the tilt variations introduced during the DRIE process. The results of the MMSE-measured tilt are validated against SAXS measurements, demonstrating that the MMSE method provides consistent and accurate tilt measurements. The non-destructive and rapid nature of MMSE, combined with the computational efficiency of neural network-based optimization, offers a powerful tool for characterizing and improving the fabrication of CAT gratings.

II. PHYSICAL FOUNDATIONS

A. Scatterometry

Scatterometry is a metrology technique used to determine the structure of a sample by analyzing the spectra of light that interacts with it.

It relies on the idea that the spectral response of light, when it interacts with a periodic structure, is unique to the structure's geometric and material properties. The scattered light carries information about the sample, which can be decoded to reconstruct the sample's physical characteristics.

One important aspect of the interaction between light and the sample is depolarization. Depolarization occurs when the polarization state of the incident light goes from fully polarized to partially polarized. This transformation to partial polarization is highly

indicative of the sample's structural properties, particularly in complex, anisotropic structures like HAR microelectronic and photonic devices.

B. Jones vectors and Mueller matrices

Jones calculus is a mathematical formalism used to describe the polarization state of light and its transformation through optical elements. Although we do not use Jones vectors directly in our analysis, discussing them provides a foundational understanding of polarization. A Jones vector represents the electric field components of fully polarized light in a given two-dimensional complex vector space

$$\mathbf{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix}. \quad (1)$$

An optical element is represented by a Jones matrix \mathbf{J} , which transforms the input Jones vector \mathbf{E}_{in} to the output Jones vector \mathbf{E}_{out} ,

$$\mathbf{E}_{out} = \mathbf{J}\mathbf{E}_{in}. \quad (2)$$

While Jones calculus is effective for fully polarized light, it does not capture partially polarized light, which manifests when light interacts with anisotropic structures like HAR microelectronic and photonic devices.

To account for depolarization, we use the Mueller matrix formalism. The Mueller matrix \mathbf{M} is a 4×4 matrix that transforms the Stokes vector of the incident light \mathbf{S}_{in} to the Stokes vector of the scattered light \mathbf{S}_{out} ,

$$\mathbf{S}_{out} = \mathbf{M}\mathbf{S}_{in}. \quad (3)$$

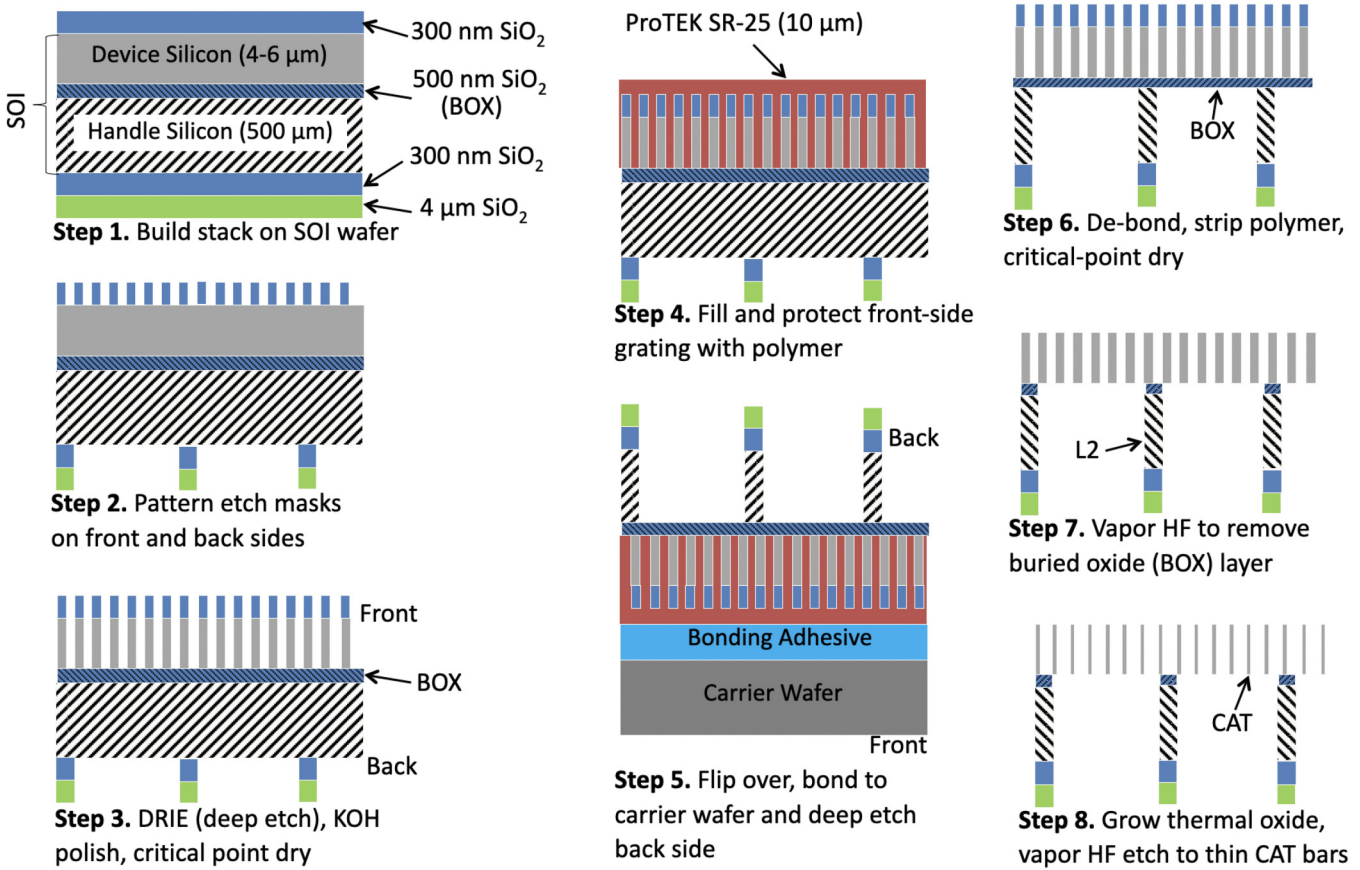


FIG. 3. Basic CAT grating fabrication steps include front side (device layer) patterning, etching the front side oxide mask, aligned back side patterning, etching back side oxide mask, DRIE of the Si device layer, wet KOH sidewall polish (Ref. 11) [Reprinted with permission from Bruccoleri *et al.*, J. Vac. Sci. Technol. B **31**, 06FF02 (2013). Copyright 2013 American Vacuum Society], front side protection and mounting to carrier, DRIE of the Si handle layer, separation from carrier, front side clean, critical-point drying, and buried oxide removal (Refs. 12 and 13) [Reprinted with permission from Bruccoleri *et al.*, J. Vac. Sci. Technol. B **34**, 06KD02 (2016). Copyright 2016, American Vacuum Society; Reprinted with permission from Heilmann *et al.*, Proc. SPIE **11444**, 114441H (2021). Copyright 2021, SPIE]. MMSE can be applied immediately after the critical DRIE of the device layer (step 3), providing prompt feedback without thinning the back side layer for SAXS.

07 January 2025 16:12:47

The Stokes vector \mathbf{S} captures light that is in a distribution of polarization states and is defined as

$$\mathbf{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} I_x + I_y \\ I_x - I_y \\ I_{45^\circ} - I_{-45^\circ} \\ I_R - I_L \end{pmatrix}. \quad (4)$$

The coordinate system is defined such that x and y are the orthogonal linear polarization directions with respect to the horizontal and vertical axes of the laboratory frame. I is the light's intensity in a particular polarization direction. 45° and -45° denote the polarization directions at $+45^\circ$ and -45° relative to the horizontal axis, and R and L represent the right and left circular polarizations, respectively.

To obtain the Mueller matrix \mathbf{M} , we use linearly independent Stokes vectors as inputs and measure the corresponding output

Stokes vectors. (The experimental setup is described in Sec. VI A.) By solving a system of linear equations, we can determine the elements of \mathbf{M} .

A common approach is to choose the four orthonormal basis states in the standard basis as the four linearly independent Stokes vectors. The output Stokes vector for each input basis state then becomes a column in the Mueller matrix.

By measuring the Mueller matrix $\mathbf{M}(\lambda)$ across a range of wavelengths, we construct a spectral tensor $\mathcal{M}(\lambda)$. This tensor captures the wavelength-dependent polarization transformation properties of the sample

$$\mathcal{M}(\lambda) = \begin{pmatrix} M_{11}(\lambda) & M_{12}(\lambda) & M_{13}(\lambda) & M_{14}(\lambda) \\ M_{21}(\lambda) & M_{22}(\lambda) & M_{23}(\lambda) & M_{24}(\lambda) \\ M_{31}(\lambda) & M_{32}(\lambda) & M_{33}(\lambda) & M_{34}(\lambda) \\ M_{41}(\lambda) & M_{42}(\lambda) & M_{43}(\lambda) & M_{44}(\lambda) \end{pmatrix}. \quad (5)$$

Using the Mueller matrix formalism to determine the structure of HAR microelectronic and photonic devices presents several challenges. The relationship between the Mueller matrix elements and the physical parameters of the sample is nonlinear. Furthermore, depolarization effects introduce further complexity, requiring sophisticated algorithms and models to accurately interpret the measured spectra. Despite these challenges, the comprehensive information provided by MMSE makes it a powerful tool for characterizing HAR structures.

III. MATHEMATICAL FORMULATION

The ultimate goal of our method is to map the measured spectra directly to the physical structure of the HAR microelectronic and photonic devices. This involves determining the structural parameters of the sample from the spectral data obtained through MMSE.

Directly mapping the spectra to the structure is challenging because the spectra are influenced by multiple interdependent parameters, each simultaneously contributing to the overall response.

Instead, we solve the forward problem using RCWA simulations and invert it. Our method is essentially a method to invert RCWA simulations that is faster and less computationally intensive than other methods in the literature, while maintaining accuracy. RCWA is a semi-analytical electromagnetic simulation method that computes the diffraction efficiencies of periodic structures by solving Maxwell's equations. It provides a way to generate theoretical spectra for a given set of structural parameters, which can then be compared with the experimental spectra.

A. RCWA simulation

RCWA discretizes the structure into layers and solves Maxwell's equations in each layer. The electric and magnetic fields within each layer are expressed as a Fourier series, and the boundary conditions are applied at the interfaces between layers. The resulting system of linear equations is solved to obtain the diffraction efficiencies.

Mathematically, RCWA can be described as follows. Let $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ be the electric and magnetic fields, respectively, in the grating structure. The fields can be expanded in terms of spatial harmonics,

$$\mathbf{E}(\mathbf{r}) = \sum_n \mathbf{E}_n e^{i\mathbf{k}_n \cdot \mathbf{r}}, \quad \mathbf{H}(\mathbf{r}) = \sum_n \mathbf{H}_n e^{i\mathbf{k}_n \cdot \mathbf{r}}, \quad (6)$$

where \mathbf{k}_n are the wave vectors of the spatial harmonics. The fields within each layer are coupled through the boundary conditions, leading to a matrix eigenvalue problem that can be solved to obtain the diffraction efficiencies.¹⁴

B. Confining physical model space

To make the problem tractable, we confine the parameter space by selecting a physical model that captures the essential features of the HAR microstructures. This reduces the dimensionality of the parameter space from a very large one that captures the

space of all possible HAR microstructures to one that is smaller, namely, less than 10. Formally, we say that our parameter space is reduced from \mathbb{R}^d to $\mathbb{R}^{d'}$, where $d' \ll d$, \mathbb{R}^d is the number of dimensions in a large space that captures all possible HAR microstructures, and $\mathbb{R}^{d'}$ only captures variations in the performance-critical parameters of our structure,

$$\mathbb{R}^d \rightarrow \mathbb{R}^{d'}. \quad (7)$$

Typical parameters in this reduced space, shown in Fig. 4, include the thickness and offset of the SiO₂ hard mask layer, the tilt angle of the grating bars, and the coefficients of a Legendre polynomial parameterization of the trench critical dimension (CD).

C. Free parameter selection and minimization of differences between RCWA and experimental spectra

Once the physical model is defined, we select the free parameters that are varied during the optimization process. Table I summarizes the performance critical parameters that we allow to vary, and their ranges of perturbation. The goal is to find the set of parameters that minimizes the mean squared error (MSE) between the RCWA-simulated spectra and the experimental MMSE spectra. This optimization problem can be formally stated as

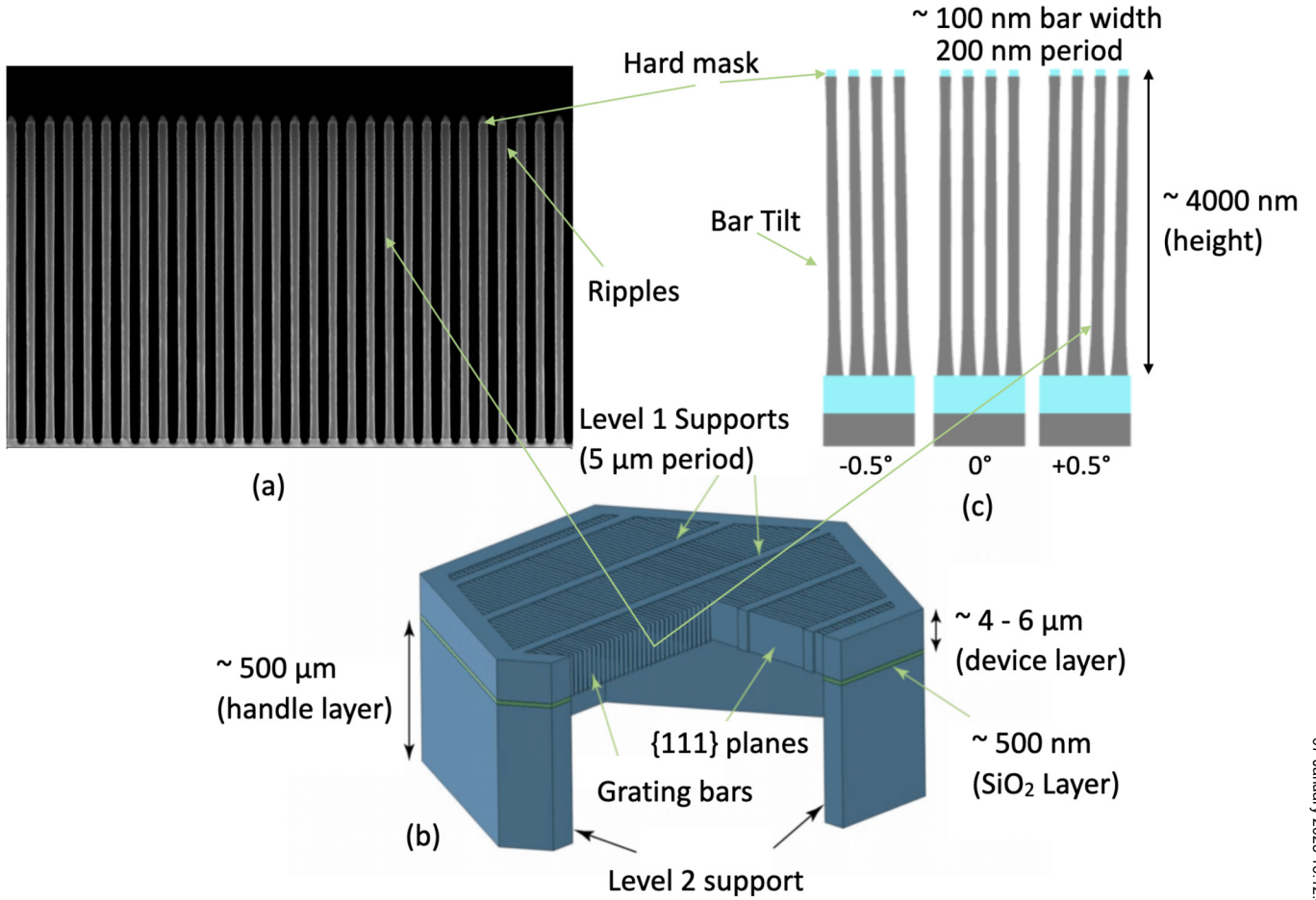
$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \|\mathcal{M}_{\text{sim}}(\mathbf{p}) - \mathcal{M}_{\text{exp}}\|^2, \quad (8)$$

where \mathbf{p} is the vector of free parameters, $\mathcal{M}_{\text{sim}}(\mathbf{p})$ is the simulated Mueller matrix spectra across wavelengths, and \mathcal{M}_{exp} is the experimental Mueller matrix spectra across wavelengths.

To demonstrate that the RCWA-simulated spectra are indeed sensitive to the performance-critical parameters, we have chosen to vary in our optimization process; we vary these parameters and visually inspect the resultant RCWA-simulated spectra. The simulated spectra have 16 different matrix elements as a function of wavelength. All the matrix elements have sensitivity to our performance-critical parameters, so we fit to all matrix elements in our method; however, we note that the off-diagonal elements of the Mueller matrix, especially in the upper-right and lower-left quadrants, are more sensitive to asymmetry, because they capture the cross-polarization effects and interactions between different polarization states. The diagonal elements generally describe the overall intensity and depolarization effects, which are often related to symmetric properties of the sample. The difference between our neural network's approximation of RCWA (the predicted spectra), and the spectra that we measure is called a "loss" function.

We carefully incorporate the above insights about the sensitivity of the Mueller matrix elements into our loss function. Specifically, we use the mean squared error between the predicted and measured spectra because this maximizes the Bayesian likelihood of a predicted spectra being the true, underlying spectra, given a measurement spectra (assuming Gaussian noise and uniform prior). Weighing the upper-right and lower-left quadrants five times more than on-diagonal elements yields the following loss

07 January 2025 16:12:47



07 January 2025 16:12:47

FIG. 4. Diagram illustrating a CAT grating on a representative 200 mm-diameter silicon wafer. (a) Cross sectional SEM image of a grating taken on a similar wafer to the one processed in this paper. (b) A model (not to scale) representing the 1 mm pitch L2 hexagonal support mesh, 5 μm period L1 support mesh, and 200 nm period CAT grating bars. (c) Depiction of a physical model used to generate RCWA data and confine the model parameter space. CAT grating bars, and top and bottom oxide layers are included in the model. The L1 and L2 support structures are not included in our current model.

function:

$$\mathcal{L} = \frac{1}{N} \sum_{\lambda=200 \text{ nm}}^{600 \text{ nm}} \sum_{i=1}^4 \sum_{j=1}^4 w_{ij} \left(M_{ij}^m(\lambda) - M_{ij}^p(\lambda) \right)^2, \quad (9)$$

where

$$w_{ij} = \begin{cases} 5, & \text{if } (i, j) \in \{(1, 4), (4, 1), (2, 4), (4, 2), (1, 3), (3, 1), (2, 3), (3, 2)\}, \\ 1, & \text{otherwise,} \end{cases}$$

N is the total number of elements summed over all wavelengths, m stands for measured, and p stands for predicted. We do not use this exact loss function, though. Instead, we use the following similar

loss function:

$$\begin{aligned} \mathcal{L} = & \frac{1}{N} \sum_{\lambda=200 \text{ nm}}^{600 \text{ nm}} \sum_{i=1}^4 \sum_{j=1}^4 \left(M_{ij}^m(\lambda) - M_{ij}^{\text{predicted}}(\lambda) \right)^2 \\ & + \sum_{\lambda=200 \text{ nm}}^{600 \text{ nm}} \left[5 \left(M_{41}^m(\lambda) - M_{14}^m(\lambda) - \left(M_{41}^p(\lambda) - M_{14}^p(\lambda) \right) \right)^2 \right. \\ & + 5 \left(M_{42}^m(\lambda) - M_{24}^m(\lambda) - \left(M_{42}^p(\lambda) - M_{24}^p(\lambda) \right) \right)^2 \\ & + 5 \left(-\left(M_{13}^m(\lambda) + M_{31}^m(\lambda) \right) - \left(-\left(M_{13}^p(\lambda) + M_{31}^p(\lambda) \right) \right) \right)^2 \\ & \left. + 5 \left(-\left(M_{32}^m(\lambda) + M_{23}^m(\lambda) \right) - \left(-\left(M_{32}^p(\lambda) + M_{23}^p(\lambda) \right) \right) \right)^2 \right]. \end{aligned} \quad (10)$$

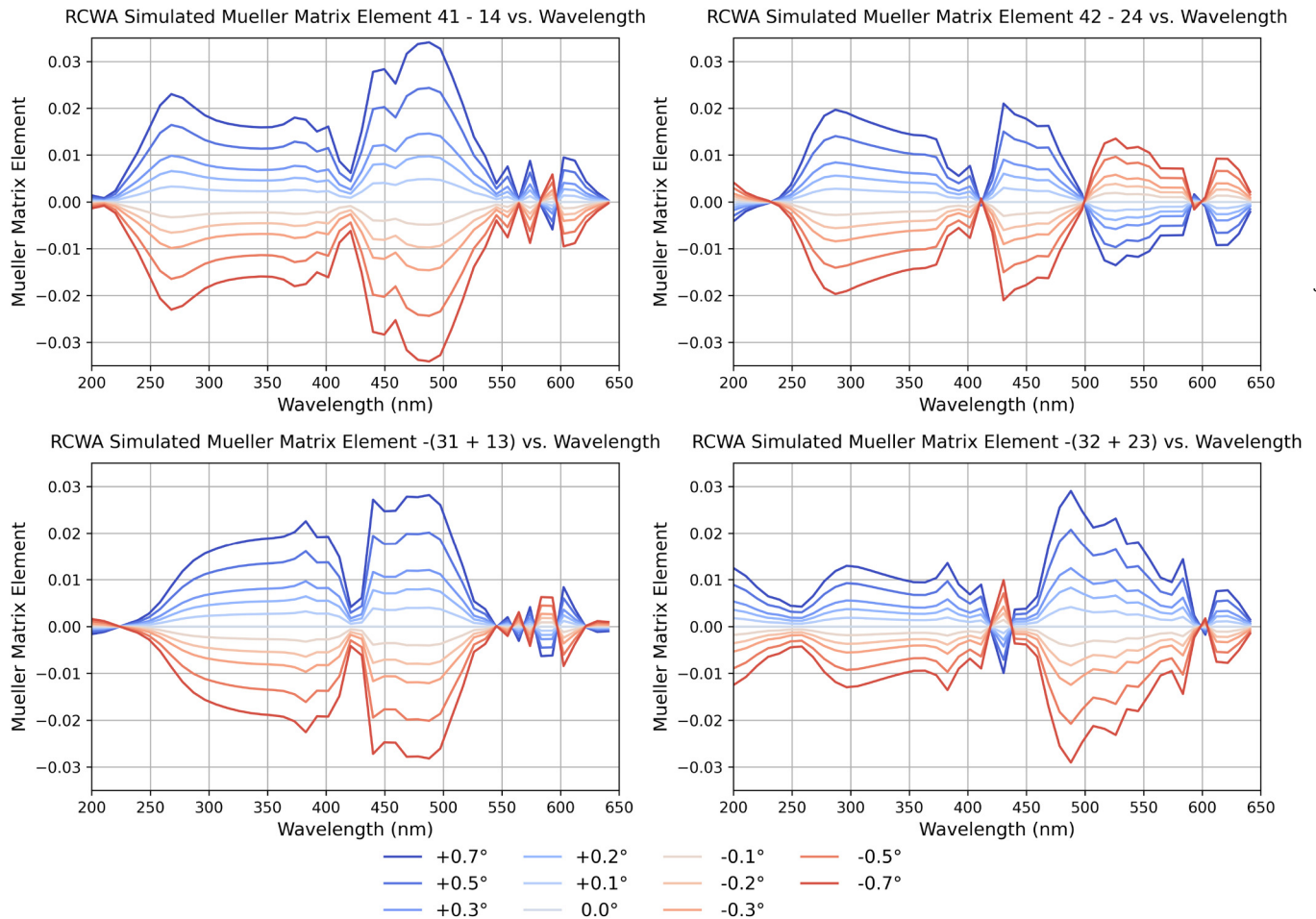
Again, N is the total number of elements summed over all wavelengths, m stands for measured, and p stands for predicted.

TABLE I. Table that summarizes the performance-critical parameters for HAR microstructures and their perturbation ranges. These parameters include the height and tilt angle of the grating bars and the coefficients of the Legendre polynomial parameterization of the trench critical dimension (CD). The input, x , to the Legendre polynomials is the height from the bottom of the trench, and the output is the width of the grating bars.

| Parameter | Description | Range (Min, Max) | Discretization |
|-----------|--|------------------|----------------|
| ht.3 | Height of the grating bar (nm) | (2500.0, 5000.0) | 1.0 |
| xtilt.4 | Bar tilt angle of the grating bar (degrees) | (-0.750, 0.750) | 0.01 |
| pw0.4 | $P_0(x)$: Constant Legendre coefficient of bar width | (50.0, 120.0) | 1.0 |
| pw1.4 | $P_1(x)$: Linear Legendre coefficient of bar width | (-75.0, 75.0) | 1.0 |
| pw2.4 | $P_2(x)$: Quadratic Legendre coefficient of bar width | (-100.0, 100.0) | 1.0 |
| pw3.4 | $P_3(x)$: Cubic Legendre coefficient of bar width | (-100.0, 100.0) | 1.0 |
| pw4.4 | $P_4(x)$: Quartic Legendre coefficient of bar width | (-100.0, 100.0) | 1.0 |

This loss function that weighs pairs of off-diagonal elements exaggerates the effect of the off-diagonals more than weighing the elements themselves, like in (9), because of the cross terms that appear in the quadratic.

Figure 5 illustrates our sensitivity analysis to tilt, where we vary tilt on the order of a few degrees and plot a linear combination of each of the upper-right and lower-left matrix element sums. The distinct change in the spectra as a function of bar tilt bolsters our



07 January 2025 16:12:47

FIG. 5. Mueller matrix elements simulated in the presence of a bar angle tilt (color coded in the range of -0.7° to $+0.7^\circ$) over the spectral range of 200–650 nm. Data were generated using the physical model illustrated in Fig. 4(c). Note the sensitivity of the off-diagonal element linear combinations to the tilt.

hypothesis that we should be able to determine the tilt of a grating bar given experimental spectra by matching it to one simulated by RCWA.

The physical model and the parameter selection help in reducing the complexity of the problem while retaining the essential characteristics of the structure. This confined parameter space allows for more efficient and accurate optimization.

IV. EXISTING PARAMETER DETERMINATION METHODS

Traditional methods for determining the parameters of HAR microelectronic and photonic structures from MMSE spectra often involve exhaustive grid searches. These methods explore the parameter space by computing the simulated spectra for every possible combination of parameters and comparing it to the experimental spectra.

All methods attempt to solve the optimization problem in (8) by finding a \mathbf{p}^* that minimizes the error, E , between the simulated and experimental spectra, over the space of \mathbf{p} , where E is defined as

$$E(\mathbf{p}) = \|\mathcal{M}_{\text{sim}}(\mathbf{p}) - \mathcal{M}_{\text{exp}}\|^2. \quad (11)$$

For instance, in a naïve grid search approach, the total computation time can be extremely high due to the exponential growth of possible parameter combinations with the number of parameters d' . If there are 10 grid points in each dimension and each RCWA computation takes 0.1 s, the total computation time for a parameter space of dimension $d' = 5$ is

$$\text{Time at Query} = \text{RCWA}_{\text{time}} \times (\text{gridpoints})^{d'}, \quad (12)$$

$$\begin{aligned} \text{Time at Query} &= 0.1 \times (10)^5 \text{ s} \\ &= 100\,000 \text{ s} \approx 28 \text{ h.} \end{aligned} \quad (13)$$

This demonstrates the impracticality of naïve grid search methods because of their computational intensity and time requirements.

The most popular alternative approach is the library method, which involves precomputing a lookup table of spectra for different parameter sets and storing this in a database. When a new experimental spectrum is measured, the closest matching precomputed spectrum is found using k-nearest neighbors (k-NNs) search.

The precomputation time for this method is similar to the grid search

$$\text{Precomputation Time} = \text{RCWA}_{\text{time}} \times (\text{gridpoints})^{d'}, \quad (14)$$

$$\begin{aligned} \text{Precomputation Time} &= 0.1 \times (10)^5 \\ &= 100\,000 \text{ s} \approx 28 \text{ h.} \end{aligned} \quad (15)$$

Once the library is built, the time to compute the square difference for a given experimental data point is significantly reduced.

TABLE II. Gradient descent for parameter optimization.

| |
|---|
| Require: Initial parameter estimates p_0 , learning rate η , stopping condition ε |
| 1: $k \leftarrow 0$ |
| 2: repeat |
| 3: Compute the gradient using (18): $\nabla E(\mathbf{p}_k)$ |
| 4: Update parameters: $\mathbf{p}_{k+1} \leftarrow \mathbf{p}_k - \eta \nabla E(\mathbf{p}_k)$ |
| 5: $k \leftarrow k + 1$ |
| 6: until $ \nabla E(\mathbf{p}_k) < \varepsilon$ |
| 7: $\mathbf{p}^* \leftarrow \mathbf{p}_k$ |
| 8: return \mathbf{p}^* |

Assuming the square difference calculation takes 0.01 s,

$$\begin{aligned} \text{Time at Query} &= 0.01 \times (10)^5 \\ &= 1000 \text{ s} \approx 17 \text{ min.} \end{aligned} \quad (16)$$

Another method is gradient descent using finite differences, which involves iteratively updating the parameter estimates to minimize the error between the simulated and experimental spectra. The update rule for gradient descent is given by

$$\mathbf{p}_{k+1} = \mathbf{p}_k - \eta \nabla E(\mathbf{p}_k), \quad (17)$$

where η is the learning rate and $\nabla E(\mathbf{p}_k)$ is the gradient of the error function. Once $\mathcal{M}_{\text{sim}}(\mathbf{p}_k)$ and \mathcal{M}_{exp} become close enough according to some stopping condition, indicated by the gradient, $\nabla E(\mathbf{p}_k)$, vanishing, \mathbf{p}_k is set to \mathbf{p}^* .

The gradient can be approximated using finite differences,

$$\frac{\partial E}{\partial p_i} \approx \frac{E(\mathbf{p} + \Delta p_i) - E(\mathbf{p})}{\Delta p_i}, \quad (18)$$

where i indexes the parameters we index over and ranges from 0 to $d' - 1$.

The computation time for each gradient step is

$$\begin{aligned} \text{Gradient Time per Step} &= \text{RCWA}_{\text{time}} \times d' \\ &= 0.1 \times 5 = 0.5 \text{ s.} \end{aligned} \quad (19)$$

For 100 iterations, the total computation time is

$$\text{Time at Query} = 0.1 \times 100 = 10 \text{ s.} \quad (20)$$

While gradient descent is faster than a full grid search, the finite difference approximation for gradients still requires multiple RCWA simulations per iteration, making it very computationally expensive (Table II).

V. OUR METHOD

We performed RCWA and machine learning analysis using proprietary software (NanoDiffract, Onto Innovation Inc., Wilmington, MA). While exact algorithm details are confidential,

the following is a description of what one could perform to achieve similar results. The key idea is to maintain the notion of gradient descending on input parameters but replace the naïvely slow gradient calculation with one that is faster by displacing RCWA.

At a high level, one could replace the computationally expensive RCWA simulations with an analytical form using trained neural networks. The neural network, once trained, can rapidly compute the Mueller matrix spectra and their gradients, enabling efficient optimization.

The gradient descent optimization process involves iteratively updating the parameter estimates to minimize the error between the simulated and experimental spectra. The update is the same as (17) but instead of expensively calculating the gradient with finite-difference, replacing RCWA with an analytical form allows us to calculate the gradient with the chain rule (which is made rapid with backpropagation)

$$\frac{\partial E}{\partial p_i} = \frac{\partial E}{\partial \mathcal{M}} \cdots \frac{\partial \mathcal{M}}{\partial p_i}. \quad (21)$$

To replace RCWA with an analytical form, we train a neural network to emulate the RCWA simulations. The neural network function $f_\omega(\mathbf{p})$ maps the input parameters \mathbf{p} to the Mueller matrix spectra \mathcal{M} ,

$$\mathcal{M} = f_\omega(\mathbf{p}). \quad (22)$$

Many flavors of a neural net will work, but we suspect a simple one will work well. The multilayer perceptron consists of multiple layers with weight matrices \mathbf{W} , which consist of unique ω 's and nonlinear activation functions σ ,

$$\mathcal{M} = f_\omega(\mathbf{p}) = \sigma(\mathbf{W}_L \sigma(\mathbf{W}_{L-1} \cdots \sigma(\mathbf{W}_1 \mathbf{p}) \cdots)). \quad (23)$$

To train the neural network, one can generate training data with an input-output form defined by (22) using RCWA simulations for various parameter sets. The loss function for training could be the mean squared error (MSE) between the RCWA-generated spectra $\mathcal{M}^{\text{RCWA}}$ and the neural network-predicted spectra \mathcal{M}^{NN} ,

$$\text{Loss} = \frac{1}{N} \sum_{n=1}^N (\mathcal{M}^{\text{RCWA}}(\mathbf{p}_n) - \mathcal{M}^{\text{NN}}(\mathbf{p}_n))^2, \quad (24)$$

where N is the number of training samples. The trained neural network can then be used to compute the spectra and their gradients efficiently.

The workflow for parameter estimation from experimental spectra using the trained neural network is summarized in Table III.

This method significantly reduces the computation time compared to traditional RCWA-based approaches, almost entirely because we replace the finite-difference gradient calculation with one that is analytical by approximating RCWA with a neural network, making it feasible for real-time parameter estimation and in-line process adjustments.

TABLE III. Neural network-driven nondestructive characterization.

| |
|--|
| Require: Physical model of sample, free parameter grid-spacing, initial parameter estimates \mathbf{p}_0 , learning rate η , stopping condition ϵ |
| 1: Capture experimental MMSE spectra from the sample |
| 2: Generate RCWA data using physical model and free parameters |
| 3: Train the neural network to approximate RCWA given this data |
| 4: Freeze the weights in the network |
| 5: $k \leftarrow 0$ |
| 6: repeat |
| 7: Compute the predicted spectra $\mathcal{M} = f_\omega(\mathbf{p}_k)$ using the neural network |
| 8: Compute the gradient: $\nabla E(\mathbf{p}_k) = \frac{\partial E}{\partial \mathcal{M}} \cdot \frac{\partial \mathcal{M}}{\partial \mathbf{p}_i}$ |
| 9: Update parameters: $\mathbf{p}_{k+1} \leftarrow \mathbf{p}_k - \eta \nabla E(\mathbf{p}_k)$ |
| 10: $k \leftarrow k + 1$ |
| 11: until $ \nabla E(\mathbf{p}_k) < \epsilon$ |
| 12: $\mathbf{p}^* \leftarrow \mathbf{p}_k$ |
| 13: return \mathbf{p}^* |

The precomputation time for our method involves generating RCWA data and training the neural network. Given RCWA data generation takes around 28 h like in (13), and neural network training takes approximately 6 h; the total pre-computation time is

$$\text{Precomputation Time} = 28 \text{ h} + 6 \text{ h} = 34 \text{ h}. \quad (25)$$

The time at query, which involves computing the predicted spectra and their gradients using the trained neural network, is significantly reduced. Assuming the gradient computation per step

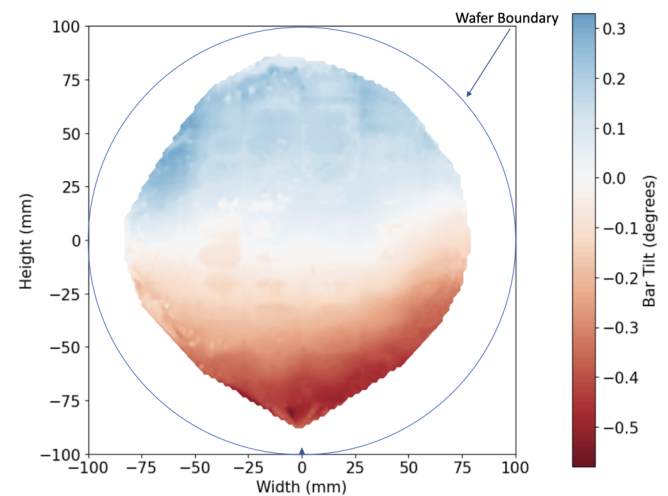


FIG. 6. Full wafer tilt map illustrating the variations in tilt across the entire wafer surface. We measured tilt at 30 208 different locations. Total time was 11 h and 26 min, yielding an average tilt measurement time of approximately 1 s. Blue regions (top of the map) indicate areas of positive tilt, while red regions (bottom of the map) indicate areas of negative tilt. The wafer boundary is labeled and points where the uncertainty in tilt is above a threshold are clipped.

07 January 2025 16:12:47

takes 0.05 s and 100 iterations are required, the total time at query is approximately:

$$\text{Time at Query} = 0.05 \times 100 = 5 \text{ s.} \quad (26)$$

VI. EXPERIMENTAL SETUP

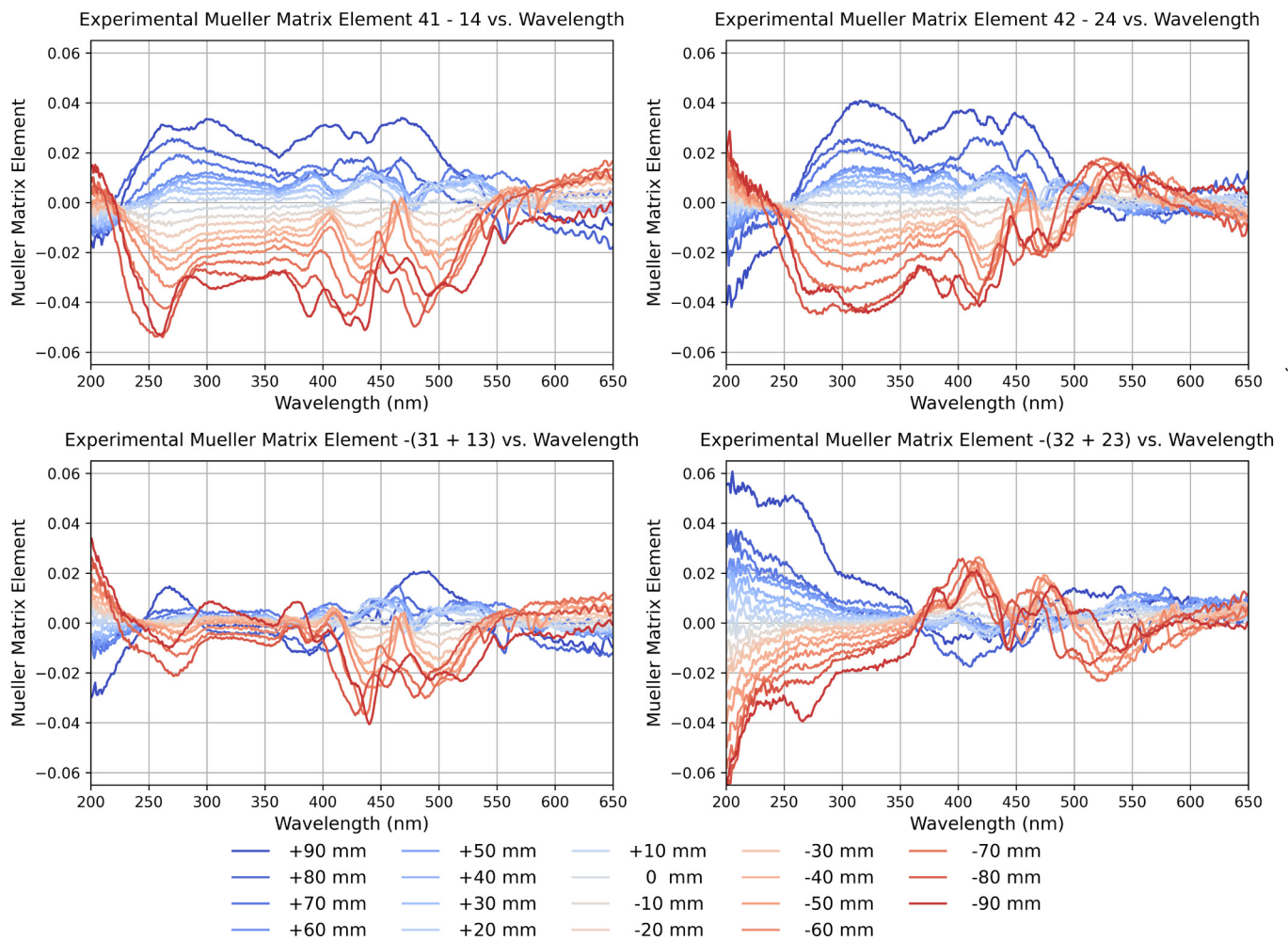
A. Setup for measuring bar tilt across wafers

To measure the bar tilt across wafers, we employ MMSE. The MMSE setup includes a light source, a polarizer, two dual-rotating compensators (one in each arm), a sample stage, an analyzer, and a detector. The light source generates a beam of known polarization, which passes through the polarizer and compensator before

interacting with the sample. The reflected light is then analyzed to determine the changes in its polarization state, providing detailed information about the sample's optical properties and structure. The setup is a commercial setup, specifically an Atlas V from Onto Innovation with RC2 ellipsometer integrated (from JA Woollam Company).

The MMSE spectra are captured over a range of wavelengths, from 200 to 650 nm, allowing us to construct the Mueller matrix $M(\lambda)$ for each measurement point on the wafer. The experimental setup is carefully calibrated to ensure accurate and repeatable measurements. The exact details of calibration are confidential, but we generally follow the methods detailed in Section five of Chen.¹⁵

The spot size of the light beam is kept small ($40 \times 40 \mu\text{m}$) to avoid mm-pitch Level 2 support structures [e.g., hexagonal support



07 January 2025 16:12:47

FIG. 7. Experimental Mueller matrix spectra along the center-line of the wafer. We specifically focus on the difference between the upper-right and lower-left quadrant matrix elements. We find that the spectra show variation across the wafer, corroborating our hypothesis that small structure variations lead to large spectra variations that we can successfully fit with machine learning. Each curve is labeled by its measurement point's vertical distance from the center of the wafer (see Fig. 8, vertical path starting at wafer notch).

structures in Fig. 4(b)]. The exact details of how this spot size is achieved are confidential, but we use custom refractive, compound lenses with multiple elements, like shaping components, that keep the spot size circular and uniform over multiple wavelengths. We use a conical geometry, where the incident light beam is parallel to the CAT grating bars with a 65° angle of incidence relative to the surface normal.

B. Data collection methods and validation using small-angle x-ray scattering

To validate the MMSE measurements, we use small-angle x-ray scattering (SAXS), a well-established technique for characterizing nanostructures. SAXS, in principle, can provide high-resolution data on the grating profile, including the bar tilt and periodicity, by analyzing the scattering patterns of x rays as they interact with the sample.

We follow the method described by Song,⁹ which involves using a collimated x-ray beam directed at the sample. The scattered x rays are detected at small angles relative to the incident beam, and the diffracted orders are analyzed as a function of incidence angle to extract bar tilt. The SAXS data serve as an accurate benchmark for validating the MMSE measurements.

By comparing the tilt angles obtained from MMSE and SAXS, we can assess the consistency and precision of our MMSE-based

characterization method. The combination of MMSE and SAXS provides a comprehensive approach for measuring and validating bar tilt across HAR photonic wafers, offering both nondestructive and high-resolution capabilities.

VII. RESULTS AND DISCUSSION

The contour plot in Fig. 6 shows the bar tilt determined by MMSE across different points on the wafer. We show that we are able to rapidly extract tilt measurements from any point on the wafer.

To validate the accuracy of these measurements, we extract measurements along a line on the wafer perpendicular to the grating bars to compare with SAXS data [Fig. 8(a)]. Collecting SAXS data is time consuming; so we do not compare SAXS data to every point on the wafer, but instead, compare data along the entire vertical length of the wafer to determine if we see alignment at both small and large bar tilts. We follow the methods described by Song,⁹ and plot the bar tilt determined by both MMSE and SAXS [Fig. 8(b)]. The close agreement between the two sets of measurements validates the accuracy of our method.

Figure 8 illustrates where exactly on the full wafer from Fig. 6 the points for SAXS were extracted. This visually shows our machine learning method's ability to measure grating bar tilt to the accuracy of SAXS, across the length of the wafer.

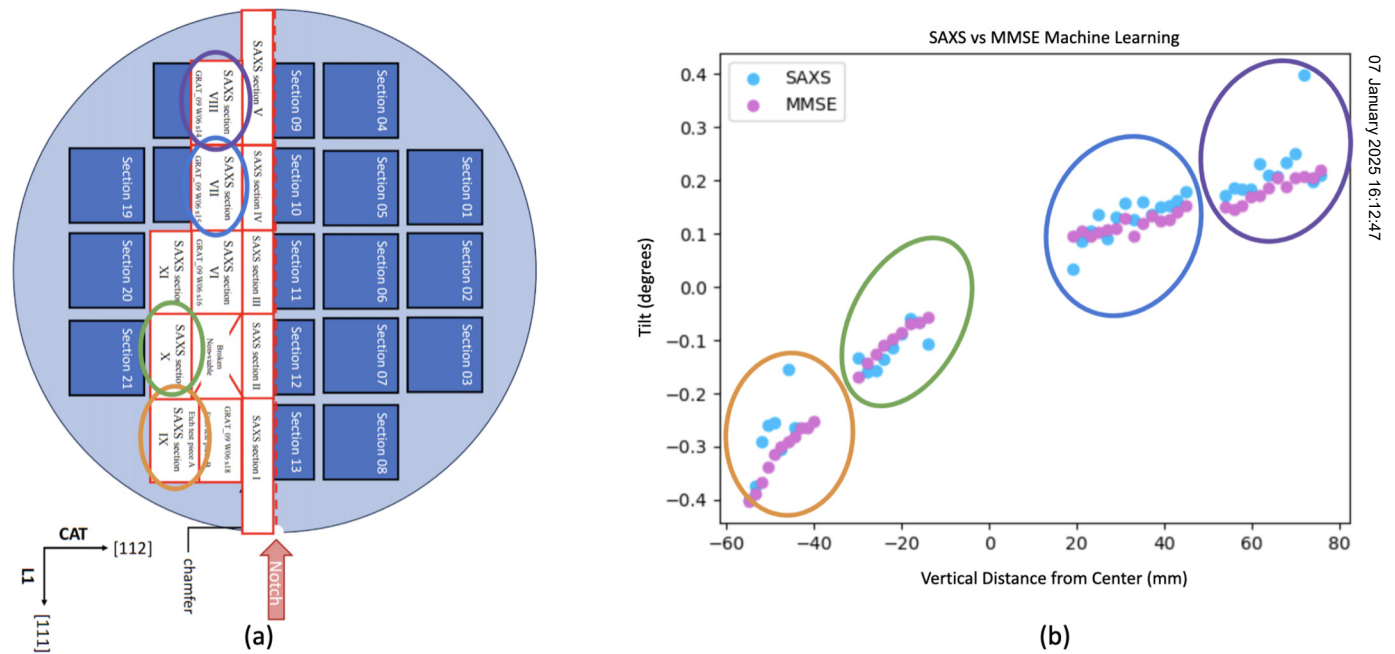


FIG. 8. Illustration of where on the wafer points were extracted for SAXS measurement, with a comparison of tilt measurements between SAXS and MMSE. (a) CAT grating wafer with sections labeled. MMSE data were taken on the entire wafer, before destructive cleaving for SAXS. Color coded circles correspond to sections used for SAXS measurement. Note that sections span the vertical length of the wafer. (b) SAXS vs MMSE machine learning tilt measurements with sections labeled by their corresponding color coded sections [circles at the bottom of the wafer in (a) correspond to circles at the left of the plot in (b), while those at the top of the wafer in (a) correspond to those at the right of the plot in (b)]. The plot shows tilt (in degrees) as a function of vertical distance from the center of the wafer (mm). SAXS measurements are represented by blue dots (those with higher variance), and MMSE measurements are represented by pink dots (those with lower variance), demonstrating the consistency and accuracy of MMSE in capturing tilt variations. It is clear that alignment is maintained across the wafer, at various degrees of both positive and negative bar tilt.

07 January 2025 16:12:47

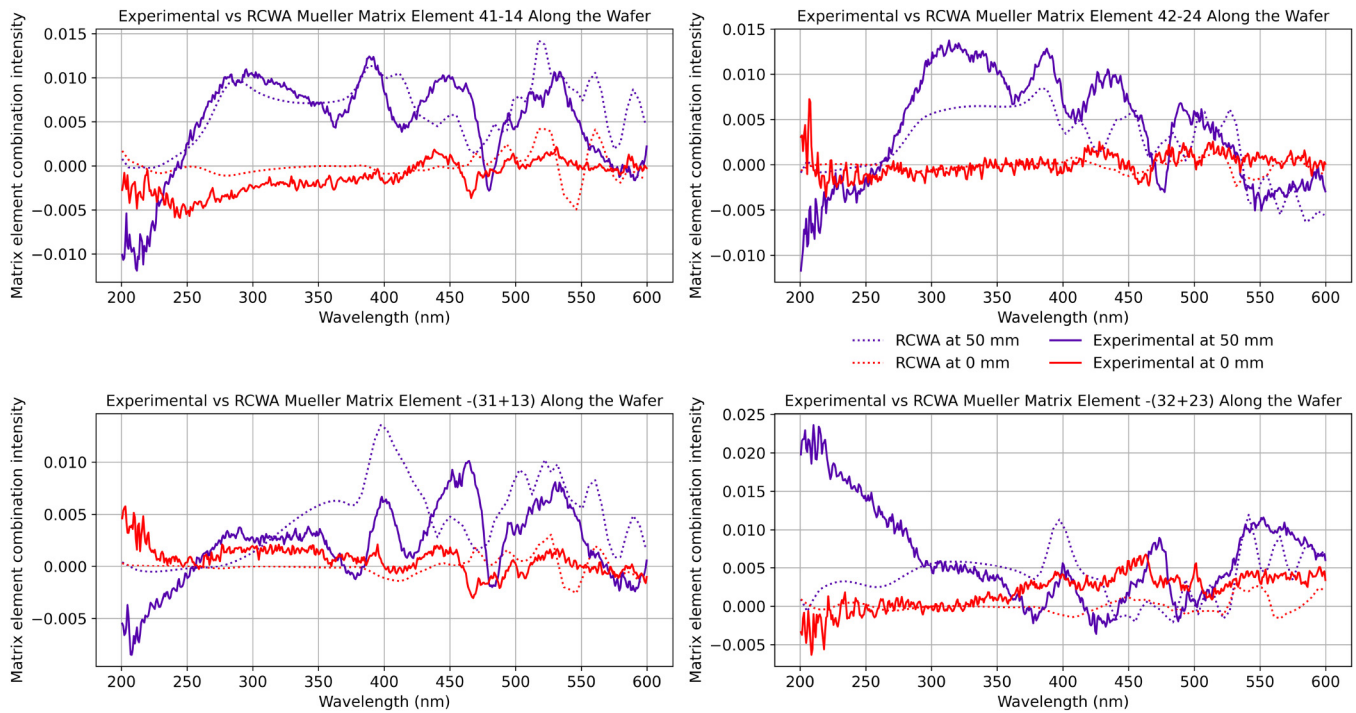


FIG. 9. Comparison between experimental spectra and RCWA emulating neural network convergence after running gradient descent optimization. The red curves (the lighter pair) represent a point at the center of the wafer and the purple curves (the darker pair) represent a point at 50 mm north of the center of the wafer. The solid curves are the experimental spectra and the dotted curves are those after convergence.

Our method essentially moves through critical-parameter-space to attempt to fit the curves in Fig. 5 to those in Fig. 7. This fitting is sped up by replacing the brute-force RCWA calculations with a neural network. Because the neural network is an analytical function, the gradient can be calculated rapidly with backpropagation in the critical-parameter-space after the weights are frozen. Figure 5 illustrates RCWA-simulated Mueller matrix spectra for upper-right and lower-left quadrant matrix element pairs. These are those that are most sensitive to asymmetry, and they gave us a first proof-of-concept that bar tilt could be accurately measured across a wafer through its effect on these Mueller matrix spectra. Figure 7 illustrates real, experimental spectra across a wafer. The variation in these spectra across the wafer provided validation that we could perform a fitting procedure that would attempt to match the spectra from Fig. 5 (or, neural network approximations to the spectra from Fig. 5 for increased speed) with the spectra from Fig. 7 to effectively measure bar tilt. Interestingly, the curves in Fig. 5 do not match those in Fig. 7 with extreme accuracy, but our method still works well in recovering performance-critical parameters (Fig. 8).

To gain further insight, we plot the experimental spectra against our model's spectra after convergence for two points along the wafer in Fig. 9. This elucidates how close our machine learning approximation of RCWA is able to recover the experimental data. Note again that the two pairs of curves are not exactly the same,

but the accuracy of our bar tilt measurements is still high. This implies that not all the variation in the experimental spectra is needed to determine the bar tilt. The approximations made by both RCWA and our neural net are enough to capture the bar tilt variance across the wafer. We, therefore, expect that a similar gradient descent algorithm that operates in a basis in which the curves are sparse, like the Fourier or Wavelet basis, would be sufficient.

For future work, we imagine we could Fourier transform the experimental spectra and pass them through a low-pass filter to remove the high frequency components. We can also try passing them iteratively through different band-pass filters to remove certain frequency components. We can then inverse Fourier transform back to the canonical basis, and see if our method is still able to recover the bar tilt with similar accuracy. This may further validate the hypothesis that only certain frequencies or components of information in the experimental spectra are needed to recover certain critical parameters. Determining which parts of the spectra are most important for measuring different structure parameters may help guide measurement techniques or machine learning architectures for rapid measurement.

Our machine learning approach, using a trained neural network to replace RCWA simulations, is not only accurate, but also significantly improves the speed of the gradient computation. Gradient computation is the bottleneck in the gradient descent fitting procedure; so replacing RCWA with an analytical form through a neural

TABLE IV. Table that summarizes notional computation time estimates of each of the prior methods compared to our method. These are back of the envelope calculations that will yield computation times with similar order of magnitudes on other systems. The time at query is the most important because it sets the limit on potential control algorithm accuracy.

| Method | Time scaling equation | Precomputation time | Time at query |
|-----------------------------|--|--|---------------|
| Our method | $\frac{1}{N} \sum_{n=1}^N (\mathcal{M}^{\text{RCWA}}(\mathbf{p}_n) - \mathcal{M}^{\text{NN}}(\mathbf{p}_n))^2$ | 30 h | 5 s |
| RCWA with finite difference | $0.1 \times d^d$ | $\text{RCWA}_{\text{time}} \times 100$ | 10 s |
| Library search | $0.1 \times 10^{d^d}$ | 28 h | 17 min |
| Naive grid search | $0.1 \times 10^{d^d}$ | 0 | 28 h |

network, and still being able to measure critical parameters by freezing network weights, is the key addition of our paper.

As shown in Table IV, which summarizes the approximate precomputation times and query times for all methods including ours, the gradient computation time per step using the neural network is reduced to 0.05 s, making the total computation time for 100 iterations approximately 5 s. This is a substantial improvement over traditional methods, which can take several minutes or up to a day. The accuracy of the machine learning approach is validated by its close agreement with the SAXS-determined tilt. The current approach focuses on a reduced parameter space to ensure tractability. However, our method can be extended to higher-dimensional parameter spaces. Future work can explore including multiple parameters like sidewall roughness and material composition. It could also be used as a sensor in a control algorithm to adjust fabrication parameters in real-time to nudge these performance-critical structure parameters to their optimal values.

VIII. CONCLUSION

Our method is not only accurate in determining bar-tilt, but also fast. By replacing the standard finite-difference gradient calculation with one that is analytical, we are able to speed up the traditional bottleneck in RCWA-based approaches to measurement. The main addition of our paper is freezing the network weights and using gradient descent on the input space after training the network, along with experimental validation. While standard methods using Mueller matrix ellipsometry surpass SAXS in their non-destructiveness and speed, they are still too slow for potential feedback control of fabrication parameters because they rely on brute-force RCWA for their fitting procedures. This paper could open the door to the development of robust control algorithms that adjust fabrication parameters in response to measurement. Control algorithms need fast feedback sensors; otherwise, the latency between measurement and reality is too high for accurate control of critical-structure parameters. We foresee future work using our method to develop such control algorithms to not only monitor but also control fabrication processes in close to real time.

ACKNOWLEDGMENTS

The authors would like to thank Charlie Settens and Jordan Cox for their help troubleshooting SAXS instrument data collection. They also thank Mariel Shapiro for providing the SEM image in Fig. 4(a) from a separate wafer. The authors are grateful to Matthew Heine and the IS&T department at MIT for their help

navigating computing resources. Finally, the authors would like to thank Mallory Whalen, Jungki Song, Bethany Levenson, James Jusuf, Tristen Wallace, Paran Culanathan, Varan Culanathan, Spencer Schneider, Richard Bao, Anish Mudide, Emma Batson, Mark Mondol, Anjelica Molnar-Fenton, and C. J. Johnson for their helpful discussions. This work was performed, in part, in the MIT.nano Characterization Facilities and supported by NASA Grant No. 80NSSC22K1904.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Shiva Mudide: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Supervision (equal); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Nick Keller:** Formal analysis (lead); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (lead). **G. Andrew Antonelli:** Conceptualization (lead); Funding acquisition (lead); Methodology (lead); Project administration (lead); Software (equal); Supervision (equal). **Geraldina Cruz:** Data curation (equal). **Julia Hart:** Data curation (equal); Investigation (equal). **Alexander R. Bruccoleri:** Conceptualization (equal); Data curation (equal); Funding acquisition (equal); Methodology (equal). **Ralf K. Heilmann:** Conceptualization (lead); Funding acquisition (lead); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Writing – review & editing (lead). **Mark L. Schattenburg:** Conceptualization (lead); Funding acquisition (lead); Methodology (equal); Project administration (lead); Resources (equal); Supervision (lead); Writing – review & editing (lead).

DATA AVAILABILITY

There are two datasets used in this paper. First is the RCWA simulation data used to train our neural network. The second is the raw experimental Mueller Matrix spectral data that we captured from our wafer. The former is available from Onto Innovation. Restrictions apply to the availability of this dataset, which were used under license for this study. Data are available from the authors upon reasonable request and with the permission of Onto

07 January 2025 16:12:47

Innovation. The latter dataset is available from the corresponding author upon reasonable request.

REFERENCES

- ¹Y. S. Shin, J. M. Park, J. Lee, and S. M. Kim, *J. Micromech. Microeng.* **19**, 065001 (2009).
- ²M. Varvara, R. Barnett, F. Avril, and P. Bennett, "A simplified test vehicle for understanding and improving tilt and its impact on the performance of inertial sensors," *2015 Transducers—2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers)*, Anchorage, AK (IEEE, New York, NY, 2015), pp. 1172–1174. doi:10.1109/TRANSDUCERS.2015.7181137.
- ³S. Hussain, A. Kumar, and S. Chakraborty, *Meas. Sci. Technol.* **25**, 082001 (2014).
- ⁴R. K. Heilmann, M. Ahn, E. M. Gullikson, and M. L. Schattenburg, *Opt. Express* **16**, 8658 (2008).
- ⁵R. K. Heilmann, M. Ahn, A. Bruccoleri, C.-H. Chang, E. M. Gullikson, P. Mukherjee, and M. L. Schattenburg, *Appl. Opt.* **50**, 1364 (2011).
- ⁶R. K. Heilmann, J. Kolodziejczak, A. R. Bruccoleri, J. A. Gaskin, and M. L. Schattenburg, *Appl. Opt.* **58**, 1223 (2019).
- ⁷R. K. Heilmann *et al.*, *Astrophys. J.* **934**, 171 (2022).
- ⁸D. Thomas, M. Muggeridge, J. Hopkins, N. Launay, H. Ashraf, and T. Barrass, *ECS Trans.* **72**, 9 (2016).
- ⁹J. Song, R. K. Heilmann, A. R. Bruccoleri, and M. L. Schattenburg, *J. Vac. Sci. Technol. B* **37**, 062917 (2019).
- ¹⁰R. K. Heilmann, A. R. Bruccoleri, E. M. Gullikson, R. K. Smith, and M. L. Schattenburg, "Soft x-ray performance and fabrication of flight-like blazed transmission gratings for the x-ray spectrometer on Arcus probe," *Proc. SPIE* **12679**, 126790L (2023).
- ¹¹A. Bruccoleri, D. Guan, P. Mukherjee, R. K. Heilmann, M. L. Schattenburg, and S. Vargo, *J. Vac. Sci. Technol. B* **31**, 06FF02 (2013).
- ¹²A. R. Bruccoleri, R. K. Heilmann, and M. L. Schattenburg, *J. Vac. Sci. Technol. B* **34**, 06KD02 (2016).
- ¹³R. K. Heilmann *et al.*, "Toward volume manufacturing of high-performance soft x-ray critical-angle transmission gratings," *Proc. SPIE* **11444**, 114441H (2021).
- ¹⁴M. G. Moharam and T. K. Gaylord, *J. Opt. Soc. Am.* **71**, 811 (1981).
- ¹⁵C. Chen, I. An, G. M. Ferreira, N. J. Podraza, J. A. Zapien, and R. W. Collins, *Thin Solid Films* **455–456**, 14 (2004).