Chapter 11 Notes: Electromagnetic Waves

Can electric and magnetic fields travel through space on their own?

At this point in our discussion, we need to pause for a moment to review the major principles of electricity and magnetism that we have considered so far. There are three fundamental experimental results on which nearly everything else has depended:

I. Force between two charged particles. If a particle with charge q_1 and another particle with charge q_2 are separated by a distance r, each particle exerts an "electrical" force on the other. The magnitudes of the forces are equal, but the directions are opposite. The forces are directed along the line connecting the two particles. The forces are attractive if the particles have opposite charges, and they are repulsive if the particles have like charges. The magnitude of the force is given by this equation, which is called "Coulomb's law":

$$F_{electrical} = k\frac{|q_1||q_2|}{r^2} \qquad k = 9\times10^9\ \frac{\mathrm{Nm}^2}{\mathrm{C}^2}$$

Everything we discussed in Chapters 1 through 8 could be traced back, in the end, to this one principle.

II. Force between two currents. If two long, straight conductors of length L are parallel to each other, separated by a distance r, and carrying currents I_1 and I_2, each conductor exerts a "magnetic" force on the other. The magnitudes of the forces are equal, but the directions are opposite. The forces are attractive if the currents flow in the same direction, and they are repulsive if the currents flow in opposite directions. The magnitude of the force is given by this equation:

$$F_{magnetic} = \frac{\mu_0}{2\pi}\frac{I_1 I_2 L}{r} \qquad \mu_0 = 4\pi\times10^{-7}\ \frac{\mathrm{N}}{\mathrm{A}^2}$$

(If the conductors are perpendicular to each other, the force is zero.) The consequences of this principle were explored in Chapter 9, where the proportionality constant a was used to represent the quantity ($\mu_0/2\pi$).

III. Magnetic Induction. If a conducting loop of area A is in the presence of a magnetic field with magnitude B, an electrical current may be "induced" in the conductor by a ***change*** in the magnitude or direction of the magnetic field, or in the shape of the loop. If the conductor obeys Ohm's law and has resistance R, and the angle between the field and the normal to the loop is θ, the induced current is given by this equation (where $\Phi \equiv BA\cos\theta$):

$$I = \frac{1}{R}\frac{\Delta\Phi}{\Delta t} \qquad [\text{"}Faraday's\quad law\text{"}]$$

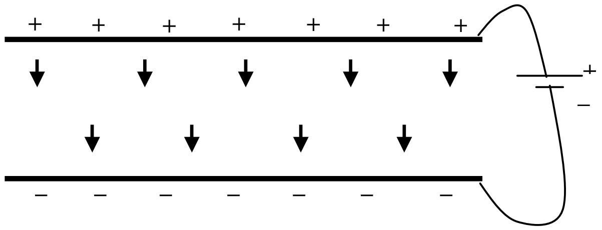
Faraday's law was discussed in Chapter 10.

Essentially everything we have covered up until now rests on these three fundamental principles. This was the status of scientific knowledge regarding electricity and magnetism around 1855, when Scottish physicist James Clerk Maxwell began a systematic reexamination of the entire subject. Maxwell set out to formulate a comprehensive mathematical theory that would explain all known electromagnetic phenomena and give the subject a rigorous and precise theoretical structure. The results of his work ultimately led to some of the most surprising and important scientific discoveries ever made. Here we will only sketch an outline of Maxwell's work; the details require mathematics that is beyond the scope of this course. (We will also describe this work in modern language. Many of the specific terms and formulations that were used by Maxwell, Faraday, and other early researchers are quite different from those in current use.)

Another perspective on Faraday's law: A changing magnetic field produces an electric field. We have seen that when the magnetic flux contained within a conducting loop is changing, an electric current is induced in the loop. In other words, charged particles are accelerated from rest, and forced to move in a flowing electric current. In order to do this, an electric force must be applied to the charges, and therefore an electric field must be present. The conclusion is inescapable that the changing magnetic flux has somehow ***created*** an electric field, which is then responsible for forcing the charges to move around the circuit. The analysis of Faraday's observations thus led to the realization that a ***changing magnetic field produces an electric field.*** In fact, it became clear that this is the underlying physical principle at work in Faraday's law.

What happens when current varies? When Maxwell began to systematically review all current knowledge about electricity and magnetism, the fundamental principles could be stated as follows: *(1) charged particles exert electric forces on each other; (2) currents exert magnetic forces on each other; (3) a changing magnetic field produces an electric field.* In the course of his analysis, Maxwell realized that something was missing from this picture, and that these three principles, by themselves, were not completely consistent with each other. The difficulty arose when he attempted to analyze situations where ***the current varied in magnitude.*** Maxwell pointed out something that no one had realized up to then: in situations where the electric current varied in magnitude, the laws of electromagnetism – as they were understood at that time – did not always give the correct prediction for the resulting magnetic field. This was a serious problem, and proved that the theory as it stood was incomplete.

An example of a system where the current varies is the *charging of a capacitor.* Up until now, we have only considered a capacitor after it had already been connected to a battery in a set-up such as this:



Here, the top plate of the capacitor has been loaded with extra positive charges, and the lower plate with negative charges. As a result, a nearly uniform electric field is created in the space between the plates (as indicated by the arrows). However, when an ***uncharged*** capacitor is first connected to a battery, it does not become charged *instantaneously*; some time is required. At first, the charges flow freely and the current in the wires is large. As time passes and as the capacitor begins to get charged up, the charges flowing in the wires encounter an increasing amount of repulsion from the extra charges that are ***already*** sitting on the plates. As a result, fewer charges flow each second and the current decreases. Eventually, the repulsion from the charges on the plates equals the forces originating from the charges on the terminals of the battery, and the current stops flowing completely. At this point, the capacitor is fully charged. During this process, the electric field magnitude in the space between the plates has been increasing as the plates acquire increasing amounts of charge.

Maxwell's proposal: Maxwell showed that if one tried to calculate the magnetic field that was present in this set-up in the usual way, one did not get a consistent answer. Depending on how the calculation was carried out, two completely different answers would result – and they could not both be correct. Maxwell realized that in order to get consistent answers one would need to identify an ***additional*** source of magnetic field, and he made a startling proposal: perhaps a varying ***electric*** field – such as the one between the capacitor plates – could produce a ***magnetic*** field! This would be a "mirror-image" of Faraday's law, which stated that a varying magnetic field produced an electric field. If he included a magnetic field contribution due to the changing electric field, Maxwell was able to formulate a theory that could always make consistent predictions. It remained, then, to test out the new theory.

Unfortunately, this was easier said than done. The predicted contribution to the magnetic field from the changing current was very small – much too small to be reliably detected with the equipment available in the 1860's. So it would not be possible to make a ***direct*** test of this prediction. Instead, Maxwell continued to analyze his new theory to see what other observable effects might be produced by the electric and magnetic fields. This process led him to make an extraordinary discovery that has had an immeasurable impact on the entire world.

Maxwell found that a peculiar chain of cause-and-effect could proceed as follows: a changing electric field creates a magnetic field – but that magnetic field is ***itself*** changing. Therefore ***it*** creates in turn an ***electric*** field – which is also changing! And so the process repeats: a changing electric field creates a changing magnetic field, and a changing magnetic field creates a changing electric field, and so on. And while this process of mutual creation proceeds, the combined electric and magnetic fields ***travel through space*** in a wave-like pattern, traveling far away from the electric charges that produced them in the first place. It is similar to the way a water wave spreads out from a rock thrown into a lake, with circular ripples spreading out from the point where the rock struck.

Properties of Electromagnetic Waves: Based on his mathematical theory of electromagnetism, Maxwell was able to determine several important properties of this predicted "electromagnetic wave":

1) The most general form of this wave would have both the electric field and the magnetic field oscillating in a regular pattern. The fields would reach a maximum magnitude, would then decrease in magnitude until they momentarily disappeared, and then they would reappear pointing in the ***opposite*** direction as their magnitude gradually increased. Then, the whole process would repeat.
2) The electric field would always point perpendicular to the magnetic field.
3) The electric and magnetic fields would vary "in phase," which is to say that they would reach maximum and minimum values at the same moment as each other.
4) The frequency of oscillation of the electric and magnetic fields could be ***any*** value; that is, it seemed that electromagnetic waves of any frequency should be possible.
5) The direction in which the wave would travel (called the direction of "propagation") would always be perpendicular to both the electric and the magnetic field.
6) The electromagnetic ("e-m") waves would carry energy, and so could transport energy from the source that produced the waves to some other point very far distant from the original source.

Finally, Maxwell was able to calculate the velocity of his predicted electromagnetic waves as they traveled through empty space, in terms of the constants k and μ_0 that appeared in Coulomb's law and the equation for the magnetic force. In 1865 Maxwell showed that the velocity of the e-m wave, which is symbolized by the letter "*c*," was given by the following equation:

$$c = \sqrt{\frac{4\pi k}{\mu_0}}$$

The values of both of these constants had been very carefully measured by this time, and so Maxwell was able to determine a precise numerical value for the velocity. He found this:

$$c = \sqrt{\frac{4\pi k}{\mu_0}} = \sqrt{\frac{(4\pi)(9\times10^{9}\ \frac{\text{Nm}^2}{\text{C}^2})}{(4\pi\times10^{-7}\ \frac{\text{N}}{\text{A}^2})}} = \sqrt{9\times10^{16}\ \frac{\text{m}^2\text{A}^2}{\text{C}^2}} = 3\times10^{8}\ \frac{\text{m}}{\text{s}}$$

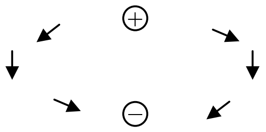
This was an extremely surprising result, to say the least, because this very high velocity ***was already known to physicists in the 1860's: it had only recently been measured to be precisely the speed of light!***

What possible connection could there be between light, and charged particles, magnets, and induced currents? Could this be just a coincidence? Certainly no one had suspected up until that time that there might be a direct connection between electromagnetic phenomena and light. Maxwell, however, felt confident that he had shown that light was simply one of many possible forms of electromagnetic wave. Although all electromagnetic waves would have to travel through a vacuum at the speed *c*, they could have any desired oscillation frequency. And so, Maxwell predicted that it should be possible to create electromagnetic waves with various different oscillation frequencies that should all have the exact same propagation speed: the speed of light. This, then, became the prediction on which the validity of Maxwell's theory would be tested.

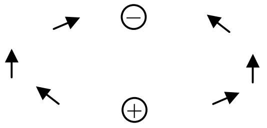
Unfortunately, it took more than 20 years for the test to be carried out; Maxwell died in 1879 before ultimate proof of his theory had been established. In 1887, Heinrich Hertz created in his laboratory an electromagnetic wave by making use of electrical circuits. The wave was created by an oscillating electric current, which caused charged particles to jump back and forth across a small gap between conducting spheres. This created a "sparking" that served as a transmitting antenna. A few meters away from the transmitting antenna, the traveling waves were detected by a "receiving" antenna of a similar type. Just as predicted, the waves were found to travel at exactly the speed of light. Maxwell had been proved correct. The early pioneers of electromagnetism could not possibly imagine the extent to which their discoveries would influence the world a hundred years later!

How are electromagnetic waves created? We have hinted at the basic process behind creation of electromagnetic waves: a ***varying*** electric current is required. In a steady current, charges are moving at a constant average velocity. However, when the current varies, the charges are *accelerated.* ***Charged particles that are undergoing acceleration radiate electromagnetic waves.*** An example of an accelerated charge is one that is oscillating back and forth, as if it were on a spring; this oscillating charge will generate an electromagnetic wave. The ***frequency*** of the oscillating electric and magnetic fields that form the wave will be equal to the oscillation frequency of the charge.

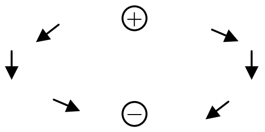
A common method of producing e-m waves is to have a positive and negative charge near each other, with each charge oscillating "out of phase" with the other. That is, when one goes up, the other goes down. Here is an illustration of such an oscillation seen at three successive moments in time (t_1, t_2, and t_3), where some of the net electric field vectors in the neighborhood of the charges are shown. (In the first diagram on the left, the positive charge is moving ***downward***, the negative charge is moving ***upward***.)



t_1



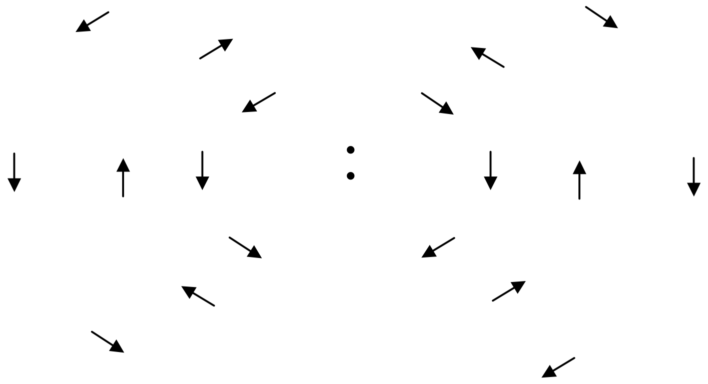
t_2



t_3

Because the charges are moving, a magnetic field is created as well. (Remember that magnetic fields are created by currents, which are nothing more than moving charges.) But these charges are also ***accelerating*** (their velocity is constantly changing as they oscillate up and down); therefore, they produce an ***electromagnetic wave***. The electric and magnetic fields (which are perpendicular to each other) travel outward away from the oscillating charges. If you were to stand in one spot to the right of the pair of charges, you would "see" the e-m wave pass you by. First you would observe a downward pointing electric field vector. A moment later the electric field vector in front of you would point upward, and then the "downward-upward" pattern would repeat.

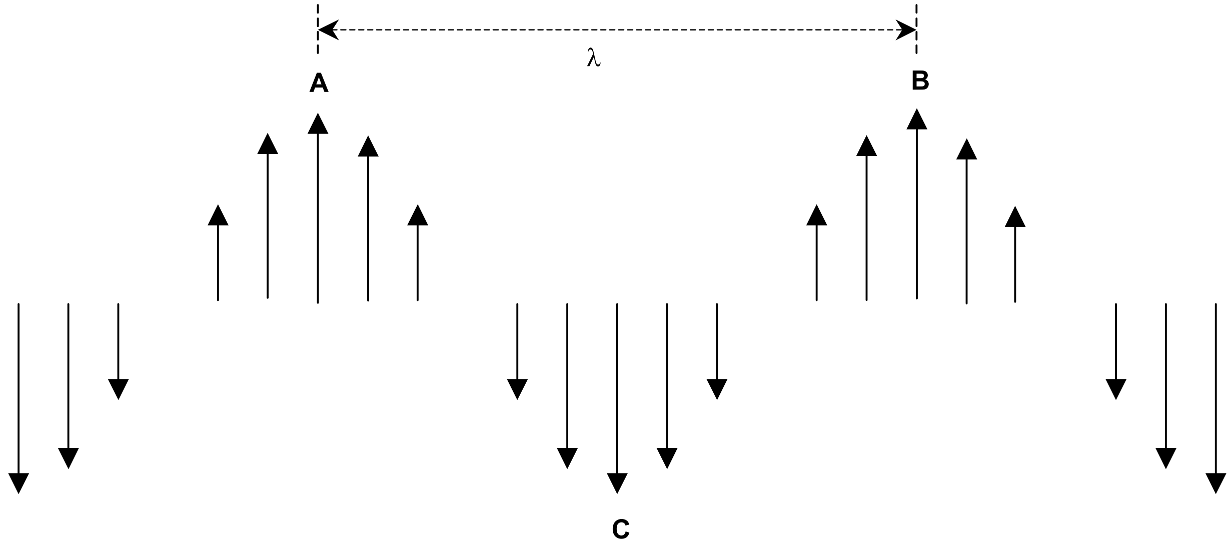
If you could stand far back and take a snapshot of the whole scene, you would observe something like this picture:



(Although here all vectors are shown with equal magnitudes, those further away from the source would actually have smaller magnitudes, as would those above and below the horizontal axis)

As time passes, the electric and magnetic fields "spread out" from the oscillating charges just as ripples spread out from a stone dropped in a pool. Of course, we have only shown a few of the electric field vectors. We could have drawn as many as we wanted, because the e-m wave spreads ***all throughout*** the space surrounding the charges. To simply matters, we will focus our attention on the electric field pattern observed along a horizontal axis drawn through the center of the pattern above.

Once again we are observing this pattern at just one moment in time, but from our perspective "far away" we can see the electric field vectors as they appear over a long stretch of the horizontal axis far from the charges. We would observe something like this (where the distance between the vectors labeled "A" and "B" is called "λ"):



(We have not shown the magnetic field vectors. They point perpendicular to the electric field vectors and so they alternate between pointing "into the page" and "out of the page.")

Notice that at a certain point, the electric field vector actually disappears; at that point, the electric field magnitude is equal to zero (and so is the magnetic field magnitude). In the diagram at the bottom of page 4 we only showed the electric field vectors that had ***maximum*** magnitude (such as the ones marked A, B and C in the diagram above). However, as the charges oscillate, the electric (and magnetic) field magnitudes vary continuously, decreasing gradually from the maximum value down to zero, and then building up again.

The pattern we see here is typical of a "transverse" wave, in which the direction of oscillation (here, "up-and-down") is ***perpendicular*** to the direction in which the wave is traveling. (This wave is traveling to the right.) We see that the pattern continuously repeats; here we have shown two full repetitions of the pattern. The horizontal distance between the vectors marked "A" and "B" is called the ***"wavelength"*** [symbol: λ]. This distance contains ***one*** full repetition of the pattern. (As we see, this is the distance between two successive maximum-magnitude vectors that both point in the same direction.) The maximum magnitude is called the ***"amplitude"*** [symbol: A].

If we were to stand in one location and watch the wave pass us by, the electric field vector directly ***in front of us*** would be observed to oscillate up and down. The time that elapses between two consecutive appearances of a maximum "upward-pointing" vector is called the ***"period"*** of the wave [symbol: T]. This is the time required for one full "up-down-up" oscillation, and so is the time needed for the complete pattern to repeat itself once. The number of full oscillations that occur each second is called the ***"frequency"*** of the wave [symbol: f]. The frequency is just the inverse of the period: $f = 1/T$.

During the time it takes for one full oscillation (which is equal to the period, T), the wave travels a distance equal to one wavelength. Since all e-m waves travel through a vacuum with speed c, we have this simple relation:

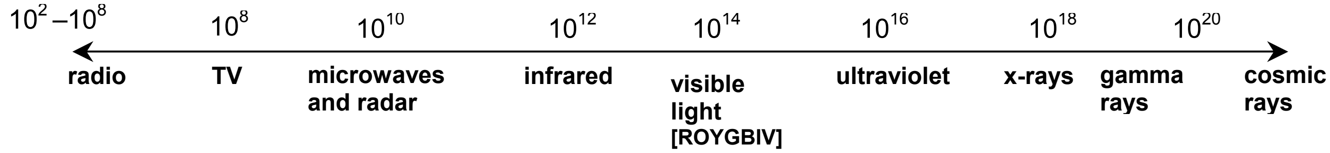
$\lambda = cT = c/f$

(When e-m waves travel through material substances, their speed is less than c; this can cause their direction of travel to change when they pass from one material to another. We'll discuss that in the next chapter.)

The "Spectrum" of Electromagnetic Waves: Although all e-m waves travel through a vacuum with the speed *c*, the frequency (and so also the wavelength) of the wave will depend on the physical process that produced it in the first place. Typically, e-m waves are produced by oscillating electric charges; it is the oscillation frequency of these charges that determines the frequency of the e-m waves they produce. There is an enormous range of different frequencies in which e-m waves are observed to occur. One way to represent this range is by a "spectrum" diagram such as the one below:

Spectrum of Electromagnetic Waves

approximate frequencies in Hz ("oscillations per second")



There is no sharp dividing line between the different types of e-m waves. Each type covers a range of frequencies and the "high end" of one type is hard to distinguish from the "low end" of the next type. Radio waves, in particular, are produced over an extremely broad range of frequencies. The different colors of visible light correspond to e-m waves with different frequencies. The lowest frequency of visible light is red, then orange, yellow, green, blue and indigo, with violet as the highest frequency (indicated on the diagram by the abbreviations "ROYGBIV").

Each type of wave is typically produced by a different source. The higher frequency waves must be produced by charges oscillating back and forth very rapidly – so rapidly, the current cannot travel very far during each oscillation period. For that reason, the ***higher*** frequency waves are typically produced by charges that oscillate over ***smaller*** distances.

Radio Waves are the lowest frequency e-m waves; they can be produced by charges oscillating through wires or conductors in ordinary electric circuits. The distances traveled by the current during one period ranges from a few meters up to many thousands of meters. The sizes of the transmitting ***antennas*** (where the waves are actually produced) are correspondingly large. The very largest antennas that have been constructed are more than 50 miles long! These are used to produce extremely low frequency radio waves to communicate with submerged submarines. (Higher frequency e-m waves are unable to penetrate below the water surface.)

Microwaves and Radar are also produced by electronic circuits, but with antennas that are typically around one meter or less in size.

Infrared waves are too high in frequency to be produced by electronic circuits. Instead, they are typically produced by electric charges oscillating in atoms and molecules. These may be due to vibrational or rotational motions of the molecules, and are related to the temperature of the material. (Higher temperatures correspond to more rapid oscillations, and therefore higher frequency waves.)

Visible light may also be produced by vibrational oscillations of atoms and molecules. In addition, visible light may be produced when electrons in atoms are rearranged into different "energy levels." (Electrons in atoms have certain specific energies; when they make a "transition" from one energy to another, they may emit or absorb e-m waves. This will be discussed in Chapter 13.)

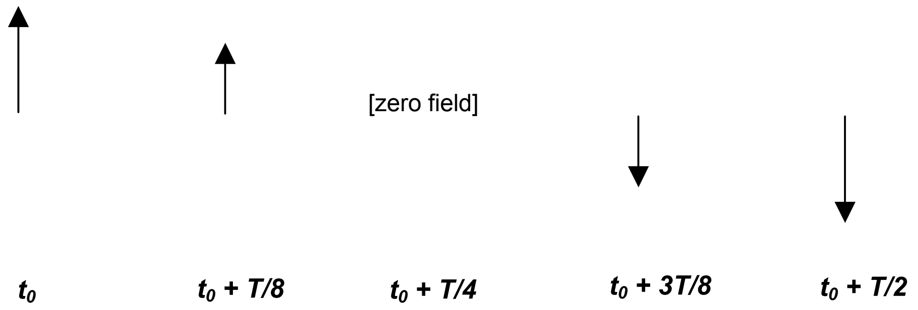
Ultraviolet light is also produced by electronic transitions in atoms. It may also be produced by ***very*** high temperature materials (such as the sun) through vibrational motions of atoms contained in the material.

X-rays may be produced by certain high-energy electronic transitions in heavier atoms (Heavier atoms have larger numbers of electrons, including ones with very high energies.) X-rays may also be produced by the rapid ***deceleration*** of high-energy electrons that occurs when they are suddenly brought to a halt by striking a metal target.

Gamma rays have such high frequencies that they are produced in the very tiny nucleus of the atom. There, electric charges may be pictured as oscillating extremely rapidly and have very high energy levels.

Cosmic rays are very high frequency gamma rays, and are produced by sources outside the solar system. The exact origin of these highest-energy e-m waves is not completely understood.

Polarization of electromagnetic waves. Let us consider how the e-m wave would appear to an observer who sees the wave approaching "head-on." As the wave approaches the observer, the electric field is seen to vary both in direction and magnitude. At the initial moment shown here (called "t_0"), the electric field is pointing "up" with its maximum magnitude; a short time later (after half of one period T has elapsed), the electric field is pointing "down" with the same magnitude. In between those times, the magnitude of the electric field would vary as shown here:



As time went on, this pattern would repeat. (The magnetic field would be in the plane of the page, and it would point perpendicular to the electric field.)

We see that the electric field oscillates up and down along a single axis; this corresponds to the vibrational motion of the charged particles that produced the e-m wave in the first place. In the case shown here, we would say that the electric field is "polarized" along the vertical axis.

If this e-m wave were a radio wave it could be detected by, for instance, a straight wire acting as an antenna. If the wire were aligned along the vertical axis ("up and down"), the electrons in the antenna would be forced into oscillatory motion as the electric field in the e-m wave passed by. A varying current would run up and down the length of the antenna, and could then be detected by a properly designed electronic circuit connected to the end of the antenna. However, if the wire antenna were aligned along the horizontal axis ("left-right"), the electrons in the antenna would oscillate along the narrow width of the wire only, and would not form a current that could flow along the length of the antenna through a detection circuit.

In the case of visible light emitted by atoms and molecules, the polarization axis of each individual atomic source is usually randomly oriented with respect to the others. Therefore light emitted (or reflected) from objects is normally "unpolarized," meaning that it is a mixture of e-m waves with every possible polarization axis. However, certain materials have been manufactured that only transmit light with ***one*** polarization axis. (This is due to the way the molecules are aligned in the material.) If one looks at an object through such a "polarizing" material, it appears slightly darkened, but otherwise completely normal. However, if ***another*** piece of polarizing material is interposed between the first polarizer and the observer, a startling effect can be observed. Depending on the orientation of the two polarizers, the object may appear normal, or noticeably darkened, or not at all! If the second polarizer is oriented so as to permit, for instance, only ***horizontally*** polarized light to pass through it, and the first polarizer is set to transmit only ***vertically*** polarized light – no light ends up getting through to the observer!

Light that is reflected from a surface is "partially" polarized, which means that it contains a disproportionate amount of one particular polarization component. For this reason, when the object is viewed through polarizing filters (such as those contained in polarizing sunglasses), the amount of light reflected off the surface may appear to be reduced (depending on the precise orientation of the filter). This can cut down the effect of glare when light strikes a highly reflective surface such as metal or water.

Interference of electromagnetic waves. When e-m waves produced by different sources are brought together, the "net" wave that results will depend on the vector sum of all of the electric and magnetic field vectors. If the different sources have the same frequency it may be possible – by careful arrangement of the paths traveled by waves from the different sources – to create regions in which the ***net*** field is always zero. This can happen if the maximum "upward" vector of one source is always located at the same point as the maximum "downward" vector of a second source. The "interference" patterns that can be produced in this way are explored in the worksheets.