Chapter 3

Geometrical Optics (a.k.a. Ray Optics

3.1 Wavefront

Geometrical optics is based on the wave theory of light, and it may be though of a tool that explain the behavior of light, and helps to predict what will happen with light in different situations. It's also the basis of constructing many types of optical devices, such as, e.g., photographic cameras, microscopes, telescopes, fiber light guides, and many others.

Geometric optics is an extensive field, there is enough material in it to "fill" an entire academic course, even a two-term course. In Ph332, out of necessity, we can only devote to this material three class hours at the mos, so we have to limit ourselves to the most fundamental aspects of this field of optics.

As mentioned, geometrical optics is based on the wave theory of light. We have already talked about waves, but only about the most simple ones, propagating in one direction along a single axis. But this is only a special case – in fact, there is a whole lot of other possible situations. A well-known scenario is a *circular wave*, which can be excited simply by throwing a stone into quiet lake water, as shown, e.g., in this Youtube clip. There are *spherical waves* – for instance, if an electronic speaker or any other small device generates a sound, the sound propagates in all direction, forming a *spherical acoustic wave*.

In addition to that, the functions describing more complicated waves are no longer as simple as those we used in Chapter 2, namely:

$$\Delta y(t) = A \sin\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t\right)$$

In the case of a circular wave spreading over a plane (e.g., on the lake surface), we need two coordinates to describe the position on the plane. So, we can use two axes, x and y, intersecting at the wave source, to describe the position. The distance r between a point of coordinates (x, y) and the wave source is:

$$r = \sqrt{x^2 + y^2}$$

We can no longer use y for displacement, but we can employ the third Cartesian coordinate, z, in this role, and we can write the wave function as:

$$\Delta z(t) = A(r) \sin\left(\frac{2\pi}{\lambda}r - \frac{2\pi}{T}t\right) = A(r) \sin\left(\frac{2\pi}{\lambda}\sqrt{x^2 + y^2} - \frac{2\pi}{T}t\right)$$

Note that the amplitude in the above equation is a function of r. Why? In Chapter 2 we treated A as a constant. O.K., but the waves in Chapter 2 were all moving in one direction and they could retain their initial energy all the way along the x axis. In a circular way, once the wave propagates, its energy spreads over longer and longer circles, so the displacement must gradually decrease with increasing r.

In the case of a three-dimensional (3D) wave, the function becomes even more complicated. Now we need all three Cartesian coordinated for describing the position in space, so that we cannot use z to express the displacement. Well, it can be done, but it depends on the type of wave we deal with. For instance, in a 3D sound wave what is displaced is the pressure, and in a 3D electromagnetic wave it's the electric and magnetic fields. So, there is no a "generic formula" for all types of 3D waves, each wave type has to be individually considered. But how we do that is not particularly relevant in the Ph332 Course, so we won't continue the discussion. we have to focus on certain elements of the wave function description, which are really important in further considerations.

Namely, we have to introduce the notion of a ***wavefront***. The name is a bit misleading, because a "front" may suggest that it's something separating an area where there is no wave, from an area where "there already is a wave". In other words, a front of an "approaching" wave – as in the case of a a tsunami approaching a coast.

The term “wavefront” was coined by researchers who studied short wave pulses, not continuous wave trains (one of them was the famous Huygens). Then, it made sense to talk about a “front”. However, today’s physics uses the same word in a different sense. A wavefront is not an actual front, it’s not “where the wave begins” an actual front. Today’s definition identifies a wavefront as an abstract notion, as something which in mathematics is termed as *locus*. The Dictionar.com defines the locus (plural:*loci*) in mathematics as:*the set of all points, lines, or surfaces that satisfy a given requirement*.

Sometimes one can come across a definition identifying a wavefront as a locus of points with the same *displacement*. OK, in some simple cases it may be a good definition – consider the aforementioned case of a circular wave on lake water: if you connected adjacent points with the same displacements, you would get a circle. Yet, this method will fail if, say, there is an oil slick near the point where the rock hits water. Oil, which is a known thing (“pour oil on troubled waters”), has an attenuating effect on water waves – i.e., it makes the displacement much smaller. So, it may happen, that you try to track the wavefront – and when you pass the borderline between the clear water, and the oil slick – even though the wave is still propagating through the oil-covered water.

There would be no such problem if the definition of the wavefront were based on the ***wave’s phase*** – i.e., the argument of the oscillating sine in the wave function. So, we will use the definition given by this Wikipedia article:*a wavefront is the locus of points characterized by propagation of positions of identical phase: propagation of a point in 1D, a curve in 2D or a surface in 3D.* A slightly different definition, perhaps a bit more “intuitive”, can be found in Encyclopedia Britannica: *Wave front, imaginary surface representing corresponding points of a wave that vibrate in unison. When identical waves having a common origin travel through a homogeneous medium, the corresponding crests and troughs at any instant are in phase; i.e., they have completed identical fractions of their cyclic motion, and any surface drawn through all the points of the same phase will constitute a wave front.*

Some important examples of wavefronts in optics are the following. The one of a spherical wave, i.e., a wave spreading uniformly from a *point source*[1] The wavefront of such a wave is a spherical surface. Another important

[1] There are no real point sources – but a wave emerging from a small source, viewed at from a distance much larger than the source size, can be considered with a good approximation as a spherical wave.

example is a *plane wave* defined by Wikipedia as a wave whose wavefronts (surfaces of constant phase) are infinite parallel planes. Defined in such a way, a plane wave is rather a mathematical idealization because nothing in real life can be infinite. If we "relax a bit" the infinity condition and let the planes be small, then a laser beam can be thought of as a "narrow plane wave". Also, a spherical light wave can be converted to a "nearly plane wave" using a lens and an experimental trick called "collimation" – again, the planes won't be infinite, but only as big as the lens is. However, in many applications laser beams or collimated beams can represent plane waves with a very good approximation.

3.2 Light Rays

Once we have defined the wavefront, we can make the next step and define the light ***rays***.

The notion of a light ray is probably as old as human civilization. How people understood light rays at very old times, we do not know. But in more recent era, they probably though of rays as of some "building blocs" of which light was made. And always, rays were thought of as *existing objects*.

In modern era, rays were demoted. We still talk about rays, but not as of existing objects, but *abstract objects*. They do not have a real existence, they exist only in our minds. Is it good, or bad? Well, many abstract things are good or very good, or very valuable. Freedom is an abstract thing. Is it worthless because it is only an abstract idea?

The light ray is an abstract concepts of a completely different category, but it's also a very valuable concept. It offers a "tool" that makes it possible to solve a whole variety of optical problems in a relatively easy way.

So, how do we define light rays? The definition is really very simple. Light rays are lines *normal to light wavefronts*. In other words, they are directions along which light waves propagate.

The art of solving optical problems based on the concept of rays is called *Ray Optics*, or, more "officially", *Geometrical Optics*. The "tool set" in ray optics is pretty small, it consists of two major rules: the Law of Reflection, and the Law of Refraction – and a few lesser rules. And, in addition, the Huygens Principle (a.k.a. the Huygens-Ffrfesnel Principle). But despite such "compact tool set", geometrical optics constitutes an extensive area of science. As mentioned, there is enough material in it for constructing a one-term or even a two-term academic course devoted exclusively to geometrical optics. However, in the present Ph332 Course we can use no more than three class hours for it. So, the scope of material discussed in the present lecture notes will not be be very wide. Therefore, it's certainly a good idea to give here references to a few other sources, which can be though of as "auxiliary texts". It is always a good thing to read about the course material – in any course! – from more than a single source. Below, there are clickable links to two other sources worth recommending:

- A long (44 pages) document titled *Basics Geometrical Optics* by Leno S. Pedrotti, CORD, Waco, Texas. Essentially, the material interesting in context of our course is on pages 73-86, the other materia go es beyond the scope of Ph332.

- A treatise Geometric Optics by by Dr. J. B. Tatum, UVIC, Ca. The material of interest is in Chapterb 1, Sections 1.1 - 1.5, Chapter 2, Sections 2.1 -29, and in Chapter 3, all secttions.

In a Power Point linked to, it is shown how light waves can be represented by "boundles of rays". The first animated slide shows a wave emerging from a point source (or a source of very small size). An equivalent situation shown in Slide 2 is a bundle of rays, each one symbolized by a straight line spreading out from the source. A "bare" line doesn't tell you what direction the wave propagates, and therefore it is a right thing to add "arrow signs" at the lines.

It there were an observer at the right side of Slide 2, looking toward the source – what would she see? A single glowing point. But if there is a bundle of rays emerging from a single point, this point is not necessarily a "real" light source. Suppose that there is a bundle of rays converging at one point and moving further, forming a bundle of *diverging* rays afterwards. The bundle would be then identical with that produced by a "real" point source. Therefore, the observer will see a point where the rays converge, even though there is no real point source at this spot! This is what we call a ***virtual image*** of a light point. We will be talking a lot more about virtual images later in this chapter, and in the following chapters.

In Slide 3, the animated wave is converging in a single point – and in Slide 4 the same situation is represented by a bundle of converging rays. If we put a screen at the point where all the rays intersect, such a bundle would produce a glowing point at this screen. In other words, not a real point source of light, but an ***image*** of such a source. Such image is called a ***real image***. So, virtual images are those that can be observed directly by eye – and real images are those that can be projected on a screen. There will be a lot more about both types of images on the pages to come.

3.3 The Reflection of Light from Mirrors and Smooth Surfaces

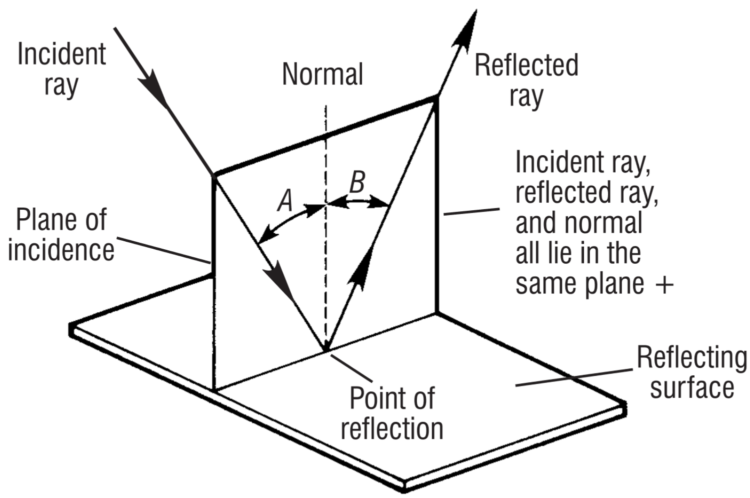

Figure 3.1: The reflection of light in the "ray representation" (from the text of Pedrotti linked above).

The Fig. 3.1 illustrates the ***Laws of Light Reflection***:

- The angle between the incident ray and the normal to the reflecting surface emerging from the point where the ray intersect the surface (the ***angle of incidence***), and the angle between the normal and the reflecting ray (the ***angle of reflection***) are equal; and
- The incident ray, the reflected ray and the normal all lie in the same plane (called the ***plane of incidence***).

3.4 Light Refraction at the Interface of Two Media

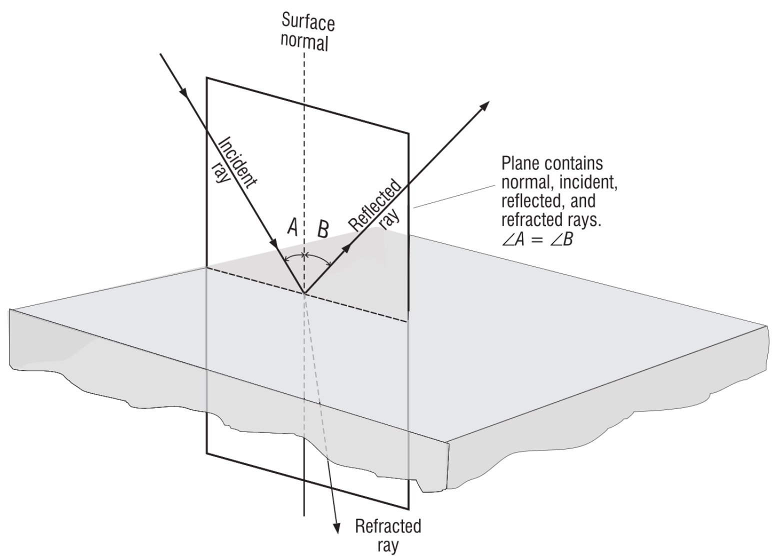

Figure 3.2: From Pedrotti's text: The reflection **and refraction** of light at a flat interface between two media.

The incident ray comes from one medium, and the refracted ray enters the other medium. The angle between the refracted ray annd the normal is called the *angle of refraction*, θ_{r}. There is a relation between the angle of incidence (θ_{i}) and the angle of refraction, known as the ***Snell's Law***. It can be derived in a relatively straightforward way, and it is instructive to do it – but one more component is needed, namely, the so-called **Huygens Principle**.

3.4.1 The Huygens Principle

A term, not used by us before, but used in the formulation of the Huygens Principle (HP) is a ***wavelet***. So, it needs an explanation. A wavelet is a tiny elementary circular or spherical wave source. Huygens reasoned that wavelets are elementary "building blocks" of all types of waves. How those building blocks act, is explained by the principle:

After LUMEN - boundless physics: *The Huygens-Fresnel principle states that every point on a wavefront is a source of wavelets. These wavelets spread out in the forward direction, at the same speed as the source wave. The new wavefront is a line tangent to all of the wavelets.*

I (Dr. Tom) am looking for a definition formulated in "more colloquial" terms – and I think a good word to use would be "reincarnation". Namely, how Huygens explained the mechanism of wave propagation: a circular or a spherical wavefront reaches certain position, then it disappears and changes into a multitude of "wavelets" – which combine their tiny "waveletfronts" (there is no such word, it's my "ad hoc creation") together, thus "reincarnating" their "maternal wavefront", in a new position slightly ahead of where the "maternal" was. And the same proces repeats for the new wavefront, and so on, an so on.

Huygens lived in the "pre-calculus era", he did not know the mathematical theory of wave motion, which emerged some time later. According to today's orthodox mathematical theory, the description of a wave motion at the fundamental level is a differential equation containing second-order derivatives with respect to time and to space variables. So, one can think that the 300+ years old Huygens non-mathematical principle can be sent to a retirement home – but no, it turns out that the Huygens Principle is ***not inconsistent*** with the sophisticated all-math theory – and it's great advantage is that it allows one to understand many phenomena of wave motion, using an approach which is much simpler than the orthodox mathematical one, and an approach that is definitely more intuitive[2]

Please click on the LUMEN - boundless physics link, there is an instructive picture showing how the HP explains the propagation of an ordinary plane wave. But we are mostly interested in explaining the refraction at an interface between two media. We will use two pictures:

[2]Here is the wave equation of something quite simple – namely, a sound wave propagating in one dimension:

$$\frac{\partial^2 p}{\partial t^2} = v^2 \frac{\partial^2 p}{\partial x^2}$$

where p is the acoustic pressure, v is the speed of sound, and "∂" is the symbol of a *partial* derivative. Many people agree that they much more prefer considerations based on the Huygens Principle.

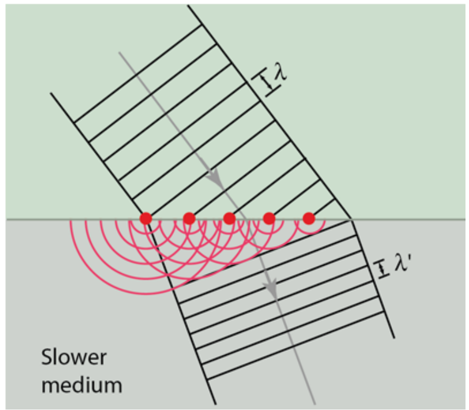

Figure 3.3: This figure is better for conceptual understanding of the refraction – the plane wave in the "new" medium is constructed by combining wavelets. But those in the "new" medium have a shorter wavelength than in the medium the incident wave comes from.

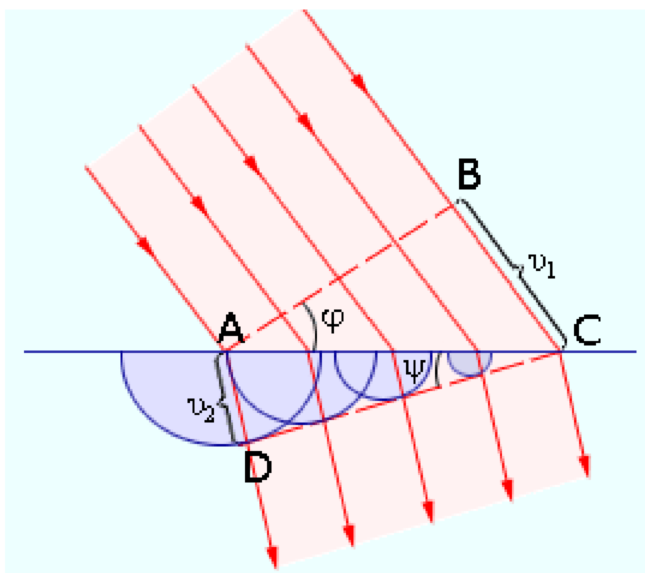

Figure 3.4: Essentially, the same picture as Fig. 3.3, but a more convenient one for deriving the Snell's Law, because the symbols we will use are already in the plot.

To derive the Snell's Law, consider a sector AB of the impinging wave which propagates in Medium 1 with speed v_1. When point A of the wavefront hits the surface, point B is still away – it will take certain time period, call it Δt, to reach the surface at point C. The distance AC is then AC$= v_1 \cdot \Delta t$.

The wavefront AB, when it reaches point A at the surface, creates a wavelet that starts spreading away with a speed of v_2. When point B of the "maternal" wavefront reaches the surface, the "waveletfront" of the said wavelet has already traveled a distance of $v_2 \cdot \Delta t$.

Now, if we call the angle between the surface and the original wavefront AB as ψ, and between the surface new "Huygens" wavefront in Medium 2 as φ.

Using simple trigonometry, one can write:

$$\mathrm{AC} = \frac{\mathrm{AC}}{\sin\psi} = \frac{v_1 \cdot \Delta t}{\sin\psi}$$

And, taking into consideration that the "new" wavefront" is tangent to wavelets in Medium 2, one can also write:

$$\mathrm{AC} = \frac{\mathrm{CD}}{\sin\varphi} = \frac{v_2 \cdot \Delta t}{\sin\varphi}$$

From which it follows that:

$$\frac{v_1 \cdot \Delta t}{\sin\psi} = \frac{v_2 \cdot \Delta t}{\sin\varphi}$$

Δt on both sides cancels out, and after some very simple algebra we get:

$$\frac{\sin\psi}{\sin\varphi} = \frac{v_1}{v_2}$$

It's almost the Snall's Law, but in the law formulation we don't use the angles between the surface and the vavefronts, ψ and φ. Instead, we use the angle between the incident ray and normal to the interface, we call it the "incident angle" θ_i. Similarly, for the refracted light, we use the angle between refracted rays and normal to the interface, and we call it θ_r.

Now, notice that the rays are perpendicular to their corresponding wavefronts, and the "normal" is perpendicular to the interface. And from high-school geometry (or from another geometry class you have taken?) recall the theorem of angles with arms perpendicular:

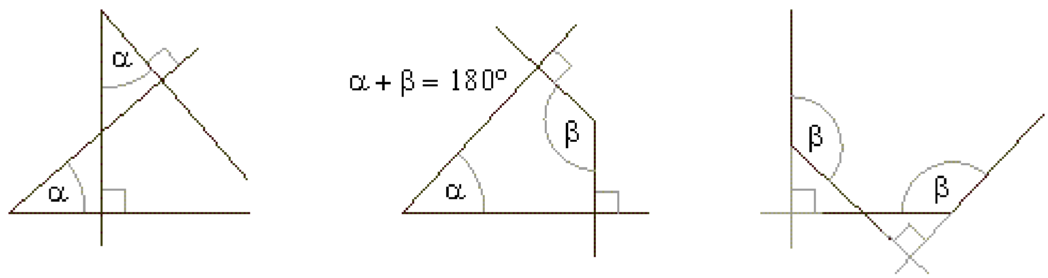

Figure 3.5: Angles with arm perpendicular are either equal, or their sum is 180°.

The angles ψ and $\theta_{\rm i}$ are both $< 90°$, so they cannot add up to 180° – there fore it must be that $\psi = \theta_{\rm i}$. For the same reason, it must be $\varphi = \theta_{\rm r}$. Therefore, we can write the result we have obtained as:

$$\frac{\sin\theta_{\rm i}}{\sin\theta_{\rm i}} = \frac{v_1}{v_2} \tag{3.1}$$

Or, in order to satisfy the famous Occam Razor, which in it's original 14th Century Latin version states: *Entia no sunt multiplicand praeter necessitatem.* Which means that *in any philosophical theory new entities should not be introduced, unless it is absolutely necessary.* And in physics, it can be reduced to: *Don't use to many symbols!*. In fact, in Eq. 3.1 we have for Medium 1, $\theta_{\rm i}$ and v_1, why not $\theta_{\rm i}$ and $v_{\rm i}$? And the same for Medium 2. Then, the Eq. 3.1 becomes:

$$\frac{\sin\theta_{\rm i}}{\sin\theta_{\rm i}} = \frac{v_{\rm i}}{v_{\rm r}} \tag{3.2}$$

This is a formulation of the Snell's Law you may find in many texts. Or, there is another one also satisfying the Occam Razor:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} \tag{3.3}$$

Now, notice that the wave propagation velocities are completely independent on the angles in the Snell's Law. Each velocity is a property of a given material. So, one can say that the v_1/v_2 ratio is a property of a given *pair of materials*. For a given pair of materials this ratio is constant, we can call it n_{1-2}. The official name of this constant is the *refractive index* – or, if you prefer, the *index of refraction*.

The refractive index values for various material can be found in numerous sources. But single materials are listed, not pairs. Why? Because, by convention, the values given are always for a "pair": vacuum - the material. So, if the table lists, for example: "for water, H_2O, $n = 1.33$, it means that $v_{\text{vacuum}}/v_{\text{water}} = 1.33$. And v_{vacuum} is the famous constant c, so the meaning is: $c/v_{\text{water}} = 1.33$.

Now, suppose that you are interested in refraction of light passing from water into diamond. You already know that $c/v_{\text{water}} = 1.33$. The tables give the value of the refractiv index for diamond as 2.417. Meaning that $c/v_{\text{diamond}} = 2.417$. So, you can write: $v_{\text{water}} = c/1.33$, and $v_{\text{diamond}} = c/2.417$. For the pair water-diamond, you need:

$$n_{\text{water}-\text{diamond}} = v_{\text{water}}/v_{\text{diamond}} = \frac{c/1.33}{c/2.417} = \frac{2.417}{1.33} = 1.817.$$

We will return to the refraction "with vengeance", when talking about the *total internal reflection*.

3.5 The Formation of Images in a Mirror

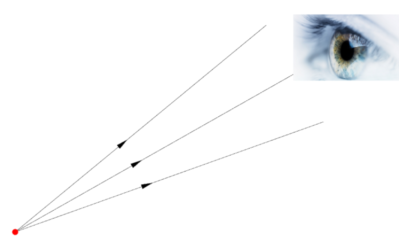

Figure 3.6: If a bundle of divergent rays, all emerging from a point source of light, reach our eyes, our visual impression – or, in plain language, how we see it – is a glowing point. Nothing surprising – after all, our senses are arranged in such a way that the image surrounding us is reproduced in our mind with maximum fidelity.

But our senses can readily be deceived. One simple way is to put a mirror at the path of the divergent ray beam. Due to the laws of reflection – in particular, the one stating that the angle of incidence, and the angle of reflection (relative to the normal to the mirror plane) are equal – as illustrated in the Fig. 3.7, the extensions of the reflected rays all intersect in at a point located "at the other side of the mirror", symmetrically to the real point light source with respect to mirror's plane. In other words, the beam of reflected rays is identical to that which would be emitted by a real source placed at the position of the image.

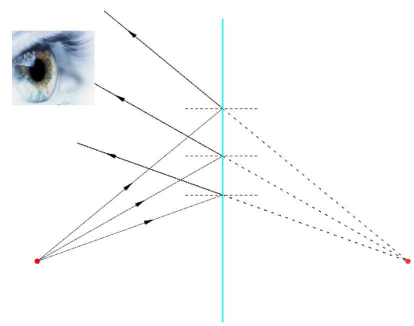

Figure 3.7:

But what about larger objects? The images we watch in mirrors in most cases are not of point-like sources, but of larger objects that *do not* emit their own light! Sure, but it can be readily explained. First, consider the screen of a TV, of a comp, uter, or of your smart phone. Using a magnifying glass, it can be easily checked that the images seen at such screens are not "continuous", but consist of *pixels* – each of which is not a *point source*, but a source small enough that from a distance we don't see it's in fact a miniature glowing square.

And larger object we look at can also be though of as consisting of pixels – in fact. each individual molecule at the object's surface acts as an individual "pixel". But, one moment, please! – you may say – it doesn't look like a good explanation, because they don't emit light, like real pixels in a TV screen! It's certainly true, they do not emit their "own" light. Most things we see do not – the word "emission" means sending out light generated by the emitting object: a candle, a LED, a light bulb. The Sun, or a star. But not Moon! Moon, as everybody know, does not make its "own" light. What we see is reflected sunlight. Right? Almost right, but not exactly – what we see is not "reflected light", but ***scattered light***. Another term for light scattering is ***diffuse reflection*** of light. As opposed to reflection from a mirror or a polished surface (often referred to as *specular reflection*, from Latin *speculum*, meaning "a mirror"), molecules at non-polished objects constitute an uneven surface. Light shone on such surfaces is scattered in all directions. Each molecule can be though of as a light source – not generating light, but producing a divergent beam of rays from light incident on them from another source (e.g., daylight, or artificial light).

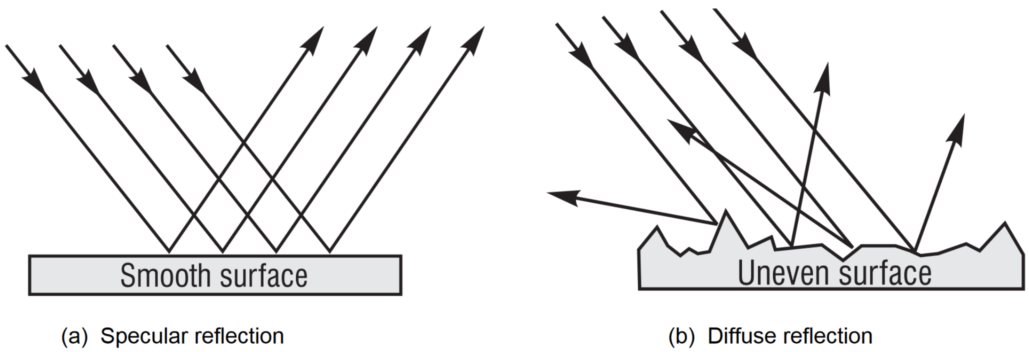

Figure 3.8: Specular reflection vs. diffuse reflection of light (from L.S. Pedrotti's text linked at Page 5).

3.6 The Formation of Images in Refracting Media

3.6.1 The Principle of Reversibility

The principle of reversibility states that if the direction of a ray of light is reversed, it will follow the same path in the opposite direction that the path it followed originally, including any changes in direction resulting from refraction or reflection. An example of the light reversibility in the case of a single refraction is shown in the Fig. 3.9.

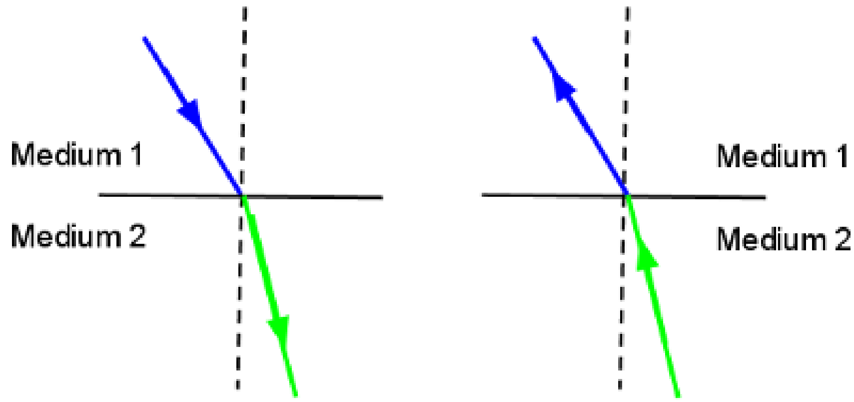

Figure 3.9: The reversibility of light passing from one medium to another (copied from https://www.meritnation.com/).

Here is a link to a nice YouTube clip explaining the principle of reversibility.

So, let's now consider a ray entering water from the air at an incident angle of θ_1. Atmospheric air has the refractive index slightly larger than 1 – namely, 1.00028. The deviation from 1, 0.00028, is so tiny that with a very good approximation we can round it down to 1. What we usually do.

The refractive index vacuum-water, or air-water, is approximately $n_{\mathrm{v}\to\mathrm{w}} = 1.33$. So, if light is incident on water surface at an angle of θ_1 (see the Fig. 3.10, left plot), we find the angle of refraction θ_2 from the Snell's law:

$$\frac{\sin\theta_1}{\sin\theta_2} = n_{\mathrm{v}\to\mathrm{w}} \quad\Longrightarrow\quad \sin\theta_2 = \frac{\sin\theta_1}{n_{\mathrm{v}\to\mathrm{w}}} \tag{3.4}$$

For instance, if $\theta_1 = 45°$, as in the Fig. 3.10, we get:

$$\sin\theta_2 = 0.70711/1.33 = 0.53166 \quad\Longrightarrow\quad \theta_2 = 32.12°$$

Now, consider a "reversed situation", shown in the right side of the Fig. 3.10. Now, the incident ray comes along the path of the refracted ray in the previous case. So, the incident angle is now θ_2. The principle of reversibility states then that the ray emerging from water follows the path of the incident ray in the previous case, so it makes an angle θ_1 with the normal. If we use the Snell's Law again, we can write:

$$\frac{\sin\theta_2}{\sin\theta_1} = n_{\mathrm{w}\to\mathrm{v}} \tag{3.5}$$

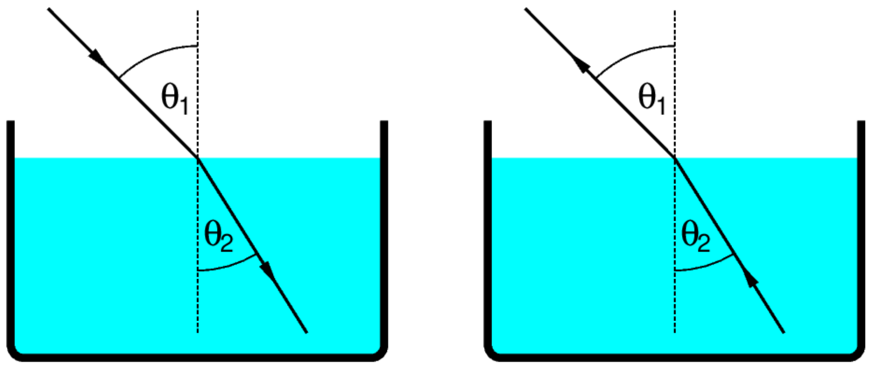

Figure 3.10: In the left plot, a ray is incident on water surface and is refracted. In the right plot, a ray is coming from underwater along the same path as the refracted ray in the left plot, but in opposite direction. So, from the Principle of Reversibility it follow that after emerging from water, the ray should follow the path of the incident ray in the left plot.

Considering that $\sin\theta_1/\sin\theta_2 = n_{\mathrm{v}\to\mathrm{w}}$, we can reach the conclusion that:

$$n_{\mathrm{w}\to\mathrm{v}} = \frac{1}{n_{\mathrm{v}\to\mathrm{w}}} \tag{3.6}$$

So, when a ray is incident from under water's surface, the refracting index to be put into the Snell's Law should be:

$$n_{\mathrm{w}\to\mathrm{v}} = 1/1.33 = 0.75$$

3.6.2 The Total Internal Reflection of Light

Let's consider light coming out of water, like in the right plot of the Fig. 3.10. Now the incident angle is θ_2, and the angle of refraction is θ_1. We already know that in a transition from water to air (vacuum) the refractive index is $n_{\mathrm{w}\to\mathrm{v}} = 0.75$. Then, according to the Snell's Law, we can "prescribe a recipe" (or, using more professional terms, set up an *algorithm*) for calculating the angle $\sin\theta_1$:

$$\frac{\sin\theta_2}{\sin\theta_1} = 0.75 \quad \Rightarrow \quad \sin\theta_1 = \frac{\sin\theta_2}{0.75} \tag{3.7}$$

OK, we can write even a more elegant algorithm, not for calculating the sine function, but directly the angle:

$$\theta_1 = \arcsin\left(\frac{\sin\theta_2}{0.75}\right) \tag{3.8}$$

(The **arcsin** function is the inverse of the **sine** function. It returns the angle whose sine is a given number; on calculators, the symbol is $\sin^{-1}$, but Dr. Tom prefers the traditional notation of arcsin).

Well, so consider a process in which the angle θ_2 starts from low values and gradually increases. When it reaches the value of 48.59°, where the sine is exactly 0.75, the sine of θ_1 becomes 1, meaning that $\theta_1 = 90°$ – see the ray plotted with a black line in the Fig. 3.11. In other words, the refracted ray "slides" on the water surface.

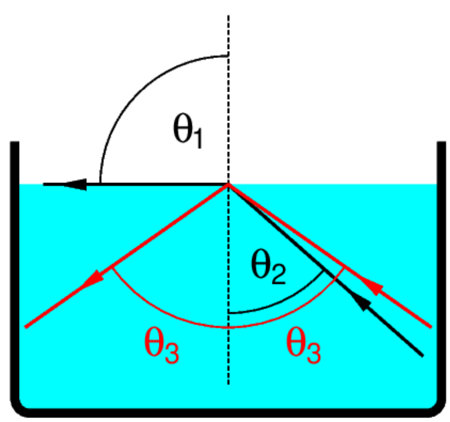

Figure 3.11: For $\theta_2 = 48.59°$, If a ray is incident at water surface "from below", and the angle of incidence $\theta_2 = 48.59°$, the angle of refraction is $\theta_1 = 90°$ (the black line). For larger angles θ_2 there are no angles θ_1 satisfying the Snell's Law – the incident ray gets totally reflected from the interface, as shown for the ray with an incident angle $\theta_3 = 55°$, plotted with a red line.

But what would happen if we keep increasing θ_2? Then the expression $\sin\theta_2/0.75$ in the Eq. 3.7 becomes larger than 1! The sine function never takes values larger than 1, so there is no such a value of θ_1 that would satisfy the Eq. 3.7. There is now way that a refracted ray can exist!

Yes, indeed, it's exactly what happens. The incident beam cannot be refracted, and enter the area above the water. But it has ***to go somewhere***. And it does – it gets "back-reflected" from the interface, satisfying the law of reflection, as shown by the red line in the Fig. 3.7.

This phenomenon is known as ***the total internal reflection***. The word "internal" comes from the fact that in most cases we deal with such reflection, it happens at the interface of a material and air (or material and vacuum), with the ray "attempting" to escape from the "inside" to the "outside".

Total internal reflection occurs when the angle of incidence is greater than a certain limiting angle, commonly referred to as the ***critical angle***. It's the incident angle for which the refracted angle is 90°. For the water-air interface, the critical angle is, as we have seen, 48.59°.

For any other material, the calculation of the material/vacuum critical angle is straightforward. For the ray entering the material (as in the left plot in the Fig. 3.10), the Snell's Law says:

$$\frac{\sin\theta_1}{\sin\theta_2} = n_{\mathrm{vac}\to\mathrm{mat}}$$

where the value of $n_{\mathrm{vac}\to\mathrm{mat}}$ is to be found from the Refractive Index Tables (given, e.g., by Wikipedia).

So, for the incident ray coming ***from the material*** (as in the right plot in the Fig. 3.10), the same Snell's Law states:

$$\frac{\sin\theta_2}{\sin\theta_1} = n_{\mathrm{mat}\to\mathrm{vac}} = \frac{1}{n_{\mathrm{vac}\to\mathrm{mat}}}$$

The θ_2 angle becomes the critical angle when $\theta_1 = 90°$, or $\sin\theta_1 = 1$, so that:

$$\sin\theta_{2\ \mathrm{crit.}} = \frac{1}{n_{\mathrm{vac}\to\mathrm{mat}}} \tag{3.9}$$

Or:

$$\theta_{2\ \mathrm{crit.}} = \arcsin\left(\frac{1}{n_{\mathrm{vac}\to\mathrm{mat}}}\right) \tag{3.10}$$

Example:

From the tables, we find that for diamond the refractive index is 2.417. We know that the tables list the $n_{\text{vac}\to\text{mat}}$ values. Therefore, for diamond $\theta_{\text{crit.}} = \arcsin(1/2.417) = 24.44°$.

Critical Angle for an Interface of Two Materials

Not always light passes from material's interior to the air (or vacuum), it may also undergo total internal reflection at an interface of two materials. For instance, an interface between diamond and water.

The Snell's Law, formulated in therms of propagation velocities, states that:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_{\text{mat.}\,1}}{v_{\text{mat.}\,2}} \tag{3.11}$$

We also know that the value $n_{\text{mat.}}$ of refractive index for a given material listed in a table is:

$$n_{\text{mat.}} = \frac{c}{v_{\text{mat.}}} \tag{3.12}$$

From which we may write that:

$$v_{\text{mat.}\,1} = c/n_{\text{mat.}\,1} \quad \text{and} \quad v_{\text{mat.}\,2} = c/n_{\text{mat.}\,2}$$

Substituting the above to the Eq. 3.11, we get:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{c/n_{\text{mat.}\,1}}{c/n_{\text{mat.}\,2}} = \frac{n_{\text{mat.}\,2}}{n_{\text{mat.}\,1}} \tag{3.13}$$

So, in order to obtain $\theta_1 = 90°$ and $\sin\theta_1 = 1$, (as in the Fig. 3.10), it must be:

$$\sin\theta_{2\,\text{crit.}} = \frac{n_{\text{mat.}\,1}}{n_{\text{mat.}\,2}} \tag{3.14}$$

So, if, for example, material 1 is water, and material 2 is diamond, then the sine of the critical angle in diamond is $\sin\theta_{2\,\text{crit.}} = 1.33/2.417 = 0.55$, and $\theta_{2\,\text{crit.}} = 33.38°$.

Important: Note that the total internal reflection can occur only in the medium with higher refractive index. If in the Eq. 3.14 $n_{\text{mat.}\,1}$ were larger than $n_{\text{mat.}\,2}$, then the equation would lose its sens.

Often in the literature one can find expressions: “optically dense medium” or “optically less dense medium”. The term “opotical density” refers simply to the refractive indices. If medium A has a refractive index of n_{A}, medium B has a refractive index of n_{B}, and $n_{\mathrm{A}} > n_{\mathrm{B}}$, then we say that “medium A has a higher optical density than medium B”.

In YouTube, there is a number of video clips explaining total internal reflection. I would recommend watching two of them: this one, and the other one. There is little verbal explanation in both, but the experiments shown “speak for themselves”, and words are really not necessary.

3.7 Underwater Objects Seen from Above the Water Surface

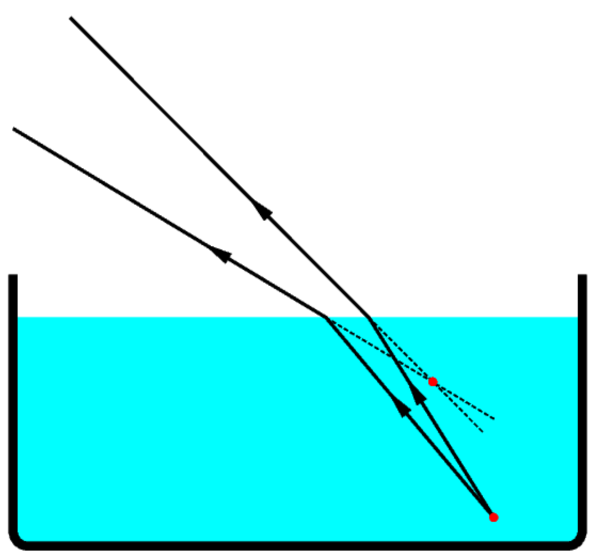

Figure 3.12: A light source under water, and the mechanism causing that the apparent position of the source is closer the surface than the real position.

The Snell’s Law offers an easy explanation for the “broken stick effect”. In the Fig. 3.11 there is a small light source in a pool. There is enough to plot just two diverging rays emitted by the source. When passing through the surface, they are refracted "more to the left", and their angular divergence increases. All angles in the Fig. 3.11 are calculated according to the Snell's Law, for $n = 1.33$. The two rays emerging from the light source make angles of 32° and 40° with the normal to the water surface. The refracted rays make, respectively, angles of 44.8° and 58.75° with the normal. Hence, the initial divergence of 8° between the two rays increases to 14.95°. Our eyes are deceived, they "locate" the glowing point where the *extensions* of the two rays emerging from water intersect – which is closer to the surface than the real light source location.

3.8 World Seen by a Fish Looking Up

In photography, there is something called a "fish-eye picture". One can take it using a special "fish-eye lens" with an angle of vision 180°. An example of such a picture, taken in a city with the camera looking straight-up, is shown below:

Figure 3.13: An example of a fish-eye photo.

Why it's called a "fish-eye vision", is explained by the pictures posted below the text in the Ph332 Course Web Page. In one of them, a fish is looking up, and the rays plotted in the figure show how the 180° angle of vision is "compressed" by the refraction to a much narrower angle.

A fish sees not only the compressed view of the "whole world above the water surface", but around it the fish sees a mirror image of the lake bottom – it is shown in the larger picture with the palms, probably taken from underwater in a shallow tropical laguna, using a standard lens: the compressed 180° vision is within the central circle, and what's outside the circle is a mirror image of the sandy bottom. Only the part of the bottom closest to the camera is not seen in the picture.