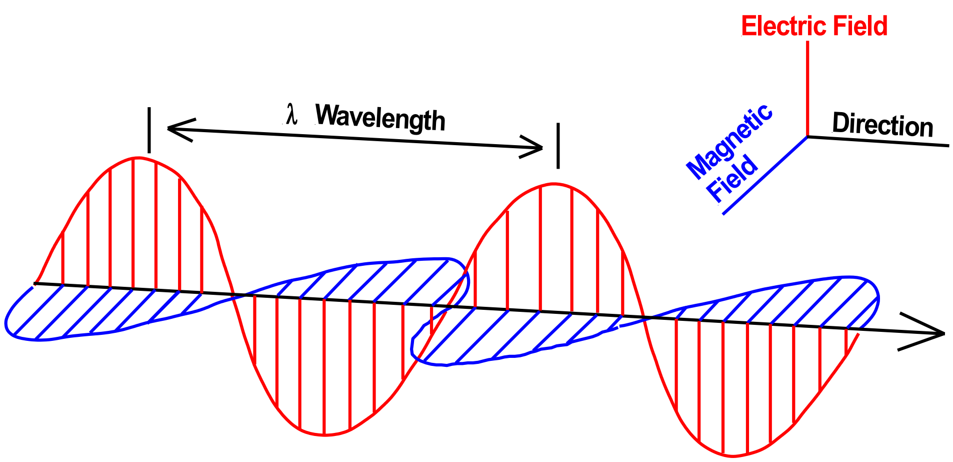
The Wave Nature of Light



www.skywave-radio.org

4 Wave Nature of Light

4.1 Debate Over the Physical Properties of Light

In 1672 Newton was elected a fellow of the British Royal Society. That same year he published his first scientific paper on light and color in the “Philosophical Transactions of the Royal Society”. In his publication Newton claimed that light was composed of tiny colored particles. Robert Hooke, also a member of the Royal Society, had published his wave theory of light and color a few years earlier. Hooke responded to Newton’s publication by claiming that what was original in Newton’s paper (Newton’s particle theory of light) was wrong and what was correct in the paper had been stolen from him (Hooke). Newton fired back with his explosive temper trying his best to humiliate Hook in public. The angry argument between Newton and Hooke lasted for years.

Dutch physicist, mathematician and astronomer Christiaan Huygens agreed with Hooke believing that light was composed of waves. Huygens published his wave theory of light in 1690, complete with the supporting mathematics. Huygens claimed that reflection, refraction, and diffraction could all be explained by his theory. According to Huygens light waves had to travel through some medium, just like water waves traveling through water and sound waves traveling through air. So he proposed that light waves traveled through an aether that exists throughout the universe.

While Newton wanted fame and recognition of his work, he despised being criticized. The easiest way to avoid being criticized was to publish nothing. Consequently, Newton did not publish “Opticks” , the full account of his optical research, until 1704, the year after the 1703 death of Hooke.

Newton, Hooke, and Huygens did all agree that the speed of light was finite. In his 1687 treatise “The Mathematical Principles of Natural Philosophy” Newton states “For it is now certain from the phenomena of Jupiter’s satellites, confirmed by the observations of different astronomers, that light is propagated in succession and requires about seven or eight minutes to travel from the Sun to the Earth” a remarkably close estimate for the speed of light. Perhaps because of Newton's already high reputation, his particle theory of light reigned until wave theory was revived in the 19th century.

4.1.1 Isaac Newton (1643 – 1727)

Isaac Newton (Figure 1) was the leading mathematician of his generation. He laid the foundation for differential and integral calculus. His work on optics, laws of motion, and gravitation made him one of the world’s top scientists.

Figure 1 Isaac Newton (source: Isaac Newton Biography)

Newton is best known for his three laws of motion and his law of universal gravitation described in his book “The Mathematical Principles of Natural Philosophy”. For nearly 300 years this publication was the foundation of classical mechanics. Newton used his law of universal gravitation to prove Kepler’s laws of planetary motion, to account for ocean tides, the trajectories of comets, and the precession of the equinoxes.

Newton studied light extensively. He introduced the term color spectrum to describe the colors produced when shinning white light through a glass prism.

Newton shares credit with Gottfried Wilhelm Leibniz for development of differential and integral calculus. He also contributed to the study of power series, the binomial theory, and approximating the roots of functions. In addition, he formulated an empirical law of cooling and made the first theoretical calculation for the speed of sound

Newton’s life can be divided into three distinct periods. The first was his boyhood days from 1643 up to 1669. Second, his highly productive period as Lucasian professor of mathematics at Cambridge University from 1669 to 1696. The third period of his life, beginning in 1696, was nearly as long as the other two combined. During this period Newton served as Master of the Royal Mint, a highly paid government position in London. However, he did not resign his positions at Cambridge University until 1701. Newton made significant contributions at the Mint. He led the Mint through a difficult period of recoinage and actively pursued measures to prevent coinage counterfeiting. As Master of the Mint, coupled with the income from his mother’s estate, Newton became a very rich man.

4.1.2 Robert Hooke (1635 – 1703)

In addition to his work on light, Robert Hooke (Figure 2) discovered the law of elasticity (the mechanics of springs) which bears his name (Hooke’s Law). Hooke’s work on elasticity resulted in his development of the balance spring which enabled him to build a portable time piece, a watch. A bitter dispute between Hooke and Huygens arose over who had in fact invented the first spring operated watch. This dispute lasted for many years after the death of the two men. However, a note dated 23 June 1670 describing a demonstration of a balance-controlled watch before the Royal Society seems to indicate that Hooke was indeed the first to invent the watch.

Figure 2 Robert Hooke (source: britainunlimited.com)

Hooke was the first to suggest that matter expands when heated and suggested that air was made of small widely spaced particles. He also proposed that heat was caused by the rapid movement of particles within matter.

In a communications to the Royal Society in 1666 Hooke explained that: 1) Heavenly bodies are attracted to each other. 2) That all bodies having a simple motion will continue to move in a straight line unless continually deflected by some extraneous force, causing them to describe a circle, an ellipse, or some other curve. 3) That the attraction between bodies increases the closer they get. In the process of developing these ideas, Hooke came close to proving that gravity followed an inverse square law. These same concepts were, of course, being studied by Newton. The fierce dispute between Hooke and Newton over who actually discovered the inverse square law first, coupled with the dispute over the wave verses particle nature of light, sent Newton into such a rage that he removed from is own work nearly all references to Hooke.

In addition to physics, Hooke did work in biology including building a microscope used in his work, studied astronomy, and was an architect working as a surveyor for the City of London. He also built a reflecting telescope which he used to study Mars, the rotation of Jupiter, and the Orion constellation.

4.1.3 Christiaan Huygens (1629 – 1695)

Christiaan Huygens (Figure 3) was the second son of a well to do distinguished Dutch family which gave him access to the highest intellectual and social circles. Later in his life Huygens became a founding member of the French Academy of Sciences.

Figure 3 Christiaan Huygens (source: Wikipedia)

Huygens invented the first working pendulum clock in 1656 which he patented the following year. The pendulum clock was a breakthrough in timekeeping and became the most accurate type of clock for nearly 300 years. While Hooke is credited with inventing the watch, Huygens was probably the first to build a watch using a spiral balance spring, which he did in 1675.

Huygens was the first to derive the formula for centripetal force and the correct laws of elastic collision. In addition, he was a pioneer in the mathematics of probability.

Huygens built his own 50 power refracting telescope which he used to discover the rings of Saturn and the Saturn moon Titan, the largest of Saturn's moons.

However, Huygens is best know for his wave theory of light which he proposed in 1678 and published in his book "Treatise on Light" in 1690. His treatise is considered the first mathematical theory of light.

4.2 Young's Double Slit Experiment

Debate over the particle verses wave nature of light continued for nearly 100 years. Because of his reputation, Newton's particle theory of light prevailed throughout most of this period, despite Huygens's compelling arguments for light waves backed up by extensive mathematics. In 1803 English physician Thomas Young (1773 – 1829) showed conclusively that light is a wave phenomena.

Figure 4 Thomas Young (source: Wikipedia)

Young was the eldest of 10 children. By the age of fourteen he had learned Greek and Latin. His primary interest was in medicine. Young attended a number of medical universities receiving his doctorate in medicine from the University of Gottingen, Germany in 1796. A year later he inherited the estate of his grand-uncle, Richard Brocklesby, allowing him to became financially independent. In 1799 he became an established physician in London. Two years later, in 1801, Young was appointed professor of physics at the Royal Institution in Westminster, England. The Royal Institution was founded in 1799 by the leading British scientists at the time, including Henry Cavendish. Distinguished scientist who worked and taught at the institute include Humphry Davy, Michael Faraday, James Dewar, William Henry Bragg and his son William Lawrence Bragg, and many others. In 1803 Young resigned as professor of physics fearing that his duties at the institute would interfere with his medical practice.

During his life Young made notable contribution to the fields of vision, light, solid mechanics, energy, physiology, language, musical harmony, and Egyptology. He was in fact instrumental in deciphering Egyptian hieroglyphs.

Interestingly, Young published many of his first academic articles anonymously to protect his reputation as a physician.

In Young's own judgment, his most important achievement was establishing the wave theory of light. This was not easy to do. He had to overcome the century-old view that light is a particle as maintained by Newton in his 1704 paper "Opticks". Young developed a number of demonstrations supporting the wave theory of light, the most important being his double-slit experiment.

To perform this experiment, Young allowed sunlight to impinge on an opaque panel in which he had cut two small pin holes or slits, S_1 and S_2, as illustrated in Figure 5. The light diffracted as it passed through the two pin holes forming two spherical waves. Interference between the two waves caused the light striking a screen to vary in intensity, as illustrated by the waveform on the far right side of Figure 5. Bright regions occurred on the screen along those points where the two waves interfered constructively (peaks in the intensity waveform). Dark areas occurred at points where the two waves interfered destructively (valleys in the intensity waveform).

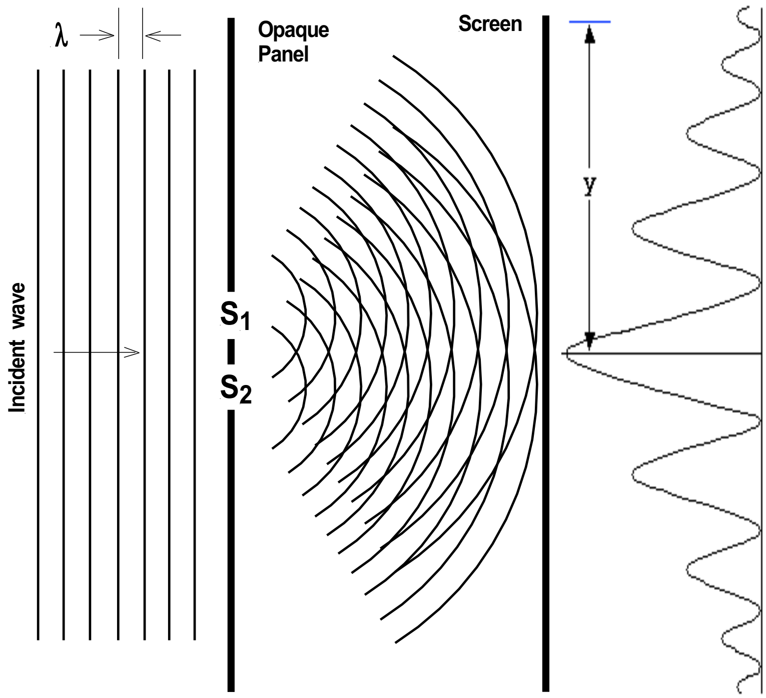


Figure 5 Young's Double-Slit Experimental Setup (source: author)

The size of the pin hole is critical. The pin hole must be approximately equal in size to the light's wavelength. If the wavelength is much larger than the pin hole, the light will simply be reflected back from the panel or absorbed by it. If the wavelength is much smaller than the pin hole, the light will pass through the panel without being diffracted.

The positions of peaks and valleys in the intensity waveform can be determined using Figure 6 where

- **d** is the distance between the two slits
- **y** is the vertical distance from the center of the screen to position P
- **θ** is the angle made by a line to P from a point midway between the slits
- **L** is the distance between the slits and the screen
- $\mathbf{r_1}$ is the distance (path length) from Slit-1 to point P on the screen
- $\mathbf{r_2}$ is the distance (path length) from Slit-2 to point P
- **δ** is the difference in path length between r_1 and r_2 .
- **m** is the order number of a peak
- $\mathbf{S_1}$ is the first slit
- $\mathbf{S_2}$ is the second slit

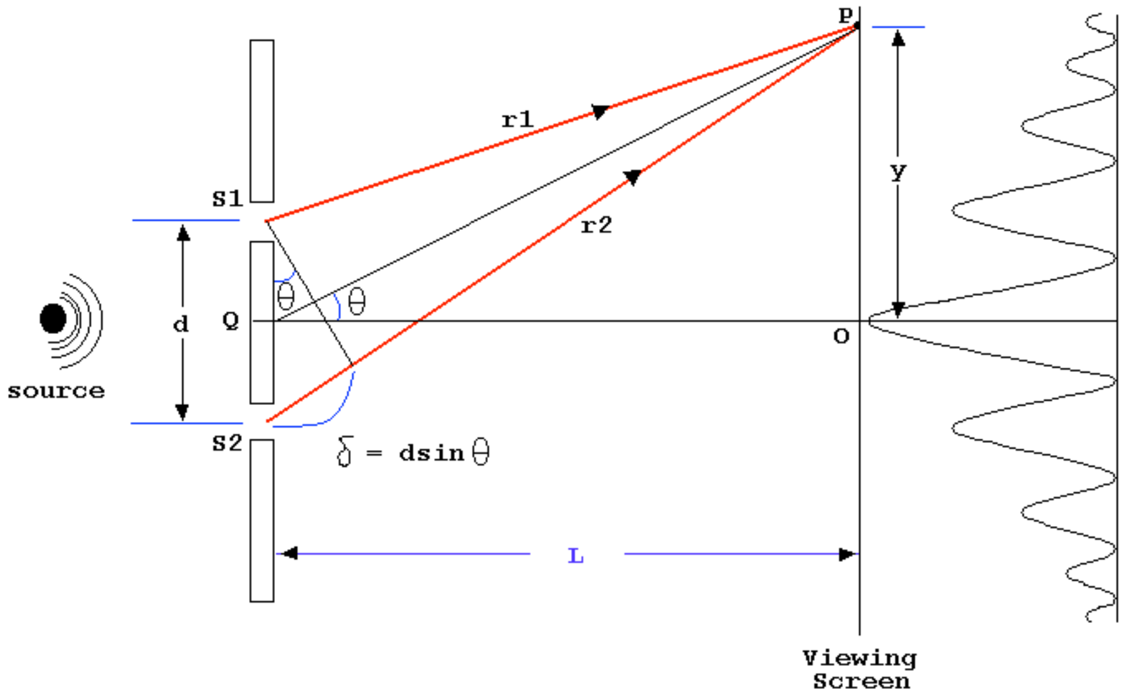


Figure 6 Positions of Interference Peaks (source: secure.math.ubc.ca)

Point O on the screen is a point midway between the two pin holes. The distance r_{2O} from pin hole S_2 to point O is equal to the distance r_{1O} from pin hole S_1 to point O . The two waves are in phase at point O since they both travel the same distance to reach that point. Consequently, they interfere constructively at point O forming the highest intensity point on the screen. At any other point P on the screen path length r_{2P} is not equal to path length r_{1P} creating a phase difference between ray r_2 and ray r_1. The magnitude of the phase difference depends on the difference δ in their path lengths.

$$\delta = r_2 - r_1 = d \sin\theta$$

Bright points occur on the screen when the path difference δ is equal to an integral multiple of the light's wavelength λ . That is, bright areas occur at

$$\delta = d \sin\theta = m\lambda$$

where $m = 0,\ \pm1,\ \pm2,\ \cdots$

At these points the interference between the two waves is constructive. The bright spot on the screen at location O corresponds to m = 0.

Black areas occur on the screen at points where the two waves are 180º out of phase. The interference at these locations is destructive, completely cancelling out the two waves. The path difference δ at these locations is an odd integer multiple of a half wavelength. That is

$$\delta = d \sin\theta = \left(n\tfrac{1}{2}\right)\lambda$$

where $n = \pm 1,\ \pm 3,\ \pm 5,\ \cdots$, or alternately

$$\delta = d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda$$

again where $m = 0,\ \pm 1,\ \pm 2,\ \cdots$

For small values of θ , the positions (y) of the bright and dark areas on the screen are

$$y_{bright} = \lambda L \frac{m}{d}$$

and

$$y_{dark} = \lambda L \frac{\left(m + \tfrac{1}{2}\right)}{d}$$

Interference patterns such as these are a characteristic of all types of waves. Particles, such as the particles of light envisioned by Newton, are incapable of creating interference patterns. The important result of this experiments is that light passing through the two pin holes **did** created interference patterns. Light must then be composed of waves.

4.3 Jean-Bernard-Leon Foucault (1819 – 1868)

In 1850 Jean Foucault (Figure 7) used a rotating mirror device to determine the relative speed of light in air versus water.

In his experiment, Foucault placed a tube of water in the path between a spinning mirror and a distant fixed mirror as illustrated in Figure 8.

Initially the tube was empty. Light reflected by the spinning mirror passed through the empty tube and was reflected by the fixed mirror at the tube's opposite end. The reflected light from the fixed mirror traveled back through the tube to the spinning mirror. However, the spinning mirror had only rotated slightly in the time taken for the light to pass through the tube and back again. Consequently, the returning beam of light was displaced slightly from the out-going light beam. The returning beam of light is labeled "Air" in Figure 8. The tube was then filled with water and the experiment repeated. This time the returning beam (labeled “Water” in Figure 8) was displaced more than when the tube was empty. That is, the mirror rotated further in the time for the light to pass through the tube of water and back again, indicating that the speed of light through water was slower than that through air.

Figure 7 Jean-Bernard-Leon Foucault (source: Wikipedia)

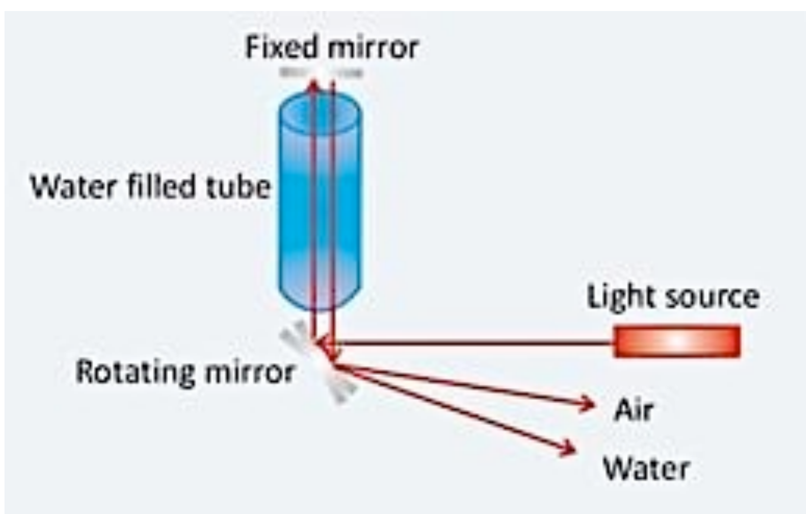
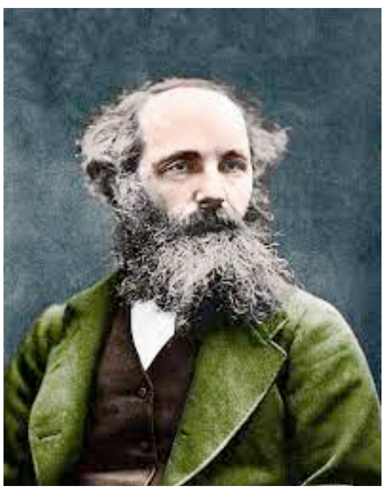


Figure 8 Foucault’s air vs water speed of light experiment
(Source: University of Virginia http://galileo.phys.virginia.edu)

Newton's particle theory of light predicted that light would travel faster through a more dense material. In contrast, the wave theory of light proposed by Christiaan Huygens and Robert Hooke predicted that light would travel slower through a dense material. For example, light would travel slower through water than air, just as Foucault discovered. Foucault's experiment showed that the wave theory of light proposed by Huygens, Hooke and Young was correct, putting an end to the long accepted particle theory of Newton.

4.4 James Clerk Maxwell (1831 – 1879)

From the early through mid 1800s it slowly became evident that electricity and magnetism, that were thought to be completely separate entities, were in fact strongly related. In 1865 James Maxwell (Figure 9) published a set of 20 equations that unified electricity and magnetics into a single theory of electromagnetics. Later Oliver Heaviside, using vector calculus, simplified Maxwell's equations into a set of 4 equations that we use today.



James Clerk Maxwell

Oliver Heaviside

Figure 9 Maxwell and Heaviside (source: Wikipedia)

In developing his set of equations, Maxwell integrated together laws of electricity, magnetism, and induction originally developed by Gauss, Faraday, and Ampere.

4.4.1 Maxwell's First Equation

Maxwell's first equation is Gauss's Law for Electricity. This law states that the total electric field E emanating from a closed surface A is equal to the total charge Q enclosed within that surface divided by the permittivity ε of the material in which the closed surface is located, as illustrated in Figure 10. The permittivity for free space is

$$\varepsilon_0 = 8.85\ x\ 10^{-12}\ \frac{(coulomb)^2}{Newton - \ meter^2}$$

In equation form, Maxwell's first equation is

$$\oint \overrightarrow{E} \cdot \overrightarrow{dA} = \frac{Q}{\varepsilon}$$

(See the chapter on Vector Analysis under Related Topics for a description of the vector notation used in the equations for this section).

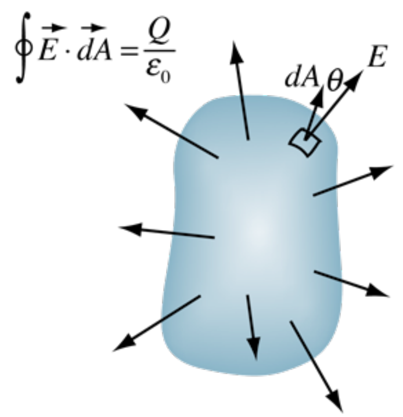


Figure 10 Gauss's Law for Electricity (source: hyperphysics)

In Figure 10, the angle θ is the angle between the electric field vector $\overrightarrow{E}$ penetrating through the surface and a vector of unit surface area $\overrightarrow{dA}$ which is perpendicular to the surface. The vector dot product is equal to

$$\vec{E} \cdot \overrightarrow{dA} = (E)(dA)\cos\theta$$

What is important about this equation is that the electric field emanating for a positive charged particle travels outward from the particle throughout the entire universe. The field from a negative particle travels inward toward the particle from everywhere in the universe as illustrated in Figure 11.

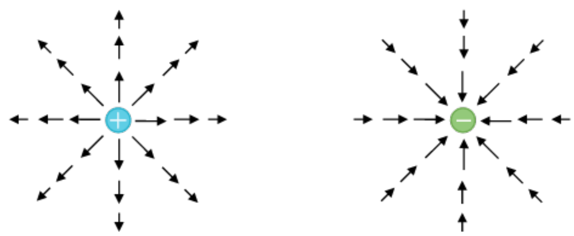


Figure 11 Field from an electric charge (source: www.khanacademy.org)

4.4.2 Maxwell's Second Equation

Maxwell's second equation is Gauss's Law for Magnetism. This equation states that the total magnetic field B emanating from a closed surface A is zero. In equation form

$$\oint \overrightarrow{B} \cdot \overrightarrow{dA} = 0$$

Unlike electric fields, magnetic fields do not begin or end on any point. Consequently, the number of magnetic field lines leaving a surface equals the number of magnetic field lines entering that surface as shown in Figure 12.

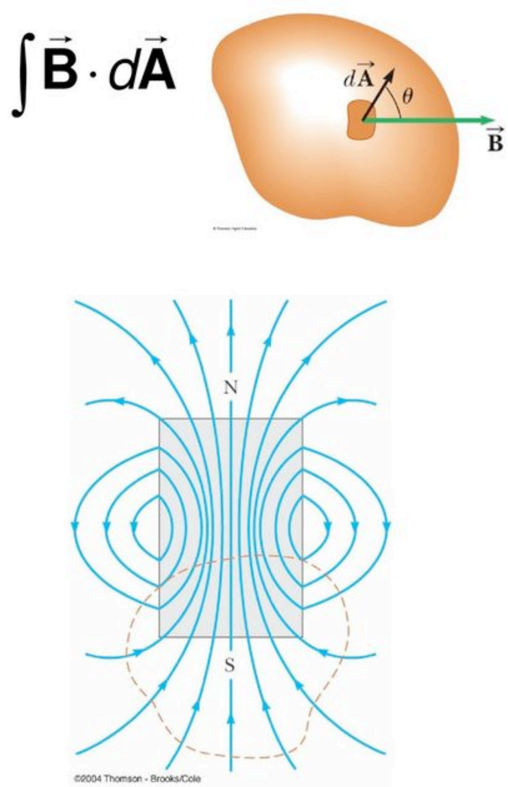


Figure 12 Gauss's Law for Magnetism (source: Thomson – Brooks/Cole)

Notice the closed surface formed by dashed lines in Figure 12. The number of magnetic field lines entering the closed surface from the sides equals the number of field lines leaving the top portion of the surface.

4.4.3 Maxwell's Third Equation

Maxwell's third equation is Faraday's Law of Induction. This equation states that a time varying magnetic field B creates an electric field E. In more formal terms, an electric field E is created along a closed path l when at least part of the path passes through a time varying magnetic flux ϕ_B. In equation form

$$\oint \vec{E} \cdot \vec{dl} = -\frac{d\phi_B}{dt}$$

For example, a bar magnetic moved in and out of a coil of wire creates an electrical current in the wire.

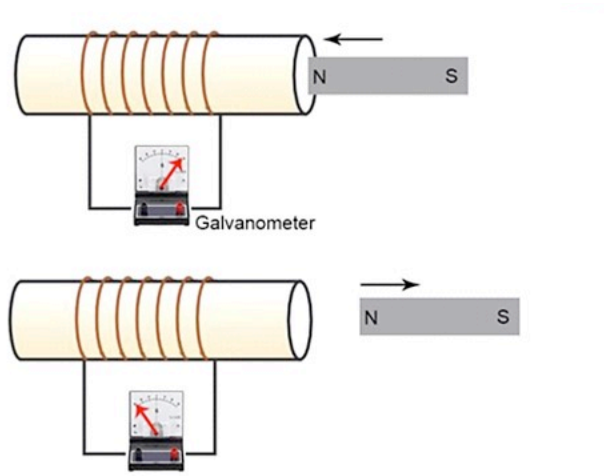


Figure 13 Faraday's Lay of Induction (source: Embib)

Notice in Figure 13 that the galvanometer points to the right when the magnet enters the coil of wire and to the left when the magnetic is removed. Thus the induced current in the coil changes direction. It is in one direction when the magnet enters the coil and in the opposite direction when the magnet is retracted. Also note that the coil of wire forms a closed path since the ends of the wire are connected together through the galvanometer.

Maxwell claimed that the inspiration for his unifying theory of electromagnetics was in fact the work done by Faraday. Faraday was one of the top experimental scientists in history, however, his mathematical skills were weak. Maxwell understood Faraday's work and was able to express it in mathematical terms.

4.4.4 Maxwell's Fourth Equation

Maxwell developed his fourth equation by expanding on Ampere's Law.

Ampere's Law states that the magnetic field B around a closed loop l is proportional to the current I passing through the loop, where the constant of proportionality is the permeability μ_0 of free space. In equation form Ampere's Law is

$$\oint \vec{B} \cdot \vec{dl} = \mu_0 I$$

which is illustrated in Figure 14.

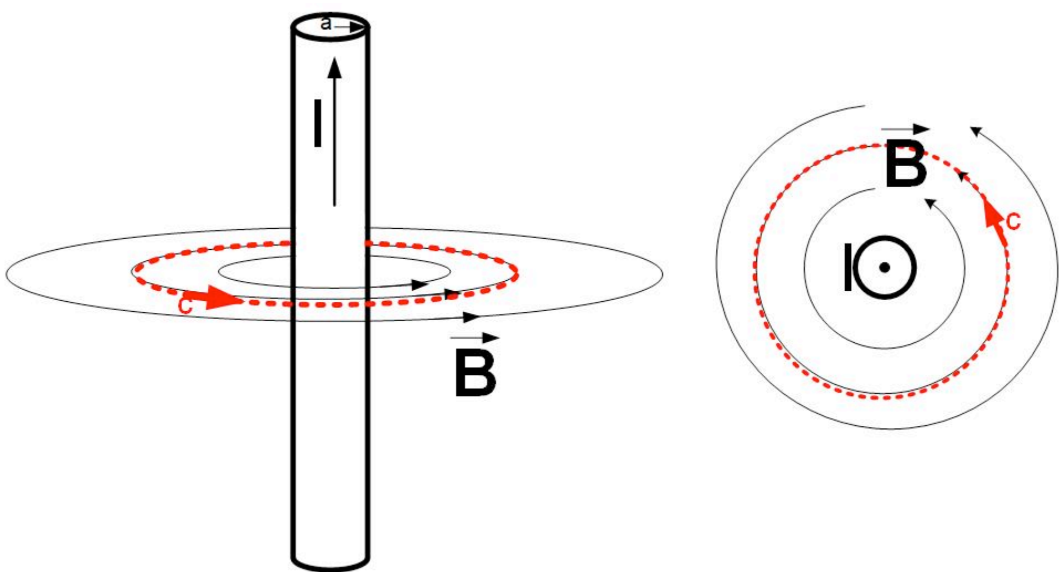


Figure 14 Ampere's Law (source: Ximera)

In Figure 14 the closed loop l is the red path c while the current I creating the magnetic field is flowing in the wire with radius "a".

The direction of the magnetic field around a current carrying wire is given by the "right hand rule" shown in Figure 15.

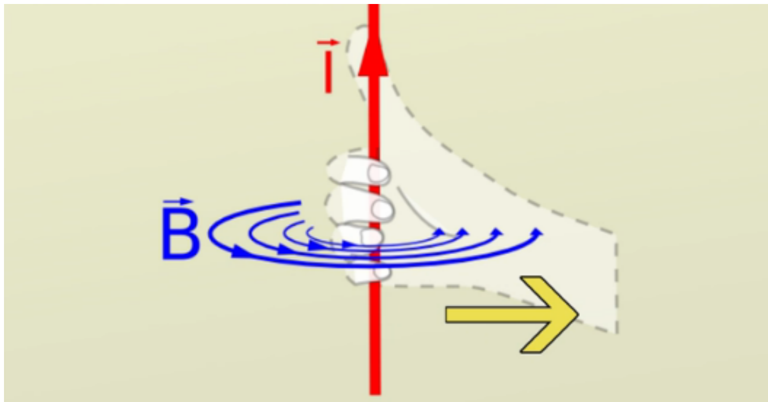


Figure 15 Right hand rule (source: Study.com)

With the thumb of your right hand pointed in the direction of the current flow I , the magnetic field produced by the current circulates around the wire in the direction of your curved fingers.

Notice that Faraday's Law of Induction and Ampere's Law are reciprocals. In Faraday's law an electric field is formed by a changing magnetic field. In Ampere's law a magnetic field is produced by a moving electrical charge (an electric current). A stationary magnetic field does not produce an electric field. Similarly, a static electric charge does not produce an magnetic field.

Maxwell had a problem with Ampere's Law. A capacitor placed in a wire loop prevents a current from flowing completely around the loop, thus preventing the formation of a magnetic field. However, a current must flow in the loop at some point to charge and discharge the capacitor, that is, to displace charge from one capacitor plate to the other. Maxwell added a second term to Ampere's law to account for the displacement current. The displacement current is

$$\varepsilon_0 \mu_0 \frac{\partial \phi_E}{dt}$$

Where

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

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

$\partial \phi_E$ = change in the electric field

Adding the displacement current to Ampere's Law produces Maxwell's fourth equation, specifically

$$\oint \vec{B} \cdot \vec{dl} = \mu_0 I + \varepsilon_0 \mu_0 \frac{\partial \phi_E}{dt}$$

4.4.5 Maxwell's Equations in Free Space

In free space, that is in a space completely devoid of all matter including the complete absence of charged particles Q, Maxwell's first equation must equal zero since Q = 0. Consequently, in free space

$$\oint \vec{E} \cdot \vec{dA} = \frac{Q}{\varepsilon} = 0$$

Maxwell's second equation is already equal to zero

$$\oint \vec{B} \cdot \vec{dA} = 0$$

In free space Maxwell's third equation remains as stated

$$\oint \vec{E} \cdot \vec{dl} = -\frac{d\phi_B}{dt}$$

In free space there are no moving electrical charges to carry a current I (I = 0), so Maxwell's fourth equation reduces to

$$\oint \vec{B} \cdot \vec{dl} = \varepsilon_0 \mu_0 \frac{\partial \phi_E}{dt}$$

Consequently, in free space Maxwell's equations reduce to just two equations

$$\oint \vec{E} \cdot \vec{dl} = -\frac{d\phi_B}{dt}$$

and

$$\oint \vec{B} \cdot \vec{dl} = \varepsilon_0 \mu_0 \frac{\partial \phi_E}{dt}$$

In free space

1. A changing electric field produces a changing magnetic field, and

2. A changing magnetic field produces a changing electric field.

Each induces the other in the complete emptiness of free space.

4.4.6 Electromagnetic Wave Equation

The wave equation for mechanical waves (sound waves, vibrating strings of musical instruments, etc.) had been known since the 18th century. This equation is

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}$$

where

y = wave amplitude

x = distance the wave has moved in the x direction

v = wave velocity

t = time

Maxwell's third and fourth equations can be put in the form of two wave equations, one for the electric field and one for the magnetic field. The wave equation for the electric field is

$$\frac{\partial^2 E}{\partial x^2} = \varepsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2} = \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2}$$

A similar perpendicular wave equation exists for the magnetic field. Consequently, an electromagnetic wave consists of two orthogonal waves, an electric wave and a magnetic wave, each perpendicular to the other and both perpendicular to the direction of wave travel as illustrated in Figure 16.

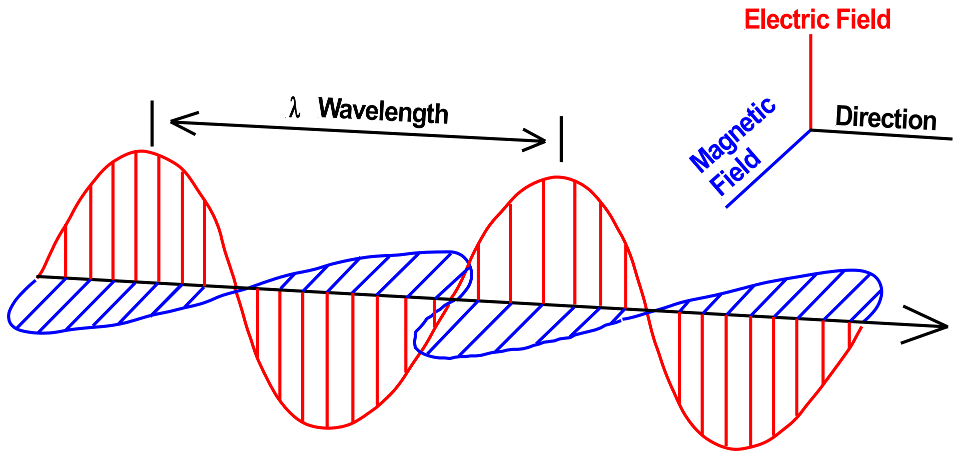


Figure 16 Electromagnetic Wave (source: author)

In the above wave equation the velocity term v is

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

Coulomb and Ampere had previously determined the values of ε_0 and μ_0 through careful static electric and magnetic measurements. Using their measurements Maxwell 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 lead 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 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.

4.5 Transverse vs Longitudinal Waves

As discussed above, electromagnetic waves are complex transverse waves in that they are composed of two inseparable orthogonal waves, an electric wave and a magnetic wave, as illustrated above in Figure 16. They are transverse waves in that they oscillate perpendicular to the direction of wave travel.

A rope tied to a tree and rapidly swung up and down produces a classic transverse wave as illustrated in Figure 17.

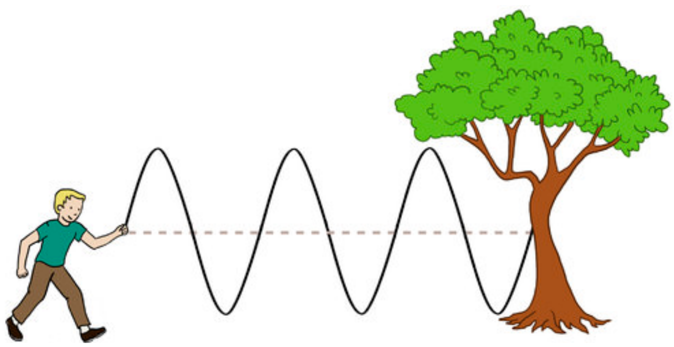


Figure 17 Using a rope to create a transverse wave (source: ck12.org)

In contrast, sound waves are longitudinal waves in that the wave oscillates back and forth in the direction of travel as illustrated in Figure 18.

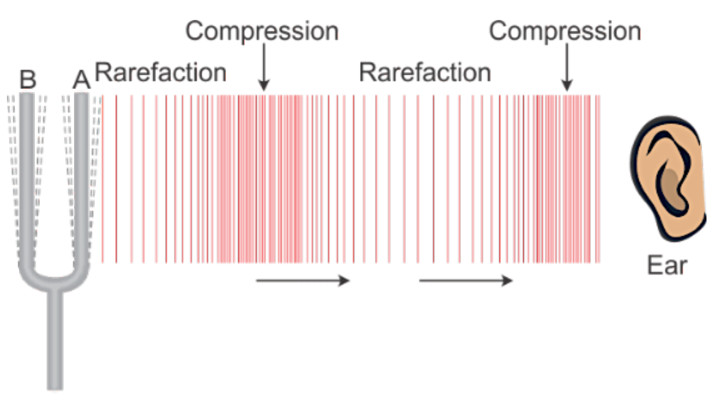


Figure 18 Longitudinal sound wave (source: tuttee.co)

In this chapter we are only concerned with transverse electromagnetic waves.

Of the two orthogonal waves, it is the electric wave that interacts most readily with the wave's surroundings. For example, it is primarily the electric field interacting with the ionosphere that causes an HF electromagnetic wave (a radio wave) to bend back to Earth. For this reason we concentrate most on the electric field in discussing the characteristics of electromagnetic waves, remembering all along that the magnetic field is always present.

4.6 Transverse Wave Equations

Figure 19 shows a transverse sinusoidal wave "frozen" in time. The equation for this wave is

$$y = y_m \sin\left[2\pi \frac{x}{\lambda}\right]$$

where

x = distance from the origin 0 in the horizontal direction

y = the displacement of the wave from the x axis

y_m = the maximum displacement, or wave amplitude

λ = the wavelength of the wave

$\left[2\pi \frac{x}{\lambda}\right]$ = wave angle in radians at a given position x along the x-axis

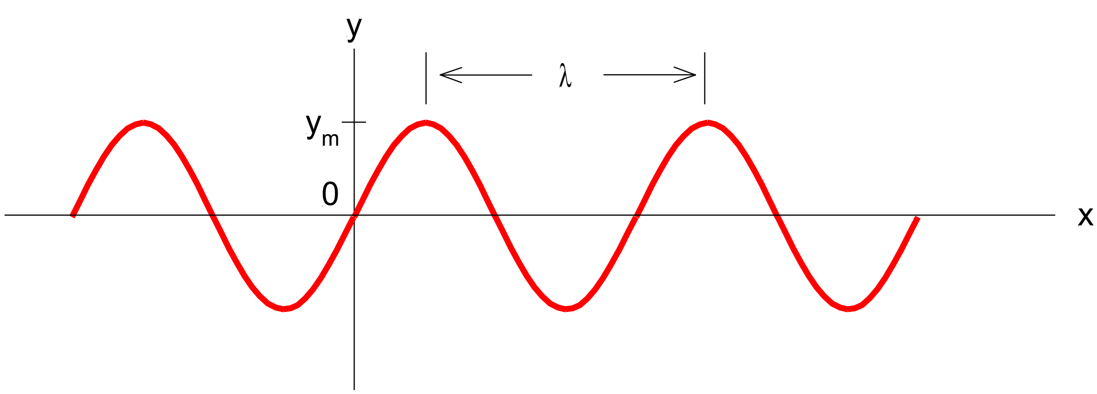


Figure 19 A periodic wave frozen in time.

Wave Amplitude y_m: Wave amplitude is the maximum displacement of the wave from the x-axis.

Wavelength λ : Wavelength is the distance in the x direction from any arbitrary point on the wave to the next corresponding point. For convenience in Figure 19 the wavelength is measured from a wave crest to the next wave crest.

At the origin (x = 0) the wave displacement is

$$y = y_m \sin\left[2\pi \frac{x}{\lambda}\right] = y_m \sin\left[2\pi \frac{0}{\lambda}\right] = y_m \sin 0 = 0$$

as shown in Figure 19. At a distance along the x-axis of $x = \frac{1}{4}\lambda$ the wave displacement is

$$y = y_m \sin\left[2\pi \frac{x}{\lambda}\right] = y_m \sin\left[2\pi \frac{1\lambda}{4\lambda}\right] = y_m \sin\frac{1}{2}\pi = y_m \cdot 1 = y_m$$

again as shown in Figure 19.

Phase Velocity: In Figure 20 the wave moves to the right at a velocity of v. To avoid deforming the wave, every point on the wave must move to the right at the same velocity. This velocity is known as the Phase Velocity.

$$\boldsymbol{phase\ velocity\ \ v} = \frac{distance\ traveled\ (x)}{t\ \ seconds}$$

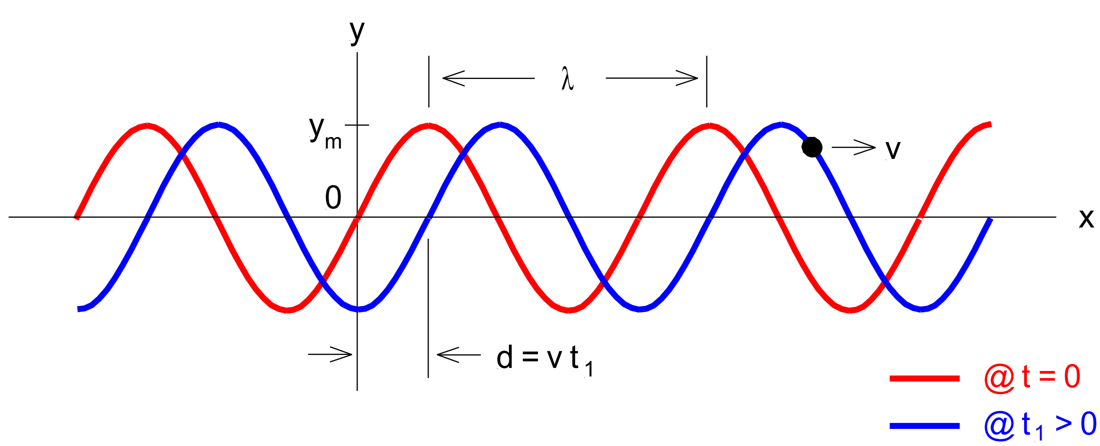


Figure 20 Traveling Wave

The phase velocity v of an electromagnetic wave in free space is the speed of light c which is

$$c = 299{,}792{,}458 \text{ m / s}$$

Traveling Wave: The equation for a sinusoidal wave moving in the x direction is

$$y = y_m \sin \frac{2\pi}{\lambda}[x - vt]$$

where the wave is propagating to the right at a phase velocity v. Note that velocity v multiplied by the time of travel t is equal to a distance. Suppose that at time t_1 in Figure 20 and at a position along the x-axis of

$$x = \left[\frac{1}{4}\lambda\right]$$

$$d = vt_1 = \left[\frac{1}{4}\lambda\right]$$

then,

$$y = y_m \sin \frac{2\pi}{\lambda}[x - vt] = y_m \sin \frac{2\pi}{\lambda}\left[\frac{\lambda}{4} - \frac{\lambda}{4}\right] = y_m \sin 0 = 0$$

as depicted by the blue wave form in Figure 20. Thus the red wave represents a particular traveling wave at time t = 0 (that is at t_0). The blue wave represents the position of the same traveling wave sometime later at $t_1 > 0$. As shown in the figure, the traveling wave has moved to the right a distance of $d = vt_1$ in the time interval $t_1 - t_0$.

For a wave propagating to the left, the equation becomes

$$y = y_m \sin \frac{2\pi}{\lambda}[x + vt]$$

That is, the term vt is added to x instead of being subtracted from it.

Phase Angle ϕ: The above equations assume that the displacement y is zero at the position x = 0 and time t = 0. This of course does not need to be the case. The general equation for a sinusoidal wave moving to the right is

$$y = y_m \sin \frac{2\pi}{\lambda}[x - vt - \phi]$$

where ϕ is the phase angle shown in Figure 21.

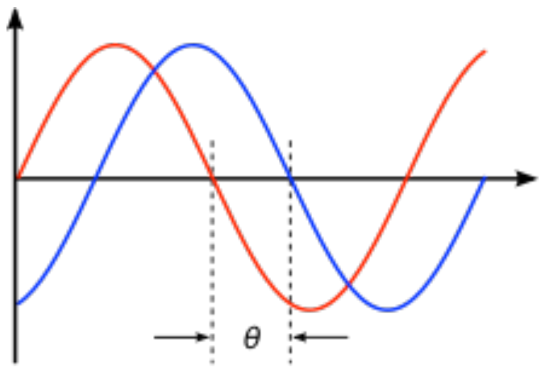


Figure 21 Phase Angle (source: Wikipedia)

If $\phi = +90° = +\frac{1}{4}\lambda$ radians, then the displacement y at x = 0 and t = 0 is

$$y = y_m \sin\frac{2\pi}{\lambda}[x - vt - \phi] = y_m \sin\frac{2\pi}{\lambda}\left[0 - 0 - \left(\frac{1}{4}\lambda\right)\right] = -y_m \sin\frac{1}{2}\pi = -y_m$$

which is the amplitude of the blue waveform in Figure 21 at position x = 0 and t = 0. That is, the blue curve is displaced to the right of the red curve by an phase angle of

$$\phi = +90° = +\frac{1}{4}\lambda\,.$$

Wave Period T: The wave period T is the time required for a wave to travel a distance of one wavelength λ at a velocity of v. Thus

$$\boldsymbol{period\ T} = \frac{\lambda}{v}\ seconds$$

Wave Frequency f : Wave frequency is the number of wave cycles that occur in one second. One wave cycle is the same as one wavelength shown in Figure 19. That is, the wave amplitude oscillates from zero, to its maximum positive amplitude, back to zero, to its maximum negative amplitude, and back to zero again in one cycle. Thus

$$\boldsymbol{frequency\ f} = \frac{1}{T}\ hertz$$

Substituting in the expression for period T and rearranging terms gives the equation for wavelength with respect to frequency and phase velocity. Thus:

$$wavelength\ \lambda = vT = v\left(\frac{1}{f}\right) = \frac{v}{f}$$

giving

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

For example, the velocity v of an electromagnetic wave in air is essentially the same as the speed of light in free space which is $c \cong 300{,}000{,}000\ meters/sec$. The length of a full wave antenna operating in the 40 meter band at a frequency of f = 7.2 MHz is

$$\lambda = \frac{c}{f} = \frac{300{,}000{,}000}{7{,}200{,}000} = \frac{300\ Mm/s}{7.2\ MHz} = 41.67\ meters = 136.67\ feet$$

The length of a typical half wavelength dipole antenna at this frequency is

$$\frac{1}{2}\lambda = \frac{1}{2}\frac{c}{f} = \frac{1}{2}\cdot\frac{300\ Mm/s}{7.2\ MHz} = 68.34\ feet$$

Angular Frequency: It is often convenient to express frequency in terms of angular frequency (ω) with units of radians per second. In this case

$$\boldsymbol{angular\ frequency\ \omega} = \frac{2\pi}{T} = 2\pi f$$

Finally, a wave number is often specified for a sinusoidal wave in which

$$\boldsymbol{wave\ number\ \ k} = \frac{2\pi}{\lambda}$$

Using these definitions the equation for a traveling sinusoidal wave moving to the right is

$$y = y_m \sin[kx - \omega t - \phi]$$

4.7 Wavefronts

In discussing waves the concept of wavefronts is very important. A wavefront is defined as a surface perpendicular to the direction of wave travel in which the phase angle of the wave is the same at all points on that surface. A wave consists of an infinite number of wavefronts, one for each phase angle and fraction thereof. Thus in discussing a wave we generally talk about a specific wavefront. For example, Figure 22 illustrates a number of wavefronts for a transverse sinusoidal wave. Notice that pink wavefronts are located along the wave such that the wave phase angle is 90^o at each pink wavefront. In addition, the pink wavefronts travel along with the wave so that at any point in time the phase angle of the wave at a pink surface is always 90^o. Similarly, the greenish wavefronts travel with the wave so that at any point in time the phase angle of the wave at a greenish surface is always 270^o.

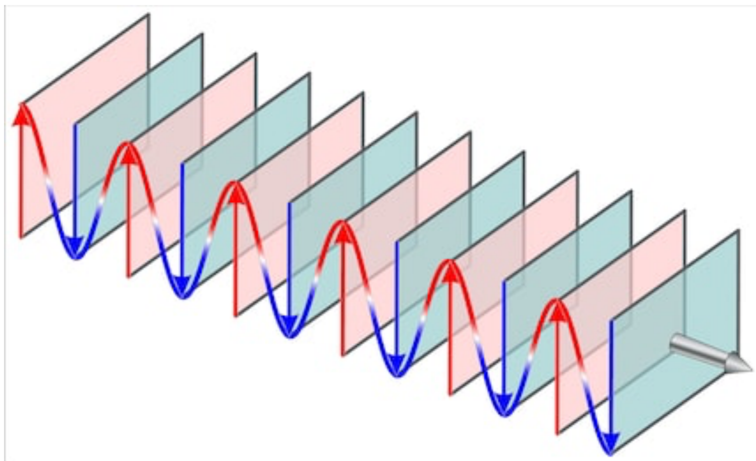


Figure 22 Wavefronts (source: shutterstock.com)

Wavefronts can have many different shapes. A plane wave travels in a single direction and has a wavefront which is a flat surface perpendicular to the direction of wave motion, as illustrated in Figure 22.

Spherical waves propagate radially outward in all directions from a single point source. In this case wavefronts are concentric spheres centered on the point source as illustrated in Figure 23. For example, an isotropic antenna on a spacecraft in deep space radiates spherically outward producing wavefronts that are spheres. (An isotropic antenna radiates equally in all directions). The intensity I of the radiated signal at a location R some distance from the antenna is

$$I = \frac{P}{4\pi R^2}$$

where P is the antenna's transmitted power. Notice that intensity follows the inverse square law for spherical wavefronts. In the above equation, the distance R is sufficiently far from the spacecraft that the spacecraft and its associated antenna appear as a point source.

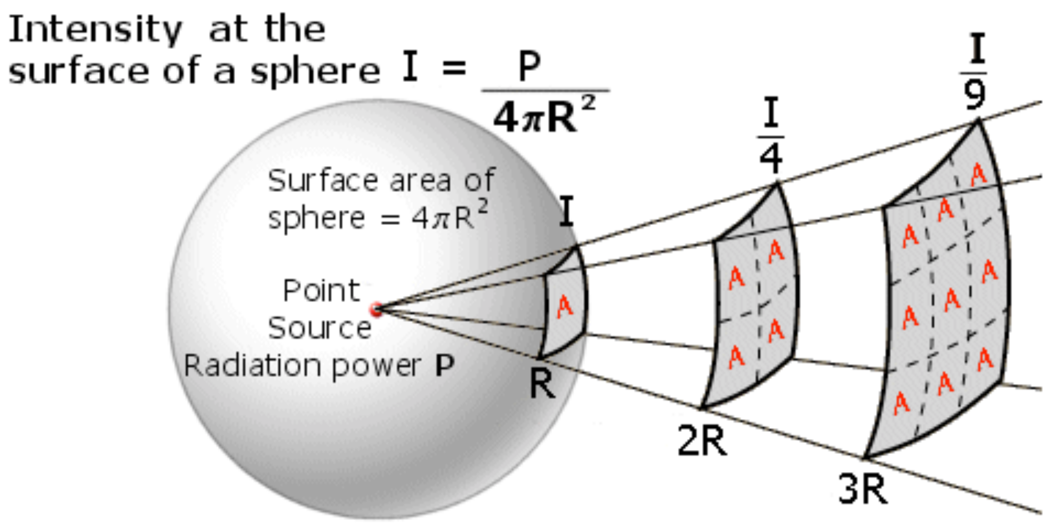


Figure 23 Spherical wavefronts (source: Blaze Labs Research)

The spherical radiation pattern of that same isotropic antenna placed on top of a 50 foot tower here on Earth would be severely distorted by its surroundings including the ground, near-by buildings, bending in the ionosphere, etc.

4.8 Wave Interference

Waves of the same frequency and phase velocity that are moving through the same medium at the same time interfere with each other. For example, two radio signals of the same frequency, both traveling at the speed of light, arriving at a receiving antenna will interfere with one another, potentially causing receiver fading.

Two interfering waves produce a resulting wave as illustrated in Figure 24. In this figure the two interfering waves are the green and blue waves. Since the two waves are the same frequency and at the same phase angle, they **interfere constructively** creating the resulting red wave. The red wave is also at the same frequency and phase angle, but has an amplitude that is the sum of the two original waves.

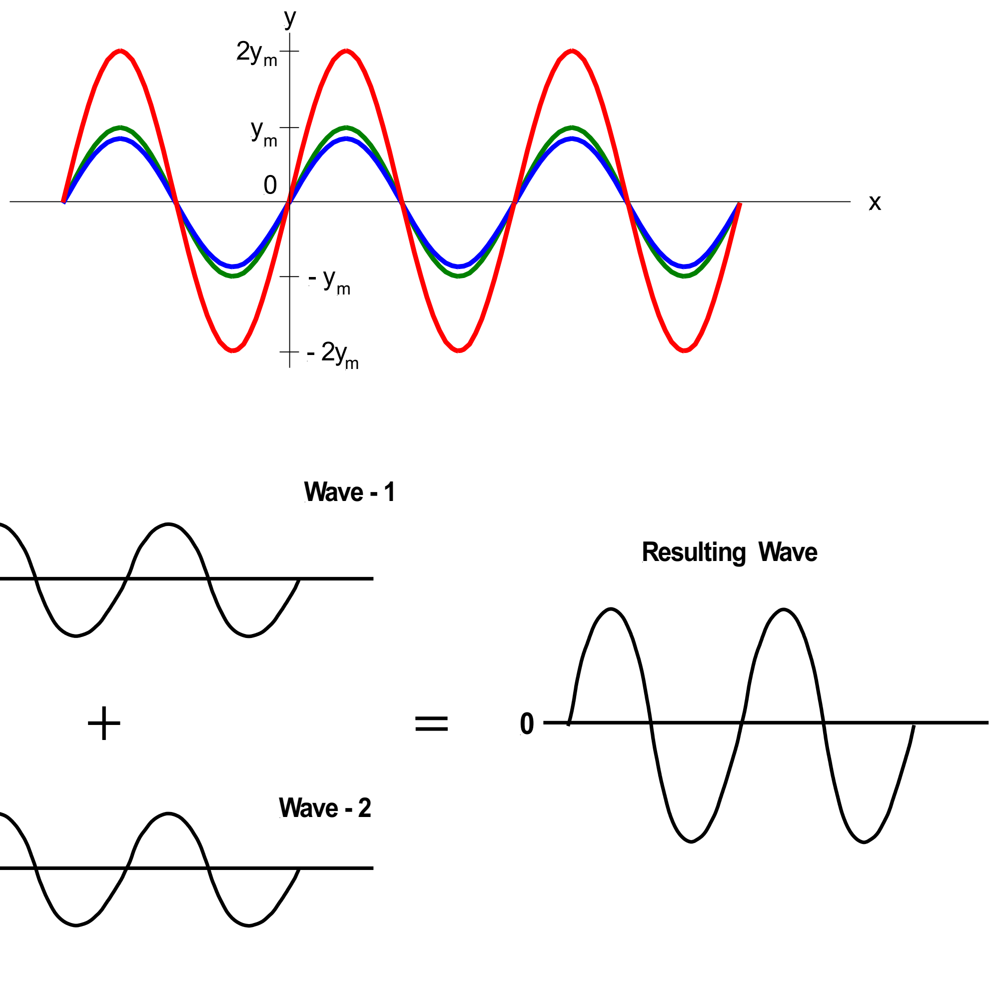
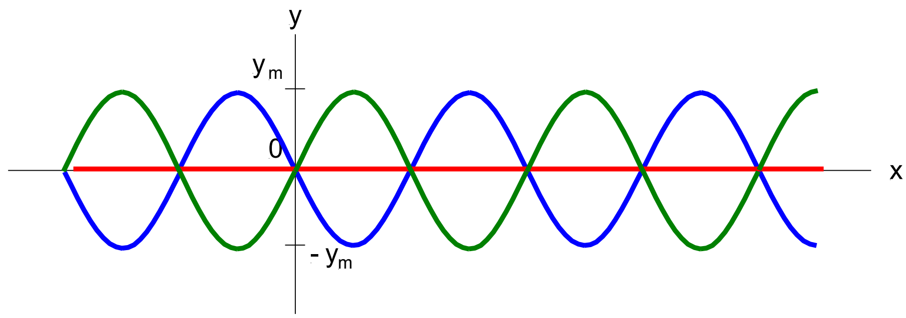


Figure 24 Constructive interference (source: author)

However, if the green and blue waves in Figure 25 are the same amplitude and frequency but 180º degrees out of phase, that is

$$\theta = 180° = \pi \, radians$$

then the amplitudes of the two waves will completely cancel producing no resulting wave. Figure 25 is an example of **destructive interference**.



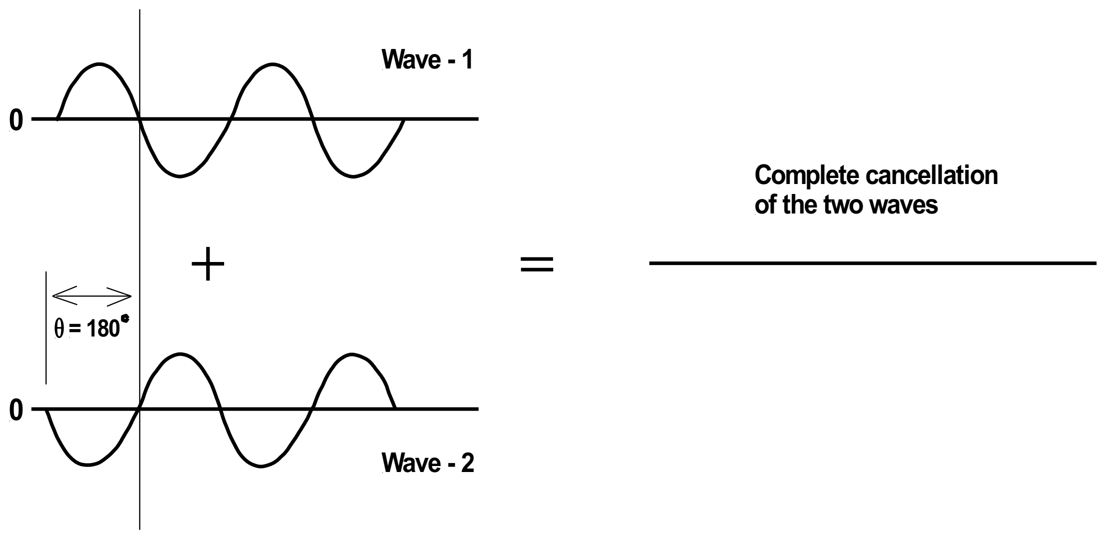


Figure 25 Destructive Interference (source: author)

The amplitude of the resulting red wave will be between these two extremes $(0 \leq y \leq 2y_m)$ if

$$amplitude\ green\ wave\ \neq amplitude\ blue\ wave$$

and the phase differences θ between the green and blue waves is

$$0 \leq \theta \leq 180°$$

The phase angle θ of the resulting red wave will also be different from that of green and blue waves as shown in Figure 26.

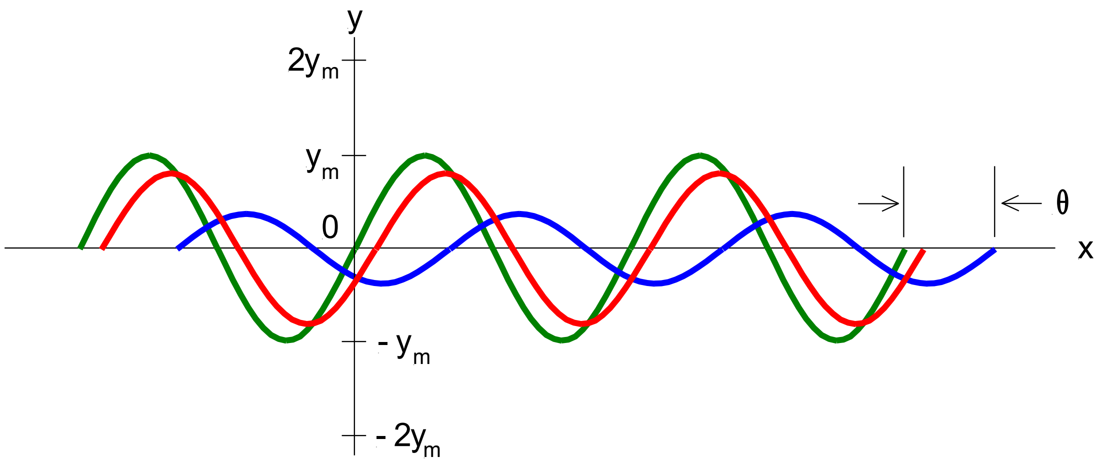


Figure 26 Two waves (blue & green) same freq. and velocity BUT different amplitudes and phase (source: author)

In general, the interference between any number of sinusoidal waves of the same frequency and phase velocity, but different amplitudes and phase angles, will produce a resulting wave which also has the same frequency and phase velocity but a different amplitude and phase angle.

The table below shows the amplitude of the resulting wave when two waves, each with an amplitude of 1, interfere. In the table the phase difference between the two waves varies from 0^0 (in phase) to 180^o (out of phase).

Phase difference between two original waves	**Amplitude of the resulting wave**	**Type of interference**
0^o	2.000	Constructive
45^o	1.848	Constructive
90^o	1.414	Constructive
118^o	0.999	Transition
135^o	0.765	Destructive
180^o	0.000	Destructive

4.8.1 Multipath Inference

It is important to note that a phase difference can develop between two radio signals if they begin as identical waves at the radio transmitter but follow different paths to the receiving destination. For example, in Figure 27 one signals travels to the destination in one hop (1F) while the other takes two hops (2F) to reach the receiver. This is known as multipath interference. If the path difference is an integer number of wavelengths

$$0, \quad \lambda, \quad 2\lambda, \quad 3\lambda, \quad etc$$

then the two waves will add constructively at the destination, that is the resulting amplitude of the received signal will be twice that of the original signal. However if the path difference is an odd number of half wavelengths

$$\frac{1}{2}\lambda, \quad \frac{3}{2}\lambda, \quad \frac{5}{2}\lambda, \quad etc$$

then the two waves will add destructively, completely canceling each other out resulting in no received signal at the destination. To solve this problem it is necessary to select an operating frequency that will only support one path to the desired destination.

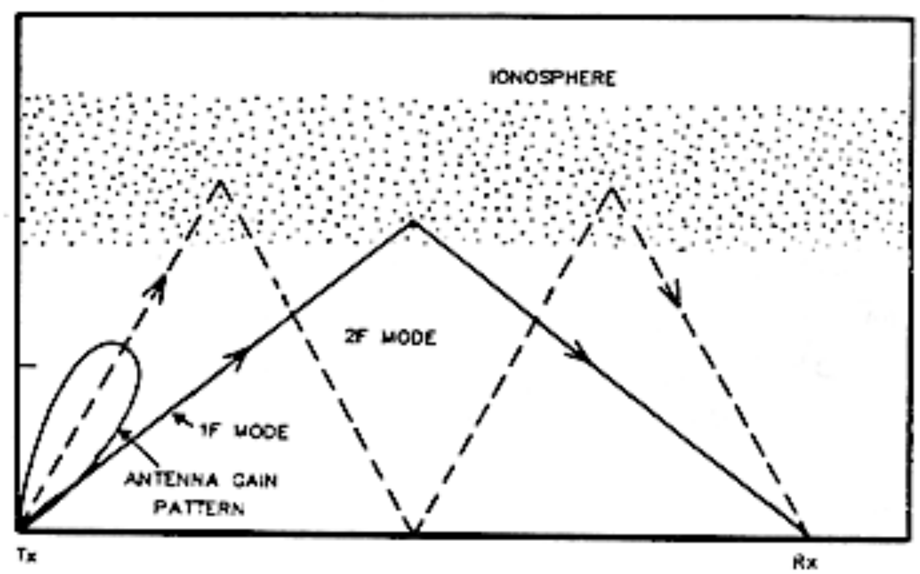


Figure 27 Multipath Problem (source: McNamara)

For example, the multipath signals shown in Figure 27 could result from a 40 meter transmission originating in Los Angeles (the Tx site) and received in Denver, Colorado (the Rx site). The two paths are the solid line labeled 1F, meaning 1 hop through the ionosphere F layer, and the higher dashed line 2F signal (2 hops through the ionosphere). Changing to a 20 meter frequency in many cases will solve the multi-path problem. On 20 meters the higher path 2F signal will likely penetrate the ionosphere and be lost to outer space. This will leave only the 1F signal which is successfully received in Denver without interference. The general guidance for illuminating multipath problems is to operate at the highest frequency that will allow communications from one location to another. In the example above, operation on 20 meters at 14.2 MHz instead of on 40 meters at 7.2 MHz.

4.8.2 Standing Waves

Suppose that two waves y_1 and y_2 of the same angular frequency ω, amplitude y_m , and speed are traveling in opposite directions. This can happen, for instance, if wave y_1 is reflected from a smooth metal surface perpendicular to the wave's direction of travel producing the reflected wave y_2. The equations for the two waves are:

$$y_1 = y_m \sin(kx - \omega t)$$

and

$$y_2 = y_m \sin(kx + \omega t)$$

The only difference in the two equations is the sign preceding the angular frequency ω resulting from the fact that the waves are traveling in opposite directions.

As before, the two waves interfere producing a resulting wave that is the sum of wave 1 and wave 2. The wave equation of the resulting wave is

$$y = y_1 + y_2 = y_m \sin(kx - \omega t) + y_m \sin(kx + \omega t)$$

Using trigonometric identities

$$\boldsymbol{y = (2y_m \sin kx) \cos \omega t}$$

This is the equation for a standing wave. The amplitude of the resulting wave is

$$y_x = 2y_m \sin kx$$

which is no longer constant but varies along the x-axis as illustrated in Figure 28. Maximum amplitude, $2y_m$, occurs at positions where

$$kx = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, etc$$

or

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, etc$$

remembering that

$$k = \frac{2\pi}{\lambda}$$

The amplitude of oscillation is zero at positions where

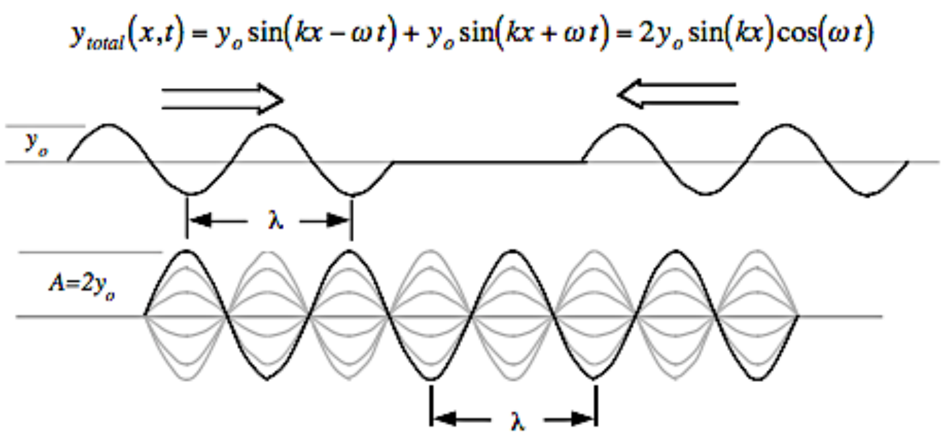
$$kx = \pi\,, 2\pi\,, 3\pi\,, etc$$

or

$$x = \frac{\lambda}{2}\,, \lambda\,, \frac{3\lambda}{2}\,, 2\lambda\,, etc$$

Locations along the wave where the amplitude is zero are called Nodes as shown in Figure 29. Positions where the amplitude is maximum are Antinodes.

Antinodes are spaced one half wavelength apart as illustrated in Figure 29. Nodes also occur every half wavelength with each node positioned one quarter wavelength from the antinodes on each side of it.



| Point A

Figure 28 Standing Wave (source: Quora)

Take a closer look at Figure 28. The figure shows a heavy black wave and a number of light gray waves. The standing wave is frozen in place. It can not move to the right or left. But it does oscillate up and down. The amplitude of the wave at Point A is a maximum in the negative direction at the time shown by the heavy black trace. A moment later the wave amplitude at Point A decreases to that of the first gray wave. It then decreases to the second gray wave and after that to zero. It then increases to the first gray wave in the positive direction, then to the next gray wave and finally to its maximum positive amplitude. It then changes direction, decreasing to zero after which it begins increasing again, this time in the negative direction. The wave continues oscillating back and forth in this manner forever. However, the wave amplitude is **always** zero at each node.

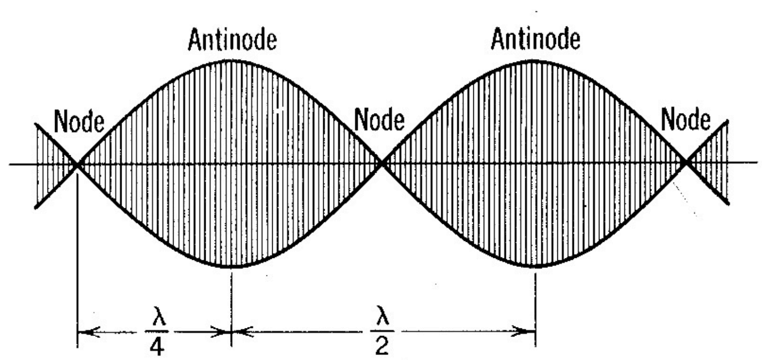


Figure 29 Standing Wave Envelope (source: Resnick & Halliday)

This is extremely important because it means that energy can not be transported along the wave in either direction. Energy can not flow past nodal points where the wave amplitude is always zero. Consequently, energy remains "standing" or frozen in the wave, oscillating up and down but not traveling to the right or left.

Figure 30 shows the electrical current distribution in a resonant half wave dipole antenna. The current on the antenna first moves to the left (arrow-1), then to the right (arrow-2), then back to the left (arrow-3) over and over again forming a standing current wave on the antenna. Nodes occur at both ends of the antenna with an antinode in the middle. Nodes have to occur at the ends since the total current (the current moving to the left plus the current moving to the right) must be zero at each end.

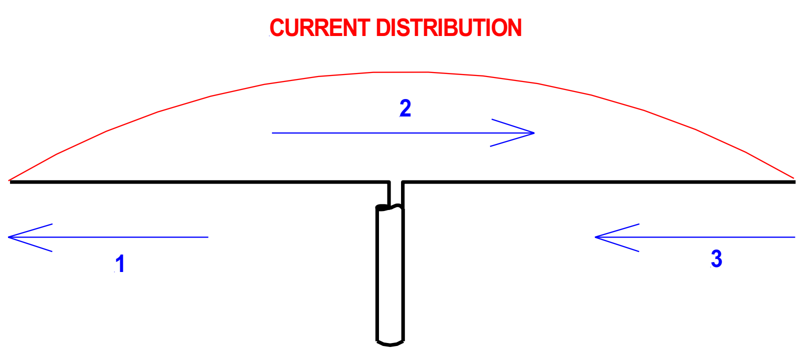


Figure 30, Current distribution on a half wave dipole antenna (source: author)

Figure 31 shows the oscillating motion of the antenna standing wave. As explained for Figure 28 above, the standing wave can not move to the right or left. It can only oscillate up and down from a maximum positive amplitude, to zero, to a maximum negative amplitude and back over and over again. The oscillating action "pumps energy" away from the antenna in the form of a radiated electromagnetic wave.

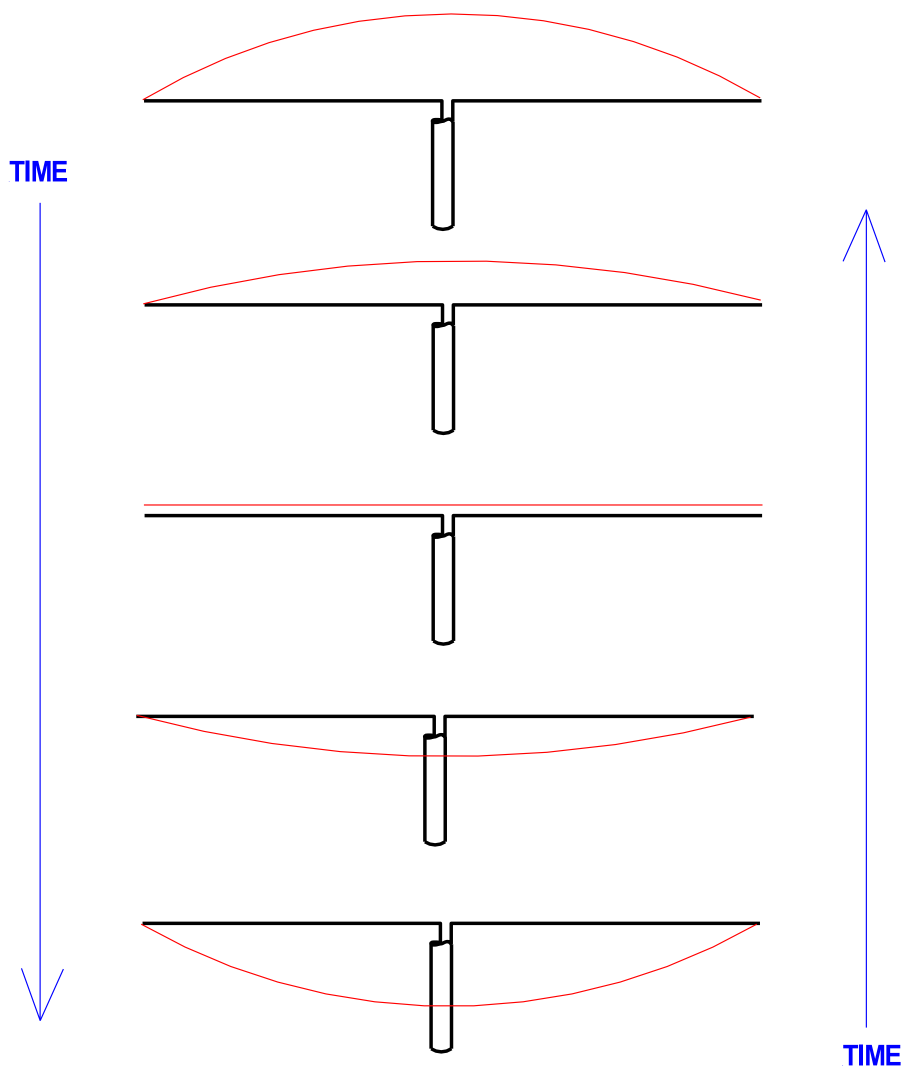


Figure 31 Oscillating antenna standing wave (source: author)

In 1888 Heinrich Hertz used standing waves to prove that electromagnetic waves, predicted 20 years earlier by James Maxwell, actually existed and that their speeds were equal to the speed of light. Hertz did this by constructing a crude spark gap transmitter shown on the left side of Figure 32. He detected the electromagnetic wave radiated from the transmitter by observing the spark that occurred across the gap of a receiving ring, proving that electromagnetic waves did in fact exist. The electromagnetic wave from the transmitter was reflected by a metal plate which Hertz placed at the far end of his laboratory. He positioned his equipment so that the transmitted and reflected waves would interfere producing a standing wave. He moved his receiving antenna back and forth between the transmitter and reflecting plate until he found a location that produced the largest spark across the gap in his receiving loop. He then moved his antenna again until he found the next position where the spark was large, thus locating two antinodes in the standing wave. Hertz knew the physical distance between two antinodes was a half wavelength. He then calculated the frequency of his transmitted signal from the electrical components used to construct the transmitter. Knowing the frequency f and wavelength λ, Hertz was able to calculate the velocity of the original and reflected waves from

$$v = \lambda f$$

a velocity which turned out to be the speed of light as predicted by Maxwell.

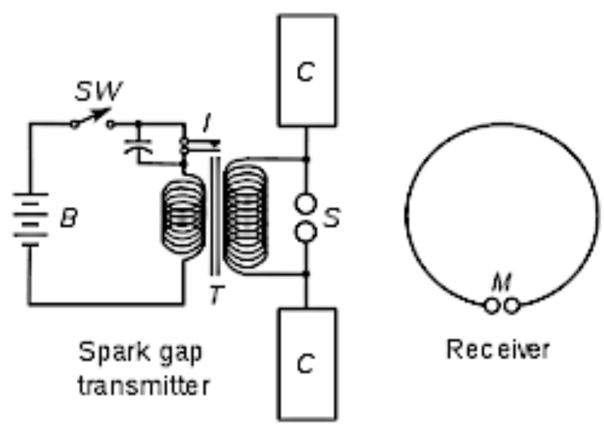


Figure 32 Hertz equipment used to detect electromagnetic waves (source: Wikipedia)

4.9 Modulated Signals and Group Velocity

A single continuous sinusoidal signal does not in itself carry any information. In order to transmit a message the signal must be modified by the information being sent. This is typically done by varying the signal's amplitude, frequency, or phase angle in a process called modulation or mixing. The mixing process produces two new signals in addition to the original signal (the carrier) and modulating signal which contains the information to be sent, for example audio from a microphone. The frequencies of the two new signals are the sum and difference of the carrier and modulating frequencies.

For example, suppose the frequency of the carrier signal f_c is 7.2 MHz and the modulating signal f_m is a variable amplitude audio signal ranging in frequency from 300 to 6,000 Hz. After the modulation process, the frequencies of the four resulting signals are

$$f_c \, , f_{c+m} \, , f_{c-m} \, , and \; f_m$$

The modulating signal is generally filtered out in the transmitting process since its frequency f_m is so low in comparison to the other three frequencies. Only the carrier frequency f_c remains constant. The upper sideband frequency f_{c+m} varies between

$$7200.300 \leq f_{c+m} \leq 7206.000 \; KHz$$

as the frequency of the modulating audio signal varies. Similarly the lower sideband frequency f_{c-m} varies between

$$7194.000 \leq f_{c+m} \leq 7199.7 \; KHz$$

The bundle consisting of these three signals is required to transmit information, an audio signal in this case, from one place to another.

Each of the three signals has its own phase velocity $(v_{c-m} \, , v_c, and \; v_{c+m})$. The bundle of these three signals, which carries the audio information, also has its own velocity known as the **group velocity** v_g. In outer space the four velocities are the same each equaling the speed of light in free space, that is c = 299,792.458 km/sec. The four velocities are also nearly the same when traveling through air since the speed of light in air is approximately equal to c. In other media the velocities are not the same because of frequency dispersion. Frequency dispersion means that the velocity of a signal through a medium varies with the signal's frequency. Consequent, the velocities of the four signals $\left(v_{c-m} \, , v_c, v_{c+m} \; and \; v_g\right)$ are all slightly different.

The group velocity is always equal to or less than c. The phase velocities of the three signals are generally equal to or less than c, but they don't have to be.

In metals and in plasmas such as the ionosphere, phase velocities can actually be faster 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 c. However, the information carrying group velocity of a modulated signal is prevented by Einstein's theory of special relativity from traveling faster than c. Likewise, all material objects are prevented from traveling faster than the speed of light.

This phenomena can actually be observed by watching waves at the beach. In Figure 33 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.

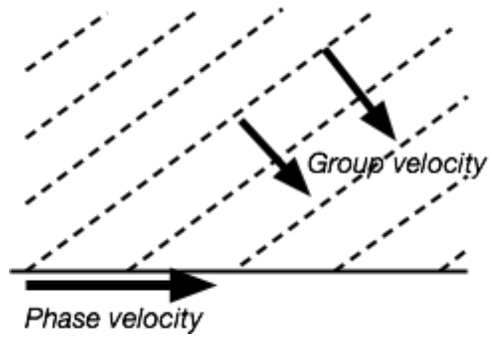


Figure 33 Phase velocity vs group velocity (source: Quora)

Figure 34 illustrates phase and group velocities in an electromagnetic wave.

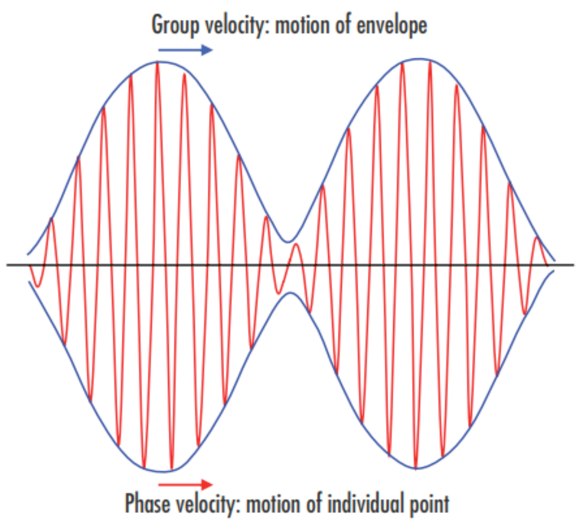


Figure 34 Phase and Group Velocities (source: www.edmundoptics.com)

In this figure the red trace is the carrier signal whose amplitude is varied by the modulating audio signal. The blue line drawn from peak to peak of the carrier signal creates an imaginary "envelope" encasing the carrier signal. This envelope helps us to visualize what is happening. The envelope is the same shape and frequency f_m of the modulating signal. The envelope contains the audio content and power being transported by the carrier signal. Consequently, the envelope travels at the group velocity which must be less than or equal to c. The phase velocity of the carrier signal may or may not be equal to c, depending on the characteristics of the medium through which the wave is traveling.

4.10 Polarization

Polarization is defined as the direction of the electric field in an electromagnetic wave. The wave in Figure 35 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 35 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 36. 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 37. A left circular polarized wave rotates in the opposite direction (counter clockwise).

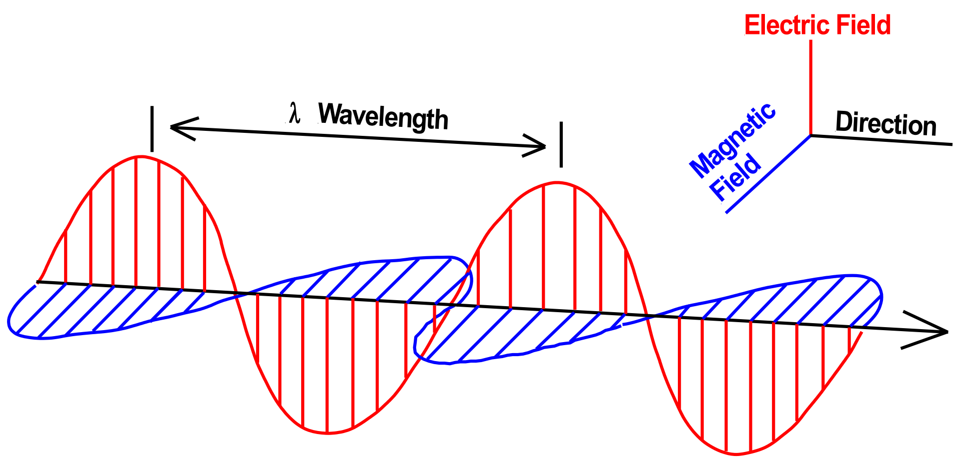


Figure 35 Vertically polarized electromagnetic wave (source: author)

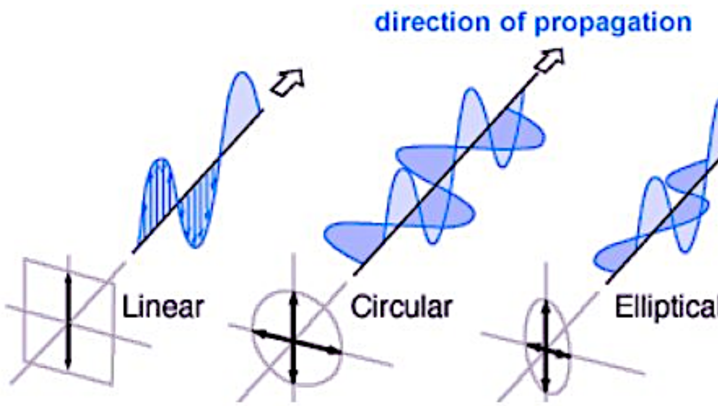


Figure 36 Typical polarizations of electromagnetic waves (source: Electronics For You)

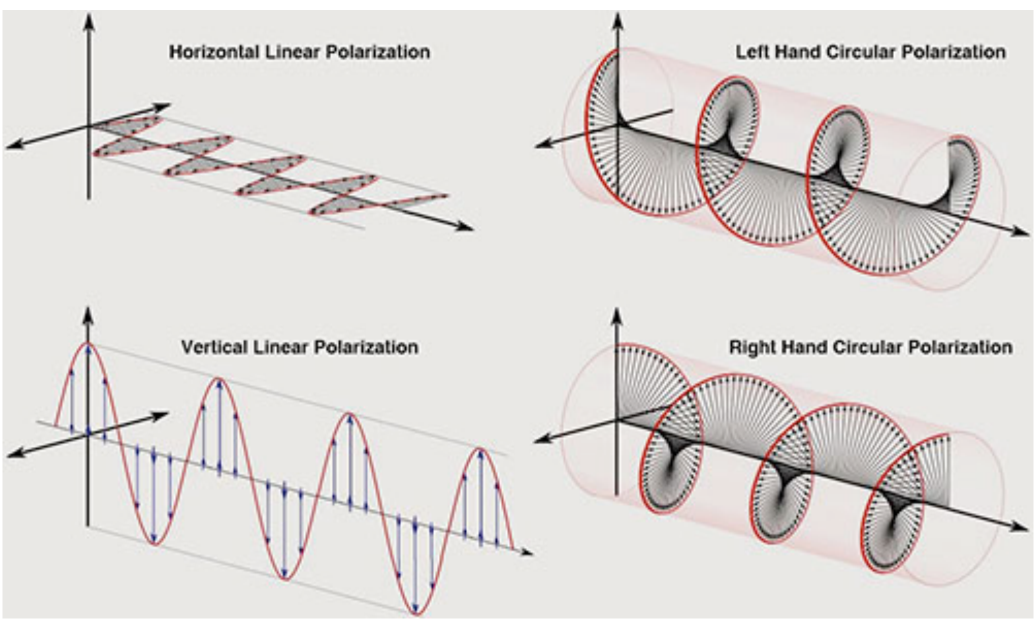


Figure 37 Linear and Circular Polarization (source: drones.stackexchange.com)

A vertical transmitting antenna radiates a linear vertically polarized electromagnetic wave as illustrated in Figure 38. The oscillating electrical current flowing up and down in the vertical (green) antenna produces the vertically oscillating electric field of the radiated wave.

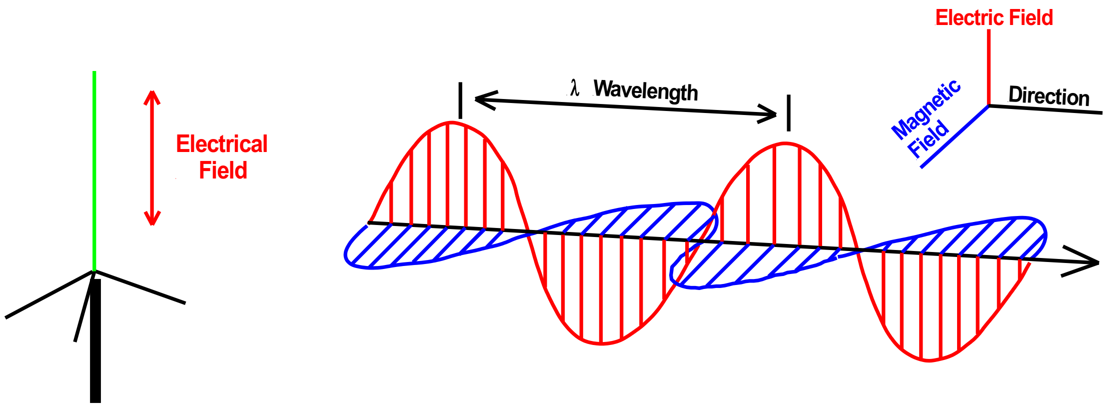


Figure 38 Vertical Antenna Polarization (source: author)

Figure 39 shows a horizontal transmitting antenna. The electrical current flowing in the green antenna wire produces an oscillating horizontal electric field in the electromagnetic wave that radiates outward from this antenna.

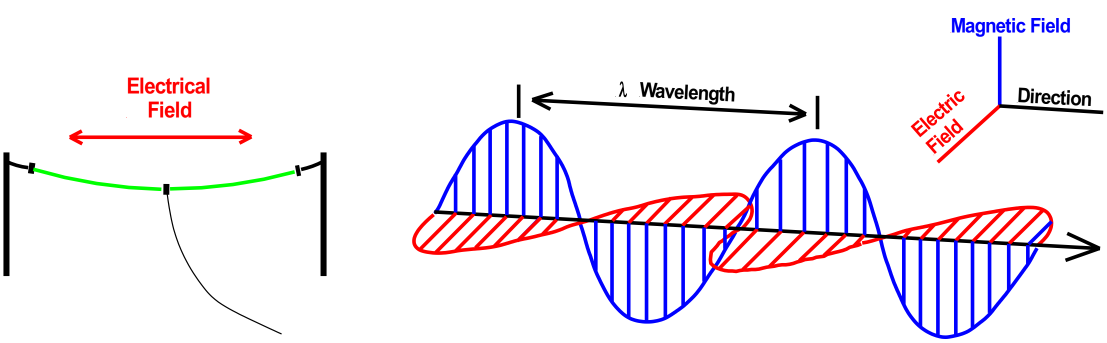


Figure 39 Horizonal antenna polarization (source: author)

The direction of polarization is very important for line-of-site (LOS) communications between two radio stations, illustrated in Figure 40. 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.

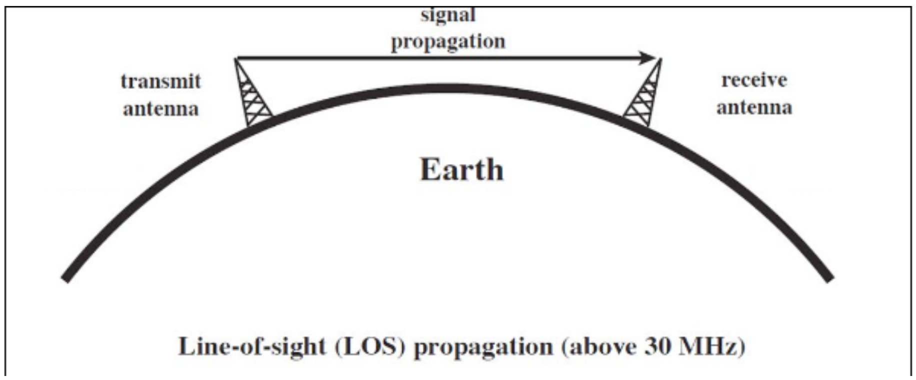


Figure 40 Line of sight radio transmission
(source: www.tutorialspoint.com/antenna_theory)

Figure 41 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. Polarization matters for line-of-site communications.

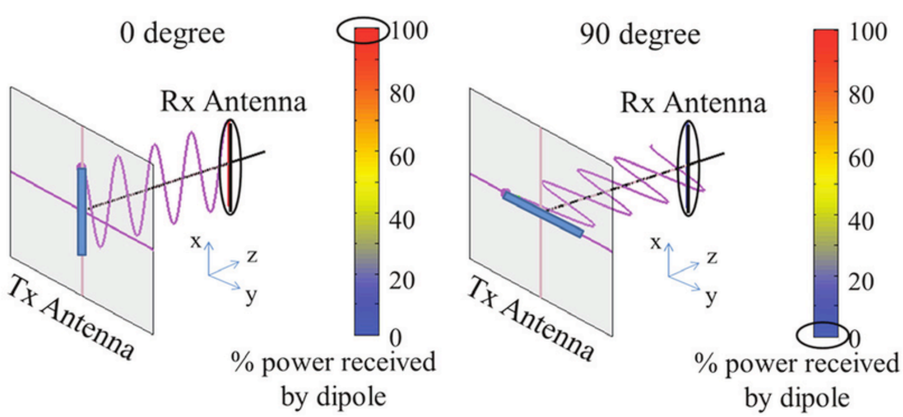


Figure 41 Polarized transmitting & receiving antennas (source: ResearchGate)

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

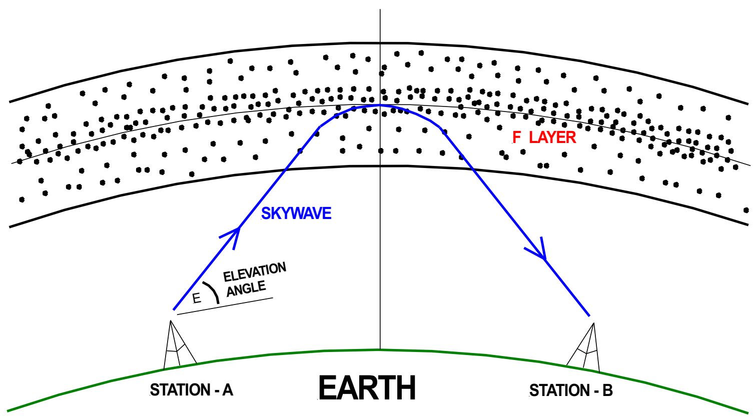


Figure 42 Radio communications through the ionosphere (source: author)

In Figure 42, high frequency 2 – 30 MHz radio waves are bent (refracted) back to Earth by the ionosphere permitting long distance over the horizon communications to occur. In the process, the radio waves become circularly polarized through interaction with the Earth's magnetic field and charged particles in the ionosphere. Consequently, for skywave communications the orientation between the transmitting and receiving antennas are generally 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 43) 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 44.

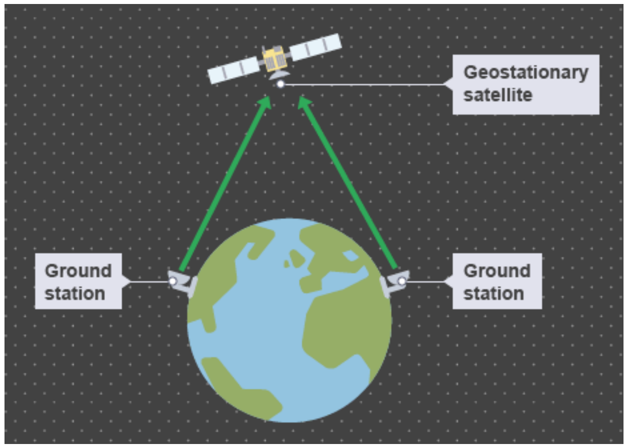


Figure 43 Line-of-site spacecraft communications (source: www.blendspace.com)

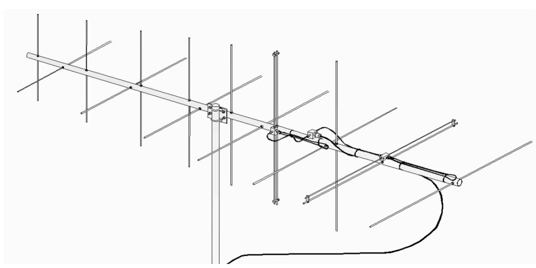


Figure 44 Circular polarized antennas (source: DX Engineering)

4.10.1 Non-polarized Light

Unlike radio waves, common sources of light, including sunlight, are unpolarized. The reason for this is that sunlight is produced by multiple sources each acting independently of the others. A ray of light from a single isolated source is still a transverse electromagnetic wave identical to that in Figure 35. However, the composite light from a multitude of sources produces a polarization that is completely random in nature as illustrated in Figure 45.

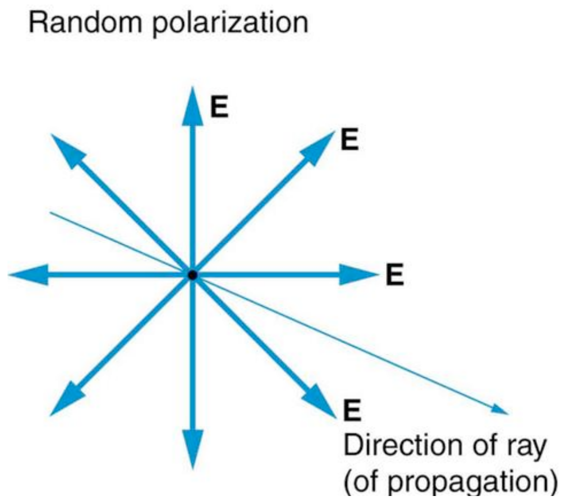


Figure 45 Unpolarized Light (source: physics stack exchange)

4.10.2 Polarizing Filters

Unpolarized light can be polarized by passing it through a polarizing filter such as that shown in Figure 46. In this figure only light that is vertically polarized passes through the filter. The rest of the unpolarized light is absorbed by the filter.

Polaroid is the trade name of the first commercially available low cost plastic polarizing material. The polarizing direction of the material is established during the manufacturing process by embedding long-chain molecules in a flexible plastic sheet. The sheet is then stretched so that the molecules aligned parallel to each other.

One of the many applications of Polaroid material is the manufacturing of Polaroid sunglasses. Polaroid sunglasses filter out most of the unpolarized sunlight permitting only that light polarized in the same direction as the plastic lenses to pass through the sunglasses.

Figure 47 shows unpolarized light passing through crossed polarizers. The first polarizer is oriented vertically allowing only vertically polarized light to pass through it. The light impinging on the second polarizer is therefore vertically polarized. However, the second polarizer is oriented horizontally blocking the vertically polarized light from passing through it. Consequently, all light is blocked from passing through crossed polarizers.

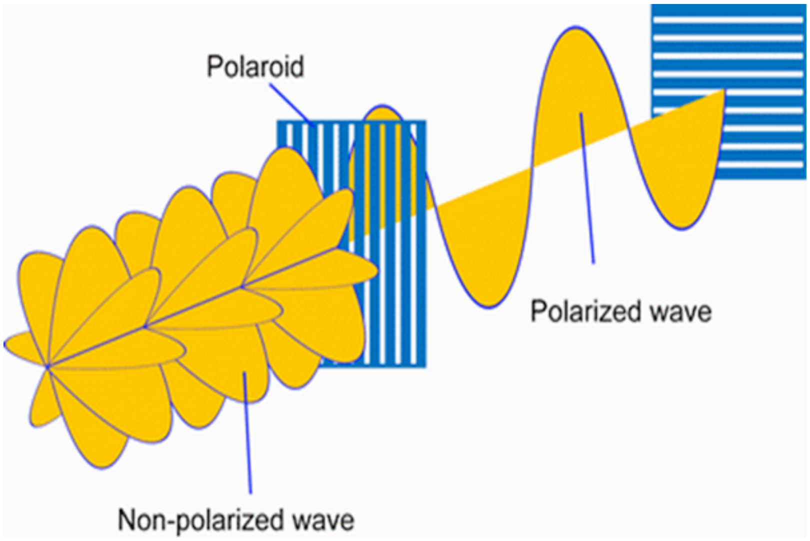


Figure 46 Polarizing non-polar light (source: quora.com)

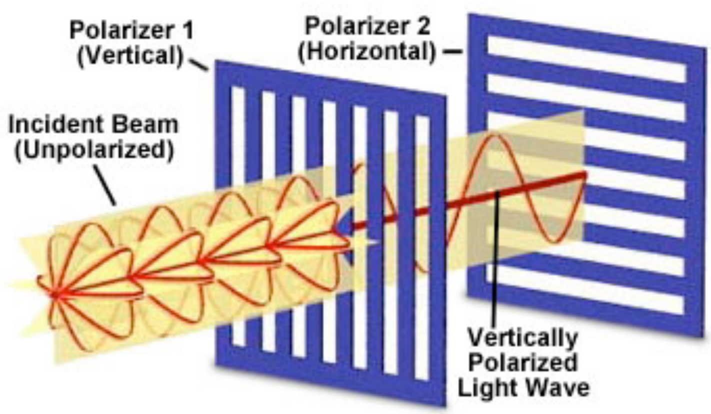


Figure 47 Light passing through crossed polarizers (source: Florida State University)

One can test whether sunglasses are polarized by looking through two pairs of glasses with one perpendicular to the other as shown in Figure 48. If both are polarized, all light will be blocked.

Figure 48 Crossed Polaroid sunglasses (source: Wikipedia)

4.11 Spectral Lines

Electrons in an atom can only occupy orbitals that are discrete distances from the nucleus. In addition, only specific orbitals are allowed. The permitted orbits are those for which the orbital angular momentum L of an electron is an integer multiple of $h/2\pi$, that is

$$L = n\frac{h}{2\pi} \qquad n = 1, 2, 3, \cdots$$

where h is Planck's constant (h = 6.62607004 × 10^{-34} m^2 kg / s) and n is called the principal quantum number.

An electron will not radiate any electromagnetic energy as long as it is in a permitted orbit. Emission or absorption of energy takes place only when an electron jumps from one allowed orbit to another (Figure 49). The radiation involved in an electron's transition from an orbit of energy E_m to another orbit with energy E_n is in the form of a photon. A photon is a packet of electromagnetic energy E_p equal to

$$E_p = hf = E_m - E_n$$

In this equation f is the frequency of the radiated energy and h is Planck's constant. (Photons will be discussed in more detail in the next chapter.)

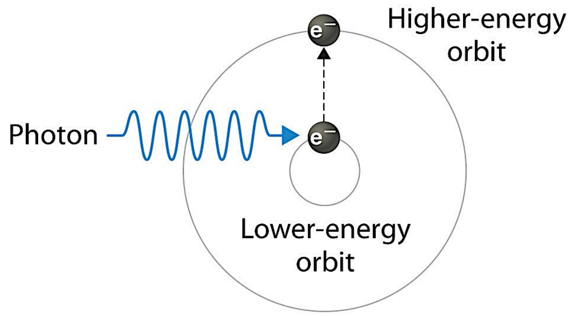


Figure 49 Photon absorption causes electron to jump to a higher energy level (source: Quora)

A hydrogen atom, for example, can absorbed energy (a photon of light) only if that energy is exactly equal to that required for the atom's electron to jump to a permitted higher orbit. The frequency of the absorbed light must therefore be:

$$f = \frac{E_m - E_n}{h}$$

Shining a broad spectrum of light through hydrogen gas ensures that within that spectrum there are photons at exactly the correct energy level (frequency) to excite a hydrogen atom electron causing it to jump to a higher energy orbital. Dark absorption lines appear in the continuous spectrum of light after it has passed through the hydrogen gas (Figure 50). Each black line corresponds to light that has been absorbed by the electron of a hydrogen atom. Since the light is absorbed, it is no longer present creating a black line or void in the spectrum. Black lines can occur only at specific frequencies, each frequency corresponding to the energy ΔE

$$f = \frac{\Delta E}{h} = \frac{E_m - E_n}{h}$$

required for an electron to jump from orbital m to orbital n.The presence of multiple absorption lines indicates that the electrons of various hydrogen atoms did not all jump to the same energy level. Some jumped to higher energy levels than others.

Atoms are most stable when their electrons are at their lowest possible energy levels, known as the ground state. Consequently, once excited, the electron of a hydrogen atom eventually drops back to its lowest energy level. In the process it emits a photon in a random direction at the same frequencies as the photon it previously absorbed. The emitted photons from many hydrogen atoms create an emission spectrum (Figure 50) which can be seen when the hydrogen gas is viewed from the proper direction, that is a direction in which the light source creating the broad spectrum is not in the background.

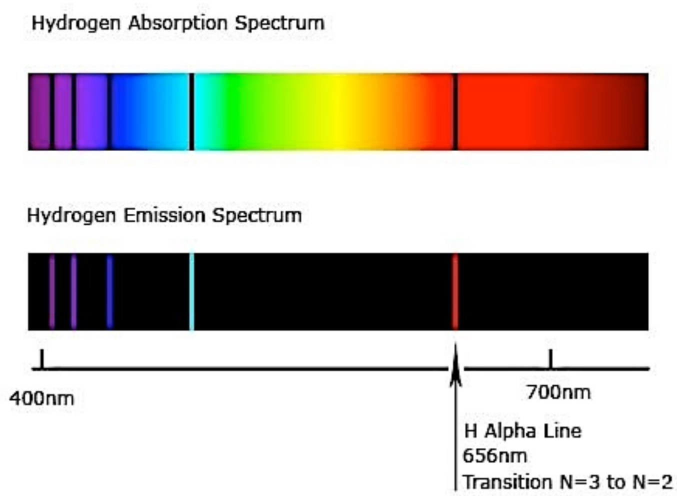


Figure 50 Hydrogen absorption and emission spectrum (source: Quora)

Shining broad band light through other atoms heated to their gaseous state creates similar absorption and emission spectrum. Each type of atom has its own unique spectrum which serves as "a finger print" identifying the type of atom. The continuous spectrum of energy from the Sun, and other stars, contain absorption lines that reveal the type of atoms present in the outer regions of the Sun, as illustrated in Figure 51. However, some of these absorption lines are the result of photons being absorbed by atoms in the Earth's atmosphere. For this reason, observing the spectrum of stars is best performed from spacecraft.

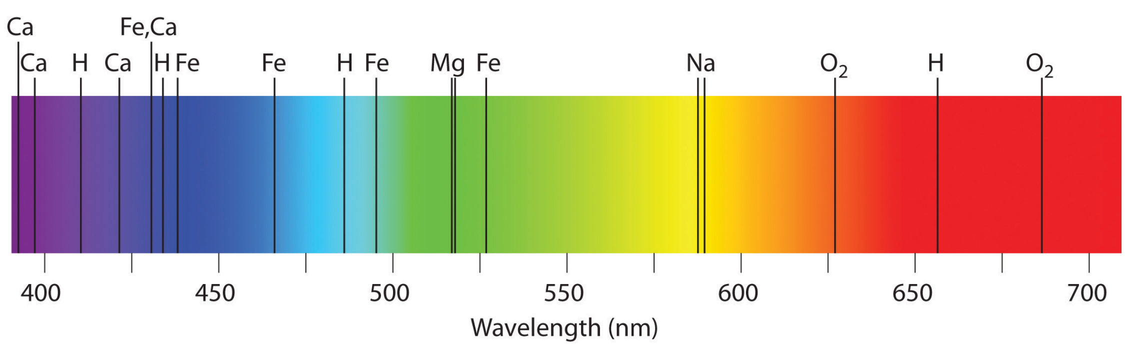


Figure 51 Absorption spectrum of the Sun (source: Continuous Spectrum)

4.12 Doppler Effect and Red / Blue Shift

The sound of an ambulance siren seems higher in frequency to a stationary observer when the ambulance is approaching than when it is standing still. In contrast, the frequency of the siren sounds lower when the ambulance is moving away. These phenomena are known as the Doppler effect. Intuitively, the sound waves from an approaching ambulance are compressed making them sound higher in frequency. When the ambulance is moving away, the sound waves are stretched out causing them to sound lower in frequency as illustrated in Figure 52.

Our intuition can be easily verified.

f_{OSA} = frequency of a sound when the sound source is approach (SA) a stationary observer (O).

f_S = frequency of a sound wave when the source (S) of the sound is at rest

λ_S = wavelength of a sound wave when the source is at rest

T_S = period of a sound wave with the source is at rest

v_S = speed of the sound source

v = speed of sound through the air

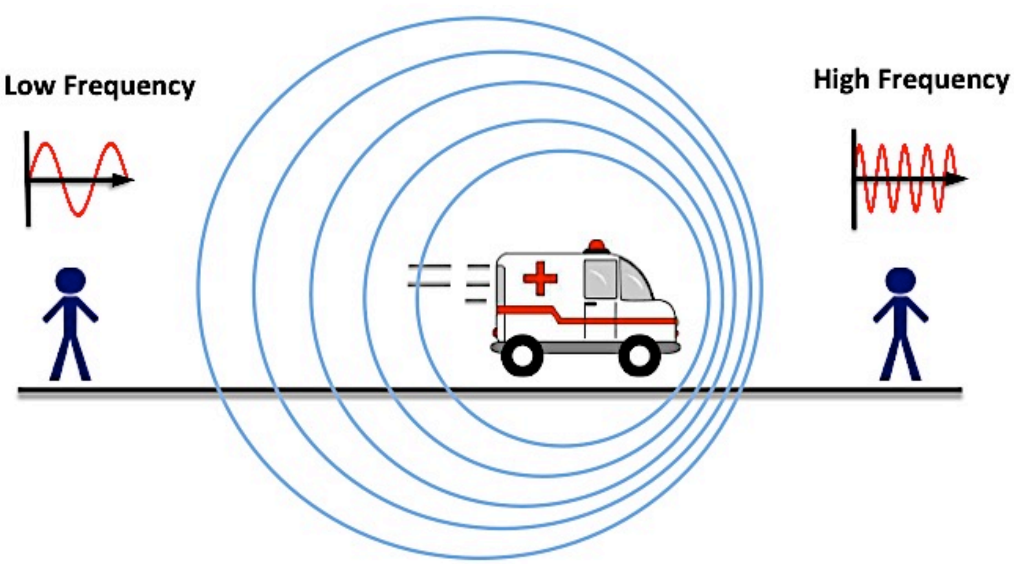


Figure 52 Doppler Effect (source: Pinterest)

The period of a sound wave when the source is at rest is

$$T_S = \frac{1}{f_S}$$

The wavelength of the sound wave with the source at rest is

$$\lambda_S = v\, T_s = \frac{v}{f_S}$$

The speed of sound through air, v , is a property of the air through which the sound is traveling and does not depend on the speed v_S at which the source of the sound is moving.

In the time period T_S the source travels toward the listener a distance of

$$d = v_S\, T_S$$

Consequently, the wavelength of the sound appears to be shorter than it actually is by an amount equal to d . Thus the wavelength perceived by the observer is

$$\lambda_{OSA} = \lambda_S - d = \frac{v}{f_S} - v_S\, T_S = v\, T_S - v_S\, T_S = (v - v_S)T_S = \frac{v - v_S}{f_S}$$

Since the speed of sound v through the air does not depend on the speed of the source, the frequency of the sound as perceived by the observer is

$$f_{OSA} = \frac{v}{\lambda_{OSA}} = \frac{v\, f_S}{(v - v_S)} = f_S\left[\frac{v}{v - v_S}\right]$$

which is higher in frequency than the sound actually produced by the sound source.

If the sound source is moving away from the listener (source receding), the distance d is **added** to the sound's wavelength giving

$$\lambda_{OSR} = \lambda_S + d = v\, T_S + v_S\, T_S = (v + v_S)T_S = \frac{v + v_S}{f_S}$$

and the frequency of the sound heard by the observer is

$$f_{OSR} = \frac{v}{\lambda_{OSR}} = \frac{v\, f_S}{(v + v_S)} = f_S\left[\frac{v}{v + v_S}\right]$$

which is lower in frequency than the actual sound emitted by the sound source.

For an observer approaching a stationary sound source the situation is slightly different.

In this case the speed of the observer as he or she approaches the sound source is v_O . Again, the speed of sound v through the air is determined by the characteristics of the air and has nothing to do with the speed of the observer. The effective speed of the sound wave as it reaches the moving observer is

$$v_a = v + v_O$$

The period of the sound wave perceived by the approaching observer is

$$T_{AO} = \frac{\lambda_S}{v_a} = \frac{\lambda_S}{v + v_O}$$

The frequency of the sound as heard by the approaching observer is

$$f_{AO} = \frac{1}{T_{AO}} = \frac{v + v_O}{\lambda_S}$$

and since

$$\lambda_S = \frac{v}{f_S}$$

$$f_{AO} = \frac{1}{T_{AO}} = \frac{v + v_O}{\lambda_S} = \frac{v + v_O}{v/f_S} = f_S \left[\frac{v + v_O}{v}\right]$$

which is higher in frequency than the sound actually produced by the sound source.

The speed of the observer is subtracted from the speed of sound if the listener is moving away (receding) from the sound source. In this case the sound heard by the observer is

$$f_{RO} = \frac{1}{T_{RO}} = \frac{v - v_O}{\lambda_S} = \frac{v - v_O}{v/f_S} = f_S \left[\frac{v - v_O}{v}\right]$$

which is lower in frequency.

The same is true for a light source, such as a distant star, moving toward or away from Earth; or, as an observer on Earth moves toward or away from a star as the Earth orbits around the Sun. In these cases the speed of sound v through air in the above equations is replaced by the speed of light c in free space, which is c = 299,792.458 km/sec. For light the equations become:

1. Light source receding from the observer:

$$f_{OSR} = f_S \left[\frac{c}{c + v_S}\right]$$

The frequency of the observed light source is lower than its actual frequency, which means that its wavelength

$$\lambda_{OSR} = \frac{c}{f_{OSR}}$$

is longer causing the light to appear redder in color than it actually is.

2. Light source approaching the observer:

$$f_{OSA} = f_S \left[\frac{c}{c - v_S}\right]$$

The frequency of the observed light source is higher than its actual frequency, which means that its wavelength

$$\lambda_{OSA} = \frac{c}{f_{OSR}}$$

is shorter causing the light to appear bluer in color than it actually is.

3. Observer receding from the light source:

$$f_{OR} = f_S \left[\frac{c - v_O}{c}\right]$$

The frequency of the observed light source is lower than its actual frequency, which means that its wavelength

$$\lambda_{OR} = \frac{c}{f_{OR}}$$

is longer causing the light to appear redder in color than it actually is.

4. Observer approaching light source:

$$f_{OA} = f_S \left[\frac{c + v_O}{c}\right]$$

The frequency of the observed light source is higher than it actually is, which means that its wavelength

$$\lambda_{OA} = \frac{c}{f_{OA}}$$

is shorter causing the light to appear bluer in color than it actually is.

It is important to reiterate that the speed of light c in the above equations is constant. Only the wavelength λ and frequency f of the light change as a light source approaches or recedes from an observer. That is

$$c = \lambda \cdot f$$

If the frequency f decreases, then the wavelength λ must increase, and visa versa.

We know that the universe is expanding because the spectrum of distant stars are red shifted. The absorption spectrum of light from a star has an expected pattern, for example that shown in Figure 53. However, the entire pattern is shifted to the right toward the red region of the spectrum if the star is moving away from Earth. Consequently, each spectral line in the pattern has a longer wavelength and lower frequency than expected. It is red shifted.

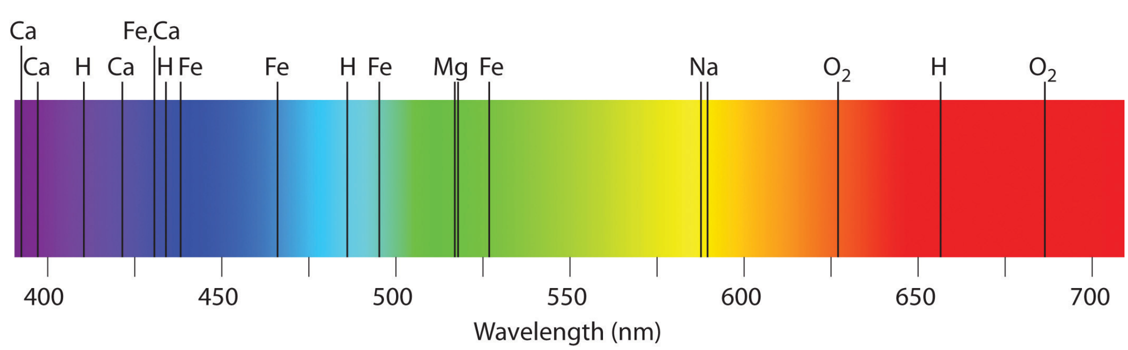


Figure 53 Typical absorption spectrum from a star (source: Continuous Spectrum)

Similarly, stars and galaxies that are approaching us, and some are, are blue shifted. That is, the stelar spectral pattern is shifted to the left toward the blue region of the spectrum.

4.13 References

J. J. O'Connor and E. F. Robertson; *Light through the ages: Ancient Greece to Maxwell*; MacTutor History of Mathematics, (University of St. Andrews, Scotland) 2002, http://www-history.mcs.st-andrews.ac.uk/HistTopics/Light_1.html

Abramowitz, Mortimer; Spring, Kenneth R.; Parry-Hill,Matthew; Flynn, Brian O.; Tchourioukanov, Kirill I. Fellers, Thomas J.; Davidson, Michael W.; *The Physics of Light and Color*; Olympus-Life Science, Olympus America Inc. https://www.olympus-lifescience.com/en/microscope-resource/primer/lightandcolor/

Setterfield, Barry; *A Brief History of c*; 1987

Halliday, David; Resnick, Robert; *Physics For Students of Science and Engineering Part I*; 1962 John Wiley & Sons, Inc.

Halliday, David; Resnick, Robert; *Physics For Students of Science and Engineering Part II*; 1962 John Wiley & Sons, Inc.

Spring, Kenneth R.; Fellers, Thomas J.; Zuckerman, Lawrence D.; Davidson, Michael; *Speed of Light*; Olympus-Life Science, Olympus America Inc.

Fowler, Michael; UVA Physics Department, *The Speed of Light,* University of Virginia, http://galileoandeinstein.physics.virginia.edu/lectures/spedlite.html

Urone, Paul Perter; *Image Formation by Mirrors*; LibreTexts

Kovacs, Peter; *History of the reading stones*; Light Microscopes

UVA Physics Department; *Newton's Particles Theory of Light-Galileo;* University of Virginia; http://galileo.phys.virginia.edu/classes/609.ral5q.fall04/Lecture:PDF/L20LIGHTII.pdf

Ling, Samuel J.; Sanny, Jeff; Moefs, William; *University Physics Volume 3*; 2018 Openstax Rice University

Plonsey, Robert; Collin, Robert E.; Principles and Applications of Electromagnetic Fields; McGraw-Hill Book Company, Inc., 1961

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

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

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

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

Yeang, Chen-Pang; Probing the Sky with Radio Waves; The University of Chicago Press, 2013

Goodman, John M.; Space Weather & Telecommunications; Springer, 2005

Zavrel, Robert J. Jr.; Antenna Physics: An Introduction; The American Radio Relay League, Inc. 2016

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

White, Ian; The VHF/UHF DX Book; DIG Publishing LTD, 1995

Dampier, W. C.; *A History of Science*; 1961 Cambridge University Press

Cajori, Florian; *A History of Physics in its Elementary Branches*; 1917, The Macmillan Company

Williams, Henry Smith; *A History of Science Part 1*; 1904 Kessinger Publishing

Williams, Henry Smith; *A History of Science Part 2*; 1904 Kessinger Publishing

Wikipedia; *History of optics*

Wikipedia: *History of glass*

Wikipedia; *Glasses*

Wikipedia; *Mirror*

Wikipedia; *al Haytham's Book of Optics*

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

Foukal, Peter; *Solar Astrophysics*; 2013 Wiley-VCH Verlag GmbH & Co.

Tyson, Neil Degrasse; Strauss, Michael A.; Gott, J. Richard; Welcome to the Universe an Astrophysical Tour; Princeton University Press; 2016

Palen, Stacy; Kay Laura; Blumenthal; Understanding Our Universe; W. W. Norton & Company, Inc., 2018

Golub, Leon; Pasachoff, Jay M.; *The Nearest Star The Surprising Science of Our Sun*; 2014 Cambridge University Press

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

Cander, Ljiljana R.; Ionospheric Space Weather; Springer Geophysics, 2019

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

Zettili, Nouredine; *Quantum Mechanics Concepts and Applications*; 2009 John Wiley & Sons, Ltd.

Masi, Marco; *Quantum Physics: an overview of a weird world*; 2018 Marco Masi

Klus, Dr. Helen; *How We Came to Know the Cosmos: Light & Matter*; 2017, The Star Garden, UK.

Friedman, Herbert; *The Astronomer's Universe Stars, Galaxies, and Cosmos*; 1990, The Commonwealth Fund Book Program

Hogan, Craig J.; *The Little Book of the Big Bang A Cosmic Primer*; 1998 Springer-Verlag New York Inc.

Sagan, Carl; *Cosmos*; 1980 Carl Sagan Productions

Piccioni, Robert L.; *Everyone's guide to Atoms Einstein and the Universe*; 2010 Robert L. Piccioni; Real Science Publishing

Starr, Chester G.; *A History of the Ancient World*; 1991, Oxford University Press

Davies, Norman; *Europe a History*; 1998, Harper Perennial

Pannekoek, A.; *A History of Astronomy*; 1961, George Allen and Unwin Ltd.

Merzbech, Uta C.; Boyer, Carl B.; *A History of Mathematics*: 2011, John Wiley & Sons, Inc.

El-Abbadi, Mostafa; *Library of Alexandria*; Encyclopedia Britannica

Williams, L. Pearce; *History of Science*; Encyclopedia Britannica

Rodrigue, Christine M; *The Geography of Mars: Lab Cassini, Richer, and Picard Experiment, 1672*; Department of Geography California State University, Long Beach, CA.; http://www.csulb.edu/~rodrigue/

Herivel, John; *Christiaan Huygens Dutch Scientist And Mathematician*; Feb 14, 2019; Encyclopedia Britannica

Rodriguez, Emily; *Robert Hooke British Scientist*; Feb 27, 2019; Encyclopedia Britannica

Westfall, Richard S; *Sir Isaac Newton English Physicist and Mathematician*; Feb 1, 2019; Encyclopedia Britannica

Crowley, Theodore; *Roger Bacon*; Encyclopedia Britannica

Jones, Alexander Raymond; *Ptolemy*; Encyclopedia Britannica

Meinwald, Constance C.; *Plato*; Encyclopedia Britannica

Amadio, Anselm H.; Kenny Anthony J. P.; *Aristotle*; Encyclopedia Britannica

Lorch, Richard;Ibn; *al-Haytham*; Encyclopedia Britannica

Van der Waerden, Bartel Leendert; Taisbak, Christian Marinus; *Euclid*; Encyclopedia Britannica

Editors of Encyclopedia Britannica; *Heron of Alexandria*; Encyclopedia Britannica

Westman, Robert S.; *Johannes Kepler*; Encyclopedia Britannica

Van Helden, Albert; *Galileo*; Encyclopedia Britannica