The Critical Angle and Beyond

Angus Macleod
Thin Film Center, Inc, Tucson, AZ

Introduction

Fill a cylindrical drinking glass with water and look down into it through the surface of the water. The wall of the glass below the water surface looks metallic with very high reflectance and it is impossible to see through it. This is the phenomenon of total internal reflection. Dip a finger in the water and touch the outside of the glass. Now you can see the surface of your wet finger through the glass. You have, by application of the water, frustrated the total internal reflection.

The effect is well known and has been for, at the very least, hundreds of years. When this phenomenon was first observed is unknown, but surely early divers, if no on else, must have observed it looking upwards towards the surface. The earliest modern account appears to be that in the book, Dioptrice, of Johannes Kepler (1571–1630) published in 1611. Clearly, it was well known and understood by the time of Isaac Newton (1642–1727). Newton refers to it in passing on the third page of his Opticks in a way that suggests no explanation was necessary: “As if Light pass out of a Glass into Air, and by being inclined more and more to the common Surface of the Glass and Air, begins at length to be totally reflected by the surface; those sorts of Rays which at like Incidences are reflected most copiously, or by inclining the Rays begin soonest to be totally reflected, are most refrlexible.”

Completely to cover the phenomenon and all its applications is utterly beyond the possible scope of a Bulletin article. We shall concentrate, therefore, on some of the fundamentals and on some implications and applications in the field of optical coatings.

Fundamentals

The simple explanation of total internal reflection is as a consequence of refraction. We imagine a plane boundary between two media, characterized by refractive indices \( n_0 \) and \( n_1 \), and a plane wave incident on the boundary from the \( n_0 \) side. The incident wave will be divided into two plane waves by the boundary. One will be reflected and one transmitted into the second medium. The directions of all three waves will be coplanar with the normal to the surface, the plane being known as the plane of incidence. Let the angle between the incident wave and the surface normal be \( \theta_i \), and between the transmitted wave and the normal, \( \theta_t \). These angles are related by the expression:

\[
\sin \theta_i = \frac{n_0 \sin \theta_t}{n_1}
\]

(1)

known by various names but probably most commonly the Law of Refraction or Snell’s Law. Equation (1) can be written in terms of \( \theta_i \) as

\[
\theta_t = \arcsin \left( \frac{n_0 \sin \theta_i}{n_1} \right)
\]

(2)

This is a well-behaved expression as long as \( n_0 \leq n_1 \). However, \( \sin \theta_1 \) can range from zero to unity, and, if \( n_0 > n_1 \) then there comes a point, as \( \sin \theta_1 \) increases, where the argument of the arcsin function becomes greater than unity and the expression can not be solved for a real \( \theta_1 \). The value of \( \theta_1 \) at which this occurs is known as the critical angle and since, for angles beyond the critical angle no wave can propagate into the second medium, all the incident power is reflected. The phenomenon is called total internal reflection, the acronym TIR being frequently used. The critical angle is given by

\[
\theta_{\text{critical}} = \arcsin \left( \frac{n_1}{n_0} \right)
\]

(3)

For crown glass with refractive index 1.52, say, bounded by air of refractive index unity, the critical angle is 41.14°, conveniently below 45°. Total internal reflection has opened up an enormous range of applications. It is an incredibly simple way of achieving a reflectance of 100%. 45° is a particularly useful angle of incidence because the light is turned through 90°, easing the design of instruments using the effect. A simple 90° isosceles prism becomes a very efficient reflector. The light can enter normally through one face and, after reflection at the hypotenuse, can exit normally through the other, having been turned through 90°. Two such prisms make a simple periscope. The light can also enter a similar prism normally through the hypotenuse and be returned in the opposite direction after two reflections. Such a prism is frequently used with another similar one as an erecting system in a telescope when the prism is known by the name Porro. Two simple prisms are shown in Figure 1. Systems based on prisms rather than plane mirrors have an advantage in being dimensionally very stable.

![Figure 1. Two similar right-angle prisms that use total internal reflection. That on the left paired with a similar one would form a simple periscope. The one on the right employs two reflections and paired with another similar one is frequently used to invert the image in a binocular telescope, when it is known as a Porro prism.](image)

Physical Principles

Let us suppose a plane harmonic wave is incident obliquely on a plane surface as sketched in Figure 2. To avoid too much clutter, only the incident and transmitted waves are shown but, of course, in reality there will be a reflected wave also. The frequency of each wave is identical and each has planes of constant phase that are separated by a distance of one wavelength measured along the direction of propagation. Although their free space wavelengths, \( \lambda \), are identical, their local wavelengths, \( \lambda/n \), and hence their spacings are different, because of the difference in refractive index.

The intersections between the planes for the two waves and the surface between the media, must line up, otherwise the boundary conditions cannot be satisfied. The actual phases associated with the planes will adjust themselves to include also the reflected wave. This leads to the relationship

\[
\frac{\lambda}{n_0 \sin \theta_i} = \frac{\lambda}{n_1 \sin \theta_t}
\]

(4)

that is simply Snell’s Law written in a different way. The velocity of the
where \( c \) is the velocity of light in vacuo. Since \( n \sin \vartheta \) will be less than unity when \( \vartheta \) is small, the velocity can be many times \( c \). This is not in any way strange. Imagine a row of marching men crossing a line slightly obliquely. Although the marching speed will be only a few miles per hour, the point of intersection can travel along the line at enormously greater speeds. It is very easy to see in this simple example that there is no violation of Special Relativity, but there are cases where a similar effect can be part of a much more involved system, and consequently more misleading. Then it can be tempting to claim such a violation. Treat any such claim with the deepest suspicion. It is bound to be incorrect.

As \( \vartheta_0 \) increases, the distance between the intersections of the constant phase planes with the surface shrinks. The minimum distance that can be supported by a progressive, plane harmonic wave in the second medium is \( \lambda n_1 \). If \( n_0 \) is greater than \( n_1 \) then that point will be reached, when the wave in the second medium must travel parallel to the surface. The corresponding angle, \( \vartheta_0 \), is known as the critical angle. Past that point, the spacing between the intersections is too small. The electric and magnetic fields, however, cannot simply stop at the boundary. They do penetrate into the second medium, where they carry power parallel to, and close to, the interface, but they decay exponentially away from it. The planes of constant phase continue to oscillate at the same frequency because the effects are still linear. The wave is said to be evanescent in the \( z \)-direction. The exponential fall along the \( z \)-axis becomes greater and greater as the interval between the planes is reduced. Since no power is transported in the \( z \)-direction into the second medium, all the incident light is reflected. This is the phenomenon of total internal reflection (TIR). The total reflection at the boundary is an interesting parameter. The frequency of the waves can be taken as \( f \) when the velocity becomes either of the expressions times frequency.

\[
\nu = \frac{f \lambda}{n \sin \vartheta} = \frac{c}{n \sin \vartheta}
\]

continued on page 16
The Critical Angle and Beyond
continued from page 15

boundary is accompanied by a phase change in reflection that varies with the angle of incidence, as indicated in Figure 3.

The mean power transported by a plane harmonic light wave is proportional to the product of the electric and magnetic field amplitudes and the cosine of their phase difference. The wave that is evanescent in the z-direction does have components of both electric and magnetic field in the x-y plane but they differ in phase by 90° and that is why they carry no power. Except at, and very close to, the critical angle, the exponential fall in the fields is very rapid. Figure 4, for example, shows the fields for s and p-polarization calculated at 45° for incidence in glass of index 1.52. The curious behavior for p-polarization on the glass side where the field does not fall to zero is because we are plotting total fields and the x and z-components of the p-field do not fall to zero together. Should, now, a plane surface be placed very close to and parallel to the first surface, it can intrude into the exponential tail of the fields. This surface will have the same pattern of intersections with the constant phase planes as the first surface. Suppose the other side of the second surface consists of material with refractive index $n_0$.

Since a progressive plane wave was responsible for the spacing of the intersections at the first surface, and since this second surface has exactly the same spacing, then a plane harmonic wave, exactly similar to the incident wave can be induced in the third medium. Unlike the evanescent wave, this one will carry power. The power has to be transported somehow across the gap by the evanescent wave and it does this by modifying the phase difference between the electric and magnetic fields. It is a rather special case of thin-film interference and is sometimes referred to as tunneling. The power is lost to the reflected wave, which is reduced so that there is no longer total reflection. The effect is known as frustrated total internal reflection, with abbreviations FTIR, or sometimes FTR, or, in spectroscopy, ATR (for attenuated total reflectance when absorption is present). Unfortunately these abbreviations are shared with others. FTIR, for example, is also used for Fourier Transform Infrared Spectroscopy. Of course the index of refraction of the final emergent medium material need not be exactly that of the incident medium in order to frustrate the total internal reflection. It is sufficient that the incidence should not be beyond the critical angle for the combination of incident medium and emergent medium.

Figure 3. The phase shifts use the usual thin-film convention. The p-shift changes to 180° at the Brewster Angle. Beyond the critical angle, both p and s-shifts vary through a total of 180°. For the ellipsometric sign convention, the p-shift should be altered throughout by 180°.

Figure 2. The mean power transported by a plane harmonic light wave is proportional to the product of the electric and magnetic field amplitudes and the cosine of their phase difference.

A fast MFC for demanding vacuum applications

The new MCV series MFC from Alicat Scientific offers a host of performance features.

- 50ms-100ms response and settling time to set-point changes.
- Integrated shut-off valve for positive shut off to 1X10⁻⁹ sec/sec He.
- Full scale flow rates of 0.5SCCM - 20SLPM.
- Local LCD display with push button controls.
- Available with a variety of analog and digital inputs and outputs, including RS-485 and PROFIBUS.
- Input power requirement of 12VDC-30VDC.
- Zero warm-up time.
- Lifetime Warranty

Contact Alicat Scientific today at 888-290-6060 or visit us on the web at www.alicatscientific.com to learn more about the MCV series MFC.
components that are eigenmodes of polarization known as  out that incident light has to be resolved into two linearly polarized  we will look at it in rather more depth. In the earlier article we pointed  theory. We discussed oblique incidence in an earlier article \[1\] but now  zero on the horizontal scale. The input power is assumed to be 1Wm\(^{-2}\).

At the critical angle, the phase differences between the electric field of the incident and reflected beams are such that there is an enhanced field across the boundary. The direction of this magnified field follows  the polarization of the incident wave, and it is frequently employed to  field across the boundary. The direction of this magnified field follows  of the incident and reflected beams are such that there is an enhanced  Figure 5. The reference axis at the surface, the directions of the light waves and the reference direction for electric field in the two polarizations.

The boundary conditions at the surface are the continuity across  of the total electric and magnetic fields parallel to the boundary, the  wave are (\sin \theta, 0, \cos \theta) for s-polarization and \(y\cos \theta\) for p-polarization. Although this is absolutely correct also for  probability is interpreting the complex angles that emerge from Snell’s Law. We therefore return to the fundamental physical principles.

We set up the usual right-handed set of reference axes, shown in  the transmitted wave gives:

\[
\eta s \cos \theta = \eta r \sin \theta
\]

This will usually be interpreted as Snell’s Law but we will simply leave it as it is. We include the possibility that \(\eta r\) may be complex.

The reflected wave we can see with very little analysis will have phase factor

\[
\exp \left[ i \left( \frac{2\pi n_0}{\lambda} (x \sin \theta - z \cos \theta) \right) \right]
\]

The transmitted wave is more complicated. From we have the coefficient of \(x\). We need the coefficient of \(z\). The direction cosines are connected by

\[
\alpha^2 + \beta^2 + \gamma^2 = 1
\]

Then

\[
\frac{2\pi n_1}{\lambda} \gamma = \frac{2\pi}{\lambda} \left[ \eta_{r0}^2 - \eta_{r1}^2 \sin^2 \theta \right]^{1/2}
\]

There are two roots and it is important that we choose the correct one. Physical reasoning comes to our assistance. This root governs the propagation of the \(z\)-component of the transmitted beam and this should be continued on page 18.
The Critical Angle and Beyond

continued from page 17

along the positive direction of the z-axis. The most general case has
$n_1$ complex, of value, say, $n_1 - ik$. The argument, therefore, has a positive
or negative real part and, if $k$ is not zero, a negative imaginary part.
In other words it is always in the third or fourth quadrant. The roots,
therefore, are in the second or fourth quadrant and we can see that a
fourth quadrant solution with its positive real part and negative imagi-

nary part has the form expected on physical principles.

Now that we have the phase factors correct, we need to see how
the tilted admittances $\eta_s$ and $\eta_p$ can be calculated. The tilted admitt-
itance is the ratio of the complex tangential magnetic field amplitude,
to the complex electric field amplitude. In $s$-polarization the electric
field amplitude is already tangential, while in $p$-polarization it is the
magnetic field that is tangential. We write these two components as $E_s$
or $E_p$ and we can calculate the other component in each case from the
vector relationships:

$$\vec{H} = y(x\hat{x} \times \vec{E})$$

and

$$\vec{E} = \frac{(x \hat{x} \times \vec{H})}{-y} = \frac{-y(x\hat{x} \times \vec{H})}{y^2}$$

(12)

where

$$y^2 = y(\alpha \ell + \beta j + \gamma k)$$

(13)

We will substitute $n_1$ free space units for $y$.

$s$-polarization magnetic: $\vec{E}_s = -y\gamma \vec{E}_n = -\left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2} \vec{E}_n$ (14) and

$p$-polarization electric: $\vec{E}_p = \frac{-y\gamma \vec{E}_n}{y^2} = \frac{\left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}{n_1} \vec{E}_n$ (15)

The minus sign in Equation (14) is simply a consequence of the arrange-
ment of coordinate axes. With positive electric field along the $y$-axis and
wave direction along the $z$-axis, the positive magnetic field direction
must be opposite to the $x$-axis. Thus we should remove the minus sign
when calculating the tilted admittance.

$$\eta_s = \frac{-\vec{H}_s}{\vec{E}_s} = \frac{\left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}{n_1}$$

(16)

and

$$\eta_p = \frac{-\vec{H}_p}{\vec{E}_p} = \frac{n_1^2}{\left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}$$

(17)

where the root must be in the fourth quadrant. Then

$$\rho_s = \frac{n_0 \cos \vartheta_0 - \left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}{n_0 \cos \vartheta_0 + \left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}$$

(18)

and

$$\rho_p = \frac{n_0 \cos \vartheta_0 + \left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}{n_0 \cos \vartheta_0 - \left[ n_1^2 - n_1^2 \sin^2 \vartheta_0 \right]^{1/2}}$$

(19)

with similar expressions for $r$. The extension to reflectance and trans-
mittance is clear. The cos $\vartheta_0$ factor in the first terms in each numerator
and denominator, makes it a little more difficult to estimate values,
especially in the admittance diagram, because we are used to having
zero reflectance when an admittance is equal to that of the incident
medium. This problem can be avoided if the numerator and denomina-
tor in the expressions are multiplied or divided, as appropriate by cos $\vartheta_0$.
This leads to what are called the modified admittances:

![Remote RF Plasma Cleaning](image)

Remote RF Plasma Cleaning
Evactron® De-Contaminators

The Evactron D-C is a remote RF plasma cleaner that uses air to
produce oxygen radicals to oxidize and remove carbon by downstream
ashing in electron microscopes and other vacuum chambers. Over 800
sold for scanning electron microscope cleaning.

LOW POWER RF GENERATOR

The new RFS-20 is the low power,
13.56 MHz RF generator used in the
Evactron D-C, available as a stand
alone unit with stable RF power
output between 0-20 Watts. It has 3
digit readout for repeatable RF
settings, reverse and forward power
readouts and remote control
connector. USD $2000 + tax and ship.

Visit [www.Evactron.com](http://www.Evactron.com) or call 1-650-369-0133 for more details
and ordering.

XEI Scientific, Inc.
We now have all the fundamental results that we need for what follows.

The Evanescent Field

Let us assume that \( n_1 \) is less than \( n_0 \sin \theta_0 \), that is we are beyond the critical angle. Then the correct root of \( \left[ n_1^2 - n_0^2 \sin^2 \theta_0 \right]^{1/2} \) will be negative imaginary. We can denote it by \(-i \hat{u}\). The phase factor of the wave is then:

\[
\exp \left( \frac{-i 2 \pi x}{\lambda} \right) \exp \left[ i \left( \frac{-2 \pi n_0 \sin \theta_0 \zeta}{\lambda} \right) \right]
\]

From Equations (12) and (13), the components of electric and magnetic field amplitudes at the boundary are then given for \( s \)-polarization by

\[
\begin{align*}
\mathbf{E}_x &= \mathbf{E}_n \\
\mathbf{E}_y &= i \xi \mathbf{E}_n \\
\mathbf{E}_z &= n_0 \sin \theta_0 \mathbf{E}_n
\end{align*}
\]

and for \( p \)-polarization by

\[
\begin{align*}
\mathbf{H}_x &= \mathbf{H}_n \\
\mathbf{H}_y &= -i \xi \mathbf{H}_n \\
\mathbf{H}_z &= n_0 \sin \theta_0 \mathbf{H}_n
\end{align*}
\]

The power flow along the \( z \)-axis is zero because of the 90° phase difference between the \( x \) and \( y \)-components. The \( y \) and \( z \)-components of electric field and magnetic field, however, are in phase (or 180° out of phase). Thus in both \( p \) and \( s \)-polarization there is power flow along the positive \( x \)-axis. The power density decays following the exponential term along the \( z \)-direction, that is away from the interface, the rate of decay being twice that of the individual fields. The implication of this for optical fibers, or optical waveguides, for example, is that part of their power is actually carried in the lower-index cladding.

Thin-Film and Associated Calculations

Thin film calculations concentrate on the behavior along the \( z \)-axis. The interest is in the tangential components of electric and magnetic field and on the phase propagation along the \( z \)-axis. So far we have assumed lossless dielectric materials but our expressions include the possibility of some absorption. Let us retain the lossless character of the incident medium but let the other materials have a complex index, \( n-ik \). Two parameters are required for the characteristic matrix of a material, the admittance of the material, \( \eta \), and the phase thickness \( \delta \). These are then given by:

\[
\begin{align*}
\eta_s &= \frac{\left[ n_1^2 - n_0^2 \sin^2 \theta_0 - i2nk \right]^{1/2}}{\cos \theta_0} \\
\eta_p &= \frac{(n-ik)^2}{\eta} \\
\delta &= \frac{2 \pi d}{\lambda} \left[ n_1^2 - n_0^2 \sin^2 \theta_0 - i2nk \right]^{1/2}
\end{align*}
\]

where the root is to be taken in the fourth quadrant. These are all that we need to complete the characteristic matrix for a thin film.

The behavior of dielectric materials beyond critical is not unlike that of perfect, lossless metal films. They present many interesting possibilities. A particularly important application of total internal reflection is in the construction of retarders. Normal birefringent retarders have a retardance that is given by:

\[
\Delta = \frac{2 \pi d}{\lambda} \left( n_0 - n_s \right)
\]

We can think of the two indices as being roughly constant with wavelength and then the retardance is inversely proportional to wavelength. It is possible to combine different materials to reduce the variation but achromatic retarders of this kind are corrected only at two or three wavelengths and still show significant variation. Reflection beyond the critical angle induces a retardation in the reflected ray and since this is an effect that depends only on the index of the material it varies much less. A popular and traditional form of retarder is the Fresnel Rhomb that, with two reflections, gives a retardation of 90°, similar to a quarterwave plate, Figure 6 and Figure 7. The Fresnel Rhomb operates over a much wider range than does a birefringent retarder, but there is still some variation due to the dispersion of the material. Generally, with normal dispersion, the retardation increases towards shorter wavelengths. Clapham and colleagues showed [2] that quite simple coatings added to the reflecting surfaces could improve the performance of the Fresnel Rhomb considerably. The examples they gave were primarily for silica rhombs. Here we use the technique to correct the glass rhomb of Figure 6 and Figure 7. A simple coating consisting in this case of 12.9nm of magnesium fluoride is sufficient to give excellent achromatic performance over the region 400 to 100nm. The angle of incidence has to be changed slightly in the coated component as indicated in Figure 7.

Figure 6. The Fresnel Rhomb gives a 90° retardation in two reflections at an angle of around 55° in glass.

Figure 7. The performance of the glass Fresnel Rhomb. The uncoated results refer to uncoated glass and the coated to an added correction layer on the surfaces consisting of 12.9nm of magnesium fluoride. (After [2]).

continued on page 20
Retarders of all kinds become much more straightforward when incidence is beyond critical. Part of the reason is that high reflectance is assured whatever dielectric coating is added to the reflecting surface. Thus the design need consider only the retardance. Usually a few layers are sufficient. Construction of retarders for angles below critical require much more complicated and difficult coatings. There is one caveat. The same problems with angular grasp that we saw in an earlier contribution with the thin-film polarizer [3] apply also in this case. The thin-film reflecting retarder does not have a wide aperture.

Another application was mentioned in [3]. At high angles of incidence the aperture of a thin-film polarizer is increased. One of the problems in the design of these devices is the poor contrast between the properties of the layers for $p$-polarization. Beyond the critical angle, however, this is no longer the case. The tilted admittances of dielectric materials become imaginary and the $s$ and $p$-values have opposite sign. This enhanced contrast eases the design problems enormously. The angles of incidence are such that the low-index material operates beyond critical while the high-index material does not. The critical angle for magnesium fluoride in glass is around 65° and so a design at around 70° is well inside critical for the low-index material. A detailed account of the design of such devices is given by Li and Dobrowolski [4].

There are many more applications of total internal reflection. We shall return to the topic in a future article.

References

Angus Macleod is currently President of the Society of Vacuum Coaters. He was born and educated in Scotland. In 1979 he moved to Tucson, AZ, where he is President of Thin Film Center, Inc. and Professor Emeritus of Optical Sciences at the University of Arizona. His best-known publication is Thin-Film Optical Filters, now in its third edition. In 2002 he received the Nathaniel H. Sugerman Memorial Award from the Society of Vacuum Coaters.

For further information, contact Angus Macleod, Thin Film Center, Inc. at angus@thinfilmcenter.com.