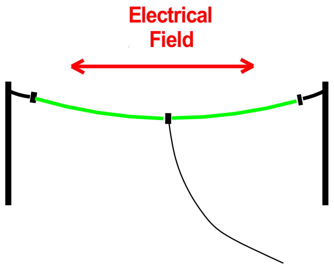
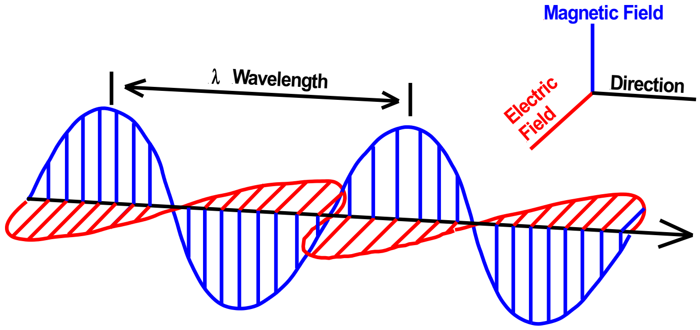
Chapter 14

Radio Wave Propagation





14 Radio Wave Propagation

Long distance over the horizon radio communications at frequencies from 2 to 30 MHz is made possible by Earth's ionosphere. Radio wave propagation through the ionosphere from one radio station to another (Figure 1) is the subject of this chapter.

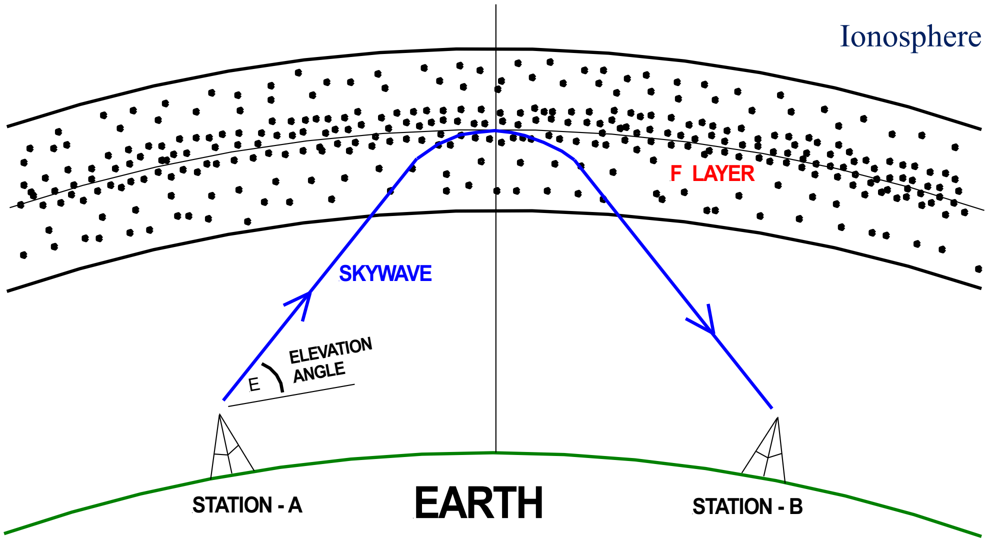


Figure 1 Skywave radio communications (source: author)

14.1 Radio Waves

A radio wave is an electromagnetic wave consisting of two orthogonal fields, a sinusoidal electric field and a perpendicular sinusoidal magnetic field. The direction of wave travel is perpendicular to both the electric and magnetic fields as illustrated in Figure 2. The two fields are in phase, that is, when the electric field reaches a peak, the magnetic field does as well. They are also inseparable. The oscillating sinusoidal electric field induces the oscillating sinusoidal magnetic field. In turn, the oscillating magnetic field induces the oscillating electric field. In an electromagnetic wave one field can not exist without the other.

The Electric Field is responsible for most of **the good**, ***and bad***, things that happen to radio waves. It is the electric field interacting with free electrons in the ionosphere that causes radio waves to bend back to Earth (Figure 3). It is the electric field interacting with Earth's surface that causes the radiation patterns of vertical and horizontal antennas to be different. The magnetic field mostly goes along for the ride. Yet, the electric and magnetic fields are inseparable

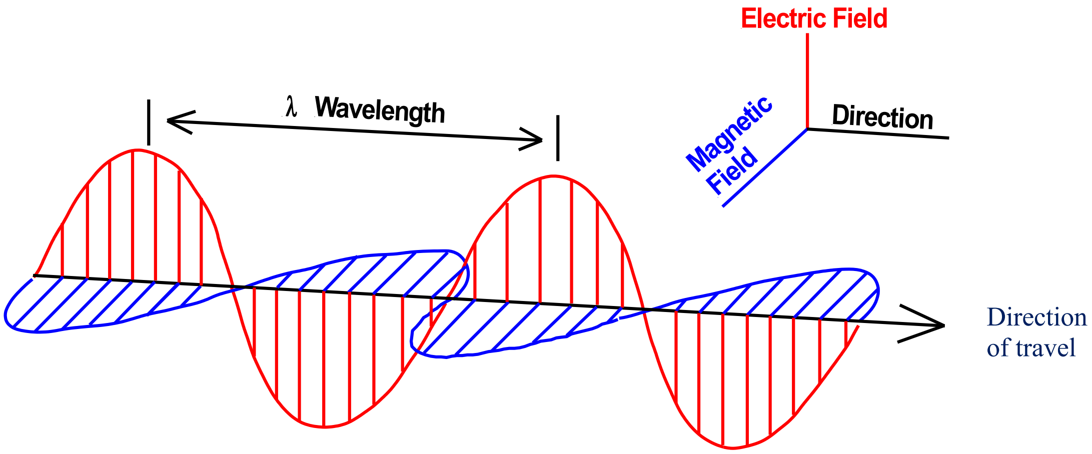


Figure 2 Electromagnetic Wave (source: author)

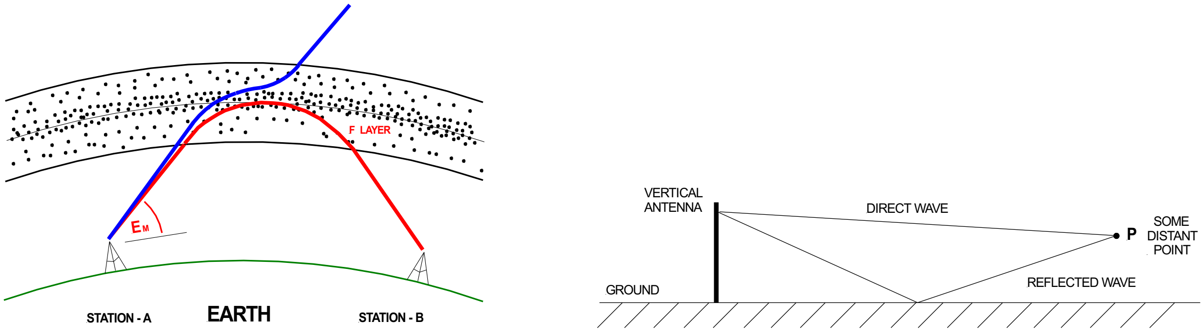


Figure 3 Radio wave interacting with the ionosphere and the ground (source: author)

The wavelength λ of an electromagnetic wave is the distance from the peak of the oscillating electric field to the next peak as illustrated in Figure 2. Since the electric and magnetic fields are locked together, the wavelength λ is also equal to the distance from the peak of the oscillating magnetic field to the next peak.

The frequency f of an electromagnetic wave is the number of electric field cycles per second, which is also equal to the number of magnetic field cycles per second.

Wavelength and frequency are related by the equation

$$\lambda = \frac{v}{f}$$

where

λ = wavelength
f = frequency, and
v = wave velocity

The wave equation for the electric field E is

$$\frac{\partial^2 E}{\partial x^2} = \varepsilon\,\mu\,\frac{\partial^2 E}{\partial t^2} = \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2}$$

where

x = wave amplitude

ε = permittivity of the medium through which the wave is traveling

μ = permeability of the medium

t = time

v = wave velocity

A similar perpendicular wave equation exists for the magnetic field.

In the above wave equation, the velocity term v is

$$v = \frac{1}{\sqrt{\varepsilon\,\mu}}$$

In the early 1800s Coulomb and Ampere both determined the values of ε_0 and μ_0 (the permittivity and permeability of free space) through careful static electric and magnetic measurements. Using their measurements Maxwell, in the 1860s, calculated the speed of an electromagnetic wave to be 284,000,000 meters per second, a value very close to the speed of light known at the time. This led Maxwell to conclude that light was an electromagnetic wave.

Using today's measurements for permittivity and permeability

ε_0 = permittivity of free space = $8.85418782 \cdot 10^{-12}\ \frac{(coulomb)^2}{Newton\ meter^2}$

μ_0 = permeability of free space = $4\pi 10^{-7}\,\frac{Newton}{(coulomb/second)^2}$

the speed of an electromagnetic wave in free space is

$$\varepsilon_0\mu_0 = 8.85418782 \cdot 10^{-12}\ x\ 4\pi 10^{-7}\ \frac{(coulomb)^2}{Newton\ meter^2}\frac{Newton}{(coulomb/second)^2}$$

$$\varepsilon_0\mu_0 = 1.11265\ x\ 10^{-17}\ \frac{1}{meter^2}\frac{second^2}{1}$$

leading to

$$v = \frac{1}{\sqrt{\varepsilon_0\mu_0}} = \frac{1}{3.33564\ x\ 10^{-9}} = 299{,}792{,}458\ \ meters/second$$

which is today's value for the speed of light.

Based on Maxwell's work we know today that radio waves and light waves are the same thing. The only difference between radio and light waves are their wavelengths as illustrated in Figure 4. Radio waves are very long while light waves are extremely short.

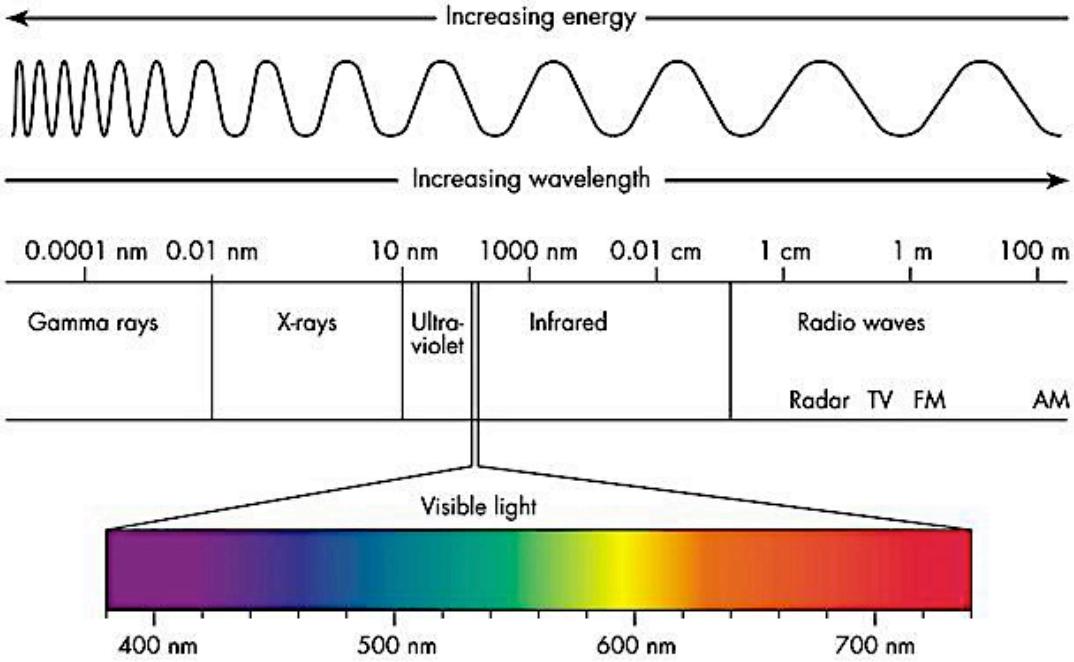


Figure 4 Electromagnetic Spectrum (source: Cyberphysics)

14.2 Reflection, Diffraction, and Refraction

Since light and radio waves are the same thing, all of the phenomena associated with light also occur with radio waves. Specifically, radio waves are

- Reflected,
- Diffracted, and
- Refracted

exactly the same as light waves.

Reflection, diffraction, and refraction are easily studied using light since we can “see” these phenomena occurring. We can not see the reflection, diffraction, and refraction of radio waves but we know that they do because radio waves and light waves are the same thing. However, we frequently “measure” the reflection, diffraction, and refraction of radio waves from and through various types of media.

14.2.1 Reflection of Light

Reflection occurs when a light wave encounters a smooth surface that does not absorb the light’s energy. For example, the surrounding scenery is clearly reflected in a lake provide the surface of the lake is perfectly smooth. However, the reflected light will be scattered in all directions producing a blurry image if the surface of the lake is rough, as illustrated in Figure 5.

Smooth lake surface Rough surface

Figure 5 Reflections from the surface of water (source: olympus-lifescience.com)

Around 60 AD Heron of Alexandria (~ 10 – 75 AD) showed by geometrical methods that light reflected from a flat polished mirror takes the shortest possible path from the object (the light source) to the mirror and finally to the observer. Using this result, he concluded that the angle of reflection from the mirror (Figure 6) must equal the angle of incidence on the grounds that this would yield the shortest path. In modern terms, the absorption and re-radiation of light by a material's electrons results in light reflecting from a smooth flat surface in all directions. However, most of the reflections cancel each other out through destructive interference resulting in the angle of the observed reflection being equal to the angle of incidence.

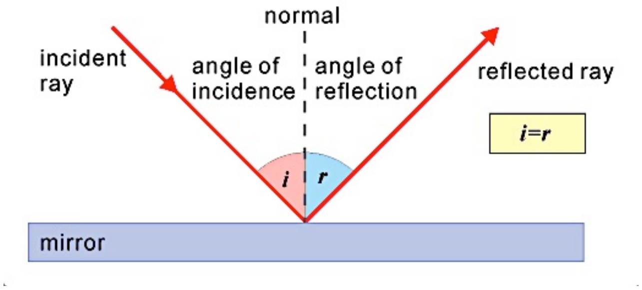


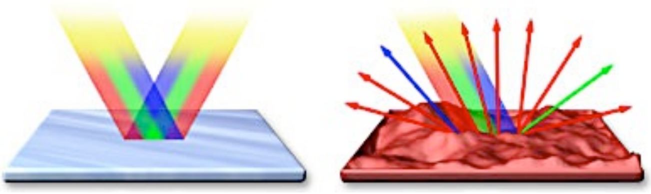
Figure 6 Angle of Incidence equals angle of Reflection (source: Quizlet.com)

Later Muslim mathematician, astronomer, and physician al-Haytham (~965 – 1040), also known as Alhazen, made the important observation that the angle of reflection is in the same plane as the angle of incidence.

As alluded to above, reflection of light can be categorized into two types:

- Specular reflection, and
- Diffuse reflection.

Specular reflection is defined as light reflected from a smooth surface such that the angle of incidence equals the angle of reflection. In contrast, diffuse reflection is produced by light scattered in all directions from a rough surface. There are far more occurrences of diffuse reflection in daily life than specular reflection. The difference between specular and diffuse reflection is illustrated in Figure 7.



Specular Reflection Diffuse Reflection

Figure 7 Specular and diffuse reflection (source: olympus-lifescience.com)

The mirror surface on the left side of Figure 7 reflects all components of white light equally, including red, green, and blue wavelengths. However, the rough reddish surface absorbs all wavelengths with the exception of red which it reflects in all directions from its rough surface.

It is largely through diffuse reflection that we see nonluminous objects around us (people, cars, houses, animals, trees, mountains, etc.). The color of an object is the color, or wavelength of light, that the object reflects. When standing in sun light a person's red shirt appears red because the shirt absorbs all wavelengths in the solar spectrum with the exception of red which it reflects.

There are two conditions required for specular reflection to occur:

1. A smooth reflecting surface is needed. That is, the average depth of surface irregularities must be substantially less than the wavelength of the incident light. Polished mirrors reflect very well. If the reflecting surface is rough, diffuse reflection will occur with light reflecting in more or less all directions.

2. The size of the reflecting surface must be much larger than the wavelength of the incident light. If this condition is not met, the incident light will again be scattered in all directions.

Reflections from a polished metal surface occur easily. The outer electrons in metal atoms are loosely held. Consequently, these electrons easily escape from their parent atoms creating a sea of free electrons throughout a metal object.

The sinusoidal electric field of a light wave (Figure 8) induces an electrical current in the free electrons at the metal's surface in the same way that a radio wave induces a current in a receiving antenna. The oscillating current in turn radiates a reflected light wave in the same way that an alternating current in a transmitting antenna radiates a radio wave. Most of the reflected light is cancelled out through destructive interference. However, some of the incident and reflected light interfere constructively forming a reflected light beam whose angle of reflection is the same as the angle of incidence. The light's sinusoidal magnetic field generally is not involved in this process.

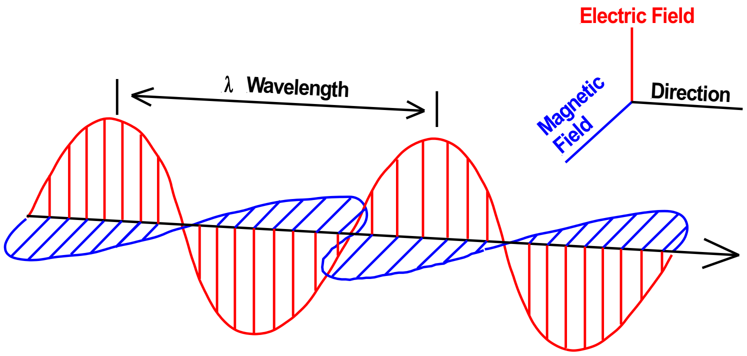


Figure 8 Sinusoidal orthogonal electric and magnetic fields of a light wave (source: author)

The conclusion reached by Heron of Alexandria and others that the angle of reflection equals the angle of incidence can be demonstrated analytically. In Figure 9 light is emitted by a source at Point A, it is reflected off a plane mirror, and arrives at Point B.

The perpendicular distance from the mirror to A and to B be the distance d with the shortest distance between the two points being D. Assume that light takes the path shown by the two red arrows. In this figure, the angle of incidence $\angle Ad$ is not necessarily equal to the angle of reflection $\angle dB$. The length of the path from A to the mirror to B is equal to

$$L = (x^2 + d^2)^{1/2} + ((D - x)^2 + d^2)^{1/2}$$

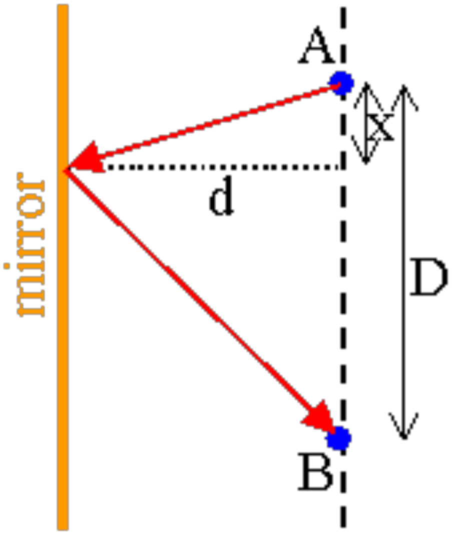


Figure 9 Reflection of light from a plane mirror (source: electron9.phys.utk.edu)

To find the shortest path from A to the mirror to B, we differentiate L with respect to x and set the result equal to zero.

$$\frac{dL}{dx} = \frac{x}{\sqrt{x^2 + d^2}} - \frac{(D - x)}{\sqrt{(D - x)^2 + d^2}} = 0$$

$$\frac{x^2}{x^2 + d^2} = \frac{D^2 + x^2 - 2Dx}{D^2 + x^2 - 2Dx + d^2}$$

$$x^2(D^2 + x^2 - 2Dx + d^2) = (D^2 + x^2 - 2Dx)(x^2 + d^2)$$

$$x^2D^2 + x^2x^2 - (2Dx)x^2 + d^2x^2 = D^2x^2 + x^2x^2 - (2Dx)x^2 + D^2d^2 + x^2d^2 - (2Dx)d^2$$

After canceling equal terms on both sides (the color coded terms) we are left with

$$0 = D^2d^2 - (2Dx)d^2$$

$$2Dd^2x = D^2d^2$$

$$2x = D$$

$$x = \frac{D}{2}$$

The shortest path from A to the mirror to B is the one for which x = D/2, or equivalently, the one for which i = r . Thus the angle of incidence i $= \angle Ad$ equals the angle of reflection r = $\angle dB$.

14.2.2 Diffraction of Light

Diffraction is a property of light that causes it to bend around sharp edges. In addition, diffraction causes light to spread out after passing through a small opening which is comparable in size to its wavelength.

Italian Jesuit priest, mathematician and physicist Francesco Maria Grimaldi (1618 – 1663) was the first to carefully study the diffraction of light. Since diffracted light seems to travel in different directions, Grimaldi named the phenomena diffraction meaning breaking up into pieces. Grimaldi showed that light spreads out and creates interference patters when it passed through a small hole. The results obtained by Grimaldi were published in 1665 after his death.

An intuitive understanding of diffraction is provided by Huygens' principles. Dutch physicist, mathematician and astronomer Christiaan Huygens (1629 – 1695) assumed that the wave front of a propagating light wave consisted of an infinite number of point sources. Light propagated from a point source as a spherical wavelet, illustrated in Figure 10. The radius of each wavelet was equal to the speed of light c multiplied by the time t since the wavelet was emitted by the point source. The infinite number of wavelets interfered with one another with the result that they all canceled out *except* in the forward direction. In the forward direction the wavelets added constructively causing the wave front to move forward (to the right in the figure) from its position at time t = 0.

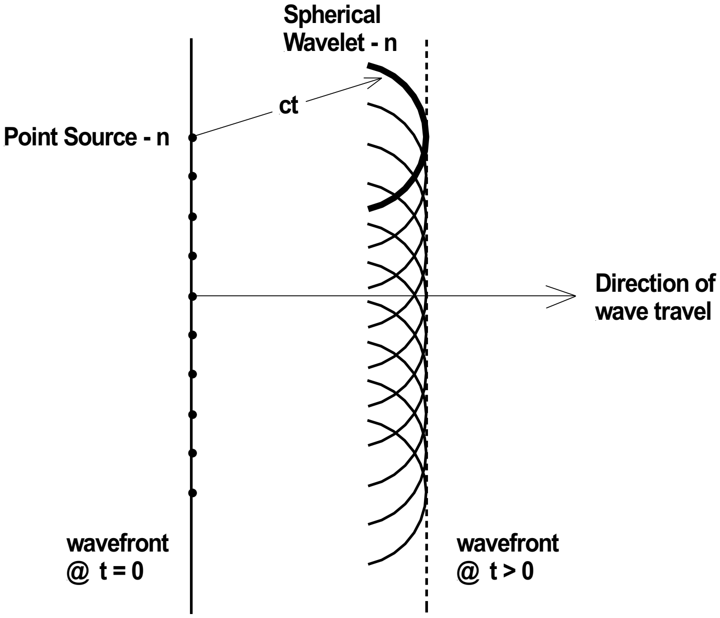


Figure 10 Huygens' Principle of Light (source: author)

In Figure 11 an incident light wave impinges on an opaque screen which contains a single pin hole a . The pin hole is one wavelength in diameter.

In accordance with Huygens principle, the wavefront approaching the screen is radiating wavelets of light from an infinite number of point sources along the wavefront. All but one of these wavelets is blocked by the screen. The one that is not blocked passes through the pin hole. This wavelet expands in a spherical manner, filling all space on the right side of the screen, since there are no other wavelets to interfere with it. The result is a complete diffraction pattern.

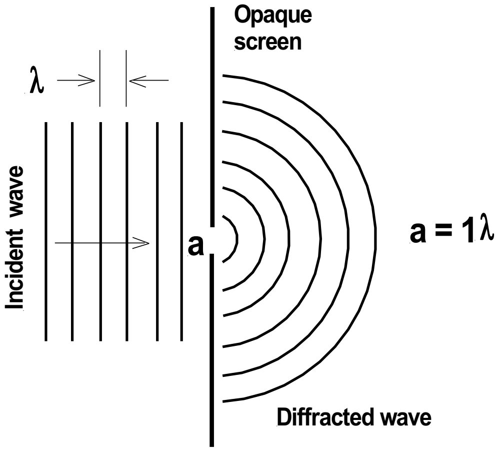


Figure 11 Diffraction through a pin hole 1λ in diameter (source: author)

In Figure 12 the pin hole is enlarged so that it is 3 wavelengths in diameter. The larger pin hole allows several wavelets to pass through. These wavelets interfere with each other on the right side of the screen forming a reduced refraction pattern.

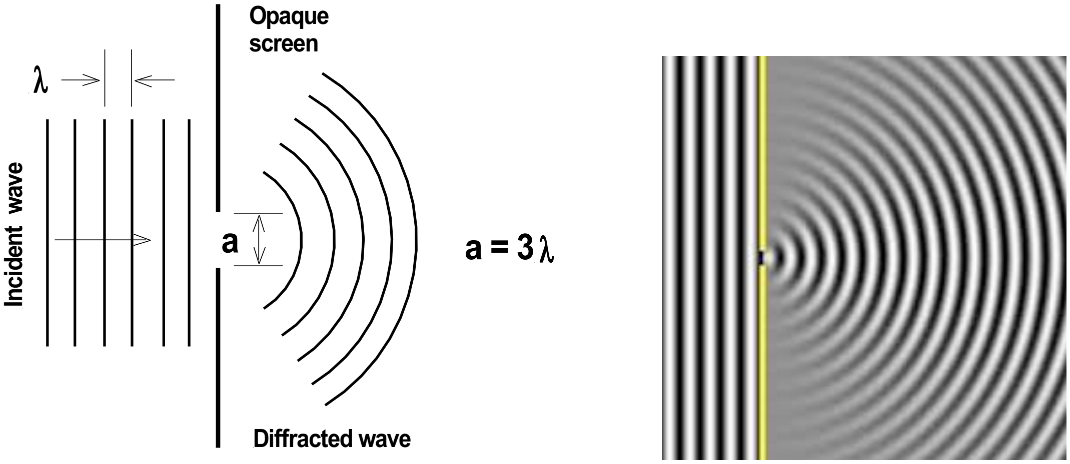


Figure 12 Diffraction through a pin hole 3λ in diameter (source: author)

In Figure 13 the pin hole is increased further in size to a diameter of 6 wavelengths, big enough to permit a large number of wavelets to pass through. All of these wavelets interfere with one another creating even a smaller refraction pattern on the right side of the screen.

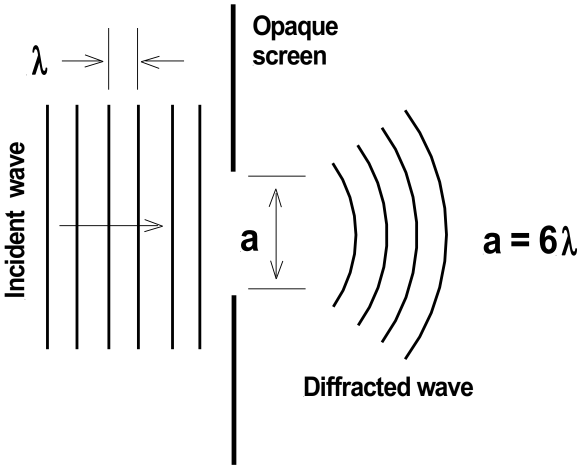


Figure 13 Diffraction through a pin hole 6λ in diameter (source: author)

In the last illustration, Figure 14, the hole is made so large that it is huge compared to a wavelength. In this case the diffraction effect completely disappears. The light wave simply passes through the hole as if the screen were not there.

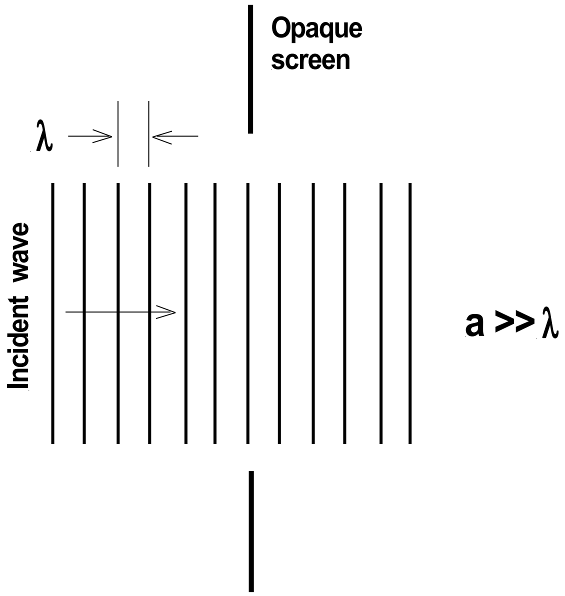


Figure 14 Diffraction through a large hole (source: author)

Diffraction of light around an object with sharp edges is illustrated in Figure 15. In this figure the bottom part of a wave front is removed by an obstacle. The top part of the wave continues on, unaffected, in the forward direction. As before, destructive interference between adjacent wavelets accounts for the wave's forward movement. However, destructive interference does not occur at the bottom of the wavefront because the lower part of the wavefront has been removed by the obstacle. Consequently, point sources of light at the bottom of the remaining wavefront propagate downward as well as straight ahead causing the light to bend around the obstacle.

Figure 16 illustrates the diffraction of a radio signal. In this figure a radio station is transmitting from the tower on the right side of the picture. A receiving station, the antenna on the left side of the figure, is in the shadow of a large hill and should not be able to hear the transmitted signal. The transmitted signal passes over the top of the hill and keeps on going in a straight line. However, some of the signal is diffracted downward by the hill top allowing the receiving station in the shadows to hear the transmission.

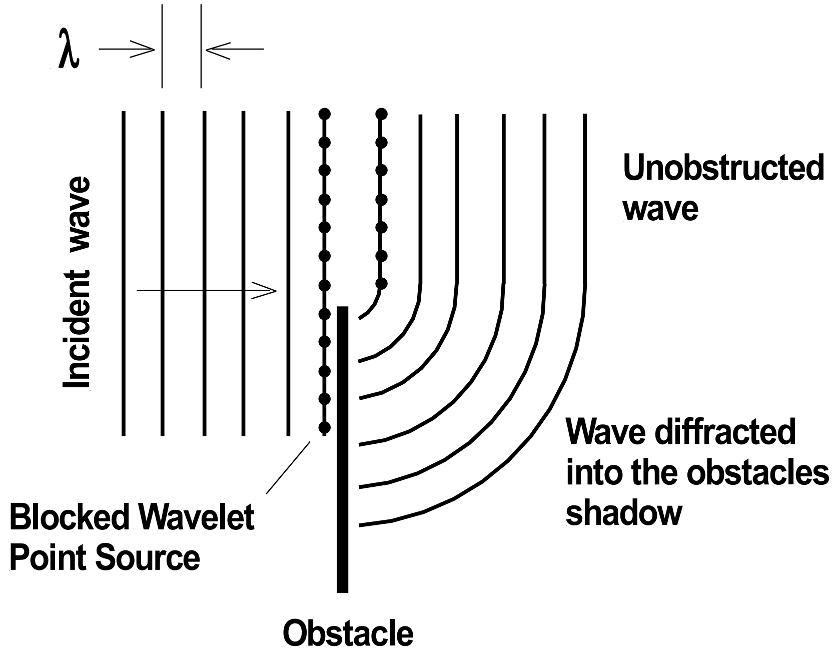


Figure 15 Diffraction of light around an obstacle (source: author)

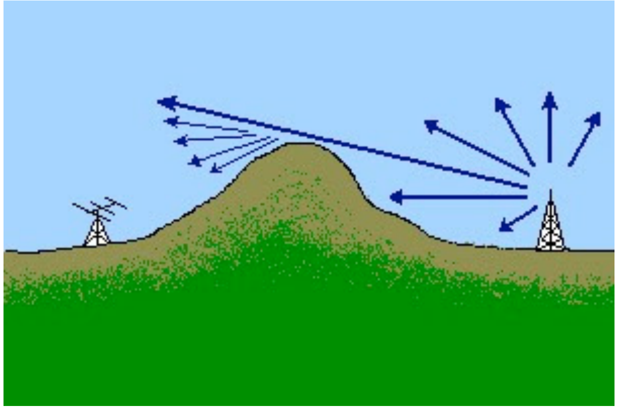


Figure 16 Diffracted radio signal (source: astrosurf)

Diffraction is a phenomenon of all types of waves. The picture in Figure 17 shows the diffraction of ocean waves through a small break in a sea wall. Diffraction is the reason that we can hear around corners. Sound waves travel from one room to another, spreading out through doorways and bending around the edge of buildings and other obstructions. Diffraction of radio signals by hills, buildings, and other structures are common occurrences at VHF radio frequencies and above. The diffraction of light is usually not noticeable because the wavelength of light is so short. However, light will diffract, as illustrated above, if an opening is extremely small or the edge of an object is sharp enough.

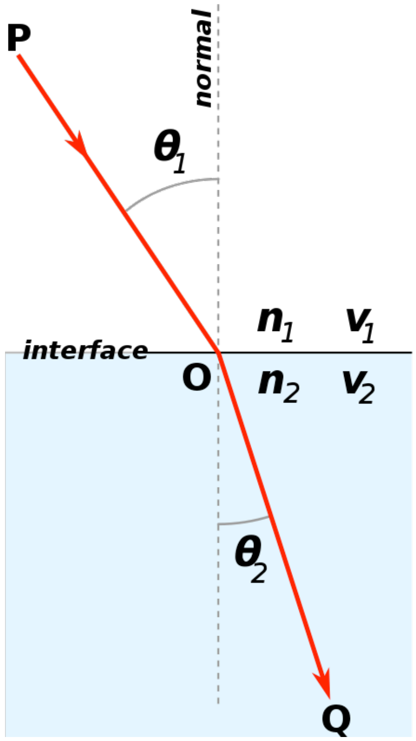
Figure 17 Diffraction of water waves (source: Verbcatcher Wikimedia)

14.2.3 Refraction of Light

Refraction is a property of light that causes it to bend at the interface between two different media as illustrated in Figures 18 and 19. Of the three phenomena reflection, diffraction, and refraction, refraction is the phenomena that is most important to us since it is the means by which HF radio signals propagate through Earth's ionosphere.

The amount of refractive bending depends on the angle of incidence. A beam of light traveling normal (perpendicular) to the interface between two media will pass straight through the interface without any bending. The amount of bending increases as the angle of incidence increases, illustrated in Figure 20. In this figure assume that the light is originating in the glass material and traveling upward through the interface into air. Within the glass the angle of incidence between the normal and the red colored ray of light is very small, about 0.3^o. The ray is bent 5^o with respect to the normal as it enters the air. The angle of incidence of the blue colored ray of light is a little larger than that of the red ray, about 30^o. However, the blue ray is bent at an angle of 45^o when it enters the air, considerably more bending than that of the red ray. The amount of bending for the purple ray (angle of incidence ~38^o) is even more, etc.

Claudius Ptolemy (~ 85 – 165 AD) was one of the first to systematically study refraction and document his results. Ptolemy described a stick appearing to bend when partially immersed in a pool of water (similar to Figure 19), and accurately recorded the angles he observed. He discovered that water and glass have different angles of refraction. However, he was unable to mathematically describe the extent to which light refracts in passing from one medium to another.



Medium 1

Medium 2

Figure 18 Refraction of a light ray (source: Wikipedia)

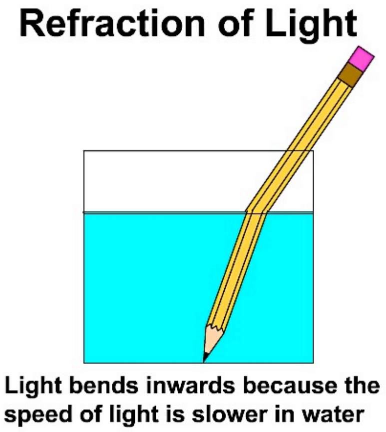


Figure 19 Pencil appears to bend in water (source: tonteraslight.blogspot.com)

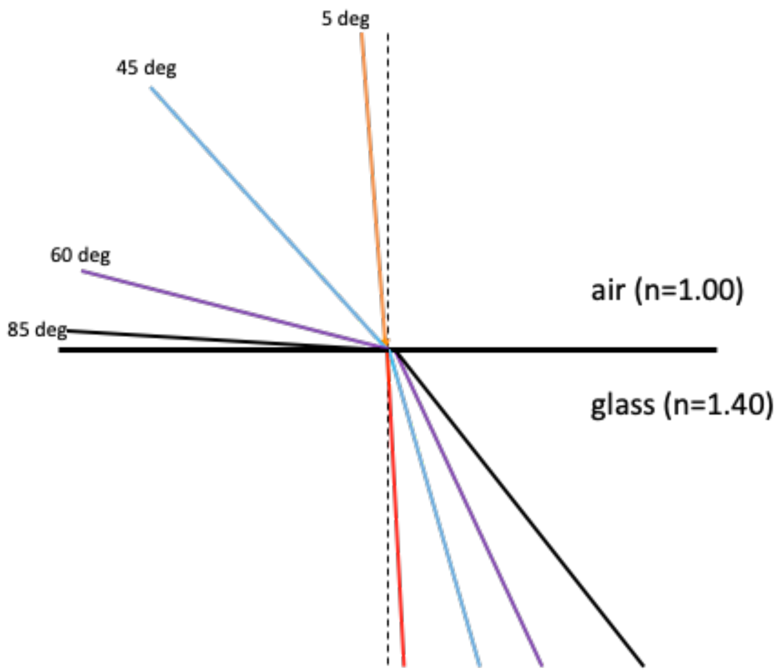


Figure 20 Degree of refraction (source: Quora)

14.2.3.1 Snell / Descartes Law of Refraction

English mathematician and astronomer Thomas Harriot (1560 – 1621) discovered the sine law of refraction for light in 1601. However, he did not publish his results. In 1621 Dutch astronomer and mathematician Willebrord Snell also discovered the sine law for refraction of light. But he didn't publish his results either. Finally, the sine law of refraction was published in 1637 by French mathematician and scientist Rene Descartes (1596 – 1650).

The Snell / Descartes law of refraction states that the amount by which light bends at the boundary between two media is given by

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

In this equation θ_1 is the angle of incidence with respect to the line perpendicular (normal) to the boundary as illustrated in Figure 18. The angle θ_2 is the angle of refraction. The refractive indices are n_1 for medium 1 and n_2 for medium 2 illustrated in Figure 18.

14.2.3.2 Huygens' Principle of Refraction

Christiaan Huygens proposed that the refractive index n of a material was related to the speed at which light travels through the substance. Huygens suggested that light travels more slowly through materials having a higher refractive index. Based on this assumption, Huygens postulated that the refractive index n_m of a material is equal to the ratio of the speed of light in a vacuum to the speed of light in the material, specifically

$$n_m = \frac{c}{v_m}$$

where c is the speed of light in a vacuum (299,792,458 m/s) and v_m is the speed of light in the material. As an illustration, the refractive index of water is 1.333 which means that light travels 1.333 times slower through water than in a vacuum.

Actually, v_m is the phase velocity of light in a material. Phase velocity is the speed at which the crest or phase of a wave travels. Phase velocity is usually different from group velocity. Group velocity is the speed at which information, consisting of many different frequencies, travels through a medium, for example the envelope of a modulated radio signal. Group velocity is also the speed at which the power carried by the wave travels. Group velocity and phase velocity are the same only in the vacuum of outer space.

Variations in the speed of light from one material to another was in fact the basis of Huygens wave theory of refraction. Huygens envisioned a wavefront to be like a line of soldiers marching along (Figure 21)

The soldiers on the left end of the line slow down as they encounter swampy terrain (yellow region in Figure 21), while the rest of the line (those in the gray region) continue marching along at their original speed. The difference in speed causes the line of soldiers to gradually bend toward the normal.

The same concept is illustrated in Figure 22. In this case the figure illustrates oblique wavefronts of light encountering the interface between air and glass. That part of a wavefront that has entered the glass slows down traveling at a speed of v_g instead of its original speed in air of v_a . In time t the wavefront in the glass covers a distance of

$$A \ to \ C = v_g \ t$$

That part of the wavefront that is still traveling through air continues at its original speed of v_a and in time t covers the much larger distance B to D. That is

$$B \ to \ D = v_a \ t$$

Consequently, the wavefront is forced to bend toward the normal.

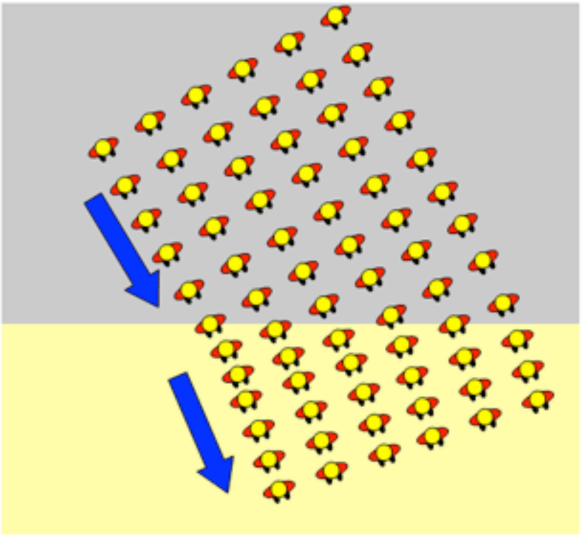


Figure 21 Huygens concept of refraction (source: esfsciencenew)

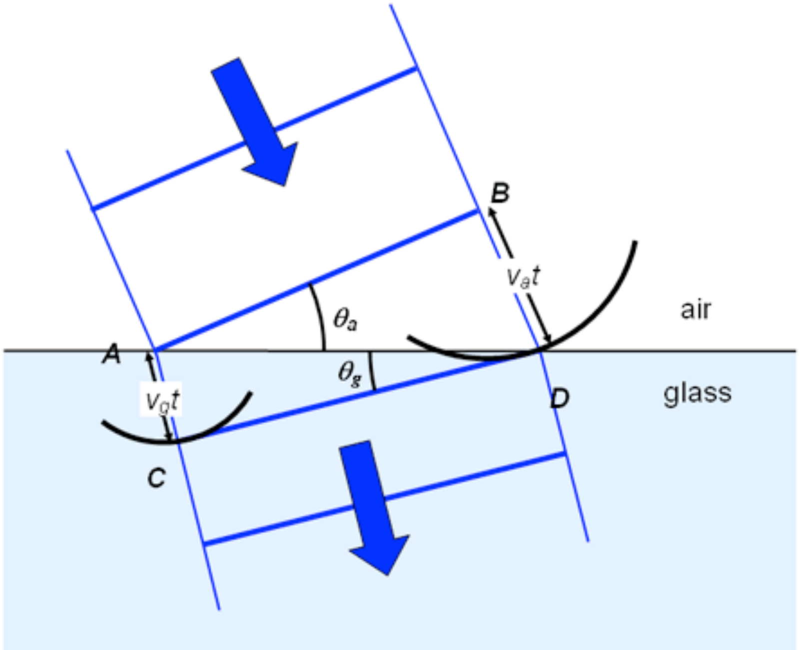


Figure 22 An illustration of refraction (source: resourcefulphysics.org)

If the blue arrows in Figure 22 are reversed, that is light traveling through the glass and into air, the wavefronts bend away from the normal as the light beam enters air. Specifically, that part of the wavefront that has entered the air travels at the much faster speed v_a covering the relatively larger distance D to B while the other end of the wavefront is still in the glass traveling at the much slower speed of v_g.

In more general terms, light bends toward the normal when passing from a material with a low refractive index to one with a higher index ($n_1 < n_2$) , as illustrated in Figure 23. In the reverse direction, light bends away from the normal when passing from a high refractive index material to a material with a smaller index ($n_1 > n_2$).

Huygens theory of refraction was initially rejected by the majority of seventeenth and eighteenth century scientists who lacked the ability to accurately measure the speed of light. It appeared to them that the speed of light was the same regardless of the material through which it passed. 150 years after Huygens' death the speed of light was measured accurately enough to prove that Huygens theories were correct.

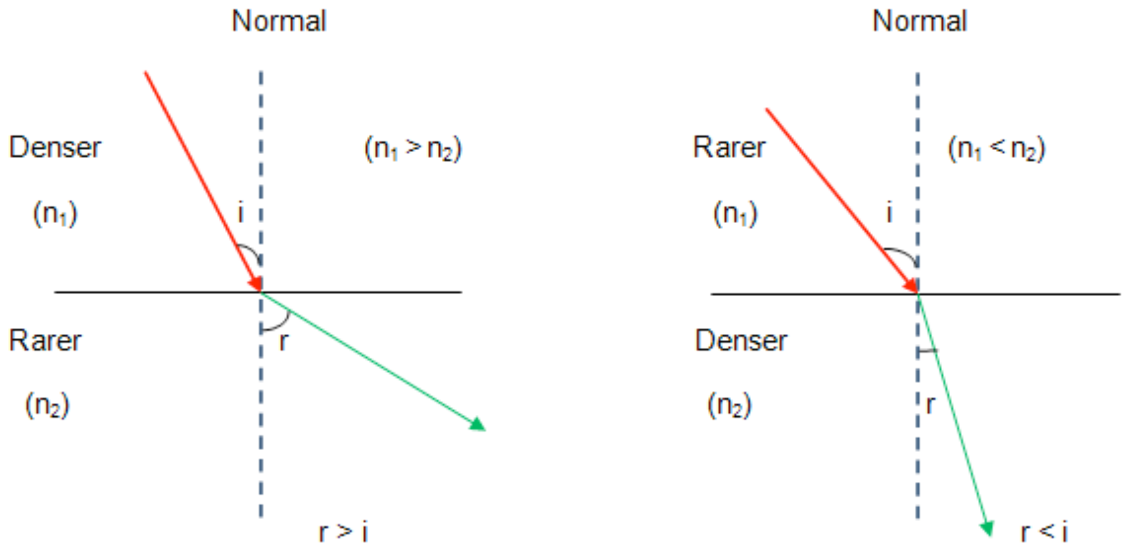
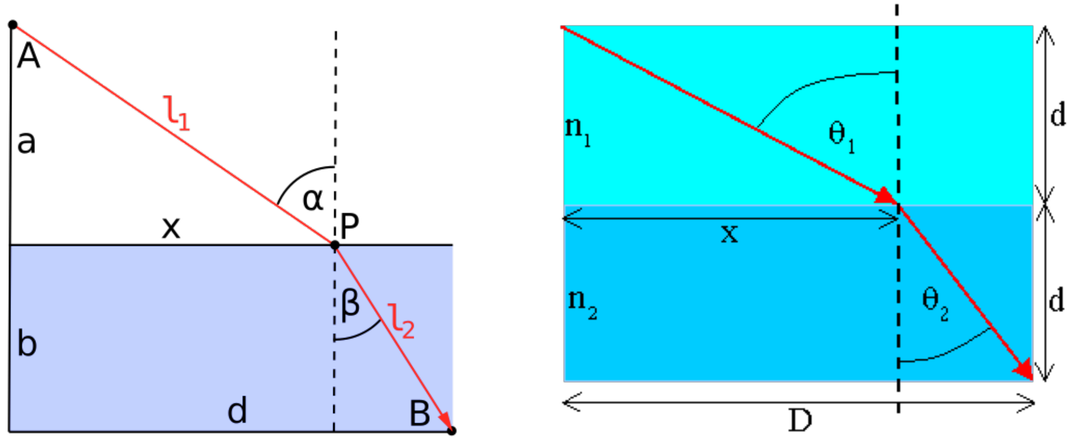


Figure 23 Bending toward and away from a line normal to boundary (source: http://buphy.bu.edu)

14.2.3.3 Deriving Snell's Law

Snell's Law can be derived using Fermat's principle of least time. Given a light source at point A in medium 1 (Figure 24a), and an observer at point B in medium 2, the refraction point P is that which minimizes the time taken by light to travel the path APB.



(a) (b)

Figure 24 Deriving Snell's law (source: electron9.phys.utk.edu)

Assume that light propagates from point A to point B across the boundary between medium 1 and medium 2 in Figure 24b

For the path shown in the figure the time required is

$$t = \frac{\sqrt{x^2 + d^2}}{c/n_1} - \frac{\sqrt{(D-x)^2 + d^2}}{c/n_2}$$

Setting dt/dx = 0 we obtain

$$\mathrm{n}_1 \frac{\mathrm{x}}{\sqrt{\mathrm{x}^2 + \mathrm{d}^2}} = \mathrm{n}_2 \frac{\mathrm{D} - \mathrm{x}}{\sqrt{(\mathrm{D}-\mathrm{x})^2 + \mathrm{d}^2}}$$

In this equation

$$\frac{\mathrm{x}}{\sqrt{\mathrm{x}^2 + \mathrm{d}^2}} = \sin\theta_1$$

and

$$\frac{\mathrm{D} - \mathrm{x}}{\sqrt{(\mathrm{D}-\mathrm{x})^2 + \mathrm{d}^2}} = \sin\theta_2$$

yielding Snell's Law

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

14.2.3.4 Speed of Light and Wavelength as a Function of Refractive Index

Snell's law of refraction can be written in two equivalent forms

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

and

$$\frac{\sin\theta_2}{\sin\theta_1} = \frac{n_1}{n_2} = \frac{v_2}{v_1}$$

The refractive index n_m for a medium is the factor by which the observed speed of light (v_m) in the medium is reduced relative to the speed of light in a vacuum ($v_0 = c$). Specifically

$$v_m = \frac{c}{n_m}$$

Similarly, the wavelength of light traveling through a medium (λ_m) is shorter than the wavelength of light in a vacuum (λ_0), that is

$$\lambda_m = \frac{\lambda_0}{n_m}$$

However, frequency f equals

$$f = \frac{v}{\lambda} = \frac{v_m}{\lambda_m} = \frac{\frac{c}{n_m}}{\frac{\lambda_0}{n_m}} = \frac{c}{\lambda_o}$$

so the frequency of light (its color) does not change as the light passes from one medium to another. For example, the color of an object in air is the same as its color in water as illustrated in Figure 25. It is only the velocity and wavelength of the light that changes.

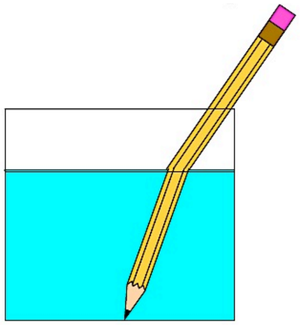


Figure 25 Refraction does not change the color of an object (source: tonteraslight.blogspot.com)

The index of refraction in a vacuum, n_0, is one since

$$n_0 = \frac{c}{v_0} = \frac{c}{c} = 1 \quad because\ \ v_0 \equiv c$$

Most gases at standard temperature and pressure have refractive indices close to 1 because of their low density. However, nearly all transparent solids and liquids have refractive indices above 1.3. For example, plastics generally have refractive indices of 1.3 to 1.7.

The refractive indices for many different types of material are shown in the table below. These values are measured at the yellow doublet D-line of sodium with a wavelength of 589 nanometers as is conventionally done. These indices can be much different for different wavelengths of light, for example higher in infrared light.

Material	Index of Refraction	State	Conditions
Vacuum	1		(by definition)
Helium	1.000036	Gas	(0°C and 1 atm)
Hydrogen	1.000132	Gas	(0°C and 1 atm)
Air	1.000277	Gas	(at STP*)
Air	1.000293	Gas	(0°C and 1 atm)
Carbon Dioxide	1.001	Gas	(0°C and 1 atm)
Liquid Helium	1.025	Liquid	(at -270°C)
Water Ice	1.31	Solid	(at 0°C)
Water	1.330	Liquid	(at 20°C)
Acetone	1.36	Liquid	(at 20°C)
Ethanol	1.361	Liquid	(at 20°C)
Kerosene	1.39	Liquid	(at 20°C)
Corn Oil	1.47	Liquid	(at 20°C)
Glycerol	1.4729	Liquid	(at 20°C)
Acrylic Glass	1.490–1.492	Solid	(at 20°C)
Benzene	1.501	Liquid	(at 20°C)
Crown Glass (pure)	1.50–1.54	Solid	(at 20°C)
Plate Glass (window glass)	1.52	Solid	(at 20°C)
Sodium Chloride (table salt)	1.544	Solid	(at 20°C)
Amber	1.55	Solid	(at 20°C)
Polycabonate	1.60	Solid	(at 20°C)
Flint Glass (pure)	1.60–1.62	Solid	(at 20°C)
Bromine	1.661	Liquid	(at 20°C)
Sapphire	1.762–1.778	Solid	(at 20°C)
Cubic Zirconia	2.15–2.18	Solid	(at 20°C)
Diamond	2.417	Solid	(at 20°C)
Silicon	3.42–3.48	Solid	(at 20°C)
Germanium	4.05–4.01	Solid	(at 20°C)

Selected refractive indices at λ=589 nm (source: Wikipedia)

14.3 Ionospheric Refraction of Radio Waves

Some media, including plasmas such as Earth's ionosphere, have index of refractions n_i which are less than 1 but greater than 0 ($0 < n_i < 1$).

The ionosphere is formed by photo-ionization. During the day, Extreme Ultra Violate (EUV) and X-ray radiation from the Sun is absorbed by the electrons of neutral atoms in the upper atmosphere, as illustrated in Figure 26. Occasionally an electron absorbs sufficiently energy to break away from its parent atom forming a free electron and a positive ion (an atom that has lost one of its electrons). There are a thousand times more neutral atoms in the upper atmosphere than ions. The ionosphere is thus very thin, wispy, and easily blown around.

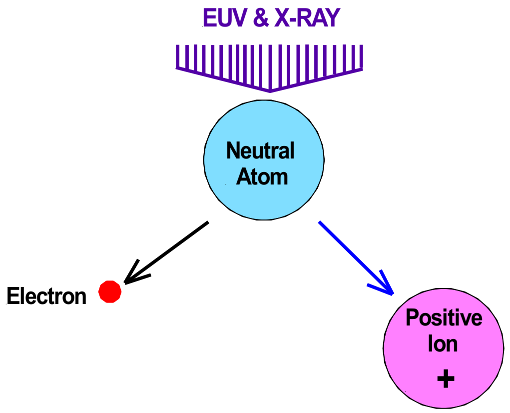


Figure 26 Photo ionization (source: author)

Radiation is intense at the top of atmosphere, as illustrated in Figure 27, but there are few atoms to ionize. As the radiation penetrates deeper into the atmosphere, it encounters more and more atoms resulting in higher levels of ionization. However, the ionization process absorbs energy from the EUV & X-ray radiation causing it to weaken in intensity. The number of atoms increase as radiation penetrates further down into the atmosphere, but levels of ionization decrease due to weakening radiation. Ionizations levels drop and eventually disappear at the bottom of the ionosphere. Consequently, the highest levels of ionization occur toward the middle of the ionosphere.

Beginning at the bottom of the ionosphere (Figure 27), electron densities increase with altitude above Earth's surface until the region of maximum density is reached somewhere near the middle of the ionosphere. Above that point electron densities decrease with altitude until the ionosphere disappears far above the Earth.

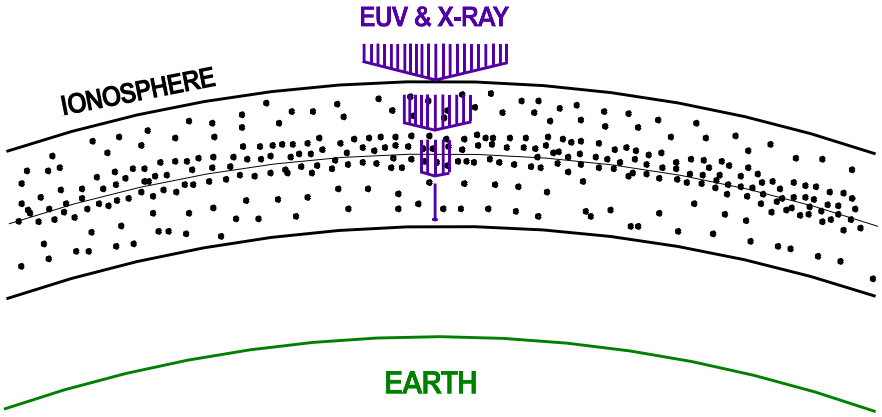


Figure 27 Formation of the ionosphere (source: author)

Since the ionospheric index of refraction n_i is greater than zero but less than one ($0 < n_i < 1$), a radio wave is bent away from the normal as it travels from air $n_a \approx 1$ into the ionosphere $n_i < n_a$. The ionosphere's index of refraction becomes smaller (closer to 0) as electron density increases. That is, the index of refraction at the bottom of the ionosphere n_{ib}, where there is little ionization, is greater than the ionospheric index of refraction n_{ic} in the highly ionized central region of the ionosphere, leading to the relationship

$$0 < n_{ic} \ll n_{ib} \leq n_a = 1$$

Consequently, a radio wave not only bends away from the normal when it enters the ionosphere, it continues to bend further and further away from the normal as it travels upward into the heavily ionized central region of the ionosphere. If a radio wave bends far enough (if it bends more than 90^{o} from the normal) it will begin traveling back to Earth as illustrated by the red trace in Figure 28. As it does so it encounters decreasing levels of ionization causing the index of refraction n_i to increase from its minimal value in the ionosphere's central region to a value of 1 (the index refraction of air $n_a = 1$) at the bottom of the ionosphere, that is

$$0 < n_{ic} \rightarrow n_i \rightarrow n_a = 1$$

Traveling from a region of a small index of refraction to one with a higher refractive index cause the radio wave to bend toward the normal. The radio wave not only bends toward the normal, it continues to bend closer and closer to the normal as it travels down through the ionosphere towards Earth's surface. In other words, the radio wave straightens out and travels to the distant Station B in Figure 28.

From this analysis it is clear that a radio wave can penetrate the ionosphere and be lost to outer space. The electron density of the ionosphere decreases with altitude from the middle to the upper edge of the ionosphere. At the upper edge ionization disappears all together. Consequently, the index of refraction must increase with altitude in the upper ionosphere until it reaches a value of one at the top of the ionosphere where ionization disappears and the vacuum of space begins. If n_{it} is the index of refraction at the top of the ionosphere then

$$n_{ic} < n_{it}$$

and

$$n_{it} \approx 1$$

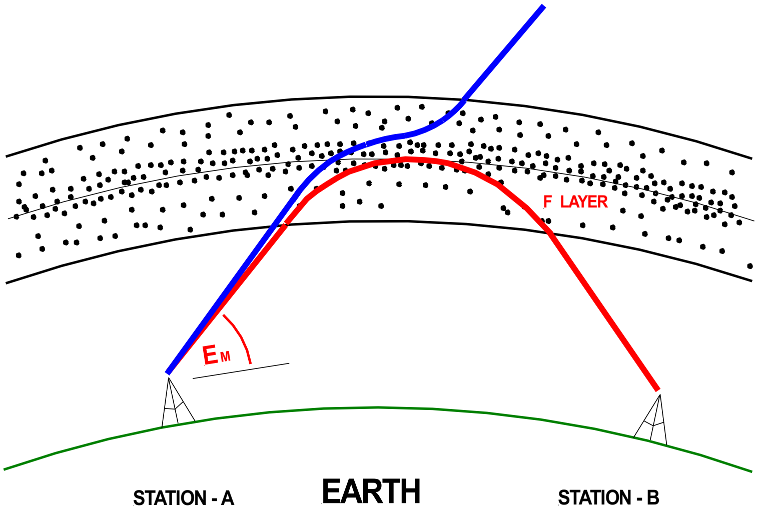


Figure 28 Ionosphere causes radio waves to be bent back to Earth (source: author)

A radio wave propagating through the upper ionosphere (blue trace in Figure 28) is continuously traveling from a region of low to a region of higher refractive index. Consequently, the radio wave is bent toward the normal, away from Earth and into outer space. A radio wave that penetrates the densest part of the ionosphere can never return to Earth. It is lost to outer space.

An index of refraction less than 1 implies that the phase velocity of a radio wave v_m can travel faster than the speed of light. Since

$$v_m = \frac{c}{n_m} \quad if \;\; n_m < 1 \quad then \;\; v_m > c$$

How is this possible? Phase velocity is simply the speed at which the crest of a wave travels. Since the wave crest does not carry any information or any of the wave's power, it can travel faster than the speed of light. However, the group velocity of a modulated radio wave carrying information, plus the power associated with the wave, is prevented by Einstein's theory of relativity from traveling faster than the speed of light. Likewise, all material objects are prevented from traveling faster than the speed of light.

This phenomenon can actually be observed by watching waves at the beach. In Figure 29 waves are approaching the shore at an oblique angle. The speed of an approaching wave is its group velocity, that is the speed associated with the movement of water as the wave passes. When the wave impacts the beach, the peak of the wave appears to race along the shoreline at a speed much fast than that of the wave itself. The difference in speed can be quite impressive, with the wave's phase velocity being considerably greater than its group velocity.

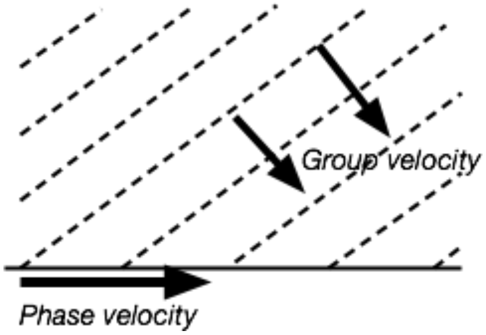


Figure 29 Phase velocity vs group velocity (source: Quora)

The phase velocity of radio waves in most materials, such as the plastic insulation of an antenna wire, is less than c. However, it can exceed c in plasmas such as the ionosphere. Only in empty space are the phase and group velocities identical, both being equal to c.

14.4 Frequency Dispersion

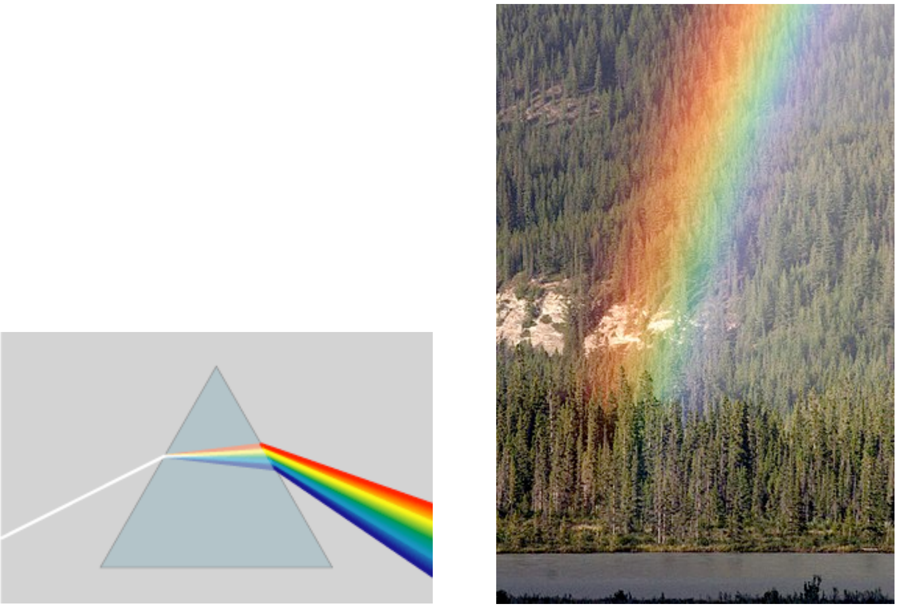
The refractive index of a material varies with the wavelength of light traveling through the material. Thus, the index of refraction n is a function of wavelength $n(\lambda)$.

This property is called dispersion. Dispersion causes prisms and water droplets to separate white light into its constituent spectral colors (Figure 30). It also causes the focal length of lenses to vary with wavelength causing chromatic aberrations which often need to be corrected in imaging systems. It was the chromatic aberration present in refracting telescopes that first lead Newton to suspect that white light was not pure.

The degree to which the refractive index varies with wavelength is given by Cauchy's equation which, in its most common form, is

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$$

where A, B, C, etc., are coefficients determined by the material through which the light is passing.



Light refracting through a prism and through water droplets forming a rainbow of colors

Figure 30 Dispersion of light (source: Wikipedia)

For visible light, the refractive indices for most transparent materials (air, glass, etc.) generally increase as the wavelength of light becomes shorter. This is called normal dispersion. The wavelength of blue light is shorter than that of red light. Consequently, the index of refraction of blue light n_{blue} is greater than that of red light n_{red} , causing blue light to be bent more than red as it passes through a prism.

$$n_{blue} > n_{red}$$

This is clearly apparent in Figure 30.

The reverse is true in material having an index of refraction less than 1, which is the situation in Earth’s ionosphere. In this case, known as anomalous dispersion, refraction increases as wavelength becomes longer, causing long wavelength radio waves to be refracted more than short wavelength signals, as illustrated in Figure 31. Long wavelength 80-meter signals (3.5 MHz) are refracted more, bending back to Earth lower in the ionosphere, than short wavelength 20-meter signals (14.0 MHz). If a radio wave passes through the densest part of the ionosphere without being refracted back to Earth (the 15-meter signal in Figure 31), it will be lost to outer space.

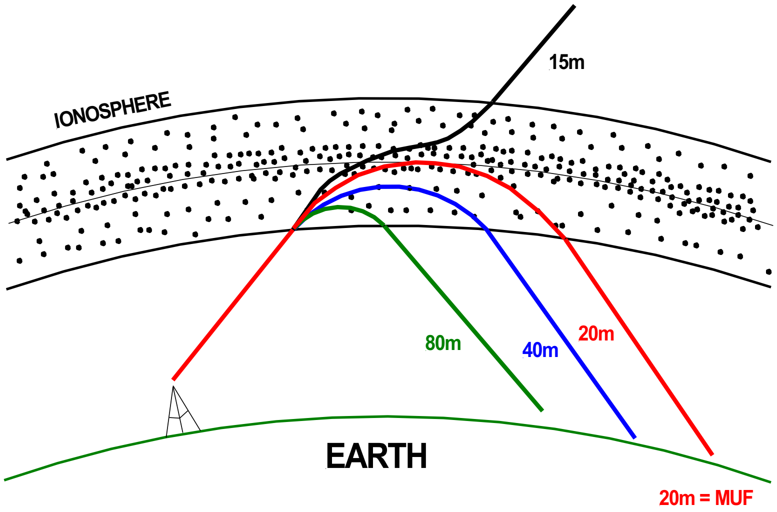


Figure 31 Anomalous dispersion in the ionosphere (source: author)

One of the serious problems resulting from dispersion is called group velocity dispersion. An information carrying signal is composed of many different frequency components. The velocity v_{f1} of a single frequency component is $v_{f1} = c/n_{f1}$ where n_{f1} is the index of refraction for that frequency as it travels through a material. Since each of the signal's frequency components has a different index of refraction n_{fm} they will all travel through the material at different speeds due to dispersion. The result is signal distortion, for example the widening of a signal pulse.

Consequently, dispersion management is extremely important in communication systems. If dispersion is too high in a fiber optic cable, for example, a group of pulses representing a bit stream will spread in time, merge, and render the bit stream useless. This limits the cable length that a signal can be sent down without regeneration.

14.5 Polarization

Polarization is defined as the direction of the electric field in an electromagnetic wave. The wave in Figure 32 is vertically polarized since its E-field oscillates up and down in the vertical direction. In this figure the E-field is the red wave while the blue wave represents the orthogonal magnetic field. Figure 32 is known as a linearly polarized wave since each of its fields oscillate in a single direction.

Electromagnetic waves can also be circular and elliptically polarized as shown in Figure 33. The E-field (as well as the magnetic field) in circular and elliptical polarized waves rotate as the wave travels. Rotation can be in either of two directions, right hand or left hand circulation. A wave that is right circular polarized rotates to the right (clockwise) with respect to the direction of wave travel as illustrated in Figure 34. A left circular polarized wave rotates in the opposite direction (counter clockwise).

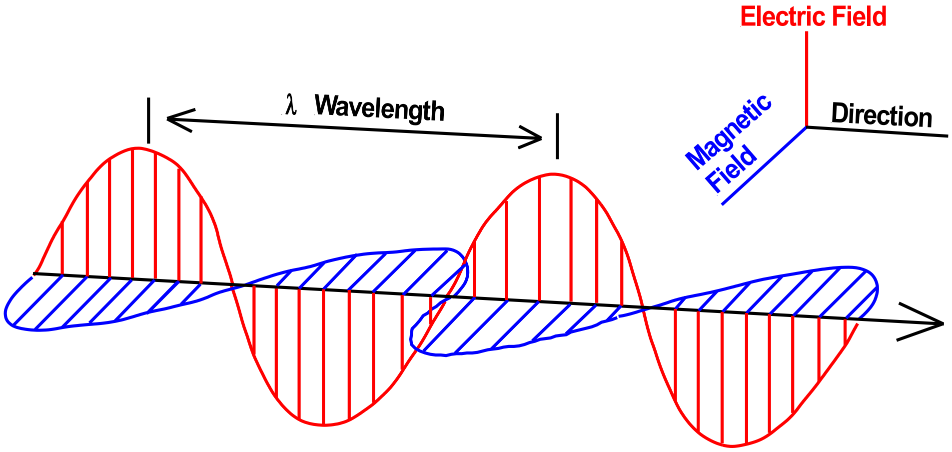


Figure 32 Vertically polarized electromagnetic wave (source: author)

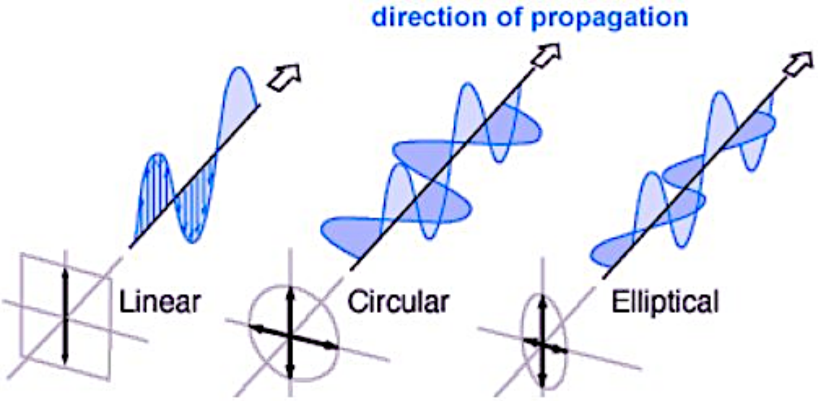


Figure 33 Typical polarizations of electromagnetic waves (source: Electronics For You)

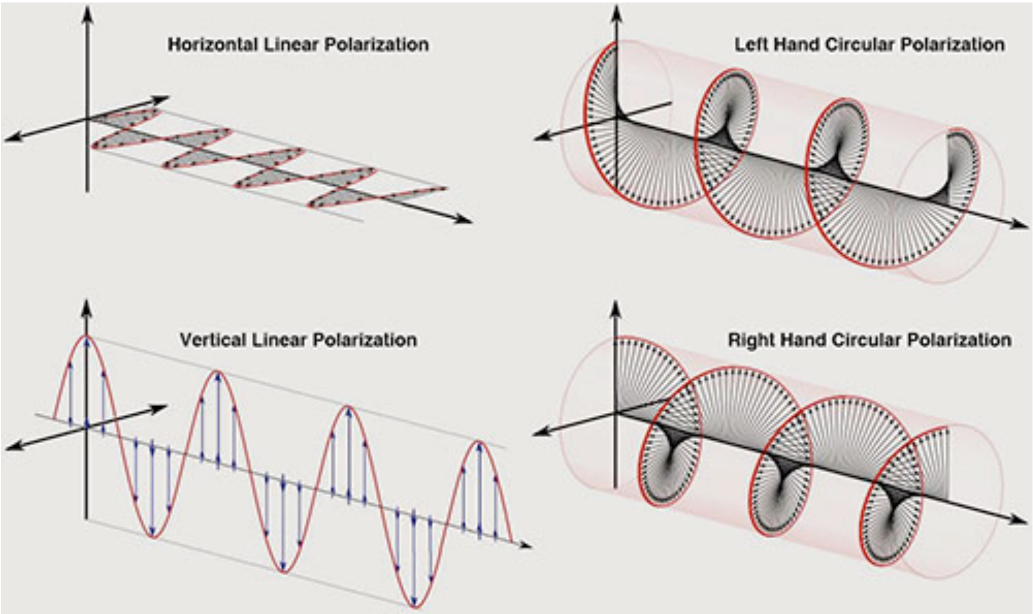


Figure 34 Linear and Circular Polarization (source: drones.stackexchange.com)

A vertical transmitting antenna radiates a linear vertically polarized electromagnetic wave as illustrated in Figure 35. The oscillating electrical current flowing up and down in the vertical green antenna produces the vertically oscillating electric field of the radiated wave.

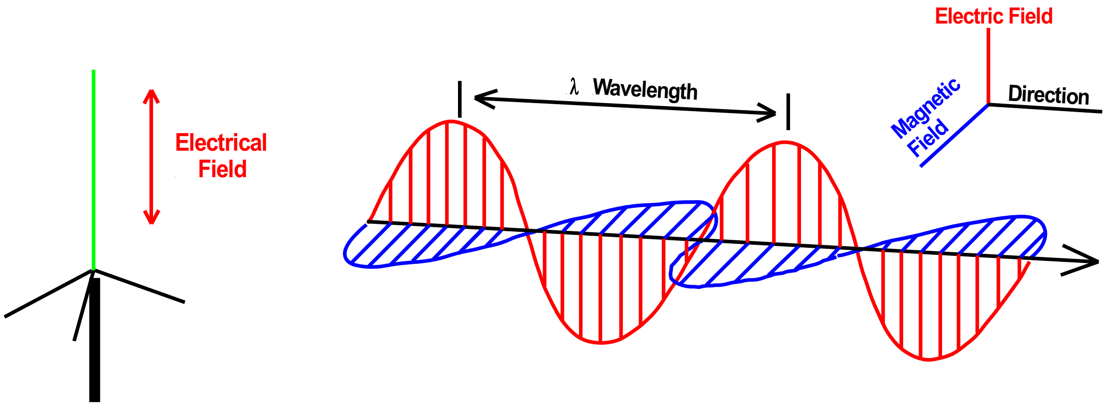


Figure 35 Vertical Antenna Polarization (source: author)

Figure 36 shows a horizontal transmitting antenna. The electrical current flowing back and forth in the green antenna wire produces an oscillating horizontal electric field in the electromagnetic wave that radiates outward from this antenna.

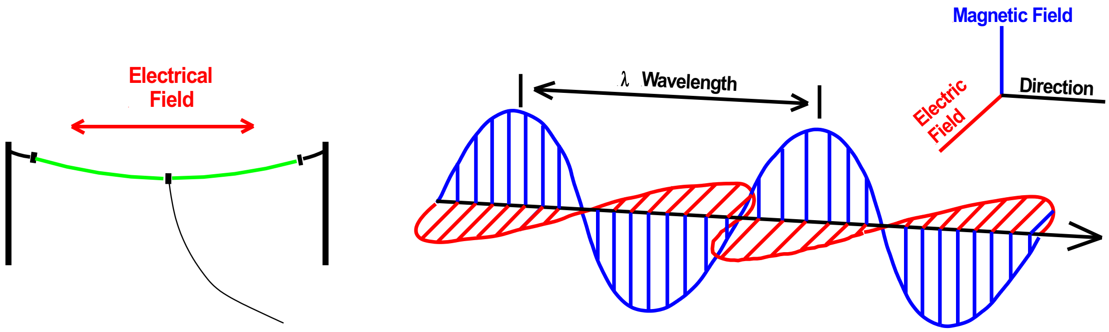


Figure 36 Horizonal antenna polarization (source: author)

The direction of polarization is very important for line-of-site (LOS) communications between two radio stations, illustrated in Figure 37. For good communications, the polarization of the receiving antenna must be the same as that of the transmitting antenna. If the transmitting station uses a vertical antenna, then a vertical antenna must also be used at the receiving site.

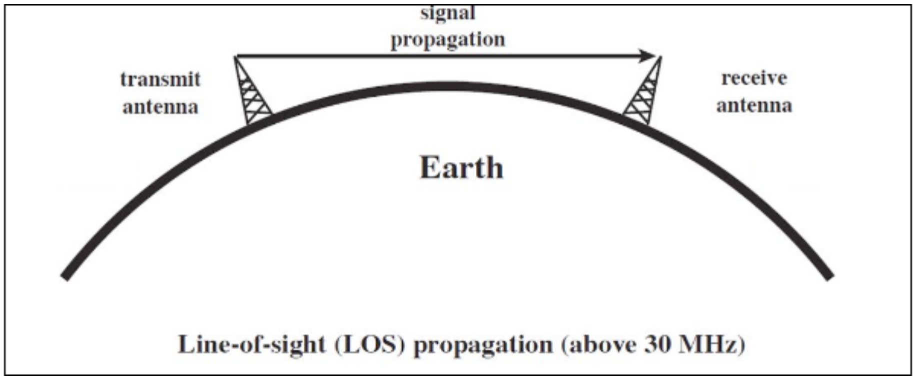


Figure 37 Line of sight radio transmission
(source: www.tutorialspoint.com/antenna_theory)

Figure 38 illustrates the importance of antenna orientation. On the left side of this figure both the transmitting and receiving antennas are vertically polarized. The vertically oscillating electric field of the electromagnetic wave induces an electrical current moving vertically up and down in the receiving antenna. Consequently, the percent of power absorbed by the receiving antenna is nearly

100%. However, on the right side of the figure the transmitting antenna is horizontal while the receiving antenna is vertical. The horizontal oscillating electric field is perpendicular to the receiving antenna and thus can not induce an electrical current in the antenna. In this case the percentage of power absorbed by the receiving antenna is essentially zero.

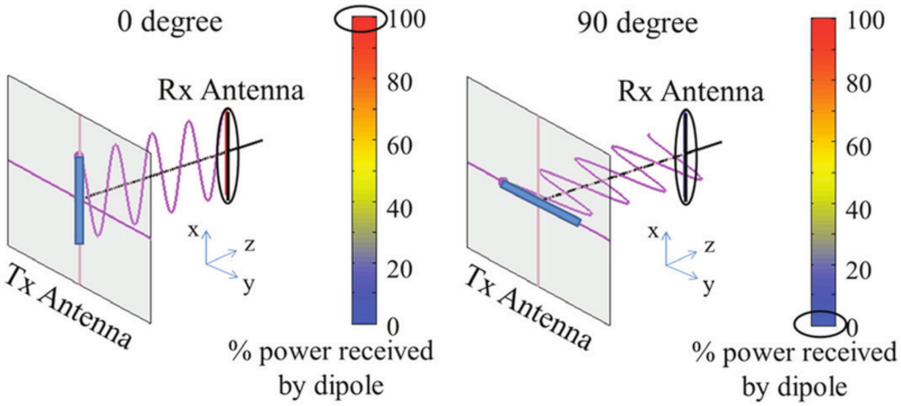


Figure 38 Polarized transmitting & receiving antennas (source: ResearchGate)

The electric field of an electromagnetic wave is a vector quantity $\vec{E}$ with a magnitude E and a direction θ with respect to the receiving antenna. The magnetic field is also a vector quantity $\vec{H}$ perpendicular to $\vec{E}$. The vector $\vec{E}$ can be partitioned into two segments E_p parallel to the receiving antenna and E_o orthogonal to the antenna. Only the parallel component E_p is capable of inducing an electrical current into the antenna. The current I induced into the antenna by the electromagnetic wave is thus proportional to $E_p = E\cos\theta$, that is

$$I \propto E\cos\theta$$

If the electric field of an arriving radio wave is parallel to the receiving antenna, the angle $\theta = 0°$ so

$$I \propto E\cos\theta = E\cos 0° = E \cdot 1 = E$$

and

$$I \propto E$$

If the arriving radio wave electric field is perpendicular to the receiving antenna, the angle $\theta = 90°$ and

$$I \propto E\cos\theta = E\cos 90° = E \cdot 0 = 0$$

in which case

$$I = 0$$

Consequently, polarization matters for line-of-site communications.

This is not the situation for long distance skywave communications through the ionosphere.

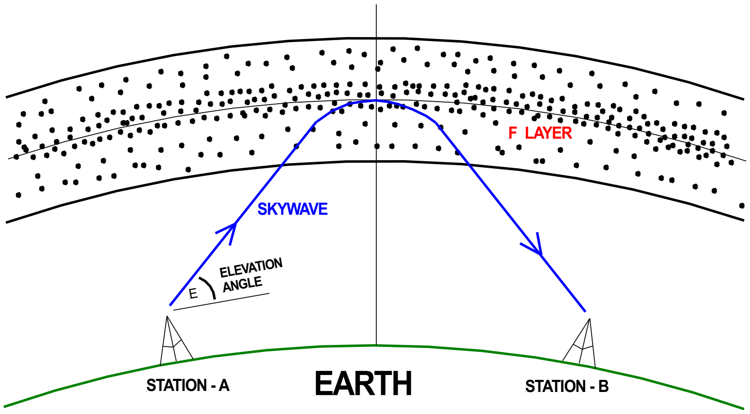


Figure 39 Radio communications through the ionosphere (source: author)

In Figure 39, high frequency 2 – 30 MHz radio waves are bent (refracted) back to Earth by the ionosphere. All linearly polarized electromagnetic waves entering the ionosphere are split into two counter-rotating circularly polarized waves, one called the Ordinary wave and the other being the Extraordinary wave. Circularly polarized waves are easily received by both vertical and horizontal antennas. Consequently, for skywave communications the orientation between the transmitting and receiving antennas are unimportant. A horizontally polarized transmitting antenna and vertical receiving antenna work just as well as if both antennas were the same polarization (both horizontal or both vertically polarized).

Communications between ground stations and spacecraft (Figure 40) is strictly line-of-site, typically at 145 MHz to 10 GHz. As discussed above, line-of-site communications requires that both transmitting and receiving stations use the same antenna orientation (both horizontal or both vertical). However, this is impossible to achieve with spacecraft whose orientation with respect to Earth is constantly changing. Consequently, communications with spacecraft typically utilizes circular polarized antennas such as those shown in Figure 41.

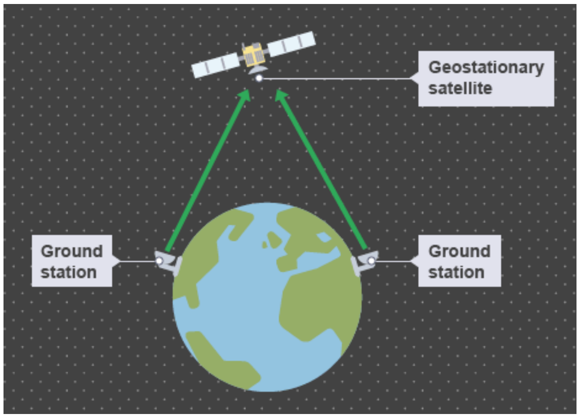


Figure 40 Line-of-site spacecraft communications (source: www.blendspace.com)

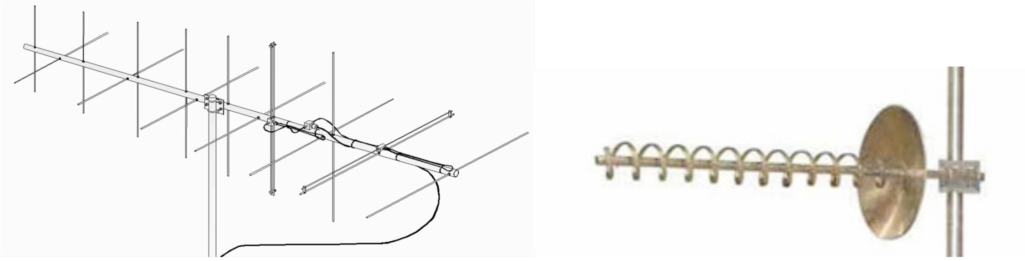


Figure 41 Circular polarized antennas (source: DX Engineering)

14.6 Ordinary and Extraordinary Waves

A plasma in the presence of a magnetic field becomes a Birefringent material. Birefringence is the optical property of a material having a refractive index that changes depending on the polarization of the light passing through the material and the light's direction of travel. Birefringence is responsible for the phenomenon of double refraction in which a ray of light incident upon a birefringent material is split by polarization into two rays taking slightly different paths through the material. The two rays are called the ordinary ray and the extraordinary ray. The ordinary ray has a refractive index of n_o whereas the refractive index of the extraordinary ray is between n_o and n_e depending on the ray's direction of travel. The magnitude of the difference in refractive indices is defined as the Birefringence

$$\Delta n = n_e - n_o$$

Propagation of the ordinary ray through the birefringent material, governed by the refractive index n_o, occurs as if birefringence was not involved. Propagation of the extraordinary wave depends on the ray's direction of travel and the associated birefringent refractive index n.

The ionosphere is a Birefringent material. When a linearly polarized radio wave enters the ionosphere, it splits into two separate waves each with a slightly different mode of propagation through the ionosphere (slightly different indices of refraction resulting in slightly different velocities and direction of travel). Both waves are circularly polarized with one wave rotating clockwise and the other counter-clockwise. Consistent with optical terminology, one wave is designated as the ordinary (O-mode) wave and the other the extraordinary (X-mode) wave.

To a large extent, the ordinary wave propagates through the ionosphere as if Earth's magnetic field was not present. The extraordinary wave propagates differently. The X-mode wave travels slower and tends to be more lossy.

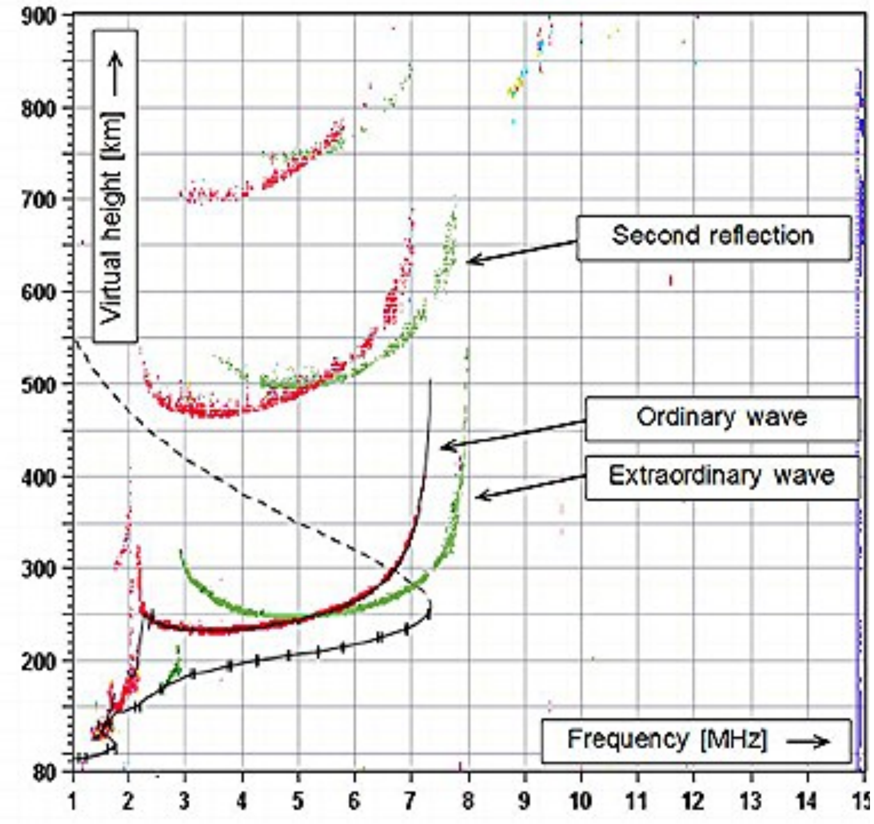


Figure 42 Ionogram showing both ordinary and extra ordinary modes (source: ResearchGate)

Ordinary and extraordinary waves are visible in ionograms such as the one shown in Figure 42. [Ionograms are covered in the Critical Frequency chapter.] The ionogram in this figure consists of two traces, one corresponding to the O-mode (red trace) and the other to the X-mode wave (green trace). The X-mode is generally the ionosonde higher frequency trace. That is, the frequency at which a X-mode signal reflects from a given height in the ionosphere is generally greater than the frequency of an O-mode signal reflecting from the same altitude. In Figure 42 a 6.7 MHz O-mode signal reflects at an altitude of 300 km while a 7.4 MHz X-mode signal reflects from the same altitude. Looked at from a different perspective, for a given frequency, the altitude at which the O-mode and X-mode signals reflect is difference. At a frequency of 3 MHz in Figure 42 the X-mode signal reflects at a higher altitude than the O-mode signal, roughly 290 verses 230 km. At a frequency of 7 MHz the situation is reversed. The O-mode signal reflects at the higher altitude, 330 verses 270 kms. The black trace in Figure 42 is the electron density profile superimposed on the ionogram.

The birefringent nature of the ionosphere has a profound effect on HF radio communications. We tend to think of HF signal propagation as being reciprocal. For example, in Figure 43 Station-A transmits a message through the ionosphere to Station-B. We presume that the response from Station-B travels back along the same path to Station-A, but of course in the opposite direction. With good radio equipment, the common belief in amateur radio is "if I can hear Station-A then Station-A can hear me". That is not at all true! Because of birefringence the propagation path through the ionosphere from Station-A to Station-B is nearly always different than the path from Station-B back to A, as illustrated in Figure 44. Note in Figure 44 that the elevation angle E_A is not the same as elevation angle E_B, that is $E_A \neq E_B$. The propagation paths between two stations are the same only if the entire path is along the equator.

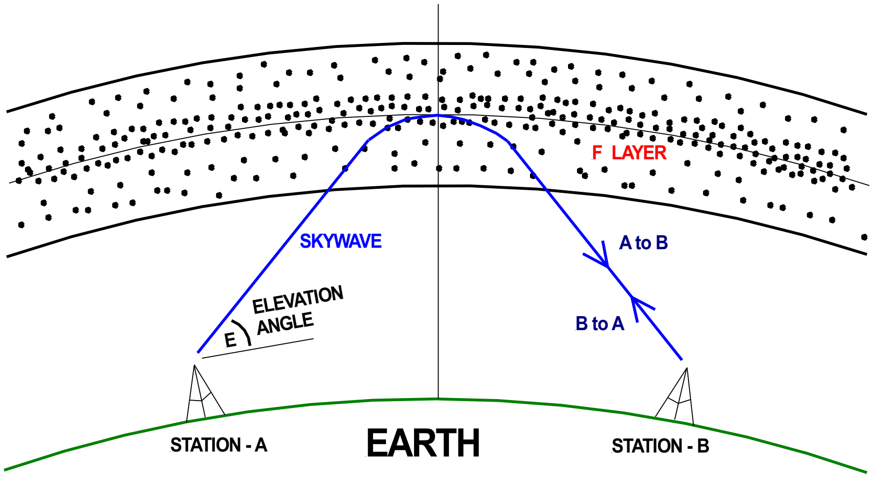


Figure 43 Reciprocal signal propagation rarely occurs (source: author)

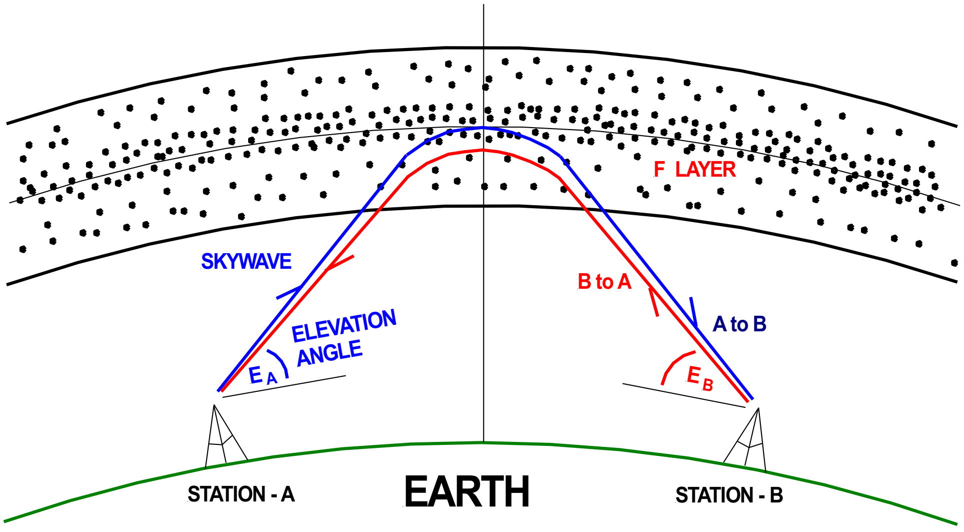


Figure 44 Non- reciprocal nature of HF signal propagation (source: author)

O and X mode waves can take radically different paths through the ionosphere, particularly at the low end of the HF frequency band in the neighborhood of 80 meters (~ 3.5 MHz). At the high end of the HF band, around 10 meters (~ 28 MHz), the O and X waves usually follow similar paths through the ionosphere. This means that the signal arriving at the receiving station is either the O mode or the X mode signal, not both. Once split, the O and X mode signals do not come back together again.

It is amazing that HF communications through the ionosphere is possible at all given that HF signals transmitted back and forth are not reciprocal. What makes HF communications possible are the relatively crude antennas that we use. We tend to think of our antennas as transmitting very narrow laser beams through the ionosphere to the distant station, and receiving a similarly narrow laser beam back from that station. If that were the case, birefringence would make ionospheric HF communications pretty much impossible, or at the very least, force us to use separate transmit and receive antennas. However, the energy radiated by our antennas are not laser beams but instead illuminate a large part of the sky as illustrated in Figure 45.

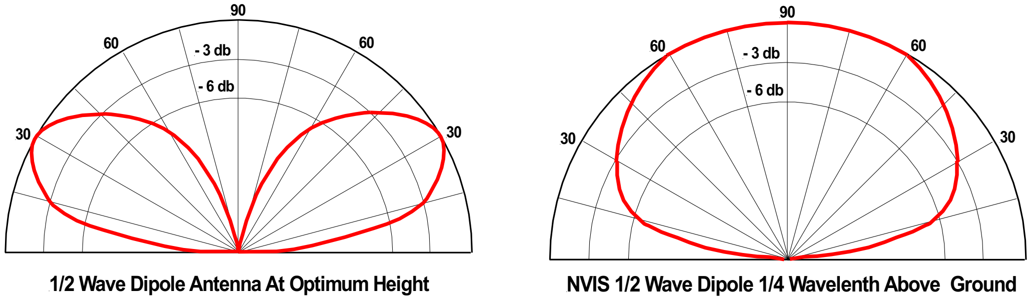


Figure 45 Elevation angels for a dipole antenna at two different heights (source: author)

The radiated energy represents an infinite number of "rays" injected into the ionosphere over a wide range of elevation angles. For example, the NVIS antenna in Figure 45 radiates rays at 90° (straight up). It also radiates rays at 60° and some at 30° and so on. All of these rays refract through the ionosphere and return to Earth over a broad range of distances from the transmitter, as illustrated in Figure 46. Some rays, if transmitted at too high of an elevation angle E, do not return to Earth at all but are lost to outer space. Signals being lost to outer space places a limit on how close a receiving station can be to the transmitting station. In Figure 46, the closest station capable of receiving transmissions from Station-A is the station located at B. Stations located closer than B can not hear Station-As transmissions. These stations are "skipped over", hence the term Skip Distance in Figure 46.

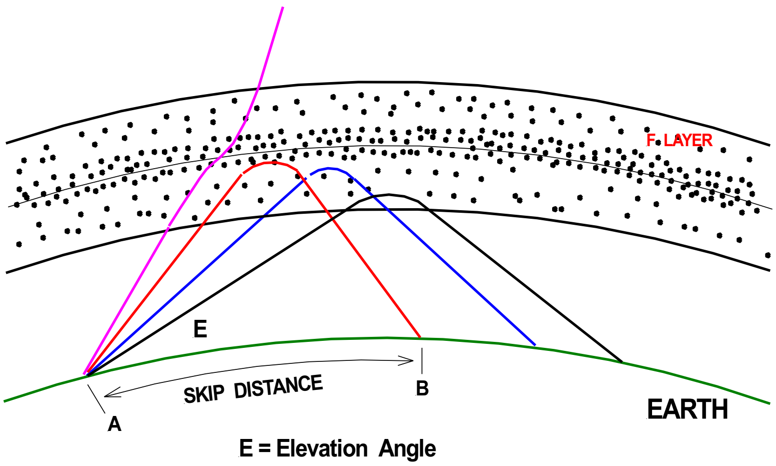


Figure 46 Radio waves return to Earth over a broad range of distances (source: author)

If we are lucky, a ray transmitted at an elevation angle E_A will be received by Station-B in Figure 44. However, is the signal received by Station-B an O-mode or an X-mode ray? In general we don't know. The signal transmitted back by Station-B follows a completely different path through the ionosphere and could be received by Station-A as either an O-mode or an X-mode signal. We know from ionograms, such as the one in Figure 42, that O-mode and X-mode signals at a given frequency refract from different altitudes in the ionosphere. We would thus expect that an X-mode signal received from Station-B would arrive at a different elevation angle E_{Bx} than the O-mode signal transmitted by Station-A. We are able to "experience" reciprocal communications, typically transmitting at one elevation angle and receiving back at a different angle, only because the radiation patterns of our antennas are so broad. Turns out for us that using relatively crude antennas is a good thing.

The fact that signal propagation is not reciprocal means that background interference experienced along the forward path (from Station A to Station B) is most likely different than interference along the reverse path. For example, Station B may experience considerable fading in the signal it receives from Station A. But Station A may detect very little fading in Station B's signal. Just because fading is occurring in one direction does not necessarily mean that it is occurring in the opposite direction.

Since O-mode and X-mode signals travel along different paths, it is extremely unlikely that both signals will arrive at the receiving station. Usually only one signal will be received, either the O-mode or the X-mode signal. This is very fortunate. The phase of the X-mode signal is continuously changing relative to the O-mode signal since it is traveling along a different path. If both O and X mode signals arrived at the receiving site, the phase difference between them would most likely cause them to interfere destructively, severely attenuating the signal heard by the receiving station. Occasionally, NVIS stations transmitting and receiving signals nearly straight up and down do receive both the O and X mode signals. However, the O-mode and X-mode signals travel slightly different distances in propagating between the relatively close transmitting and receiving NVIS stations. The difference in distance traveled causes one of the signals to arrive at the receiving station slightly later than the other signal, typically a few milliseconds later. The slight time delay often causes the received signal to sound hollow or echoey.

To make matters more interesting, in the northern hemisphere O-mode signals diverge toward the north magnetic pole while X-mode signals diverge toward the equator. The divergence is more extreme in the polar regions but exists to some extent everywhere, except exactly along the magnetic equator. This has some interesting consequences. An X-mode signal transmitted from Los Angeles California to Fairbanks Alaska could end up instead in Hawaii. While the X-mode transmission from Los Angeles, initially headed for Fairbanks, bent back to Hawaii, O-mode transmissions from Hawaii to Los Angeles could follow a more direct route experiencing less fading along the way.

We assume that our radio transmissions follow Great Circle paths from the transmitting to the receiving station. That is not necessarily true as illustrated above. O-mode signals more closely follow Great Circle paths and are generally stronger than X-mode signals. However, Great Circle paths become less meaningful for signals passing through the polar regions.

References

Nichols, Eric P.; “Propagation and Radio Science”; The American Radio Relay League, Inc. 2015

Hunsucker R. D.; Hargreaves, J. K.; “The High-Latitude Ionosphere and its Effects on Radio Propagation”; Cambridge University Press 2003

Davies, Kenneth; “Ionospheric Radio”; Peter Peregrinus Ltd.,1990

McNamara, Leo F.; “The Ionosphere: Communications, Surveillance, and Direction Finding”; Krieger Publishing Company, 1991

Yeang, Chen-Pang; “Probing The Sky With Radio Waves”; The University of Chicago Press, 2013

Devoldere, John; “Low-Band DXing” fourth edition; ARRL, 2005

Levis, Curt A. ; Johnson, Joel T.; and Teixeira, Fernando L.; “Radiowave Propagation Physics and Applications”; John Wiley & Sons, Inc., 2010

Goodman, John M.; “Space Weather & Telecommunications”; Springer Science+Business Media Inc. 2005

UCAR Center for Science Education (UCAR SciEd); https://scied.ucar.edu/learning-zone/atmosphere/

Khazanov, George V.; “Space Weather Fundamentals”; CRC Press, 2016

Foukal, Peter; “Solar Astrophysics third edition”; Whiley-VCH Publishing Company, 2013

Ganushkina, N. Yu.; Liemohn, M. W.; Dubyagin. S.; “Current Systems in the Earth’s Magnetosphere”; AGU, March 8, 2018

Gallagher, Dr. D.L.; “The Earth’s Plasmasphere”; Space Plasma Physics, Marshall Space Flight Center, Huntsville, Al, September 05, 2018

Moore, T. E., Morwitz, J. L.; “Stellar Ablation of Planetary Atmospheres”; August 9, 2007

Yau, Andrew W.; Abe, Takumi; Peterson, W. K.; “The Polar Wind: recent Observations”; Department of Physics and Astronomy, University of Calgary

Carroll, Bradley W. and Ostlie, Dale A.; “An Introduction to Modern Astrophysics”; Addison-Wesley Publishing Company Inc., 1996

Cander, Ljiljana R.; “Ionospheric Space Weather”; Springer Nature Switzerland AG 2019

Moldwin, Mark; “An Introduction to Space Weather”; Cambridge University Press, 2008

Campbell, Wallace H.; “Introduction to Geomagnetic Fields”; Cambridge University Press, 2003

Golub, Leon and Pasachoff, Jay M.; “Nearest Star The Surprising Science of Our Sun second edition”; Cambridge University Press, 2014

Loff, Sarah: ”Explorer and Early Satellites”; National Aeronautics and Space Administration, Aug 3, 2017

Minzner, R. A.; “The 1976 Standard Atmosphere Above 86 km Altitude” NASA Goddard Space Flight Center, 1976