Research Article

applied optics

Interfacial surface roughness determination by coherence scanning interferometry using noise compensation

HIROKAZU YOSHINO,[1,2,*] JOHN MICHAEL WALLS,[1] AND ROGER SMITH[1]

[1]Loughborough University, Leicestershire LE11 3TU, UK
[2]Taylor Hobson Ltd., Leicestershire LE4 9JD, UK
*Corresponding author: hirokazu.yoshino@ametek.com

The capability of coherence scanning interferometry has been extended recently to include the determination of the interfacial surface roughness between a thin film and a substrate when the surface perturbations are less than ~10 nm in magnitude. The technique relies on introducing a first-order approximation to the helical complex field (HCF) function. This approximation of the HCF function enables a least-squares optimization to be carried out in every pixel of the scanned area to determine the heights of the substrate and/or the film layers in a multi-layer stack. The method is fast but its implementation assumes that the noise variance in the frequency domain is statistically the same over the scanned area of the sample. This results in reconstructed surfaces that contain statistical fluctuations. In this paper we present an alternative least-squares optimization method, which takes into account the distribution of the noise variance-covariance in the frequency domain. The method is tested using results from a simulator and these show a significant improvement in the quality of the reconstructed surfaces.

1. INTRODUCTION

Surface metrology of form, flatness, roughness, and smoothness are important for quality assurance in many types of manufacturing. It is a particular issue for optical components or optical coatings where control of features in the nanometer or sub-nanometer range is required. Stylus-based surface profilometry is the conventional technique used to provide two-dimensional surface metrology. However, this technique can cause modification of the surface under measurement. *Coherence scanning interferometry* (CSI), previously referred to as *scanning white light interferometry*, is a well established, non-contact method that provides true three-dimensional measurements [1]. Optical microscopy provides lateral images without height information. *Scattermeters* measure the proportion of specular to diffuse reflection to calculate the root mean square roughness [2]. In comparison, CSI measures absolute heights at each pixel in the field of view with sub-nanometer vertical resolution [3]. This allows all the various surface roughness parameters measured using stylus profilometry to be computed using CSI [4].

CSI measures surface topography by locating and connecting peak positions in the interference pattern, referred to as an *interferogram*, at each pixel over the scanned area to reconstruct the measured surface. One of the prerequisites for the test surface is that the surface should have identical amplitude reflection coefficients over the field of view; otherwise, a phase shift on reflection occurs, which results in an erroneous vertical profile. Even if the refractive index is unchanged over the measurement area, problems can occur in the CSI measurement for transparent/semi-transparent thin films with thicknesses of $< \sim 1.5$ μm. This is because the interferogram has multiple peaks corresponding to the interfacial surfaces in the thin film assembly. The peaks may be superimposed depending on the thin film structure. In the case where the films are $> \sim 1.5$ μm, it is possible to detect and separate the peaks to reproduce the interfacial topographies [5–7]. However, this is not the case with thin films of a few hundred nanometers in thickness. A Fourier transform of the interferogram can be performed for investigations in this thickness regime. Subtle changes in phase and amplitude are compared with those synthesized mathematically [8–11].

One of the methods defined in the frequency domain approach uses a theory based on the *helical complex field* (HCF)

function [12,13] and its extensions [14,15]. *Interfacial surface roughness* (ISR) has been determined using this method by introducing a first-order approximation to the HCF function, which enables fast real-time computation [3,16]. However, the method can introduce spurious surface roughness, as shown in a previous study [15].

This paper presents a methodology to reduce the spurious roughness which can occur with the existing ISR method and improves its computation stability for a wide variety of samples with no additional hardware, no changes in measurement procedure, and little extra computational effort.

2. THEORY

A. Helical Complex Field Function

The determined and synthesized HCF functions, which have a reflection coefficient averaged over the incident angle of the light of amplitude $\bar{r}$ [15], are expressed as follows:

$$\mathrm{HCF}^d(\nu; \mathbf{d}) = \bar{r}_{\mathrm{ref}}(\nu) \cdot \frac{\mathcal{F}[I]_{SB+}}{\mathcal{F}[I_{\mathrm{ref}}]_{SB+}},$$

$$\mathrm{HCF}^s(\nu, \mathbf{d}) = \bar{r}(\nu, \mathbf{d}) \cdot \exp(j4\pi\nu\Delta z_{\mathrm{HCF}} \cos\bar{\theta}), \quad \textbf{(1)}$$

where the thin film assembly consists of a substrate represented by its subscript sub and the thickness of the L film layers expressed by $\mathbf{d} = [d_{\mathrm{sub}}, d_1, ..., d_L]^{\mathsf{T}}$. Note that d_{sub} is not used in Eq. (1) but Δd_{sub} is considered in the following discussion. The determined HCF function, HCF^d, is given by the positive sideband of the Fourier transform of the actual interference signal I obtained from a test sample divided by that of a known reference material I_{ref}. Whereas the synthesized HCF function, HCF^s, is derived from a mathematical model of the test thin film structure [12]. Note that given any signal s, the positive sideband of the Fourier transform of the signal is termed $\mathcal{F}[s]_{SB+}$. The averaged incident angle of the CSI instrument is denoted by $\bar{\theta}$ and ν represents the light frequency. The unknown parameter Δz_{HCF} is associated with the surface height [15].

Assuming that the thin films are completely flat over the field of view, the set of film thicknesses is numerically determined to be $\mathbf{d} = \hat{\mathbf{d}}$ together with Δz_{HCF} by minimizing the squared error between the two functions in Eq. (1). Normally the interference signals over M test sample pixels and M_{ref} reference sample pixels, typically a few hundred, are averaged to have each signal with less noise; i.e., $I = 1/M\sum_i I^i$ is effectively used in Eq. (1) and so is $I_{\mathrm{ref}} = 1/M_{\mathrm{ref}}\sum_i I_{\mathrm{ref}}^i$.

B. First-Order Approximation to the Synthesized HCF Function

Let ϵ_{px} be the noise induced in the interference signal at any pixel I_{px} with its interfacial surface perturbation $\Delta\mathbf{d}$, then the determined and the synthesized HCF functions are, respectively,

$$\mathrm{HCF}^d_{\mathrm{px}} = \bar{r}_{\mathrm{ref}}(\nu) \cdot \frac{\mathcal{F}[I_{\mathrm{px}} + \epsilon_{\mathrm{px}}]_{SB+}}{\mathcal{F}[I_{\mathrm{ref}}]_{SB+}},$$

$$\mathrm{HCF}^s_{\mathrm{px}}(\nu; \hat{\mathbf{d}} + \Delta\mathbf{d}) \approx \overline{\mathrm{HCF}^d} + j4\pi\nu \cos\bar{\theta} \cdot \overline{\mathrm{HCF}^d}$$

$$\times \left\{ \Delta d_{\mathrm{sub}} + \sum_{l=1}^{L} G_l(\nu; \hat{\mathbf{d}})\Delta d_l \right\},$$
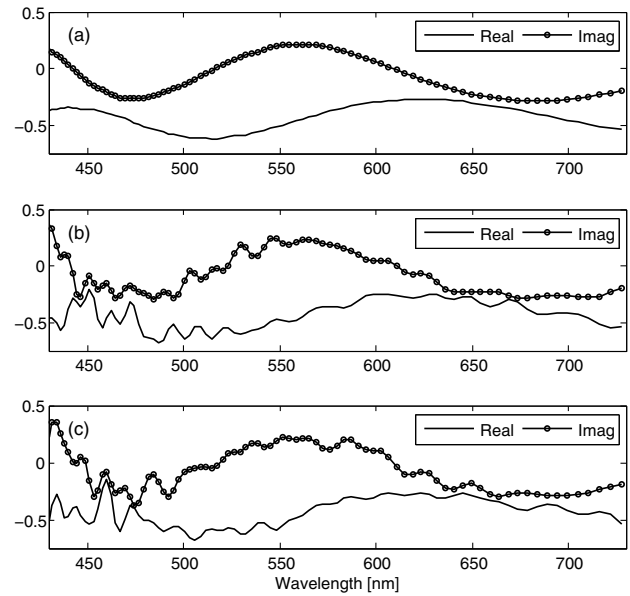

Fig. 1. Determined HCF function of a 520 nm SiO_2 thin film on a Si substrate: (a) The global determined HCF function $\overline{\mathbf{HCF^d}}$, obtained from the full 21×21 matrix of four pixels; (b) the HCF function $\mathbf{HCF^d_{px}}$ determined from four pixels at the edge of the measurement area; (c) the locally determined HCF function $\mathbf{HCF^d_{px}}$ at the center of the measurement area.

where

$$G_l(\nu; \hat{\mathbf{d}}) = 1 + \frac{1}{4\pi\nu\,\cos\bar{\theta}}\frac{\partial\chi(\hat{\mathbf{d}})}{\partial d_l}, \quad \arg(\bar{r}) = \chi,$$

$$\overline{\mathrm{HCF}^d} = \bar{r}_{\mathrm{ref}}(\nu) \cdot \frac{\mathcal{F}[E[I_{\mathrm{px}} + \epsilon_{\mathrm{px}}]]_{SB+}}{\mathcal{F}[I_{\mathrm{ref}}]_{SB+}}. \quad \textbf{(2)}$$

Note that the condition $E[I_{\mathrm{px}} + \epsilon_{\mathrm{px}}] = E[I_{\mathrm{px}}]$ holds. It follows that $\overline{\mathrm{HCF}^d}$ is smooth over the wavelength range of interest, as shown in Fig. 1(a).

For the computations based on Eq. (2), the expressions for the HCF functions need to be re-written in a spectrally discrete manner ($\boldsymbol{\nu} = [\nu_1, \nu_2, ..., \nu_m]^{\mathsf{T}}$) as follows:

$$\mathbf{HCF^d_{px}} = [\mathrm{HCF}^d_{\mathrm{px}}(\nu_1), \mathrm{HCF}^d_{\mathrm{px}}(\nu_2), ..., \mathrm{HCF}^d_{\mathrm{px}}(\nu_m)]^{\mathsf{T}},$$

$$\mathbf{HCF^s_{px}} \approx \overline{\mathbf{HCF^d}} + \mathrm{Diag}[\overline{\mathbf{HCF^d}}]\mathbf{G}\Delta\mathbf{d}. \quad \textbf{(3)}$$

Accordingly the expression is re-written as a linear inverse problem with noise $\boldsymbol{\varepsilon}_o$ in the frequency domain:

$$\mathbf{HCF^d_{px}} \simeq \overline{\mathbf{HCF^d}} + \mathrm{Diag}[\overline{\mathbf{HCF^d}}]\mathbf{G}\Delta\mathbf{d} + \boldsymbol{\varepsilon}_o,$$

where

$$\mathbf{G} = j4\pi \cos\overline{\theta} \begin{bmatrix} \nu_1 & \nu_1 \cdot G_1(\nu_1) & \cdots & \nu_1 \cdot G_L(\nu_1) \\ \nu_2 & \nu_2 \cdot G_1(\nu_2) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \nu_m & \nu_m \cdot G_1(\nu_m) & \cdots & \nu_m \cdot G_L(\nu_m) \end{bmatrix},$$

$$\mathrm{Diag}[\overline{\mathbf{HCF^d}}] = \begin{bmatrix} \overline{\mathrm{HCF}^d}(\nu_1) & 0 & \cdots & 0 \\ 0 & \overline{\mathrm{HCF}^d}(\nu_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \overline{\mathrm{HCF}^d}(\nu_m) \end{bmatrix},$$

$$\overline{\mathbf{HCF^d}} = [\overline{\mathrm{HCF}^d}(\nu_1), \overline{\mathrm{HCF}^d}(\nu_2), \ldots, \overline{\mathrm{HCF}^d}(\nu_m)]^\mathsf{T},$$

$$\boldsymbol{\varepsilon}_o = [\varepsilon_{o1}, \varepsilon_{o2}, \ldots, \varepsilon_{om}]^\mathsf{T} \sim N(\boldsymbol{\varepsilon}_o|\mathbf{0}, \sigma_o^2 \mathbf{I}). \quad (4)$$

Note that the noise ε_{oi} existing in the frequency domain is assumed to follow a normal distribution $N(\varepsilon_{oi}|0, \sigma_o^2)$ with the mean at zero and variance σ_o for $i \in \mathbb{N}$. Given a *random vector* $\mathbf{a}$, the operators $E[\mathbf{a}]$ and $E[(\mathbf{a} - E[\mathbf{a}])(\mathbf{a} - E[\mathbf{a}])^\mathsf{T}]$ are understood to be the *ensemble* and the *variance-covariance matrix* of the random vector respectively. The variance-covariance matrix of the noise $\boldsymbol{\varepsilon}_o$ is assumed to be $\sigma_o^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

C. Interfacial Surface Roughness Determination by the Present ISR Method

Using the discrete expressions introduced in the previous section, a merit function $J_{\mathrm{px}} = \|\mathbf{HCF^d_{px}} - \mathbf{HCF^s_{px}}\|^2$ is minimized for every pixel with respect to $\Delta\mathbf{d}$ such that $\boldsymbol{\varepsilon}_o \sim N(\boldsymbol{\varepsilon}_o|\mathbf{0}, \sigma_o^2 \mathbf{I})$. The solution $\hat{\Delta}\mathbf{d}$ is equivalent to that established by the *maximum likelihood estimation* under the assumption that each element of the noise $\boldsymbol{\varepsilon}_o$ is stochastically independent and has the same variance, i.e., the relation $E[(\boldsymbol{\varepsilon}_o - E[\boldsymbol{\varepsilon}_o])(\boldsymbol{\varepsilon}_o - E[\boldsymbol{\varepsilon}_o])^\mathsf{T}] = \sigma_o^2 \mathbf{I}$ holds. Then, the solution of the linear inverse problem in Eq. (4) is given analytically by [15,17,18]

$$\hat{\Delta}\mathbf{d} = (\mathbf{G}^\mathsf{T}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{u}, \quad (5)$$

where

$$\mathbf{u} = \{\mathrm{Diag}[\overline{\mathbf{HCF^d}}]\}^{-1}[\mathbf{HCF^d_{px}} - \overline{\mathbf{HCF^d}}].$$

As expressed in Eq. (5), the vector $\mathbf{u}$ is an observed signal while $\Delta\mathbf{d}$ is an unknown original signal to be estimated. The problem can be re-expressed as

$$\mathbf{u} = \mathbf{G}\Delta\mathbf{d} + \boldsymbol{\varepsilon}, \quad \text{where } \boldsymbol{\varepsilon} = \{\mathrm{Diag}[\overline{\mathbf{HCF^d}}]\}^{-1}\boldsymbol{\varepsilon}_o, \quad (6)$$

where the noise $\boldsymbol{\varepsilon}$ and its variance σ^2 have been re-defined for simplification. Accordingly, the problem and the merit function are simplified such that $\min J^\dagger_{\mathrm{px}} = \|\mathbf{u} - \mathbf{G}\Delta\mathbf{d}\|^2$ subject to $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\varepsilon}|\mathbf{0}, \sigma^2 \mathbf{I})$.

D. ISR Methodology with Noise Compensation

Although the existing ISR method [15,16,19] has been used to determine the surface topography of a layer buried under a transparent thin oxide film, the method can induce spurious roughness caused by system and environmental noise in the signal. This is illustrated in Fig. 1 for data taken from a

520 nm SiO_2 thin film. If the optimization process averages over the full scanned area as in Fig. 1(a), the result is a smoothly varying HCF function. However, if the HCF function is determined locally, as shown in Figs. 1(b) and 1(c), there is a different functionality.

Consider that we have a variance-covariance matrix Σ for the noise, then the optimal solution given by the least-squares error method in Eq. (5) is not valid. This is because the probability distribution of an observed signal $\mathbf{u}$ now follows $N(\mathbf{u}|\mathbf{G}\Delta\mathbf{d}, \Sigma)$. It follows that we need to modify the merit function to deal with this different probability distribution, otherwise the least-squares method would lead to an erroneous solution or become unstable. Without the modification, the ISR method measures a higher interfacial surface roughness.

1. ISR with Noise Compensation (ISR-NC) Determination

Let $p(\mathbf{u})$ be the probability density function (PDF) of the observed signal $\mathbf{u}$, then using the assumption that the spectral noise $\boldsymbol{\varepsilon}$ follows the normal distribution $N(\boldsymbol{\varepsilon}|\mathbf{0}, \Sigma)$, the PDF also is a normal distribution. Therefore, the PDF and its log likelihood function $L(\Delta\mathbf{d})$ are expressed as follows [18]:

$$p(\mathbf{u}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{u} - \mathbf{G}\Delta\mathbf{d})^\mathsf{T}\Sigma^{-1}(\mathbf{u} - \mathbf{G}\Delta\mathbf{d})\right],$$

$$L(\Delta\mathbf{d}) = \log p(\mathbf{u}) = -\frac{1}{2}(\mathbf{u} - \mathbf{G}\Delta\mathbf{d})^\mathsf{T}\Sigma^{-1}(\mathbf{u} - \mathbf{G}\Delta\mathbf{d}) + C, \quad (7)$$

where C is a constant independent of $\mathbf{u}$ and $\Delta\mathbf{d}$. Maximization of the log-likelihood function with respect to $\Delta\mathbf{d}$ is equivalent to minimization of the following merit function $J^\ddagger_{\mathrm{px}}$:

$$\underset{\Delta\mathbf{d}}{\text{minimize}} J^\ddagger_{\mathrm{px}} = (\mathbf{u} - \mathbf{G}\Delta\mathbf{d})^\mathsf{T}\Sigma^{-1}(\mathbf{u} - \mathbf{G}\Delta\mathbf{d}),$$

$$\text{subject to } \boldsymbol{\varepsilon} \sim N(\boldsymbol{\varepsilon}|\mathbf{0}, \Sigma). \quad (8)$$

As with the existing ISR method, the optimal solution for this linear inverse problem $\hat{\Delta}\mathbf{d}$ is obtained analytically [18] with the variance-covariance matrix as follows [17,18,20]:

$$\hat{\Delta}\mathbf{d} = (\mathbf{G}^\mathsf{T}\Sigma^{-1}\mathbf{G})^{-1}\mathbf{G}^\mathsf{T}\Sigma^{-1}\mathbf{u}. \quad (9)$$

In statistical signal processing, this method is often referred to as *pre-whitening* [18]. One of the benefits of this method is that the multi-correlation (covariance) of the noise is also considered when minimizing the merit function. This means that the ISR with noise compensation (ISR-NC) method puts more importance on the wavelength domains with smaller noise variance when determining an optimal solution. The way in which the variance-covariance matrix is calculated is described in the following section.

2. Determination of the Variance-Covariance Matrix of the Noise

The variance-covariance matrix of the noise Σ is determined from the reference measurement with a known material. First the reference global interferogram I_{ref} is computed by averaging over the interferogram at the i-th pixel I^i_{ref} resulting in $\overline{\mathbf{HCF^d_{ref}}}$. The HCF function at the i-th pixel on the reference sample $\mathbf{HCF^{d,i}_{ref}}$ is also obtained.

Defining $\boldsymbol{\varepsilon}_i = \mathbf{HCF^{d,i}_{ref}} - \overline{\mathbf{HCF^d_{ref}}}$ as the noise at the i-th pixel its variance-covariance matrix Σ_o can be computed to match the expression in Eq. (9) as shown in Eq. (10). Note

that Σ_o is determined by computing the *sample variance* and *covariance* over an area corresponding to $M_{\rm ref}$ pixels (normally ~from 20×20 to 50×50):

$$\Sigma_o = E[(\mathbf{HCF_{ref}^d} - \overline{\mathbf{HCF_{ref}^d}})(\mathbf{HCF_{ref}^d} - \overline{\mathbf{HCF_{ref}^d}})^{\mathsf{T}}]$$

$$\simeq \frac{1}{M_{\rm ref}}\sum_{i=1}^{M_{\rm ref}}(\mathbf{HCF_{ref}^{d,i}} - \overline{\mathbf{HCF_{ref}^d}})(\mathbf{HCF_{ref}^{d,i}} - \overline{\mathbf{HCF_{ref}^d}})^{\mathsf{T}},$$

$$\therefore \Sigma \simeq \mathrm{Diag}[\overline{\mathbf{HCF^d}}]^{-1}\Sigma_o\{\mathrm{Diag}[\overline{\mathbf{HCF^d}}]^{-1}\}^{\mathsf{T}}, \tag{10}$$

where

$$\overline{\mathrm{HCF}^d_{\rm ref}}(\nu) = \bar{r}_{\rm ref}(\nu) \cdot \frac{\mathcal{F}[I_{\rm ref}]_{SB+}}{\mathcal{F}[I_{\rm ref}]_{SB+}} = \bar{r}_{\rm ref}(\nu),$$

$$\mathrm{HCF}^{d,i}_{\rm ref}(\nu) = \bar{r}_{\rm ref}(\nu) \cdot \frac{\mathcal{F}[I^i_{\rm ref}]_{SB+}}{\mathcal{F}[I_{\rm ref}]_{SB+}} + \varepsilon_i,$$

$$I_{\rm ref} \simeq \frac{1}{M_{\rm ref}}\sum_{i=1}^{M_{\rm ref}} I^i_{\rm ref}.$$

Note that no additional process is required to obtain Σ since the measurement of a reference sample is a prerequisite for the existing HCF-based techniques [13–15].

Figures 2 and 3 show the actual variance-covariance matrix of the noise and its variances (diagonal elements) obtained from a flat silicon surface used as a reference. It is clear that the noise variance is larger as the wavelength approaches its limits for both the real and imaginary parts. As expected, the noise variance is not constant over the spectral range of interest.
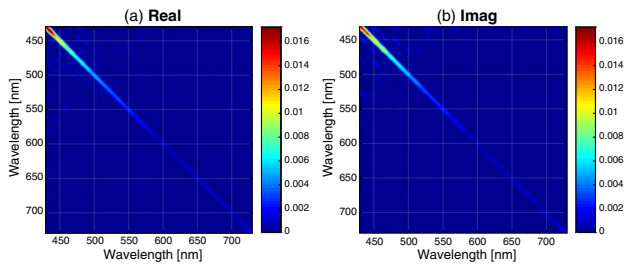

Fig. 2. Noise variance-covariance matrix Σ_o from a silicon reference sample with $M_{\rm ref} = 21 \times 21$ pixels (actual CSI measurement): (a) real part, (b) imaginary part (color available online).
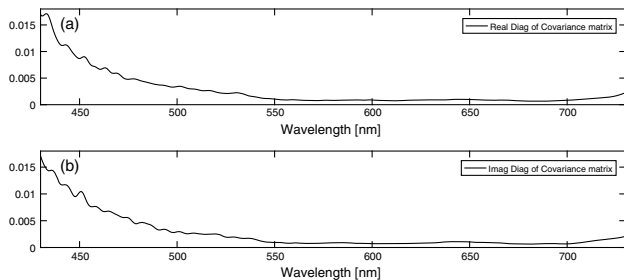

Fig. 3. Noise variance in the frequency domain (actual CSI measurement). The diagonal element of the (a) real and (b) imaginary parts of the noise variance-covariance matrix illustrated in Fig. 2.

3. COMPUTER SIMULATION

In this section a comparison is made between the existing ISR method and the ISR-NC method. Due to the approximations made there could be a discrepancy between the approximated HCF function and the original even if the interference signals are free from noise. Thus, the performance of each method is compared as follows: the ISR method free of noise (ISR-NF) the ISR method with noise (ISR), and the ISR method with noise compensation (ISR-NC).

A. Simulation Setup

For the model, we assume that a nanometer-sized feature is buried under a thin film as shown in Fig. 4. The number of pixels used for the reference measurement $M_{\rm ref}$ is fixed as $32 \times 32 = 1024$ throughout the simulations. White Gaussian noise is added to the interferogram (time domain) and the light intensity is tuned to produce profiles similar to those obtained from the experimental data (Figs. 2 and 3) resulting in the noise variances shown in Fig. 5. Comparing the actual noise variances with those simulated as shown in Figs. 3 and 5, the Gaussian noise used in the simulations is considered to reasonably reproduce the actual noise in the frequency domain. Note that the wavelength range for the model is set from 400 to 730 nm, similar to that used previously [15].

Table 1 shows the results from all the models tested. To simulate a high-performance instrument having a $4M$ pixel camera, such as the CCI HD (Taylor Hobson), the signals will be averaged over every four pixels to create an interference signal, so that 1024 signals in each measurement become 256 averaged signals.

B. Results and Analysis

Comparisons between the noise-free ISR (ISR-NF), the ISR, and the noise robust ISR (ISR-NC) methods are made by examining the height of the buried substrate and the surface roughness (Sq) of the top surface and buried layer interface.
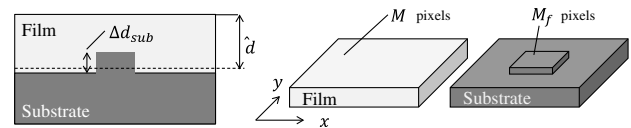

Fig. 4. Schematic drawing of the model. The number of the pixels in the measurement area is M together with the pixels having the feature is M_f. Note that the global film thickness $\hat{d} \simeq d$.
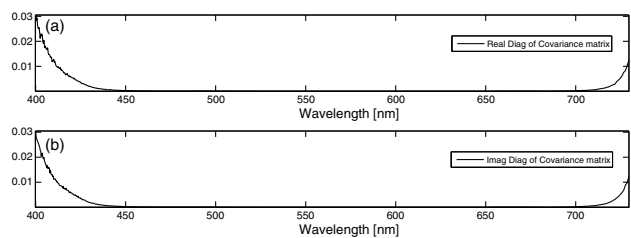

Fig. 5. Noise variance in the frequency domain: The diagonal element of the (a) real and (b) imaginary parts of the noise variance-covariance matrix given to the simulations. Note that the signal-to-noise (S/N) ratio is 2000.

Table 1.   Simulation Conditions Together with the Number of Pixels Allocated for the "Global" and "Featured" Areas

Sim #	Substrate Type	Film Type	Film Thickness (nm)	Feature (nm)	Noise S/N[a]	M[b]	M_f[b]
1-1	Si	SiO$_2$	309, 514, 720, 925	5	10^3	36^2	8^2
1-2	Si	ZrO$_2$	203, 339, 464, 610	5	10^3	36^2	8^2
2-1	Si	SiO$_2$	514	2.5, 5, 10, 20	10^3	36^2	8^2
2-2	Si	ZrO$_2$	339	2.5, 5, 10, 20	10^3	36^2	8^2
3-1	Si	SiO$_2$	514	5	10^4–10^2	36^2	8^2
3-2	Si	ZrO$_2$	339	5	10^4–10^2	36^2	8^2
4-1	Si, SiC, BK7, Ge	SiO$_2$	514	5	10^3	36^2	8^2
4-2	Si, SiC, BK7, Ge	ZrO$_2$	339	5	10^3	36^2	8^2
5-1	Si	SiO$_2$, ZrO$_2$, Ta$_2$O, AZO	350	5	10^3	36^2	8^2
5-2	Si	SiO$_2$, ZrO$_2$, Ta$_2$O, AZO	700	5	10^3	36^2	8^2

[a]The S/N level interval is set reasonably so that its effect on performance can be evaluated, particularly for 3-1 and 3-2. The noise added to the signals in the time domain follows a normal distribution.

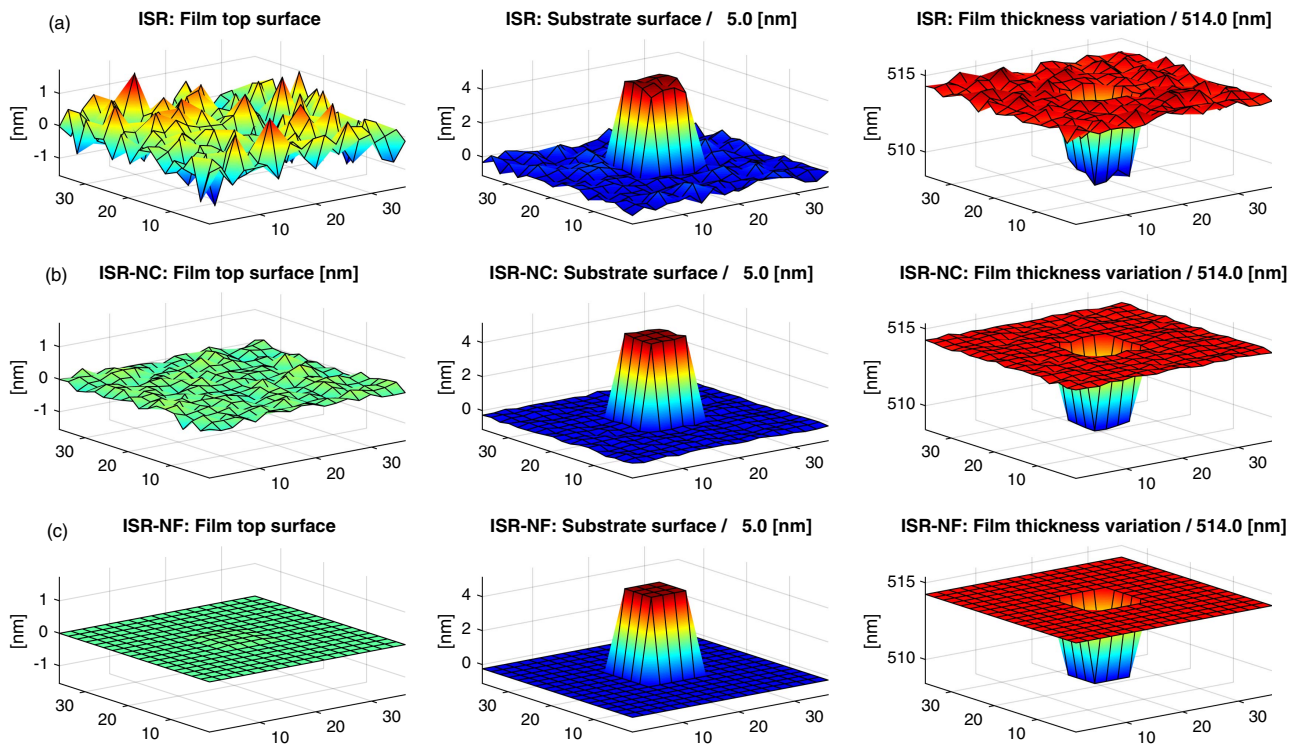[b]The number of pixels is reduced by a quarter after averaging the signal over four pixels.

Fig. 6.   Comparisons between the three computational methods on the sample: SiO$_2$ (thickness s = 514 nm) on a Si substrate; the feature height is 5 nm: (a) the ISR method (with noise), (b) the ISR-NC method (with noise), (c) the ISR-NF method (noise free). The S/N ratio is set at 10^2 to correspond to Sim 3-1 in Table 1 (color available online).

Figure 6 shows three-dimensional images of the resulting computations using the three methods. The ISR-NC method yields the smoothest surfaces, which is also the case for all the models tested since the method is free from noise.

1. Simulation 1: Performance Sensitivity to Variation in Thin Film Thickness
Thin films, of SiO$_2$ and ZrO$_2$, of varying thickness were investigated while other parameters remained unchanged, as shown in Table 1. The thicknesses in Table 1 correspond to

odd integer multiples (3, 5, 7, 9) of the *quarter-wavelength optical thickness* (QWOT).

Although the feature heights determined by the ISR and ISR-NC methods were very similar, the height variance of the ISR-NC method was always less than that provided by the ISR method.

The surface roughness (Sq) of the top surface and the substrate determined by the ISR-NC method are smoother than those of the ISR method, as shown in Fig. 7. In addition, one simulation (Sim 1-2 with a 610 nm ZrO$_2$ film) did not work

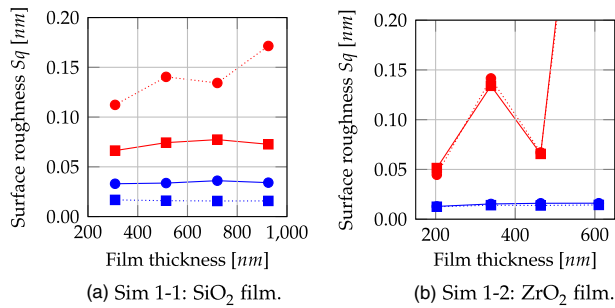(a) Sim 1-1: SiO$_2$ film.

(b) Sim 1-2: ZrO$_2$ film.

Fig. 7. Surface roughness (Sq) as a function of film thickness: Red circles, top ISR; red squares, Sub ISR; blue circles, top ISR-NC; blue squares, sub ISR-NC.

properly with the ISR method due to an inaccurate approximation. This is discussed later in Section 4.

2. Simulation 2: The Effect of Feature Height on Performance

The ISR methods, including ISR-NC, use a first-order Taylor expansion to the HCF function to make the problem linear [15,16,19]. This requires that the perturbation of the interfacial surface topography is "small." In this experiment, the noise is fixed at $S/N = 10^3$ and we evaluate the performance of the methods as a function of feature height with the fixed thin film thicknesses of 514 nm for SiO$_2$ and 339 nm for ZrO$_2$.

Figure 8 shows that the ISR methods with or without noise work well up to ~10 nm in feature height for the SiO$_2$ film and up to ~5 nm for the ZrO$_2$ film. The ISR-NC method gives a reasonable approximation up to ~10 nm for ZrO$_2$ film. These results, however, do not necessarily prove the superiority of the ISR-NC method. The root cause of the deterioration in performance, which is basically proportional to the feature height, is the quality of the first-order approximation to the HCF function of interest. The HCF function is poorly approximated in some wavelength regions which have a high noise variation, as shown in Fig. 5.

3. Simulation 3: Noise Compensation Performance

In this set of simulations, the performance of each method as a function of the signal-to-noise (S/N) ratio is investigated. The other parameters remain unchanged, as shown in Table 1.

The ISR-NF, ISR, and ISR-NC methods give similar mean feature height values regardless of the noise level for the SiO$_2$
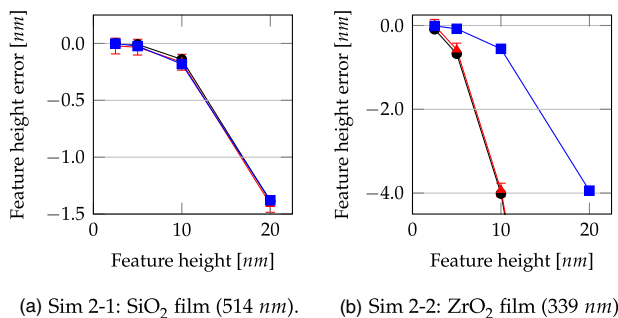

(a) Sim 2-1: SiO$_2$ film (514 nm).

(b) Sim 2-2: ZrO$_2$ film (339 nm).

Fig. 8. Feature height sensitivity as a function of feature height: black circles, ISR-NF; red triangles, ISR; blue squares, ISR-NC.

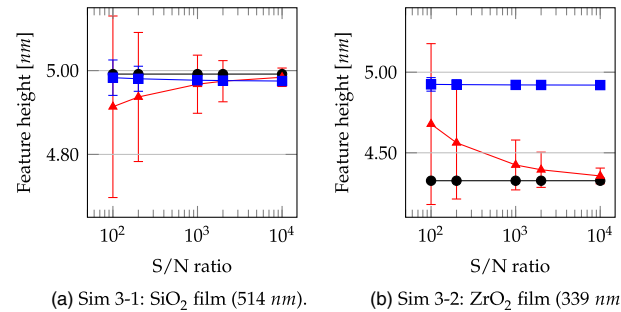(a) Sim 3-1: SiO$_2$ film (514 nm).

(b) Sim 3-2: ZrO$_2$ film (339 nm).

Fig. 9. Signal-to-noise ratio sensitivity to the determined feature height: black circles, ISR-NF; red triangles, ISR; blue squares, ISR-NC.

(a) Sim 3-1: SiO$_2$ film (514 nm).

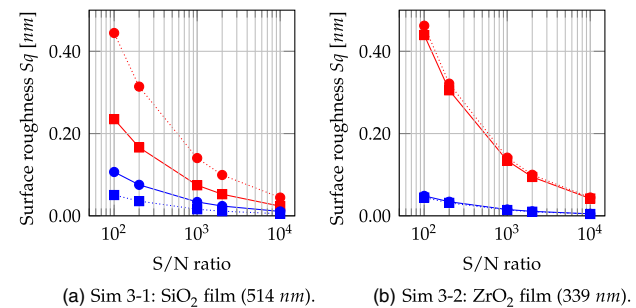(b) Sim 3-2: ZrO$_2$ film (339 nm).

Fig. 10. Surface roughness (Sq) as a function of the S/N ratio: red circles, top ISR; red squares, sub ISR; blue circles, top ISR-NC; blue squares, sub ISR-NC.

film but this is not the case for ZrO$_2$ film. The ISR method determines the height as 4.3 nm compared to the actual value of 5 nm, as shown in Fig. 9(b). As in the previous Section 3.B.2, this is due to a poor first-order approximation of the amplitude reflection coefficient in the smaller variance wavelength regions.

The thin film and substrate surface roughnesses determined by the ISR and ISR-NC methods are proportional to the increase in the S/N ratio, as shown in Fig. 10. However, the level of surface roughness determined by the ISR-NC method is lower for all noise levels.

4. Simulation 4: Effect of Substrate Materials

All the variables except for the substrate material are unchanged in the simulations 4-1 and 4-2, as shown in Table 1. The substrate materials used are Si, SiC, BK7 glass, and Ge.

Similar to previous results in Sections 3.B.1–3.B.3, the ISR-NC method resulted in a smaller variance in feature height in both simulations 4-1 and 4-2. The film and substrate surfaces determined by the ISR-NC method are about an order of magnitude smoother than those from the ISR method.

5. Simulation 5: Effect of the Type of Deposited Film

Simulations 5-1 and 5-2 shown in Table 1 investigate the effect of different film materials for 350 nm and 700 nm thickness films. The ISR-NC method again provided more accurate buried surface topographies together with smaller variances.

Table 2. Corresponding Effective Quarter-Wavelength Optical Thickness Values at the Wavelength of 600 nm

Film Material	Index of Refraction at 600 nm	Film Thickness	
		350 nm	700 nm
SiO_2	1.46	3.4	6.8
ZrO_2	2.21	5.2	10.3
Ta_2O_5	2.12	5.0	9.9
AZO	1.83	4.3	8.5

The reconstructed surfaces were about an order of magnitude smoother irrespective of the thin film material using the ISR-NC method.

Table 2 shows the effective QWOT values of the films used in the simulations. Usually films with thickness greater than QWOT × 3 are considered to have enough features in the frequency domain for the HCF theory to work [15]. It follows that there should not be much difference in the simulated
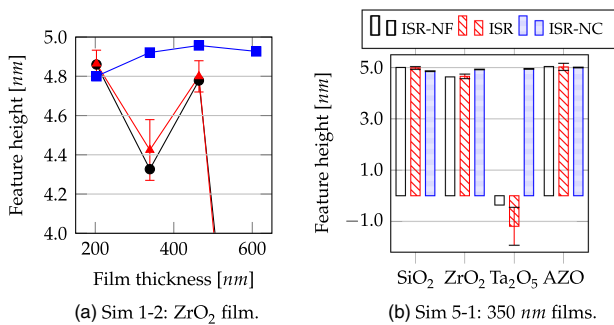
performance of the various films unless the first-order approximation to the HCF function is sufficiently accurate. However, the ISR method did not work for the Ta_2O_5 film in simulation 5-1 while the ISR-NC method did. This issue will be discussed in the following Section 4.

4. DISCUSSION

As shown in Sections 3.B.1, 3.B.2 and 3.B.5, the ISR method does not work optimally resulting in erroneous buried feature heights such as those from simulation 2-2 with a feature height >10 nm, simulation 1-2 with a ZrO_2 600 nm film, and simulation 5-1 with a Ta_2O_5 350 nm film, as shown in Figs. 8(b) and 11, respectively. All these simulations show that the corresponding surfaces determined by the ISR-NC method are more accurately represented than those using the noise-free ISR method. The root cause of this problem lies in an inaccurate approximation to the HCF function. The lack of accuracy of the HCF function arises when the perturbation (feature height) of the interfacial topography is too large ($\sim > 10$ nm) and when the approximated spectral amplitude reflection coefficient locally deviates from the true value resulting in a spike.

Consider first the simulation 2-2, which has a 20 nm feature height. If we compare the true HCF function with its first-order approximation, then Fig. 12(a) shows that the first-order approximation does not hold, especially in the wavelength region between 400 and 475 nm. Fig. 12(b) shows the difference between the true HCF function HCF_{px}^d (without noise) and the approximated estimates by each method HCF_{px}^s (in the presence of noise). Prior knowledge of the noise variance-covariance matrix Σ allows the ISR-NC method to put less importance on the value of the HCF function in the specific wavelength domains where the noise is large, i.e., from ~400 to ~450 nm and from ~700 to ~730 nm, as shown in Fig. 5. This is

Fig. 11. Erroneous determined feature height (originally set as 5 nm): black circles, ISR-NF; red triangles, ISR; blue squares, ISR-NC.

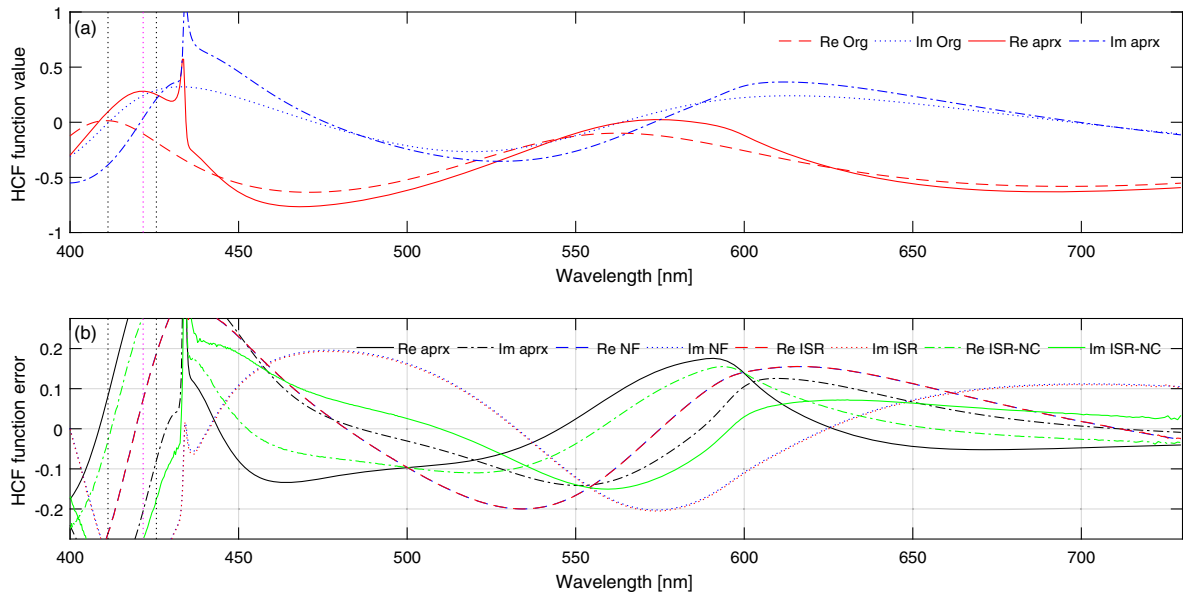(a) Sim 1-2: ZrO_2 film. (b) Sim 5-1: 350 nm films.

Fig. 12. HCF functions generated at the feature pixel (simulation 2-2 with 20 nm feature height): (a) true HCF function (without noise) denoted by "Org" and its first-order approximation by "aprx"; (b) spectral difference between the true HCF function HCF_{px}^d and the HCF functions produced by each method HCF_{px}^s (noise occurs in both ISR and ISR-NC, and NF stands for ISR-NF) (color available online).
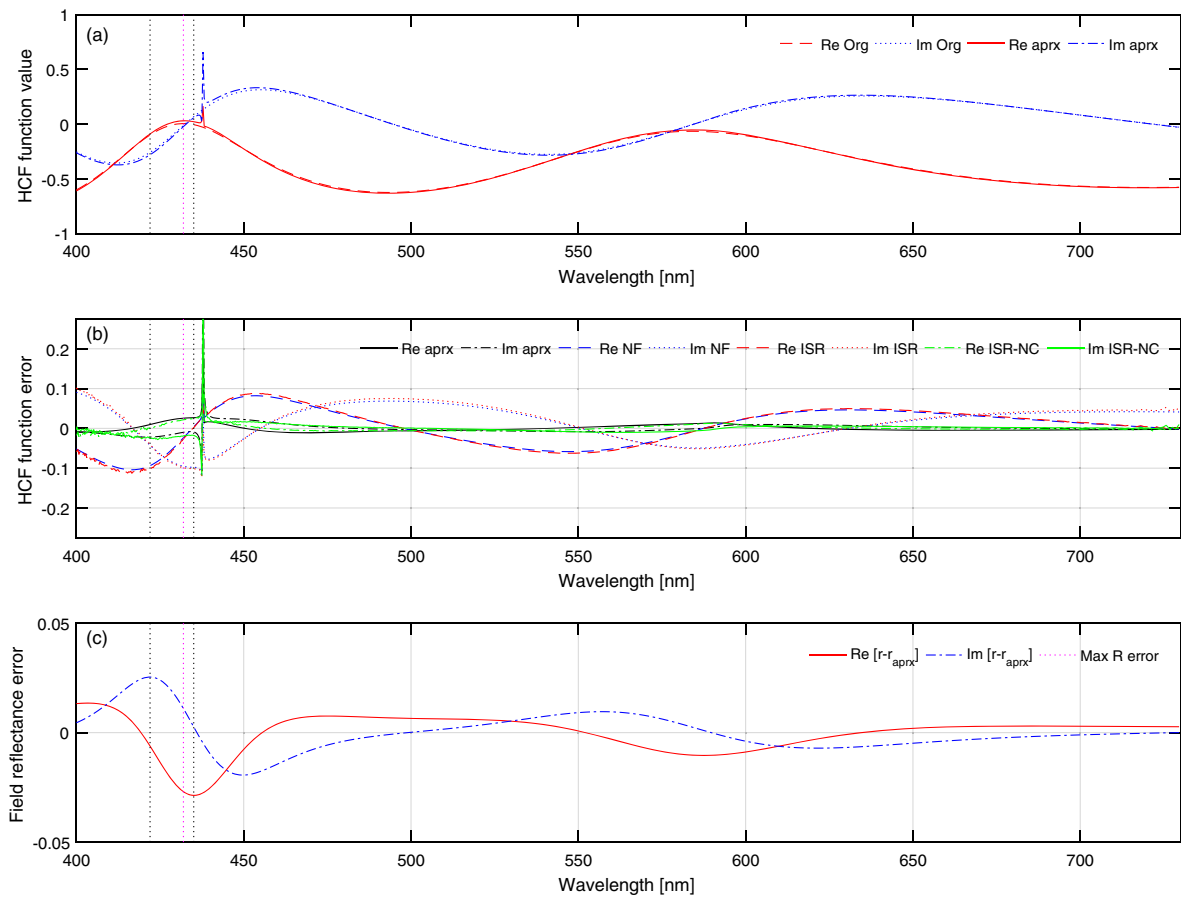
Fig. 13. HCF functions generated at the feature pixel (simulation 5-1 with Ta_2O_5 film): (a) the true HCF function (without noise) denoted by "Org" and its first-order approximation by "aprx"; (b) the spectral difference between the true HCF function HCF_{px}^d and the HCF functions produced by each method HCF_{px}^s (noise exists for ISR and ISR-NC, and NF stands for ISR-NF); (c) the spectral difference between the real and imaginary parts of the true amplitude reflection coefficient and its first-order approximation. Note that the dotted lines (black and pink) represent the maximum deviations of the real ($Re[r - r_{aprx}]$), imaginary ($Im[r - r_{aprx}]$), and the reflectivity R, respectively (color available online).

not the case for the ISR method and is the reason why the ISR-NC method provides more accurate determinations.

The second cause of inaccurate surface reconstruction observed in simulations 1-2 and 5-1 is due to an inaccurate approximation to the amplitude reflection coefficient. Consider simulation 5-1 using the Ta_2O_5 thin film. The first-order approximation to the HCF function is successful, as shown in Fig. 13(a), except for a spike observed at ∼440 nm wavelength denoted by "aprx". The solution provided by the ISR method defined in Eq. (5) is influenced by this spike, which reduces the fitting performance as shown in Fig. 13(b). The residual of $\|HCF_{px}^s - HCF_{px}^d\|^2$ at 435 nm wavelength is relatively small for the ISR method whereas that given by the ISR-NC method is large. It follows that the ISR-NC method does not attempt to fit the spike feature due to the noise variance-covariance matrix Σ.

To confirm this further, an improvement in the performance of the ISR method was achieved by reducing the wavelength region used for numerical optimization to avoid the region in which the spikes occur.

To achieve a good fit between the determined and synthesized HCF functions in the frequency domain, there are two options: (1) using the ISR method with wavelength domains having less noise variance, such as from 430 to 700 nm in the examples above, or (2) using the ISR-NC method. The latter option enables the measurement of thinner films to be more stable owing to the wider wavelength domain for curve-fitting, irrespective of the noise characteristics.

5. CONCLUSIONS

Present methods for interfacial surface roughness measurement using CSI can be classified into two types: those that compute surface topographies in the time domain and those that determine surface topographies in the frequency domain. The methods belonging to the first group are used for films over ∼1.5 μm in thickness whereas those in the second group are able to deal with thin films less than ∼1.5 μm. The frequency domain methods usually use the least-squares optimization to fit the mathematical model to the measurement signal. However, the basic assumption for the method is that the noise is normally distributed and thus least-squares is not always suitable. In the examples above, the noise variance of the HCF function was not constant over the wavelength domain from

430 to 730 nm. The noise variance is always larger at the ends of the spectral region of interest where the light intensity is low. Therefore the noise variance-covariance matrix should be used in the numerical optimization of the ISR method. Such a matrix is obtained from the measurement of a known flat reference material and will vary depending on the environmental situation and the particular light source used.

Although the ISR method using the HCF function successfully determined the roughness of the thin film top surfaces and buried surfaces [15,16], spurious surface roughness in the determined substrate surfaces could be observed. This paper has presented an effective solution to that problem by introducing the noise variance-covariance matrix, which only involves a small computation when measuring the reference surface. Measurement of the reference surface is required anyway to counteract unknown changes in the phase and amplitude of the light provided by the optical system of the CSI instrument. Using these signals at the same time for the noise analysis is a beneficial side effect.

The reproducibility of the ISR-NC method was better than the existing ISR method for all the computer simulations in the presence of noise for determination of interfacial topography and surface roughness (Sq). The method was also effective over a wide wavelength range, thus allowing use of more features of the HCF function for the curve-fitting and hence better reproducibility. The noise used in the computer simulations is realistic since the noise variance-covariance matrix obtained from the flat silicon surface in Figs. 2 and 3 is similar to the noise in Fig. 5. Incorporation of noise compensation to the ISR method will improve the measurement accuracy.

REFERENCES

1. P. de Groot, "Coherence scanning interferometry," in *Optical Measurement of Surface Topography* (Springer, 2011), Chap. 9, pp. 187–208.
2. T. R. Thomas, "Other measurement topics," in *Rough Surfaces*, 2nd ed. (Imperial College, 1999), Chap. 3, pp. 35–61.
3. H. Yoshino, R. Smith, J. M. Walls, and D. Mansfield, "The development of thin film metrology by coherence scanning interferometry," Proc. SPIE **9749**, 97490P(2016).
4. F. Blateyron, "The areal field parameters," in *Characterisation of Areal Surface Texture* (Springer, 2013), pp. 15–43.
5. B. S. Lee and T. C. Strand, "Profilometry with a coherence scanning microscope," Appl. Opt. **29**, 3784–3788 (1990).
6. P. J. De Groot and X. C. de Lega, "Transparent film profiling and analysis by interference microscopy," Proc. SPIE **7064**, 70640I (2008).
7. A. Bosseboeuf and S. Petitgrand, "Application of microscopic interferometry techniques in the MEMS field," Proc. SPIE **5145**, 1–16 (2003).
8. S. Kim and G. Kim, "Method for measuring a thickness profile and a refractive index using white-light scanning interferometry and recording medium therefor," U.S. patent 6,545,763 (8 April 2003).
9. P. J. de Groot, "Interferometry method for ellipsometry, reflectometry, and scatterometry measurements, including characterization of thin film structures," U.S. patent 7,403,289 (22 July 2008).
10. Y. Ghim and S. Kim, "Spectrally resolved white-light interferometry for 3D inspection of a thin-film layer structure," Appl. Opt. **48**, 799–803 (2009).
11. I. Abdulhalim, "Spectroscopic interference microscopy technique for measurement of layer parameters," Meas. Sci. Technol. **12**, 1996–2001 (2001).
12. D. Mansfield, "Apparatus for and a method of determining characteristics of thin-layer structures using low-coherence interferometry," WO patent PCT/GB2005/002,783 (19 January 2006).
13. D. Mansfield, "The distorted helix: thin film extraction from scanning white light interferometry," Proc. SPIE **6186**, 61860O (2006).
14. H. Yoshino, P. M. Kaminski, R. Smith, J. M. Walls, and D. Mansfield, "Refractive index determination by coherence scanning interferometry," Appl. Opt. **55**, 4253–4260 (2016).
15. H. Yoshino, A. Abbas, P. M. Kaminski, R. Smith, J. M. Walls, and D. Mansfield, "Measurement of thin film interfacial surface roughness by coherence scanning interferometry," J. Appl. Phys. **121**, 105303 (2017).
16. D. Mansfield, "Extraction of film interface surfaces from scanning white light interferometry," Proc. SPIE **7101**, 71010U (2008).
17. T. Lewis and P. Odell, "A generalization of the Gauss-Markov theorem," J. Am. Stat. Assoc. **61**, 1063–1066 (1966).
18. K. Sekihara, *Introduction to Statistical Signal Processing* (Kyouritsu Shuppan, 2011).
19. D. I. Mansfield, "Apparatus for and a method of determining surface characteristics," U.S. patent 13/352,687 (12 July 2012).
20. A. Albert, "The Gauss-Markov theorem for regression models with possibly singular covariances," SIAM J. Appl. Math. **24**, 182–187 (1973).