1

Light propagation

1.1 Background history

The history of optics is filled with examples of unique uses and situations that are beyond the scope of this book which is intended for the student on a first year optics course. However, a brief review is necessary to show how man has been trying to understand and describe light over the last 2500 years.

The word "optics" originated in a book on visual perception written by Euclid some 2000 years ago. Euclid developed geometrical theories to account for the observation of images by mirrors. Some names that come to mind in the history of optics are Ptolemy, Bacon, Brahe, Kepler, and more recently Newton, Huygens, Fermat, Young, and Einstein.

There is a story by Archimedes (212 BC) that the Greeks defended Syracuse (in modern-day Sicily) from the Roman fleet by reflecting sunlight with the soldiers' shields and burning the ships' sails by focusing the intense heat of the Sun's rays.

Muslims in the thirteenth century were purported to have the ability to create a burning mirror to use for burning cities (in the Holy Land). Roger Bacon, a monk under Pope Clement IV, was motivated by this threat to study optics as a weapon of war. He developed similar devices for the Christian crusaders battling the Muslims.

Ptolemy of Alexandria, a Greek from Egypt (about AD 190), knew that two transparent substances, glass and water, had indices of refraction of 3/2 and 4/3, respectively. These were calculated by casting shadows of objects illuminated by the Sun into water and glass.

Ibn Al-Haytham of Cairo (Khan, 2007) made probably the most precise measurements of the index of refraction in the tenth century, while Europe was still in the Dark Ages. His scientific experiments were the best to that date, and were marveled at until modern times.

Willebrord Snell in the seventeenth century empirically wrote the refraction law that bears his name, but could not explain the relationship, because light was thought to be composed of corpuscular particles. This confusion gave rise to many explanations of why light bent toward the normal of a surface in a denser medium.

Isaac Newton's contributions to the dispersion of light through a prism were some of the greatest and yet worst work in optics. He did this work before the age of 26, and was thrown out of London's Science Academy for his revolutionary approach to optics. He found light to be composed of many colors, and made a prism system to display these colors. The fact that he had incorrectly concluded that white light could not be focused with glass, due to dispersion, set back the development of the achromat doublet by many decades. This delay in the development of the achromat was a consequence of his great dominance in the field of optics, with the widely held assumption that if Newton said it, it must be so!

Modern optics is driven by optical systems modeled after the human eye. In fact, it might turn out that the modeling and copying of the human eye's functions may have been carried too far. In most systems of detection, the optical configuration for forming images and the associated signal processing techniques all mimic the human eye. Although geometrical optical systems typically mimic the eye as an imaging system, other optical systems in nature do not. These include compound eyes, polarization sensing eyes, as well as color sensing in the insect world.

This book covers only the geometrical aspects of optics, which can be thought of as the lowest level in the hierarchy of optics. The assumptions that light travels in a straight line and that all equations are linear will be held throughout this textbook. We will consider optics to be mainly confined to light radiation that is detectable by the human eye (i.e. the visible spectrum as opposed to the entire electromagnetic spectrum).

1.2 Nature of light

What is light? That question can be very difficult to answer. In fact, throughout the ages, optics theory has bounced back and forth between corpuscular quanta and wave models. Present-day scientists' description of light depends on the application on which they are working. The formation of light from heat, e.g. from a fire or an incandescent light bulb, is described as being due to excited atoms, and this can be explained in classical terms.

The atoms which are thermally excited have electrons which are "bumped" into higher energy orbits from which they decay to lower energy orbits. During this process, the electrons release a quantum of radiation with a frequency (υ)

of radiation in proportion to the energy (E) released. The atoms of the substance being heated have a large range of discrete energy level orbitals, and the decay from these orbitals to lower energy states releases a continuum of energies. However, the planet model of the atom is not complete, because "strictly speaking," since the electrons use energy when orbiting the nucleus, the atom should eventually collapse.

The relationship between energy and the frequency of a quantum of light emitted is

$$E = h\upsilon, \tag{1.1}$$

where h is Planck's constant and υ is the frequency of light.

Light propagates into space whenever a charged particle is accelerated or decelerated. A common example is the X-ray machine used in radiology departments, in which a beam of electrons is focused onto an anode with high voltage (50 kV). When the electrons are stopped at the anode, X-rays are emitted at the speed of light.

Light bulbs give off a continuum of visible light because of the many different electron energy levels decaying to a continuum of lower energy levels, thus, energy quanta of many frequencies are emitted. The heated tungsten filament produces white light, which is light that contains all frequencies of light from zero to infinity. However, there is radiation being given off at frequencies above and below that which the human eye can detect. The electromagnetic spectrum has been classically divided up into regions by energy level; however, the exact dividing points are not well defined. The main spectral regions of interest are shown in Table 1.1.

Table 1.1. *Various common names of spectral regions, with the approximate center frequency.*

Wave name	υ cycles s^{-1}
Gamma rays	$\sim 3\ (10^{24})$
X-rays	$3\ (10^{16})$
Ultraviolet	$8\ (10^{14})$
Visible	$6\ (10^{14})$
Infrared	$3\ (10^{12})$
Microwave	$3\ (10^{11})$
UHF	$3\ (10^{8})$
VHF	$3\ (10^{8})$
FM	10^{8}
AM	10^{6}
Audio	10^{4}

The visible spectrum, to which we humans respond, is between 4 (10^{14}) and 7.5 (10^{14}) Hz. Monochromatic light, such as that from a laser, has a center frequency with a very narrow bandwidth. For example, a HeNe laser has a center frequency of 4.74 (10^{14}) Hz, while a laser diode (InGaAs) has a frequency of 4.47 (10^{14}) Hz.

Example 1.1

What is the energy of a photon from a laser diode of frequency 4.47(10^{14}) Hz?

$$E = h\upsilon = 6.6\left(10^{-34}\right) \times 4.47\left(10^{14}\right) \text{ joules}$$

$$E = 2.95\left(10^{-19}\right) \text{ joules} = \frac{2.95\left(10^{-19}\right) \text{ joules}}{1.6(10^{-19}) \text{ coulombs}} = 1.7 \text{ electron volts.}$$

The previous discussion assumes light to be made up of particles or quanta of energy ($h\upsilon$). An alternative approach is to consider light as an electromagnetic (EM) wave. From many physics observations, it is concluded that whenever an electric charge is accelerated, a wave is emitted (similar to our photon model). The waves that are formed consist of electric and magnetic fields that propagate at the speed of light.

James Maxwell logically coined the term “electromagnetic waves.” An EM wave is a self-propagating wave consisting of electric and magnetic fields fluctuating together. Maxwell developed equations describing these EM waves, and derived the wave equation, which is an expression that describes their propagation.

Maxwell’s equations also predicted how fast these waves would move, i.e. their velocity. He found that the velocity is dependent on two constants of the medium (permittivity and permeability), and that the velocity itself is also a constant – a revolutionary conclusion. Maxwell discovered that light, in fact all electromagnetic radiation, produces an electrical field that travels at a constant velocity, in air, of about 3 (10^8) m s^{-1}. A changing electric field (**E**) induces a changing magnetic field (**H**), as shown in Figure 1.1.

Most waves encountered in nature, e.g. water waves, propagate in a medium; however, EM waves can also propagate in a vacuum. The EM wave keeps itself going through its own internal mechanism, so once launched, the EM wave no longer depends on its source, the accelerated charge, and propagates in a straight line in a homogeneous medium. It propagates on its own, and carries some characteristics of the source which generated it.

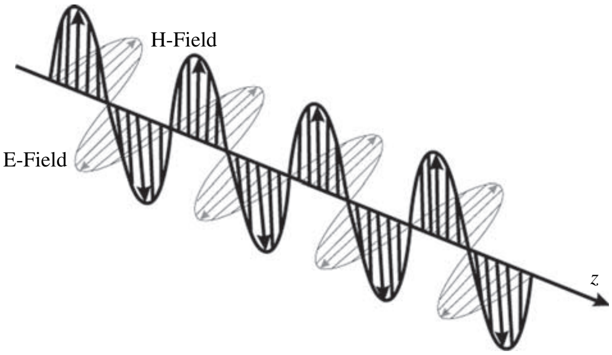

Figure 1.1 A diagram of an EM wave.

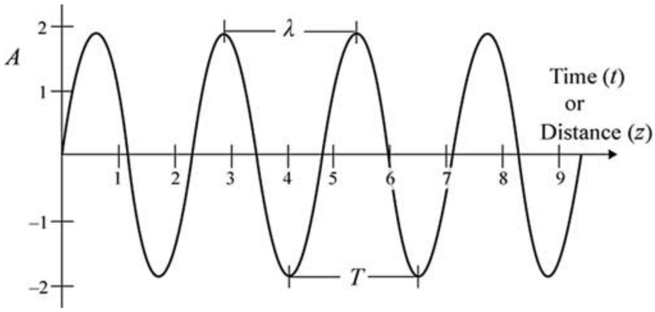

Figure 1.2 A sinusoidal EM wave plotted in time or distance.

Mathematically, we represent light as a sinusoidal electric field propagating through time (t) and space (z) as

$$|\mathbf{E}(z, t)| = A \sin(kz - ct), \tag{1.2}$$

where A is the amplitude, k is the wave number, z is the axial distance, c is the speed of light and t is the time. Figure 1.2 shows a sine wave plotted from Equation (1.2) in time (t) or distance (z). Pick a position in space (z-fixed), and watch the light wave as it passes this position. The wave would be modulated in a sinusoidal way, as shown in time (t).

For a fixed position in space (z constant), the amplitude of the light wave varies sinusoidally with t. For an instantaneous time t (a snapshot in time), the light wave's intensity would be sinusoidal over space (i.e. in the z variable). Either variable is correctly modeled as a sine wave.

Recall that the frequency is constant for a given monochromatic light source. However, the velocity of the wave may change as we propagate through different media. The speed of light in a vacuum (free space) is approximately $3\,(10^8)$ m s^{-1}

or 186 287 miles per second. We will follow the nearly universal convention of representing the vacuum velocity of light as "*c*," which is believed to come from the Latin word *celeritas* (speed).

The mathematical representation of a wave, shown in Equation (1.2), is sinusoidal in two variables, time and distance. If the wave is plotted versus time (t), one cycle is the time period T, and if it is plotted versus distance (z), also shown in Figure 1.2, one cycle is the wavelength (λ) of that EM wave. The frequency (υ) of the wave is the reciprocal of the period:

$$\upsilon = 1/T. \tag{1.3}$$

The velocity of the EM wave, $3\ (10^8)\ \mathrm{m\,s^{-1}}$, is the distance it travels in one period (λ) divided by the time it takes to move one period (T), so

$$c = \lambda/T, \tag{1.4}$$

which can be rewritten in terms of frequency using Equation (1.3),

$$c = \lambda\upsilon. \tag{1.5}$$

The velocity of light in free space is considered the fastest velocity known. Sunlight takes about 8 minutes to reach the Earth from the sun. Light could travel between Los Angeles and New York about 62 times in a second. In a homogeneous medium, light travels in straight lines called rays. This ray concept is a fundamental description of light, albeit one which oversimplifies what is really propagating.

Example 1.2

Find the velocity of light for a laser diode that has a frequency of $4.47(10^{14})$ Hz and a wavelength of 670 nm (red).

$$\upsilon = 4.47\left(10^{14}\right)\ \text{hertz};$$
$$\lambda = 670\ \text{nm (red)},$$
$$c = 4.47\left(10^{14}\right)(670)\left(10^{-9}\right) \approx 3\left(10^{8}\right)\ \mathrm{m\,s^{-1}}.$$

The visible spectrum of light lies between 400 nm and 700 nm in wavelength. The convention is to define that spectrum in terms of wavelength; however, the description in terms of frequency should be used. This is because the frequency of light does not change once generated.

Very crudely, we can divide the visible region of the electromagnetic spectrum into three parts: red, yellow, and blue (600–700 nm, 500–600 nm, and

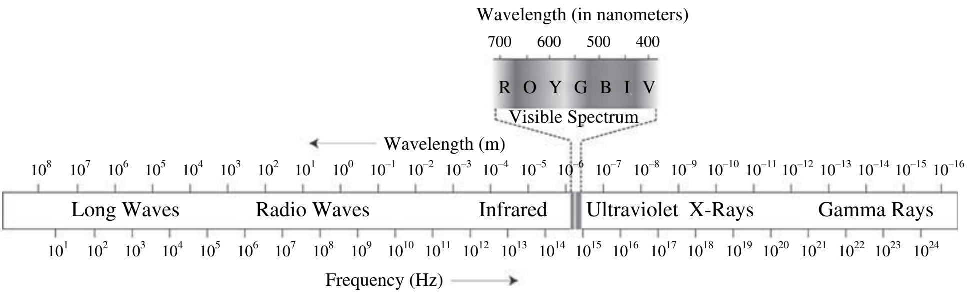

Figure 1.3 The EM spectrum.

400–500 nm, respectively) to represent human color sensitivity. (We could explain this sensitivity, as per Charles Darwin, by noting the need for our ancestors to know when the bananas were ripe.)

In this visible range, red light has the longest wavelength (lowest frequency) of 700 nm or a frequency of 4.3 (10^{14}) Hz. Red is the least energetic region of the visible spectrum. If we assume an atom has a diameter of about 0.65 nm, then the red wavelength is about the length of 1000 atoms laid side by side. Blue/violet on the other hand, which has the shortest wavelength (400 nm) and is the most energetic light wave to which the eye responds, has a wavelength corresponding to about 600 atoms.

When all the wavelengths of light are present for the entire spectrum (400–700 nm), white light is observed by humans. The wavelengths between 400 and 700 nm each form one of the colors red, orange, yellow, green, blue, indigo, and violet, making rainbows visible in the sky to humans. The visible spectrum is a very small part of the electromagnetic spectrum as seen in Figure 1.3.

A key point here, used throughout the book, is that these EM waves are modeled geometrically as straight lines, thus producing rays for the study of geometrical optics. The field of geometrical optics manipulates these rays to form images and illuminations or to transfer information.

A wave is produced by an accelerating charge, such as an electron in an atom changing energy levels. This gives a quantum of energy, $h\upsilon$, so very simply, one would conclude that each EM wave would have an energy of $h\upsilon$. Here are the three main points of our conception of EM waves so far:

- Atoms give off photons of energy ($h\upsilon_i$).
- Each wave can be thought of as an EM ray with energy of $h\upsilon_i$. This is wrong, but conceptually acceptable at this point.
- Many photons (N_i) give off $\Sigma_i N_i h\upsilon_i$ energy. Conceptually, this is sufficient for a very preliminary observation, but again is wrong. We can think of light as traveling in waves, with each wave having the energy $Nh\upsilon$, where N is the number of photons in the wave.

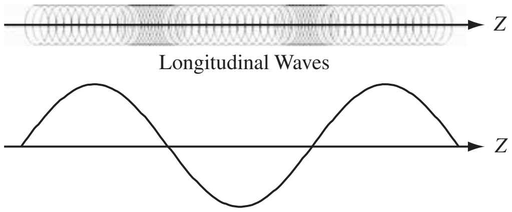

Figure 1.4 Wave types.

There are two types of waves: transverse waves, also known as EM waves, which have been discussed in this chapter, and longitudinal waves. A pictorial representation of these wave types is shown in Figure 1.4

Longitudinal waves need a medium in which to propagate, and without such a medium, their energy is lost. Sound waves are an example of longitudinal waves. Note that sound waves, seismic waves, and other kinds of waves that require matter in which to propagate travel much slower than the speed of light.

In the case of EM waves, the energy is contained in the electric and magnetic fields, which can exist in a vacuum. In fact, they propagate fastest in a vacuum. In other media, the velocity is less than the speed of light in a vacuum because the atoms making up the material are excited and relaxed, slowing the fields. The energy propagates perpendicular to the $\mathbf{E} \times \mathbf{H}$ field direction, as shown in Figure 1.1.

1.3 Wavefronts and rays

Geometrical optics represents the EM wave as a vector pointing in the direction of propagation: a straight line representation, called a ray. This model is somewhat misleading and incorrect, but for the most part, the ray model may be used in the context of geometrical optics to produce useful results.

There are two types of radiation sources in geometrical optics: point and extended. A point source, such as for starlight, may be thought of as a source from which rays emanate in all directions. See Figure 1.5. The rays are actually propagating into 4π steradians, or into three-dimensional space. (Solid geometry will be discussed later.) The ray is simply the path followed by a single photon of light, or an imaginary line drawn in the direction the wave is traveling.

In a homogeneous isotropic medium, the ray paths are straight lines which have varying amplitudes of both the **E**- and **H**-fields, as shown in Figure 1.1. If we connect all the peak values of the EM wave that are a distance of 100 peaks

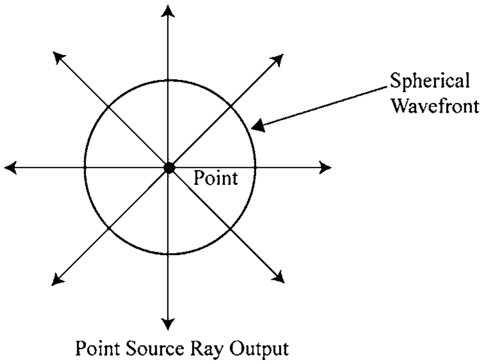

Figure 1.5 Rays propagating from a point source of radiation.

from the source, we would produce a spherical surface (see Figure 1.5). The points on this spherical surface would be at equal distances from the point source, and these distances would equal the radius of the sphere. The rays are the radii of the spherical wavefronts, and are perpendicular to these wavefronts. One may think of the wavefronts as being at each crest or trough of the EM wave which is emanating from a point source. The expression for a spherical wave is:

$$\frac{A}{r}\mathrm{e}^{\mathrm{i}\varphi}\mathrm{e}^{\mathrm{i}(\mathbf{k}\bullet\mathbf{r}-\omega t)}, \tag{1.6}$$

where $k = 2\pi/\lambda$ is the wave number, φ is the phase, $\omega = 2\pi/T$, and r is the distance from the point source to the wavefront.

At a boundary of two homogeneous media (such as air and glass), the ray direction changes suddenly, but the ray remains a straight line in each medium. However, if the medium were not homogeneous, e.g. it is like our atmosphere in which the density changes with altitude, the ray would bend continuously.

Even if the rays are changing direction, the wavefront is always perpendicular to the ray. A combination of the rays or the sum of several rays forms a beam of light such as a search light, which is represented by many rays.

A wavefront is, therefore, a set of points with equal phase located at regular intervals from the source of light. Phase is the relationship of the sinusoidal period of the EM wave. A wave emanates from a point source in all directions as a spherical wavefront, centered at the source, as shown in Figure 1.5. It is important to note that the optical path length relative to the source is constant over the wavefront.

As this EM wave propagates (at 186 282 miles per second), the spherical wavefront becomes a plane surface at large distances. Thus, in this case, we have a series of plane waves (see Figure 1.6).

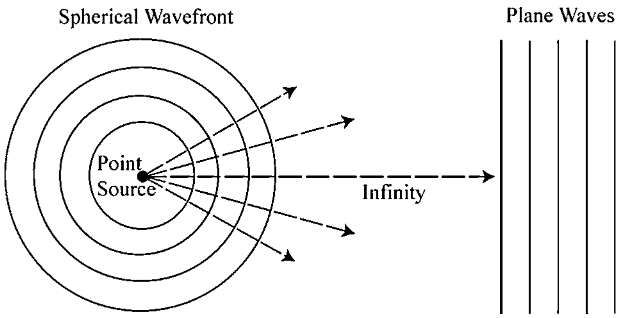

Figure 1.6 Spherical wavefronts emanating from a point source which become plane waves as the radius becomes infinite.

1.4 Index of refraction

When a charged particle is accelerated it emits EM radiation. EM radiation is described according to Maxwell's equations (Maxwell, 1865), which are beyond the scope of this book. If this radiation has the correct energy, it will be in the visible spectrum (light as we know it). These light waves follow descriptions derived by Maxwell in his equations of EM light waves for a time-varying field (electric or magnetic) (Born and Wolf, 1959). The result, after some minor manipulation, is the wave equation for EM light waves in a charge-free homogeneous medium:

$$\nabla^2 \mathbf{E} - \mu_\mathrm{m}\, \varepsilon_\mathrm{m} \frac{\partial \mathbf{E}}{\partial t} = 0, \tag{1.7}$$

where $\mathbf{E}$ is the time varying electric field, μ_m is the permeability of the medium, and ε_m is the permittivity of the medium

The corresponding speed of light in the medium is

$$v_\mathrm{m} = \frac{1}{\sqrt{\mu_\mathrm{m}\, \varepsilon_\mathrm{m}}}. \tag{1.8}$$

For the case of free space (vacuum), the permeability and permittivity are well known:

$$\mu_0 = 4\pi(10^{-7})\,\mathrm{N\,s^2\,C^{-2}},$$

$$\varepsilon_0 = 8.85(10^{-12})\,\mathrm{C^2\,N^{-1}m^{-2}},$$

which gives (using Equation (1.8)) the speed of light (c) as $2.99792458 \times 10^8\ \mathrm{m\,s^{-1}}$ in free space. The speed of light is most often approximated to

$$c \approx 3(10^8)\,\mathrm{m\,s^{-1}}. \tag{1.9}$$