Large-angle in-plane light scattering from rough surfaces

Tansel Karabacak, Yiping Zhao, Matthew Stowe, Bill Quayle, Gwo-Ching Wang, and Ton-Ming Lu

An in-plane light scattering setup that is capable of measuring large azimuthal scattering angles is presented. This type of measurement makes it easier to probe large $k_i$ at a fixed $k'_i$ value ($k_i$ and $k'_i$ are momentum transfer vectors parallel and perpendicular to the surface, respectively). Therefore the system allows us to explore small lateral scale and large vertical roughness ($\sim l$, the wavelength of the probe beam) of a rough surface. In-plane intensity measurements from a rough backside Si wafer and a Cu thin-film surface are reported. The structure factor that is related to surface roughness parameters is obtained from the measured in-plane intensity profiles. Both scalar (Beckmann–Kirchhoff) and vector (Rayleigh–Rice) theories have been applied to interpret the experimental data. The roughness parameters obtained from the scattering measurements are compared with those measured by atomic-force microscopy. © 2000 Optical Society of America


1. Introduction

The characterization of random rough surfaces by light scattering has received considerable attention in recent years.1–3 Although the most direct way to probe a rough surface quantitatively is to measure the surface morphology by use of real space imaging techniques such as atomic-force microscopy (AFM) and scanning-tunneling microscopy (STM), light scattering does have some advantages over imaging: It is a noncontact method and can have a large sampling size. With proper design, it can be used as a real-time monitoring tool.4–7

The general diffraction coordinate and the reciprocal space geometry for light scattering are shown in Figs. 1(a) and 1(b), respectively. The $x$–$y$ plane is the sample plane, the $z$ direction is the sample surface normal, and the $x$–$z$ plane is the plane of incidence. The incident angle is $\theta_i$, the polar angle of the scattered beam is $\theta_s$, and the azimuthal angle is $\phi_s$. For diffraction from a random rough surface, the momentum transfer $k$ that is due to scattering can be decomposed into two parts, $k_\parallel$, the momentum transfer parallel to the sample surface, and $k_\perp$, the momentum transfer perpendicular to the sample surface:

$$k_\parallel = k(\cos \theta_s + \cos \theta_i)$$

$$k_x = k(\sin \theta_i \cos \phi_s - \sin \theta_s)$$

$$k_y = k \sin \theta_s \sin \phi_s$$

where $k = 2\pi/\lambda$ is the wave number. Basically there are two different detection schemes for a scattering measurement: a moving sample scheme and a moving detector scheme. In the moving sample scheme, the position of the detector is fixed, and the sample rotates with respect to a certain axis. As the sample rotates, the detector detects signals associated with the rotation angles, generating angular-dependent diffraction profiles. In the moving detector scheme, the position of the sample is fixed, and the detector rotates with respect to a certain axis.

For the moving sample scheme there are basically two types of light-scattering geometry: out of plane and in plane, as shown in Fig. 2. For the in-plane geometry the detector rotates in a plane parallel to the sample plane ($\theta_s = \theta_i = \text{constant}$); the rotation axis is the $z$ axis. Still using the reflection direction as a reference point, we have

$$k_\perp = 2k \cos \theta$$

$$k_x = k \sin \theta(\cos \Delta \phi - 1)$$
Clearly, in this case \( k_1 \) is independent of the rotation angle, and only \( k_i = (k_x^2 + k_y^2)^{1/2} \) changes with the rotation angle. Figure 3(a) plots \( k_1 \) and \( k_i \) as functions of the rotation angle at \( \theta = 45^\circ \) in the range \( |\Delta \theta| \leq 10^\circ \) for the in-plane geometry. In the in-plane geometry the range of the rotation angle is from \(-180^\circ\) to \(180^\circ\). The range of \( k_i \) is from 0 to \( 2k \sin \theta \), which is determined by the incident angle \( \theta \). The greater \( \theta \), the larger the \( k_i \) range.

For the out-of-plane geometry the detector rotates in the incident plane, i.e., the \( x-z \) plane (\( \phi_x = 0 \)); the rotation axis is the \( y \) axis. Usually the reflection direction serves as a reference point, i.e., \( \theta_x = \theta + \Delta \phi \).

After a rotation, the momentum transfer changes into

\[
k_\perp = k[\cos(\theta + \Delta \theta) + \cos \theta],
\]

\[
k_x = k[\sin(\theta + \Delta \theta) - \sin \theta],
\]

\[
k_y = 0.
\]

Figure 3(b) plots \( k_\perp \) and \( k_x \) as functions of the rotation angle at \( \theta = 45^\circ \) in the range \( |\Delta \theta| \leq 10^\circ \) for the out-of-plane geometry. Clearly both \( k_x \) and \( k_\perp \) change. The range of the rotation angle \( \Delta \theta \) for the out-of-plane geometry is from \( \pi/2 - \theta \) to \( \pi/2 + \theta \). Therefore the range of \( |k_\perp| \) is from 0 to \( \max[k(1 - \sin \theta), k|\sin(\pi/2 + 2\theta) - \sin \theta|] \).

Using either geometry makes it possible to obtain the diffraction structure factor, which can provide the surface roughness information. So far, in most of the literature on light scattering the out-of-plane geometry has been used (see, for example, Ref. 8 and references therein). In only a few papers has the in-plane geometry been considered.5,9–12 However, as we compare Figs. 3(a) and Fig. 3(b) for the out-of-plane geometry, both \( k_\perp \) and \( k_x \) change with the rotation angle, whereas for the in-plane geometry only \( k_\parallel \) changes with the rotation angle. Therefore we
would expect it to be easier to analyze the diffraction profile obtained with an in-plane diffraction geometry than that of the out-of-plane geometry. In this paper we show that the in-plane diffraction geometry has a considerable advantage over the traditional out-of-plane geometry in terms of data acquisition and the analysis of roughness parameters. Using the in-plane geometry, one can reach the limit of the lateral resolution for an optical technique. Also, one can detect the vertical roughness of a surface as large as the wavelength of the incident beam.

We performed detailed in-plane measurements of a backside Si wafer and a Cu sample. We obtained the roughness parameters from the measured scattering profiles and determined the roughness exponent $\alpha$ from the power-law behavior in the tail parts of the diffracted intensity. We later extracted the $\alpha$ values and the other roughness parameters (vertical root-mean-square roughness $w$ and lateral correlation length $\xi$) by using an AFM and compared them with those obtained by the in-plane light scattering technique.

2. Diffraction Intensity

In general, the diffraction intensity can be written as

$$I(k) = FP(k) \ast S(k),$$  \hspace{1cm} (10)

where $F$ is a function that depends on the scattering angles ($\theta_i, \phi_i$) and the physical properties of the sample such as the refractive index, the wavelength, and the polarization of the incident beam. $P(k)$ is the system point-spread function, or the system response function; the asterisk represents convolution. Therefore the system response determines the angular resolution or the reciprocal space resolution of the system. $S(k)$ is the diffraction structure factor. If the local slope of the surface is small, i.e., $|\nabla z| \ll 1$, according to the Kirchhoff approximation the structure factor $S(k)$ can be written as

$$S(k) = \int C(k_{\perp}, r) \exp(i \mathbf{k}_{\parallel} \cdot \mathbf{r}) d^2r,$$  \hspace{1cm} (11)

i.e., the structure factor is a Fourier transform of the height difference function of the surface, $C(k_{\perp}, r) = \langle \exp(i \mathbf{k}_{\parallel} \cdot \mathbf{r}) \rangle$, where $z(\mathbf{r})$ is the height at lateral distance vector $\mathbf{r}$ and the average is carried out all over the surface through $\mathbf{r}$. $C(k_{\perp}, r)$ is a measure of the statistical phase correlations between surface positions. For a statistically rough surface, the relative surface heights, $z(r) - z(0)$, usually obey a Gaussian distribution. The height difference function has the form

$$C(k_{\perp}, r) = \exp \left[ -k_{\perp}^2 H(r) \right],$$  \hspace{1cm} (12)

where $H(r) = \langle (z(r) - z(0))^2 \rangle$ is the height–height correlation function. In the case of self-affine, homogeneous, and isotropic rough surfaces, $H(r)$ can be expressed as

$$H(r) = H(\xi) = 2w^2f(r/\xi),$$ \hspace{1cm} (13)

where $w = \langle (z(r) - \bar{z})^2 \rangle^{1/2}$ is the interface width and $\bar{z}$ is the average surface height. The function $f(r/\xi)$ is a scaling function that has the following properties within the short-range and the long-range regions:

$$f(r/\xi) = \begin{cases} (r/\xi)^{2\alpha} & r \ll \xi \\ 1 & r \gg \xi \end{cases}.$$ \hspace{1cm} (14)

where $\alpha$ is the roughness exponent; it describes how wiggly the surface is and is related to the surface fractal dimension $D_s$ by $\alpha = 3 - D_s$ (note that $\alpha$ must satisfy $0 \leq \alpha \leq 1$). As an example, one can use a phenomenological function, $f(r/\xi) = 1 - \exp(-r/\xi^{2\alpha})$, which satisfies Eq. (14). For a surface with a Gaussian or an exponential autocorrelation function [defined as $\langle z(\mathbf{r})z(0) \rangle$], the value of $\alpha$ is 1 or 0.5, respectively. The lateral correlation length, $\xi$, is the distance within which the surface heights of any two points are correlated. The parameters $w$, $\xi$, and $\alpha$ characterize the major statistical properties of a self-affine surface.

The structure factor $S(k)$ can be divided into two parts. One is a sharp central $\delta$ function associated with the long-range flatness of the surface; another one contains a broad diffuse component and is connected to the short-range roughness of the surface:

$$S(k) = S_\delta(k_{\parallel}, k_{\perp}) + S_{\text{diff}}(k_{\parallel}, k_{\perp}),$$  \hspace{1cm} (15)

where $S_\delta(k_{\parallel}, k_{\perp}) = (2\pi)^2 \exp(-\Omega)\delta(k_{\parallel})$ and $\Omega = (k_{\perp}w)^2$. Using the generic form of the height–height correlation function [Eq. (13)], we can write the diffuse part $S_{\text{diff}}(k_{\parallel}, k_{\perp})$ as

$$S_{\text{diff}}(k_{\parallel}, k_{\perp}) = \exp(-\Omega) \int d^2r \times \exp[\Omega(1 - f(r/\xi))] \exp(i \mathbf{k}_{\parallel} \cdot \mathbf{r}).$$ \hspace{1cm} (16)

Here we see that the diffuse structure factor is sensitive to the short-range properties of rough surfaces.

At this point we can see one of the motivations for using an in-plane measurement technique. It can be seen from Eqs. (15) and (16) that the $\delta$ peak intensity is a function only of $\Omega$ and that the diffuse profile is a function of both $\Omega$ and $k_{\parallel}$. If we can measure the structure factor of a surface in such a way that we keep $k_{\parallel}$ constant, then we can directly obtain the lateral roughness parameters $\xi$ and $\alpha$ from the diffuse profile. The vertical roughness parameter $w$ can be obtained from the $\delta$ peak intensity at different $k_{\parallel}$ values. As we discussed above, the in-plane scattering geometry provides such an opportunity. For the out-of-plane geometry, inasmuch as both $k_{\parallel}$ and $k_{\perp}$ change with the change of the polar angle, Eq. (16) becomes too complicated to be analyzed. As we approach the region where $\Omega \ll 1$, the dif-
fuse part of the structure factor takes a form that is directly proportional to the power-spectrum density (PSD) function of the surface:

\[ S_{\text{diff}} = (2\pi)k_\perp^2 \exp(-\Omega)\text{PSD}(k), \]  

where the PSD is defined to be

\[ \text{PSD}(k_x, k_y) = \frac{1}{A} \frac{1}{2\pi} \left| \int \int z(x, y) \exp[i(k_x x + k_y y)] \, dx \, dy \right|^2, \]

where \( A \) is the area of the scatterer and \( z(x, y) \) is the surface height function. Therefore in this limit we obtain

\[ S(k) = (2\pi)^2 \exp(-\Omega)\delta(k) + (2\pi)k_\perp^2 \exp(-\Omega)\text{PSD}(k). \]

In addition, a power-law behavior of the form \( k_\perp^{-2-2\alpha} \) always exists in the large \( k_\perp \) regime \([k_\perp \gg \max(1, \Omega)^{1/2}]\) for all values of \( \Omega \) and has the form

\[ S_{\text{diff}}(k, k_\perp) = 2\pi e^{-2\Omega} B k_\perp^{-2-2\alpha}, \]

where \( B = 2^{2\alpha+1} \alpha \Gamma(\alpha + 1)/\Gamma(\alpha - 1). \) This power-law relation also holds for surfaces with non-Gaussian height distributions.

One can therefore obtain the roughness parameter \( \alpha \) by calculating the slope taken from the tail part of the log–log plot of \( S_{\text{diff}}(k, k_\perp) \) versus \( k_\perp \), where \( k_\perp \) is kept constant (the slope equals \(-2-2\alpha\)). Here we see another advantage of the in-plane measurement technique. Because \( k_\perp \) is kept constant during the in-plane measurement and we know that there is no contribution from the \( \delta \) part away from \( k_\perp = 0 \) [see Eq. (15)], we can easily extract \( \alpha \) from the log–log plot of the structure factor \( S \) versus \( k_\perp \). Then the only thing needed is an efficient in-plane measurement technique that has the ability to provide large \( k_\perp \) values.

One of the challenging parts of our study was to extract the structure factor from the intensity profile in the case of scattering from a considerably rough surface. To do so, one needs to know the angular- and the material-properties-dependent factor \( F \) in Eq. (10). Starting from different assumptions, one can find different factors. Here we consider two possible types of prefactor \( F \) that originate from scalar and from vector theories.

A. Scalar (Beckmann–Kirchhoff) Theory

In the scalar theory the predicted diffuse intensity can be expressed in a series form. As we approach the limit \( \Omega \ll 1 \), only the first term of the series survives. In this case the relation between the diffuse part of the structure factor, \( S_{\text{diff}} \), and the intensity, \( I \), can be shown to be

\[ S_{\text{diff}}(k) \propto \frac{k_\perp^2 \exp(-\Omega)\lambda^4}{F_B^2(\cos \theta_x + \cos \theta_y)^2} I(k), \]

where \( F_B \) is a function of the scattering angle and the reflectivity \( R \) of the sample: \( F_B = F_B(\theta_x, \theta_y, \phi_x, R) \). Explicitly, it is defined as

\[ F_B = \frac{b + (a k_x + c k_y)/k_\perp}{2 \cos \theta_i}, \]

where \( k = 2\pi/\lambda \) and

\[ a = (1 - R)\sin \theta_i + (1 + R)\sin \theta_x \cos \phi_x, \]

\[ b = (1 + R)\cos \theta_i - (1 - R)\cos \theta_x, \]

\[ c = (1 + R) \sin \theta_x \sin \phi_x. \]

As \( \Omega \) gets larger, relation (21) should begin to fail because of the extra terms that come from the series expansion. Also, shadowing and multiple scattering effects are some of the basic concerns in the Beckmann–Kirchhoff theory.22,23 As these effects become significant, we should expect the Beckmann–Kirchhoff theory to work worse at large incident and scattering angles.

B. Vector (Rayleigh–Rice) Theory

Using the Rayleigh–Rice theory in the limit \( \Omega \ll 1 \), one can obtain a relation between \( S_{\text{diff}} \) and \( I \) that includes the polarization effects and the sample's dielectric constant \( \epsilon \) (Ref. 8):

\[ S_{\text{diff}}(k) \propto \frac{k_\perp^2 \exp(-\Omega)\lambda^4}{Q^2 \cos^2 \theta_i \cos \theta_i} I(k), \]

where \( Q = Q(\theta_x, \theta_y, \phi_x, \epsilon) \) depends on the polarization states of the incident light and the reflected light in the detector. For the case of right or left circularly polarized incident light and a polarization-insensitive detector, \( Q \) is found to be (see Appendix A for detailed calculations)

\[ Q = |q_{sll}|^2 + |q_{spl}|^2 + |q_{lll}|^2 + |q_{psl}|^2, \]

where the elements \( q_{ijkl} \) are given in Eqs. (A12) below.

The vector Kirchhoff theory is another light-scattering theory that is worth mentioning. For small surface roughness and a perfectly conducting sample (and with the half-space Green's function reformulation), the vector Kirchhoff theory gives the same angular factors \( q_{ijkl} \) that one would obtain from the Rayleigh–Rice vector theory [see Eqs. (28) for the values of angular factors in this limit]. In the case of scattering from finite conductors, however, the vector Kirchhoff theory does not have a complete solution that would relate the intensity and the PSD at any scattering angle.21

The relations can be simplified if we consider the in-plane scattering case, in which \( \theta_x = \theta_x \). That is, the parameters \( k_\perp \) (therefore \( \Omega \)), \( \theta_x \), and \( \theta_y \) are kept constant throughout the measurements. Then Eq. (10) also is simplified to the following result:

\[ S(k) \propto FI(k), \]

1 September 2000 / Vol. 39, No. 25 / APPLIED OPTICS 4661
where $F$ has the values

$$F = \begin{cases} \frac{1}{F_B^2} & \text{from scalar theory with } F_B = F_B(\phi_s, R) \\ 1/Q & \text{from vector theory with } Q = Q(\phi_s, \varepsilon) \end{cases}$$

(26)

One of our basic aims was to find a suitable prefactor that would link the experimentally obtained intensity data to the corresponding structure factor. If a correct prefactor is found, one can test the reliability of the calculated structure factor by examining some characteristic properties of the rough surface, such as the power-law behavior mentioned above. We can calculate the $\alpha$ value from the power-law behavior at the large $k_i$ regime and compare it with that obtained from other scanning probe techniques such as AFM and STM. (Note that one can calculate the surface parameters $w$, $\xi$, and $\alpha$ from the real space height–height correlation function obtained from AFM or STM.)

### 3. Experiments

To achieve large $k_i$ measurements, we designed an in-plane light-scattering setup as shown in Figs. 4 and Fig. 5(a). The incident light was a He–Ne laser with a wavelength of 632.8 nm. A spatial filter was used to improve the quality of the incident beam. A focusing lens was placed in front of and at the back of the spatial filter to focus the light entering the spatial filter and to collimate the light after the filter. A right circularly polarized incident light was achieved with a linear polarizer and a quarter-wave plate (right circular polarization is in the clockwise direction when viewed in the direction of propagation). Then the beam was sent through two apertures to reduce the stray light further. The beam was directed to the sample by two high-precision coated mirrors; we could achieve different incident angles by adjusting the position of the upper mirror [see Fig. 5(a)]. After two reflections from these two mirrors we still had right circularly polarized incident light.

The reflected specular beam from the sample was absorbed by a black hole. When necessary, neutral-density filters were also used to reduce the intensity of the incident light so that it would not saturate the photodetector. The detector that we used was a Si-photodiode detector with a 1.1 mm × 1.1 mm square-shaped detecting area. We kept the detector arm length (~0.5 m) long enough that we were able to reach a reasonable angular resolution (~0.12°). The detector arm was able to rotate 360° about the sample normal in the plane parallel to the sample surface. Also, the detector arm position was adjustable over the polar angle $\theta_s$, so one could set $\theta_i = \theta_s$, which is the condition for our in-plane measurements. The rotation was performed by a stepper motor. In addition, the stage was isolated by black-painted boards so that the measurement would not be affected by stray light scattered by other optical components.

The scattered intensity was measured with the help of lock-in techniques (see Fig. 4). The chopper provided reference signals for the two lock-in amplifiers that we used. The intensity of the scattered light from the sample was measured with the photodiode detector, which was placed at the end of the arm. We also placed another Si-photodiode detector adjacent (see Fig. 4) to the chopper blade, so we were able to monitor the intensity fluctuations of the incident beam. Therefore the scattered intensity could be readjusted by the incident intensity fluctuations simultaneously during the experiment, and the effects of incident intensity fluctuations were minimized. Data acquisition was controlled by a computer through the use of a LabView program.
The LabView program was capable of controlling the two lock-in amplifiers and the stepper motor through two RS232 serial ports and a National Instruments data-acquisition card, respectively. The detector arm could be rotated by use of the optional step angles of 0.1°, 0.2°, and 0.5°.

Before the data were taken, the whole system was optimized to give a maximum signal-to-noise ratio. Then we performed the in-plane measurements on the samples including two backside Si wafers and one electroplated Cu film. Using the configuration shown in Figs. 2(a) and 5(a), we swept the detector through angle $\phi_s$ from $-\pi$ to $\pi$. The detector arm was rotated with a step angle of 0.1°. At each step the computer program adjusted the lock-in amplifier to the highest sensitivity, and 10 readings were accumulated to improve the signal-to-noise ratio. The average intensity value was recorded, along with its standard deviation. We repeated the in-plane measurements with incident angles ($\theta_i$) of 48°, 75°, and 83° for the backside Si wafer and of 52° for Cu. Because of the saturation of intensity in the detector near the specular beam, we performed additional near-specular beam measurements with a neutral-density filter and multiplied the measured intensity data by the factor $1/\text{transmissivity of the filter}$.

For comparison, we also performed out-of-plane detection from the backside Si wafer, as shown in Fig. 5(b). As we discussed in Section 1, the only difference in this geometry from that for in-plane detection is that the detector rotates in the plane of incidence.

Real space measurements were also performed by the AFM technique. AFM scans were obtained with a Park Scientific Instruments Auto Probe CP with Si$_3$N$_4$ tips. The typical radius and the tip side angle were approximately 10 nm and 10°, respectively. The sampling AFM images for the backside Si wafer and the Cu film are shown in Figs. 6(a) and 6(b), respectively. The roughness parameters $\alpha$, $w$, and $\xi$ obtained from the height–height correlation function analysis are listed in Table 1.

### 4. Results and Discussions

#### A. In-Plane Diffraction Profiles

In Fig. 7 we show the measured in-plane intensity profiles of the backside Si sample for incident angles 48°, 75°, and 83°. As we discussed in Section 1, all three profiles are symmetric. The diffraction profiles have two components, for incident angles 75° and 83°, a very sharp central peak, and a broad diffuse profile, which is consistent with the prediction of Eq. (15). The central peak intensity for the 75° incident angle is smaller than that of the 83° incident angle because as $\theta_i$ decreases, $k_l$ increases, resulting in increasing $\Omega$ for the same rough surface. As $\theta_i$ decreases further, as for the case of $\theta_i = 48^\circ$, the sharp central peak of the profile disappears, and the diffuse profile becomes much broader than in the other two cases. These are the general features predicted by the diffraction theory.$^{10,13}$

#### B. Comparison of In-Plane and Out-Plane Diffraction Profiles

Figure 8 shows the diffraction profile obtained from the out-of-plane diffraction geometry for the same backside Si wafer at 75° incident angle. Clearly the diffraction profile obtained by the out-of-plane setup is highly asymmetric, as discussed in Section 1.

Because, for the out-of-plane geometry, both $k_\perp$ and $k_l$ change with the rotation angle, Eq. (16) becomes too complicated to be analyzed. It is only when $\Omega = (k_\perp w)^2 < 1$ (also $w/\xi < 1$) that Eq. (17) is valid. Then the change of $k_\perp$ acts only as an angular factor, which is separated from the PSD. Therefore one should have $w < \lambda$ to satisfy the $\Omega < 1$ condition. For example, during an out-of-plane measurement $k_\perp$ may take values as high as the order of $-4\pi/\lambda$; and, for $\lambda = 0.6328$ μm, we should have $w < 500$ Å to make sure that at least $\Omega < 1$.

However, for the in-plane geometry, because $\Omega$ is fixed for a fixed incident angle, the analysis is much simpler. For small $\Omega$, the diffuse part of the struc-

---

**Table 1. Calculated $\alpha$, $w$, and $\xi$ Values from AFM Measurements for a Backside Si Wafer and for an Electroplated Cu Film**

<table>
<thead>
<tr>
<th>Material</th>
<th>$\alpha$ (μm)</th>
<th>$w$ (μm)</th>
<th>$\xi$ (μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>0.89 ± 0.02</td>
<td>0.29 ± 0.01</td>
<td>3.75 ± 0.03</td>
</tr>
<tr>
<td>Cu</td>
<td>0.72 ± 0.01</td>
<td>0.0273 ± 0.0005</td>
<td>0.50 ± 0.01</td>
</tr>
</tbody>
</table>

1 September 2000 / Vol. 39, No. 25 / APPLIED OPTICS 4663
ture factor is simply proportional to the PSD of the surface. Also, with an in-plane geometry, the measurement of \( w \) does not require that \( w/\xi \ll 1 \) as long as \( w/y_1 \) is satisfied.\(^{10,22,23}\) Therefore the in-plane geometry can expand the measurement of the vertical roughness range to as high as \( w/l \) with a detector of good dynamic range and with large incident angles providing small \( k' \).\(^{24}\) By changing the incident angle, one can obtain the diffraction structure factor for different \( \Omega \) values.

C. Extraction of Diffraction Structure Factors for a Backside Si Wafer

To obtain the diffraction structure factor, we corrected the intensity profiles by using the correction factor \( F \) that was presented in relation (25) and Eq. (26). Both scalar and vector theories were applied to describe \( F \). Then, based on approximation (20), we extracted the roughness exponent \( \alpha \) from the tails of the structure factors. The extracted values of \( \alpha \) were then compared with that obtained from the AFM measurements. As we performed the intensity correction, the experimentally obtained front (flat) side Si wafer reflectivities \( (R = \{0.336, 0.322, 0.444\}) \) for \( \theta_i = \{48^\circ, 75^\circ, 83^\circ\} \) and theoretical complex dielectric constant value \( (\varepsilon = 15.07 + i0.15) \) were used.\(^{25}\)

Figure 9(a) shows the corrected in-plane diffraction profiles for the backside Si wafer at an incident angle of 83\(^\circ\) for both scalar and vector theories. Note that we have labeled the structure factors obtained from Rayleigh’s vector and Beckmann’s scalar theories, respectively, \( S_Q \) (curve b) and \( S_{FB} \) (curve c) are obtained from the Rayleigh vector and the Beckmann scalar theories, respectively. (b) In-plane light-scattering normalized intensity profile and normalized corrected structure factors for the scattering from an electroplated Cu film with a 52\(^\circ\) incident angle.

Fig. 7. In-plane intensity profiles of a Si backside for incident angles 83\(^\circ\), 75\(^\circ\), and 48\(^\circ\).

Fig. 8. Diffraction profile obtained from the out-of-plane diffraction geometry for the same backside Si wafer as for Fig. 7 at a 75\(^\circ\) incident angle.
of the profile at a large \(k_{||}\) region gives the \(\alpha\) value with a slope of \(-2 - 2\alpha\). As \(0 \leq \alpha \leq 1\), the value of the slope should be \(-2\) to \(-4\). Figure 10 shows that only the vector theory gives a reasonable slope value of approximately \(-3.84\).

Tables 1 and 2 list comparisons of the extracted \(\alpha\), \(w\), and \(\xi\) values obtained by AFM measurements and values from the in-plane light-scattering measurements for different incident angles. Roughness parameters \(w\) and \(\xi\) obtained from the diffraction profile agree well with the AFM measurements. We can also see that, for large incident angles, only the \(\alpha\) values obtained from the structure factors based on the vector theory \((S_{V})\) appear to be very close to the values obtained from the AFM analysis. At the large incident angle of 83°, which gives the best \(\alpha\) value from the vector theory analysis, we have \(\Omega = (k_{\perp}/w)^2 \approx 0.5\). This suggests that the \(\Omega \ll 1\) condition of relation (23) can be relaxed to \(\Omega < 1\). Zhao et al. reported a similar observation in their paper.\(^{10}\)

As \(\Omega\) gets larger (\(-1.5\) and \(-14.8\) for the incident angles of 75° and 48°, respectively), however, the vector theory also begins to fail.

We also checked the effect of multiple scattering and shadowing effects in our experiments for backside Si wafers. These effects can play a role especially in the large incident and scattering angle profiles predicted by the scalar theory.\(^{22,23}\) But even at \(\theta_i = 83°\) we have \(\tan^{-1}(w/\xi) \approx 4°\), which is still smaller than the grazing angle of \((90° - 83°) = 7°\). Therefore surface slopes are not deep enough to cause severe multiple scattering. We followed Wagner’s formula to estimate the shadowing effects.\(^{22}\) Again, even at the incident angle of 83°, the ratio of the apparent interface width to the true interface width is \(-0.99993\). That is to say, shadowing effects also can be neglected in our measurements. Therefore the failure of the scalar theory should be due to the large values of \(\Omega\) invalidating relation (21).

D. Scattering from a Cu Film

For scattering from a metal surface, one would expect the scalar and the vector theories to give similar results. Figure 9(b) shows the in-plane light-scattering intensity profile and the corrected structure factors for the scattering from an electroplated Cu sample with a 52° incident angle. We can see that the corrected structure factors obtained from the scalar and the vector theories are very close. In fact, because Cu is a good conductor, we find a reflectivity of \(R \approx 1\). Therefore, with \(\theta_i = \theta_s = \theta\), we get

\[
F_B^2 = \frac{1 + \cos^2 \theta - \sin^2 \theta \cos \phi_s}{2 \cos^2 \theta}
\]

for an in-plane setup.

In the limit of a perfect conductor, \(\epsilon\) approaches infinity, and polarization-dependent elements of \(Q\), defined in Eq. (A12) below, can be simplified as

\[
\begin{align*}
q_w &= -\cos \phi_s, \\
q_w &= (\sin \phi_s/\cos \theta), \\
q_{ps} &= -(\sin \phi_s/\cos \theta), \\
q_{pp} &= (\sin^2 \theta - \cos \phi_s)/\cos^2 \theta.
\end{align*}
\]

Because for circularly polarized light we have \(Q = |q_{ss}|^2 + |q_{pp}|^2 + |q_{ps}|^2 + |q_{pp}|^2\), a detailed calculation reveals that \(Q \propto F_B^2\), i.e., the scalar and the vector theories should give the same correction factor in the limit of a perfectly conducting surface.

Table 2 lists the \(\alpha\) values for Cu obtained from the diffraction tails after the angular corrections. They are quite close to the value obtained from AFM. From Table 2 we can see that the other roughness parameters, \(w\) and \(\xi\), obtained from the diffraction profile are also consistent with those in Table 1 obtained from the AFM measurements.

Table 2. Calculated \(\alpha\), \(w\), and \(\xi\) Values From the In-Plane Light-Scattering Measurements for a Backside Si Wafer and for an Electroplated Cu Film

<table>
<thead>
<tr>
<th>Light Scattering</th>
<th>Material</th>
<th>(\theta_i) (deg)</th>
<th>From (S_Q)</th>
<th>From (S_{FB})</th>
<th>(w) ((\mu)m)</th>
<th>(\xi) ((\mu)m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>48</td>
<td>2.10 ± 0.06</td>
<td>2.12 ± 0.03</td>
<td>0.20 ± 0.01</td>
<td>3.96 ± 0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.84 ± 0.02</td>
<td>1.22 ± 0.02</td>
<td>1.58 ± 0.02</td>
<td>0.0340 ± 0.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>0.92 ± 0.02</td>
<td>1.58 ± 0.02</td>
<td>0.0340 ± 0.0005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>52</td>
<td>0.73 ± 0.01</td>
<td>0.73 ± 0.01</td>
<td>1.10 ± 0.05</td>
<td>0.0340 ± 0.0005</td>
<td></td>
</tr>
</tbody>
</table>

1 September 2000 / Vol. 39, No. 25 / APPLIED OPTICS 4665
In conclusion, we have shown in detail the advantage of using the in-plane method for measuring light scattering for a rough surface. The momentum transfer perpendicular to the sample surface \( k_z \) is fixed, so the calculation of the structure factor from an intensity profile and therefore data analysis is much easier than for the out-of-plane geometry. Under the condition that \( w/\xi \ll 1 \), the in-plane geometry can expand the measurement of the vertical roughness range to as high as \( w \sim \lambda \).

Also, our experimental results show that both the scalar and the vector theories work well for a perfectly conducting and small roughness film. But, in the case of a much rougher semiconductor sample, the vector theory works better as long as we have \( \Omega = (k_z w)^2 < 1 \). None of the \( \alpha \) values obtained from the Beckman–Kirkhoff scalar theory is consistent with the AFM results. The shadowing and multiple scattering effects are expected to be insignificant. Therefore the scalar theory is believed to fail because the value of \( \Omega \) at which relation (21) becomes invalid is insufficiently small.

Appendix A

Barrick, in his work on radar cross sections for rough surfaces,\(^{26}\) presented the polarization dependence of the scattering matrix elements. The scattering matrix is a tool with which incident and scattered field vector components can be related and is represented as follows:

\[
\begin{bmatrix}
E_s^1 \\
E_s^2
\end{bmatrix} = [S]
\begin{bmatrix}
E_i^1 \\
E_i^2
\end{bmatrix},
\]

where \( E_1 \) and \( E_2 \) are the complex electric field components along the orthogonal basis vector directions. The scattering matrix has the elements

\[
[S] = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix},
\]

where each of the \( a_{ij} \) elements above is a complex number (has both amplitude and phase) and relates the \( j \)th component of the incident field vector to the \( i \)th component of the scattered field vector.

The most familiar vector basis is the one formed by unit vectors \( \hat{p} \) and \( \hat{s} \), where \( \hat{p} \) lies in a plane containing the \( z \) axis and the propagation vector \( \mathbf{k} \) (see Fig. 1) and \( \hat{s} \) lies in the plane that is perpendicular to the \( \hat{p} \) plane. Any electric field vector can be represented on this basis as

\[
\mathbf{E} = E_s \mathbf{s} + E_p \mathbf{p} = E_s \hat{s} + E_p \exp(i\delta) \hat{p},
\]

where \( \delta \) is the time phase angle by which \( \mathbf{E}_s \) leads \( \mathbf{E}_p \). On this basis of linearly polarized field vectors, we can rewrite Eq. (A1) as

\[
\begin{bmatrix}
E_s^1 \\
E_s^2
\end{bmatrix} = \begin{bmatrix}
a_{ss} & a_{sp} \\
a_{ps} & a_{pp}
\end{bmatrix}
\begin{bmatrix}
E_i^1 \\
E_i^2
\end{bmatrix}.
\]

Right and left circularly polarized unit basis vectors \( \mathbf{R} \) and \( \mathbf{L} \), respectively, also form an orthogonal set (here we use the Institute of Electrical and Electronics Engineers standards, in which right circular polarization is in the clockwise direction as viewed in the direction of propagation and left circular polarization is in the counterclockwise direction). We can represent any electric field on this basis as

\[
\mathbf{E} = \mathbf{E}_R + \mathbf{E}_L = \mathbf{E}_R \hat{R} + \mathbf{E}_L \exp(i\Psi) \hat{L},
\]

where \( \Psi \) is the time phase angle by which \( \mathbf{E}_L \) leads \( \mathbf{E}_R \). Therefore, on the basis of circularly polarized field vectors, we can rewrite Eq. (A1) as

\[
\begin{bmatrix}
E_R^s \\
E_S^s
\end{bmatrix} = \begin{bmatrix}
a_{RR} & a_{RL} \\
a_{LR} & a_{LL}
\end{bmatrix}
\begin{bmatrix}
E_R^i \\
E_S^i
\end{bmatrix}.
\]

Matrix notation provides a convenient representation of the transformations of a wave between the linearly and the circularly polarized systems. These transformations are

\[
\begin{bmatrix}
E_R \\
E_S
\end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & -i \\
1 & +i
\end{bmatrix}
\begin{bmatrix}
E_s \\
E_p
\end{bmatrix},
\]

\[
\begin{bmatrix}
E_s \\
E_p
\end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 \\
+1 & -i
\end{bmatrix}
\begin{bmatrix}
E_R \\
E_L
\end{bmatrix}.
\]

Also, the individual matrix elements of the circular scattering matrix are related to the linearly polarized elements by

\[
a_{RR} = \frac{-a_{pp} + a_{ss} - i(a_{sp} + a_{ps})}{2},
\]

\[
a_{RL} = \frac{+a_{pp} + a_{ss} + i(a_{sp} - a_{ps})}{2},
\]

\[
a_{LR} = \frac{+a_{pp} + a_{ss} - i(a_{sp} - a_{ps})}{2},
\]

\[
a_{LL} = \frac{-a_{pp} + a_{ss} + i(a_{sp} + a_{ps})}{2}.
\]

The scattering cross section \( \sigma \) is defined to be \( 4\pi \) times the power delivered per unit solid angle in the direction of the detector divided by the power per unit area incident at the target surface. Therefore we represent the scattering cross section for an \( l \)-polarized incident and a \( k \)-polarized scattered wave as

\[
\sigma_{kl} = 4\pi r^2 \left( \frac{E_k^* \cdot E_l^{s*}}{E_l^* \cdot E_l^{s*}} \right),
\]

where \( r \) is the distance from target surface to detector. Scattering cross-section elements are related to those of scattering matrix \([S]\) by

\[
\sigma_{kl} = 4\pi r^2 |a_{kl}|^2.
\]

Barrick has used the Rice model to obtain the scattering cross section for the diffuse part of the scattered wave from a rough surface. The model is a perturbation technique for far fields (large \( r \)) and assumes that the roughness is small \( (\Omega < 1) \), the surface slopes are relatively small \( (w/\xi < 1) \), and the roughness is isotropic. For an incident beam that
has circular shape and uniform intensity, the diffuse part of the average scattering cross section per unit illuminated surface area is (see Fig. 1 for the angles of scattering)

\[
\gamma_{kl}^{\text{diff}} = \frac{\langle \sigma_{kl}^{\text{diff}} \rangle}{(A/\cos \theta_s)} = 4\pi \left(\frac{2\pi}{\lambda}\right)^4 \cos^2 \theta_i \cos^2 \theta_s |q_{kl}|^2 \text{PSD}(k_s), \tag{A11}
\]

where \(A\) is the illuminated area when the incident beam comes normally to surface, \(\lambda\) is the wavelength of the wave, and \(k_s = (k_x^2 + k_y^2)^{1/2}\), where \(k_x\) and \(k_y\) are defined in Eq. (2). Scattering angles \(\theta_i\), \(\theta_s\), and \(\phi_s\) are shown in Fig. 1. The quantity \(q_{kl}\) is directly proportional to the scattering matrix elements. Therefore, after we know all \(q_{kl}\) for a particular type of polarization system (e.g., linear or circular), we can calculate them for the other type of polarization system through transformations given in Eqs. (A8). For the case of a linear polarization system (e.g., \(s\)- and \(p\)-polarization states of incident wave and detector sensitivity) and for a nonmagnetic target surface (\(\mu = 1\)), Barrick’s results give

Later we can find the magnitudes of complex scattering matrix elements \(a_{kl}\) from Eq. (A10). After we know \(a_{kl}\), we can set \(q_{kl} = |a_{kl}| \exp(i\phi_{kl})\) and calculate the magnitudes of scattered fields from Eq. (A6) and by use of the definition \(|\mathbf{E}_k|^2 = (\mathbf{E}_k^* \cdot \mathbf{E}_k)^{1/2}\). The last equation that we get relates incident and scattered field magnitudes through the incident and scattering angles, wavelength of the beam, PSD of the surface, and polarization- and permittivity-dependent quantities. Because the experimentally measured intensity is proportional to the square of the field amplitude, after some algebra we find that

\((i)\) For the right circularly polarized incident light \((\mathbf{E} = E_R \mathbf{L})\) with intensity \(I \propto |E_R|^2\), the diffuse part of the measured scattered intensity at constant \(r\) is proportional to

\[
I'(\mathbf{k}) \propto |E_{RR}|^2 + |E_{LR}|^2 \propto I'Q_R \cos^2 \theta_s \cos \theta \lambda^4 \text{PSD}(k_s),
\]

where \(Q_R = |q_{ss}|^2 + |q_{pp}|^2 + |q_{sp}|^2 + |q_{ps}|^2\), and \(q_{kl}\) is given by Eqs. (A12).

\[
q_{ss} = \frac{(1 - \varepsilon) \cos \phi_s}{(\cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i})(\cos \theta_s + \sqrt{\varepsilon - \sin^2 \theta_s})},
\]

\[
q_{sp} = \frac{(\varepsilon - 1) \sqrt{\varepsilon - \sin^2 \theta_i} \sin \phi_s}{(\cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i})(\cos \theta_s + \sqrt{\varepsilon - \sin^2 \theta_s})},
\]

\[
q_{ps} = \frac{\varepsilon - 1)(\sqrt{\varepsilon - \sin^2 \theta_i} \sin \phi_s}{(\cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i})(\cos \theta_s + \sqrt{\varepsilon - \sin^2 \theta_s})},
\]

\[
q_{pp} = \frac{(\varepsilon - 1)(\sqrt{\varepsilon - \sin^2 \theta_i} \sin \phi_s)}{(\cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i})(\cos \theta_s + \sqrt{\varepsilon - \sin^2 \theta_s})},
\]

A12

where \(\varepsilon\) is the relative permittivity of the surface. Barrick also gives the elements \(q_{kl}\) for the circular polarization system (e.g., \(R\)- and \(L\)-polarization states of incident wave and detector sensitivity), which are related to linear polarization elements by

\[
q_{RR} = \frac{-q_{pp} + q_{ss} + i(q_{sp} + q_{ps})}{2},
\]

\[
q_{RL} = \frac{-q_{pp} + q_{ss} - i(q_{sp} - q_{ps})}{2},
\]

\[
q_{LR} = \frac{-q_{pp} + q_{ss} + i(q_{sp} - q_{ps})}{2},
\]

\[
q_{LL} = \frac{-q_{pp} + q_{ss} - i(q_{sp} + q_{ps})}{2}. \tag{A13}
\]

Therefore we can substitute elements of Eqs. (A13) into Eq. (A11) to get the elements of \(\langle \sigma_{kl}^{\text{diff}} \rangle\).

\((ii)\) For the left circularly polarized incident light \((\mathbf{E} = E_L \mathbf{L})\) with intensity \(I \propto |E_L|^2\), the diffuse part of the measured scattered intensity at constant \(r\) is proportional to

\[
I'(\mathbf{k}) \propto |E_{RL}|^2 + |E_{LR}|^2 \propto I'Q_L \cos^2 \theta_s \cos \theta \lambda^4 \text{PSD}(k_s),
\]

where \(Q_L = |q_{ss}|^2 + |q_{pp}|^2 + |q_{sp}|^2 + |q_{ps}|^2\), which is just equal to \(Q_R\) in relation (A14).

Therefore the scattered wave intensity has the same polarization-dependent factor \((Q = Q_L = Q_R)\) whether the incident beam is right or left circularly polarized.

This study was supported by National Science Foundation grant 9971265; M. Stowe and B. Quayle
were supported by a Research Experience for Undergraduates grant.

References and Note
24. One can calculate $w$ from a single in-plane structure factor profile by using the relation $R_S = f d^2 (k_x, k_y)/(d^2 k_x k_y) = \exp(-\Omega)$. But see Ref. 10 and Chap. 3 of Ref. 13 for the detailed methods of $w$ and calculation from a structure factor.