Electromagnetic Waves and Lasers

Wayne D Kimura

Chapter 1

What are electromagnetic waves?

1.1 Electromagnetic (EM) radiation

Electromagnetic radiation refers to a type of energy that is able to propagate through space. If this space is a vacuum, then this radiation travels at the speed of light, i.e., approximately 3×10^8 m s^{-1}. As we will explain later, this radiation is characterized as having both electric and magnetic fields associated with it, which is why this radiation is referred to as 'electromagnetic'.

The concept of waves is something we are all familiar with from walking along the ocean beach and watching the waves breaking on the sand. It is readily apparent there are many different kinds of water waves—fast ones, slow ones; ripples on a pond, giant tsunamis. Electromagnetic radiation shares many of the same traits as water waves; thus, it is highly convenient to combine the two concepts into the theory of electromagnetic waves (or EM waves for short).

This book will provide a broad overview of EM waves, both their theory and practical applications, with a special emphasis on lasers. This will not only give the reader tangible examples of how the theory is manifested in real life, but also practical knowledge about lasers and their operation and usage. The latter can be useful to those involved with using lasers. As a short treatise on this subject matter, this book is not intended to delve deeply into the details of EM waves or lasers. A bibliography is provided at the end of each chapter of this book and referenced throughout it for those who wish to explore in more depth the topics covered in this book. Rather, the aim of this book is to provide a quick synopsis, which will allow the reader to gain a competent general understanding of EM waves and lasers.

While lasers are rather specialized devices and are becoming more common place (e.g., laser pointers), EM waves permeate our everyday lives, 24/7, and are literally penetrating your body as you are reading this book. In order to gain an appreciation of the ubiquity of EM waves, it is helpful to first define two of their most fundamental characteristics—wavelength, represented by λ, and frequency, represented by ν or f. Figure 1.1 depicts a sinusoidal wave, which can represent any wave

doi:10.1088/978-1-6817-4613-5ch1

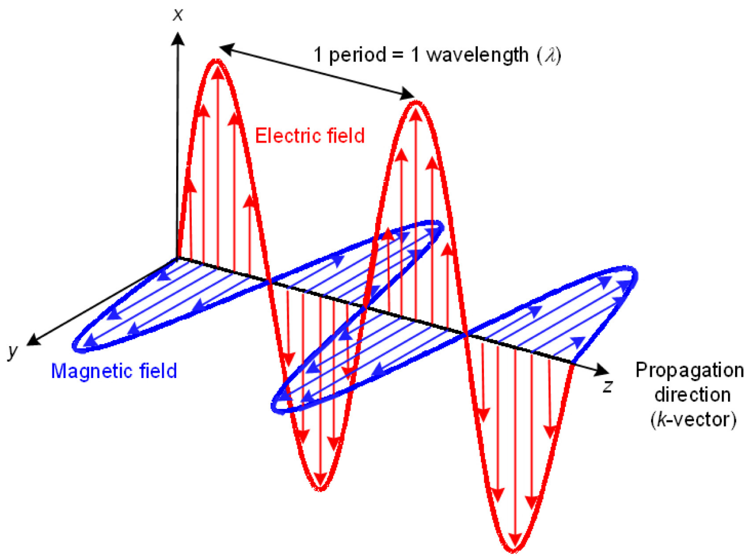


Figure 1.1. Schematic representation of an EM wave.

in general (e.g., EM, sound, water), where the z-axis denotes the direction of propagation of the wave and typically has the units of either time or position. As explained later, this direction of propagation is often indicated by the so-called k-vector of the wave. EM waves consist of both oscillating electric and magnetic field components, which are plotted along the x-axis and y-axis in figure 1.1, respectively. Note that the magnetic field is oriented orthogonal to the electric field.

The distance between two adjacent peaks or two adjacent valleys in figure 1.1 is the wavelength of the wave λ in units of length if the z-axis is distance, or the period of the wave T in units of time if the z-axis is time. How many times per second the wave oscillates is the frequency of the wave ν, and this frequency depends on how fast the wave is moving, i.e., its velocity v, where $v = \nu\lambda$. Unfortunately, the characters v and ν look almost the same, so this can be quite confusing. Fortunately, when dealing with EM waves, we typically use the velocity of light in a vacuum, denoted by $c \approx 3 \times 10^8$ m s^{-1}, as our reference velocity. Therefore, the frequency of the EM wave can be expressed as $\nu = c/\lambda$. Since c is a large number, this implies that for small values of λ, ν can be huge; or, conversely, for low values of ν, λ can be huge. As we will show shortly, in table 1.1, the wavelengths for EM waves can range from 10^{-11} m to thousands of kilometers. That is over 19 orders of magnitude!

As mentioned, the z-axis in figure 1.1 can have units of time or position. In the latter form, it is easier to visualize the concept of water waves moving along, say, a wharf where distance represents the position along the wharf. However, if you were standing still on the wharf, then the waves would be passing by you as a function of time. Note that in either case, the wave is the same; it is just how it is being plotted that is different. When characterizing EM waves, you will often encounter plots where the abscissa is either time or distance depending on what is being discussed.

We can now review the tremendously wide range, or spectrum, of EM radiation. Table 1.1 summarizes the EM spectrum as characterized by wavelength, the name of the spectral range, example of sources or applications, and, where applicable, typical lasers that emit light within that part of the spectrum. Some lasers emit light at only one wavelength while others do so over a range of wavelengths. Therefore, table 1.1 is only meant to provide a general idea where various lasers fit within the EM spectrum.

The shortest EM waves in table 1.1 are gamma rays (γ-rays) with wavelengths less than the size of an atom. As explained later, the energy in the EM wave is inversely proportional to λ; hence, γ-rays are one of the most energetic forms of EM radiation. Radioactive elements commonly emit γ-rays when they decay into smaller elements. Since γ-rays more easily pass through objects than x-rays, detecting γ-rays is one way of finding hidden, clandestine nuclear bomb materials [1].

Hard x-rays are commonly used in medicine and for security screening. A relatively easy way to generate hard x-rays is by bombarding a target, such as tungsten, with high-energy electrons to create x-rays via a process called bremsstrahlung [2]. Since electrons have an electric field, it is possible to have them emit EM radiation. In the case of bremsstrahlung, this radiation is emitted when the electrons are suddenly decelerated by collisions with the tungsten atoms. As we shall see later, there are other ways electrons can be used to create EM radiation.

Soft x-rays have less energy than hard x-rays and consequently they do not penetrate materials as easily. However, sometimes this can be a good trait because, unlike hard x-rays, which are energetic enough to disrupt molecules such as DNA, soft x-rays can more gently probe these molecules without damaging them. Although there are soft x-ray sources, these tend to have limited output power. This situation is beginning to change as x-ray lasers are now being realized [3, 4]. However, as we shall explain later when discussing how lasers work, good resonator mirrors are always needed, but these are more difficult to fabricate at x-ray wavelengths because the x-rays want to pass through the mirrors rather than reflect off them!

Ultraviolet (UV) light from the sun is responsible for tanning of our skin, but UV light is also important for photolithography where the short wavelength of UV light enables creating tiny features on integrated circuits [5]. As we will show later, the resolution of optical instruments, such as photolithography machines, is directly related to the wavelength of the EM radiation. Excimer lasers [6, 7] are a powerful source of UV light and are commonly used as the light source for photolithography.

As you probably remember from your early school days, the visible portion of the EM spectrum is one of the narrowest, and yet it fills our lives with so many different colors! And each of these colors are simply slightly different wavelengths of the light, e.g., blue is around 475 nm, green is about 510 nm, yellow is around 570 nm, orange is about 590 nm and red is around 650 nm. What this implies is if we could 'see' the colors of the rest of the EM spectrum, which are much broader than the visible spectrum, we would be inundated with an unbelievable number of different colors. This is why the rest of the EM spectrum is so useful because each one of those colors can be used for different purposes, just as the colors mixed on a painter's palette can

Table 1.1. The EM spectrum and example of laser sources.

Name of EM wave	Typical wavelength	Example of sources or applications	Example of lasers (see text for details)
Gamma rays	10^{-11} m	Radioactive materials	
Hard x-rays	10^{-10} m (1 Å)	Medical x-rays, airport security	
Soft x-rays	10^{-9} m (1 nm)	x-ray microscopy and holography	Plasma-based x-ray lasers
Ultraviolet (UV)	10^{-7} m (100 nm)	Tanning booths, black lights	Excimer
Visible	400–700 nm	LEDs, light bulbs	HeNe, dye, ion, SHG, semiconductor, OPO
Near infrared (NIR)	10^{-6} m (1 μm)	Fiber-optic communication	Nd:YAG, Ti-sapphire, semiconductor, fiber, OPO, chemical
Mid-infrared (MIR)	10^{-5} m (10 μm)	Clothes iron	CO_2, Raman scattering
Far-infrared (FIR)	10^{-4} m (100 μm)	Long wave or dark heaters	CO_2-laser-pumped methanol
Terahertz (THz)	10^{-3} m (1 mm)	Airport security scanners	FEL
Microwaves	10^{-2} m (1 cm)	Radar, microwave ovens	
Radio (UHF)	10 cm	Cell phones, GPS, Wi-Fi, Bluetooth	
Radio (VHF)	1 m	TV and FM broadcasting, radio-controlled devices, satellite communication,	
Radio (LF)	10^{3} m (1 km)	AM broadcasting, navigation	
Radio (ULF)	10^{5} m (100 km)	Magnetosphere, earth-mode communication	
Radio (SLF/ELF)	10^{7}–10^{8} m	Submarine communication, 60 Hz electric power grid	

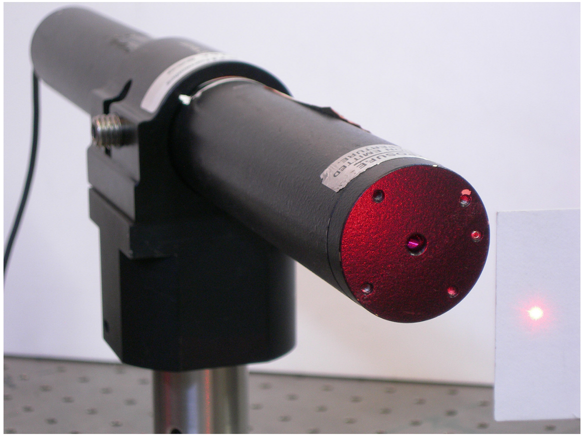


Figure 1.2. Photograph of HeNe laser emitting light at 632.8 nm. Its red laser beam is exiting the hole in the center of the cylinder and striking a white card. The red light on the card is so bright, the camera is saturated and displays the image as a white spot.

create beautiful paintings with many nuances of colors—all from just the short band in the visible spectrum.

There are also a large number of different lasers that emit light in the visible spectrum. The most common ones are those emitting red light in bar code scanners, the red or green light of laser pointers, and the ones used in laser light shows. More powerful visible lasers include helium–neon (HeNe) (λ = 632.8 nm) [8], and argon ion (λ = 488.0 nm and 514.5 nm) [8]. Figure 1.2 shows a photograph of a commercial HeNe laser. As explained later, through a mechanism called second harmonic generation (SHG), it is possible to take laser light emitted at a longer wavelength, such as by a Nd:YAG laser that emits near-infrared light at 1.06 μm, and convert it into visible laser light at 532.0 nm.

Examples of near-infrared (NIR) sources include your TV remote control and the semiconductor lasers used for fiber optic telecommunication [6, 9, 10]. Lasers that emit light in the NIR include Nd:YAG [6, 11], which can be used for cutting materials; Ti:Sapphire [12] and optical parametric oscillators (OPO) [13, 14], which are highly tunable over different NIR wavelengths making them useful for scientific research; semiconductor lasers (e.g., AlGaAs, InGaAs, and InGaAsP) [6, 9, 10], which are used in CD and DVD players; and chemical lasers (HF/DF, chemical oxygen iodine laser (COIL)) [6, 15, 16], which are lasers powered by chemical reactions and not electricity.

While we cannot see mid-infrared (MIR) light, we can sense it as heat. Thus, a clothes iron is a great source of MIR light, as well as our bodies; infrared-sensitive surveillance video cameras are actually detecting the MIR radiation emitted from our warm bodies. One of the most common types of MIR lasers used in industrial processing are carbon dioxide (CO_2) lasers (λ = 10.6 μm) [6, 8, 16]. For the same reason CO_2 is a greenhouse gas that absorbs MIR radiation, it is also a great emitter of MIR radiation. These lasers are capable of emitting kilowatts of power for cutting and welding materials. We mentioned it is possible to convert NIR laser light into shorter visible light using SHG. Using a different process, called Raman scattering (see section 1.3.8), it is possible to convert laser light into longer wavelengths such as MIR [17].

So-called ‘dark heaters’ are sources of far-infrared (FIR) and, thus, we can still sense FIR as heat. One way to generate FIR is by using a CO_2 laser to excite organic molecules, such as methanol, which emits FIR radiation [16].

Terahertz (THz) waves, also called T-waves or mm-waves because the wavelength is literally millimeters, are capable of penetrating certain nonmetal materials, much like x-rays, but are much safer to use because they cannot disrupt molecules [18]. This is because the wavelength of THz light (see section 1.2.4) is so long that the atoms cannot absorb it and, therefore, the THz light passes through the material. This is why they are routinely used in airport scanners for screening passengers. A free electron laser (FEL) can easily generate powerful THz light [6, 19]. An FEL is a rather special type of laser because it is continuously tunable over a wide spectral range from x-rays [20] to THz. Therefore, it could have been listed under x-rays, UV, visible, NIR, MIR, and FIR.

We are all familiar with microwaves because of microwave ovens, but microwaves are also commonly used in radars [21]. High-power microwaves are used to accelerate electrons and protons to very high energies in linear accelerators and storage rings [22]. As an aside, the predecessor to the laser was the maser, which produced a beam of microwave radiation [23].

When the EM radiation has 10 cm and longer wavelengths, we are entering the realm of radio waves [24]. Table 1.1 only gives a sampling of the many radio bands that are available. We utilize radio waves during our daily lives in multiple ways, from cell phones, GPS, and Wi-Fi connections, to watching TV and listening to the radio. These are generally line-of-sight radio transmissions where an unobscured pathway between the transmitter and receiver provides the strongest signal. (It is also possible to receive signals via scattering of the radio wave off objects, but the signal is much weaker.) This is why cell phone towers are so prolific, in order to ensure that cell phone users have a line-of-sight pathway to at least one tower at all times no matter where the user is located. However, when the wavelength becomes huge, i.e., >100 km, then obstructions such as buildings are small compared to the wavelength and the EM wave is able to propagate literally around the obstacles. These extra-long wavelengths can also penetrate through the earth and ocean waters, which is handy for communicating with submerged submarines [25]. Unfortunately, these extra-long wavelengths also correspond to very low frequencies, i.e., a few Hertz, so that data transmission is very slow. Interestingly, one of the largest sources of SLF waves is from our 60 Hz power grid. The national power grid in the US is like a gigantic planetary EM wave source!

1.2 Characterization of EM waves

Besides wavelength and frequency, there are other ways to characterize EM waves [2, 26–28]. The type of EM wave depicted in figure 1.1 is called a transverse-electromagnetic (TEM) wave because the electric and magnetic field components are oriented in the transverse direction relative to the direction of wave propagation. This form of EM wave is one of the most common ones when working with lasers.

The electric field in the EM wave is typically abbreviated as *E*-field and has units of V m^{-1}.[1] The magnetic field is typically abbreviated as *H*-field with units of A m^{-1}; however, you will sometimes see it abbreviated as *B*-field. Unfortunately, this can be confusing because, as explained later, in Maxwell's equations, B represents magnetic flux, i.e., $B = \mu_0 H$, where $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1} is the permeability of free space and B has the units of Webers m^{-2} or, equivalently, Tesla (T). (1 Webers m^{-2} = 10 000 gauss = 1 T.) Thus, the magnetic flux from a magnet can be rightly referred to as the *B*-field; however, strictly speaking, it is incorrect to call the magnetic field component in an EM wave the *B*-field.

Assuming the EM wave is propagating in a vacuum, we can mathematically model the sinusoidal *E*-field and *H*-field in figure 1.1 as

$$E_x(z, t) = E_{\max} \cos\left[2\pi\left(ft - \frac{z}{\lambda}\right)\right] = E_{\max} \cos\left[\omega\left(t - \frac{z}{c}\right)\right], \tag{1.1a}$$

$$H_y(z, t) = H_{\max} \cos\left[2\pi\left(ft - \frac{z}{\lambda}\right)\right] = H_{\max} \cos\left[\omega\left(t - \frac{z}{c}\right)\right], \tag{1.1b}$$

$$E_{\max} = cH_{\max} = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} H_{\max}, \tag{1.1c}$$

where $\omega = 2\pi f$ and $\varepsilon_0 = 8.854 \times 10^{-12}$ farads m^{-1} is the permittivity of free space. This shows explicitly how the fields vary both in time and position, and how the magnitude of the *H*-field is directly linked with the magnitude of the *E*-field. Moreover, in equation (1.1c) we utilize the identity that the speed of light $c = (\mu_0 \varepsilon_0)^{-1/2}$.

Working with trigonometric functions, such as in equations (1.1a) and (1.1b), can be mathematically awkward at times. Hence, you will often see equations (1.1a) and (1.1b) rewritten as

$$E_x(z, t) = E_{\max} e^{j\omega t} e^{jkz} \quad ; \quad H_x(z, t) = H_{\max} e^{j\omega t} e^{jkz}, \tag{1.2}$$

where Euler's identity, $e^{jX} = \cos X + j\sin X$, has been used and it is assumed the real part is taken, i.e., $\mathrm{Re}[e^{kX}]$ is implied but generally not shown in the equation. In addition, in equation (1.2) the wave number $k = 2\pi/\lambda$ is introduced and has typical units of cm^{-1}. This k is the same as the *k*-vector noted earlier and, since it is associated with the *z*-direction, we now understand why the *k*-vector relates to the direction of propagation of the EM wave.

[1] SI units are generally used throughout this book.

1.2.1 Maxwell's equations

All the physics associated with electromagnetism are eloquently expressed in Maxwell's equations [2, 26],

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \text{ (Faraday's Law of Induction)}, \tag{1.3}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}, \text{ (Ampere's Law)}, \tag{1.4}$$

$$\nabla \cdot \boldsymbol{D} = \rho, \text{ (Gauss' Law for Electric Field)}, \tag{1.5}$$

$$\nabla \cdot \boldsymbol{B} = 0, \text{ (Gauss' Law for Magnetic Field)}, \tag{1.6}$$

where variables in bold represent vectors. You will also often find Maxwell's equations expressed in their integral form rather than the derivative (point) form of equations (1.3)–(1.6). For a Cartesian coordinate system with unit vectors $\boldsymbol{u}_x$, $\boldsymbol{u}_y$, and $\boldsymbol{u}_z$ in the x, y, and z directions, respectively, the curl is defined by

$$\nabla \times \boldsymbol{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\boldsymbol{u}_x + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\boldsymbol{u}_y + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\boldsymbol{u}_z, \tag{1.7}$$

and the divergence is defined by

$$\nabla \cdot \boldsymbol{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}. \tag{1.8}$$

Note, the curl yields a vector, which is why there are vector quantities on the right-side of equations (1.3) and (1.4), while the divergence yields a scalar, which is why there are scalar quantities on the right-side of equations (1.5) and (1.6). As mentioned, $\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and magnetic field intensities measured in V m^{-1} and A m^{-1}, respectively; $\boldsymbol{D}$ is the electric flux density or electric displacement measured in coulomb m^{-2}; $\boldsymbol{B}$ is the magnetic flux density or magnetic induction measured in weber m^{-2} or Tesla; and $\boldsymbol{D} = \varepsilon_0 \boldsymbol{E}$ and $\boldsymbol{B} = \mu_0 \boldsymbol{H}$ in a vacuum. $\boldsymbol{J}$ is electric current density and ρ is volume charge density.

Equation (1.3) is Faraday's law of induction, which shows that a time-varying magnetic flux will create an electric field. An electric generator operates on this principle, where moving a magnetic field across wires induces current to flow through the wire. Conversely, equation (1.4) is Ampere's law as amended by Maxwell to include a time-varying displacement current, which indicates that either a current density and/or a time-varying displacement current will create a magnetic field. A classic example of this is that sending current through a wire creates a magnetic field around the wire.

This complementary role of equations (1.3) and (1.4) can be used to help understand why EM waves propagate. For example, a radio antenna has a time-varying current flowing through it created by a transmitter circuit. Through equation (1.4) this generates a time-varying magnetic field that emanates from the antenna. Through equation (1.3) this then creates a time-varying current, which once again generates a time-varying magnetic field, and the process repeats itself. Thus, the resultant EM wave always satisfies Maxwell's equations.

Equations (1.5) and (1.6) are Gauss's law for electric and magnetic fields, respectively. The physical interpretation of equation (1.5) is that electric charge can give rise to electric flux density. Gauss's law comes in handy when analyzing the fields from a distribution of stationary or moving electrons. In equation (1.6), we see the right-hand side is zero. The implication of this is that magnetic charges do not exist, i.e., there are no monopole magnetic sources. Put another way, all magnetic field lines must be closed or terminate on north and south pole points; they cannot simply emanate from a point out to infinity as the electric field does from an electron.

Maxwell's equations are very powerful and are the basis for understanding and mathematically characterizing many aspects of EM wave theory. In keeping with the primary aim of this book to provide the reader with a quick synopsis of EM wave theory, we will not derive these other relationships. Instead we will simply state them and encourage the reader to consult the bibliography for more details.

1.2.2 Helmholtz wave equation

The first key equation that can be directly derived from Maxwell's equations is the Helmholtz wave equation [27], which describes the propagation of an EM wave

$$\nabla^2 \boldsymbol{E} = \mu\varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2}, \tag{1.9}$$

where the Laplacian in Cartesian coordinates is defined as

$$\nabla^2 E = \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2}, \tag{1.10}$$

and $\mu = \mu_r\mu_0$, where μ_r is the relative permeability, and $\varepsilon = \varepsilon_r\varepsilon_0$, where ε_r is the relative permittivity. The relative permittivity and permeability account for the dielectric and magnetic susceptibility properties, respectively, of the medium in which the wave propagates. In that regards, equation (1.9) is valid for an EM wave propagating through any medium. If the medium is a vacuum, then the coefficient on the right-hand side of equation (1.9) becomes $\mu_0\varepsilon_0 = c^2$.

Notice that equation (1.9) is a function of both space (x, y, z) and time (t). By using the technique of separation of variables, it is assumed the solution of equation (1.9) can be expressed as the product of two functions—one dependent on only space and one on only time. This means the Helmholtz equation for only the spatially dependent function becomes

$$(\nabla^2 + k^2)\boldsymbol{E} = 0, \tag{1.11}$$

where k is the wave number introduced earlier. Equation (1.11) is important because it describes how the EM wave can change while traveling through space, for example being focused to a spot or creating interference patterns.

To illustrate how equation (1.11) can be used, when dealing with laser beams propagating in, say, the z-direction, the transverse profile of the beam in the x and y directions generally changes relatively slowly as a function of z. This allows one to make the approximation of ignoring the $\partial^2/\partial z^2$ term in the Laplacian in equation (1.10), which considerably eases solving equation (1.11). This is referred to as the paraxial approximation [29]. The various equations given later in this book assume the paraxial approximation.

Extending this concept further, equation (1.11) assumes another useful approximation in which the E-field components of x and y are only a function of time and z. Hence, the EM wave is 'flat' in the transverse direction and is called a uniform plane wave. Although, as we shall see, a laser beam cannot stay perfectly collimated forever, there are regions along the beam where it is quite collimated and within these regions the light can be treated like a flat plane wave. This makes it easier to analyze the characteristics of the beam and understand effects such as diffraction.

1.2.3 Poynting vector

As another application of Maxwell's equations, we can derive the amount of power flow of an EM wave. The result is the Poynting vector [27, 29]:

$$\boldsymbol{P} = \boldsymbol{E} \times \boldsymbol{H}, \tag{1.12}$$

which represents the instantaneous power density in W m^{-2} that flows in the direction of vector $\boldsymbol{P}$. If we assume an EM wave as described by equation (1.1), then the time-average power density that arises from applying equation (1.12) is

$$P_{\text{average}} = \frac{1}{2}\frac{E_{\text{max}}^2}{\eta_0}, \tag{1.13}$$

where $\eta_0 = [\mu_0/\varepsilon_0]^{1/2} = 377$ ohms is the intrinsic impedance of free space and E_{max} is the peak of the EM electric field. If the wave is propagating through a medium, then the generalized form of the intrinsic impedance, $\eta = [\mu/\varepsilon]^{1/2}$, would be used in equation (1.13).

Equation (1.13) shows that the electric field scales as the square-root of the peak power. In certain applications, it is the electric field of the laser beam that is important. For example, it is the electric field that interacts with the electrons and, if the field is high enough, it can ionize the atom by removing an electron from the atom. In section 1.3.3, we will give an example of how equation (1.13) can be used to estimate the conditions needed to break down air by simply focusing a laser beam using a lens.

1.2.4 Wave-particle duality of EM radiation

We conclude this subsection on the basic characteristics of EM waves by discussing a fundamental dichotomy. As mentioned, using the wave analogy to characterize EM radiation is very convenient, but it does have limitations. One of the most important limitations relates to the wave/particle dual nature of EM radiation [2, 29]. By this we mean that at times EM radiation behaves like waves and at other times it behaves like particles of energy, which we call photons. As an example, atoms can absorb or emit photons. To conceptualize this duality, it can be helpful to think of the photon not as a solid particle but as a small packet of waves with a frequency corresponding to the wavelength of the light. Indeed, the amount of energy in a photon is given by

$$E_{\text{photon}} = h\nu = \hbar\omega, \tag{1.14}$$

where h is Planck's constant (6.626×10^{-34} m^2 kg s^{-1}), ν is the frequency of the wave packet, $\hbar = h/2\pi$, and ω is the radian frequency of the wave packet. This shows that the shorter the wavelength (i.e., the higher the frequency), the higher the energy in the photon. This is why γ-rays are so energetic.

The notion of light consisting of photons also provides some conceptual benefits. For example, it is easier to visualize a laser beam as consisting of a stream of photons that has a well-defined boundary at the edge of the laser beam, just like a jet stream of water has a well-defined boundary. Thus, it is tempting to think the photons might act like the water molecules. However, in reality the edge of a laser beam is not as well defined as you might think, and, as we will show, the wave-like nature of the light cannot be ignored.

1.3 Basic properties of EM waves

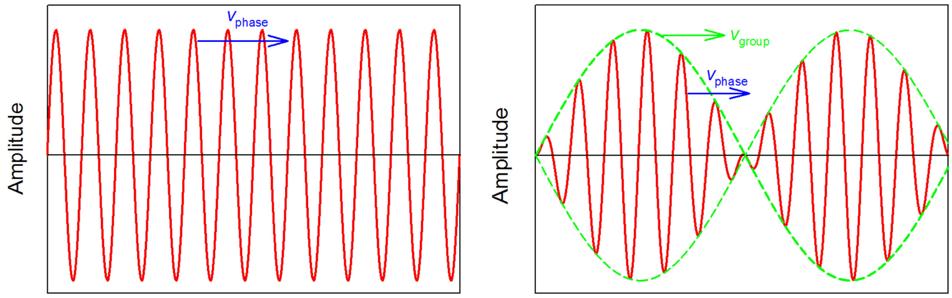
Often textbooks will assume EM waves with infinite transverse extents, i.e., infinitely wide plane waves, in order to facilitate the mathematics. Here we take a slightly different approach where we discuss the basic properties of EM waves as manifested by laser beams. The goal is to provide a more direct connection with the concepts being discussed and how they are evident in the real world.

1.3.1 Phase and group velocity

The first very basic property of laser light is the velocity of the photons. In a vacuum, the velocity is c, the speed of light. If the photons are traveling through a dielectric, then their velocity v_{photon} is less than the speed of light. This amount of reduction is defined by the index of refraction of the dielectric n, where $v_{\text{photon}} = c/n$ [2]. For example, if the laser light is traveling through quartz (fused silica) that has an index of refraction of 1.46, then the velocity of the photons is 1.46 times less than c, i.e., $v_{\text{photon}} \approx 2 \times 10^8$ m s^{-1}. This also means the EM wave velocity is less by this amount. If one looks at an EM wave propagating along the longitudinal direction z, as depicted in figure 1.3(a), the velocity of any point on the wave is called the phase velocity, v_{phase}, which is equal to the velocity of the photon, i.e., $v_{\text{phase}} = v_{\text{photon}}$. The phase velocity is often of importance for laser beams because it is the phase relationship between the waves within beams that gives rise to phenomena such as interference effects or mode-locking, which will be discussed in sections 1.3.6 and 2.5, respectively. More precisely, the phase velocity is given by [29]

$$v_{\text{phase}} = \frac{c}{n} = \frac{\lambda}{T} = \frac{\omega}{k}, \tag{1.15}$$

where λ and ω are their values within the dielectric medium, and recall that $k = 2\pi/\lambda$.



(a) (b)

Figure 1.3. (a) Pure EM wave where the wave travels at phase velocity v_{phase}. (b) Modulated EM wave where the wave still travels at phase velocity v_{phase}, but the envelope travels at group velocity v_{group}, where v_{group} depends on the dispersion characteristics of the medium the wave is traveling through and, hence, v_{group} does not necessarily equal v_{phase}.

There can be times when the envelope of the EM wave is modulated, as illustrated in figure 1.3(b). The velocity of the envelope is called the group velocity and is defined as [29]

$$v_{\text{group}} = \frac{\partial \omega}{\partial k}. \tag{1.16}$$

To better understand group velocity, how this modulation can arise, and how movement of this envelope can occur, figure 1.4(a) shows two waves (red and blue waveforms) where the red wave has a slightly higher frequency than the blue one. If you sum these waves together, the resultant wave is the modulated green one shown below the two waves. Where the two waves are approximately in phase, the green wave is nearly twice the amplitude of the two waves, and where they are approximately in anti-phase, as indicated by the vertical black line, the green wave is close to zero amplitude. Note, unlike interfering a laser beam with itself (see section 1.3.6) where it is possible to have phase points where the combined wave is exactly twice the amplitude of the original beam or exactly canceled out, because the red and blue waves have slightly different frequencies, in general there will rarely be phase points where they exactly double the amplitude or exactly cancel out.

Now suppose the phase of the blue wave shifts slightly with respect to the red wave. This is illustrated in figure 1.4(b). The black vertical line is still at its original position on the red wave and indicates the amount of shift is about a quarter of a wavelength. The sum of these two waves is again the green modulated wave and its envelope, as indicated by the black dashed lines, has moved to the right with respect to the red dashed envelope in figure 1.4(a). The red dashed envelope has been reproduced in figure 1.4(b) to help see the movement of the black dashed envelope. In figure 1.4(c), the phase of the blue wave has been shifted more so that at the black vertical line the two waves are approximately in phase, resulting in the envelope of the green wave being near its maximum. Thus, we see that a shifting of the phase between waves can cause the envelope of the combined waves to move with a

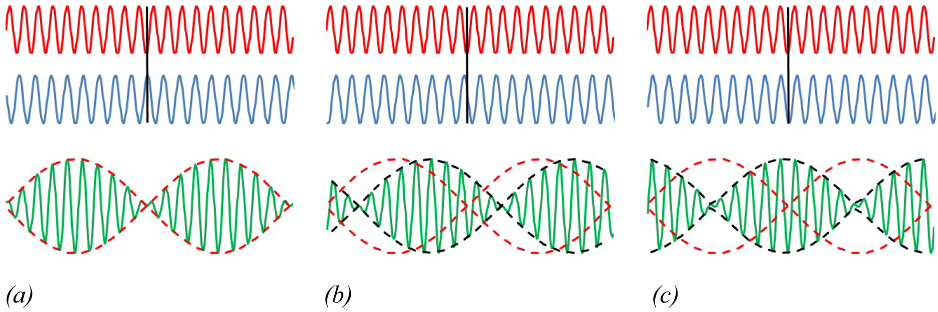


Figure 1.4. Example of group velocity. (a) Two waves of slightly different frequencies (red and blue waveforms) summed together to create a modulated wave (green waveform) whose envelope is indicated by red dashed lines. (b) Same as (a) except the blue waveform is slightly shifted forward in phase with respect to the red wave, resulting in the envelope of the green modulated wave moving to the right. (c) Further shifting in phase of the blue wave with respect to the red one.

velocity equal to the group velocity, i.e., in proportion to the change of ω with respect to k.

The example in figure 1.4 begs two questions: first, how is it possible to have two slightly different wavelengths that are combined and, second, how is it possible that their phases can slip with respect to each other? The answer to the second question is the phenomenon called dispersion where, depending on the frequency of the light, its velocity through a medium will change [2]. This is equivalent to saying that the index of refraction of the medium changes depending on the wavelength of the light. The index of refraction of materials as a function of wavelength is readily available, and generally the refractive index changes slowly versus wavelength except when the wavelength nears a resonance of the material [30, 31]. At the resonance, the material absorbs the light and the index of refraction no longer behaves in a simple manner [2].

The answer to the first question is that if the light source is not monochromatic, i.e., it is not a single wavelength, or the light source is extremely short in pulse length, then there will be a range of different wavelengths propagating though the medium. These different wavelengths or frequencies will travel at slightly different velocities through the medium depending on the dispersion characteristics of the medium, thereby resulting in a combined wave whose group velocity is given by equation (1.16).

As discussed in Chapter 2, laser light emission is essentially monochromatic even though there is a finite range of frequencies being emitted centered at the laser line, called the linewidth or bandwidth of the laser emission. The linewidth is generally very small so that the index of refraction is basically constant and there is negligible dispersion. Even if the laser emits multiple longitudinal modes (see section 2.5) that each have slightly different frequencies, the frequency range is again so small that dispersion is typically not an issue. Nonetheless, dispersion can have practical ramifications, such as the exact focal position of a lens being a function of the wavelength of the light being focused by the lens. This is why polychromatic, compound lenses are made in order to provide sharp, i.e., well-focused, images over the entire visible spectrum.

Ultrashort laser pulses are a different matter. Some lasers are capable of emitting pulses with extremely short pulse duration, e.g., femtoseconds (fs = 10^{-15} s) [32, 33] with some now reaching down to attoseconds (at = 10^{-18} s) [34]. Although the laser light may be centered at a particular wavelength, we know from Fourier analysis that a pulse can be decomposed into a spectrum of different frequencies whose extent or range of frequencies is inversely proportional to the time duration of the pulse. Hence, these ultrashort laser pulses have very large frequency ranges, and these different frequency components can have appreciably different phase velocities depending on the dispersion characteristics of the dielectric medium the laser pulse is traveling through. One must avoid sending these ultrashort pulses through any transmissive optic, such as a window or lens, because dispersion will cause the different frequency components of the pulse to exit the optic at different times, thereby increasing the time duration of the pulse.

1.3.2 EM wave propagation

The next basic property of EM waves as manifested by laser beams is how the beam propagates through space [2, 29]. Although laser beams are often thought of as being 'pencil-beams' of light, in reality the laser beam can never stay perfectly collimated as it propagates. Before we explore this issue in more detail, we must first examine the typical transverse intensity profile of a laser beam, where intensity is defined as power per unit area and has units such as W cm^{-2}. If we assume the laser beam is cylindrically symmetric in its intensity profile, then, as illustrated in figure 1.5(a), the cross-section of a typical laser beam has a Gaussian intensity profile $I(r)$ [27, 29], in which

$$I(r) = I_0 e^{-2r^2/w^2}, \tag{1.17}$$

where w is the radius of the Gaussian profile corresponding to the $1/e$ point, also referred to as the spot size of the Gaussian beam [29]. The factor of 2 in the exponent of equation (1.17) arises because it was assumed that the amplitude of the electric field of the Gaussian beam is given by $E(r) = E_0 \exp(-r^2/w^2)$. The intensity, which is equivalent to power per unit area, is found by squaring the amplitude (see equation (1.13)). This also implies that w represents the radius of the electric field and not the radius of the laser beam intensity.

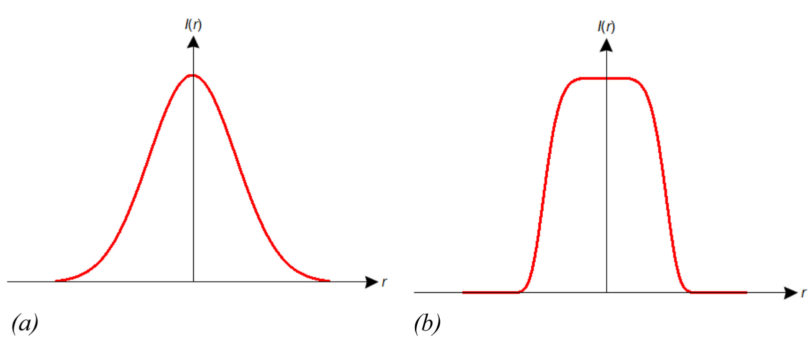


Figure 1.5. (a) Generic Gaussian beam profile. (b) Example of super-Gaussian beam profile.

Some lasers, such as HeNe lasers emitting red light at 632.8 nm, generate laser beams that fit a Gaussian profile very well. However, depending on other factors, such as how the laser medium is excited, the beam output may be an elliptically shaped Gaussian profile where w in equation (1.17) may have very different values in the x and y directions. Or, the output may have an almost flat-top profile, as depicted in figure 1.5(b), and is referred to as a super-Gaussian beam [29]. Thus, in the equations that follow, where we characterize the various properties of the laser beams, it should be kept in mind that these only precisely apply to pure Gaussian beams. And, since pure Gaussian beams are actually uncommon, the equations are an approximation of what you will encounter in real life.

This brings us to answering the question of how do laser beams propagate through space. Figure 1.6 shows schematically a Gaussian laser beam traveling in the z-direction with a transverse intensity profile given by equation (1.17). At $z = 0$, the beam has its minimum radius, which is called the waist of the beam, with a spot size value of $w = w_0$. This waist might be located at the output of the laser, i.e., immediately on the output side of its output resonator mirror (see figure 2.5), or it might be at the focal point of a lens that is focusing the laser beam. This is also the position along the length of the laser beam where it is a uniform plane wave; in fact, it is the only point where it is a plane wave. At any other position along z in both the positive and negative directions, the EM wave begins to acquire a curvature and deviates from being a perfect plane wave. Fortunately, this can be a gradual process where the laser beam can be considered essentially a plane wave within a distance called the Rayleigh range [29], as indicated in figure 1.6.

By definition, the Rayleigh range is

$$z_\mathrm{R} = \frac{\pi w_0^2}{\lambda}. \tag{1.18}$$

It specifically represents the distance where the area of the laser beam doubles. The Rayleigh range also marks approximately the division point between the near-field or Fresnel region (distances less than z_R) and the far-field or Fraunhofer region (distances larger than z_R) [35]. Within the near-field, the laser beam behaves much like a collimated beam consisting of wave fronts that can be approximated as plane waves; whereas, in the far-field, the beam consists of spherical waves that diverge with distance and grow larger in radius. This phenomenon is called self-diffraction [29], and the divergence angle of the far-field beam (see figure 1.6) is given by

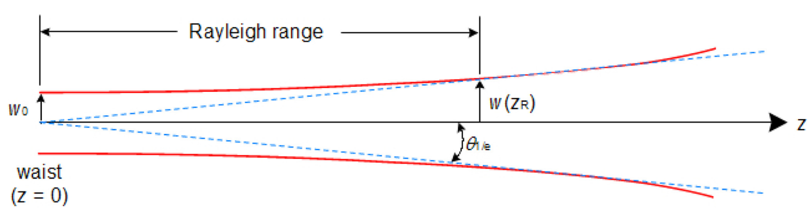


Figure 1.6. Defining propagation parameters for a Gaussian laser beam.

$$\theta_{1/e} = \frac{\lambda}{\pi w_0} = \frac{w_0}{z_R}, \tag{1.19}$$

where the angle $\theta_{1/e}$ is defined at the $1/e$ point of the Gaussian beam (see equation (1.17)).

Equation (1.18) implies a very important property of laser beams. If you want a laser beam to stay well collimated over a long distance (i.e., large value for z_R), the waist should be as large as possible and/or the shortest possible laser wavelength should be used. Since there is often no choice on the laser wavelength, this only gives the waist size to control. But, the diameter of the laser beam being emitted by the laser is also typically a fixed parameter. This is why expansion telescopes are utilized at the output of the laser to expand the diameter of the laser beam in order to achieve a longer Rayleigh range. For example, if $\lambda = 632.8$ nm and $w_0 = 2$ mm, then $z_R = 20$ m; but if $w_0 = 20$ cm, then $z_R = 200$ km!

Knowing the waist size of the laser beam, it is possible to calculate the radius of the Gaussian beam as a function of z in the far-field through the approximate expression [29]

$$w(z) \approx \frac{w_0 z}{z_r} = \frac{\lambda z}{\pi w_0}. \tag{1.20}$$

1.3.3 Focusing of Gaussian beams

We mentioned in conjunction with figure 1.6 that the waist can also be at the focal plane of a lens. In fact, a positive-focal-length lens could be positioned on the right side of figure 1.6 with a beam of light coming towards the lens from the far right. The beam outline shown in red in figure 1.6 would represent the focusing of this beam of light to a focal radius equal to w_0. This is more clearly illustrated in figure 1.7 where a laser beam with incoming radius w_i is being focused by a positive (plano-convex) lens with focal length f. For the sake of this discussion, let us assume that this laser beam consists of plane waves entering the lens. The lens causes these waves to exit

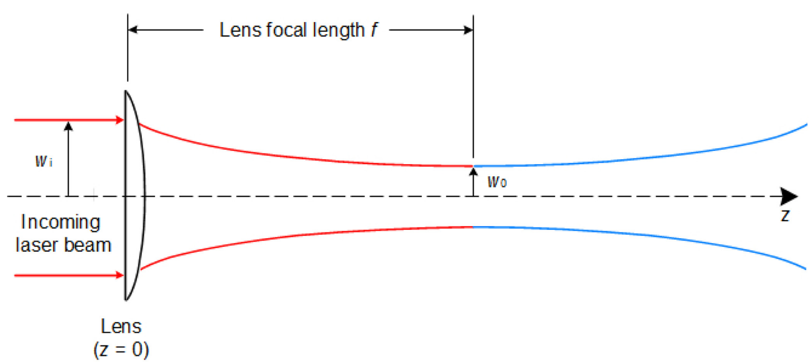


Figure 1.7. Focusing of laser beam by lens.

with the curvature of the exit surface of the lens. It turns out this curvature is the same as the curvature in the far-field of the beam shown in figure 1.6. Hence, a lens can be viewed as a means for transforming an EM wave from the far-field to the near-field, i.e., focusing the beam down to a small spot at its focal plane. As the beam travels past the focal plane it expands and returns to the far-field (shown in blue in figure 1.7). Note, at the focal plane of the lens, the focused light has not only its smallest radius w_0, but at this point the wave is once again a plane wave.

With this realization, equation (1.20) can also represent approximately the focusing properties of the lens if we replace z with f, the focal length of the lens. Thus, the waist spot size of a laser beam focused by a lens with focal length f is given by [29]

$$w_0 \approx \frac{\lambda f}{\pi w_\mathrm{i}}, \tag{1.21}$$

where w_i is the radius of the incoming laser beam at the lens.

In section 1.2.3 with regard to the Poynting vector, we mentioned how equation (1.13) can be used to estimate the conditions needed to break down air by focusing a laser beam. Suppose you have a Nd:YAG laser with an output pulse energy of 1 J, pulse length of 6 ns, and beam radius of w_i = 5 mm. Its peak power is then 1 J/6 ns = 1.7×10^8 W = 170 MW. To cause laser-induced breakdown of air at this wavelength and pulse duration requires roughly 3×10^{10} V m^{-1} of electric field. From equation (1.13), the power density needed to achieve this field is 1.2×10^{18} W m^{-2}. This implies that our laser beam needs to be focused down to an area no greater than (170 MW)/(1.2×10^{18} W m^{-2}) = 1.4×10^{-10} m^2. Hence, the waist spot size needs to be equal to or smaller than $w_0 = [(1.4 \times 10^{-10}\ \mathrm{m}^2)/\pi]^{1/2} = 6.7 \times 10^{-6}$ m = 6.7 μm. To achieve this spot size, equation (1.21) indicates for w_i = 5 mm and λ = 1.06 μm that the focal length of the lens should be equal to or shorter than 10 cm. If you sent this laser beam through this lens, then at 10 cm from the lens you will see bright flashes of light suspended in mid-air and hear loud snapping sounds every time the laser fires and breaks down the air!

Returning again to equation (1.21), as mentioned, laser beams are often not precisely Gaussian so that expressions such as equation (1.21) become less accurate in their approximation as the beam profile deviates more from being purely Gaussian. Moreover, the variable w in these equations are the $1/e$ values for the Gaussian E-field; therefore, it is the $1/e^2$ value of the intensity profile that needs to be measured, i.e., the radius where the intensity is 86% down from the peak. The $1/e^2$ intensity level for a Gaussian beam is shown in figure 1.8. It can be inconvenient or awkward to measure the 86% intensity point. Instead the width of the Gaussian profile is often measured at the point where the intensity is one-half its maximum value. As illustrated in figure 1.8, this is called the full-width-at-half-maximum (FWHM) and is equal to $2r_\mathrm{FWHM}$, where r_FWHM is the radius at FWHM. An alternative width that is also straightforward to measure is at the 99% points (see figure 1.8), which is essentially the same as measuring the total beam diameter. This is equal to $2r_{99\%}$, where $r_{99\%}$ is the radius at the 99% point. It is then possible to

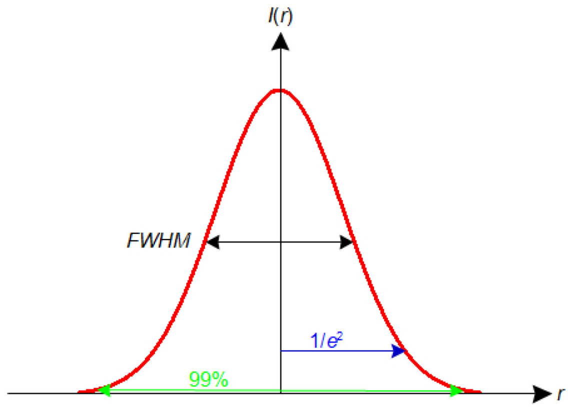


Figure 1.8. Different conventions for measuring width of a Gaussian profile. FWHM stands for full-width-at-half-maximum. Sometimes you will see HWHM, which is half-width-at-half-maximum, and is the radius of the Gaussian beam at the half-maximum point.

use these more easily measured quantities to calculate the w radius for the Gaussian profile. Specifically, $w = 1.70r_{\mathrm{FWHM}} = 0.659r_{99\%}$ [36].

As another way of calculating the waist spot size of a lens, we note the f-number or speed of a lens is defined as [37]

$$f\# = \frac{f}{D} = \frac{1}{2NA}, \tag{1.22}$$

where D is the diameter of the lens and $NA = D/2f$ is called the numerical aperture of the lens [37]. It is implied that the lens is gathering all the light impinging on it and focusing it to a spot. For the purposes of calculating the spot size of a laser beam focused by the lens, let us assume the entire laser beam fills up the diameter of this lens. (In reality, as explained in section 1.3.7, we would not want to entirely fill up the lens aperture with the laser beam because of diffraction effects caused by the edge of the aperture.) In that regards, D represents the diameter that contains, say, 99% of the laser beam power. This is a convenient parameter to work with because generally one is interested in the beam diameter that contains essentially all the beam power and, as mentioned, this 99% diameter is also easier to measure in the laboratory.

Notice that the numerical aperture of the lens is inversely related to the $f\#$. Thus, a large $f\#$ where the focal length of the lens is long compared to the lens diameter corresponds to a small NA. This is generally the regime one would like to operate in because the approximations for the Gaussian beam optics become progressively less accurate as the $f\#$ becomes smaller and NA becomes larger, i.e., short focal length lenses.

The diameter of the focused spot containing 99% of the beam power is then [36]

$$2r_{99\%} \approx \frac{\lambda}{NA} = 2f\#\lambda = \frac{2f\lambda}{D}. \tag{1.23}$$

We should emphasize that equation (1.23) is an approximation that becomes less accurate for small *f*#. This means a *f*# of order 1 does not mean you can achieve a waist spot size of order λ. In general, the spot size will be much larger than λ because typical laser beams are not pure Gaussian ones. This brings us to another characteristic of laser beams.

1.3.4 Non-ideal Gaussian beams

As mentioned, when a laser beam deviates from being a pure Gaussian, e.g., the super-Gaussian in figure 1.5(b), the approximations in the preceding equations not only become less accurate, but the actual results are also not as good as the ideal results. For example, besides a non-Gaussian beam being unable to focus to as small a spot size as a pure Gaussian beam, a non-Gaussian beam will also tend to spread out faster than a pure Gaussian beam, i.e., its Rayleigh range will be shorter than predicted by equation (1.18). This is why pure Gaussian beams are also called diffraction-limited beams because its amount of spreading as a function of propagation distance is only limited by equation (1.18).

To help quantify the degree to which a non-Gaussian beam deviates from a pure Gaussian one, equations (1.18), (1.19), and (1.20) can be modified as follows [38]:

$$z_{\mathrm{R}} = \frac{\pi w_0^2}{M^2\lambda}; \quad \theta_{1/e} = \frac{M^2\lambda}{\pi w_0}; \quad w(z) \approx \frac{M^2\lambda z}{\pi w_0}. \tag{1.24}$$

where M^2 is called the M^2 factor, or beam quality factor, and is always greater than or equal to unity. When $M^2 = 1$, the beam is a diffraction-limited, pure Gaussian beam. Real laser beams tend to have $M^2 > 1$ and this is often empirically determined, for example by measuring the achieved waist spot size when the beam is focused by a lens and comparing this with the spot size predicted by equation (1.20). The ratio of the two spot sizes yields the value for M^2. Good quality laser beams will have M^2 values close to 1; poor quality beams may be many times diffraction-limited in their characteristics.

1.3.5 Beam coherence

The preceding characteristics related to Gaussian beams, e.g., the ability to focus to very small spot sizes, are related to another important property of laser beams that makes them unique in the family of EM waves. This is the coherence of the laser beam [27, 29]. As explained later, one special attribute of lasers is their ability to generate very pure colors of light, i.e., emit radiation at a single wavelength (monochromatic), unlike, say, an incandescent light bulb that emits light over a broad wavelength range. However, producing monochromatic light by itself does not necessarily yield diffraction-limited Gaussian beams. For example, a light emitting diode (LED), which generates light at one wavelength, emits incoherent radiation. Coherence is needed wherein all the EM waves within the laser beam are synchronized in phase with each other. This coherence is illustrated in figure 1.9(a) where the EM waves are all propagating perfectly in phase, that is their peaks and

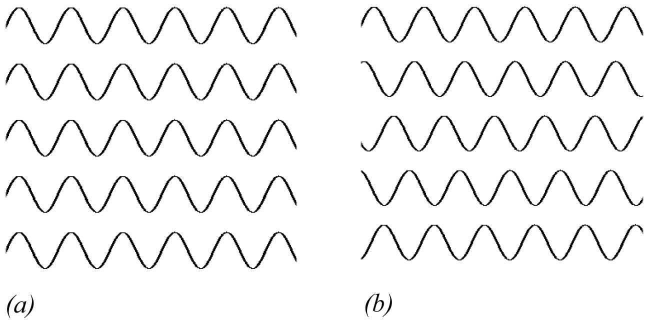


Figure 1.9. (a) EM waves coherently in phase. (b) EM waves incoherently out of phase.

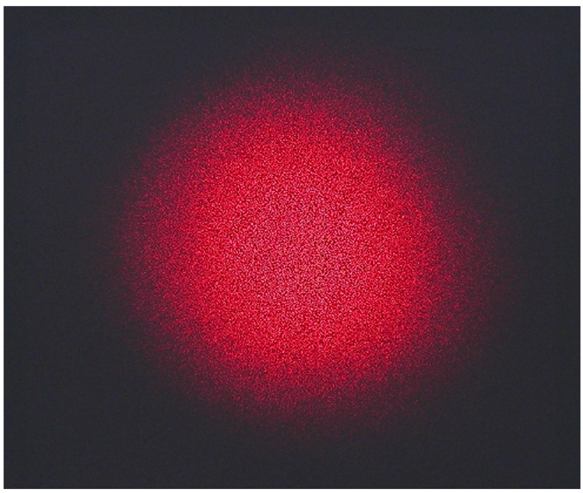


Figure 1.10. Photograph of laser spot from HeNe laser ($\lambda = 632.8$ nm) shining on a wall. The mottled pattern is called speckle and is due to the coherence of the beam, which causes the laser light scattering off the wall to interfere with itself.

valleys are all aligned with each other. Incoherence occurs when the phase relationship between the EM waves is completely random, as depicted in figure 1.9(b). It is also possible to have partial coherence, where the waves are somewhat in phase [28].

Thus, a diffraction-limited laser beam has a high degree of coherence. A poor quality laser beam with a large M^2 value may have partial coherence. Indeed, until the invention of lasers, it was difficult to create truly coherent light sources. Ironically, once a coherent beam has been generated, it is actually difficult to destroy the coherence. An example of this can be seen whenever one shines, say, the red beam ($\lambda = 632.8$ nm) from a HeNe laser onto a wall (see figure 1.10). If you examine the laser spot carefully, you will see a speckle pattern consisting of bright and dark spots within the laser beam image on the wall [39]. These spots will seem to move around, especially if you look at the laser spot from different angles. What is

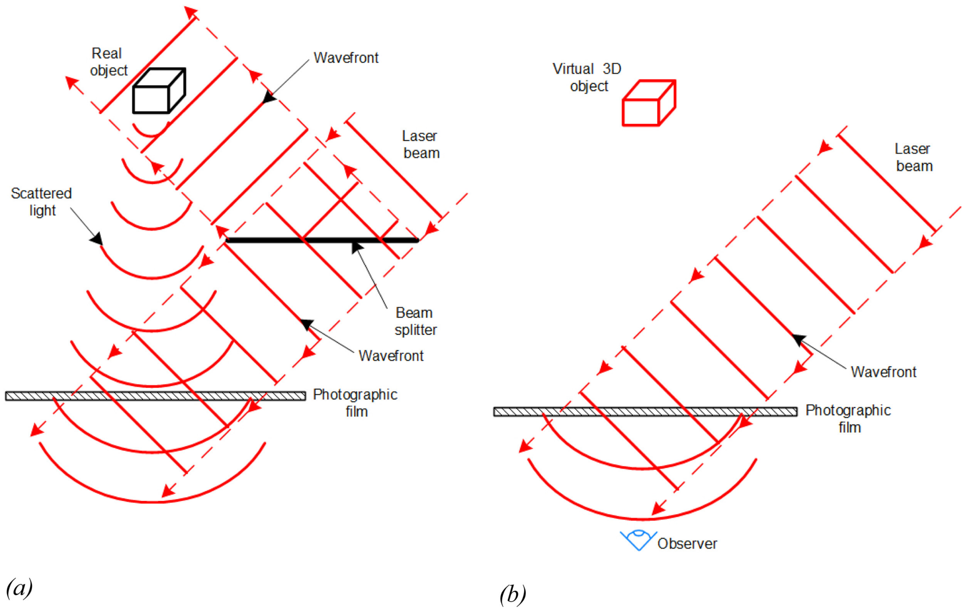


Figure 1.11. Basic scheme for creating and viewing a hologram. (a) Light from a laser beam scatters off the object and at the same time illuminates the photographic film, which creates an interference pattern on the film, thereby creating the hologram. (b) The photographic film (hologram) is illuminated by the laser beam and, due to the interference pattern on the hologram, the observer sees the same scattered light pattern that was created by the object, thereby creating a virtual 3D image of the object.

not obvious is that anyone else looking at the same laser spot will see a different speckle pattern than the one you are observing. This is because the speckle pattern is caused by the EM waves from the laser beam scattering off the atoms on the walls at different angles. Since the waves are coherent before they strike the wall, when they reflect off the wall their phases with respect to each other shift because they are scattering at different angles. This means the waves create an interference pattern, where the peaks of one wave might line up with the valley of another wave, thereby canceling each other out and creating a 'dark' spot. Depending on where you are standing, the angles of the scattered light are different. This is why each observer sees a different speckle pattern.

Coherence not only gives rise to the basic properties of Gaussian beams, it is also important for many applications. An example is holography [40], whose basic concept is illustrated in figure 1.11. The first step is to imprint a hologram, which will be the piece of photograph film depicted in figure 1.11(a). The output from a laser, say a HeNe laser, is sent through a beam-splitter so that a portion of the beam illuminates the object (probe beam) and the other portion is directed at the photographic film (reference beam). Also shown in figure 1.11(a) are the wavefronts of the beams that are oriented perpendicular to the direction of the beam propagation. The wavefronts of the probe beam scatter off the object like water waves reflecting off a pier. When these scattered wavefronts reach the position of the photographic film, they coherently interfere with the wavefronts of the reference.

Due to the coherence between the two sets of wavefronts, there will be regions of high light intensity where the wavefronts add together and regions of low light intensity where the wavefronts cancel each other out. This produces an interference pattern that is recorded by the photographic film and will look like random regions of dark and light spots on the film. In fact, it is not necessary to use color film, black and white film is fine.

The procedure for viewing the hologram is given in figure 1.11(b). A HeNe laser beam is directed at the film. When the laser light passes through the film, the film scatters the light to create the same scattered light distribution pattern that was produced when the real object was scattering light towards the photographic film. To an observer viewing this scattered light pattern, it looks exactly like what they would see had they been present when the film was being exposed. Thus, the observer will see a virtual image of the object. The 3D aspect of the hologram occurs because when the film was exposed in figure 1.11(a), the scattered light from the object came from all directions and angles with respect to viewing the object. When the laser light scatters off the film in figure 1.11(b), it reproduces these same directions and angles. Therefore, if the observer looks at the hologram from a different direction, they will see the light that the object would have scattered in that direction. To the observer it will appear they are looking at the object from a different angle, thereby making the object appear 3D.

1.3.6 Interferometry using laser beams

The concept of interference was introduced when explaining how a hologram works and the importance of coherence. This concept will now be explored further by discussing interferometry [29]. Although there are many different types of laser interferometers, the basic concept is to split the beam into two beams and then recombine the beams at some point in space so that they create an interference pattern consisting of bright and dark lines, called fringes. Figure 1.12 shows the basic layout for a Michelson interferometer. A laser beam (colored in blue) enters from the left. To make a useful interferometric image, you generally need a laser beam with a finite width, as depicted in figure 1.12. Achieving the desired width can be accomplished by using an expansion telescope to enlarge the laser beam emitted by the laser.

The laser beam is directed towards a 50/50 beam splitter. This is generally an optic whose substrate material transmits the laser beam with low absorption, for example fused silica in the case of visible and NIR light. It also has highly polished surfaces to minimize scattering of the laser light from the surfaces. One surface of the beam splitter has a multilayer dielectric coating [41, 42] deposited on it that has been designed to reflect 50% of the laser light at 45° to the surface of the beam splitter and transmit the remainder of the light through the optic. The other side of the beam splitter also has a multilayer dielectric coating, except this coating has been designed not to reflect any of the laser light. It is referred to as an anti-reflection coating (AR-coating). The AR-coating prevents the back surface of the beam splitter from reflecting the laser beam, which would cause multiple images at the output of the

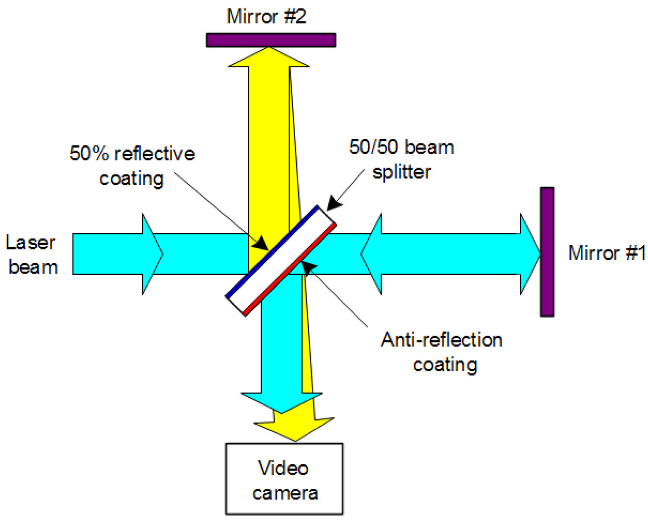


Figure 1.12. Basic layout for Michelson interferometer.

interferometer. This is important in order to achieve clear interference fringes from the interferometer. (Section 2.8.1 explains more about optical coatings.)

In figure 1.12, the blue laser beam that passes through the beam splitter is reflected by Mirror #1 directly upon itself so that when the reflected beam reaches the beam splitter, it is reflected downward off the 50% reflective coating towards an image detector, in this case a video camera. We shall refer to this blue beam as the reference beam. Note, both Mirrors 1 and 2 have their reflective coating on the front surface of the mirror so that the laser beam does not pass through the mirror substrate.

The 50% of the laser beam that is reflected by the beam splitter is directed towards Mirror #2. This beam has been colored yellow to differentiate it from the beam that passes through the beam splitter. Mirror #2 reflects the yellow beam back towards the beam splitter, but the mirror has been intentionally adjusted so that the yellow beam is slightly skewed relative to the blue beam as the two beams overlap each other at the video camera. The reason this is important is illustrated in figure 1.13, which shows schematically the EM waves of the blue and yellow laser beams in figure 1.12. For the purposes of this discussion we can assume that the colored bars represent the positive peaks of the EM waves. If the two laser beams are exactly parallel to each other, then there is no variation in the overlap of the EM waves across the widths of the beams, as shown in figure 1.13(a). If the peaks of the yellow beam happen to overlap in the negative valleys of the blue beam, it will cancel out the electric field of the blue beam (so-called destructive interference). The video camera would then detect no light. Conversely, if the peaks of the two beams overlap each other (so-called constructive interference), then the camera will only see a uniform bright image. This is referred to as operating on the zero order of the interferometer.

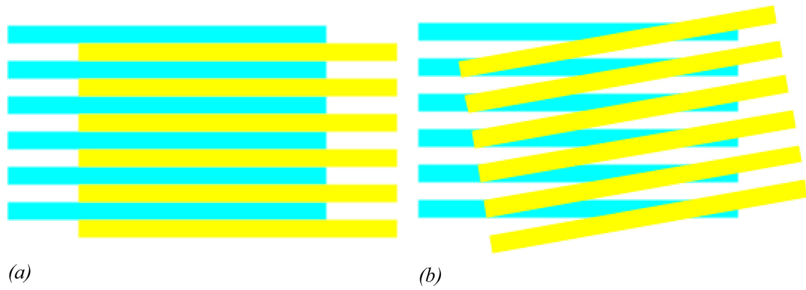
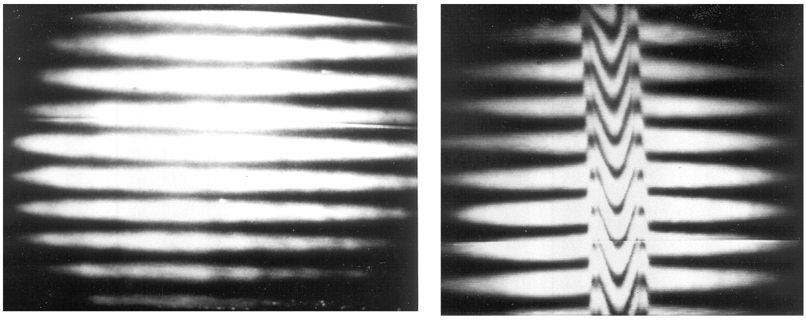


Figure 1.13. Cartoon of interfering waves where the colored bars represent the peak of the wave. (a) If the two laser beams are exactly parallel to each other, there are no fringes. (b) If the two laser beams are slightly skewed, there are fringes.

On the other hand, if the beams are slightly skewed, then as depicted in figure 1.13(b), the EM waves cross each other, creating regions in space where the peaks sometimes overlap each other or sometimes overlap the valleys. The camera will thus see a regular linear pattern of bright or dark lines. These are the fringes seen by the camera. Notice in figure 1.13(b) that the horizontal distance over which, say, the peaks overlap each other depends on the angle of the yellow beam relative to the blue one. As the angle becomes larger, the horizontal distance gets smaller, which means the fringes become narrower in width. Thus, skewing the beams permits operating at higher orders on the interferometer where the higher the order, the narrower the fringes.

Figure 1.13(b) also implies that to achieve the sharpest contrast in the fringe pattern requires utilizing a monochromatic light source consisting of perfect plane waves. Lasers are the closest to being such a source, but they are not perfect. As mentioned, there is generally a finite bandwidth of wavelengths over which the laser emits light, so it is only partially monochromatic. Furthermore, even a collimated laser beam still has some wave front curvature, i.e., the fronts are not perfect plane waves. In addition, if the optics in the interferometer (i.e., mirrors and beam splitter) have any slight surface curvature, then this will further reduce the contrast. This is why when specifying the optical quality of these optics, you generally want surface flatnesses significantly less than the laser wavelength (e.g., surface flatness deviation $<\lambda/10$).

Figure 1.14(a) is a photograph of the fringe pattern produced by a laser interferometer used to probe an electrical arc discharge [43]. In this case, the tilt of the mirrors of the interferometer was adjusted to produce a horizontal fringe pattern. The fringes appear wider in the middle of the photograph due to the laser intensity being brighter in the middle since the laser beam had roughly a super-Gaussian profile. The arc is located in one branch of the interferometer. When the arc is present (see figure 1.14(b)), it ionizes the gas within the arc and creates a hot plasma. This plasma has a different index of refraction compared to the surrounding gas. Therefore, laser light passing through the arc experiences a phase shift relative to the surrounding gas and results in the fringe pattern shifting within the region where the arc is present.



(a) (b)

Figure 1.14. Photographs showing examples of interference fringe pattern produced in a laser interferometer used to probe an electrical arc discharge. (a) Fringe pattern with no arc present. (b) Fringe pattern with arc oriented vertically in photograph.

These photographs illustrate how a laser interferometer is able to provide 2D and even 3D information. For example, in figure 1.14(b) we see the fringe pattern within the arc tends to shift up at the edges of the arc and shift down in the middle. This behavior indicates that the plasma distribution within the arc is somewhat like a hollow tube with higher plasma density at its walls. If you assume the arc is circularly symmetric, which is a reasonable assumption, then it is possible to transform the 2D information from this fringe pattern into a 3D plasma distribution using a process called Abel transformation [44, 45].

Operating an interferometer on zero order with no fringe pattern can be useful when using the interferometer as a means to precisely measure changes in distance. For example, suppose Mirror #2 is stationary but Mirror #1 is allowed to move, e.g., it might be sitting on a metal bar that contracts or expands in its length depending on how much you heat the bar. This will cause Mirror #1 to move left or right in figure 1.12 by a very small amount, e.g., microns. This movement can be detected by setting the interferometer to zero-order and using a photodetector, such as a photodiode, to sense when the light intensity increases and decreases. If you are using a HeNe laser (λ = 0.6328 μm) as the light source for the interferometer, then this implies you will be able to detect movements as small as 0.6328 μm/2 = 0.3164 μm. The factor of two arises because when the light reflects off Mirror #1, it travels twice the distance that Mirror #1 has moved.

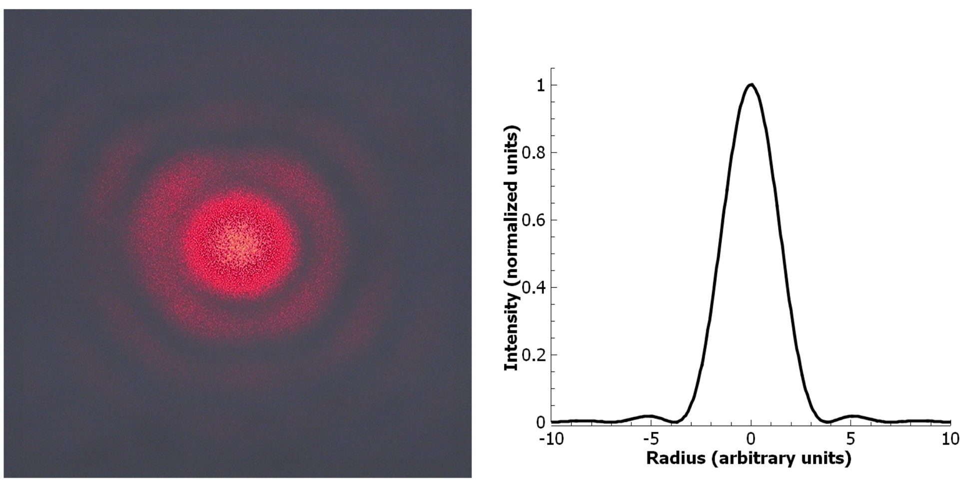
The AR-coating on the beam splitter in figure 1.12 provides an opportunity to discuss another aspect of EM waves. This coating is created by depositing thin multiple layers of two different dielectric materials on the beam splitter surface that have different index of refraction values at the laser wavelength. Laser light reflects off these layers; however, because of their different index of refraction, the phase velocity of the wave within each material is different. It is possible by controlling the thickness of the layers to have the reflected light from one dielectric layer be shifted 180° out of phase (π-phase shift) relative to light reflected by the other dielectric layer (see section 2.8.1). Then, in a similar fashion as shown in figure 1.13(a), the peaks of one reflected wave overlap the valleys of the other reflected wave resulting in zero light intensity. Thus, there is no reflected light.

In reality, because perfect laser beams and perfect AR-coatings do not exist, there is some reflected light, but it is generally greatly reduced. By judicious design of the multilayer AR-coatings, it is now possible to make broadband coatings that work well over a wide wavelength range. This has been applied, for example, on AR-coated sunglasses, where the coating needs to work over the visible sunlight spectrum.

1.3.7 Diffraction

Thus, having coherence permits many useful applications. However, it can also be a bane. Diffraction occurs when the EM wave interferes with itself [28]. Classic examples are the single and double slit experiments, where a symmetric diffraction pattern is created when the laser beam passes through the slits. Another example is knife-edge diffraction, where the laser beam scrapes the sharp edge and creates an asymmetric diffraction pattern. In each case it is the wave nature of the EM wave that best explains the creation of the diffraction pattern, i.e., the edges are a source of scattered EM waves that constructively or destructively interfere with the non-scattered EM wave.

Figure 1.15(a) shows an example of diffraction that occurs when a HeNe laser beam is sent through a pin-hole whose diameter is much smaller than the laser beam diameter. A series of concentric rings is formed, called an Airy pattern [46]. The light that forms these rings is being scattered at an angle from the edges of the pin-hole and, hence, the light that forms these rings emanates as a cone of radiation from the



(a) *(b)*

Figure 1.15. (a) Photograph showing diffraction of HeNe laser light when the laser beam passes through a hole whose diameter is much smaller than the laser beam diameter. (b) Cross-section of Airy pattern intensity profile, similar to the cross-section profile of the beam in (a).

pin-hole. The intensity profile of an Airy pattern is depicted in figure 1.15(b) and would be similar to what would be measured along any diameter of the image shown in figure 1.15(a). The rings are much fainter than the main central lobe.

Now suppose instead of a pin-hole we pass the HeNe laser beam through an aperture (iris) whose diameter is only slightly smaller than the laser beam. What we observe downstream of the iris is shown in figure 1.16. Notice that, unlike the fairly uniform laser spot image shown in figure 1.10, we see a series of bright and dark rings inside the laser spot. These are called Fresnel rings [29] and they are due to diffraction caused by the laser light scraping the inside edge of the iris. What is surprising is that these Fresnel rings can still occur even when the iris diameter is considerably bigger than the apparent diameter of the laser beam. I say 'apparent' because the laser beam is a Gaussian one and, technically speaking, the tails of a Gaussian extend to infinity. Although the diameter of the laser beam certainly does not extend to infinity, it can still have sufficient intensity near the edge of the beam, which cannot be seen, to cause diffraction effects beyond $r_{99\%}$. As a practical matter, to completely avoid creating Fresnel rings, the iris radius should be larger than about $3r_{99\%}$.

Fresnel rings can be a nuisance when they arise during the construction of, say, an interferometer because the rings will superimpose themselves on the fringe pattern. There can be, however, an even more sinister problem. Under the right conditions, through constructive interference, the Fresnel rings can concentrate themselves at the very center of the laser beam, thereby creating an intense spot of laser power. This spot is called the Spot of Arago [29] and an example is shown in figure 1.17.

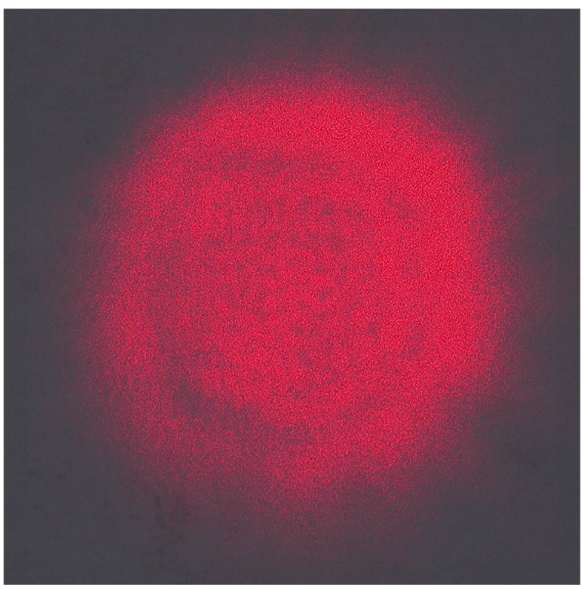


Figure 1.16. Photograph showing diffraction of HeNe laser beam when the beam passes through an iris whose diameter is slightly larger than the laser beam.

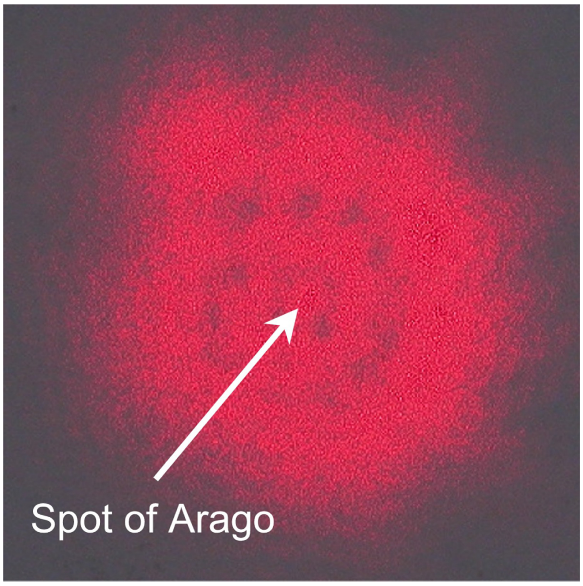


Figure 1.17. Photograph of HeNe laser beam at a position downstream of an iris showing a Spot of Arago that has been created in the center.

Although not very bright in this example, the laser intensity in this spot can be high enough to damage an optic that happens to be located where this spot has formed.

Diffraction either by the edges of, say, an aperture or self-diffraction of a Gaussian beam ultimately affects the ability to distinguish features associated with the EM wave, for example in discerning the profiles of two laser beams lying close to each other. This is related to the resolution of the optical system and it can be very important in many applications. As a simple example, the angular resolution θ of the Airy pattern shown figure 1.15 is given by [46]

$$\theta = 1.22\frac{\lambda}{D}, \tag{1.25}$$

where θ is in radians, λ is the wavelength of the light, and D the diameter of the aperture or lens with λ and D measured in the same units, e.g., meters. This implies for a fixed value of D that finer angular resolution occurs at shorter wavelengths. Hence, if one wants to resolve tiny features using an optical probe, it is better to use as short a wavelength as possible. In an analogous manner, equation 1.21 shows that the waist spot size of a focused Gaussian beam is directly proportional to the wavelength of the light. Hence, for a fixed focal length and beam input size, a smaller focal spot is possible by using a shorter wavelength. This is why a Blu-Ray disc is able to contain much more digital data than a regular DVD because it utilizes a laser diode emitting light at 405 nm rather than the laser diodes in DVD players that emit light at 650 nm. This permits writing and resolving smaller pits in the disc that represent the digital information.

1.3.8 Light scattering

The laser beam images in figures 1.10 and 1.15–1.17 also demonstrate the simple phenomenon of light scattering, in this case off a wall to enable obtaining a photograph using a camera. Here it can be more convenient to refer to the photons scattering off the molecules in the wall rather than the EM wave. Usually the photons are elastically scattered, i.e., the energy of the photon does not change and, thus, by equation (1.14), their frequency does not change. This is called Rayleigh scattering [2].

Figure 1.18(a) illustrates what is happening during Rayleigh scattering from the viewpoint of the energy of the photon and the molecule interacting with the photon. A photon with a frequency ν_0, corresponding to an energy $h\nu_0$ (see equation (1.14)), impinges on the molecule, which is at its lowest energy state E_0. The photon causes the molecule to be excited momentarily into a higher energy state corresponding to the energy of the photon. This higher state is called a virtual state because it is generally not associated with one of the allowed energy states of the molecule. What usually happens is that the molecule relaxes back down to its original lowest energy state and reemits the photon with the same energy as the incoming photon, i.e., $h\nu_0$; therefore, the reemitted light is at the same wavelength as the incoming light. However, this remission is generally in a random new direction relative to the incoming photon's direction, which results in a scattered light distribution.

Sometimes, though, the molecule relaxes down to an energy level that is higher than the ground state, i.e., level E_1 shown in figure 1.18(b). Now the reemitted photon has a lower energy than the incoming photon, i.e.,. $h\nu_1 < h\nu_0$, which implies the reemitted light is at a lower frequency corresponding to a longer wavelength. As depicted in figure 1.18(c), a complementary process is also possible where the molecule happens to be in the excited state E_1, the incoming photon excites the molecule to a virtual level, and the molecule relaxes down to the ground state E_0. In this case, the reemitted photon has a higher energy than the incoming one, i.e., $h\nu_1 > h\nu_0$, which means the reemitted light has a shorter wavelength than the incoming one. These are examples of inelastic scattering because the energy of the reemitted photon is different than the incoming one.

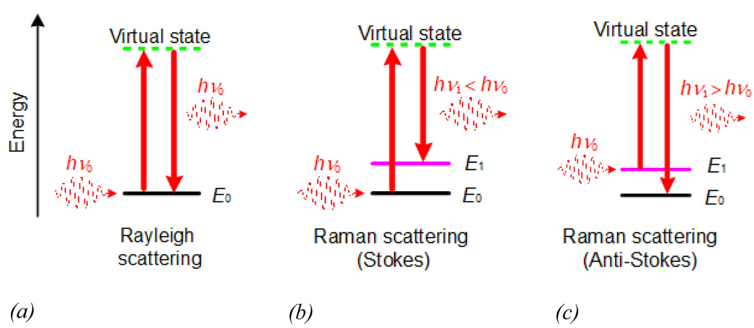


Figure 1.18. Rayleigh scattering versus Raman scattering. (a) Rayleigh scattering. (b) Raman scattering with Stokes emission. (c) Raman scattering with anti-Stokes emission.

The two processes shown in figures 1.18(b) and 1.18(c) are referred to as Raman scattering [17], and when the reemitted light is at a longer wavelength, i.e., figure 1.18(b), it is called Stokes scattering and when it is at a shorter wavelength, i.e., figure 1.18(c), it is called anti-Stokes scattering. Stokes scattering tends to occur more often than anti-Stokes scattering because most molecules are typically in the ground state and only a few in excited states. Because the probability of having Stokes or anti-Stokes scattering occurring is very small, it was not until the invention of lasers, with their ability to emit copious amounts of photons at one wavelength and maintain a relatively small beam diameter over long distances (i.e., long Rayleigh range), did it become possible to generate appreciable Stokes or anti-Stokes radiation. And, even then, special Raman cells are generally used, which are typically long tubes (e.g., ~2 m) with windows at their ends in which the laser beam enters on one end and the Stokes or anti-Stokes radiation exit the opposite end. The tubes are filled with a gas such as hydrogen that serves as the Raman medium. The gas is at very high pressures (e.g., 10 atm) in order to maximize the number of molecules that can interact with the laser photons.

The concept of photons interacting with the energy levels of a medium is one of the key aspects of how lasers work. This is a convenient segue into the next chapter, which will discuss lasers in more detail.

Chapter 1 Summary

1) The propagation of EM waves can be mathematically characterized by using the Helmholtz wave equation, as derived from Maxwell's equations.
2) The time-average power density flow in an EM wave is given by the Poynting vector and is equal to P_{average} (W m^{-2}) = (1/2) $E_{\max}{}^2$ (V m^{-1})/377 Ω.
3) Photons can be conceptualized as compact packets of waves with wavelength λ and frequency ν, and an energy equal to $h\nu$.
4) The phase velocity of an EM wave is $v_{\text{phase}} = c/n = \omega/k$, where c is the speed of light, n is the index of refraction, ω is the radian frequency of the light, and $k = 2\pi/\lambda$. The phase velocity corresponds to how fast the wave is moving in any medium.
5) The group velocity of EM waves is $v_{\text{group}} = \partial\omega/\partial k$ and corresponds to how fast the envelope moves, consisting of combined waves with different frequencies.
6) Laser beams have typically Gaussian profiles with an E-field $1/e$-radius equal to a spot size w; although, some lasers emit beams with super-Gaussian profiles.
7) The Rayleigh range is $z_R = \pi w_0{}^2/\lambda$, where w_0 is the minimum waist size of the beam. This describes how far a laser beam can stay approximately collimated. Beyond the Rayleigh range, the beam begins to spread out with an angle $\theta_{1/e} = \lambda/\pi w_0$.
8) A lens can focus a laser beam down to a focus waist spot size of $w_0 = \lambda f/\pi w_i$, where f is the focal length of the lens and $2w_i$ is the diameter of the laser beam entering the lens. However, a more practical formula is $2r_{99\%} = 2f\lambda/D$,

where $r_{99\%}$ is the focus diameter containing 99% of the laser beam and D is the diameter of the beam entering the lens that contains 99% of the beam.

9) Often real laser beams do not have pure Gaussian profiles and this can be accounted for in the equations by including the M^2 value for the real beam, where $M^2 \geqslant 1$.
10) Coherence of the laser beam arises when the EM waves of the beam are in phase with each other. It is a distinctive characteristic of laser beams, which results in effects such as speckle, diffraction, and interference; it also enables applications such as interferometry and holography.
11) Interferometry is typically where a laser beam is split into two beams using a beamsplitter and the two beams are then combined together again to create a fringe pattern. If something happens to change the phase of one beam relative to the other beam, then there will be a change in the fringe pattern.
12) The coherence of the laser beam leads to diffraction effects. If the laser beam scrapes an edge or an aperture, scattered light from the edge interferes with the laser beam and causes fringe patterns. For example, this means a Gaussian beam passing through a circular aperture that is slightly smaller in diameter than the beam will create Fresnel rings and the beam will no longer have a smooth Gaussian profile.
13) Generally, the resolution of an optical system depends directly on the wavelength of the light. Therefore, to achieve the highest resolution or the smallest spot size, it is best to use short-wavelength light sources.
14) Two basic types of light scattering are elastic scattering, also called Rayleigh scattering, where the wavelength of the scattered photon does not change, and inelastic scattering, where the wavelength of the scattered photon either decreases or increases. An example of inelastic scattering is Raman scattering, where the availability of intense laser beams has made it possible to generate appreciable power in Stokes beams (longer scattered λ) or anti-Stokes beams (shorter scattered λ).

References

[1] Prueta J., McNabb D P, Hagmann C A, Hartemann F V and Barty C P J 2006 Detecting clandestine material with nuclear resonance fluorescence *J. Appl. Phys.* **99** 123102

[2] Jackson J D 1999 *Classical Electrodynamics* 3rd edn (New York: Wiley)

[3] Benware B R, Macchietto C D, Moreno C H and Rocca J J 1998 Demonstration of a high average power tabletop soft x-ray laser *Phys. Rev. Lett.* **81** 5804–7

[4] Elton R C 1990 *X-ray Lasers* (Boston, MA: Academic)

[5] Leuschner R and Pawlowski G 2013 *Photolithography Materials Science and Technology* (Wiley Online Library)

[6] Injeyan H and Goodno G D 2011 *High-Power Laser Handbook* (New York: McGraw Hill)

[7] Rhodes C K 1979 *Excimer Lasers* (Berlin: Springer)

[8] Cherrington B E 1979 *Gaseous Electronics and Gas Lasers* (Oxford: Pergamon)

[9] Numai T 2015 *Fundamentals of Semiconductor Lasers* (Berlin: Springer)

[10] Baranov A and Tournie E 2013 *Semiconductor Lasers: Fundamentals and Applications* (Cambridge:Woodhead Publishing)
[11] Koechner W 2006 *Solid-State Laser Engineering* 6th edn (Berlin: Springer)
[12] Razali W and Bidin N 2012 *Ti:Sapphire Laser Oscillator: Basic Principle and Construction* (Düsseldorf: LAP Lambert)
[13] Tang A M and Cheng L K 1995 *Fundamentals of Optical Parametric Processes and Oscillations* (Amsterdam: Harwood)
[14] Zernike F and Midwinter J E 1973 *Applied Nonlinear Optics* (New York: Wiley)
[15] Basov N G, Bashkin A S, Igoshin V I, Oraevsky A N and Shcheglov V A 1990 *Chemical Lasers* (Berlin: Springer)
[16] Cheo P K 1987 *Handbook of Molecular Lasers* (New York: Marcel Dekker)
[17] Long D A 2002 *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (Wiley Online Library)
[18] Song H and Nagatsuma T 2015 *Handbook of Terahertz Technologies Devices and Applications* (Singapore: Pan Stanford)
[19] Marshall T C 1985 *Free-Electron Lasers* (New York: Macmillan)
[20] Ishikawa T *et al* 2012 A compact x-ray free-electron laser emitting in the sub-ångström region *Nat. Photon.* **6** 540–4
[21] Skolnik M I 2008 *Radar Handbook* 3rd edn (McGraw Hill)
[22] Wangler T P 2008 *RF Linear Accelerators* 2nd edn (Wiley Online Library)
[23] Gordon J P, Zeiger H J and Townes C H 1955 The Maser—New type of microwave amplifier, frequency standard, and spectrometer *Phys. Rev.* **99** 1264–74
[24] Barclay L W 2003 *Propagation of Radiowaves* 2nd edn (Stevenage: Institution of Engineering and Technology)
[25] Barr R, Llanwyn Jones D and Rodger C J 2000 ELF and VLF radio waves *J. Atmos. Solar-Terrestrial Phys.* **62** 1689–718
[26] Iskander M F 2013 *Electromagnetic Fields and Waves* (Long Grove, IL: Waveland Press)
[27] Someda C G 2006 *Electromagnetic Waves* 2nd edn (Boca Raton, FL: CRC Press)
[28] Born M and Wolf E 1999 *Principles of Optics* 7th edn (Cambridge: Cambridge University Press)
[29] Siegman A E 1986 *Lasers* (Mill Valley, CA: University Science Books)
[30] Palik E D 1985 *Handbook of Optical Constants of Solids* (Boston, MA: Academic)
[31] Polyanskiy M Index of refraction calculator: https://refractiveindex.info/.
[32] Rulliere C 2004 *Femtosecond Laser Pulses: Principles and Experiments* 2nd edn (Berlin: Springer)
[33] Steinmeyer G, Sutter D H, Gallmann L, Matushek N and Keller U 1999 Frontiers in ultrashort pulse generation: pushing the limits in linear and nonlinear optics *Science* **286** 1507-12
[34] Krausz F and Ivanov M 2009 Attosecond physics *Rev. Mod. Phys.* **81** 163–234
[35] Goodman J W 1968 *Introduction to Fourier Optics* (New York: McGraw-Hill)
[36] Friedman E and Miller J L 2004 *Photonics Rules of Thumb, Optics, Electro-Optics, Fiber Optics, and Lasers* 2nd Ed. (New York: McGraw-Hill)
[37] Greivenkamp J E 2004 *Field Guide to Geometrical Optics* (Bellingham, WA: SPIE)
[38] Siegman A E Defining and measuring laser beam qualilty *Solid-State Lasers—New Developments and Applications* ed M Inguscio and R Wallenstein 1993 (New York: Plenum)
[39] Dainty J C 1975 *Laser Speckle and Related Phenomena* (Berlin: Springer)

[40] Hariharan P 2002 *Basics of Holography* (Cambridge: Cambridge University Press)
[41] Stenzel O 2014 *Optical Coating* (Berlin: Springer)
[42] Baumeister P 2004 *Optical Coating Technology* (Bellingham, WA: SPIE)
[43] Kimura W D, Kushner M J, Crawford E A and Byron S R 1986 Laser interferometric measurements of a laser preionization triggered spark column *IEEE Trans. Plasma Sci.* ***PS* 14** 246–55
[44] Pretzler G 1991 A new method for numerical abel-inversion *Z. Naturforsch* **46a** 639–41
[45] Kalal M and Nugent K 1988 Abel inversion using fast fourier transforms *Appl. Opt.* **27** 1956–9
[46] Greivenkamp J E 2004 *Field Guide to Geometrical Optics* (Bellingham, WA: SPIE)