Two-dimensional optics : diffraction and dispersion of surface plasmons

Chimento, P.F.

2

A subwavelength slit as a quarter-wave retarder

We have experimentally studied the polarization-dependent transmission properties of a nanoslit in a gold film as a function of its width. The slit exhibits strong birefringence and dichroism. We find, surprisingly, that the transmission of the polarization parallel to the slit only disappears when the slit is much narrower than half a wavelength, while the transmission of the perpendicular component is reduced by the excitation of surface plasmons. We exploit the slit's dichroism and birefringence to realize a quarter-wave retarder.

2.1 Introduction

This chapter was previously published as: Chimento, Kuzmin, et al. (2011).

The study of the transmission of light through small perforations in metal films has a venerable history[1] and has important applications in the field of optical data storage.[2] It dates back to the middle of the nineteenth century when Fizeau described the polarizing effect of wedge-shaped scratches in such films.[3]

[1] Lord Rayleigh, 1897; Bethe, 1944; Bouwkamp, 1954; Jones and Richards, 1954.

[2] Bouwhuis et al., 1985.

[3] Fizeau, 1861.

This field has recently come back to center stage following the observation that, at a specific set of wavelengths, the transmission of a thin metal film containing a regular two-dimensional array of subwavelength apertures is much larger than elementary diffraction theory predicts.[4] This phenomenon of extraordinary optical transmission, which is commonly attributed to surface plasmons traveling along the corrugated interface, has spawned many studies of thin metal films carrying variously-shaped corrugations and perforations. These include holes with circular, cylindrical, or rectangular cross sections,[5] either individually or in arrays, and elongated slits.[6] The polarization of the incident light is an important parameter, in particular when the width of the hole or slit is subwavelength in one or both directions. The case of a slit which is long in one dimension

[4] Ebbesen et al., 1998.

[5] Astilean, Lalanne, and Palamaru, 2000.

[6] Takakura, 2001; Yang and Sambles, 2002; Suckling et al., 2004.

and subwavelength in the other seems particularly simple, as elementary waveguide theory predicts that it acts as a perfect polarizer when the slit width is less than about half the wavelength of the incident light.

TE: transverse electric
TM: transverse magnetic

For infinitely long slits, one can define TE and TM polarized modes. The TM mode's electric field vector is perpendicular to the long axis of the slit, and the TE mode has its electric field vector parallel to the long axis. In standard waveguide models, the metal is usually assumed to be perfect, so that the continuity equation for the electric field implies that its parallel component must be zero at the metallic boundaries. In a slit geometry, this implies that TE-polarized light incident on such a slit will not be transmitted by the structure if the wavelength λ of the incident light is larger than twice the slit width w. This width is commonly referred to as the cutoff width. The TM-polarized mode, on the other hand, can propagate unimpeded through the slit, the effective mode index increasing steadily as the width is reduced.[7] For this reason one expects very narrow slits in metal films to act as perfect polarizers.[8]

[7] Astilean et al., 2000; Takakura, 2001.
[8] Fizeau, 1861.

While the perfect metal model is an excellent approximation for wavelengths in the mid to far infrared or microwave domains, the model is too naïve when the wavelength of the incident light is smaller, because of the dispersion in the permittivity of metals. As a consequence, in the visible part of the spectrum the TE mode cutoff width of real metals like silver and aluminum is slightly smaller than $\lambda/2$,[9] and the cutoff is more gradual. Although the TM mode propagates through the slit, it couples to surface plasmon modes on the front and back surfaces of the slit,[10] which act as a loss channel. Since these losses are heavily dependent on the slit width,[11] the transmitted intensity of the TM mode is more dependent on this width than the perfectly conducting waveguide model predicts.

[9] Schouten, Visser, Lenstra, and Blok, 2003; Schouten, Visser, Gbur, Lenstra, and Blok, 2004.
[10] Schouten, Kuzmin, et al., 2005.
[11] Lalanne, Hugonin, and Rodier, 2006; Baudrion et al., 2008; Kihm, Lee, Kim, Kang, and Park, 2008.

Here we demonstrate that, for thin metal films, such a nanoslit also acts as an optical retarder, and that the TE/TM transmission ratio is around unity well below the cutoff width, approaching zero only when the slit is extremely narrow. We have employed these properties to turn such a slit into a quarter-wave retarder.

2.2 *Description of experiment*

In the experiment, shown schematically in Fig. 2.1, we illuminate an array of ten 10 μm by 50–500 nm slits with a laser beam at λ = 830 nm, at normal incidence (see Fig. 2.2.) For all practical purposes, each slit's length can be considered infinite compared to its width and the laser wavelength. The slits are milled through a 200 nm thick gold film

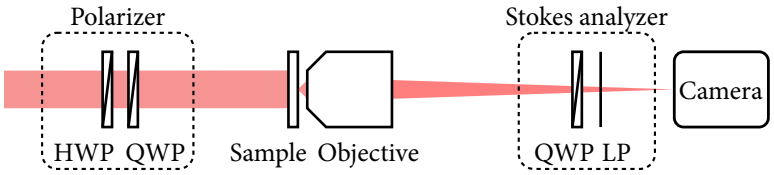

Figure 2.1: Sketch of the experimental setup. HWP: half-wave plate, QWP: quarter-wave plate, LP: linear polarizer. The sample (see Fig. 2.2) is illuminated on the gold side, using light with a controlled polarization. The transmitted light's polarization is analyzed for each pixel of a CCD camera. The Stokes analyzer consists of a quarter-wave plate and a linear polarizer, which can be rotated independently of each other under computer control to any desired orientation.

using a focused Ga^+ ion beam. The slits' widths increase stepwise from 50 nm, well below the cutoff width for TE-polarized light, to 500 nm, at which value the lowest TE mode can propagate through the slit. The film is deposited on a 0.5 mm thick Schott D263T borosilicate glass substrate, covered by a 10 nm titanium adhesion layer which damps surface plasmons, ensuring that their propagation length is negligibly short on the gold-glass interface. The laser beam width at the sample is approximately 4 mm so that, effectively, the structure is illuminated homogeneously with a flat wavefront. The light transmitted by the structure is imaged on a CCD camera (Apogee Alta U1) by means of a 0.65 NA microscope objective. The polarization of the light incident on the structure is controlled by a combination of half-wave and quarter-wave plates, enabling us to perform the experiment with a variety of input polarizations.

CCD: charge-coupled device
NA: numerical aperture

We analyze the polarization by measuring the Stokes parameters of the light transmitted through each slit, using a quarter-wave plate and a linear polarizer. We define the Stokes parameters according to the following standard convention: S_0 is the total intensity, S_1 is the intensity of the horizontal component (TE) minus the intensity of the vertical component (TM), S_2 is the intensity of the diagonal (45° clockwise) component minus the intensity of the anti-diagonal (45° counterclockwise) component, and S_3 is the intensity of the right-handed circular component minus the intensity of the left-handed circular component. Since the transmitted light is fully polarized, it is convenient to use the *normalized* Stokes parameters $s_1 = S_1/S_0$, $s_2 = S_2/S_0$, and $s_3 = S_3/S_0$, so that each ranges from −1 to +1.

$$
\begin{aligned}
S_0 &= I_{\text{total}} \\
S_1 &= I_H - I_V \\
S_2 &= I_D - I_A \\
S_3 &= I_R - I_L
\end{aligned}
$$

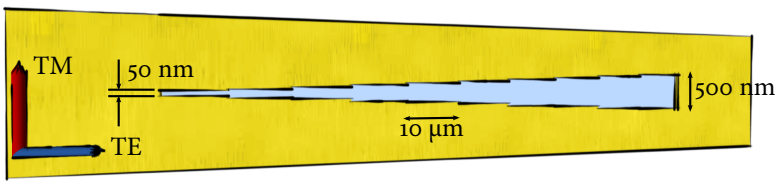

Figure 2.2: Sketch of the sample. It consists of a 200 nm gold film sputtered on top of a glass substrate. Note that the vertical scale is greatly exaggerated compared to the horizontal scale. Adapted from Kuzmin (2008, p. 78).

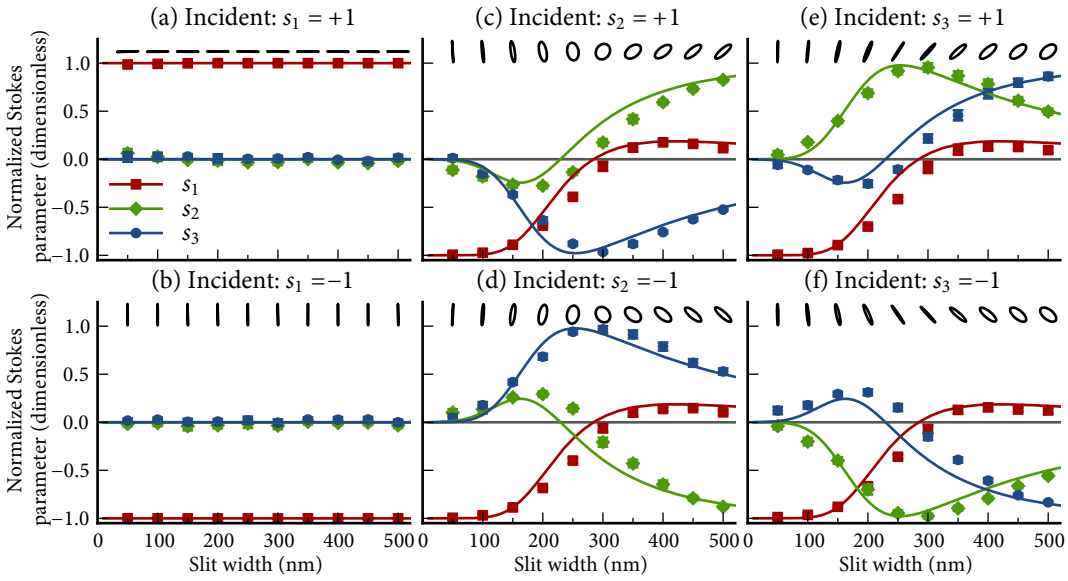

Figure 2.3: Normalized Stokes parameters of the light transmitted through the slit, for illumination with (a) horizontal linear polarization ($s_1 = +1$), (b) vertical linear polarization ($s_1 = -1$), (c) diagonal linear polarization ($s_2 = +1$), (d) antidiagonal linear polarization ($s_2 = -1$), (e) left-handed circular polarization ($s_3 = +1$), and (f) right-handed circular polarization ($s_3 = -1$). The polarization ellipses above each graph provide a quick visual indication of the polarization state of the transmitted light. The solid lines represent the results of our model, described later on in section 2.4, based on simple waveguide theory.

2.3 *Results and interpretation*

THE FULL STOKES ANALYSIS of the transmitted light, for each of the six basic Stokes input polarizations ($s_{1,2,3} = \pm 1$), is shown in Fig. 2.3. Figures 2.3a and 2.3b confirm that the TE and TM directions are the slit's eigenpolarizations. However, each has its own damping and propagation constant, as we will show. In the general case, a slit is therefore both dichroic and birefringent, both properties depending on the slit width w.

Figures 2.3c–f show the variation of the Stokes parameters of the transmitted light when the incident light is not polarized along one of the slits' eigenpolarizations. In all cases, s_1 is seen to go to -1 as the slit gets narrower, reflecting the fact that very narrow slits transmit only TM-polarized light.

Let us examine Figs. 2.3c–d more closely, where the incident wave is diagonally linearly polarized ($s_2 = \pm 1$). As the slit width w is reduced from 500 to 300 nm, the transmitted light gradually becomes more and more elliptically polarized, while the main axis of the polarization ellipse remains oriented along the polarization direction of the incident light; see the line of polarization ellipses in each frame. As w is reduced further to around 250 nm, the transmitted polarization assumes a more circular form. For narrower slits, the polarization ellipse orients itself essentially vertically, reflecting the fact that the polarization becomes more linear, ultimately

being purely TM-polarized at $w = 50$ nm. In Figs. 2.3e–f, with circular input polarization, a similar process happens as w is reduced, except that the transmitted polarization changes gradually from almost circular to linear, before becoming nearly TM-polarized at $w = 50$ nm.

We note that there is a point in Figs. 2.3e–f, around $w \approx 250$ nm, where circular polarization is transformed into linear polarization. This implies that the slit acts as a quarter-wave retarder, albeit with unequal losses for the fast and slow axes. Because of the inequality of these losses, the incident diagonal polarization in Figs. 2.3c–d is not transformed into a perfectly circular polarization. However, a properly oriented linear polarization incident on a $w \approx 250$ nm slit whose orientation compensates for the differential loss, *will* be transformed into circular polarization. Experiments on other slits have shown that the measured dichroism is highly dependent on the slit parameters, such as milling depth,[12] and the incident wavelength.[13] Realizing an ideal quarter-wave retarder therefore requires either careful design and manufacture of the slit, or serendipity.

[12] Bosman, 2011, p. 13.

[13] Bogers, 2012, p. 30.

As expected, the curves of s_2 and s_3 as a function of w flip their sign when the sign of the incident Stokes parameter is flipped. When the incident light's s_2 and s_3 are exchanged, on the other hand, so are s_2 and s_3 in the transmitted light. The curve of s_1 remains the same for all non-s_1 incident polarizations. The results shown in Fig. 2.3 can all be represented in one figure by plotting the measured Stokes parameters on the Poincaré sphere. Reducing the slit width then traces out a path of the transmitted polarization state over the Poincaré sphere's surface, as shown in Fig. 2.4.

In order to analyze our experimental data, we write the incident field as a Jones vector, preceded by an arbitrary complex amplitude such that the TE component is real and positive:

$$\mathbf{E}_{\text{in}} = \tilde{A} \begin{bmatrix} E_{\text{TE}} \\ E_{\text{TM}}\, e^{i\psi} \end{bmatrix}, \qquad \text{with } E_{\text{TE}}, E_{\text{TM}} \geq 0. \tag{2.1}$$

We express the transmission properties of the slit as a Jones matrix. Its off-diagonal elements are zero, because the TE and TM directions are the slit's eigenpolarizations, and the diagonal elements represent the complex amplitude transmission. The output field is then the Jones vector:

$$\mathbf{E}_{\text{out}} = \begin{bmatrix} t_{\text{TE}} & 0 \\ 0 & t_{\text{TM}} \end{bmatrix} \mathbf{E}_{\text{in}}. \tag{2.2}$$

First, it is instructive to calculate the transmission T_{TE} and T_{TM} in order to get an idea of the slit's dichroism. Here, we define the transmission $T = |t|^2$

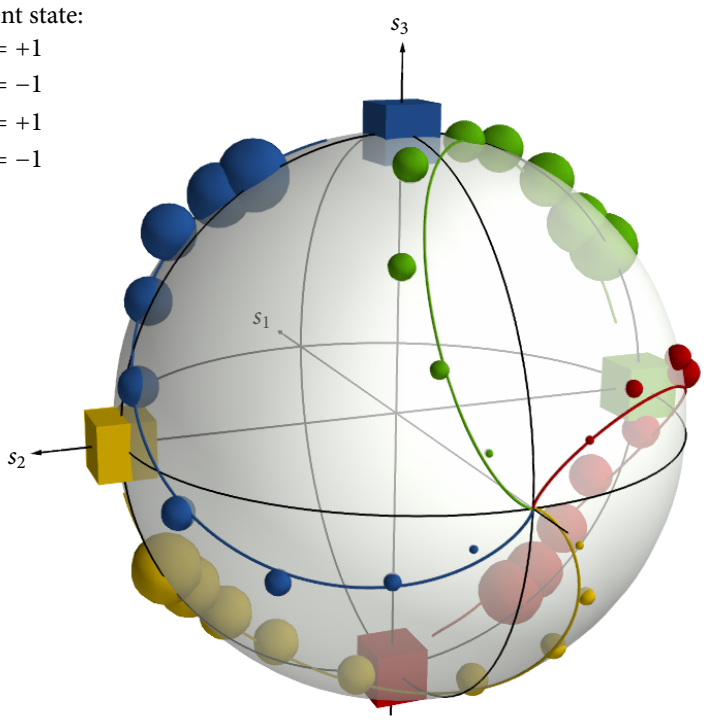

Figure 2.4: Path of the transmitted polarization state over the Poincaré sphere as the slit width decreases. The incident polarization state starts at one of the poles or equatorial points, represented by the boxlike markers. The spherical markers, with size proportional to the slit width, mark the transmitted polarization state as it travels over the sphere's surface. The solid lines are the predictions of our model.

as the ratio of power emerging from a slit to power incident on the area of the slit. It can be calculated from the unnormalized Stokes parameter S_1 for incident light with $s_1 = \pm 1$. T_{TE} and T_{TM} are plotted in Fig. 2.5, normalized so that $T_{\mathrm{TE}} = 1$ at $w = 500$ nm. As the slit width w is decreased, we see that the TE and TM transmission also decrease until $w \approx 350$ nm. When w is further reduced, the TM transmission goes through a minimum at $w \approx 150$ nm, where the light-surface plasmon coupling is maximum.[14] It increases again when the slit width gets even smaller, whereas the TE transmission goes through a gradual cutoff, becoming negligible only for the narrowest slits. Apparently, a narrow slit in a *thin* metal film is not such a good polarizer as often assumed.

[14] Lalanne et al., 2006.

In order to calculate the phase lag $\Delta\varphi$ between the TE and TM-polarized components of the transmitted field, we write the normalized Stokes

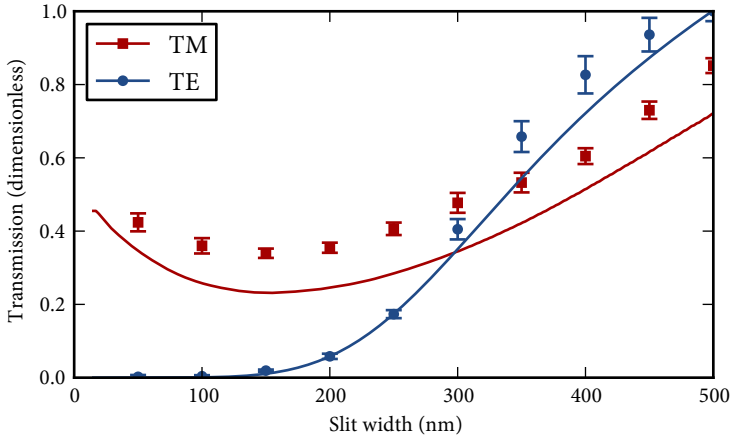

Figure 2.5: Dichroism of a subwavelength slit. The points show the measured transmission for TM and TE-polarized incident light as a function of the slit width w, normalized to the TE transmission at $w = 500$ nm. The solid lines show our model's result for the slit transmission according to (2.7) and (2.8).

parameters in terms of (2.2):

$$s_1 = -\frac{T_R E_{\mathrm{TM}}^2 - E_{\mathrm{TE}}^2}{T_R E_{\mathrm{TM}}^2 + E_{\mathrm{TE}}^2}, \tag{2.3}$$

$$s_2 = \frac{2\sqrt{T_R} E_{\mathrm{TM}} E_{\mathrm{TE}}}{T_R E_{\mathrm{TM}}^2 + E_{\mathrm{TE}}^2} \cos(\Delta\varphi - \psi), \tag{2.4}$$

$$s_3 = -\frac{2\sqrt{T_R} E_{\mathrm{TM}} E_{\mathrm{TE}}}{T_R E_{\mathrm{TM}}^2 + E_{\mathrm{TE}}^2} \sin(\Delta\varphi - \psi), \tag{2.5}$$

where $T_R = |t_{\mathrm{TM}}/t_{\mathrm{TE}}|^2$ is shorthand for the transmission ratio, E_{TM} and E_{TE} are the transmitted fields, and ψ is the TM–TE phase lag; see (2.1). We calculate $\Delta\varphi$ from our measured Stokes parameters using (2.3), (2.4), and (2.5). We see in Fig. 2.6 that $\Delta\varphi$ decreases almost linearly with increasing slit width. It passes through a value of $\pi/2$ at $w \approx 250$ nm. Although the retardation equals $\lambda/4$, the 250 nm slit does not act as an ideal quarter-wave retarder because the amplitudes of the TE and TM-polarized components of the transmitted light are not equal, as noted earlier.

Figure 2.5 illustrates the slit's dichroism and Fig. 2.6 its birefringence. The effect that we observe in Fig. 2.3 as the slit width is decreased from 500 to 300 nm can be explained in terms of increasing birefringence and small dichroism in that range. Below 300 nm, dichroism becomes more important, and consequently, the main axis of the polarization ellipse rotates. The dichroism observed here was also suggested by calculations by Nugrowati, Pereira, and van de Nes,[15] where ultrashort TE pulses were shown to experience lower propagation speeds than TM pulses through a slit in an aluminum layer.

If the slit width is further decreased past the surface plasmon-induced

[15] Nugrowati et al., 2008.

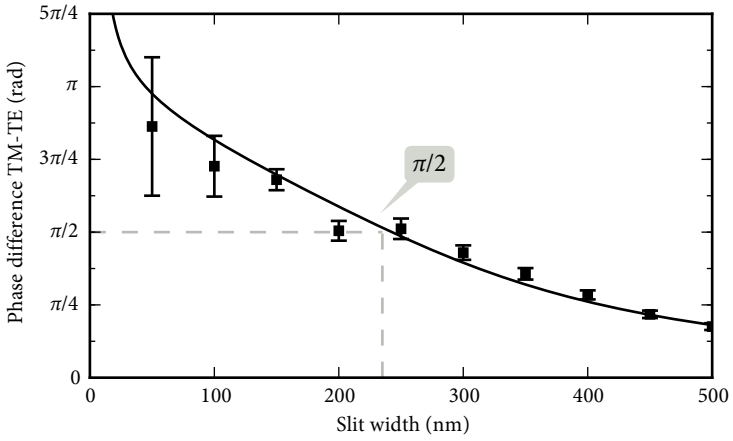

Figure 2.6: Birefringence of a subwavelength slit. The points represent the measured phase difference between the TM and TE modes as a function of the slit width. They are obtained from a fit of the various Stokes parameters of Fig. 2.3. The solid line shows the calculated phase difference according to (2.9). At a certain slit width, indicated by the arrow, the phase difference reaches $\pi/2$ and the slit acts as a quarter-wave retarder.

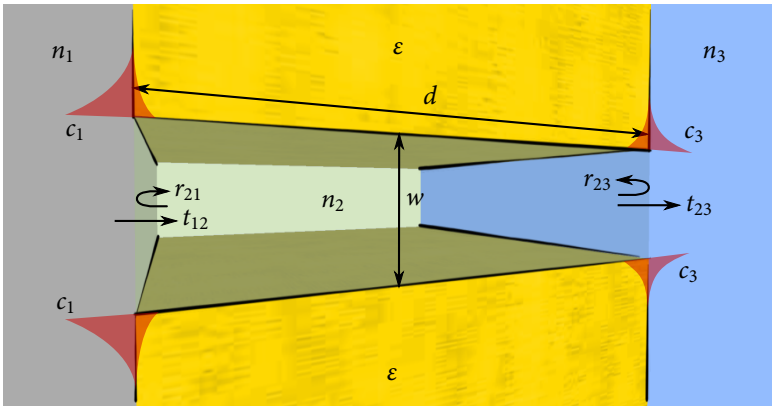

Figure 2.7: Cross-section of our model slit. The relevant physical quantities are illustrated. The indices of refraction $n_{1,2,3}$ are depicted using differently colored materials, although they could well be the same material in an experiment. The evanescent tails sketched in red represent surface plasmons. The light is transmitted through the slit from left to right.

minimum at $w \approx 150$ nm, the dichroic effect becomes even larger. The TE-polarized component of the transmitted light becomes weaker and weaker, while the TM component grows, causing the polarization ellipse to collapse to a vertical line. We see that the waveguide's TE cutoff does not resemble a sharp cutoff at $w = \lambda/2$ at all, but rather a gradual one.

2.4 *Waveguide model*

We have made our computer code for this model available (Chimento, 2013b).

WE WILL NOW PROCEED to explain these experimental results by modeling the slit as a simple lossy waveguide. Our metallic slit forms a rectangular waveguide with one dimension of the rectangle much larger than the other. For that reason we can effectively describe each slit as a step-index planar waveguide, with its walls made of a metal with relative per-

mittivity ε. Inside the waveguide, the solutions to Maxwell's equations separate into TE and TM modes, each with a complex propagation constant β. Although the equations are in closed form, we must calculate the propagation constants for each mode, β_{TE} and β_{TM}, numerically.[16]

[16] Snyder and Love, 1983.

For the TM and TE modes, we calculate complex reflection and transmission coefficients r_{21}, t_{12}, r_{23}, and t_{23} (see Fig. 2.7) using the Fresnel equations at normal incidence, substituting the effective mode index for the index of medium 2. The effective mode index is calculated by dividing the propagation constant by k_0. As shown in Fig. 2.7, the index 1 indicates the medium from which the light is incident (air), 2 the waveguide, and 3 the medium into which the transmitted light emerges (glass in our experiment). This simplification avoids calculating overlap integrals between the guided mode and the modes outside the waveguide, but still describes the observed phenomena quite well. We can then treat the waveguide as a Fabry-Pérot interferometer and calculate each mode's complex transmission through a waveguide of length d,

$$t_{123} = \frac{t_{12}t_{23}e^{i\beta d}}{1 - r_{21}r_{23}e^{2i\beta d}}, \tag{2.6}$$

which gives for the transmission

$$T_{\mathrm{TE}} = \frac{n_3}{n_1}|t_{123}^{\mathrm{TE}}|^2, \tag{2.7}$$

$$T_{\mathrm{TM}} = \frac{n_3}{n_1}|t_{123}^{\mathrm{TM}}|^2 - 2|c_1|^2 - 2|c_3|^2. \tag{2.8}$$

Here, c_1 and c_3 are the coupling constants of the slit system to a surface plasmon mode traveling in one direction away from the slit on the interface with medium 1 or 3, respectively. Numerical values for these parameters can be calculated[17] using Eq. (20) of Lalanne et al.,[18] which gives an approximate analytical model for the coupling of a slit mode to a surface plasmon mode. As an illustration of the important role these surface plasmon coupling constants play in the phenomenon described here, the TM transmission modelled with and without coupling to surface plasmons is shown in Fig. 2.8. The TE mode does not couple to surface plasmons.

[17] Calculating c_1 and c_3 requires evaluating an integral with poles close to the real axis. Common adaptive quadrature algorithms for numerical integration cannot handle it, yielding a garbage answer without obviously failing. Gaussian quadrature works for numerically evaluating the integral.

[18] Lalanne et al., 2006.

[19] Lalanne et al., 2006.

It is interesting to note in Fig. 2.8 that the surface plasmon coupling coefficients on both sides exhibit a maximum at $nw/\lambda \approx 0.23$ and a minimum at $nw/\lambda \approx 1$, as predicted by Lalanne et al.,[19] where n is the index of refraction of the medium outside the slit on each respective side. These two curves added together yield a maximum in the surface plasmon excitation, and therefore a dip in the TM transmission, at around $w \approx 150$ nm. Even though this dip is not at $nw/\lambda \approx 0.23$ as Lalanne et al. predict, it is caused by two plasmon excitation maxima that do follow the prediction.

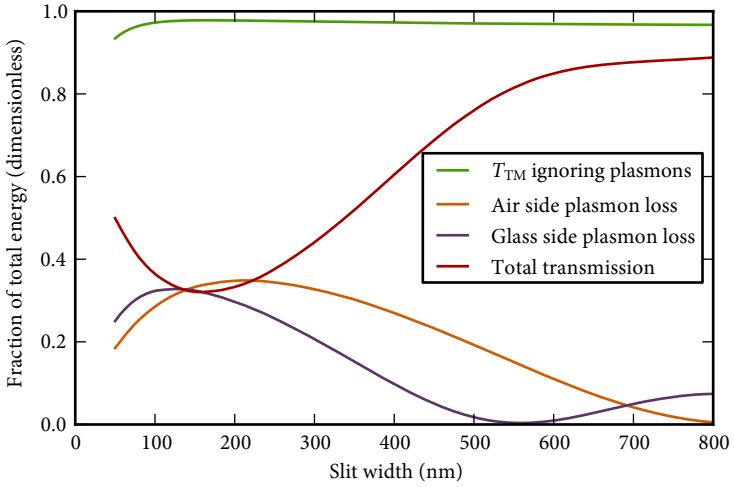

Figure 2.8: Calculated effect of surface plasmons on the transmission of TM-polarized light as a function of the slit width w. The green line shows the calculated TM transmission neglecting coupling to surface plasmons, based on waveguide theory alone, i.e. $(n_3/n_1)|t_{123}^{TM}|^2$. The orange line shows the total fraction of energy $2|c_1|^2$ converted to surface plasmons on the illuminated (air) side of the sample according to Lalanne et al. (2006). Likewise, the purple line shows the fraction $2|c_3|^2$ converted to surface plasmons on the unilluminated (glass) side. Finally, the red line shows the total TM transmission according to (2.8). In these calculations, we disregard the numerical aperture of the imaging system.

In our model we ignore the thin titanium adhesion layer present between the gold and the glass. According to the model, the $|c_3|$ coefficient for a thick titanium layer would be slightly higher than that of the gold layer. However, we expect that the layer is too thin to have any effect on the coupling between the slit TM mode and surface plasmons. It does not prevent the light from scattering into the surface plasmon mode, but only ensures that the surface plasmon mode is very lossy.

Our model exhibits good agreement with the measurements, despite the fact that it does not contain any fitting parameters. The slit's gradual TE cutoff is predicted well, and can be ascribed to gold not being a perfect conductor at this wavelength, and to the considerable dispersion of the reflection coefficients r_{12} and r_{23} around cutoff. The model also predicts a plasmon-related TM transmission dip at the right slit width. In Fig. 2.5, we compare these calculated values to our measurements. In our calculations, we took the finite NA and its influence on the TM and TE transmission into account, which is explained in section Appendix 2.A.

The complex transmission also gives us the relative phase delay between the TM and TE modes:

$$\Delta\varphi = \arg t_{123}^{TM} - \arg t_{123}^{TE} \pmod{2\pi}. \tag{2.9}$$

This phase difference is plotted in Fig. 2.6 and compared to the values calculated from our measurements using (2.3), (2.4), and (2.5). The values predicted by our simple model for the phase delay exhibit excellent agreement with the measurements.

The model presented here suggests exploring the parameter space in order to design slits that act as non-dichroic quarter-wave retarders. The requirements are that the TM and TE transmission are equal taking into account the TM loss to surface plasmons, and that the phase difference is $\pi/2$. All these requirements are influenced by the metal permittivity $\varepsilon(\lambda)$, the slit width w, and the film thickness d.

Our experimental results contradict a recently published proposal for a quarter-wave retarder using perpendicular metallic nanoslits,[20] in which the width of the slits is varied purely to control the TM transmission. Varying the width of the slit also changes the TE transmission of the incident light and the phase difference between the TM and TE components.

[20] Khoo, Li, and Crozier, 2011.

2.5 *Summary*

We have studied the transmission properties of a subwavelength slit milled in a 200 nm thick gold-metal film as a function of the slit width (50–500 nm), and of the polarization of the incident radiation (at $\lambda = 830$ nm). As the slit width is decreased, the transmission of the TE mode diminishes quite gradually until it becomes very small at a slit width of about $\lambda/8$, reminiscent of the phenomenon of waveguide cutoff. In contrast, the transmission of the TM mode does not vanish. Instead, it exhibits a minimum associated with the efficient excitation of surface plasmons.

Moreover, we have studied the birefringence of this subwavelength slit and found that the phase lag between the TM mode and TE mode passes through a value of $\pi/2$, so that a properly dimensioned slit can act as a quarter-wave retarder. We have successfully explained our experimental results with a simple waveguide model.

Appendix 2.A Reciprocity of the slit transmission

This section is an appendix that did not appear in the published paper.

WE STATED EARLIER IN THIS CHAPTER that we corrected our model for the finite NA of the detector in our experiment. The necessity of this correction was brought to our attention by an apparent violation of reciprocity in the experiment. In the experiments described in the foregoing sections, we illuminated the sample on the gold side (hereafter the 'forward' configuration), but when we turned the sample around and performed the experiment again while illuminating it from the glass side (the 'reverse' configuration), the results were different!

This is, of course, not really a violation of reciprocity, but it is caused by the detector's NA. The exit aperture of the slit is subwavelength, so it radiates in all directions, but not uniformly. The slit's scattering profile depends on the shape of the mode inside the slit, and also on the medium that the slit scatters into. Therefore, not all the radiation that actually exits the slit is emitted into the cone of angles that the detector can collect. Which fraction is collected by the detector depends on the circumstances, meaning that the two configurations cannot be compared directly without correcting for this effect.

We assume that the detector is situated in air, with an index of refraction n_4. In the forward configuration, $n_1 = n_4$ and n_3 is the index of the glass substrate. Conversely, in the reverse configuration, n_1 is the index of the glass substrate and $n_3 = n_4$.

THE SLIT SUPPORTS ONE TM AND ONE TE MODE. We treat the slit as a parallel-plate waveguide with metal walls in which we assume that higher-order modes do not propagate. The metal wall boundary is at $x = \pm w/2$. We call the complex amplitudes of the waveguide modes inside the slit $\tilde{\mathbf{E}}_{\mathrm{TM}}(x, z)$ and $\tilde{\mathbf{E}}_{\mathrm{TE}}(x, z)$. They depend on the permittivity ε, the index of the slit material n_2, and the slit width w.

To find the angular scattering profile of the modes, we take the Fourier transform of the mode profile at the exit aperture of the slit: $\tilde{\mathbf{E}}_{\mathrm{TM}}(x, d)$ and $\tilde{\mathbf{E}}_{\mathrm{TE}}(x, d)$. This gives us the scattered electric field amplitude $\tilde{\mathbf{F}}$ as a function of transverse wavenumber k_x. This is appropriate if the collection objective is in the far field of the slit. We estimate the Fresnel number N_F using typical values for our experiment,

$$N_F = \frac{a^2}{L\lambda} = \frac{w^2}{4L\lambda} \approx \frac{(5 \times 10^{-7})^2}{4 \cdot 1 \times 10^{-3} \cdot 8 \times 10^{-7}} \approx \frac{25 \times 10^{-14}}{32 \times 10^{-10}} \ll 1, \qquad (2.10)$$

which justifies the assumption of Fraunhofer diffraction at 1 mm distance from the slit. L is in this case the working distance of the objective.

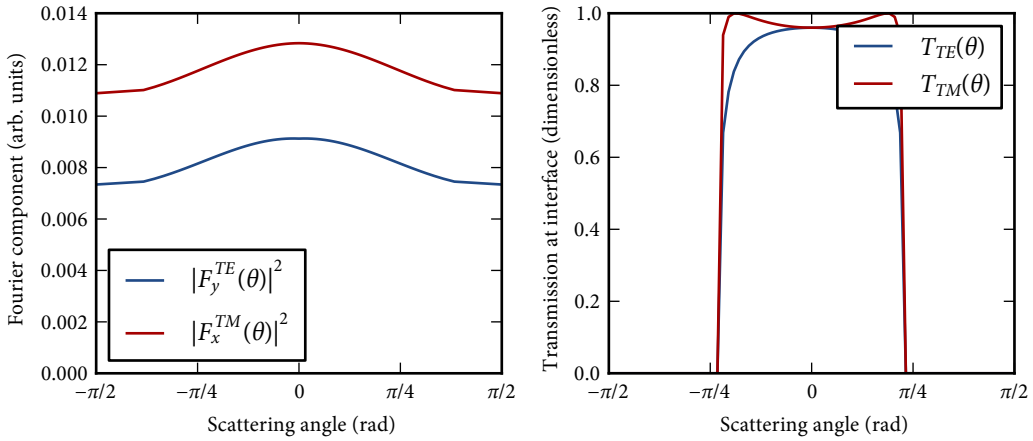

Figure 2.9: (a) Scattering profile of a 250 nm wide slit, as a function of angle. (b) Transmission of scattered light at the n_3–n_4 interface as a function of scattering angle. Note the Brewster angle at the point where the TM transmission reaches unity. Other parameters: $\lambda = 800$ nm, $n_2 = 1.0$, $n_3 = 1.5$, $n_4 = 1.0$.

The angle θ of the corresponding plane wave component is equal to

$$\theta = \arcsin(k_x/n_3 k_0), \tag{2.11}$$

where $k_0 = 2\pi/\lambda$ is the wavenumber in free space. We plot the angular scattering profiles for a 250 nm wide slit in Fig. 2.9a, calculated numerically by fast Fourier transform.

In the forward configuration (light incident on the air side) described above, we have to take into account the Fresnel losses at the n_3–n_4 (glass-air) interface. Part of the scattered energy never leaves the glass substrate, due to total internal reflection. The TE and TM components are also transmitted differently, since there is a Brewster angle for TM.

For the transmission $T = 1 - R$, we write:

$$T_{\mathrm{TE}}(\theta) = 1 - \left(\frac{n_3 \cos\theta - n_4 \sqrt{1 - (\frac{n_3}{n_4}\sin\theta)^2}}{n_3 \cos\theta + n_4 \sqrt{1 - (\frac{n_3}{n_4}\sin\theta)^2}} \right)^2 \tag{2.12}$$

$$T_{\mathrm{TM}}(\theta) = 1 - \left(\frac{n_3 \sqrt{1 - (\frac{n_3}{n_4}\sin\theta)^2} - n_4 \cos\theta}{n_3 \sqrt{1 - (\frac{n_3}{n_4}\sin\theta)^2} + n_4 \cos\theta} \right)^2 \tag{2.13}$$

We plot these transmission profiles in Fig. 2.9b. They are independent of the slit width, or indeed any of the slit parameters. Also note that for $n_3 = n_4$, $T = 1$, as it should be since there is no interface in that case.

The finite NA of the detector means that not all of the scattered light is collected. Light outside a maximum acceptance angle $\theta_{\max}$ misses the detector. Due to Snell's law, $NA = n_3 \sin\theta_{\max}$, and therefore

$$\theta_{\max} = \arcsin(NA/n_3), \tag{2.14}$$

no matter what medium n_4 the detector is actually in.

The detector signal, then, must be corrected by a factor

$$C_i = \frac{\displaystyle\int_{-\theta_{\max}}^{\theta_{\max}} T_i(\theta) \left|\tilde{\mathbf{F}}_i(n_3 k_0 \sin\theta)\right|^2 d\theta}{\displaystyle\int_{-\pi/2}^{\pi/2} \left|\tilde{\mathbf{F}}_i(n_3 k_0 \sin\theta)\right|^2 d\theta}, \tag{2.15}$$

where $i \in \{\text{TE}, \text{TM}\}$, which is the ratio of energy entering the detector to the total energy scattered from the slit. This factor depends on λ, ε, n_2, n_3, n_4, w, and NA. (It does not depend on d or n_1; the factor containing d can be taken outside the integrals and divided away, and n_1 only influences the coupling to surface plasmon modes on the front interface.)

In the experiment described in this chapter, we take λ, ε, n_2, and n_4 as design parameters, so we will not examine their influence on the correction factor. The index n_3 can take one of two values: for reverse illumination, $n_3 = n_4$, whereas for forward illumination $n_3 \neq n_4$. Since we are only interested in the ratio of TM to TE transmission, we will examine the ratio of the two correction factors $C_{\text{TM}}/C_{\text{TE}}$ in four situations: forward and reverse illumination, and for low and high NA.

We examine the ratio because the values of C_{TM} and C_{TE} individually are not particularly surprising; a high-NA detector will collect more of the light in both forward and reverse illumination, obviously. Reverse illumination also causes more light to enter the detector, no matter the NA because there is no total internal reflection loss at the n_3–n_4 interface, so $T_i(\theta) = 1$ in (2.15).

In Fig. 2.10a we plot the ratio $C_{\text{TM}}/C_{\text{TE}}$ for reverse illumination, calculated for two different numerical apertures. We see that the numerical aperture does not influence the shape of the curve very much.

Forward illumination, for which the ratio of the correction factors is plotted in Figs. 2.10b, is quite different. If the detector has low NA, it behaves much the same as in reverse illumination, with slightly more difference between TM and TE. In the high-NA case, on the other hand, the detector picks up more TM light than TE light for almost every slit width: there is about a 6% difference. A look at Fig. 2.9 suggests that this happens when the NA is large enough that $\theta_{\max}$ is large enough that it includes the Brewster angle at the n_3–n_4 interface, at which all the TM light is transmitted and not all the TE light.

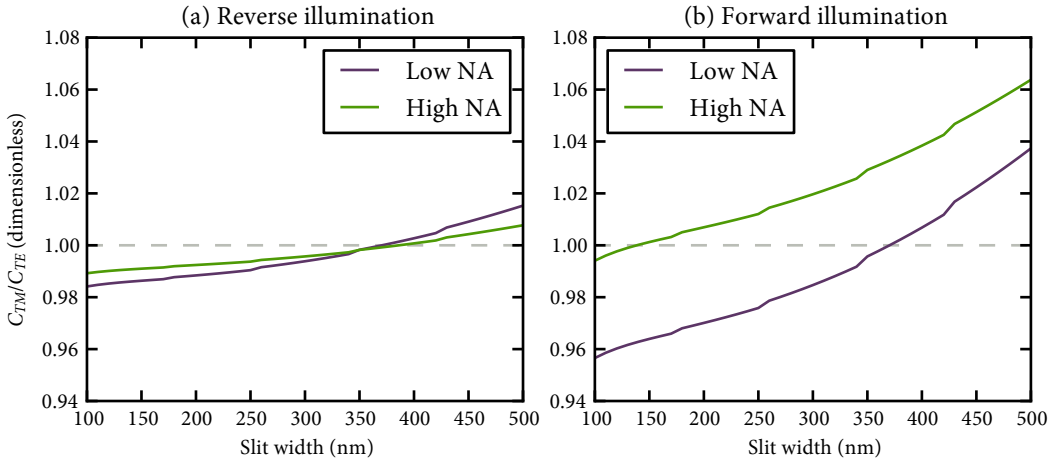

Figure 2.10: Ratio of the two NA correction factors $C_{\mathrm{TM}}/C_{\mathrm{TE}}$: less than 1 means that the detector picks up more TE light, more than 1 means the detector picks up more TM light. It is plotted for two different NA values: low ($NA = 0.2$, purple), and high ($NA = 0.8$, green). (a) Reverse illumination ($n_1 = 1.5$, $n_2 = n_3 = n_4 = 1.0$); (b) forward illumination ($n_1 = n_2 = n_4 = 1.0$, $n_3 = 1.5$). As usual, we take $\lambda = 800$ nm.