Roof prisms and phase coatings

Michael J. Hurben

October 19, 2024

Abstract

Roof prisms are common components of widely-available optical instruments such as binoculars. Unless they are treated with specialized coatings, however, they degrade resolution due to combined effects from (a) different phase shifts associated with s- and p-polarizations upon total internal reflection, and (b) the novel geometry of the dual light paths. Here we present a full treatment of this problem, beginning with a derivation of the phase shift expressions from basic electrodynamics. The deleterious effect on resolution is then demonstrated by using a Jones matrix approach to calculate how the Fraunhoffer diffraction pattern is altered by the presence of a 90° roof prism. Finally, mitigation of this aberration by means of phase-compensation coatings is described, using a transfer matrix method to calculate the phase shift offset for single and triple (Herpin equivalent) films.

1 Introduction

Binoculars are a pair of aligned refracting telescopes and must meet various requirements to ensure quality performance. A sizable field of view, adequate magnification, and comfortable eye relief are examples of important features that a good design will provide. Fortunately all of these attributes can be realized using the elegant Keplerian telescope, which places positive lenses at either end of an open tube. But a price is paid for this simple approach, for the image will be upside-down and left-right reversed. Additional lenses can be employed to correct for this, but then a much longer device is the result, like the archaic mariner's spyglass, which is not amenable for compact

binoculars. (For reference on these optics fundamentals, the reader is referred to a standard text such as Hecht [1].)

A better solution is to use multiple reflections to undo the reversal caused by the lenses. The surfaces might be mirrors or the insides of a glass prism. The latter is far preferable because it may utilize total internal reflection (TIR), in which none of the light's energy is lost. This is not the case for even the best mirrors, which cannot reflect with 100% efficiency.

To get both up-down and left-right correction we might employ a Porro prism, which uses two glass blocks arranged to reflect from four inner surfaces. But this geometry also acts to shift the optical path sideways, resulting in the iconic Z-shape that many binocular models feature. This makes for a bulkier device and is not well-suited for a sealed housing.

Certain types of roof prism can do the same work as the Porro, but without displacing the overall light path. It is a neat trick, requiring a more sophisticated geometry. Central to their operation is a right-angle junction of two reflecting surfaces. These are incorporated into the prism by substituting them for one flat side, resulting in a peaked, roof-like appearance. Incident light is split into two parts that follow different paths: one half strikes one side of the roof, then the second; vice versa for the other half. The end result is a reversal of the light in a direction perpendicular to the roof line.

Every "straight-through" binocular body holds these clever devices between the objective lenses and eyepieces, undoing the image inversion of the Keplerian telescope and a giving the instrument a streamlined shape amenable to waterproof sealing and the filling with a dry gas to prevent condensation. But there is a catch.

These prisms must be made to exacting standards, as even a slight misalignment of the surfaces will wreck the performance in the form of image doubling. They are precision instruments and priced accordingly. But even for a perfectly machined prism, an insidious complication is yet lurking: there are phase shifts upon reflection which conspire with the roof geometry to cause degraded resolution [2]. Understanding why this occurs, and how it can be fixed, is not trivial, and a comprehensive technical review on this interesting problem is warranted. Recent papers specific to roof prism diffraction effects [3], as well as the closely related problem of Airy pattern deterioration by corner-cube reflectors [4] [5] are relevant, but succinct and geared towards a more specialized audience.

In the first part of this treatment, the phase shifts that accompany TIR will be derived from basic electromagnetism. In the second section, we quan-

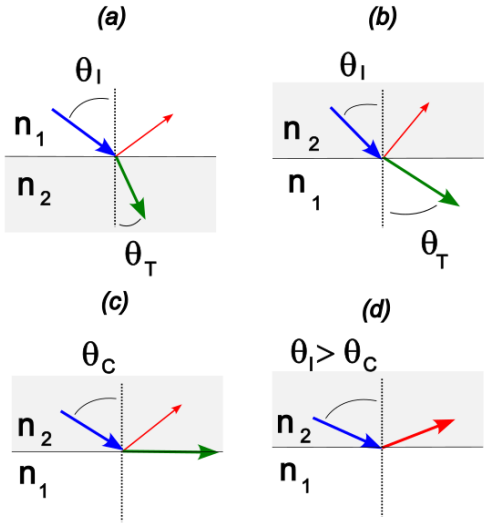

Figure 1: Graphical examples of Snell's law. Of particular interest is the TIR resulting when the critical angle is exceeded (d).

tify the impact on resolution by calculating how a roof prism distorts a single-slit diffraction pattern. In the final part, we will see how the judicious choice of dielectric coatings can mitigate the effect.

2 Phase shifts from total internal reflection

Total internal reflection (TIR) is often introduced in the context of Snell's law, which relates the angles of incidence and transmission at the interface between two media. These are characterized by different refractive indices, n_1 and n_2 (and as a result, different velocities v_1 and v_2 and wavenumbers k_1 and k_2). We will write Snell's law as

$$n_1 \sin\theta_I = n_2 \sin\theta_T. \tag{1}$$

The subscripts on the angles refer to incident and transmitted rays, with the light moving from the medium with index n_1 into the other with n_2. The reflected light always has the same angle as the incident. Various examples of refraction which illustrate this relationship are shown in Figure 1. In each

of these, $n_2 > n_1$, so in (a) we have light going from a fast to slow medium. Of greater interest is the situation in which light moves from a slower to a faster medium, such as (b). There will be a critical angle for the incident light, denoted θ_C, such that $\theta_T = \pi/2$. This then yields $\sin\theta_C = n_1/n_2$, shown in (c). If incident light arrives at an angle larger than this (d), there is no angle θ_T that can satisfy Snell's law, and TIR must result. With zero transmission, all of the incident light is reflected.

To approach this more rigorously, we consider a transverse, sinusoidal electromagnetic wave incident on a dielectric boundary, and determine the form of the reflected and transmitted waves. Doing so requires identifying the constraints on the electric and magnetic fields on either side of the interface. This leads to the Fresnel equations, as developed in any undergraduate text on optics [1] or electricity and magnetism [6]. What is often neglected, however, is a derivation of the phase shift that occurs upon TIR. This shift depends upon the refractive indices, the angle of incidence, and the polarization of the incident light, and is important for various optical devices.

We start by ascribing sinusoidal dependence for the electric field in time and space in the usual way, namely

$$E_I = E_{I0} e^{i(\mathbf{k}_I \cdot \mathbf{r} - \omega t)}, \tag{2}$$

$$E_R = E_{R0} e^{i(\mathbf{k}_R \cdot \mathbf{r} - \omega t - \delta)}, \tag{3}$$

and

$$E_T = E_{T0} e^{i(\mathbf{k}_T \cdot \mathbf{r} - \omega t)}. \tag{4}$$

Figure 2 shows the standard geometry and nomenclature for the problem. The $x - y$ plane defines the surface interface, while the $y - z$ plane is the plane of incidence. Because the electric and magnetic fields can have any orientation that is perpendicular to the propagation direction, we must consider fields that lie along the x-direction as well as those in the $y - z$ plane. The two cases are shown in (b) and (c). The case of a transverse electric field (TE) is also called "s-polarized" while a transverse magnetic (TM) field is the "p-polarized" case, per convention.

The boundary conditions that express how the fields change as we cross from one medium into the other are found from Maxwell's equations. We assume that our materials are non-magnetic and will differ only in terms of

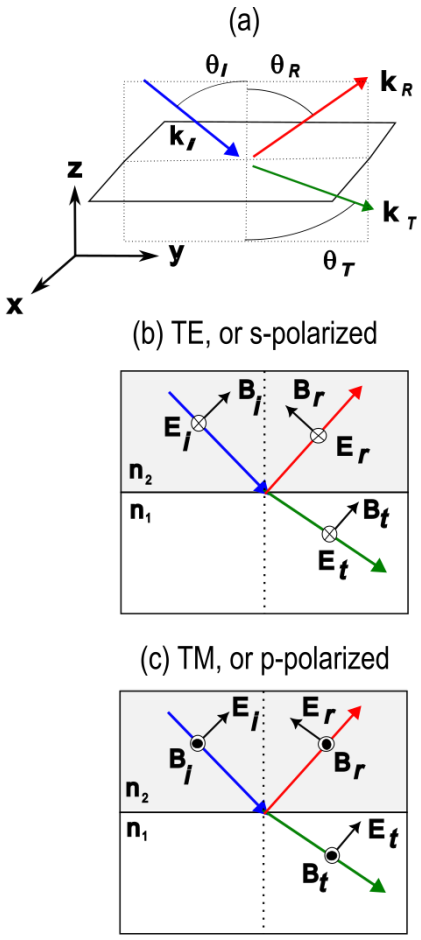

Figure 2: Coordinate axes and geometries used to describe light striking a boundary.

refractive index n and permittivity ϵ. There is no free charge or electrical current. In such a case, Maxwell's equations expressed as surface and line integrals are:

$$\oint \epsilon \vec{E} \cdot d\vec{A} = 0, \qquad \oint \vec{B} \cdot d\vec{A} = 0, \tag{5}$$

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_B}{dt}, \qquad \oint \frac{\vec{B}}{\mu} \cdot d\vec{s} = \epsilon \frac{d\Phi_E}{dt}. \tag{6}$$

Figure 3 shows the surfaces and paths chosen to make these integrals trivial.

For the first two cases of (a) and (b), the cylinder is made so flat that only the top and bottom contribute to the integrals. For the line integrals, the paths in (c) and (d) are made to barely extend above and below the boundary, so there is no net flux passing through. The boundary conditions are then:

$$\epsilon_1 E_{1z} = \epsilon_2 E_{2z}, \qquad E_{1y} = E_{2y}, \qquad E_{1x} = E_{2x}, \tag{7}$$

$$B_{1z} = B_{2z}, \qquad B_{1y} = B_{2y}, \qquad B_{1x} = B_{2x}. \tag{8}$$

These are nice and simple results, and something should be noted here: it is easy (and wrong) to picture TIR as an electromagnetic wave simply bouncing off a boundary without any penetration. Per Maxwell's equations, that is simply not possible. The fields cannot be non-zero on one side of the boundary and yet totally vanish on the other, TIR or no TIR. The slight-of-hand nature employs here is an evanescent wave, which provides a non-zero field beyond the boundary but carries none of the energy away.

For the case of s-polarization, or the TE mode, the E field is restricted to the x-direction, while the B field has both $y-$ and $z-$components. We can make use of the latter by noting $E = cB/n$, allowing us to eliminate the magnetic field term. The boundary conditions then become:

$$E_{I0} + E_{R0} e^{-i\delta} = E_{T0}, \tag{9}$$

$$n_1 (E_{I0} - E_{R0} e^{-i\delta}) \cos\theta_I = n_2 E_{T0} \cos\theta_T. \tag{10}$$

These combine to give:

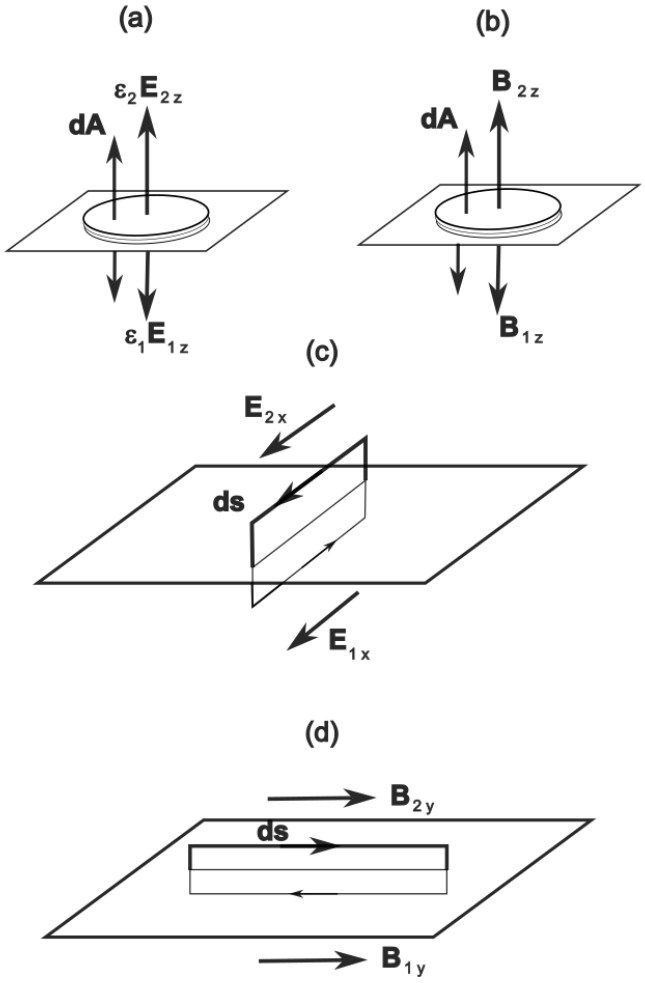

Figure 3: Surfaces and contours used to derive boundary conditions from Maxwell's equations.

$$E_{I0}(n_1 \cos\theta_I - n_2 \cos\theta_T) = E_{R0}e^{-i\delta}(n_1 \cos\theta_I + n_2 cos\theta_T). \tag{11}$$

In order to reduce the clutter, we will write the ratio of indices as

$$n = \frac{n_2}{n_1}, \tag{12}$$

and define

$$a = \cos\theta_I, \tag{13}$$

$$b = in\cos\theta_T. \tag{14}$$

Why did we introduce the i term? It is because $\cos\theta_T$ is imaginary for TIR, and that is the only regime we care about. We can see this by writing Snell's law as:

$$\cos^2\theta_T = 1 - \sin^2\theta_T = 1 - n^2\sin^2\theta_I, \tag{15}$$

so

$$\cos\theta_T = i\sqrt{n^2\sin^2\theta_I - 1}. \tag{16}$$

The term inside the radical is positive as long as the critical angle is exceeded because $n\sin\theta_I \geq n\sin\theta_C = 1$.

The simplified version is

$$\frac{E_{R0}e^{-i\delta}}{E_{I0}} = \frac{a - ib}{a + ib}. \tag{17}$$

Now the beautiful part: if we expand the complex exponential, we can group the real and imaginary parts of this equation as follows:

$$\frac{E_{R0}\cos\delta}{E_{I0}} = \frac{a^2 - b^2}{a^2 + b^2}, \tag{18}$$

$$\frac{E_{R0}\sin\delta}{E_{I0}} = \frac{2ab}{a^2 + b^2}. \tag{19}$$

Square and add these and we get the square of the reflection coefficient:

$$r^2 = \frac{E_{R0}^2}{E_{I0}^2} = 1. \tag{20}$$

Divide them and we have an expression for the phase shift:

$$\tan\delta = \frac{2ab}{a^2 - b^2}. \tag{21}$$

For the case of the TM mode, or p-polarization, the steps taken are identical, but we start with slightly different in boundary conditions, because the B field is restricted to the x-direction, while the E field has both y- and z-components. The conditions are:

$$n_1 E_{I0} + n_1 E_{R0} e^{-i\delta} = n_2 E_{T0}, \tag{22}$$

$$(E_{I0} - E_{R0} e^{-i\delta}) \cos\theta_I = E_{T0} \cos\theta_T. \tag{23}$$

Eliminating the transmitted field terms, we have:

$$E_{I0}(n\cos\theta_I - \cos\theta_T) = E_{R0} e^{-i\delta}(n\cos\theta_I + cos\theta_T). \tag{24}$$

We will play the same trick as before, but write the slightly different terms:

$$\alpha = n\cos\theta_I, \tag{25}$$

$$\beta = i\cos\theta_T. \tag{26}$$

We will then arrive at a similar result for the phase:

$$\tan\delta = \frac{2\alpha\beta}{\alpha^2 - \beta^2}. \tag{27}$$

Note this can also be written as

$$\tan\delta = \frac{2(\beta/\alpha)}{1 - (\alpha/\beta)^2}, \tag{28}$$

which brings to mind the half-angle identity for the tangent:

$$\tan x = \frac{2\tan(x/2)}{1 - \tan^2(x/2)}. \tag{29}$$

So here then are the final expressions for the phase shifts. For the TM-mode or p-polarization,

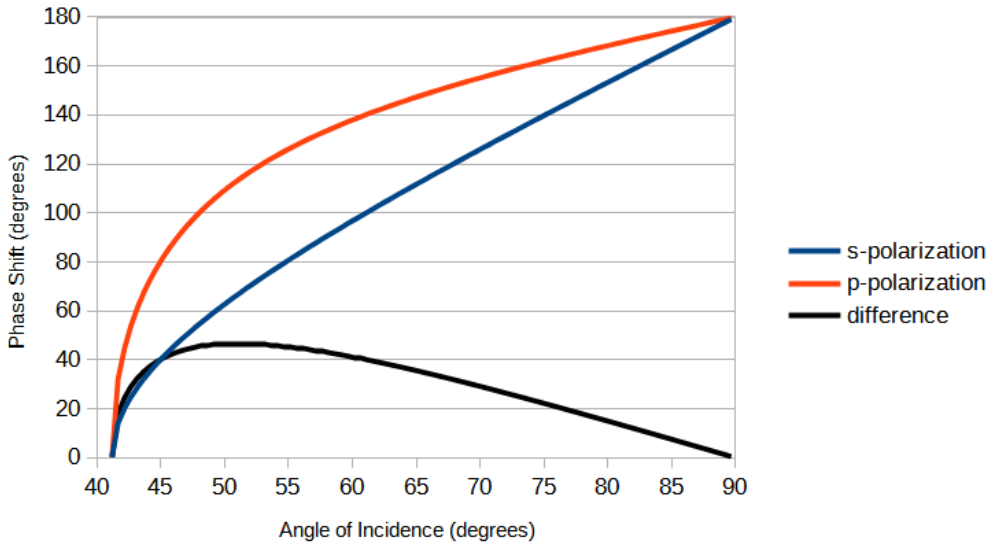

Figure 4: Phase shift dependence on angle of incidence and polarization, for TIR with a refractive index ratio of 1.52.

$$\delta_P = 2 \arctan\left(\frac{n\sqrt{n^2 \sin^2\theta_I - 1}}{\cos\theta_I}\right). \tag{30}$$

For the TE-mode, or s-polarization,

$$\delta_S = 2 \arctan\left(\frac{\sqrt{n^2 \sin^2\theta_I - 1}}{n\cos\theta_I}\right). \tag{31}$$

Figure 4 shows the reflected light phase shift as a function of incident angle for the case of the two polarizations, where we have chosen $n_2 = 1.52$ and $n_1 = 1$, corresponding to a typical glass bordered by air. The critical angle is just over 41°. The phase shifts behave somewhat similarly, but they are not the same except at the critical angle and at the limiting case of a glancing incident angle. Other than at these endpoints, the p-polarization always produces a larger phase shift. We also plot the difference between the shifts as this will be important later.

Let's consider why these curves behave as they do. When TIR occurs, all of the incident energy of the incoming wave is reflected, but the elec-

tric field just across the interface cannot simply vanish. For there to be no energy transmission, the field must decay exponentially as we move into the air beyond the prism. How rapidly it attenuates depends on the angle, polarization, and ratio of refractive indices.

We look first at the angle dependence for the s-polarization because the behavior of the electric field is simpler. In Figure 5(a) we illustrate the case of incidence at just less than the critical angle, so that no TIR occurs. The figure shows the wave fronts, the lines of constant amplitude in the field. In (b) we have the case of just exceeding the critical angle, where the wavefronts in the air region must eventually decay to nothing. In making this transition where we barely changed the angle, we began with wavefronts extending to negative infinity, so it is reasonable that the exponential decay will be a long one and will meet the boundary with a low slope. This means that little phase shift in the reflected wave is needed to ensure that the total electric field changes smoothly across the boundary, as shown in (c).

As the incidence angle increases, it becomes more difficult for the changing electric field to penetrate the boundary; less momentum is being directed downward. The exponential decay, then, will become more pronounced, reaching the boundary with a substantial negative slope. The reflected wave will need to be shifted so that the superposed waves in the glass meet the boundary with the same slope, as seen in (d).

The p-polarized light generates a bigger shift for the same incidence angle than s-polarized light because of the electrical polarization of the glass and the boundary geometry. If the electric field is parallel to the boundary, as in the case of s-polarized waves, each atom becomes polarized in response, but as these dipoles are lying end to end they cancel one another out along the boundary. With the field perpendicular, the dipoles in the glass bulk again cancel, but at the surface we come to the ends of the final polar molecules. There is therefore a net accumulation of charge here, and the field across the interface must be larger in order to remain continuous with the field boosted by the discontinuity. The electric field that decays away on the air side is stronger than it would be for s-polarization, and as a result, we need a bigger phase shift for the reflected wave in order to ensure the smooth boundary conditions are met.

When linearly polarized light undergoes TIR, then, it will emerge with an elliptical polarization, while elliptically polarized light will in general be reflected into a different elliptical configuration. Multiple reflections will rapidly lead to more complex changes to the light, as we will see when looking

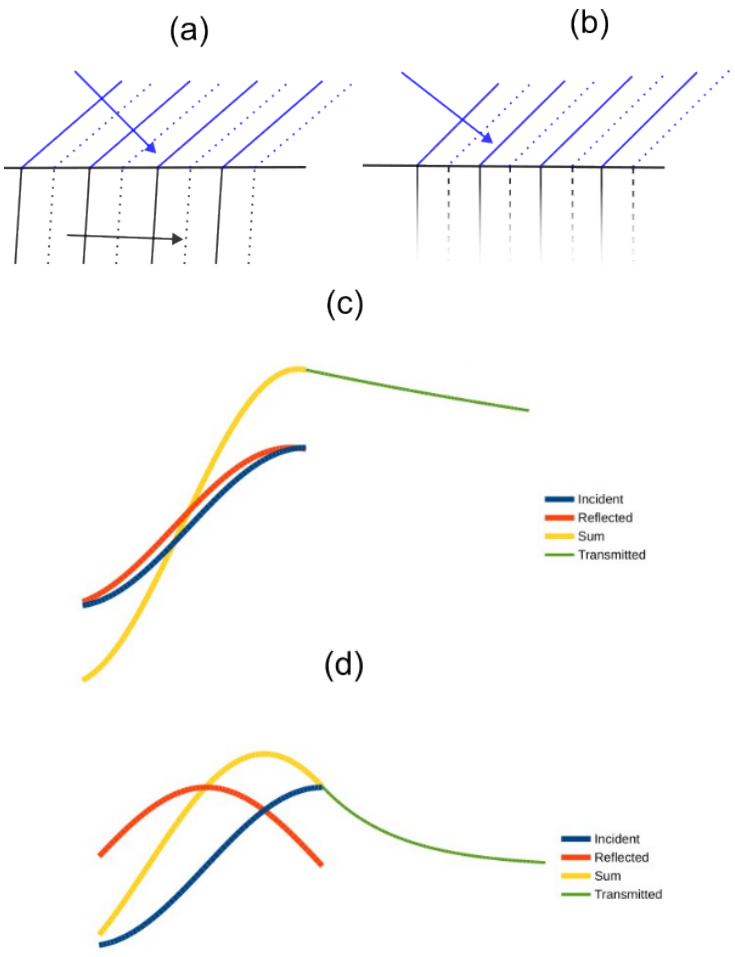

Figure 5: Behavior of wavefronts when less than (a) and greater than (b) the critical angle. Diagrams (c) and (d) illustrate how the shape of the exponential decay dictates the phase shift upon reflection needed to ensure that the boundary conditions are met.

at roof prism performance in the following section.

3 Roof prism impact on the diffraction pattern

Resolution is key to the performance for any optical system, and the resolving power of an ideal instrument will be limited only by the inherent diffraction properties of light. It is important therefore to calculate how a roof prism will affect the diffraction pattern of light passing through it. Early workers grew accustomed to seeing anomalous results, which were known to occur if the angle between roof surfaces did not meet high tolerance requirements. Still, even when prisms were fabricated to the most exacting standards, they showed degraded resolution in the form of a broadened, or even doubled, central line in the single slit pattern. It was realized that the phase shifts from TIR, together with the roof prism optical paths, were the culprit.[2]

In this section we will calculate the far-field (Fraunhoffer) diffraction pattern for monochromatic, unpolarized light of wavelength λ from a slit of width 2w, which passes through an Amici roof prism. While an Amici is not employed in binoculars, it shares with the Abbe-König and Schmidt-Pechan prisms a 90° roof oriented at 45° to the incoming rays, dividing the incoming cylinder of light into two different paths.

We will consider two cases, the first having the slit in the same plane as the line made by the two roof surfaces, as shown in Figure 6(a). This will be referred to as the parallel case, while the other, the perpendicular case, will have the slit oriented horizontally, as shown in Figure 6(b). There are three frames which we will need to work with: the frame 0, which is the base coordinate system, plus frames associated with surfaces 1,and 2, as shown in Figure 6(c). Figure (d) shows the view of these surfaces when looking into the prism along the y-axis. Frame 0 is useful for the incident and emerging light, while 1 and 2 are amenable to working with the two roof surfaces. We will construct all of these shortly. For the parallel slit orientation, the slit lies along the z-axis, while for the perpendicular it will lie along the x-axis.

For either case, a ray falling perpendicular on the prism has a propagation direction (indicated by the unit vector $\hat{I}$) having only a y-component, and it will emerge along the $-z$ direction. Because the angles involved in Fraunhoffer diffraction are small, we will make the approximation that all light arriving at the roof surfaces moves only in the y-direction.

Our specific task is then to develop an expression for the intensity pro-

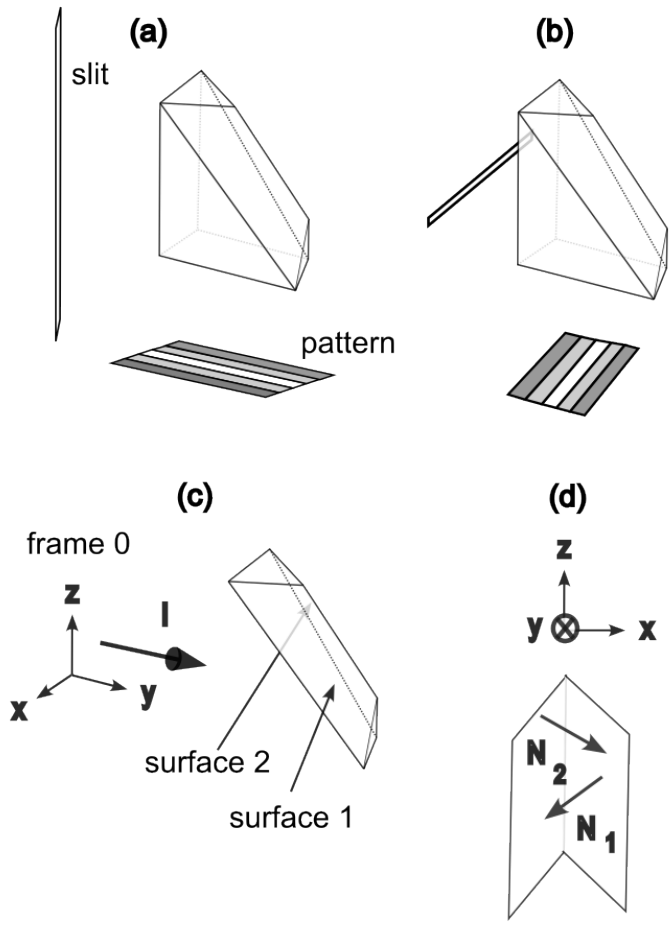

Figure 6: Geometries used for the diffraction pattern calculations.

file, along a direction perpendicular to the image of the slit after emerging form the prism. As intensity is proportional to the squared magnitude of the electric field strength, we start there, summing over all of the source contributions from the aperture and accounting for any additional modifications made by the prism. For each geometry we will be working in one-dimension. For the parallel case, the diffraction integral for the final electric field is

$$\vec{E}_F = \int \mathcal{T}\vec{E}_0 e^{2\pi i x X/\lambda Y} dx. \tag{32}$$

Here, $\mathcal{T}$ is a 2×2 Jones matrix which accounts for the reflections and phase shifts in the prism. We will develop this matrix explicitly below. X is the distance along which the image pattern varies, and Y is the optical distance from source to image. Note that for small angles, $\sin\theta_I \approx X/Y$.

The perpendicular case features an identical expression, except for the trivial swap of x and z, and the more important fact that $\mathcal{T}$ will be different because due to different light paths taken. However at this point, we will focus explicitly on the parallel geometry. When that is completed, we will consider the perpendicular case and use symmetry considerations to show that the diffraction pattern is not affected.

The diffraction integral above is a vector expression, so we must consider the polarization for the source field $\vec{E}_0$. It will be necessary to treat the light as polarized during parts of the derivation below, but we will eventually move to an expression for unpolarized light by integrating over all linear polarization angles.

For now, we express the electric field as a two-vector with components in vertical and horizontal directions perpendicular to the propagation direction. For simplicity we will assume linear polarization and write

$$\vec{E}_0 = E_0 \begin{bmatrix} \cos\psi \\ \sin\psi \end{bmatrix}. \tag{33}$$

The effect of the prism is captured by the generic matrix $\mathcal{T}$. But we must account for the fact that there are two different paths and hence two Jones matrices at play. We handle this by noting that with the prism in the center of the light path, the integration can be broken into two parts corresponding to paths 1 and 2 as:

$$\vec{E}_F = \int_{-w}^{0} \mathcal{T}_1\vec{E}_0 e^{2\pi i x X/\lambda Y} dx + \int_{0}^{w} \mathcal{T}_2\vec{E}_0 e^{2\pi i x X/\lambda Y} dx. \tag{34}$$

Nothing in the Jones transformations will depend upon the variable x so long as the we are dealing with a narrow diffraction pattern. The integrals are simple, giving:

$$\vec{E}_F = \frac{Y\lambda}{2\pi i X}\left[\mathcal{T}_1\vec{E}_0(1-e^{-2\pi i w X/\lambda Y}) - \mathcal{T}_2\vec{E}_0(1-e^{2\pi i w X/\lambda Y})\right]. \tag{35}$$

This can be further consolidated as:

$$\vec{E}_F = \frac{\sin(\pi w X/Y\lambda)}{\pi X/Y\lambda}\left[\mathcal{T}_1\vec{E}_0 e^{-i\pi w X/\lambda Y} + \mathcal{T}_2\vec{E}_0 e^{i\pi w X/\lambda Y}\right]. \tag{36}$$

For small angles, $X/Y \approx \sin\theta \approx \theta$, and

$$\vec{E}_F = w\,\mathrm{sinc}(\pi w\theta/\lambda)\left[\mathcal{T}_1\vec{E}_0 e^{-i\pi w\theta/\lambda} + \mathcal{T}_2\vec{E}_0 e^{i\pi w\theta/\lambda}\right]. \tag{37}$$

We will write $\eta = \pi w\theta/\lambda$ going forward. Also, we can move from polarized to unpolarized light for this expression, as follows. In vector form,

$$\vec{E}_F = \begin{bmatrix} E_{FV} \\ E_{FH}\end{bmatrix} = E_0 w\,\mathrm{sinc}(\eta)\left[\mathcal{T}_1 e^{-i\eta} + \mathcal{T}_2 e^{i\eta}\right]\begin{bmatrix}\cos\psi \\ \sin\psi\end{bmatrix}. \tag{38}$$

The subscripts "FV" and "FH" refer to the vertical and horizontal components of the final field. These are

$$\begin{aligned} &E_{FV} = E_0 w\,\mathrm{sinc}(\eta)\times \\ &\left[\left[(\mathcal{T}_1)_{11}e^{-i\eta} + (\mathcal{T}_2)_{11}e^{i\eta}\right]\cos\psi + \left[(\mathcal{T}_1)_{12}e^{-i\eta} + (\mathcal{T}_2)_{12}e^{i\eta}\right]\sin\psi\right], \end{aligned} \tag{39}$$

and

$$\begin{aligned} &E_{FH} = E_0 w\,\mathrm{sinc}(\eta)\times \\ &\left[\left[(\mathcal{T}_1)_{21}e^{-i\eta} + (\mathcal{T}_2)_{21}e^{i\eta}\right]\cos\psi + \left[(\mathcal{T}_1)_{22}e^{-i\eta} + (\mathcal{T}_2)_{22}e^{i\eta}\right]\sin\psi\right]. \end{aligned} \tag{40}$$

We will now integrate over all orientations of linear polarizations for each of these, in order to get the average unpolarized field strengths, denoted $\mathcal{E}$, corresponding to each direction, according to

$$\mathcal{E} = \frac{1}{\pi}\int_0^\pi E(\psi)d\psi. \tag{41}$$

In both cases the $\cos\psi$ term yields 0 and the $\sin\psi$ term gives 2 under integration, so

$$\mathcal{E}_{FV} = \frac{2E_0 w\,\mathrm{sinc}(\eta)}{\pi}\left[(\mathcal{T}_1)_{12}e^{-i\eta} + (\mathcal{T}_2)_{12}e^{i\eta}\right], \tag{42}$$

$$\mathcal{E}_{FH} = \frac{2E_0 w\,\mathrm{sinc}(\eta)}{\pi}\left[(\mathcal{T}_1)_{22}e^{-i\eta} + (\mathcal{T}_2)_{22}e^{i\eta}\right]. \tag{43}$$

The intensity, which goes as the squared magnitude, comes from squaring and adding the two total fields for the vertical and horizontal directions.

$$I = \frac{\epsilon_0 c}{2}|E_F|^2 = \frac{\epsilon_0 c}{2}\left[|\mathcal{E}_{FV}|^2 + |\mathcal{E}_{FH}|^2\right]. \tag{44}$$

We can see where this is leading. When we include the field expressions, the intensity will be seen to have the familiar $\mathrm{sinc}^2(\eta)$ form, but modified by an expression which will depend on the coefficients from the Jones matrices.

We note that the matrix elements will themselves be complex, and this in an appropriate time to determine what they are. To start, we define path 1 as that having light first striking surface 1 followed by the second surface. Path 2 goes in the opposite order. Although we removed explicit polarization dependence in the expressions for the averaged field strengths above, we will still need to work with linearly polarized light in order to develop the Jones matrix elements, since the phase shifts depend on the division of light into p- and s-polarizations.

We will consider path 1 first. In terms of physical changes to the light, all that occurs are two sets of phase shifts, which are captured by very simple expressions. For the first reflection, the effect is given by:

$$\mathcal{P}_1 = \begin{bmatrix} e^{i\delta_{P1}} & 0 \\ 0 & e^{i\delta_{S1}} \end{bmatrix}. \tag{45}$$

We note here that multiplying both components of the electric field by the same phase term will change nothing (as long as we do so for path 2 also, which we will.) Hence, we should look to multiply such complex exponential terms by another factor it helps simplify matters. If we multiply by $\exp(-i(\delta_P + \delta_S)/2)$, this matrix becomes

$$\mathcal{P}_1 = \begin{bmatrix} e^{i\delta_1} & 0 \\ 0 & e^{-i\delta_1} \end{bmatrix}, \tag{46}$$

where

$$\delta_1 = \frac{\delta_{P1} - \delta_{S1}}{2}. \tag{47}$$

The second reflection is similarly represented by

$$\mathcal{P}_2 = \begin{bmatrix} e^{i\delta_2} & 0 \\ 0 & e^{-i\delta_2} \end{bmatrix}, \tag{48}$$

where

$$\delta_2 = \frac{\delta_{P2} - \delta_{S2}}{2}. \tag{49}$$

Unfortunately, we cannot simply multiply these two matrices in order to capture what occurs in the prism, because the validity of these phase shift matrices depends on us having decomposed the electric field into components which lie in the p-and s-directions specific to the reflecting surfaces. These orientations are different for the two surfaces, and we will need to account for this using rotation matrices of the form

$$\mathcal{R} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix}. \tag{50}$$

To find the elements requires that we trace through the geometry of the surfaces and light. Figure 7 shows the course of path 1, beginning with light along the incident direction $\hat{I} = \hat{y}$, which meets surface 1 at an angle θ_{I1} relative to the normal vector N_1. The reflected light moves in the $\hat{R}$ direction, reflects from surface 2 at an angle θ_{I2} relative to $\hat{N}_2$, and leaves the prism along the final direction, $\hat{F}$.

The surface normal vectors can be found from a careful inspection of the geometry. It is easily verified that the following vectors are orthonormal and correspond to a 45° inclination of the roof line:

$$\hat{N}_1 = \begin{bmatrix} -1/\sqrt{2} \\ -1/2 \\ -1/2 \end{bmatrix}, \tag{51}$$

$$\hat{N}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/2 \\ -1/2 \end{bmatrix}. \tag{52}$$

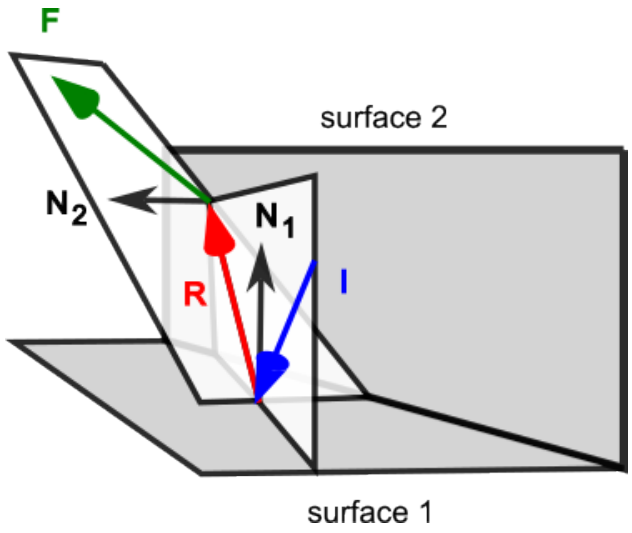

Figure 7: Path 1 through the roof prism.

The first angle of incidence is given by:

$$\cos\theta_{I1} = |\hat{I} \cdot \hat{N}_1| = \frac{1}{2}. \tag{53}$$

The reflected light direction must lie in the same plane as $\hat{I}$ and $\hat{N}_1$, and can be seen to be given by:

$$\hat{R} = \hat{I} - 2(\hat{I} \cdot \hat{N}_1)\hat{N}_1 = \hat{I} + \hat{N}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/2 \\ -1/2 \end{bmatrix}. \tag{54}$$

For the second angle of incidence, we have

$$\cos\theta_{I2} = |\hat{R} \cdot \hat{N}_2| = |\hat{I} \cdot \hat{N}_2| = \frac{1}{2}. \tag{55}$$

The angle of incidence on each surface is therefore always 60°. This means that the p- and s-orientation phase shifts are the same for both surfaces as well, as they depend only on angle of incidence and refractive indices.

The final light direction is:

$$\hat{F} = \hat{R} - 2(\hat{R} \cdot \hat{N}_2)\hat{N}_2 = \hat{R} + \hat{N}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}. \tag{56}$$

In order to work out the rotation matrices for the field components, we need to construct unit vectors amenable for working with the p- and s-polarizations for each surface. We will denote these directions according to vertical and horizontal, respectively, to be consistent with the treatment above. For path 1, we'll first need the $\hat{H}_1$ direction which we define as being perpendicular to both $\hat{I}$ and $\hat{N}_1$:

$$\hat{H}_1 = \frac{\hat{I} \times \hat{N}_1}{\sin\theta_{I1}} = \frac{2}{\sqrt{3}} \begin{bmatrix} -1/2 \\ 0 \\ 1/\sqrt{2} \end{bmatrix}. \tag{57}$$

Then, the vertical unit vector for the incident light is

$$\hat{V}_1 = \hat{H}_1 \times \hat{I} = \frac{2}{\sqrt{3}} \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ -1/2 \end{bmatrix}. \tag{58}$$

Note that after reflection, the $\hat{H}_1$ direction is maintained for the s-component, but the vertical component is shifted so as to be perpendicular to $\hat{R}$. This will be denoted with a prime:

$$\hat{V}_1' = \hat{H}_1 \times \hat{R} = \frac{2}{\sqrt{3}} \begin{bmatrix} -1/2\sqrt{2} \\ -3/4 \\ -1/4 \end{bmatrix}. \tag{59}$$

For the second surface we have similar expressions:

$$\hat{H}_2 = \frac{\hat{R} \times \hat{N}_2}{\sin\theta_{I2}} = \frac{2}{\sqrt{3}} \begin{bmatrix} -1/2 \\ -1/\sqrt{2} \\ 0 \end{bmatrix}, \tag{60}$$

$$\hat{V}_2 = \hat{H}_2 \times \hat{R} = \frac{2}{\sqrt{3}} \begin{bmatrix} 1/2\sqrt{2} \\ -1/4 \\ -3/4 \end{bmatrix}, \tag{61}$$

$$\hat{V}_2' = \hat{H}_2 \times \hat{F} = \frac{2}{\sqrt{3}} \begin{bmatrix} 1/\sqrt{2} \\ -1/2 \\ 0 \end{bmatrix}. \tag{62}$$

For the initial state, we will use the (x, y, z) reference frame in which the vertical and horizontal directions are given by $\hat{V}_0 = \hat{z}$ and $\hat{H}_0 = \hat{x}$,

respectively. The first transformation we will need must rotate into the coordinate system for surface 1. We can use the approach of [3] and write:

$$\mathcal{R}_{0,1} = \begin{bmatrix} \hat{V}_1 \cdot \hat{V}_0 & \hat{V}_1 \cdot \hat{H}_0 \\ -\hat{V}_1 \cdot \hat{H}_0 & \hat{V}_1 \cdot \hat{V}_0 \end{bmatrix} = -\frac{1}{\sqrt{3}} \begin{bmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & 1 \end{bmatrix}. \tag{63}$$

The next Jones matrix is for the phase shifts on surface 1, so we have

$$\mathcal{P}_1\mathcal{R}_{0,1} = -\frac{1}{\sqrt{3}} \begin{bmatrix} e^{i\delta} & 0 \\ 0 & e^{-i\delta} \end{bmatrix} \begin{bmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & 1 \end{bmatrix} = -\frac{1}{\sqrt{3}} \begin{bmatrix} e^{i\delta} & \sqrt{2}e^{i\delta} \\ -\sqrt{2}e^{-i\delta} & e^{-i\delta} \end{bmatrix}. \tag{64}$$

Next we need the rotation matrix for moving from the surface 1 frame to surface 2, which is

$$\mathcal{R}_{1,2} = \begin{bmatrix} \hat{V}_1' \cdot \hat{V}_2 & \hat{H}_1' \cdot \hat{V}_2 \\ -\hat{H}_1' \cdot \hat{V}_2 & \hat{V}_1' \cdot \hat{V}_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & -2\sqrt{2} \\ 2\sqrt{2} & 1 \end{bmatrix}. \tag{65}$$

This gives

$$\mathcal{R}_{1,2}\mathcal{P}_1\mathcal{R}_{0,1} = -\frac{1}{3\sqrt{3}} \begin{bmatrix} e^{i\delta} + 4e^{-i\delta} & \sqrt{2}e^{i\delta} - 2\sqrt{2}e^{-i\delta} \\ 2\sqrt{2}e^{i\delta} - \sqrt{2}e^{-i\delta} & e^{-i\delta} + 4e^{i\delta} \end{bmatrix}. \tag{66}$$

With the next phase change matrix this becomes

$$\mathcal{P}_2\mathcal{R}_{1,2}\mathcal{P}_1\mathcal{R}_{0,1} = -\frac{1}{3\sqrt{3}} \begin{bmatrix} e^{2i\delta} + 4 & \sqrt{2}e^{2i\delta} - 2\sqrt{2} \\ 2\sqrt{2} - \sqrt{2}e^{-2i\delta} & e^{-2i\delta} + 4 \end{bmatrix}. \tag{67}$$

One more rotation is needed to get into directions vertical and horizontal for the final direction $\hat{F}$, where $\hat{V}_F = \hat{y}$ and $\hat{H}_F = \hat{x}$. This means rotating back into the 0-frame we started with. Using the same approach as above, this is found to be:

$$\mathcal{R}_{2,0} = -\frac{1}{\sqrt{3}} \begin{bmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & 1 \end{bmatrix}. \tag{68}$$

The complete transformation for path 1 is then

$$\begin{aligned} \mathcal{T}_1 &= \mathcal{R}_{2,0}\mathcal{P}_2\mathcal{R}_{1,2}\mathcal{P}_1\mathcal{R}_{0,1} \\ &= \frac{1}{9} \begin{bmatrix} e^{2i\delta} - 2e^{-2i\delta} + 8 & \sqrt{2}(e^{2i\delta} + e^{-2i\delta}) + 2\sqrt{2} \\ -\sqrt{2}(e^{2i\delta} + e^{-2i\delta}) - 2\sqrt{2} & e^{-2i\delta} - 2e^{2i\delta} + 8 \end{bmatrix}. \end{aligned} \tag{69}$$

Simplifying further:

$$\mathcal{T}_1 = \frac{1}{9}\begin{bmatrix} 2i\sin(2\delta) - e^{-2i\delta} + 8 & 2\sqrt{2}(\cos(2\delta)+1) \\ -2\sqrt{2}(\cos(2\delta)+1) & -2i\sin(2\delta) - e^{2i\delta} + 8 \end{bmatrix}. \tag{70}$$

The construction of the Jones transformation for path 2 is done in an identical manner, which will not be reproduced here. One needs the reverse rotation matrices, $\mathcal{R}_{0,2}$, $\mathcal{R}_{2,1}$, and $\mathcal{R}_{1,0}$, which differ from their forward versions only in reversing the signs of the off-diagonal elements. The resulting full matrix is found to differ from $\mathcal{T}_1$ only in that way:

$$\begin{aligned} \mathcal{T}_2 &= \mathcal{R}_{1,0}\mathcal{P}_1\mathcal{R}_{2,1}\mathcal{P}_2\mathcal{R}_{0,2} \\ &= \frac{1}{9}\begin{bmatrix} 2i\sin(2\delta) - e^{-2i\delta} + 8 & -2\sqrt{2}(\cos(2\delta)+1) \\ 2\sqrt{2}(\cos(2\delta)+1) & -2i\sin(2\delta) - e^{2i\delta} + 8 \end{bmatrix}. \end{aligned} \tag{71}$$

Now we are ready to return to the expression for the intensity. Denoting as $f(\eta)$ the terms which modulate the squared sinc function, we have

$$f(\eta) = \left[|(\mathcal{T}_1)_{12}e^{-i\eta} + (\mathcal{T}_2)_{12}e^{i\eta}|^2 + |(\mathcal{T}_1)_{22}e^{-i\eta} + (\mathcal{T}_2)_{22}e^{i\eta}|^2\right]. \tag{72}$$

With the Jones matrix elements this is

$$\begin{aligned} f(\eta) = &\frac{32}{81}(\cos(2\delta)+1)^2\sin^2(\eta) \\ &+\frac{4}{81}(65 - 16\cos(2\delta) + 8\sin^2(2\delta))\cos^2(\eta). \end{aligned} \tag{73}$$

The η dependence can be consolidated in terms of $\cos(2\eta)$ alone, according to

$$A\sin^2(\eta) + B\cos^2(\eta) = \frac{A+B}{2} + \frac{(B-A)\cos(2\eta)}{2}. \tag{74}$$

And grinding through the terms, we finally arrive at

$$f(\eta) = 2 + [\frac{38}{27} - \frac{64}{81}\cos(2\delta) - \frac{16}{81}\cos(4\delta)]\cos(2\eta). \tag{75}$$

It was verified that this gives the same spatial dependence modulating the squared sinc function as is given by Mahan and Price[7], which they reduced

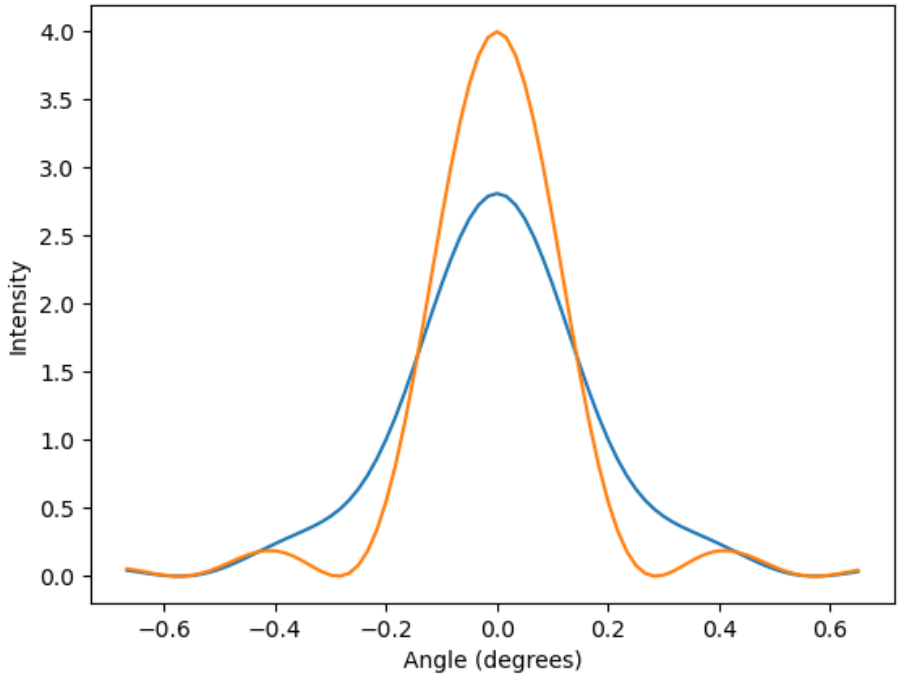

Figure 8: Calculated diffraction patterns for a slit to wavelength ratio of 200. The red curve is the normal pattern. The blue curve shows the broadening caused by the roof prism.

further so as to eliminate δ in favor of refractive index. * We prefer here to keep explicit δ dependence as we will next be looking at how to eliminate the line broadening by changing δ.

The final expression for the intensity profile is

$$I = I_0 \,\mathrm{sinc}^2(\eta) f(\eta), \tag{76}$$

where we have absorbed the constant terms into I_0.

In Figure 8 we plot both the nominal diffraction pattern for the case of no roof prism, along with the results from the above calculation using a refractive index of 1.57. The significant broadening in the width of the

*The Mahan and Price expression is surprisingly compact: $f(\eta) = 2[1 + (1 - 24n^4 - 6n^2)/(3n^2 - 1)]^2 \cos(2\eta)$

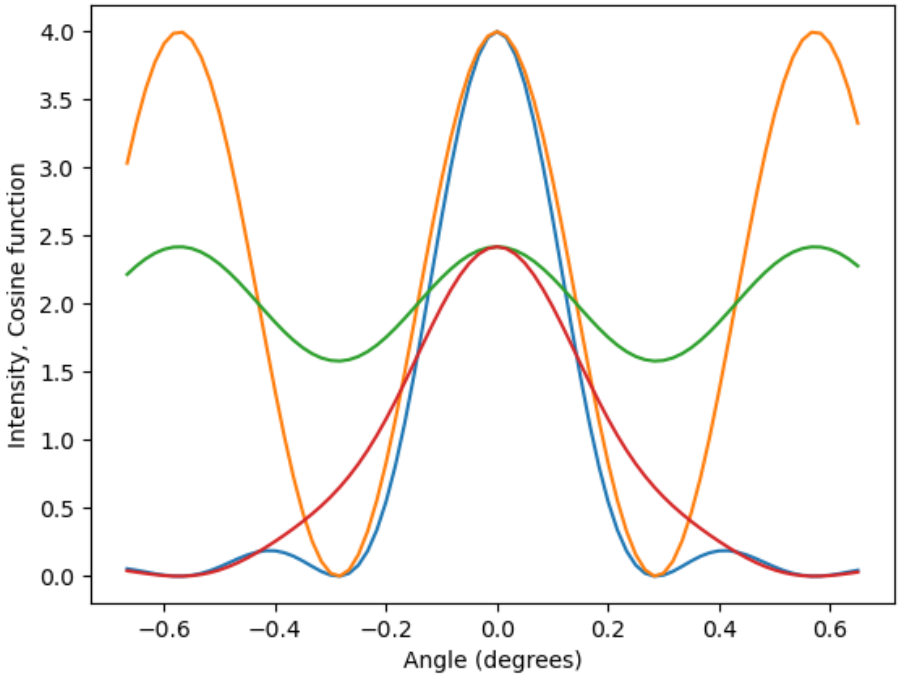

Figure 9: Diffraction patterns along with the cosine-based functions which cause the distortion. The orange curve is the cosine function when the phase shift is nominal, leading to the undistorted pattern shown in blue. The green curve is the worst case, with a phase difference of zero; the red curve shows the broadened diffraction peak that results.

central line and the adjacent lines is obvious. Here the intensity is plotted in arbitrary units, and we used a value of 0.02 for the ratio of slit width to wavelength. The x-axis is in degrees, and specifies η, which is proportional to the angle θ as is customarily used for the display of such patterns.

The degradation of resolution can in fact be made worse. It is worth exploring the function $f(\eta)$ in order to see how it is affected. Differentiating equation (75) with respect to δ reveals that the extreme values will be found for $\delta = 0, \pi/2, \pi, ...$ and so on. In Figure 9, we plot the $f(\eta)$ function along with the squared-sinc term for several limiting cases.

It may come as a surprise that according to this analysis, setting the p-

and s-polarization phase shifts equal to one another leaves the diffraction pattern degraded. Instead we must aim to set $2\delta = \pi$ or an odd integer multiple of π. From the definition of δ, the means we will need $\delta_P - \delta_S = \pm m\pi/2$, where $m = 1, 3, 5, ...$ This is a consequence of a tacit choice of sign convention used for the Fresnel equation for r for p-polarization. Historically there has been no agreement as to the "best" coordinate orientation used to describe the reflected wave. Compare, for example, equations 4.40 in Hecht [1] and 9.109 in Griffiths [6], and note that they differ in sign. To this point we have chosen the convention used by Mahan and Price [7] for consistency with that treatment. As we will see in the next section, early work on the topic of mitigating the problem confusingly employed the opposite approach.

Such is the result when the slit is parallel to the roof prism's center line. What if we now rotate the slit by 90°, so that the diffraction pattern is rotated by the same amount, as shown in Figure 6(b)? We can immediately see that a calculation isn't necessary. Light from either half of the slit aperture must travel along both paths, so there can be no new spatial dependence introduced by the prism that could affect the result. Hence the curious patterns seen by early workers when imaging a cross-shaped source, which would broaden along one direction but not the other [2].

4 Mitigation using coatings

Finally we consider how the diffraction pattern might be brought back to its normal sinc-squared form, corresponding to the best possible resolution. As shown above, we will need to tune the difference in the p- and s-polarization phase shifts to one of several specific values. We will restrict our search to only those solutions which maintain TIR. Therefore we will not consider depositing metal layers, but only additional dielectrics between the glass and air.

Mauer [9] demonstrated that by using layers with carefully chosen refractive indices and thicknesses, it is possible to alter the phase difference adequately, at least for a range of incident angles and wavelengths. Here we will roughly follow the approach of that work, but with the details of the calculation spelled out in greater detail.

Recall that the difference in phase shifts affects the diffraction pattern via equation (75), in the form of $\cos(2\delta)$ and $\cos(4\delta)$ terms, where

$$
\begin{aligned}
2\delta &= \delta_P - \delta_S \\
&= 2\arctan\left(\frac{n\sqrt{n^2\sin^2\theta_I - 1}}{\cos\theta_I}\right) - 2\arctan\left(\frac{\sqrt{n^2\sin^2\theta_I - 1}}{n\cos\theta_I}\right).
\end{aligned} \tag{77}
$$

We can use

$$
\arctan A - \arctan B = \arctan\left(\frac{A-B}{1+AB}\right), \tag{78}
$$

to write

$$
2\delta = 2\arctan\left(\frac{\cos^2\theta_I\,(n-1/n)\,\sqrt{n^2\sin^2\theta_I - 1}}{\cos^2\theta_I + n^2\sin^2\theta_I - 1}\right). \tag{79}
$$

Setting $\theta_I = 60°$, as that corresponds to the case of normal incidence on the Amici prism as worked out above, this becomes

$$
2\delta = 2\arctan\left(\frac{(n-1/n)\sqrt{3n^2-4}}{2+4\sqrt{3n^2-4}}\right). \tag{80}
$$

We first note that changing the index of the prism itself will have only a small effect. The range of indices available for real materials is limited, with most glasses having an index below 1.9, and for a 60° incidence, the index must be above 1.154 to maintain TIR. At these endpoints, the value of 2δ changes from 0 to about 0.56 radians, far from the nearest desired value of π.

Consider next the effect of adding a single dielectric layer with an index less than that of the glass, with the intention of creating TIR at the new glass/layer interface. Recall that the index value n in equation (80) is a ratio, which will be $n = n_{glass}/n_{layer}$. The coating will need an index between 1 and about 1.35 so as to keep TIR at 60°, leading to a phase shift delta that will vary only between 0 and about 0.33.

If the coating does not produce TIR at the glass/layer boundary, we will obtain some partial reflection at this interface, with the rest of the light transmitted into the dielectric. TIR may occur at the layer/air boundary, but it is not immediately clear if this new configuration will change the phase offset sufficiently. To determine exactly what transpires, we must work through the full solution for the reflected electric field resulting from

plane wave oblique incidence from glass into a dielectric layer of index n_1 and thickness d_1, followed by air.

The introduction of one of more dielectric coatings complicates the situation in a non-trivial way, as there will be a wavelength sensitivity to the results because of the finite layer thickness(es). A convenient way to handle this is the transfer matrix method (TMM) [8], which is similar in spirit to that of the Jones matrices employed above. Here, each layer and each boundary in a planar stack of multiple dielectric materials is represented by a relatively simple 2×2 matrix. To analyze the entire structure, we construct a single 2×2 transformation by multiplying the individual matrices in the appropriate order.

The TMM applied here used the following conventions. We assume incident light arrives from the left, characterized by an $\exp(-ikx)$ type term, and that the left side of the boundary has index n_i while the right side as n_j. The boundary transition is then characterized by the matrix $\mathcal{D}_{i,j}$, where

$$\mathcal{D}_{i,j} = \frac{1}{t_{i,j}} \begin{bmatrix} 1 & r_{i,j} \\ r_{i,j} & 1 \end{bmatrix}, \tag{81}$$

where for s-polarized light, the Fresnel equations give

$$r_{i,j} = \frac{n_i \cos\theta_i - n_j \cos\theta_j}{n_i \cos\theta_i + n_j cos\theta_j}, \tag{82}$$

and

$$t_{i,j} = 1 + r_{i,j}. \tag{83}$$

For p-polarization,

$$r_{i,j} = \frac{n_j \cos\theta_i - n_i \cos\theta_j}{n_j \cos\theta_i + n_i cos\theta_j}, \tag{84}$$

and

$$t_{i,j} = \frac{n_i}{n_j}(1 + r_{i,j}). \tag{85}$$

We emphasize that here we are choosing the same sign convention as used by Mauer [9] in order to stay consistent with that work. This differs from that used by Mahan and Price [7], as discussed in the previous section. The upshot is a difference of π radians must be applied in order to have agreement.

This means that below we will be looking for the more intuitively satisfying condition of a zero phase offset to provide the solution for our degraded image.

We account for the layer thickness with a propagation matrix of the form

$$\mathcal{P}_i = \begin{bmatrix} e^{-i\Phi_i} & 0 \\ 0 & e^{i\Phi_i} \end{bmatrix}, \tag{86}$$

where Φ_i is the phase shift upon traversing the ith layer and is proportional to the layer thickness divided by the wavelength.

The total transfer matrix that connects the left side (in our case the glass of the prism, denoted with index 0) with the far right side (air, denoted by index N) is given by

$$\mathcal{M} = \mathcal{D}_{0,1}\mathcal{P}_1\mathcal{D}_{1,2} \times \ldots \mathcal{P}_{N-1}\mathcal{D}_{N-1,N}. \tag{87}$$

We consider first the case of a single dielectric layer, with an index of 1.316, which ensures that a 60° incidence at the glass($n = 1.517$)/dielectric boundary results in partial reflection and transmission. Now we vary the film thickness and calculate the resulting phase angles for the reflected wave. In order to stay consistent with the approach in [9], we treat the independent variable as the phase thickness, expressed in degrees. Results for the phase shifts for s- and p-polarizations are plotted in Figure 10.

We see that a thickness of 90° (a quarter-wave, accounting for the longer path due to non-normal incidence) brings the difference to the desired value of 0°. Let's break down why this is happening. From Snell's law, we see that the transmission angle in the dielectric is a steep 86.7°, which will also be the angle of reflection at the air boundary. Equation (80) then gives a phase difference of less than 0.3°. The final offset will depend on the phasor addition with the portion of the light reflected from the first boundary, hence the thickness dependence, from which we can select the optimal value.

This thickness, of course, will not be nominal for light at other wavelengths. In Figure 11 we augment the results by including curves corresponding to the extremes of the visible spectrum (380 and 750 nm) as well as the midpoint wavelength (565 nm). Here we plot the difference between s- and p-polarization phases. The x-axis phase is based on the 565nm wavelength. Choosing to tune to zero phase shift for this central wavelength, we will obtain a phase offset at 380 nm of about 3.6°, while at 380 nm it will be near −1.6°. These are small offsets: plotting the diffraction pattern for

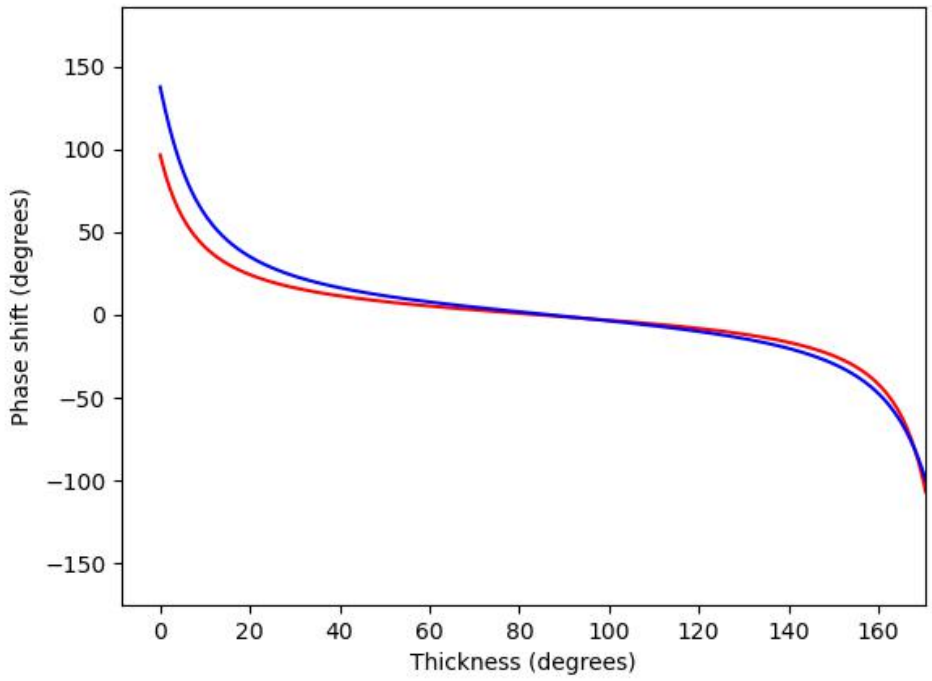

Figure 10: Calculated phase shifts in degrees versus layer thickness, as a phase expressed in degrees. Red is for the s-polarization, blue is the p-polarization. A phase thickness of 90° (corresponding to a quarter-wave) produces zero phase shift offset.

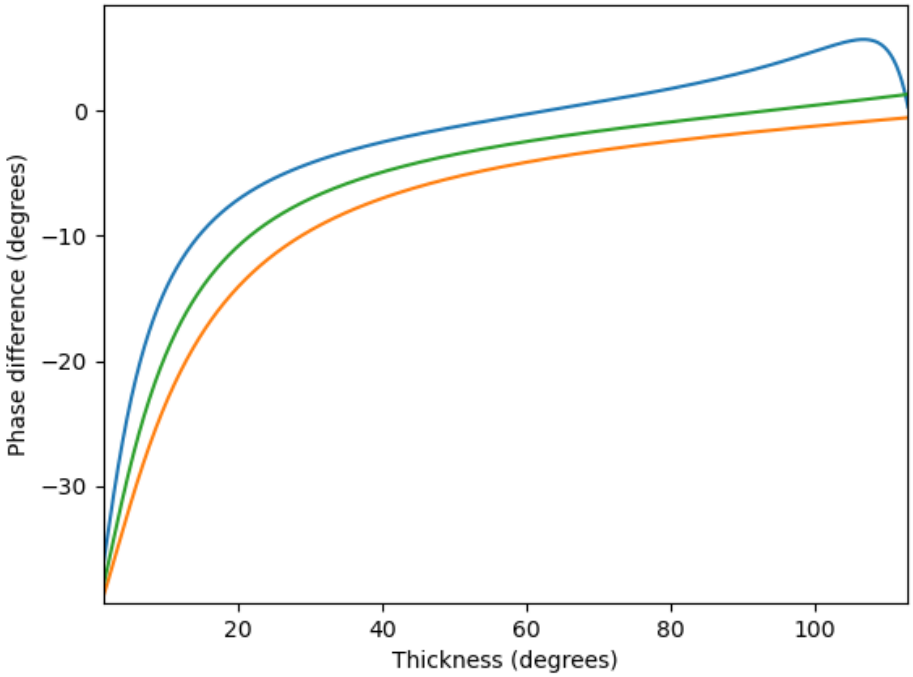

Figure 11: Difference in phase between p- and s-polarizations in degrees versus phase thickness in degrees based on a 565nm wavelength. Blue is for 380nm, green is for 565nm, and red is for 750nm.

the previous section does not produce discernible differences, even using a logarithmic axis.

The drawback to this solution is that it requires the incidence angle not exceed 60°, because then, TIR will occur at the glass/layer interface, ruining the effect. This limitation to a single layer approach was noted by Mauer[9], who proposed the use of multiple layers to get around it. Specifically, he studied the effects of three layer systems arranged as a "Herpin equivalent"[8] to a single layer. This approach uses a central layer of index n_Q and thickness t_Q sandwiched between two identical layers each of index n_P and thickness t_P, which behaves in some ways as a single layer, while providing more degrees of freedom for tuning the phase shift, wavelength dependence, and span of allowable incident angles.

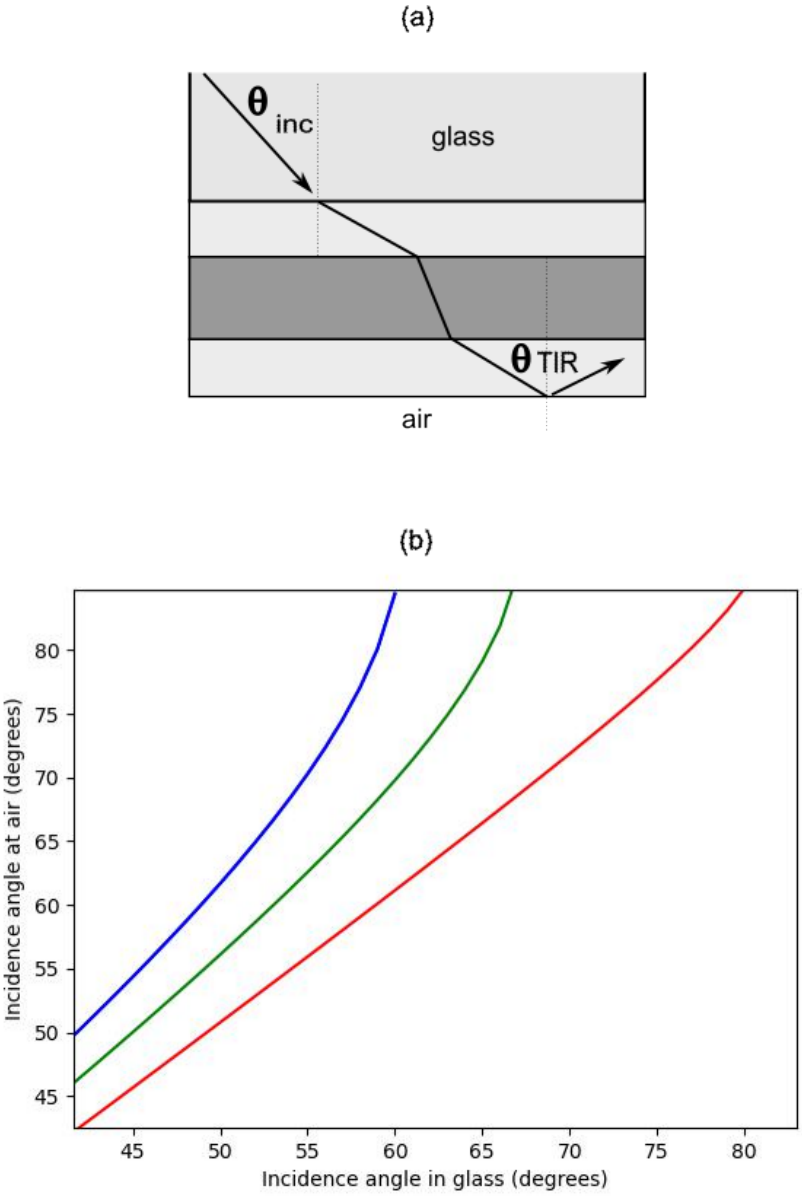

Figure 12: (a) Series of three films as a Herpin equivalent, lying between the glass and air. (b) Calculated incidence at air angle versus incidence at glass, both in degrees. The blue curve is for a single film; the other two are for three-layer systems which allow a greater range of incident angles at the glass boundary.

Figure 12 illustrates how this can help increase the range of incidence angles. Diagram (a) shows a schematic of a three layer system. Diagram (b) shows how the incidence angle at the air interface changes as a function of the angle at the glass boundary. The blue curve is for a single layer, using the same parameters that were used to generate Figure 11. TIR occurs at the air layer as long as the angle of incidence at the glass is greater than about 42°, but we can only go up to 60°. The green curve is for a three layer system, with the proximal layer having $n_P = 1.4$ and the center layer having $n = 2.3$. We can now move out to larger angles, although we pay a price in that the angle at the air boundary is less steep than for the single layer, meaning the p- and s- phase shift offset is larger. Such is the trade-off that needs to be made. The red curve uses $n_P = 1.5$, increasing the angle of incidence range, but again at the price of less phase correction.

The TMM approach is easily applied to such a multi-layer system, and was used to reproduce the curves published by Mauer These are shown in Figure 13, which corresponds to Figure 2 in [9]. Here the thickness of the outer two layers was set to 3/7th the thickness of the inner layer, and then the total thickness, in phase units, was swept over. A region where the phase shifts curves are nearly tangent, between 200° and 300° is evident, indicating a potential solution space, which Mauer then examines for wavelength and incident angle dependence.

This demonstrates the general strategy for phase compensation via dielectric coatings, but the Mauer design is an early effort and would be far from optimal, because the range of total phase thickness with minimal phase shift, and hence the range of wavelengths, is not very broad. As with other multilayer coatings, broadband performance will require a larger number of carefully tuned layers. As for the specific numbers of layers and their compositions as employed in modern binocular prisms, there is a paucity of published information. This is expected, if inconvenient, since optics manufacturers would be understandably loathe to share proprietary information or trade secrets in such a highly competitive market. We can, however, look to the older patent literature to help build a multilayer model with good spectral performance.

Ito and Noguchi [10] provide a range of coating designs which are based on nine layer structures comprised of three different Herpin equivalent trilayers. By modifying and extending the Mauer TMM analysis, the phase shift as a function of wavelength can be calculated for nine layers. An optimization routine for finding a nominal set of layer thicknesses which minimize the phase

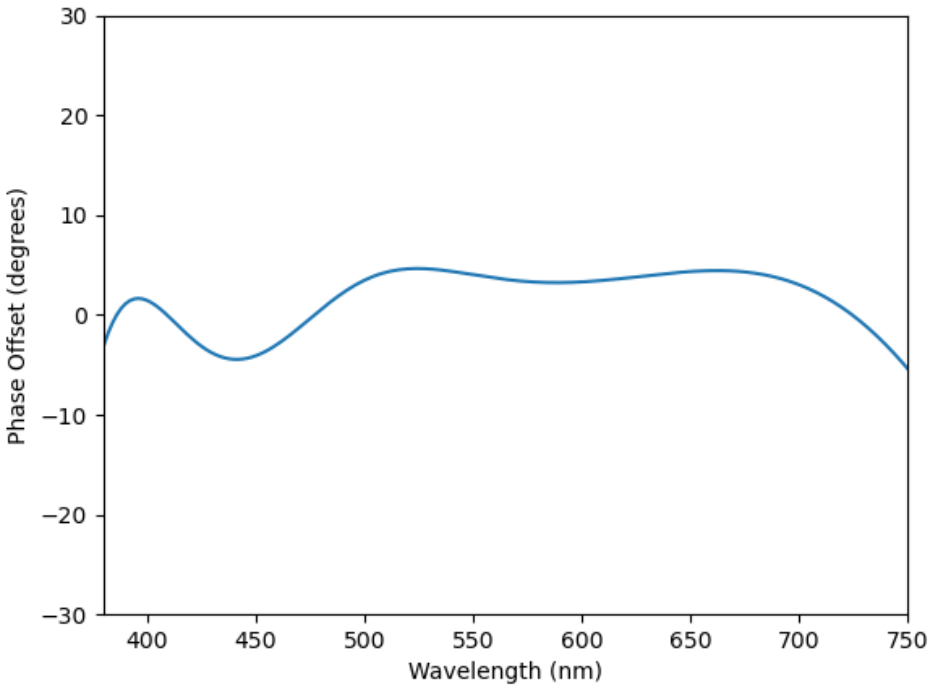

Figure 13: Phase difference in degrees versus wavelength in nm, for a nine-layer coating described above. The magnitude of the phase shift is less than 5 degrees over the visible range.

shift over the visible range was also written to explore the design space cited in the patent. A specific result is shown in Figure 14 below, where we are plotting the difference in p- and s-polarization phase for a multilayer reflection as a function of wavelength (in nm). This is based on an incidence angle of 49°, a glass index of 1.5181, followed by a series of layer having thicknesses (in nm) of $33.27, 105.53, 54.57, 163.13, 75.76, 96.81, 118.5, 223.14$, and 126.15, and indices of $2.055, 1.388, 2.055, 1.4693, 2.055, 1.4693, 2.055, 1.388$, and 2.055. This represents a design far more applicable for practical use in a roof prism than that from the early Mauer proposals.

It should be noted that changes in the incidence angle will cause the phase shift to grow for this design, so even a nine-layer coating leaves room for further refinement. As the Ito and Noguchi patent is nearly a quarter of a

century old, it is expected that state-of-the-art phase coatings have expanded the design space even further.

5 Conclusions

The problem of resolution loss caused by roof prisms, and its mitigation via dielectric coatings, is not a topic that is widely addressed, despite the ubiquity of roof prisms in modern optical devices. Nontechnical explanations endemic to manufacturers websites provide only facile overviews, at best. To our knowledge this is the only treatment that provides a detailed overview of this curious problem and its solution.

Working through this topic would be a valuable exercise appropriate for a senior-level undergraduate, for various reasons. First, scant attention is paid to the phase shifts that accompany TIR, even though a derivation of the Fresnel equations is expected to be undertaken in any course on electromagnetism. Second, the connection between the phase shifts and the exponential decay associated with the evanescent wave "behind" the TIR can be intuitively illustrated via a model involving transverse waves on strings in the presence of critical damping, and is a good case study in how phase shifts upon reflection are a consequence of boundary conditions. It also provides a neat optical analogue to tunneling into a finite potential, an integral effect in quantum mechanics that can be difficult to envision. Third, the ambiguity in the sign convention for the p-polarization reflection coefficient can be highlighted here; an unawareness of this subtlety may lead to unwelcome future surprises. Fourth, the ray tracing and application of the Jones matrices present tractable yet challenging exercises that hone skills for setting up non-trivial three-dimensional problems and carefully working through matrix manipulations. Fifth, developing the expression for the modified Fraunhoffer pattern requires good understanding of diffraction and polarization; topics which are often presented sequentially. A problem requiring the use of both concepts may be of pedagogical value. Finally, it provides a good learning vehicle for the practical application of transfer matrices.

References

[1] E. Hecht, "Optics", 5th ed., Pearson (2016).

[2] A.I. Mahan, "Some newly solved and some unsolved problems in optics," J. Wash. Acad. Sci. 44, No. 6, pp. 165-194 (1954).

[3] L. Jinjun, S. Xueping, and Z. Weibing, "The splitting mechanism of zero order diffraction pattern by roof prisms," J. Opt. 42, 367–375 (2013).

[4] T.W. Murphy Jr and S.D Goodrow, "Polarization and far-field diffraction patterns of total internal reflection corner cubes," Appl Opt. 52(2):117-26 (2013).

[5] S. Xueping, W. Liu, and J. Lu, "Analysis and correction of far-field diffraction pattern for corner-cube reflector," J. Opt. 46 (2017).

[6] D.J. Griffiths, "Introduction to Electrodynamics," 4th ed., Cambridge University Press (2017).

[7] A.I. Mahan and E E. Price, "Diffraction Pattern Deterioration by Roof Prisms," J. Opt. Soc. Am. 40, 664-686 (1950).

[8] A. Sarangan, "Optical thin film design," 1st ed., CRC Press (2020).

[9] P. Mauer, "Phase Compensation of Total Internal Reflection," J. Opt. Soc. Am. 56, 1219-1221 (1966).

[10] Ito T. and Noguchi, M. "Viewing optical instrument having roof prism and a roof prism," US Patent US6304395B1 (2001).